Thierry Roncalli

Handbook of Financial Risk Management – Companion Book

Introduction

This companion book contains the solutions of the tutorial exercises which are found at the end of each chapter. Additional materials (datasets, codes, figures and slides) concerning the *Handbook of Risk Management* are available at the following internet web page:

`http://www.thierry-roncalli.com/RiskManagementBook.html`

Contents

[1]The following exercise is taken from Chapters 1 and 2 of Jolliffe (2002).

Part I

Risk Management in the Financial Sector

Chapter 2

Market Risk

2.4.1 Calculating regulatory capital with the Basel I standardized measurement method

1. (a) In the maturity approach, long and short positions are slotted into a maturity-based ladder comprising fifteen time-bands. The time bands are defined by disjoint intervals $]M^{-}, M^{+}]$. The risk weights depend on the time band t and the value of the coupon[1]:

$K(t)$	0.00%	0.20%	0.40%	0.70%	1.25%	1.75%	2.25%
M^{+}_{BC}	1M	3M	6M	1Y	2Y	3Y	4Y
M^{+}_{SC}	1M	3M	6M	1Y	1.9Y	2.8Y	3.6Y
$K(t)$	2.75%	3.25%	3.75%	4.50%	5.25%	6.00%	8.00%
M^{+}_{BC}	5Y	7Y	10Y	15Y	20Y	$+\infty$	
M^{+}_{SC}	4.3Y	5.7Y	7.3Y	9.3Y	10.6Y	12Y	20Y

These risk weights apply to the net exposure on each time band. For reflecting basis and gap risks, the bank must also include a 10% capital charge to the smallest exposure of the matched positions. This adjustment is called the '*vertical disallowance*'. The Basel Committee considers a second adjustment for horizontal offsetting (the '*horizontal disallowance*'). For that, it defines 3 zones (less than 1 year, one year to four years and more than four years). The offsetting can be done within and between the zones. The adjustment coefficients are 30% within the zones 2 and 3, 40% within the zone 1, between the zones 1 and 2, and between the zones 2 and 3, and 100% between the zones 1 and 3. To compute mathematically the required capital, we note $\mathcal{L}^{\star}(t)$ and $\mathcal{S}^{\star}(t)$ the long and short nominal positions for the time band t. $t = 1$ corresponds to the first time band $[0, 1\mathrm{M}]$, $t = 2$ corresponds to the second time band $]1\mathrm{M}, 3\mathrm{M}[$, etc. The risk weighted positions for the time band t are defined as $\mathcal{L}(t) = K(t) \times \mathcal{L}^{\star}(t)$ and $\mathcal{S}(t) = K(t) \times \mathcal{S}^{\star}(t)$. The required capital for the overall net open position is then equal to:

$$\mathcal{K}^{\mathrm{OP}} = \left| \sum_{t=1}^{15} \mathcal{L}(t) - \sum_{t=1}^{15} \mathcal{S}(t) \right|$$

The matched position $\mathcal{M}(t)$ for the time band t is equal to $\min(\mathcal{L}(t), \mathcal{S}(t))$. We deduce that the additional capital for the vertical disallowance is:

$$\mathcal{K}^{\mathrm{VD}} = 10\% \times \sum_{t=1}^{13} \mathcal{M}(t)$$

[1] Coupons 3% or more are called big coupons (or BC) and coupons less than 3% are called small coupons (SC). When the maturity is greater than 20Y, $K(t)$ is equal to 6.00% for big coupons and 12.50% for small coupons.

$\mathcal{N}(t) = \mathcal{L}(t) - \mathcal{S}(t)$ is the net exposure for the time band t. We then define the net long and net short exposures for the three zones as follows:

$$\begin{aligned}\mathcal{L}_i &= \sum_{t\in\Delta_i} \max(\mathcal{N}(t), 0) \\ \mathcal{S}_i &= -\sum_{t\in\Delta_i} \min(\mathcal{N}(t), 0)\end{aligned}$$

where $\Delta_1 = [0, 1\mathrm{Y}]$, $\Delta_2 =]1\mathrm{Y}, 4\mathrm{Y}]$ and $\Delta_3 =]4\mathrm{Y}, +\infty]$. We define $\mathcal{CF}_{i,j}$ as the exposure of the zone i that can be carried forward to the zone j. We then compute the additional capital for the horizontal disallowance:

$$\begin{aligned}\mathcal{K}^{\mathrm{HD}} &= 0.4 \times \min(\mathcal{L}_1, \mathcal{S}_1) + 0.3 \times \min(\mathcal{L}_2, \mathcal{S}_2) + 0.3 \times \min(\mathcal{L}_3, \mathcal{S}_3) + \\ &\quad 0.4 \times \mathcal{CF}_{1,2} + 0.4 \times \mathcal{CF}_{2,3} + \mathcal{CF}_{1,3}\end{aligned}$$

The regulatory capital for the general market risk is the sum of the three components:

$$\mathcal{K} = \mathcal{K}^{\mathrm{OP}} + \mathcal{K}^{\mathrm{VD}} + \mathcal{K}^{\mathrm{HD}}$$

(b) For each time band, we report the long, short, matched and net exposures:

Time band	$\mathcal{L}^\star(t)$	$\mathcal{S}^\star(t)$	$K(t)$	$\mathcal{L}(t)$	$\mathcal{S}(t)$	$\mathcal{M}(t)$	$\mathcal{N}(t)$
3M-6M	100	50	0.40%	0.40	0.20	0.20	0.20
7Y-10Y	10	50	3.75%	0.45	2.25	0.45	−1.80

The capital charge for the overall open position is:

$$\begin{aligned}\mathcal{K}^{\mathrm{OP}} &= \lvert 0.40 + 0.45 - 0.20 - 2.25 \rvert \\ &= 1.6\end{aligned}$$

whereas the capital for the vertical disallowance is:

$$\begin{aligned}\mathcal{K}^{\mathrm{VD}} &= 10\% \times (0.20 + 0.45) \\ &= 0.065\end{aligned}$$

We now compute the net long and net short exposures for the three zones:

zone	1	2	3
$\mathcal{L}_i$	0.20	0.00	0.00
$\mathcal{S}_i$	0.00	0.00	1.80

It follows that there is no horizontal offsetting within the zones. Moreover, we notice that we can only carry forward the long exposure $\mathcal{L}_1$ to the zone 3 meaning that:

$$\begin{aligned}\mathcal{K}^{\mathrm{HD}} &= 40\% \times 0.00 + 30\% \times 0.00 + 30\% \times 0.00 + \\ &\quad 40\% \times 0.00 + 40\% \times 0.00 + 100\% \times 0.20 \\ &= 0.20\end{aligned}$$

We finally deduce that the required capital is:

$$\begin{aligned}\mathcal{K} &= 1.6 + 0.065 + 0.20 \\ &= \$1.865 \text{ mn}\end{aligned}$$

2. (a) We have:

Stock	3M	Exxon	IBM	Pfizer	AT&T	Cisco	Oracle
$\mathcal{L}_i$	100	100	10	50	60	90	
$\mathcal{S}_i$		50					80
$\mathcal{N}_i$	100	50	10	50	60	90	-80

We deduce that the capital charge for the specific risk is equal to \$35.20 mn:

$$\mathcal{K}^{\text{Specific}} = 8\% \times \sum_{i=1}^{7} |\mathcal{N}_i| = 8\% \times 440 = 35.20$$

(b) The total net exposure $\sum_{i=1}^{7} \mathcal{N}_i$ is equal to \$280 mn, meaning that the capital charge for the general market risk is equal to \$22.40 mn:

$$\mathcal{K}^{\text{General}} = 8\% \times 280 = 22.40$$

(c) To hedge the market risk of the portfolio, the investor can sell \$280 mn of S&P 500 futures contracts[2]. In this case, the capital charge for the general market risk is equal to zero. However, this new exposure implies an additional capital charge for the specific risk:

$$\mathcal{K}^{\text{Specific}} = 35.20 + 4\% \times 280 = 35.20 + 11.20 = 46.40$$

Let $\mathcal{S}$ be the short exposure on S&P 500 futures contracts. We have:

$$\begin{aligned} \mathcal{K} &= \mathcal{K}^{\text{Specific}} + \mathcal{K}^{\text{General}} \\ &= (35.20 + 4\% \times \mathcal{S}) + 8\% \times |280 - \mathcal{S}| \end{aligned}$$

We notice that there is a trade-off between the capital charge for the specific risk which is an increasing function of $\mathcal{S}$ and the capital charge for the general market risk which is a decreasing function of $\mathcal{S}$ for $\mathcal{S} \leq 280$. Another expression of $K^{(\text{total})}$ is:

$$\mathcal{K} = \begin{cases} 57.60 - 4\% \times \mathcal{S} & \text{if } \mathcal{S} \leq 280 \\ 12.80 + 12\% \times \mathcal{S} & \text{otherwise} \end{cases}$$

We verify that the minimum is reached when $\mathcal{S}$ is exactly equal to 280 (see Figure 2.1).

3. (a) Under SMM, we have:

$$\mathcal{K}^{\text{SMM}} = 8\% \times \mathcal{N}_w$$

(b) The 10-day Gaussian value-at-risk is equal to:

$$\begin{aligned} \text{VaR}_{99\%}(w; \text{ten days}) &= 2.33 \times \frac{\mathcal{N}_w \times \sigma(w)}{\sqrt{260}} \times \sqrt{10} \\ &= 0.457 \times \mathcal{N}_w \times \sigma(w) \end{aligned}$$

We deduce that the required capital is approximately equal to:

$$\begin{aligned} \mathcal{K}^{\text{IMA}} &\approx (3 + \xi) \times \text{VaR}_{99\%}(w; \text{ten days}) \\ &= (3 + \xi) \times 0.457 \times \mathcal{N}_w \times \sigma(w) \end{aligned}$$

Because $\xi \leq 1$, it follows that:

$$\mathcal{K}^{\text{IMM}} \leq 1.828 \times \mathcal{N}_w \times \sigma(w)$$

[2]We assume that the beta of the portfolio with respect to the S&P 500 index is equal to one.

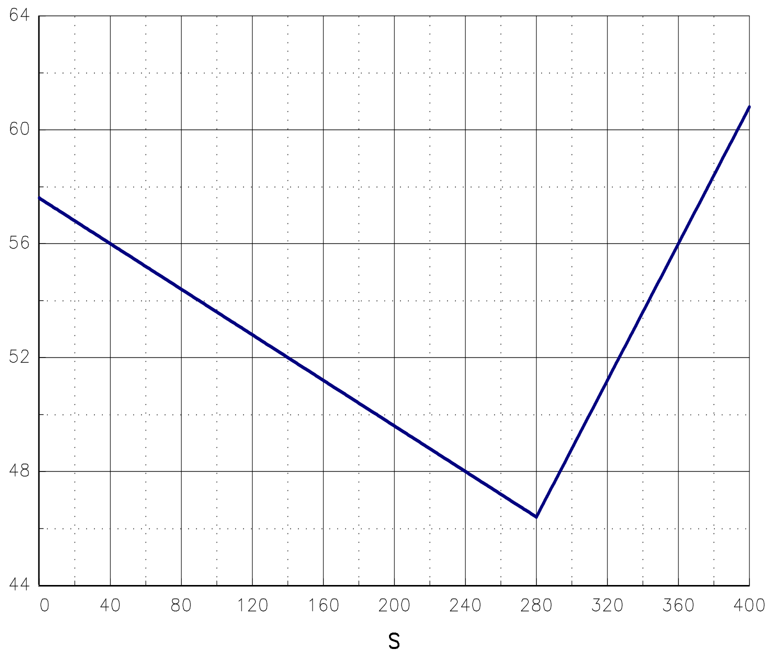

FIGURE 2.1: Capital charge $\mathcal{K}$ with respect to $\mathcal{S}$

(c) A sufficient condition for $\mathcal{K}^{\text{IMM}} \leq \mathcal{K}^{\text{SMM}}$ is:

$$\begin{aligned} & 1.828 \times \mathcal{N}_w \times \sigma(w) \leq 8\% \times \mathcal{N}_w \\ \Leftrightarrow \quad & \sigma(w) \leq \frac{8\%}{1.828} \\ \Leftrightarrow \quad & \sigma(w) \leq 4.37\% \end{aligned}$$

(d) The annualized volatility of the portfolio must be lower than 4.37%. This implies that long equity exposures induce more required capital under IMM than under SMM. Indeed, the volatility of directional equity portfolios is generally higher than 12%. In order to obtain equity portfolios with such lower volatility, the portfolio must be long/short, meaning that the directional risk must be (partially) hedged.

4. (a) The bank is exposed to foreign exchange and commodity risks with spot and forward positions. Contrary to stocks or many equity products, these exposures include a maturity pattern. For instance, the \$100 mn EUR long position has not the same maturity than the \$100 mn EUR short position, implying that the bank cannot match the two positions.

   (b) We first consider the FX risk. We have $\mathcal{N}_{\text{EUR}} = 100 - 100 = 0$, $\mathcal{N}_{\text{JPY}} = 50 - 100 = -50$, $\mathcal{N}_{\text{CAD}} = 0 - 50 = -50$ and $\mathcal{N}_{\text{Gold}} = 50 - 0 = 50$. We deduce that the aggregated long and short positions are $\mathcal{L}_{\text{FX}} = 0$ and $\mathcal{S}_{\text{FX}} = 100$. It follows that the required capital is:

$$\begin{aligned} \mathcal{K}^{\text{FX}} &= 8\% \times \left(\max\left(\mathcal{L}_{\text{FX}}, \mathcal{S}_{\text{FX}}\right) + \left|\mathcal{N}_{\text{Gold}}\right|\right) \\ &= 8\% \times (100 + 50) \\ &= \$12 \text{ mn} \end{aligned}$$

For the commodity risk, we exclude the Gold position, because it is treated as a foreign exchange risk. We have:

$$\begin{aligned}\mathcal{K}^{\text{Commodity}} &= 15\% \times \sum_{i=5}^{7} |\mathcal{L}_i - \mathcal{S}_i| + 3\% \times \sum_{i=5}^{7} (\mathcal{L}_i + \mathcal{S}_i) \\ &= 15\% \times (50 + 20 + 20) + 3\% \times 390 \\ &= 13.50 + 11.70 \\ &= \$25.20 \text{ mn}\end{aligned}$$

We finally obtain:

$$\begin{aligned}\mathcal{K} &= \mathcal{K}^{\text{FX}} + \mathcal{K}^{\text{Commodity}} \\ &= \$33.20 \text{ mn}\end{aligned}$$

5. (a) Under the maturity ladder approach, the bank should spread long and short exposures of each currency to seven time bands: 0-1M, 1M-3M, 3M-6M, 6M-1Y, 1Y-2Y, 2Y-3Y, 3Y+. For each time band, the capital charge for the basis risk is equal to 1.5% of the matched positions (long and short). Nevertheless, the residual net position of previous time bands may be carried forward to offset exposures in next time bands. In this case, a surcharge of 0.6% of the residual net position is added at each time band to cover the time spread risk. Finally, a capital charge of 15% is applied to the global net exposure (or the residual unmatched position) for directional risk. To compute mathematically the required capital, we note $\mathcal{L}_i(t)$ and $\mathcal{S}_i(t)$ the long and short positions of the commodity i for the time band t. $t = 1$ corresponds to the first time band $[0, 1\text{M}]$ and $t = 7$ corresponds to the last time band $]3\text{Y}, +\infty[$. The cumulative long and short exposures are $\mathcal{L}_i^+(t) = \mathcal{L}_i^+(t-1) + \mathcal{L}_i(t)$ with $\mathcal{L}_i^+(0) = 0$ and $\mathcal{S}_i^+(t) = \mathcal{S}_i^+(t-1) + \mathcal{S}_i(t)$ with $\mathcal{S}_i^+(0) = 0$. The cumulative matched position is $\mathcal{M}_i^+(t) = \min\left(\mathcal{L}_i^+(t), \mathcal{S}_i^+(t)\right)$. We deduce that the matched exposition for the time band t is equal to $\mathcal{M}_i(t) = \mathcal{M}_i^+(t) - \mathcal{M}_i^+(t-1)$ with $\mathcal{M}_i^+(0) = 0$. The value of the carried forward $\mathcal{CF}_i(t)$ can be obtained recursively by reporting the unmatched positions at time t which can be offset in the times bands τ with $\tau > t$. The residual unmatched position is $\mathcal{N}_i = \max\left(\mathcal{L}_i^+(7), \mathcal{S}_i^+(7)\right) - \mathcal{M}_i^+(t)$. We finally deduce that the required capital is the sum of the individual capital charges:

$$\mathcal{K}_i = 1.5\% \times \left(\sum_{t=1}^{7} 2 \times \mathcal{M}_i(t)\right) + 0.6\% \times \left(\sum_{t=1}^{6} \mathcal{CF}_i(t)\right) + 15\% \times \mathcal{N}_i$$

We notice that the matched position $\mathcal{M}_i(t)$ is multiplied by 2, because we apply the capital charge 1.5% to the long and short matched positions.

(b) We compute the cumulative positions $\mathcal{L}_i^+(t)$ and $\mathcal{S}_i^+(t)$ and deduce the matched expositions $\mathcal{M}_i(t)$:

Time band	t	$\mathcal{L}_i(t)$	$\mathcal{S}_i(t)$	$\mathcal{L}_i^+(t)$	$\mathcal{S}_i^+(t)$	$\mathcal{M}_i(t)$	$\mathcal{CF}_i(t)$
0−1M	1	500	300	500	300	300	200
1M−3M	2	0	900	500	1 200	200	700
3M−6M	3	0	0	500	1 200	0	700
6M−1Y	4	1 800	100	2 300	1 300	800	600
1Y−2Y	5	300	600	2 600	1 900	600	300
2Y−3Y	6	0	100	2 600	2 000	100	200
3Y+	7	0	200	2 600	2 200	200	0

The sum of matched positions is equal to 2 200. This means that the residual unmatched position is 400 (2 600 − 2 200). At time band $t = 1$, we can carry forward 200 of long position in the next time band. At time band $t = 2$, we can carry forward 700 of short position in the times band $t = 4$. This implies that $\mathcal{CF}_i(3) = 700$ and $\mathcal{CF}_i(4) = 700$. At time band $t = 4$, the residual unmatched position is equal to 1 000 (1 800 − 100 − 700). However, we can only carry 600 of this long position in the next time bands (300 for $t = 5$, 100 for $t = 6$ and 200 for $t = 1$). At the end, we verify that the residual position is 400, that is the part of the long position at time band $t = 4$ which can not be carried forward (1 000 − 600). We also deduce that the sum of carried forward positions is 2 700. It follows that the required capital is[3]:

$$\begin{aligned}\mathcal{K}_i &= 1.5\% \times 4\,400 + 0.6\% \times 2\,700 + 15\% \times 400 \\ &= \$142.20\end{aligned}$$

2.4.2 Covariance matrix

1. (a) We have:

$$\sigma_A = \sqrt{\Sigma_{1,1}} = \sqrt{4\%} = 20\%$$

For the other stocks, we obtain $\sigma_B = 22.36\%$ and $\sigma_C = 24.49\%$.

(b) The correlation is the covariance divided by the product of volatilities:

$$\rho(R_A, R_B) = \frac{\Sigma_{1,2}}{\sqrt{\Sigma_{1,1} \times \Sigma_{2,2}}} = \frac{3\%}{20\% \times 22.36\%} = 67.08\%$$

We obtain:

$$\rho = \begin{pmatrix} 100.00\% & & \\ 67.08\% & 100.00\% & \\ 40.82\% & -18.26\% & 100.00\% \end{pmatrix}$$

2. (a) Using the formula $\Sigma_{i,j} = \rho_{i,j}\sigma_i\sigma_j$, it follows that:

$$\Sigma = \begin{pmatrix} 1.00\% & & \\ 1.00\% & 4.00\% & \\ 0.75\% & 0.00\% & 9.00\% \end{pmatrix}$$

(b) We deduce that:

$$\begin{aligned}\sigma^2(w) &= 0.5^2 \times 1\% + 0.5^2 \times 4\% + 2 \times 0.5 \times 0.5 \times 1\% \\ &= 1.75\%\end{aligned}$$

and $\sigma(w) = 13.23\%$.

(c) It follows that:

$$\begin{aligned}\sigma^2(w) &= 0.6^2 \times 1\% + (-0.4)^2 \times 4\% + 2 \times 0.6 \times (-0.4) \times 1\% \\ &= 0.52\%\end{aligned}$$

and $\sigma(w) = 7.21\%$. This long/short portfolio has a lower volatility than the previous long-only portfolio, because part of the risk is hedged by the positive correlation between stocks A and B.

[3]The total matched position is equal to $2 \times 2\,200 = 4\,400$ (long + short).

(d) We have:

$$\begin{array}{rcl}\sigma^2(w) & = & 150^2 \times 1\% + 500^2 \times 4\% + (-200)^2 \times 9\% + \\ & & 2 \times 150 \times 500 \times 1\% + \\ & & 2 \times 150 \times (-200) \times 0.75\% + \\ & & 2 \times 500 \times (-200) \times 0\% \\ & = & 14\,875\end{array}$$

The volatility is equal to \$121.96 and is measured in USD contrary to the two previous results which were expressed in %.

3. (a) We have:

$$\mathbb{E}[R] = \beta\mathbb{E}[\mathcal{F}] + \mathbb{E}[\varepsilon]$$

and:

$$R - \mathbb{E}[R] = \beta(\mathcal{F} - \mathbb{E}[\mathcal{F}]) + \varepsilon - \mathbb{E}[\varepsilon]$$

It follows that:

$$\begin{array}{rcl}\operatorname{cov}(R) & = & \mathbb{E}\left[(R - \mathbb{E}[R])(R - \mathbb{E}[R])^\top\right] \\ & = & \mathbb{E}\left[\beta(\mathcal{F} - \mathbb{E}[\mathcal{F}])(\mathcal{F} - \mathbb{E}[\mathcal{F}])\beta^\top\right] + \\ & & 2 \times \mathbb{E}\left[\beta(\mathcal{F} - \mathbb{E}[\mathcal{F}])(\varepsilon - \mathbb{E}[\varepsilon])^\top\right] + \\ & & \mathbb{E}\left[(\varepsilon - \mathbb{E}[\varepsilon])(\varepsilon - \mathbb{E}[\varepsilon])^\top\right] \\ & = & \sigma^2_{\mathcal{F}}\beta\beta^\top + D\end{array}$$

We deduce that:

$$\sigma(R_i) = \sqrt{\sigma^2_{\mathcal{F}}\beta_i^2 + \tilde{\sigma}_i^2}$$

We obtain $\sigma(R_A) = 18.68\%$, $\sigma(R_B) = 26.48\%$ and $\sigma(R_C) = 15.13\%$.

(b) The correlation between stocks i and j is defined as follows:

$$\rho(R_i, R_j) = \frac{\sigma^2_{\mathcal{F}}\beta_i\beta_j}{\sigma(R_i)\,\sigma(R_j)}$$

We obtain:

$$\rho = \left(\begin{array}{rrr}100.00\% & & \\ 94.62\% & 100.00\% & \\ 12.73\% & 12.98\% & 100.00\%\end{array}\right)$$

4. (a) We have:

$$\begin{array}{rcl}\mu(Z_i) & = & \mathbb{E}[X_i Y_i] \\ & = & \mathbb{E}[X_i]\,\mathbb{E}[Y_i] \\ & = & \mu_i(X)\,\mu_i(Y)\end{array}$$

because X_i and Y_i are independent. For the covariance, we obtain:

$$\begin{aligned}\operatorname{cov}(Z_i, Z_j) &= \mathbb{E}\left[(X_i Y_i)(X_j Y_j)\right] - \mathbb{E}\left[X_i Y_i\right]\mathbb{E}\left[X_j Y_j\right] \\ &= \mathbb{E}\left[X_i X_j\right]\mathbb{E}\left[Y_i Y_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_j\right] \\ &= \left(\operatorname{cov}(X_i, X_j) + \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\right)\times \\ &\quad \left(\operatorname{cov}(Y_i, Y_j) + \mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_j\right]\right) - \\ &\quad \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_j\right] \\ &= \operatorname{cov}(X_i, X_j)\operatorname{cov}(Y_i, Y_j) + \\ &\quad \operatorname{cov}(X_i, X_j)\,\mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_j\right] + \\ &\quad \operatorname{cov}(Y_i, Y_j)\,\mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] \\ &= \Sigma_{i,j}(X)\,\Sigma_{i,j}(Y) + \Sigma_{i,j}(X)\,\mu_i(Y)\,\mu_j(Y) + \\ &\quad \Sigma_{i,j}(Y)\,\mu_i(X)\,\mu_j(X)\end{aligned}$$

To obtain this formula, we use the fact that $X_i X_j$ and $Y_i Y_j$ are independent. In a matrix form, we find that:

$$\begin{aligned}\mu(Z) &= \mu(X)\circ\mu(Y) \\ \Sigma(Z) &= \Sigma(X)\circ\Sigma(Y) + \\ &\quad \Sigma(X)\circ\mu(Y)\circ\mu(Y)^\top + \\ &\quad \Sigma(Y)\circ\mu(X)\circ\mu(X)^\top\end{aligned}$$

(b) Using the numerical values, we obtain[4] $\mu(Z) = \mathbf{0}$ and:

$$\Sigma(Z) = \begin{pmatrix} 0.333\% & & \\ 0.250\% & 1.333\% & \\ 0.188\% & 0.000\% & 3.000\% \end{pmatrix}$$

The expression of the P&L is:

$$\Pi(w) = 150 Z_1 + 500 Z_2 - 200 Z_3$$

We find that $\mu(\Pi) = 0$ and $\sigma(\Pi) = 69.79$. We deduce that the Gaussian VaR with a 99% confidence level is equal to \$162.36. For the Monte Carlo method, we use the following steps: *(i)* we first simulate the random variate X with the Cholesky algorithm; *(ii)* we then simulate Y with a uniform random generator; *(iii)* we calculate the components $Z_i = X_i Y_i$; *(iv)* we finally deduce the P&L. With one million of simulations, we find that the Monte Carlo VaR is equal to \$182.34. We explain this result because the distribution of $\Pi(w)$ is far to be normal as illustrated in Figure 2.2.

2.4.3 Risk measure

1. (a) We have:

$$\mathrm{VaR}_\alpha(L) = \inf\left\{\ell : \Pr\left\{L \geq \ell\right\} \geq \alpha\right\}$$

and:

$$\mathrm{ES}_\alpha(L) = \mathbb{E}\left[L \mid L \geq \mathrm{VaR}_\alpha(L)\right]$$

[4] We remind that the mean and the variance of the distribution $\mathcal{U}(0,1)$ is $1/2$ and $1/12$.

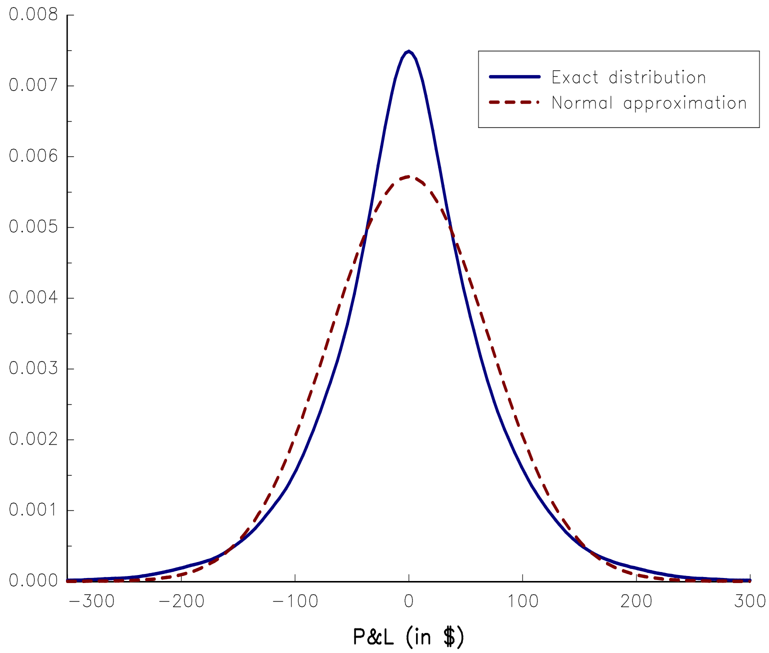

FIGURE 2.2: Probability distribution of the P&L

(b) We assume that $\mathbf{F}$ is continuous. It follows that $\text{VaR}_\alpha(L) = \mathbf{F}^{-1}(\alpha)$. We deduce that:

$$\begin{aligned} \text{ES}_\alpha(L) &= \mathbb{E}\left[L \mid L \geq \mathbf{F}^{-1}(\alpha)\right] \\ &= \int_{\mathbf{F}^{-1}(\alpha)}^{\infty} x \frac{f(x)}{1 - \mathbf{F}\left(\mathbf{F}^{-1}(\alpha)\right)} \,\mathrm{d}x \\ &= \frac{1}{1-\alpha} \int_{\mathbf{F}^{-1}(\alpha)}^{\infty} x f(x) \,\mathrm{d}x \end{aligned}$$

We consider the change of variable $t = \mathbf{F}(x)$. Because $\mathrm{d}t = f(x)\,\mathrm{d}x$ and $\mathbf{F}(\infty) = 1$, we obtain:

$$\text{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 \mathbf{F}^{-1}(t) \,\mathrm{d}t$$

(c) We have:

$$f(x) = \theta \frac{x^{-(\theta+1)}}{x_-^{-\theta}}$$

The non-centered moment of order n is[5]:

$$\begin{aligned}\mathbb{E}\left[L^n\right] &= \int_{x_-}^{\infty} x^n \theta \frac{x^{-(\theta+1)}}{x_-^{-\theta}} \,\mathrm{d}x \\ &= \frac{\theta}{x_-^{-\theta}} \int_{x_-}^{\infty} x^{n-\theta-1} \,\mathrm{d}x \\ &= \frac{\theta}{x_-^{-\theta}} \left[\frac{x^{n-\theta}}{n-\theta}\right]_{x_-}^{\infty} \\ &= \frac{\theta}{\theta-n} x_-^n\end{aligned}$$

We deduce that:

$$\mathbb{E}\left[L\right] = \frac{\theta}{\theta-1} x_-$$

and:

$$\mathbb{E}\left[L^2\right] = \frac{\theta}{\theta-2} x_-^2$$

The variance of the loss is then:

$$\operatorname{var}(L) = \mathbb{E}\left[L^2\right] - \mathbb{E}^2\left[L\right] = \frac{\theta}{(\theta-1)^2(\theta-2)} x_-^2$$

x_- is a scale parameter whereas θ is a parameter to control the distribution tail. We have:

$$1 - \left(\frac{\mathbf{F}^{-1}(\alpha)}{x_-}\right)^{-\theta} = \alpha$$

We deduce that:

$$\mathrm{VaR}_\alpha(L) = \mathbf{F}^{-1}(\alpha) = x_-(1-\alpha)^{-\theta^{-1}}$$

We also obtain:

$$\begin{aligned}\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha} \int_{\alpha}^{1} x_-(1-t)^{-\theta^{-1}} \,\mathrm{d}t \\ &= \frac{x_-}{1-\alpha} \left[-\frac{1}{1-\theta^{-1}}(1-t)^{1-\theta^{-1}}\right]_{\alpha}^{1} \\ &= \frac{\theta}{\theta-1} x_-(1-\alpha)^{-\theta^{-1}} \\ &= \frac{\theta}{\theta-1} \mathrm{VaR}_\alpha\end{aligned}$$

Because $\theta > 1$, we have $\frac{\theta}{\theta-1} > 1$ and:

$$\mathrm{ES}_\alpha(L) > \mathrm{VaR}_\alpha(L)$$

(d) We have:

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_{\mu+\sigma\Phi^{-1}(\alpha)}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$$

[5] The moment exists if $n \neq \theta$.

By considering the change of variable $t = \sigma^{-1}(x - \mu)$, we obtain:

$$\begin{aligned}
\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} (\mu + \sigma t)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)\,\mathrm{d}t \\
&= \frac{\mu}{1-\alpha}\left[\Phi(t)\right]_{\Phi^{-1}(\alpha)}^{\infty} + \\
&\quad \frac{\sigma}{(1-\alpha)\sqrt{2\pi}}\int_{\Phi^{-1}(\alpha)}^{\infty} t\exp\left(-\frac{1}{2}t^2\right)\,\mathrm{d}t \\
&= \mu + \frac{\sigma}{(1-\alpha)\sqrt{2\pi}}\left[-\exp\left(-\frac{1}{2}t^2\right)\right]_{\Phi^{-1}(\alpha)}^{\infty} \\
&= \mu + \frac{\sigma}{(1-\alpha)\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\Phi^{-1}(\alpha)\right]^2\right) \\
&= \mu + \frac{\sigma}{(1-\alpha)}\phi\left(\Phi^{-1}(\alpha)\right)
\end{aligned}$$

Because $\phi'(x) = -x\phi(x)$, we have:

$$\begin{aligned}
1 - \Phi(x) &= \int_x^{\infty} \phi(t)\,\mathrm{d}t \\
&= \int_x^{\infty}\left(-\frac{1}{t}\right)(-t\phi(t))\,\mathrm{d}t \\
&= \int_x^{\infty}\left(-\frac{1}{t}\right)\phi'(t)\,\mathrm{d}t
\end{aligned}$$

We consider the integration by parts with $u(t) = -t^{-1}$ and $v'(t) = \phi'(t)$:

$$\begin{aligned}
1 - \Phi(x) &= \left[-\frac{\phi(t)}{t}\right]_x^{\infty} - \int_x^{\infty}\frac{1}{t^2}\phi(t)\,\mathrm{d}t \\
&= \frac{\phi(x)}{x} + \int_x^{\infty}\frac{1}{t^3}(-t\phi(t))\,\mathrm{d}t \\
&= \frac{\phi(x)}{x} + \int_x^{\infty}\frac{1}{t^3}\phi'(t)\,\mathrm{d}t
\end{aligned}$$

We consider another integration by parts with $u(t) = t^{-3}$ and $v'(t) = \phi'(t)$:

$$\begin{aligned}
1 - \Phi(x) &= \frac{\phi(x)}{x} + \left[\frac{\phi(t)}{t^3}\right]_x^{\infty} - \int_x^{\infty} -\frac{3}{t^4}\phi(t)\,\mathrm{d}t \\
&= \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} - \int_x^{\infty}\frac{3}{t^5}\phi'(t)\,\mathrm{d}t
\end{aligned}$$

We continue to use the integration by parts with $v'(t) = \phi(t)$. At the end, we obtain:

$$\begin{aligned}
1 - \Phi(x) &= \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + 3\frac{\phi(x)}{x^5} - 3\cdot 5\frac{\phi(x)}{x^7} + \\
&\quad 3\cdot 5\cdot 7\frac{\phi(x)}{x^9} - \ldots \\
&= \frac{\phi(x)}{x} + \frac{1}{x^2}\sum_{n=1}^{\infty}(-1)^n\left(\prod_{i=1}^{n}(2i-1)\right)\frac{\phi(x)}{x^{2n-1}} \\
&= \frac{\phi(x)}{x} + \frac{\Psi(x)}{x^2}
\end{aligned}$$

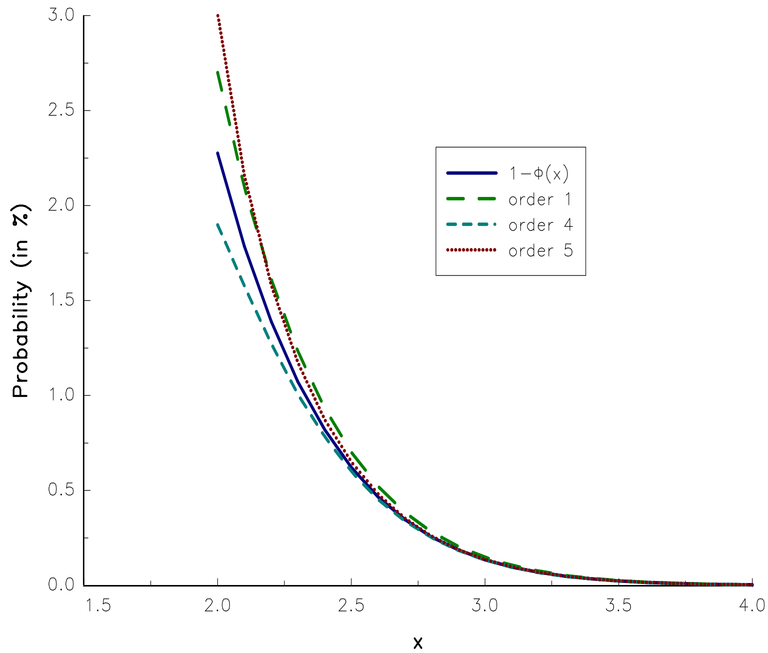

FIGURE 2.3: Approximation of $1 - \Phi(x)$

We have represented the approximation in Figure 2.3. We finally deduce that:

$$\phi(x) = x(1 - \Phi(x)) - \frac{\Psi(x)}{x}$$

By using the previous expression of $\mathrm{ES}_\alpha(L)$, we obtain with $x = \Phi^{-1}(\alpha)$:

$$\begin{aligned}
\mathrm{ES}_\alpha(L) &= \mu + \frac{\sigma}{(1-\alpha)}\phi\left(\Phi^{-1}(\alpha)\right) \\
&= \mu + \frac{\sigma}{(1-\alpha)}\phi(x) \\
&= \mu + \frac{\sigma}{(1-\alpha)}\left(\Phi^{-1}(\alpha)(1-\alpha) - \frac{\Psi\left(\Phi^{-1}(\alpha)\right)}{\Phi^{-1}(\alpha)}\right) \\
&= \mu + \sigma\Phi^{-1}(\alpha) - \sigma\frac{\Psi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)\Phi^{-1}(\alpha)} \\
&= \mathrm{VaR}_\alpha(L) - \sigma\frac{\Psi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)\Phi^{-1}(\alpha)}
\end{aligned}$$

We deduce that $\mathrm{ES}_\alpha(L) \to \mathrm{VaR}_\alpha(L)$ because:

$$\lim_{\alpha\to 1}\frac{\Psi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)\Phi^{-1}(\alpha)} = 0$$

(e) For the Gaussian distribution, the expected shortfall and the value-at-risk coincide for high confidence level α. It is not the case with the Pareto distribution, which has a fat tail. The use of the Pareto distribution can then produce risk measures which may be much higher than those based on the Gaussian distribution.

2. (a) We have:

$$\begin{aligned}\mathcal{R}(L_1+L_2) &= \mathbb{E}[L_1+L_2]=\mathbb{E}[L_1]+\mathbb{E}[L_2]=\mathcal{R}(L_1)+\mathcal{R}(L_2)\\ \mathcal{R}(\lambda L) &= \mathbb{E}[\lambda L]=\lambda\mathbb{E}[L]=\lambda\mathcal{R}(L)\\ \mathcal{R}(L+m) &= \mathbb{E}[L-m]=\mathbb{E}[L]-m=\mathcal{R}(L)-m\end{aligned}$$

We notice that:

$$\mathbb{E}[L]=\int_{-\infty}^{\infty}x\,\mathrm{d}\mathbf{F}(x)=\int_0^1\mathbf{F}^{-1}(t)\,\mathrm{d}t$$

We deduce that if $\mathbf{F}_1(x)\geq\mathbf{F}_2(x)$, then $\mathbf{F}_1^{-1}(t)\leq\mathbf{F}_2^{-1}(t)$ and $\mathbb{E}[L_1]\leq\mathbb{E}[L_2]$. We conclude that $\mathcal{R}$ is a coherent risk measure.

(b) We have:

$$\begin{aligned}\mathcal{R}(L_1+L_2) &= \mathbb{E}[L_1+L_2]+\sigma(L_1+L_2)\\ &= \mathbb{E}[L_1]+\mathbb{E}[L_2]+\\ &\quad\sqrt{\sigma^2(L_1)+\sigma^2(L_2)+2\rho(L_1,L_2)\sigma(L_1)\sigma(L_2)}\end{aligned}$$

Because $\rho(L_1,L_2)\leq 1$, we deduce that:

$$\begin{aligned}\mathcal{R}(L_1+L_2) &\leq \mathbb{E}[L_1]+\mathbb{E}[L_2]+\\ &\quad\sqrt{\sigma^2(L_1)+\sigma^2(L_2)+2\sigma(L_1)\sigma(L_2)}\\ &\leq \mathbb{E}[L_1]+\mathbb{E}[L_2]+\sigma(L_1)+\sigma(L_2)\\ &\leq \mathcal{R}(L_1)+\mathcal{R}(L_2)\end{aligned}$$

We have:

$$\begin{aligned}\mathcal{R}(\lambda L) &= \mathbb{E}[\lambda L]+\sigma(\lambda L)\\ &= \lambda\mathbb{E}[L]+\lambda\sigma(L)\\ &= \lambda\mathcal{R}(L)\end{aligned}$$

and:

$$\begin{aligned}\mathcal{R}(L+m) &= \mathbb{E}[L-m]+\sigma(L-m)\\ &= \mathbb{E}[L]-m+\sigma(L)\\ &= \mathcal{R}(L)-m\end{aligned}$$

If we consider the convexity property, we notice that:

$$\begin{aligned}\mathcal{R}(\lambda L_1+(1-\lambda)L_2) &\leq \mathcal{R}(\lambda L_1)+\mathcal{R}((1-\lambda)L_2)\\ &\leq \lambda\mathcal{R}(L_1)+(1-\lambda)\mathcal{R}(L_2)\end{aligned}$$

We conclude that $\mathcal{R}$ is a convex risk measure.

3. We have:

ℓ_i	0	1	2	3	4	5	6	7	8
$\Pr\{L=\ell_i\}$	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$\Pr\{L\leq\ell_i\}$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

(a) We have $\mathrm{VaR}_{50\%}(L) = 3$, $\mathrm{VaR}_{75\%}(L) = 6$, $\mathrm{VaR}_{90\%}(L) = 7$ and:

$$\begin{aligned}\mathrm{ES}_{50\%}(L) &= \frac{3 \times 10\% + \ldots + 8 \times 10\%}{60\%} = 5.5 \\ \mathrm{ES}_{75\%}(L) &= \frac{6 \times 10\% + \ldots + 8 \times 10\%}{30\%} = 7.0 \\ \mathrm{ES}_{90\%}(L) &= \frac{7 \times 10\% + 8 \times 10\%}{20\%} = 7.5\end{aligned}$$

(b) We have to build a bivariate distribution such that:

$$\mathbf{F}_1^{-1}(\alpha) + \mathbf{F}_2^{-1}(\alpha) < \mathbf{F}_{1+2}^{-1}(\alpha)$$

To this end, we may use the Makarov inequalities. For instance, we may consider an ordinal sum of the copula $\mathbf{C}^+$ for $(u_1, u_2) \leq (\alpha, \alpha)$ and another copula $\mathbf{C}_\alpha$ for $(u_1, u_2) > (\alpha, \alpha)$ to produce a bivariate distribution which does not satisfy the subadditivity property. By taking for example $\alpha = 70\%$ and $\mathbf{C}_\alpha = \mathbf{C}^-$, we obtain the following bivariate distribution[6]:

ℓ_i	0	1	2	3	4	5	6	7	8	$p_{2,i}$
0	0.2									0.2
1		0.1								0.1
2			0.1							0.1
3				0.1						0.1
4					0.1					0.1
5						0.1				0.1
6									**0.1**	0.1
7								**0.1**		0.1
8							**0.1**			0.1
$p_{1,i}$	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

We then have:

ℓ_i	0	2	4	6	8	10	14
$\Pr\{L_1 + L_2 = \ell_i\}$	0.2	0.1	0.1	0.1	0.1	0.1	0.3
$\Pr\{L_1 + L_2 \leq \ell_i\}$	0.2	0.3	0.4	0.5	0.6	0.7	1.0

Because $\mathbf{F}_1^{-1}(80\%) = \mathbf{F}_2^{-1}(80\%) = 6$ and $\mathbf{F}_{1+2}^{-1}(80\%) = 14$, we obtain:

$$\mathbf{F}_1^{-1}(80\%) + \mathbf{F}_2^{-1}(80\%) < \mathbf{F}_{1+2}^{-1}(80\%)$$

2.4.4 Value-at-risk of a long/short portfolio

We note $S_{A,t}$ (resp. $S_{B,t}$) the price of stock A (resp. B) at time t. The portfolio value is:

$$P_t(w) = w_A S_{A,t} + w_B S_{B,t}$$

where w_A and w_B are the number of stocks A and B. We deduce that the P&L between t and $t+1$ is:

$$\begin{aligned}\Pi(w) &= P_{t+1} - P_t \\ &= w_A(S_{A,t+1} - S_{A,t}) + w_B(S_{B,t+1} - S_{B,t}) \\ &= w_A S_{A,t} R_{A,t+1} + w_B S_{B,t} R_{B,t+1} \\ &= W_{A,t} R_{A,t+1} + W_{B,t} R_{B,t+1}\end{aligned}$$

[6]We have $p_{1,i} = \Pr\{L_1 = \ell_i\}$ and $p_{2,i} = \Pr\{L_2 = \ell_i\}$.

where $R_{A,t+1}$ and $R_{B,t+1}$ are the asset returns of A and B between t and $t+1$, and $W_{A,t}$ and $W_{B,t}$ are the nominal wealth invested in stocks A and B at time t.

1. We have $W_{A,t} = +2$ and $W_{B,t} = -1$. The P&L (expressed in USD million) has the following expression:
$$\Pi(w) = 2R_{A,t+1} - R_{B,t+1}$$
We have $\Pi(w) \sim \mathcal{N}\left(0, \sigma^2(\Pi)\right)$ with:
$$\begin{aligned} \sigma(\Pi) &= \sqrt{(2\sigma_A)^2 + (-\sigma_B)^2 + 2\rho_{A,B} \times (2\sigma_A) \times (-\sigma_B)} \\ &= \sqrt{4 \times 0.20^2 + (-0.20)^2 - 4 \times 0.5 \times 0.20^2} \\ &= \sqrt{3} \times 20\% \\ &\simeq 34.64\% \end{aligned}$$
The annual volatility of the long/short portfolio is then equal to \$346 400. We consider the square-root-of-time rule to calculate the daily value-at-risk:
$$\begin{aligned} \mathrm{VaR}_{99\%}(w; \text{one day}) &= \frac{1}{\sqrt{260}} \times \Phi^{-1}(0.99) \times \sqrt{3} \times 20\% \\ &= 5.01\% \end{aligned}$$
The 99% value-at-risk is then equal to \$50 056.

2. We use the historical data to calculate the scenarios of asset returns $(R_{A,t+1}, R_{B,t+1})$. We then deduce the empirical distribution of the P&L with the formula $\Pi(w) = 2R_{A,t+1} - R_{B,t+1}$. Finally, we calculate the empirical quantile. With 250 scenarios, the 1% decile is between the second and third worst cases:
$$\begin{aligned} \mathrm{VaR}_{99\%}(w; \text{one day}) &= -\left[-56\,850 + \frac{1}{2}(-54\,270 - (-56\,850))\right] \\ &= 55\,560 \end{aligned}$$
The probability to lose \$55 560 per day is equal to 1%. We notice that the difference between the historical VaR and the Gaussian VaR is equal to 11%.

3. If we assume that the average of the last 60 VaRs is equal to the current VaR, we obtain:
$$\mathcal{K}^{\mathrm{IMA}} = m_c \times \sqrt{10} \times \mathrm{VaR}_{99\%}(w; \text{one day})$$
$\mathcal{K}^{\mathrm{IMA}}$ is respectively equal to \$474 877 and \$527 088 for the Gaussian and historical VaRs. In the case of the standardized measurement method, we have:
$$\begin{aligned} \mathcal{K}^{\mathrm{Specific}} &= 2 \times 8\% + 1 \times 8\% \\ &= \$240\,000 \end{aligned}$$
and:
$$\begin{aligned} \mathcal{K}^{\mathrm{General}} &= |2 - 1| \times 8\% \\ &= \$80\,000 \end{aligned}$$
We deduce that:
$$\begin{aligned} \mathcal{K}^{\mathrm{SMM}} &= \mathcal{K}^{\mathrm{Specific}} + \mathcal{K}^{\mathrm{General}} \\ &= \$320\,000 \end{aligned}$$

The internal model-based approach does not achieve a reduction of the required capital with respect to the standardized measurement method. Moreover, we have to add the stressed VaR under Basel 2.5 and the IMA regulatory capital is at least multiplied by a factor of 2.

4. If $\rho_{A,B} = -0.50$, the volatility of the P&L becomes:

$$\begin{aligned}\sigma(\Pi) &= \sqrt{4 \times 0.20^2 + (-0.20)^2 - 4 \times (-0.5) \times 0.20^2} \\ &= \sqrt{7} \times 20\%\end{aligned}$$

We deduce that:

$$\frac{\mathrm{VaR}_\alpha(\rho_{A,B} = -50\%)}{\mathrm{VaR}_\alpha(\rho_{A,B} = +50\%)} = \frac{\sigma(\Pi; \rho_{A,B} = -50\%)}{\sigma(\Pi; \rho_{A,B} = +50\%)} = \sqrt{\frac{7}{3}} = 1.53$$

The value-at-risk increases because the hedging effect of the positive correlation vanishes. With a negative correlation, a long/short portfolio becomes more risky than a long-only portfolio.

5. The P&L formula becomes:

$$\Pi(w) = W_{A,t} R_{A,t+1} + W_{B,t} R_{B,t+1} - (\mathcal{C}_{A,t+1} - \mathcal{C}_{A,t})$$

where $\mathcal{C}_{A,t}$ is the call option price. We have:

$$\mathcal{C}_{A,t+1} - \mathcal{C}_{A,t} \simeq \boldsymbol{\Delta}_t (S_{A,t+1} - S_{A,t})$$

where $\boldsymbol{\Delta}_t$ is the delta of the option. If the nominal of the option is USD 2 million, we obtain:

$$\begin{aligned}\Pi(w) &= 2R_A - R_B - 2 \times 0.5 \times R_A \\ &= R_A - R_B \qquad (2.1)\end{aligned}$$

and:

$$\begin{aligned}\sigma(\Pi) &= \sqrt{0.20^2 + (-0.20)^2 - 2 \times 0.5 \times 0.20^2} \\ &= 20\%\end{aligned}$$

If the nominal of the option is USD 4 million, we obtain:

$$\begin{aligned}\Pi(w) &= 2R_A - R_B - 4 \times 0.5 \times R_A \\ &= -R_B \qquad (2.2)\end{aligned}$$

and $\sigma(\Pi) = 20\%$. In both cases, we have:

$$\begin{aligned}\mathrm{VaR}_{99\%}(w; \text{one day}) &= \frac{1}{\sqrt{260}} \times \Phi^{-1}(0.99) \times 20\% \\ &= \$28\,900\end{aligned}$$

The value-at-risk of the long/short portfolio (2.1) is then equal to the value-at-risk of the short portfolio (2.2) because of two effects: the absolute exposure of the long/short portfolio is higher than the absolute exposure of the short portfolio, but a part of the risk of the long/short portfolio is hedged by the positive correlation between the two stocks.

6. We have:

$$\Pi(w) = W_{A,t} R_{A,t+1} - \left(\boldsymbol{\mathcal{C}}_{B,t+1} - \boldsymbol{\mathcal{C}}_{B,t}\right)$$

and:

$$\boldsymbol{\mathcal{C}}_{B,t+1} - \boldsymbol{\mathcal{C}}_{B,t} \simeq \boldsymbol{\Delta}_t \left(S_{B,t+1} - S_{B,t}\right)$$

where $\boldsymbol{\Delta}_t$ is the delta of the option. We note x the nominal of the option expressed in USD million. We obtain:

$$\begin{aligned} \Pi(w) &= 2R_A - x \times \boldsymbol{\Delta}_t \times R_B \\ &= 2R_A - \frac{x}{2} R_B \end{aligned}$$

We have[7]:

$$\begin{aligned} \sigma^2(\Pi) &= 4\sigma_A^2 + \frac{x^2\sigma_B^2}{4} + 2\rho_{A,B} \times (2\sigma_A) \times \left(-\frac{x}{2}\sigma_B\right) \\ &= \frac{\sigma_A^2}{4}\left(x^2 - 8\rho_{A,B} x + 16\right) \end{aligned}$$

Minimizing the Gaussian value-at-risk is equivalent to minimizing the variance of the P&L. We deduce that the first-order condition is:

$$\frac{\partial\, \sigma^2(\Pi)}{\partial\, x} = \frac{\sigma_A^2}{4}\left(2x - 8\rho_{A,B}\right) = 0$$

We deduce that the minimum VaR is reached when the nominal of the option is $x = 4\rho_{A,B}$. We finally obtain:

$$\begin{aligned} \sigma(\Pi) &= \frac{\sigma_A}{2}\sqrt{16\rho_{A,B}^2 - 32\rho_{A,B}^2 + 16} \\ &= 2\sigma_A\sqrt{1 - \rho_{A,B}^2} \end{aligned}$$

and:

$$\begin{aligned} \mathrm{VaR}_{99\%}(w; \text{one day}) &= \frac{1}{\sqrt{260}} \times 2.33 \times 2 \times 20\% \times \sqrt{1 - \rho_{A,B}^2} \\ &\simeq 5.78\% \times \sqrt{1 - \rho_{A,B}^2} \end{aligned}$$

If $\rho_{A,B}$ is negative (resp. positive), the exposure x is negative meaning that we have to buy (resp. to sell) a call option on stock B in order to hedge a part of the risk related to stock A. If $\rho_{A,B}$ is equal to zero, the exposure x is equal to zero because a position on stock B adds systematically a supplementary risk to the portfolio.

2.4.5 Value-at-risk of an equity portfolio hedged with put options

1. Let $R = (R_A, R_B)$ be the random vector of stock returns. We remind that[8]:

$$\operatorname{cov}(R) = \sigma^2(R_I)\,\beta\beta^\top + D$$

where $\beta = (\beta_A, \beta_B)$ and D is the covariance matrix of idiosyncratic risks.

[7] Because $\sigma_A = \sigma_B = 20\%$.

[8] See Exercise 2.4.2 on page 8.

(a) We deduce that $\sigma(R_j) = \sqrt{\beta_j^2 \sigma^2(R_I) + \tilde{\sigma}_j^2}$. We obtain

$$\sigma(R_A) = \sqrt{0.5^2 \times 4\% + 3\%} = 20\%$$

and:

$$\sigma(R_B) = \sqrt{1.5^2 \times 4\% + 7\%} = 40\%$$

The cross-correlation is:

$$\rho(R_A, R_B) = \frac{\sigma^2(R_I)\beta_A\beta_B}{\sigma(R_A)\sigma(R_B)} = \frac{4\% \times 0.5 \times 1.5}{20\% \times 40\%} = 37.5\%$$

(b) To find the correlation between the stocks and the index, we can proceed in two different ways.

i. We consider the random vector $R = (R_A, R_B, R_I)$. The formula $\operatorname{cov}(R) = \sigma^2(R_I)\beta\beta^\top + D$ is still valid with $\beta_3 = \beta_I = 1$ and $D_{3,3} = \tilde{\sigma}_I^2 = 0\%$. We obtain:

$$\rho(R_A, R_I) = \frac{\sigma^2(R_I)\beta_A\beta_I}{\sigma(R_A)\sigma(R_I)} = \frac{4\% \times 0.5 \times 1}{20\% \times 20\%} = 50\%$$

and:

$$\rho(R_B, R_I) = \frac{\sigma^2(R_I)\beta_B\beta_I}{\sigma(R_B)\sigma(R_I)} = \frac{4\% \times 1.5 \times 1}{40\% \times 20\%} = 75\%$$

ii. The definition of beta β_j is:

$$\beta_j = \frac{\operatorname{cov}(R_j, R_I)}{\sigma^2(R_I)} = \frac{\rho(R_j, R_I)\sigma(R_j)\sigma(R_I)}{\sigma^2(R_I)}$$

It follows that:

$$\rho(R_j, R_I) = \frac{\sigma(R_I)}{\sigma(R_j)}\beta_j$$

We retrieve the previous formula.

(c) The correlation matrix is then equal to:

$$\rho = \begin{pmatrix} 100.0\% & & \\ 37.5\% & 100.0\% & \\ 50.0\% & 75.0\% & 100.0\% \end{pmatrix}$$

We deduce that the covariance matrix Σ is:

$$\Sigma = \begin{pmatrix} 4\% & & \\ 3\% & 16\% & \\ 2\% & 6\% & 4\% \end{pmatrix}$$

2. Let $w = (w_A, w_B, w_I)$ be the composition of the portfolio. The expression of the P&L between t and $t+h$ is:

$$\begin{aligned} \Pi(w) &= w_A(S_{A,t+h} - S_{A,t}) + w_B(S_{B,t+h} - S_{B,t}) + w_A(S_{I,t+h} - S_{I,t}) \\ &= w_A S_{A,t} R_{A,t+h} + w_B S_{B,t} R_{B,t+h} + w_I S_{I,t} R_{I,t+h} \\ &= W_A R_{A,t+h} + W_B R_{B,t+h} + W_I R_{I,t+h} \\ &= W^\top R_{t+h} \end{aligned}$$

where W_j is the current wealth invested in asset j, $W = (W_A, W_B, W_I)$ is the vector of dollar notionals and $R_{t+h} = (R_{A,t+h}, R_{B,t+h}, R_{I,t+h})$ is the random vector of asset returns.

(a) We have $W_A = 400$, $W_B = 500$ and $W_I = 250$. We deduce that:

$$\begin{aligned}\sigma^2(\Pi) &= W^\top \Sigma W \\ &= 400^2 \times 4\% + 500^2 \times 16\% + 250^2 \times 4\% + \\ &\quad 2 \times 400 \times 500 \times 3\% + 2 \times 400 \times 250 \times 2\% + \\ &\quad 2 \times 500 \times 250 \times 6\%\end{aligned}$$

We find that $\sigma(\Pi)$ is equal to \$282 67. Using the square-root-of-time rule[9], it follows that:

$$\begin{aligned}\text{VaR}_{99\%}(w; \text{ten days}) &= \Phi^{-1}(99\%) \times 282.67 \times \sqrt{\frac{2}{52}} \\ &= \$128.96\end{aligned}$$

(b) The 99% quantile corresponds to the 2.6th order statistic of the sample. The historical value-at-risk is then the interpolated value between the second and third largest losses:

$$\begin{aligned}\text{VaR}_{99\%}(w; \text{one day}) &= 55.23 - (2.6 - 2) \times (55.23 - 52.06) \\ &= \$53.33\end{aligned}$$

We deduce that the 10-day VaR is:

$$\begin{aligned}\text{VaR}_{99\%}(w; \text{ten days}) &= \sqrt{10} \times \text{VaR}_{99\%}(w; \text{one day}) \\ &= \$168.64\end{aligned}$$

(c) If we assume that the average of the last 60 VaRs is equal to the current VaR, we obtain:

$$\mathcal{K}^{\text{IMA}} = m_c \times \text{VaR}_{99\%}(w; \text{ten days})$$

$\mathcal{K}^{\text{IMA}}$ is respectively equal to \$387 and \$506 for the Gaussian and historical VaRs. In the case of the standardized measurement method, we have[10]:

$$\begin{aligned}\mathcal{K}^{\text{Specific}} &= (W_A + W_B) \times 8\% + W_I \times 4\% \\ &= 900 \times 8\% + 250 \times 4\% \\ &= \$82\end{aligned}$$

and:

$$\begin{aligned}\mathcal{K}^{\text{General}} &= |W_A + W_B + W_I| \times 8\% \\ &= \$92\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathcal{K}^{\text{SMM}} &= \mathcal{K}^{\text{Specific}} + \mathcal{K}^{\text{General}} \\ &= \$174\end{aligned}$$

[9] We use the following correspondence: 10 days is equivalent to 2 weeks and one year is equivalent to 52 weeks.

[10] We assume that the specific capital charge for an equity index is 4%.

3. Let x be the number of put options. The expression of the P&L becomes:

$$\Pi(w) = W_A R_{A,t+h} + W_B R_{B,t+h} + W_I R_{I,t+h} + x\left(\boldsymbol{\mathcal{P}}_{t+h} - \boldsymbol{\mathcal{P}}_t\right)$$

where $\boldsymbol{\mathcal{P}}_t$ is the value of the put option at time t. Under the delta approach, we have:

$$\begin{aligned}\boldsymbol{\mathcal{P}}_{t+h} - \boldsymbol{\mathcal{P}}_t &\simeq \boldsymbol{\Delta}_t\left(S_{I,t+h} - S_{I,t}\right) \\ &= \boldsymbol{\Delta}_t S_{I,t} R_{I,t+h}\end{aligned}$$

We deduce that:

$$\Pi(w) = W_A R_{A,t+h} + W_B R_{B,t+h} + \left(W_I + x\boldsymbol{\Delta}_t S_{I,t}\right) R_{I,t+h}$$

Using the numerical values, we obtain:

$$\Pi(w) = 400 R_{A,t+h} + 500 R_{B,t+h} + (250 - 12.5x) \times R_{I,t+h}$$

(a) To hedge 50% of the index exposure, the number of put options must satisfy the following equation:

$$250 - 12.5x = 125$$

The portfolio manager must purchase 10 put options. In this case, the expression of the P&L becomes:

$$\Pi(w) = 400 R_{A,t+h} + 500 R_{B,t+h} + 125 R_{I,t+h}$$

and the 10-day Gaussian VaR is equal to \$119.43.

(b) We have:

$$\begin{aligned}\Pi(w) &= 400 R_{A,t+h} + 500 R_{B,t+h} + (250 - 12.5x) \times R_{I,t+h} \\ &= 400\left(\beta_A R_{I,t+h} + \varepsilon_{A,t+h}\right) + 500\left(\beta_B R_{I,t+h} + \varepsilon_{B,t+h}\right) + \\ &\quad (250 - 12.5x) \times R_{I,t+h} \\ &= (1200 - 12.5x) \times R_{I,t+h} + 400 \times \varepsilon_{A,t+h} + 500 \times \varepsilon_{B,t+h}\end{aligned}$$

As the index return is not correlated with the idiosyncratic risks, minimizing the VaR is equivalent to minimizing the beta exposure in the index:

$$x = \frac{1200}{12.5} = 96$$

The purchase of 96 put options is required to remove the directional risk. In this case, the P&L reduces to:

$$\Pi(w) = 400 \times \varepsilon_{A,t+h} + 500 \times \varepsilon_{B,t+h}$$

and its volatility becomes:

$$\sigma(\Pi) = \sqrt{400^2 \times 3\% + 500^2 \times 7\%} = \$149.33$$

We deduce that the minimum 10-day VaR is equal to \$68.13. In Figure 2.4, we show the evolution of the VaR with the number of purchased options. We verify that the minimum is reached for $x = 96$.

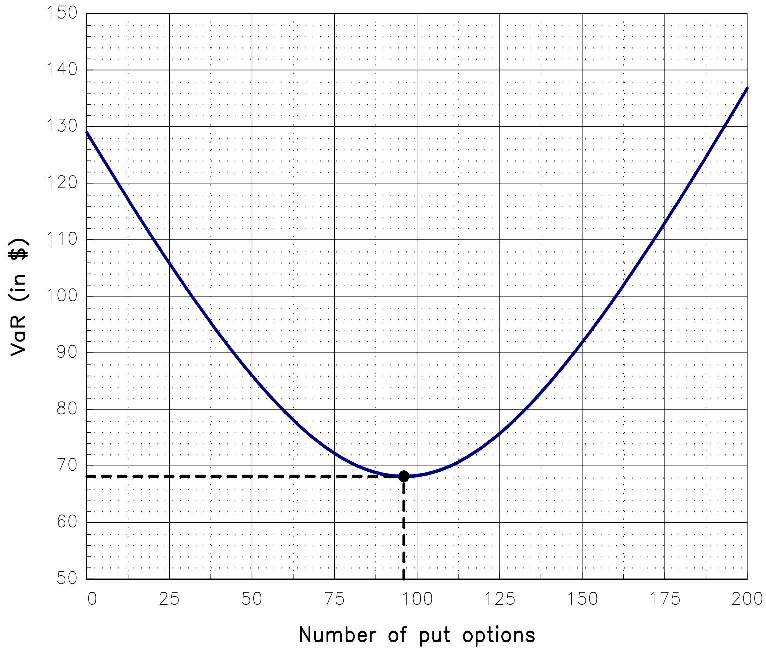

FIGURE 2.4: Value of the 10-day VaR with respect to the number of purchased options

2.4.6 Risk management of exotic derivatives

Let $\boldsymbol{\mathcal{C}}_t$ be the option price at time t. The P&L of the trader between t and $t+1$ is:

$$\Pi = -\left(\boldsymbol{\mathcal{C}}_{t+1} - \boldsymbol{\mathcal{C}}_t\right)$$

The formulation of the exercise suggests that there are two main risk factors: the price of the underlying asset S_t and the implied volatility Σ_t. We then obtain:

$$\Pi = C_t\left(S_t, \Sigma_t\right) - C_{t+1}\left(S_{t+1}, \Sigma_{t+1}\right)$$

1. We have:

$$\begin{array}{rcl} \Pi & = & C_t\left(S_t, \Sigma_t\right) - C_{t+1}\left(S_{t+1}, \Sigma_{t+1}\right) \\ & \approx & -\boldsymbol{\Delta}_t\left(S_{t+1} - S_t\right) - \frac{1}{2}\boldsymbol{\Gamma}_t\left(S_{t+1} - S_t\right)^2 - \boldsymbol{\upsilon}_t\left(\Sigma_{t+1} - \Sigma_t\right) \end{array}$$

   Using the numerical values of $\boldsymbol{\Delta}_t$, $\boldsymbol{\Gamma}_t$ and $\boldsymbol{\upsilon}_t$, we obtain:

$$\begin{array}{rcl} \Pi & \approx & -0.49 \times (97 - 100) - \frac{1}{2} \times 0.02 \times (97 - 100)^2 \\ & = & 1.47 - 0.09 \\ & = & 1.38 \end{array}$$

   We explain the P&L by the sensitivities very well.

2. We have:

$$\begin{array}{rcl} \Pi & = & C_{t+1}\left(S_{t+1}, \Sigma_{t+1}\right) - C_{t+2}\left(S_{t+2}, \Sigma_{t+2}\right) \\ & \approx & -\boldsymbol{\Delta}_{t+1}\left(S_{t+2} - S_{t+1}\right) - \frac{1}{2}\boldsymbol{\Gamma}_{t+1}\left(S_{t+2} - S_{t+1}\right)^2 - \\ & & \boldsymbol{\upsilon}_{t+1}\left(\Sigma_{t+2} - \Sigma_{t+1}\right) \end{array}$$

Using the numerical values of $\boldsymbol{\Delta}_{t+1}$, $\boldsymbol{\Gamma}_{t+1}$ and $\boldsymbol{\upsilon}_{t+1}$, we obtain:

$$\begin{aligned}\Pi &\approx -0.49 \times 0 - \frac{1}{2} \times 0.02 \times 0^2 - 0.38 \times (22 - 20) \\ &= -0.76\end{aligned}$$

To compare this value with the true P&L, we have to calculate $\boldsymbol{\mathcal{C}}_{t+1}$:

$$\begin{aligned}\boldsymbol{\mathcal{C}}_{t+1} &= \boldsymbol{\mathcal{C}}_t - (\boldsymbol{\mathcal{C}}_t - \boldsymbol{\mathcal{C}}_{t+1}) \\ &= 6.78 - 1.37 \\ &= 5.41\end{aligned}$$

We deduce that:

$$\begin{aligned}\Pi &= \boldsymbol{\mathcal{C}}_{t+1} - \boldsymbol{\mathcal{C}}_{t+2} \\ &= 5.41 - 6.17 \\ &= -0.76\end{aligned}$$

Again, the sensitivities explain the P&L very well.

3. We have:

$$\begin{aligned}\Pi &= C_{t+2}(S_{t+2}, \Sigma_{t+2}) - C_{t+3}(S_{t+3}, \Sigma_{t+3}) \\ &\approx -\boldsymbol{\Delta}_{t+2}(S_{t+3} - S_{t+2}) - \frac{1}{2}\boldsymbol{\Gamma}_{t+2}(S_{t+3} - S_{t+2})^2 - \\ &\quad \boldsymbol{\upsilon}_{t+2}(\Sigma_{t+3} - \Sigma_{t+2})\end{aligned}$$

Using the numerical values of $\boldsymbol{\Delta}_{t+2}$, $\boldsymbol{\Gamma}_{t+2}$ and $\boldsymbol{\upsilon}_{t+2}$, we obtain:

$$\begin{aligned}\Pi &\approx -0.44 \times (95 - 97) - \frac{1}{2} \times 0.018 \times (95 - 97)^2 - \\ &\quad 0.38 \times (19 - 22) \\ &= 0.88 - 0.036 + 1.14 \\ &= 1.984\end{aligned}$$

The P&L approximated by the Greek coefficients largely overestimate the true value of the P&L.

4. We notice that the approximation using the Greek coefficients works very well when one risk factor remains constant:

   (a) Between t and $t+1$, the price of the underlying asset changes, but not the implied volatility;

   (b) Between $t+1$ and $t+2$, this is the implied volatility that changes whereas the price of the underlying asset is constant.

   Therefore, we can assume that the bad approximation between $t+2$ and $t+3$ is due to the cross effect between S_t and Σ_t. In terms of model risk, the P&L is then exposed to the vanna risk, meaning that the Black-Scholes model is not appropriate to price and hedge this exotic option.

2.4.7 P&L approximation with Greek sensitivities

1. We note $\boldsymbol{\mathcal{C}}_t = C_t(S_t, \Sigma_t, T)$ meaning that the option price depends on the current price S_t, the implied volatility Σ_t and the maturity date T. The delta of the option is the first derivative of $\boldsymbol{\mathcal{C}}_t$ with respect to S_t.

$$\boldsymbol{\Delta}_t = \frac{\partial\, C_t(S_t, \Sigma_t, T)}{\partial\, S_t}$$

whereas the gamma is the second derivative:

$$\boldsymbol{\Gamma}_t = \frac{\partial^2\, C_t(S_t, \Sigma_t, T)}{\partial\, S_t^2}$$

The theta of the option is the first derivative of $\boldsymbol{\mathcal{C}}_t$ with respect to the time t. We have:

$$\boldsymbol{\Theta}_t = \partial_t\, C_t(S_t, \Sigma_t, T)$$

For the vega coefficient, we have:

$$\boldsymbol{\upsilon}_t = \frac{\partial\, C_t(S_t, \Sigma_t, T)}{\partial\, \Sigma_t}$$

2. Let r_t and b_t be the interest rate and the cost-of-carry parameter. We note $\tau = T - t$ the residual maturity. The Black-Scholes formula is:

$$\boldsymbol{\mathcal{C}}_t = S_t e^{(b_t - r_t)\tau}\Phi(d_1) - K e^{-r_t\tau}\Phi(d_2)$$

with:

$$\begin{aligned} d_1 &= \frac{1}{\Sigma_t\sqrt{\tau}}\left(\ln\frac{S_t}{K} + b_t\tau\right) + \frac{1}{2}\Sigma_t\sqrt{\tau} \\ d_2 &= d_1 - \Sigma_t\sqrt{\tau} \end{aligned}$$

To calculate the Greek coefficients, we need the following preliminary result:

$$K e^{-r_t\tau}\phi(d_2) = S_t e^{(b_t - r_t)\tau}\phi(d_1) \tag{2.3}$$

Indeed, we have:

$$\begin{aligned} \phi(d_2) &= \phi\left(d_1 - \Sigma_t\sqrt{\tau}\right) \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(d_1 - \Sigma_t\sqrt{\tau}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(d_1^2 - 2d_1\Sigma_t\sqrt{\tau} + \Sigma_t^2\tau\right)\right) \\ &= \phi(d_1)\exp\left(d_1\Sigma_t\sqrt{\tau} - \frac{1}{2}\Sigma_t^2\tau\right) \\ &= \phi(d_1)\exp\left(\ln\frac{S_t}{K} + b_t\tau\right) \\ &= \frac{S_t}{K}e^{b_t\tau}\phi(d_1) \end{aligned}$$

The derivation of Equation (2.3) is then straightforward. We deduce that:

$$\begin{aligned} \boldsymbol{\Delta}_t &= e^{(b_t - r_t)\tau}\Phi(d_1) + S_t e^{(b_t - r_t)\tau}\phi(d_1)\frac{\partial\, d_1}{\partial\, S_t} - K e^{-r_t\tau}\phi(d_2)\frac{\partial\, d_2}{\partial\, S_t} \\ &= e^{(b_t - r_t)\tau}\Phi(d_1) + S_t e^{(b_t - r_t)\tau}\phi(d_1)\frac{\partial\, d_1}{\partial\, S_t} - K e^{-r_t\tau}\phi(d_2)\frac{\partial\, d_2}{\partial\, S_t} \end{aligned}$$

We have:

$$\frac{\partial\, d_1}{\partial\, S_t} = \frac{\partial\, d_2}{\partial\, S_t} = \frac{1}{S_t \Sigma_t \sqrt{\tau}}$$

We finally obtain:

$$\begin{aligned}
\boldsymbol{\Delta}_t &= e^{(b_t - r_t)\tau} \Phi(d_1) + \frac{S_t e^{(b_t - r_t)\tau} \phi(d_1) - K e^{-r_t \tau} \phi(d_2)}{S_t \Sigma_t \sqrt{\tau}} \\
&= e^{(b_t - r_t)\tau} \Phi(d_1)
\end{aligned}$$

The expression of the gamma is therefore:

$$\boldsymbol{\Gamma}_t = \frac{e^{(b_t - r_t)\tau} \phi(d_1)}{S_t \Sigma_t \sqrt{\tau}}$$

To calculate the theta, we first calculate the derivative of d_1 and d_2 with respect to τ:

$$\begin{aligned}
\frac{\partial\, d_1}{\partial\, \tau} &= -\frac{1}{2\Sigma_t \tau \sqrt{\tau}} \ln \frac{S_t}{K} + \frac{b_t}{2\Sigma_t \sqrt{\tau}} + \frac{\Sigma_t}{4\sqrt{\tau}} \\
\frac{\partial\, d_2}{\partial\, \tau} &= -\frac{1}{2\Sigma_t \tau \sqrt{\tau}} \ln \frac{S_t}{K} + \frac{b_t}{2\Sigma_t \sqrt{\tau}} - \frac{\Sigma_t}{4\sqrt{\tau}}
\end{aligned}$$

We deduce then:

$$\begin{aligned}
\partial_\tau\, C_t(S_t, \Sigma_t, T) &= (b_t - r_t) S_t e^{(b_t - r_t)\tau} \Phi(d_1) + r_t K e^{-r_t \tau} \Phi(d_2) + \\
&\quad S_t e^{(b_t - r_t)\tau} \phi(d_1) \frac{\partial\, d_1}{\partial\, \tau} - K e^{-r_t \tau} \phi(d_2) \frac{\partial\, d_2}{\partial\, \tau} \\
&= (b_t - r_t) S_t e^{(b_t - r_t)\tau} \Phi(d_1) + r_t K e^{-r_t \tau} \Phi(d_2) + \\
&\quad S_t e^{(b_t - r_t)\tau} \phi(d_1) \left(\frac{\partial\, d_1}{\partial\, \tau} - \frac{\partial\, d_2}{\partial\, \tau} \right) \\
&= (b_t - r_t) S_t e^{(b_t - r_t)\tau} \Phi(d_1) + r_t K e^{-r_t \tau} \Phi(d_2) + \\
&\quad S_t e^{(b_t - r_t)\tau} \phi(d_1) \left(\frac{\Sigma_t}{4\sqrt{\tau}} + \frac{\Sigma_t}{4\sqrt{\tau}} \right) \\
&= (b_t - r_t) S_t e^{(b_t - r_t)\tau} \Phi(d_1) + r_t K e^{-r_t \tau} \Phi(d_2) + \\
&\quad \frac{1}{2\sqrt{\tau}} S_t \Sigma_t e^{(b_t - r_t)\tau} \phi(d_1)
\end{aligned}$$

The expression of the theta coefficient is then the opposite of $\partial_\tau\, C_t(S_t, \Sigma_t, T)$. For the vega coefficient, we obtain:

$$\begin{aligned}
\boldsymbol{\upsilon}_t &= S_t e^{(b_t - r_t)\tau} \phi(d_1) \frac{\partial\, d_1}{\partial\, \Sigma_t} - K e^{-r_t \tau} \phi(d_2) \frac{\partial\, d_2}{\partial\, \Sigma_t} \\
&= S_t e^{(b_t - r_t)\tau} \phi(d_1) \left(\frac{\partial\, d_1}{\partial\, \Sigma_t} - \frac{\partial\, d_2}{\partial\, \Sigma_t} \right)
\end{aligned}$$

We have:

$$\begin{aligned}
\frac{\partial\, d_1}{\partial\, \Sigma_t} &= -\frac{1}{\Sigma_t^2 \sqrt{\tau}} \left(\ln \frac{S_t}{K} + b_t \tau \right) + \frac{1}{2}\sqrt{\tau} \\
\frac{\partial\, d_2}{\partial\, \Sigma_t} &= -\frac{1}{\Sigma_t^2 \sqrt{\tau}} \left(\ln \frac{S_t}{K} + b_t \tau \right) - \frac{1}{2}\sqrt{\tau}
\end{aligned}$$

It follows that:

$$\boldsymbol{\upsilon}_t = S_t \sqrt{\tau} e^{(b_t - r_t)\tau} \phi(d_1)$$

3. (a) To calculate $\mathcal{C}_0$, $\mathbf{\Delta}_0$, $\mathbf{\Gamma}_0$ and $\mathbf{\Theta}_0$, we consider the Back-Scholes formulas with $b_t = r_t$. We have $\mathcal{C}_t = S_t \Phi(d_1) - K e^{-r_t \tau} \Phi(d_2)$, $\mathbf{\Delta}_t = \Phi(d_1)$, $\mathbf{\Gamma}_t = \phi(d_1) / (S_t \Sigma_t \sqrt{\tau})$, $\mathbf{\Theta}_t = -r_t K e^{-r_t \tau} \Phi(d_2) - S_t \Sigma_t \phi(d_1) / (2\sqrt{\tau})$ with the following numerical values: $S_t = 100$, $\Sigma_t = 20\%$, $\tau = 1$ and $r_t = 5\%$. We notice that the option price is a decreasing function of the strike, because it is a convex function of the strike. The delta and gamma coefficients are positive. The delta is a decreasing function with respect to K. This is not the case of the gamma. The theta of the call option is negative, because the time value decreases with the residual maturity.

   (b) We apply the Black-Scholes formula $\mathcal{C}_t = S_t \Phi(d_1) - K e^{-r_t \tau} \Phi(d_2)$ with $S_t = 102$, $\Sigma_t = 20\%$, $r_t = 5\%$ and $\tau = 1 - 1/252$ because the residual maturity is one year minus one trading day. We deduce that the P&L of a long position on this option is $\Pi = \mathcal{C}_1 - \mathcal{C}_0$:

K	80	95	100	105	120
Π	1.852	1.464	1.285	1.099	0.589

   (c) We obtain the following results:

K	80	95	100	105	120
$\Pi^{\mathbf{\Delta}}$	1.857	1.456	1.274	1.084	0.574
$\Pi^{\mathbf{\Delta}+\mathbf{\Gamma}}$	1.871	1.489	1.311	1.124	0.608
$\Pi^{\mathbf{\Delta}+\mathbf{\Theta}}$	1.839	1.432	1.249	1.060	0.556
$\Pi^{\mathbf{\Delta}+\mathbf{\Gamma}+\mathbf{\Theta}}$	1.853	1.465	1.287	1.100	0.590

   The approximation of the P&L by the Greek sensitivities is very accurate.

   (d) We obtain the following results:

K	80	95	100	105	120
Π	45.386	42.001	40.026	37.596	28.090
$\Pi^{\mathbf{\Delta}}$	44.575	34.939	30.568	26.027	13.785
$\Pi^{\mathbf{\Delta}+\mathbf{\Gamma}}$	52.424	54.058	52.182	48.877	33.412
$\Pi^{\mathbf{\Delta}+\mathbf{\Theta}}$	42.186	31.793	27.361	22.888	11.445
$\Pi^{\mathbf{\Delta}+\mathbf{\Gamma}+\mathbf{\Theta}}$	50.036	50.912	48.975	45.739	31.071

   In this case, the approximation of the P&L by the Greek sensitivities is not very good. Indeed, the remaining maturity is now six months meaning that (1) the theta effect is not well measured and (2) the price of the underlying asset has changed significantly. In this situation, the delta P&L is overestimated.

2.4.8 Calculating the non-linear quadratic value-at-risk

1. We have:

$$\begin{aligned}
\mathbb{E}\left[X^{2n}\right] &= \int_{-\infty}^{+\infty} x^{2n} \phi(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x^{2n-1} x \phi(x) \, \mathrm{d}x
\end{aligned}$$

Using the integration by parts formula, we obtain[11]:

$$\begin{aligned}\mathbb{E}\left[X^{2n}\right] &= \left[-x^{2n-1}\phi(x)\right]_{-\infty}^{+\infty} + (2n-1)\int_{-\infty}^{+\infty} x^{2n-2}\phi(x)\,\mathrm{d}x \\ &= (2n-1)\int_{-\infty}^{+\infty} x^{2n-2}\phi(x)\,\mathrm{d}x \\ &= (2n-1)\,\mathbb{E}\left[X^{2n-2}\right]\end{aligned}$$

We deduce that $\mathbb{E}\left[X^2\right] = 1$, $\mathbb{E}\left[X^4\right] = (2\times 2-1)\,\mathbb{E}\left[X^2\right] = 3$, $\mathbb{E}\left[X^6\right] = (2\times 3-1)\,\mathbb{E}\left[X^4\right] = 15$ and $\mathbb{E}\left[X^8\right] = (2\times 4-1)\,\mathbb{E}\left[X^4\right] = 105$. For the odd moments, we obtain:

$$\begin{aligned}\mathbb{E}\left[X^{2n+1}\right] &= \int_{-\infty}^{+\infty} x^{2n+1}\phi(x)\,\mathrm{d}x \\ &= 0\end{aligned}$$

because $x^{2n+1}\phi(x)$ is an odd function.

2. Let $\boldsymbol{\mathcal{C}}_t$ be the value of the call option at time t. The P&L is equal to:

$$\Pi(w) = \boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t$$

where h is the holding period[12]. We also have $S_{t+h} = (1 + R_{t+h})\,S_t$ where R_{t+h} is the asset return. We notice that the daily volatility is equal to:

$$\sigma = \frac{32.25\%}{\sqrt{260}} = 2\%$$

We deduce that $R_{t+h} \sim \mathcal{N}(0, 4\text{ bps})$.

(a) We have:

$$\begin{aligned}\Pi(w) &\approx \boldsymbol{\Delta}_t\,(S_{t+h} - S_t) \\ &= \boldsymbol{\Delta}_t S_t R_{t+h}\end{aligned}$$

It follows that $\Pi(w) \sim \mathcal{N}\left(0, \boldsymbol{\Delta}_t^2\sigma^2 S_t^2\right)$ where σ is the volatility of R_{t+h} and:

$$\mathrm{VaR}_\alpha(w;h) = \Phi^{-1}(\alpha)\,|\boldsymbol{\Delta}_t|\,\sigma S_t$$

The numerical application gives $\mathrm{VaR}_\alpha(w;h) = 2.33$ dollars.

(b) In the case of the delta-gamma approximation, we obtain:

$$\begin{aligned}\Pi(w) &\approx \boldsymbol{\Delta}_t\,(S_{t+h} - S_t) + \frac{1}{2}\boldsymbol{\Gamma}_t\,(S_{t+h} - S_t)^2 \\ &= \boldsymbol{\Delta}_t R_{t+h} S_t + \frac{1}{2}\boldsymbol{\Gamma}_t R_{t+h}^2 S_t^2\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbb{E}[\Pi] &= \mathbb{E}\left[\boldsymbol{\Delta}_t R_{t+h} S_t + \frac{1}{2}\boldsymbol{\Gamma}_t R_{t+h}^2 S_t^2\right] \\ &= \frac{1}{2}\boldsymbol{\Gamma}_t\mathbb{E}\left[R_{t+h}^2\right] S_t^2 \\ &= \frac{1}{2}\boldsymbol{\Gamma}_t\sigma^2 S_t^2\end{aligned}$$

[11] because $\phi'(x) = -x\phi(x)$.

[12] Here, h is equal to one day.

and:

$$\begin{aligned}\mathbb{E}\left[\Pi^2\right] &= \mathbb{E}\left[\left(\boldsymbol{\Delta}_t R_{t+h} S_t + \frac{1}{2}\boldsymbol{\Gamma}_t R_{t+h}^2 S_t^2\right)^2\right] \\ &= \mathbb{E}\left[\boldsymbol{\Delta}_t^2 R_{t+h}^2 S_t^2 + \boldsymbol{\Delta}_t \boldsymbol{\Gamma}_t R_{t+h}^3 S_t^3 + \frac{1}{4}\boldsymbol{\Gamma}_t^2 R_{t+h}^4 S_t^4\right]\end{aligned}$$

We have $R_{t+h} = \sigma X$ with $X \sim \mathcal{N}(0,1)$. It follows that:

$$\mathbb{E}\left[\Pi^2\right] = \boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{3}{4}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4$$

because $\mathbb{E}[X] = 0$, $\mathbb{E}\left[X^2\right] = 1$, $\mathbb{E}\left[X^3\right] = 0$ and $\mathbb{E}\left[X^4\right] = 3$. The standard deviation of the P&L is then:

$$\begin{aligned}\sigma(\Pi) &= \sqrt{\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{3}{4}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4 - \left(\frac{1}{2}\boldsymbol{\Gamma}_t \sigma^2 S_t^2\right)^2} \\ &= \sqrt{\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4}\end{aligned}$$

Therefore, the Gaussian approximation of the P&L is:

$$\Pi(w) \sim \mathcal{N}\left(\frac{1}{2}\boldsymbol{\Gamma}_t \sigma^2 S_t^2, \boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4\right)$$

We deduce that the Gaussian value-at-risk is:

$$\mathrm{VaR}_\alpha(w;h) = -\frac{1}{2}\boldsymbol{\Gamma}_t \sigma^2 S_t^2 + \Phi^{-1}(\alpha)\sqrt{\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4}$$

The numerical application gives $\mathrm{VaR}_\alpha(w;h) = 2.29$ dollars.

(c) Let $L = -\Pi$ be the loss. We recall that the Cornish-Fisher value-at-risk is equal to (FRM, page 88):

$$\mathrm{VaR}_\alpha(w;h) = \mu(L) + \mathfrak{z}(\alpha;\gamma_1(L),\gamma_2(L)) \cdot \sigma(L)$$

with:

$$\begin{aligned}\mathfrak{z}(\alpha;\gamma_1,\gamma_2) &= z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\gamma_1 + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\gamma_2 - \\ &\quad \frac{1}{36}\left(2z_\alpha^3 - 5z_\alpha\right)\gamma_1^2 + \cdots\end{aligned}$$

and $z_\alpha = \Phi^{-1}(\alpha)$. γ_1 et γ_2 are the skewness and excess kurtosis of the loss L. We have seen that:

$$\Pi(w) = \boldsymbol{\Delta}_t \sigma S_t X + \frac{1}{2}\boldsymbol{\Gamma}_t \sigma^2 S_t^2 X^2$$

with $X \sim \mathcal{N}(0,1)$. Using the results in Question 1, we have $\mathbb{E}[X] = \mathbb{E}\left[X^3\right] = \mathbb{E}\left[X^5\right] = \mathbb{E}\left[X^7\right] = 0$, $\mathbb{E}\left[X^2\right] = 1$, $\mathbb{E}\left[X^4\right] = 3$, $\mathbb{E}\left[X^6\right] = 15$ and $\mathbb{E}\left[X^8\right] = 105$. We deduce that:

$$\begin{aligned}\mathbb{E}\left[\Pi^3\right] &= \mathbb{E}\left[\boldsymbol{\Delta}_t^3 \sigma^3 S_t^3 X^3 + \frac{3}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 X^4\right] + \\ &\quad \mathbb{E}\left[\frac{3}{4}\boldsymbol{\Delta}_t \boldsymbol{\Gamma}_t^2 \sigma^5 S_t^5 X^5 + \frac{1}{8}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6 X^6\right] \\ &= \frac{9}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 + \frac{15}{8}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[\Pi^4\right] &= \mathbb{E}\left[\left(\boldsymbol{\Delta}_t \sigma S_t X + \frac{1}{2}\boldsymbol{\Gamma}_t \sigma^2 S_t^2 X^2\right)^4\right] \\ &= 3\boldsymbol{\Delta}_t^4 \sigma^4 S_t^4 + \frac{45}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + \frac{105}{16}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8\end{aligned}$$

The centered moments are then:

$$\begin{aligned}\mathbb{E}\left[\left(\Pi - \mathbb{E}\left[\Pi\right]\right)^3\right] &= \mathbb{E}\left[\Pi^3\right] - 3\mathbb{E}\left[\Pi\right]\mathbb{E}\left[\Pi^2\right] + 2\mathbb{E}^3\left[\Pi\right] \\ &= \frac{9}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 + \frac{15}{8}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6 - \frac{3}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 - \\ &\quad \frac{9}{8}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6 + \frac{2}{8}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6 \\ &= 3\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 + \boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[\left(\Pi - \mathbb{E}\left[\Pi\right]\right)^4\right] &= \mathbb{E}\left[\Pi^4\right] - 4\mathbb{E}\left[\Pi\right]\mathbb{E}\left[\Pi^3\right] + 6\mathbb{E}^2\left[\Pi\right]\mathbb{E}\left[\Pi^2\right] - \\ &\quad 3\mathbb{E}^4\left[\Pi\right] \\ &= 3\boldsymbol{\Delta}_t^4 \sigma^4 S_t^4 + \frac{45}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + \frac{105}{16}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8 - \\ &\quad 9\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 - \frac{15}{4}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8 + \\ &\quad \frac{3}{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + \frac{9}{8}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8 - \frac{3}{16}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8 \\ &= 3\boldsymbol{\Delta}_t^4 \sigma^4 S_t^4 + 15\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + \frac{15}{4}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8\end{aligned}$$

It follows that the skewness is:

$$\begin{aligned}\gamma_1\left(L\right) &= -\gamma_1\left(\Pi\right) \\ &= -\frac{\mathbb{E}\left[\left(\Pi - \mathbb{E}\left[\Pi\right]\right)^3\right]}{\sigma^3\left(\Pi\right)} \\ &= -\frac{3\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 + \boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6}{\left(\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4\right)^{3/2}} \\ &= -\frac{6\sqrt{2}\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t \sigma^4 S_t^4 + 2\sqrt{2}\boldsymbol{\Gamma}_t^3 \sigma^6 S_t^6}{\left(2\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4\right)^{3/2}}\end{aligned}$$

whereas the excess kurtosis is:

$$\begin{aligned}\gamma_2\left(L\right) &= \gamma_2\left(\Pi\right) \\ &= \frac{\mathbb{E}\left[\left(\Pi - \mathbb{E}\left[\Pi\right]\right)^4\right]}{\sigma^4\left(\Pi\right)} - 3 \\ &= \frac{3\boldsymbol{\Delta}_t^4 \sigma^4 S_t^4 + 15\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + \frac{15}{4}\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8}{\left(\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4\right)^2} - 3 \\ &= \frac{12\boldsymbol{\Delta}_t^2 \boldsymbol{\Gamma}_t^2 \sigma^6 S_t^6 + 3\boldsymbol{\Gamma}_t^4 \sigma^8 S_t^8}{\left(\boldsymbol{\Delta}_t^2 \sigma^2 S_t^2 + \frac{1}{2}\boldsymbol{\Gamma}_t^2 \sigma^4 S_t^4\right)^2}\end{aligned}$$

Using the numerical values, we obtain $\mu(L) = -0.0400$, $\sigma(L) = 1.0016$, $\gamma_1(L) = -0.2394$, $\gamma_2(L) = 0.0764$, $\mathfrak{z}(\alpha; \gamma_1, \gamma_2) = 2.1466$ and $\mathrm{VaR}_\alpha(w; h) = 2.11$ dollars. The value-at-risk is reduced with the Cornish-Fisher approximation because the skewness is negative whereas the excess kurtosis is very small.

3. (a) We have:

$$\begin{aligned} Y &= X^\top A X \\ &= \left(\Sigma^{-1/2} X\right)^\top \Sigma^{1/2} A \Sigma^{1/2} \left(\Sigma^{-1/2} X\right) \\ &= \tilde{X}^\top \tilde{A} \tilde{X} \end{aligned}$$

with $\tilde{A} = \Sigma^{1/2} A \Sigma^{1/2}$, $\tilde{X} \sim \mathcal{N}\left(\tilde{\mu}, \tilde{\Sigma}\right)$, $\tilde{\mu} = \Sigma^{-1/2}\mu$ and $\tilde{\Sigma} = I$. We deduce that:

$$\begin{aligned} \mathbb{E}[Y] &= \tilde{\mu}^\top \tilde{A} \tilde{\mu} + \operatorname{tr}\left(\tilde{A}\right) \\ &= \mu^\top A \mu + \operatorname{tr}\left(\Sigma^{1/2} A \Sigma^{1/2}\right) \\ &= \mu^\top A \mu + \operatorname{tr}(A\Sigma) \end{aligned}$$

and:

$$\begin{aligned} \operatorname{var}(Y) &= \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y] \\ &= 4\tilde{\mu}^\top \tilde{A}^2 \tilde{\mu} + 2\operatorname{tr}\left(\tilde{A}^2\right) \\ &= 4\mu^\top A \Sigma A \mu + 2\operatorname{tr}\left(\Sigma^{1/2} A \Sigma A \Sigma^{1/2}\right) \\ &= 4\mu^\top A \Sigma A \mu + 2\operatorname{tr}\left((A\Sigma)^2\right) \end{aligned}$$

(b) For the moments, we obtain:

$$\begin{aligned} \mathbb{E}[Y] &= \operatorname{tr}(A\Sigma) \\ \mathbb{E}\left[Y^2\right] &= (\operatorname{tr}(A\Sigma))^2 + 2\operatorname{tr}\left((A\Sigma)^2\right) \\ \mathbb{E}\left[Y^3\right] &= (\operatorname{tr}(A\Sigma))^3 + 6\operatorname{tr}(A\Sigma)\operatorname{tr}\left((A\Sigma)^2\right) + 8\operatorname{tr}\left((A\Sigma)^3\right) \\ \mathbb{E}\left[Y^4\right] &= (\operatorname{tr}(A\Sigma))^4 + 32\operatorname{tr}(A\Sigma)\operatorname{tr}\left((A\Sigma)^3\right) + \\ & \quad 12\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^2 + 12(\operatorname{tr}(A\Sigma))^2\operatorname{tr}\left((A\Sigma)^2\right) + \\ & \quad 48\operatorname{tr}\left((A\Sigma)^4\right) \end{aligned}$$

It follows that the first and second centered moments are $\mu(Y) = \operatorname{tr}(A\Sigma)$ and $\operatorname{var}(Y) = 2\operatorname{tr}\left((A\Sigma)^2\right)$. For the third centered moment, we have:

$$\begin{aligned} \mathbb{E}\left[(Y - \mathbb{E}[Y])^3\right] &= \mathbb{E}\left[Y^3\right] - 3\mathbb{E}\left[Y^2\right]\mathbb{E}[Y] + 2\mathbb{E}^3[Y] \\ &= (\operatorname{tr}(A\Sigma))^3 + 6\operatorname{tr}(A\Sigma)\operatorname{tr}\left((A\Sigma)^2\right) + \\ & \quad 8\operatorname{tr}\left((A\Sigma)^3\right) - 3(\operatorname{tr}(A\Sigma))^3 - \\ & \quad 6\operatorname{tr}\left((A\Sigma)^2\right)\operatorname{tr}(A\Sigma) + 2(\operatorname{tr}(A\Sigma))^3 \\ &= 8\operatorname{tr}\left((A\Sigma)^3\right) \end{aligned}$$

The skewness is then equal to:

$$\begin{aligned}\gamma_1(Y) &= \frac{8\operatorname{tr}\left((A\Sigma)^3\right)}{\left(2\operatorname{tr}\left((A\Sigma)^2\right)\right)^{3/2}} \\ &= \frac{2\sqrt{2}\operatorname{tr}\left((A\Sigma)^3\right)}{\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^{3/2}}\end{aligned}$$

We obtain for the fourth centered moment:

$$\begin{aligned}\mathbb{E}\left[(Y-\mathbb{E}[Y])^4\right] &= \mathbb{E}\left[Y^4\right]-4\mathbb{E}\left[Y^3\right]\mathbb{E}[Y]+6\mathbb{E}\left[Y^2\right]\mathbb{E}^2[Y]- \\ &\quad 3\mathbb{E}^4[Y] \\ &= (\operatorname{tr}(A\Sigma))^4+32\operatorname{tr}(A\Sigma)\operatorname{tr}\left((A\Sigma)^3\right)+ \\ &\quad 12\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^2+48\operatorname{tr}\left((A\Sigma)^4\right) \\ &\quad 12(\operatorname{tr}(A\Sigma))^2\operatorname{tr}\left((A\Sigma)^2\right)-4(\operatorname{tr}(A\Sigma))^4- \\ &\quad 24(\operatorname{tr}(A\Sigma))^2\operatorname{tr}\left((A\Sigma)^2\right)- \\ &\quad 32\operatorname{tr}\left((A\Sigma)^3\right)\operatorname{tr}(A\Sigma)+6(\operatorname{tr}(A\Sigma))^4+ \\ &\quad 12\operatorname{tr}\left((A\Sigma)^2\right)(\operatorname{tr}(A\Sigma))^2-3(\operatorname{tr}(A\Sigma))^4 \\ &= 12\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^2+48\operatorname{tr}\left((A\Sigma)^4\right)\end{aligned}$$

It follows that the excess kurtosis is:

$$\begin{aligned}\gamma_2(Y) &= \frac{12\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^2+48\operatorname{tr}\left((A\Sigma)^4\right)}{\left(2\operatorname{tr}\left((A\Sigma)^2\right)\right)^2}-3 \\ &= \frac{12\operatorname{tr}\left((A\Sigma)^4\right)}{\left(\operatorname{tr}\left((A\Sigma)^2\right)\right)^2}\end{aligned}$$

4. We have:

$$\Pi(w)=w^\top\left(\boldsymbol{\mathcal{C}}_{t+h}-\boldsymbol{\mathcal{C}}_t\right)$$

where $\boldsymbol{\mathcal{C}}_t$ is the vector of option prices.

(a) The expression of the P&L is:

$$\begin{aligned}\Pi(w) &\approx w^\top\left(\boldsymbol{\Delta}_t\circ(S_{t+h}-S_t)\right) \\ &= w^\top\left((\boldsymbol{\Delta}_t\circ S_t)\circ R_{t+h}\right) \\ &= \tilde{\boldsymbol{\Delta}}_t^\top R_{t+h}\end{aligned}$$

where $\tilde{\boldsymbol{\Delta}}_t$ is the vector of delta exposures in dollars:

$$\tilde{\boldsymbol{\Delta}}_{i,t}=w_i\boldsymbol{\Delta}_{i,t}S_{i,t}$$

Because $R_{t+h} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, it follows that $\Pi \sim \mathcal{N}\left(0, \tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t\right)$. We deduce that the Gaussian value-at-risk is:

$$\mathrm{VaR}_\alpha(w; h) = \Phi^{-1}(\alpha)\sqrt{\tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t}$$

The risk contribution of option i is then equal to:

$$\begin{aligned}
\mathcal{RC}_i &= w_i \frac{\Phi^{-1}(\alpha)\left(\Sigma\tilde{\boldsymbol{\Delta}}_t\right)_i \boldsymbol{\Delta}_{i,t} S_{i,t}}{\sqrt{\tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t}} \\
&= \Phi^{-1}(\alpha)\frac{\tilde{\boldsymbol{\Delta}}_{i,t} \cdot \left(\Sigma\tilde{\boldsymbol{\Delta}}_t\right)_i}{\sqrt{\tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t}}
\end{aligned}$$

(b) In the case of the delta-gamma approximation, we obtain:

$$\begin{aligned}
\Pi(w) &\approx w^\top\left(\boldsymbol{\Delta}_t \circ (S_{t+h} - S_t)\right) + \\
&\quad \frac{1}{2} w^\top \left(\boldsymbol{\Gamma}_t \circ (S_{t+h} - S_t) \circ (S_{t+h} - S_t)^\top\right) w \\
&= \tilde{\boldsymbol{\Delta}}_t^\top R_{t+h} + \frac{1}{2} R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h}
\end{aligned}$$

where $\tilde{\boldsymbol{\Gamma}}_t$ is the matrix of gamma exposures in dollars:

$$\tilde{\boldsymbol{\Gamma}}_{i,j,t} = w_i w_j \boldsymbol{\Gamma}_{i,j,t} S_{i,t} S_{j,t}$$

We deduce that:

$$\begin{aligned}
\mathbb{E}[\Pi] &= \mathbb{E}\left[\tilde{\boldsymbol{\Delta}}_t^\top R_{t+h} + \frac{1}{2} R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h}\right] \\
&= \frac{1}{2}\mathbb{E}\left[R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h}\right] \\
&= \frac{1}{2}\operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)
\end{aligned}$$

and:

$$\begin{aligned}
\operatorname{var}(\Pi) &= \mathbb{E}\left[(\Pi - \mathbb{E}[\Pi])^2\right] \\
&= \mathbb{E}\left[\left(\tilde{\boldsymbol{\Delta}}_t^\top R_{t+h} + \frac{1}{2} R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h} - \frac{1}{2}\operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(\tilde{\boldsymbol{\Delta}}_t^\top R_{t+h}\right)^2\right] + \frac{1}{4}\mathbb{E}\left[\left(R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h} - \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)\right)^2\right] + \\
&\quad \mathbb{E}\left[\left(\tilde{\boldsymbol{\Delta}}_t^\top R_{t+h}\right)\left(R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h} - \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)\right)\right] \\
&= \mathbb{E}\left[\left(\tilde{\boldsymbol{\Delta}}_t^\top R_{t+h}\right)^2\right] + \frac{1}{4}\operatorname{var}\left(R_{t+h}^\top \tilde{\boldsymbol{\Gamma}}_t R_{t+h}\right) \\
&= \tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t + \frac{1}{2}\operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)
\end{aligned}$$

Therefore, the Gaussian approximation of the P&L is:

$$\Pi(w) \sim \mathcal{N}\left(\frac{1}{2}\operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right), \tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t + \frac{1}{2}\operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)\right)$$

We deduce that the Gaussian value-at-risk is:

$$\mathrm{VaR}_\alpha(w; h) = -\frac{1}{2}\operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right) + \Phi^{-1}(\alpha)\sqrt{\tilde{\boldsymbol{\Delta}}_t^\top \Sigma \tilde{\boldsymbol{\Delta}}_t + \frac{1}{2}\operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)}$$

(c) If the portfolio is delta neutral, $\boldsymbol{\Delta}_t$ is equal to zero and we have:

$$\Pi \simeq \frac{1}{2} R_{t+h}^{\top} \tilde{\boldsymbol{\Gamma}}_t R_{t+h}$$

Let $L = -\Pi$ be the loss. Using the formulas of Question 3(b), we obtain:

$$\mu(L) = -\frac{1}{2} \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)$$

$$\sigma(L) = \sqrt{\frac{1}{2} \operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)}$$

$$\gamma_1(L) = -\frac{2\sqrt{2} \operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^3\right)}{\left(\operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)\right)^{3/2}}$$

$$\gamma_2(L) = \frac{12 \operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^4\right)}{\left(\operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)\right)^{2}}$$

We have all the statistics to compute the Cornish-Fisher value-at-risk.

(d) We notice that the previous formulas obtained in the multivariate case are perfectly coherent with those obtained in the univariate case. When the portfolio is not delta neutral, we could then postulate that the skewness is[13]:

$$\gamma_1(L) = -\frac{6\sqrt{2}\tilde{\boldsymbol{\Delta}}_t^{\top} \Sigma \boldsymbol{\Gamma}_t \Sigma \tilde{\boldsymbol{\Delta}}_t + 2\sqrt{2} \operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^3\right)}{\left(2\tilde{\boldsymbol{\Delta}}_t^{\top} \Sigma \tilde{\boldsymbol{\Delta}}_t + \operatorname{tr}\left(\left(\tilde{\boldsymbol{\Gamma}}_t \Sigma\right)^2\right)\right)^{3/2}}$$

In fact, it is the formula obtained by Britten-Jones and Schaeffer (1999).

5. (a) Using the numerical values, we obtain $\mu(L) = -78.65$, $\sigma(L) = 88.04$, $\gamma_1(L) = -2.5583$ and $\gamma_2(L) = 10.2255$. The value-at-risk is then equal to 0 for the delta approximation, 126.16 for the delta-gamma approximation and -45.85 for the Cornish-Fisher approximation. We notice that we obtain an absurd result in the last case, because the distribution is far from the Gaussian distribution (high skewness and kurtosis). If we consider a smaller order expansion:

$$\mathfrak{z}(\alpha; \gamma_1, \gamma_2) = z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\gamma_1 + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\gamma_2$$

the value-at-risk is equal to 171.01.

(b) In this case, we obtain 126.24 for the delta approximation, 161.94 for the delta-gamma approximation and -207.84 for the Cornish-Fisher approximation. For the delta approximation, the risk decomposition is:

Option	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
1	50.00	0.86	42.87	33.96%
2	20.00	0.77	15.38	12.19%
3	30.00	2.27	67.98	53.85%
$\mathcal{R}(w)$			126.24	

[13] You may easily verify that we obtained this formula in the case $n = 2$ by developing the different polynomials.

For the delta-gamma approximation, we have:

Option	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^{\star}$
1	50.00	4.06	202.92	125.31%
2	20.00	1.18	23.62	14.59%
3	30.00	1.04	31.10	19.21%
$\mathcal{R}(w)$			161.94	

We notice that the delta-gamma approximation does not satisfy the Euler decomposition.

2.4.9 Risk decomposition of the expected shortfall

1. We have:

$$L(w) = -R(w) = -w^\top R$$

It follows that:

$$L(w) \sim \mathcal{N}\left(-\mu(w), \sigma^2(w)\right)$$

with $\mu(w) = w^\top \mu$ and $\sigma(w) = \sqrt{w^\top \Sigma w}$.

2. The expected shortfall $\mathrm{ES}_\alpha(w;h)$ is the average of value-at-risks at level α and higher:

$$\mathrm{ES}_\alpha(w;h) = \mathbb{E}\left[L \mid L \geq \mathrm{VaR}_\alpha(w;h)\right]$$

We know that the value-at-risk is:

$$\mathrm{VaR}_\alpha(w;h) = -w^\top \mu + \Phi^{-1}(\alpha)\sqrt{w^\top \Sigma w}$$

We deduce that:

$$\mathrm{ES}_\alpha(w;h) = \frac{1}{1-\alpha}\int_{x^-}^{\infty} \frac{x}{\sigma(w)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x+\mu(w)}{\sigma(w)}\right)^2\right) \mathrm{d}x$$

where $x^- = -\mu(w) + \Phi^{-1}(\alpha)\sigma(w)$. With the change of variable $t = \sigma(w)^{-1}(x+\mu(w))$, we obtain:

$$\begin{aligned}
\mathrm{ES}_\alpha(w;h) &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} \left(-\mu(w) + \sigma(w)\,t\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t \\
&= -\frac{\mu(w)}{1-\alpha}\left[\Phi(t)\right]_{\Phi^{-1}(\alpha)}^{\infty} + \\
&\quad \frac{\sigma(w)}{(1-\alpha)\sqrt{2\pi}}\int_{\Phi^{-1}(\alpha)}^{\infty} t\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t \\
&= -\mu(w) + \frac{\sigma(w)}{(1-\alpha)\sqrt{2\pi}}\left[-\exp\left(-\frac{1}{2}t^2\right)\right]_{\Phi^{-1}(\alpha)}^{\infty} \\
&= -\mu(w) + \frac{\sigma(w)}{(1-\alpha)\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\Phi^{-1}(\alpha)\right]^2\right)
\end{aligned}$$

The expected shortfall of portfolio w is then:

$$\mathrm{ES}_\alpha(w;h) = -w^\top \mu + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)}\sqrt{w^\top \Sigma w}$$

3. The vector of marginal risk is defined as follows:

$$\begin{aligned}\mathcal{MR} &= \frac{\partial \operatorname{ES}_\alpha (w;h)}{\partial w} \\ &= -\mu + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \frac{\Sigma w}{\sqrt{w^\top \Sigma w}}\end{aligned}$$

We deduce that the risk contribution $\mathcal{RC}_i$ of the asset i is:

$$\begin{aligned}\mathcal{RC}_i &= w_i \times \mathcal{MR}_i \\ &= -w_i \mu_i + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \frac{w_i \times (\Sigma w)_i}{\sqrt{w^\top \Sigma w}}\end{aligned}$$

It follows that:

$$\begin{aligned}\sum_{i=1}^n \mathcal{RC}_i &= \sum_{i=1}^n -w_i \mu_i + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \frac{w_i \times (\Sigma w)_i}{\sqrt{x^\top \Sigma x}} \\ &= -w^\top \mu + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \frac{w^\top (\Sigma w)}{\sqrt{w^\top \Sigma w}} \\ &= \operatorname{ES}_\alpha (w;h)\end{aligned}$$

The expected shortfall then verifies the Euler allocation principle.

4. We have:

$$L(w) = -\sum_{i=1}^n w_i R_i = \sum_{i=1}^n L_i$$

with $L_i = -w_i R_i$. We know that:

$$\begin{aligned}\mathcal{RC}_i &= \mathbb{E}\left[L_i \mid L \geq \operatorname{VaR}_\alpha (w;h)\right] \\ &= \frac{\mathbb{E}\left[L_i \cdot \mathbb{1}\left\{L \geq \operatorname{VaR}_\alpha (w;h)\right\}\right]}{\mathbb{E}\left[\mathbb{1}\left\{L \geq \operatorname{VaR}_\alpha (w;h)\right\}\right]} \\ &= \frac{\mathbb{E}\left[L_i \cdot \mathbb{1}\left\{L \geq \operatorname{VaR}_\alpha (w;h)\right\}\right]}{1-\alpha}\end{aligned}$$

We deduce that:

$$\mathcal{RC}_i = -\frac{w_i}{1-\alpha} \mathbb{E}\left[R_i \cdot \mathbb{1}\left\{R(w) \leq -\operatorname{VaR}_\alpha (w;h)\right\}\right]$$

We know that the random vector $(R, R(w))$ has a multivariate normal distribution:

$$\begin{pmatrix} R \\ R(w) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ w^\top \mu \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma w \\ w^\top \Sigma & w^\top \Sigma w \end{pmatrix}\right)$$

We deduce that:

$$\begin{pmatrix} R_i \\ R(w) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_i \\ w^\top \mu \end{pmatrix}, \begin{pmatrix} \Sigma_{i,i} & (\Sigma w)_i \\ (\Sigma w)_i & w^\top \Sigma w \end{pmatrix}\right)$$

Let $I = \mathbb{E}\left[R_i \cdot \mathbb{1}\left\{R(w) \leq -\operatorname{VaR}_\alpha (w;h)\right\}\right]$. We note f the density function of the random vector $(R_i, R(w))$ and $\rho = \Sigma_{i,i}^{-1/2} \left(w^\top \Sigma w\right)^{-1/2} (\Sigma w)_i$ the correlation between R_i and $R(w)$. It follows that:

$$\begin{aligned}I &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbb{1}\left\{s \leq -\operatorname{VaR}_\alpha (w;h)\right\} \cdot r f(r,s) \,\mathrm{d}r \,\mathrm{d}s \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{-\operatorname{VaR}_\alpha (w)} r f(r,s) \,\mathrm{d}r \,\mathrm{d}s\end{aligned}$$

Let $t = (r - \mu_i) / \sqrt{\Sigma_{i,i}}$ and $u = \left(s - w^\top \mu\right) / \sqrt{w^\top \Sigma w}$. We deduce that[14]:

$$I = \int_{-\infty}^{+\infty} \int_{-\infty}^{\Phi^{-1}(1-\alpha)} \frac{\mu_i + \sqrt{\Sigma_{i,i}} t}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{t^2 + u^2 - 2\rho t u}{2\left(1-\rho^2\right)}\right) \mathrm{d}t \, \mathrm{d}u$$

By considering the change of variables $(t, u) = \varphi(t, v)$ such that $u = \rho t + \sqrt{1-\rho^2} v$, we obtain[15]:

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} \int_{-\infty}^{g(t)} \frac{\mu_i + \sqrt{\Sigma_{i,i}} t}{2\pi} \exp\left(-\frac{t^2 + v^2}{2}\right) \mathrm{d}t \, \mathrm{d}v \\ &= \mu_i \int_{-\infty}^{+\infty} \int_{-\infty}^{g(t)} \frac{1}{2\pi} \exp\left(-\frac{t^2 + v^2}{2}\right) \mathrm{d}t \, \mathrm{d}v + \\ &\quad \sqrt{\Sigma_{i,i}} \int_{-\infty}^{+\infty} \int_{-\infty}^{g(t)} \frac{t}{2\pi} \exp\left(-\frac{t^2 + v^2}{2}\right) \mathrm{d}t \, \mathrm{d}v + \\ &= \mu_i I_1 + \sqrt{\Sigma_{i,i}} I_2 \end{aligned}$$

where the bound $g(t)$ is defined as follows:

$$g(t) = \frac{\Phi^{-1}(1-\alpha) - \rho t}{\sqrt{1-\rho^2}}$$

For the first integral, we have[16]:

$$\begin{aligned} I_1 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \left(\int_{-\infty}^{g(t)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) \mathrm{d}v\right) \mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\Phi^{-1}(1-\alpha) - \rho t}{\sqrt{1-\rho^2}}\right) \phi(t) \, \mathrm{d}t \\ &= 1 - \alpha \end{aligned}$$

The computation of the second integral I_2 is a little bit more tedious. Integration by parts with the derivative function $t\phi(t)$ gives:

$$\begin{aligned} I_2 &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\Phi^{-1}(1-\alpha) - \rho t}{\sqrt{1-\rho^2}}\right) t\phi(t) \, \mathrm{d}t \\ &= -\frac{\rho}{\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \phi\left(\frac{\Phi^{-1}(1-\alpha) - \rho t}{\sqrt{1-\rho^2}}\right) \phi(t) \, \mathrm{d}t \\ &= -\frac{\rho}{\sqrt{1-\rho^2}} \phi\left(\Phi^{-1}(1-\alpha)\right) \int_{-\infty}^{+\infty} \phi\left(\frac{t - \rho\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}}\right) \mathrm{d}t \\ &= -\rho\phi\left(\Phi^{-1}(1-\alpha)\right) \end{aligned}$$

[14]Because we have $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$.

[15]We use the fact that $\mathrm{d}t \, \mathrm{d}v = \sqrt{1-\rho^2} \, \mathrm{d}t \, \mathrm{d}u$ because the determinant of the Jacobian matrix containing the partial derivatives $D\varphi$ is $\sqrt{1-\rho^2}$.

[16]We use the fact that:

$$\mathbb{E}\left[\Phi\left(\frac{\Phi^{-1}(p) - \rho T}{\sqrt{1-\rho^2}}\right)\right] = p$$

where $T \sim \mathcal{N}(0, 1)$.

We could then deduce the value of I:

$$\begin{aligned} I &= \mu_i (1-\alpha) - \rho\sqrt{\Sigma_{i,i}}\phi\left(\Phi^{-1}(1-\alpha)\right) \\ &= \mu_i (1-\alpha) - \frac{(\Sigma w)_i}{\sqrt{w^\top \Sigma w}}\phi\left(\Phi^{-1}(\alpha)\right) \end{aligned}$$

We finally obtain that:

$$\mathcal{RC}_i = -w_i\mu_i + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \frac{w_i \times (\Sigma w)_i}{\sqrt{w^\top \Sigma w}}$$

We obtain the same expression as found in Question 3. Nevertheless, the conditional representation is more general than the Gaussian formula, because it is valid for any probability distribution.

2.4.10 Expected shortfall of an equity portfolio

1. We have:

$$\begin{aligned} \Pi &= 4\left(P_{A,t+h} - P_{A,t}\right) + 3\left(P_{B,t+h} - P_{B,t}\right) \\ &= 4P_{A,t}R_{A,t+h} + 3P_{B,t}R_{B,t+h} \\ &= 400 \times R_{A,t+h} + 600 \times R_{B,t+h} \end{aligned}$$

where $R_{A,t+h}$ and $R_{B,t+h}$ are the stock returns for the period $[t, t+h]$. We deduce that the variance of the P&L is:

$$\begin{aligned} \sigma^2(\Pi) &= 400 \times (25\%)^2 + 600 \times (20\%)^2 + \\ &\quad 2 \times 400 \times 600 \times (-20\%) \times 25\% \times 20\% \\ &= 19\,600 \end{aligned}$$

We deduce that $\sigma(\Pi) = \$140$. We know that the one-year expected shortfall is a linear function of the volatility:

$$\begin{aligned} \mathrm{ES}_\alpha(w; \text{one year}) &= \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \times \sigma(\Pi) \\ &= 2.34 \times 140 \\ &= \$327.60 \end{aligned}$$

The 10-day expected shortfall is then equal to \$64.25:

$$\begin{aligned} \mathrm{ES}_\alpha(w; \text{ten days}) &= \sqrt{\frac{10}{260}} \times 327.60 \\ &= \$64.25 \end{aligned}$$

2. We have:

$$\Pi_s = 400 \times R_{A,s} + 600 \times R_{B,s}$$

We deduce that the value Π_s of the daily P&L for each scenario s is:

s	1	2	3	4	5	6	7	8
Π_s	−36	−10	−24	−26	−12	−30	−14	−16
$\Pi_{s:250}$	−36	−30	−26	−24	−16	−14	−12	−10

The value-at-risk at the 97.5% confidence level correspond to the 6.25$^{\text{th}}$ order statistic[17]. We deduce that the historical expected shortfall for a one-day time horizon is equal to:

$$\begin{aligned}
\mathrm{ES}_{\alpha}(w;\text{one day}) &= -\mathbb{E}\left[\Pi \mid \Pi \leq -\mathrm{VaR}_{\alpha}(\Pi)\right] \\
&= -\frac{1}{6}\sum_{s=1}^{6}\Pi_{s:250} \\
&= \frac{1}{6}(36+30+26+24+16+14) \\
&= 24.33
\end{aligned}$$

By considering the square-root-of-time rule, it follows that the 10-day expected shortfall is equal to \$76.95.

2.4.11 Risk measure of a long/short portfolio

We have:

$$\begin{aligned}
\Pi_{t,t+h} &= 2\left(P_{A,t+h}-P_{A,t}\right)-5\left(P_{B,t+h}-P_{B,t}\right) \\
&= 2P_{A,t}R_{A,t+h}-5P_{B,t}R_{B,t+h} \\
&= 100\times\left(R_{A,t+h}-R_{B,t+h}\right)
\end{aligned}$$

where $R_{A,t+h}$ and $R_{B,t+h}$ are the stock returns for the period $[t,t+h]$.

1. We deduce that the (annualized) variance of the P&L is:

$$\begin{aligned}
\sigma^2\left(\Pi_{t,t+260}\right) &= 100^2\times(25\%)^2+100^2\times(20\%)^2- \\
&\quad 2\times 100^2\times 12.5\%\times 25\%\times 20\% \\
&= 900
\end{aligned}$$

We have $\sigma\left(\Pi_{t,t+260}\right)=\$30$. It follows that the 10-day standard deviation is equal to:

$$\begin{aligned}
\sigma\left(\Pi_{t,t+10}\right) &= \sqrt{\frac{10}{260}}\times\sigma\left(\Pi_{t,t+260}\right) \\
&= \$5.883
\end{aligned}$$

(a) We obtain:

$$\begin{aligned}
\mathrm{VaR}_{99\%}(w;\text{ten days}) &= \Phi^{-1}(99\%)\times\sigma\left(\Pi_{t,t+10}\right) \\
&= \$13.69
\end{aligned}$$

(b) We have:

$$\mathrm{ES}_{\alpha}(w;\text{ten days})=\frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}\times\sigma\left(\Pi_{t,t+10}\right)$$

and:

$$\begin{aligned}
\frac{\phi\left(\Phi^{-1}(97.5\%)\right)}{1-97.5\%} &= 2.3378 \\
&\approx 2.34
\end{aligned}$$

The 10-day expected shortfall is then equal to \$13.75.

TABLE 2.1: Order statistic $\Pi_{s:250}$ of the daily P&L

s	1	2	3	4	5	6	7	8	9	10
$\Pi_{s:250}$	−6.3	−6.0	−5.1	−4.8	−4.6	−4.5	−4.3	−4.3	−4.0	−3.9

2. Given the historical scenario s, the one-day simulated P&L is equal to:

$$\begin{aligned} \Pi_s &= 100 \times (R_{A,s} - R_{B,s}) \\ &= 100 \times D_s \end{aligned}$$

The order statistic $\Pi_{s:250}$ of the daily P&L is given in Table 2.1.

(a) We deduce that the one-day value-at-risk at the 99% confidence level corresponds to the 2.5^{nd} order statistic:

$$\text{VaR}_{99\%}(w; \text{one day}) = -\left(\frac{-6.0 - 5.1}{2}\right) = \$5.55$$

It follows that:

$$\text{VaR}_{99\%}(w; \text{ten days}) = \sqrt{10} \times \text{VaR}_{99\%}(w, \text{one day}) = \$17.55$$

(b) The value-at-risk at the 97.5% confidence level correspond to the 6.25^{th} order statistic. We deduce that the historical expected shortfall for a one-day time horizon is equal to:

$$\begin{aligned} \text{ES}_{97.5\%}(w; \text{one day}) &= -\frac{1}{6}\sum_{s=1}^{6} \Pi_{s:250} \\ &= \frac{1}{6}(6.3 + 6.0 + 5.1 + 4.8 + 4.6 + 4.5) \\ &= \$5.22 \end{aligned}$$

By considering the square-root-of-time rule, it follows that the 10-day expected shortfall is equal to \$16.50.

(c) In Basel II, the capital charge is equal to:

$$\begin{aligned} \mathcal{K} &= 3 \times \text{VaR}_{99\%}(w; \text{ten days}) \\ &= \$52.65 \end{aligned}$$

In Basel 2.5, the capital charge becomes:

$$\begin{aligned} \mathcal{K} &= 3 \times \text{VaR}_{99\%}(w; \text{ten days}) + 3 \times \text{SVaR}_{99\%}(w; \text{ten days}) \\ &= \$157.96 \end{aligned}$$

where SVaR is the stressed value-at-risk. In Basel III, we obtain:

$$\begin{aligned} \mathcal{K} &= 2 \times \text{SES}_{99\%}(w; \text{ten days}) \\ &= \$65.99 \end{aligned}$$

where SES is the stressed expected shortfall.

[17] We have $2.5\% \times 250 = 6.25$.

2.4.12 Kernel estimation of the expected shortfall

1. We have:

$$\begin{aligned}\mathbb{E}\left[X \cdot \mathbb{1}\left\{X \leq x\right\}\right] &= \int_{-\infty}^{\infty} t \cdot \mathbb{1}\left\{t \leq x\right\} \cdot \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}\left(\frac{t - x_i}{\boldsymbol{h}}\right) \,\mathrm{d}t \\ &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{x} \frac{t}{h} \mathcal{K}\left(\frac{t - x_i}{\boldsymbol{h}}\right) \,\mathrm{d}t\end{aligned}$$

We consider the change of variable $u = \boldsymbol{h}^{-1}\left(t - x_i\right)$:

$$\begin{aligned}\int_{-\infty}^{x} \frac{t}{h} \mathcal{K}\left(\frac{t - x_i}{\boldsymbol{h}}\right) \,\mathrm{d}t &= \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \left(x_i + \boldsymbol{h}\, u\right) \mathcal{K}\left(u\right) \,\mathrm{d}u \\ &= \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} x_i \mathcal{K}\left(u\right) \,\mathrm{d}u + \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \boldsymbol{h}\, u \mathcal{K}\left(u\right) \,\mathrm{d}u\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[X \cdot \mathbb{1}\left\{X \leq x\right\}\right] = &\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} x_i \mathcal{K}\left(u\right) \,\mathrm{d}u + \\ &\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \boldsymbol{h}\, u \mathcal{K}\left(u\right) \,\mathrm{d}u\end{aligned}$$

2. We have:

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} x_i \mathcal{K}\left(u\right) \,\mathrm{d}u &= \frac{1}{n} \sum_{i=1}^{n} x_i \left(\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \mathcal{K}\left(u\right) \,\mathrm{d}u\right) \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i \mathcal{I}\left(\frac{x - x_i}{\boldsymbol{h}}\right)\end{aligned}$$

3. Since we have:

$$\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} u \mathcal{K}\left(u\right) \,\mathrm{d}u = \Big[u \mathcal{I}\left(u\right)\Big]_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} - \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \mathcal{I}\left(u\right) \,\mathrm{d}u$$

we deduce that:

$$\left|\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} u \mathcal{K}\left(u\right) \,\mathrm{d}u\right| \leq \left(\frac{x - x_i}{\boldsymbol{h}}\right) \cdot \mathcal{I}\left(\frac{x - x_i}{\boldsymbol{h}}\right)$$

It follows that:

$$\begin{aligned}\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \boldsymbol{h} u \mathcal{K}\left(u\right) \,\mathrm{d}u &= \boldsymbol{h} \int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} u \mathcal{K}\left(u\right) \,\mathrm{d}u \\ &\rightarrow 0 \text{ when } \boldsymbol{h} \rightarrow 0\end{aligned}$$

4. Finally, we obtain the result:

$$\mathbb{E}\left[X \cdot \mathbb{1}\left\{X \leq x\right\}\right] \approx \frac{1}{n} \sum_{i=1}^{n} x_i \mathcal{I}\left(\frac{x - x_i}{\boldsymbol{h}}\right)$$

We conclude that:

$$\begin{aligned}
\mathrm{ES}_{\alpha}(w;h) &= \frac{1}{(1-\alpha)}\mathbb{E}\left[L(w)\cdot\mathbb{1}\left\{L(w)\geq \mathrm{VaR}_{\alpha}(w;h)\right\}\right] \\
&= -\frac{1}{(1-\alpha)}\mathbb{E}\left[\Pi(w)\cdot\mathbb{1}\left\{\Pi(w)\leq -\mathrm{VaR}_{\alpha}(w;h)\right\}\right] \\
&\approx -\frac{1}{(1-\alpha)}\left(\frac{1}{n_S}\sum_{s=1}^{n_S}\Pi_s\mathcal{I}\left(\frac{-\mathrm{VaR}_{\alpha}(w;h)-\Pi_s}{\boldsymbol{h}}\right)\right)
\end{aligned}$$

because we have $\Pi(w) = -L(w)$.

Chapter 3

Credit Risk

3.4.1 Single and multi-name credit default swaps

1. We have $\mathbf{F}(t) = 1 - e^{-\lambda t}$, $\mathbf{S}(t) = e^{-\lambda t}$ and $f(t) = \lambda e^{-\lambda t}$. We know that $\mathbf{S}(\boldsymbol{\tau}) \sim \mathcal{U}_{[0,1]}$. Indeed, we have:

$$\begin{aligned}
\Pr\{U \leq u\} &= \Pr\{\mathbf{S}(\boldsymbol{\tau}) \leq u\} \\
&= \Pr\left\{\boldsymbol{\tau} \geq \mathbf{S}^{-1}(u)\right\} \\
&= \mathbf{S}\left(\mathbf{S}^{-1}(u)\right) \\
&= u
\end{aligned}$$

It follows that $\boldsymbol{\tau} = \mathbf{S}^{-1}(U)$ with $U \sim \mathcal{U}_{[0,1]}$. Let u be a uniform random variate. Simulating τ is then equivalent to transform u into t:

$$t = -\frac{1}{\lambda}\ln u$$

2. (a) The premium leg is paid quarterly. The coupon payment is then equal to:

$$\begin{aligned}
\mathcal{PL}(t_m) &= \Delta t_m \times \boldsymbol{s} \times N \\
&= \frac{1}{4} \times 150 \times 10^{-4} \times 10^{6} \\
&= \$3\,750
\end{aligned}$$

In case of default, the default leg paid by protection seller is equal to:

$$\begin{aligned}
\mathcal{DL} &= (1 - \boldsymbol{\mathcal{R}}) \times N \\
&= (1 - 40\%) \times 10^{6} \\
&= \$600\,000
\end{aligned}$$

The corresponding cash flow chart is given in Figure 3.1. If the reference entity does not default, the P&L of the protection seller is the sum of premium interests:

$$\Pi^{\text{seller}} = 8 \times 3\,750 = \$30\,000$$

If the reference entity defaults in one year and two months, the P&L of the protection buyer is[1]:

$$\begin{aligned}
\Pi^{\text{buyer}} &= (1 - \boldsymbol{\mathcal{R}}) \times N - \sum_{t_m < \boldsymbol{\tau}} \Delta t_m \times \boldsymbol{s} \times N \\
&= (1 - 40\%) \times 10^{6} - \left(4 + \frac{2}{3}\right) \times 3\,750 \\
&= \$582\,500
\end{aligned}$$

[1] We include the accrued premium.

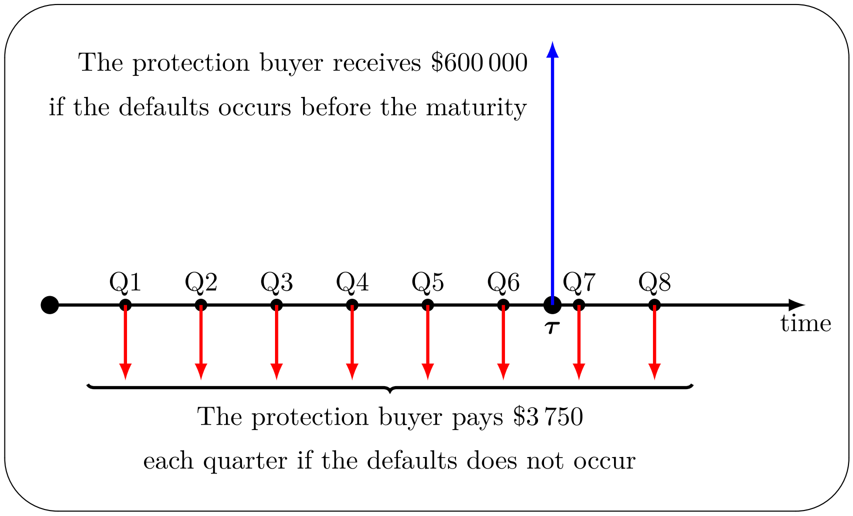

FIGURE 3.1: Cash flow chart of the CDS contract

(b) Using the credit triangle relationship, we have:

$$\mathcal{S} \simeq (1 - \mathcal{R}) \times \lambda$$

We deduce that[2]:

$$\begin{aligned} \mathrm{PD} &\simeq \lambda \\ &\simeq \frac{\mathcal{S}}{1 - \mathcal{R}} \\ &= \frac{150 \times 10^{-4}}{1 - 40\%} \\ &= 2.50\% \end{aligned}$$

(c) We denote by $\mathcal{S}'$ the new CDS spread. The default probability becomes:

$$\begin{aligned} \mathrm{PD} &= \frac{\mathcal{S}'}{1 - \mathcal{R}} \\ &= \frac{450 \times 10^{-4}}{1 - 40\%} \\ &= 7.50\% \end{aligned}$$

The protection buyer is short credit and benefits from the increase of the default probability. His mark-to-market is therefore equal to:

$$\begin{aligned} \Pi^{\mathrm{buyer}} &= N \times (\mathcal{S}' - \mathcal{S}) \times \mathrm{RPV}_{01} \\ &= 10^6 \times (450 - 150) \times 10^{-4} \times 1.189 \\ &= \$35\,671 \end{aligned}$$

[2] We recall that the one-year default probability is approximately equal to λ:

$$\begin{aligned} \mathrm{PD} &= 1 - \mathbf{S}(1) \\ &= 1 - e^{-\lambda} \\ &\simeq 1 - (1 - \lambda) \\ &\simeq \lambda \end{aligned}$$

The offsetting mechanism is then the following: the protection buyer B transfers the agreement to C, who becomes the new protection buyer; C continues to pay a premium of 150 bps to the protection seller A; in return, C pays a cash adjustment of \$35 671 to B.

3. (a) For a given date t, the credit curve is the relationship between the maturity T and the spread $\mathcal{s}_t(T)$. The credit curve of the reference entity #1 is almost flat. For the entity #2, the spread is very high in the short-term, meaning that there is a significative probability that the entity defaults. However, if the entity survive, the market anticipates that it will improve its financial position in the long-run. This explains that the credit curve #2 is decreasing. For reference entity #3, we obtain opposite conclusions. The company is actually very strong, but there are some uncertainties in the future[3]. The credit curve is then increasing.

   (b) If we consider a standard recovery rate (40%), the implied default probability is 2.50% for #1, 10% for #2 and 1.33% for #3. We can consider a short credit position in #2. In this case, we sell the 5Y protection on #2 because the model tells us that the market default probability is over-estimated. In place of this directional bet, we could consider a relative value strategy: selling the 5Y protection on #2 and buying the 5Y protection on #3.

4. (a) Let $\boldsymbol{\tau}_{k:n}$ be the k^{th} default among the basket. FtD, StD and LtD are three CDS products, whose credit event is related to the default times $\boldsymbol{\tau}_{1:n}$, $\boldsymbol{\tau}_{2:n}$ and $\boldsymbol{\tau}_{n:n}$.

   (b) The default correlation ρ measures the dependence between two default times $\boldsymbol{\tau}_i$ and $\boldsymbol{\tau}_j$. The spread of the FtD (resp. LtD) is a decreasing (resp. increasing) function with respect to ρ.

   (c) To fully hedge the credit portfolio of the 3 entities, we can buy the 3 CDS. Another solution is to buy the FtD plus the StD and the LtD (or the third-to-default). Because these two hedging strategies are equivalent, we deduce that:

$$\mathcal{s}_1^{\text{CDS}} + \mathcal{s}_2^{\text{CDS}} + \mathcal{s}_3^{\text{CDS}} = \mathcal{s}^{\text{FtD}} + \mathcal{s}^{\text{StD}} + \mathcal{s}^{\text{LtD}}$$

   (d) We notice that the default correlation does not affect the value of the CDS basket, but only the price distribution between FtD, StD and LtD. We obtain a similar result for CDO[4]. In the case of the subprime crisis, all the CDO tranches have suffered, meaning that the price of the underlying basket has dropped. The reasons were the underestimation of default probabilities.

3.4.2 Risk contribution in the Basel II model

1. (a) The portfolio loss L follows a Gaussian probability distribution:

$$L(w) \sim \mathcal{N}\left(0, \sqrt{w^\top \Sigma w}\right)$$

   We deduce that:

$$\mathrm{VaR}_\alpha(w) = \Phi^{-1}(\alpha)\sqrt{w^\top \Sigma w}$$

[3] An example is a company whose has a monopoly because of a strong technology, but faces a hard competition because technology is evolving fast in its domain (e.g. Blackberry at the end of 2000s).

[4] The junior, mezzanine and senior tranches can be viewed as FtD, StD and LtD.

(b) We have:

$$\begin{aligned}\frac{\partial\,\mathrm{VaR}_\alpha(w)}{\partial\,w} &= \frac{\partial}{\partial\,w}\left(\Phi^{-1}(\alpha)\left(w^\top\Sigma w\right)^{\frac{1}{2}}\right)\\ &= \Phi^{-1}(\alpha)\frac{1}{2}\left(w^\top\Sigma w\right)^{-\frac{1}{2}}(2\Sigma w)\\ &= \Phi^{-1}(\alpha)\frac{\Sigma w}{\sqrt{w^\top\Sigma w}}\end{aligned}$$

The marginal value-at-risk of the i^{th} credit is then:

$$\mathcal{MR}_i = \frac{\partial\,\mathrm{VaR}_\alpha(w)}{\partial\,w_i} = \Phi^{-1}(\alpha)\frac{(\Sigma w)_i}{\sqrt{w^\top\Sigma w}}$$

The risk contribution of the i^{th} credit is the product of the exposure by the marginal risk:

$$\begin{aligned}\mathcal{RC}_i &= w_i\times\mathcal{MR}_i\\ &= \Phi^{-1}(\alpha)\frac{w_i\times(\Sigma w)_i}{\sqrt{x^\top\Sigma x}}\end{aligned}$$

(c) By construction, the random vector $(\varepsilon, L(w))$ is Gaussian with:

$$\left(\begin{array}{c}\varepsilon\\ L(w)\end{array}\right)\sim\mathcal{N}\left(\left(\begin{array}{c}\mathbf{0}\\ 0\end{array}\right),\left(\begin{array}{cc}\Sigma & \Sigma w\\ w^\top\Sigma & w^\top\Sigma w\end{array}\right)\right)$$

We deduce that the conditional distribution function of ε given that $L(w)=\ell$ is Gaussian and we have:

$$\mathbb{E}\left[\varepsilon\mid L(w)=\ell\right] = \mathbf{0}+\Sigma w\left(w^\top\Sigma w\right)^{-1}(\ell-0)$$

We finally obtain:

$$\begin{aligned}\mathbb{E}\left[\varepsilon\mid L(w)=\mathbf{F}^{-1}(\alpha)\right] &= \Sigma w\left(w^\top\Sigma w\right)^{-1}\Phi^{-1}(\alpha)\sqrt{w^\top\Sigma w}\\ &= \Phi^{-1}(\alpha)\frac{\Sigma w}{\sqrt{w^\top\Sigma w}}\\ &= \frac{\partial\,\mathrm{VaR}_\alpha(w)}{\partial\,w}\end{aligned}$$

The marginal VaR of the i^{th} credit is then equal to the conditional mean of the individual loss ε_i given that the portfolio loss is exactly equal to the value-at-risk.

2. (a) EAD_i is the exposure at default, LGD_i is the loss given default, $\boldsymbol{\tau}_i$ is the default time and T_i is the maturity of the credit i. We have:

$$\left\{\begin{array}{l}w_i=\mathrm{EAD}_i\\ \varepsilon_i=\mathrm{LGD}_i\times\mathbb{1}\left\{\boldsymbol{\tau}_i<T_i\right\}\end{array}\right.$$

The exposure at default is not random, which is not the case of the loss given default.

(b) We have to make the following assumptions:

i. the loss given default LGD_i is independent from the default time $\boldsymbol{\tau}_i$;

ii. the portfolio is infinitely fine-grained meaning that there is no exposure concentration:

$$\frac{\text{EAD}_i}{\sum_{i=1}^n \text{EAD}_i} \simeq 0$$

iii. the default times depend on a common risk factor X and the relationship is monotonic (increasing or decreasing).

In this case, we have:

$$\mathbb{E}\left[\varepsilon_i \mid L = \mathbf{F}^{-1}(\alpha)\right] = \mathbb{E}\left[\text{LGD}_i\right] \times \mathbb{E}\left[D_i \mid L = \mathbf{F}^{-1}(\alpha)\right]$$

with $D_i = \mathbb{1}\{\boldsymbol{\tau}_i < T_i\}$.

(c) It follows that:

$$\begin{aligned} \mathcal{RC}_i &= w_i \times \mathcal{MR}_i \\ &= \text{EAD}_i \times \mathbb{E}\left[\text{LGD}_i\right] \times \mathbb{E}\left[D_i \mid L = \mathbf{F}^{-1}(\alpha)\right] \end{aligned}$$

The expression of the value-at-risk is then:

$$\begin{aligned} \text{VaR}_\alpha(w) &= \sum_{i=1}^n \mathcal{RC}_i \\ &= \sum_{i=1}^n \text{EAD}_i \times \mathbb{E}\left[\text{LGD}_i\right] \times \mathbb{E}\left[D_i \mid L = \mathbf{F}^{-1}(\alpha)\right] \end{aligned}$$

(d) i. We have

$$\begin{aligned} \mathbb{E}\left[Z_i Z_j\right] &= \mathbb{E}\left[\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i\right)\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_j\right)\right] \\ &= \rho \end{aligned}$$

ρ is the constant correlation between assets Z_i and Z_j.

ii. We have:

$$\begin{aligned} p_i &= \Pr\{\tau_i \leq T_i\} \\ &= \Pr\{Z_i \leq B_i\} \\ &= \Phi(B_i) \end{aligned}$$

iii. It follows that:

$$\begin{aligned} p_i(x) &= \Pr\{Z_i \leq B_i \mid X = x\} \\ &= \Pr\left\{\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i \leq B_i \mid X = x\right\} \\ &= \Pr\left\{\varepsilon_i \leq \frac{B_i - \sqrt{\rho}X}{\sqrt{1-\rho}} \,\middle|\, X = x\right\} \\ &= \Phi\left(\frac{B_i - \sqrt{\rho}x}{\sqrt{1-\rho}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}x}{\sqrt{1-\rho}}\right) \end{aligned}$$

(e) Under the assumptions $(\mathcal{H})$, we know that:

$$\begin{aligned} L &= \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times p_i\left(X\right) \\ &= \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times \Phi\left(\frac{\Phi^{-1}\left(p_i\right) - \sqrt{\rho} X}{\sqrt{1-\rho}}\right) \\ &= g\left(X\right) \end{aligned}$$

with $g'(x) < 0$. We deduce that:

$$\begin{aligned} \mathrm{VaR}_\alpha\left(w\right) = \mathbf{F}^{-1}\left(\alpha\right) &\Leftrightarrow \Pr\left\{g\left(X\right) \leq \mathrm{VaR}_\alpha\left(w\right)\right\} = \alpha \\ &\Leftrightarrow \Pr\left\{X \geq g^{-1}\left(\mathrm{VaR}_\alpha\left(w\right)\right)\right\} = \alpha \\ &\Leftrightarrow \Pr\left\{X \leq g^{-1}\left(\mathrm{VaR}_\alpha\left(w\right)\right)\right\} = 1-\alpha \\ &\Leftrightarrow g^{-1}\left(\mathrm{VaR}_\alpha\left(w\right)\right) = \Phi^{-1}\left(1-\alpha\right) \\ &\Leftrightarrow \mathrm{VaR}_\alpha\left(w\right) = g\left(\Phi^{-1}\left(1-\alpha\right)\right) \end{aligned}$$

It follows that:

$$\begin{aligned} \mathrm{VaR}_\alpha\left(w\right) &= g\left(\Phi^{-1}\left(1-\alpha\right)\right) \\ &= \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times p_i\left(\Phi^{-1}\left(1-\alpha\right)\right) \end{aligned}$$

The risk contribution $\mathcal{RC}_i$ of the ith credit is then:

$$\begin{aligned} \mathcal{RC}_i &= \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times p_i\left(\Phi^{-1}\left(1-\alpha\right)\right) \\ &= \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times \Phi\left(\frac{\Phi^{-1}\left(p_i\right) - \sqrt{\rho}\Phi^{-1}\left(1-\alpha\right)}{\sqrt{1-\rho}}\right) \\ &= \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times \Phi\left(\frac{\Phi^{-1}\left(p_i\right) + \sqrt{\rho}\Phi^{-1}\left(\alpha\right)}{\sqrt{1-\rho}}\right) \end{aligned}$$

3. (a) We note Ω the event $X \leq g^{-1}\left(\mathrm{VaR}_\alpha\left(w\right)\right)$ or equivalently $X \leq \Phi^{-1}\left(1-\alpha\right)$. We have:

$$\begin{aligned} \mathrm{ES}_\alpha\left(w\right) &= \mathbb{E}\left[L \mid L \geq \mathrm{VaR}_\alpha\left(w\right)\right] \\ &= \mathbb{E}\left[L \mid g\left(X\right) \geq \mathrm{VaR}_\alpha\left(w\right)\right] \\ &= \mathbb{E}\left[L \mid X \leq g^{-1}\left(\mathrm{VaR}_\alpha\left(w\right)\right)\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times p_i\left(X\right) \mid \Omega\right] \\ &= \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times \mathbb{E}\left[p_i\left(X\right) \mid \Omega\right] \end{aligned}$$

(b) It follows that:

$$\begin{aligned}
\mathbb{E}\left[p_i(X) \mid \Omega\right] &= \mathbb{E}\left[\Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}X}{\sqrt{1-\rho}}\right) \middle| \Omega\right] \\
&= \int_{-\infty}^{\Phi^{-1}(1-\alpha)} \Phi\left(\frac{\Phi^{-1}(p_i)}{\sqrt{1-\rho}} + \frac{-\sqrt{\rho}}{\sqrt{1-\rho}}x\right) \times \\
&\qquad \frac{\phi(x)}{\Phi\left(\Phi^{-1}(1-\alpha)\right)}\,\mathrm{d}x \\
&= \frac{\Phi_2\left(\Phi^{-1}(1-\alpha), \Phi^{-1}(p_i); \sqrt{\rho}\right)}{1-\alpha} \\
&= \frac{\mathbf{C}\left(1-\alpha, p_i; \sqrt{\rho}\right)}{1-\alpha}
\end{aligned}$$

where $\mathbf{C}$ is the Gaussian copula. We deduce that:

$$\mathcal{RC}_i = \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times \frac{\mathbf{C}\left(1-\alpha, p_i; \sqrt{\rho}\right)}{1-\alpha}$$

(c) If $\rho = 0$, we have:

$$\begin{aligned}
\Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) &= \Phi\left(\Phi^{-1}(p_i)\right) \\
&= p_i
\end{aligned}$$

and:

$$\begin{aligned}
\frac{\mathbf{C}\left(1-\alpha, p_i; \sqrt{\rho}\right)}{1-\alpha} &= \frac{(1-\alpha)\,p_i}{1-\alpha} \\
&= p_i
\end{aligned}$$

The risk contribution is the same for the value-at-risk and the expected shortfall:

$$\begin{aligned}
\mathcal{RC}_i &= \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right] \times p_i \\
&= \mathbb{E}\left[L_i\right]
\end{aligned}$$

It corresponds to the expected loss of the credit. If $\rho = 1$ and $\alpha > 50\%$, we have:

$$\begin{aligned}
\Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) &= \lim_{\rho \to 1} \Phi\left(\frac{\Phi^{-1}(p_i) + \Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \\
&= 1
\end{aligned}$$

If $\rho = 1$ and α is high ($\alpha > 1 - \sup_i p_i$), we have:

$$\begin{aligned}
\frac{\mathbf{C}\left(1-\alpha, p_i; \sqrt{\rho}\right)}{1-\alpha} &= \frac{\min\left(1-\alpha; p_i\right)}{1-\alpha} \\
&= 1
\end{aligned}$$

In this case, the risk contribution is the same for the value-at-risk and the expected shortfall:

$$\mathcal{RC}_i = \mathrm{EAD}_i \times \mathbb{E}\left[\mathrm{LGD}_i\right]$$

However, it does not depend on the unconditional probability of default p_i.

4. Pillar 2 concerns the non-compliance of assumptions ($\mathcal{H}$). In particular, we have to understand the impact on the credit risk measure if the portfolio is not infinitely fine-grained or if asset correlations are not constant.

3.4.3 Calibration of the piecewise exponential model

1. We have:

$$\mathbf{S}(t) = \Pr\{\boldsymbol{\tau} \geq t\} = 1 - \mathbf{F}(t)$$

and:

$$f(t) = \partial_t \mathbf{F}(t) = -\partial_t \mathbf{S}(t)$$

2. The function $\lambda(t)$ is the instantaneous default rate:

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta \to 0^+} \frac{1}{\Delta} \Pr\{t \leq \boldsymbol{\tau} \leq t + \Delta \mid \tau \geq t\} \\
&= \lim_{\Delta \to 0^+} \frac{1}{\Delta} \frac{\Pr\{t \leq \boldsymbol{\tau} \leq t + \Delta\}}{\Pr\{\boldsymbol{\tau} \geq t\}} \\
&= \frac{1}{\Pr\{\boldsymbol{\tau} \geq t\}} \lim_{\Delta \to 0^+} \frac{\Pr\{t \leq \boldsymbol{\tau} \leq t + \Delta\}}{\Delta} \\
&= \frac{f(t)}{\mathbf{S}(t)}
\end{aligned}$$

In the case of the exponential model, we obtain:

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

3. Since $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$, it follows that:

$$\begin{aligned}
s(T) &= \frac{(1 - \mathcal{R}) \times \int_0^T e^{-rt} f(t)\,\mathrm{d}t}{\int_0^T e^{-rt} \mathbf{S}(t)\,\mathrm{d}t} \\
&= \frac{(1 - \mathcal{R}) \times \int_0^T e^{-rt} \lambda e^{-\lambda t}\,\mathrm{d}t}{\int_0^T e^{-rt} e^{-\lambda t}\,\mathrm{d}t} \\
&= \lambda \times (1 - \mathcal{R})
\end{aligned}$$

4. (a) We define the survival function as follows:

$$\mathbf{S}(t) = \begin{cases} e^{-\lambda_1 t} & \text{if } t \leq 3 \\ e^{-3\lambda_1 - \lambda_2 (t-3)} & \text{if } 3 < t \leq 5 \\ e^{-3\lambda_1 - 2\lambda_2 - \lambda_3 (t-5)} & \text{if } t > 5 \end{cases}$$

We deduce that:

$$f(t) = \begin{cases} \lambda_1 e^{-\lambda_1 t} & \text{if } t \leq 3 \\ \lambda_2 e^{-3\lambda_1 - \lambda_2 (t-3)} & \text{if } 3 < t \leq 5 \\ \lambda_3 e^{-3\lambda_1 - 2\lambda_2 - \lambda_3 (t-5)} & \text{if } t > 5 \end{cases}$$

We verify that the hazard rate is a piecewise constant function:

$$\lambda(t) = \frac{f(t)}{\mathbf{S}(t)} = \begin{cases} \lambda_1 & \text{if } t \leq 3 \\ \lambda_2 & \text{if } 3 < t \leq 5 \\ \lambda_3 & \text{if } t > 5 \end{cases}$$

(b) Let $(t_1^\star, t_2^\star, t_3^\star)$ be the knots of the piecewise exponential model[5]. We note that $\mathbf{S}(t) = \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t-t_{m-1}^\star\right)}$ and $f(t) = \lambda_m \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t-t_{m-1}^\star\right)}$. When $T \in \left[t_{m-1}^\star, t_m^\star\right[$, it follows that:

$$\begin{aligned} \mathcal{s}(T) &= \frac{(1-\boldsymbol{\mathcal{R}}) \times \int_0^T e^{-rt} f(t)\,\mathrm{d}t}{\int_0^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t} \\ &= \frac{(1-\boldsymbol{\mathcal{R}}) \times \left(\int_0^{t_{m-1}^\star} e^{-rt} f(t)\,\mathrm{d}t + \int_{t_{m-1}^\star}^T e^{-rt} f(t)\,\mathrm{d}t\right)}{\left(\int_0^{t_{m-1}^\star} e^{-rt}\mathbf{S}(t)\,\mathrm{d}t + \int_{t_{m-1}^\star}^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t\right)} \end{aligned}$$

We introduce the following notation with $T \leq t_m^\star$:

$$\begin{aligned} \mathcal{I}\left(t_{m-1}^\star, T\right) &= \int_{t_{m-1}^\star}^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t \\ &= \mathbf{S}\left(t_{m-1}^\star\right) \int_{t_{m-1}^\star}^T e^{-rt} e^{-\lambda_m\left(t-t_{m-1}^\star\right)}\,\mathrm{d}t \\ &= \mathbf{S}\left(t_{m-1}^\star\right) e^{\lambda_m t_{m-1}^\star} \int_{t_{m-1}^\star}^T e^{-(r+\lambda_m)t}\,\mathrm{d}t \\ &= \mathbf{S}\left(t_{m-1}^\star\right) e^{\lambda_m t_{m-1}^\star} \left[\frac{e^{-(r+\lambda_m)t}}{-(r+\lambda_m)}\right]_{t_{m-1}^\star}^T \\ &= \mathbf{S}\left(t_{m-1}^\star\right) e^{\lambda_m t_{m-1}^\star} \frac{e^{-(r+\lambda_m)t_{m-1}^\star} - e^{-(r+\lambda_m)T}}{(r+\lambda_m)} \\ &= \mathbf{S}\left(t_{m-1}^\star\right) \frac{e^{-rt_{m-1}^\star} - e^{-rT} e^{-\lambda_m\left(T-t_{m-1}^\star\right)}}{(r+\lambda_m)} \end{aligned}$$

We obtain the following cases:

i. If $T \leq 3$, we have:

$$\begin{aligned} \mathcal{s}(T) &= \frac{(1-\boldsymbol{\mathcal{R}}) \times \int_0^T e^{-rt} f(t)\,\mathrm{d}t}{\int_0^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t} \\ &= \frac{(1-\boldsymbol{\mathcal{R}}) \times \lambda_1 \times \mathcal{I}(0,T)}{\mathcal{I}(0,T)} \\ &= (1-\boldsymbol{\mathcal{R}}) \times \lambda_1 \end{aligned}$$

ii. If $3 < T \leq 5$, we have:

$$\begin{aligned} \mathcal{s}(T) &= \frac{(1-\boldsymbol{\mathcal{R}}) \times \left(\int_0^3 e^{-rt} f(t)\,\mathrm{d}t + \int_3^T e^{-rt} f(t)\,\mathrm{d}t\right)}{\left(\int_0^3 e^{-rt}\mathbf{S}(t)\,\mathrm{d}t + \int_3^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t\right)} \\ &= (1-\boldsymbol{\mathcal{R}}) \times \frac{\lambda_1 \mathcal{I}(0,3) + \lambda_2 \mathcal{I}(3,T)}{\mathcal{I}(0,3) + \mathcal{I}(3,T)} \end{aligned}$$

iii. If $T > 5$, we have:

$$\mathcal{s}(T) = (1-\boldsymbol{\mathcal{R}}) \times \frac{\lambda_1 \mathcal{I}(0,3) + \lambda_2 \mathcal{I}(3,5) + \lambda_3 \mathcal{I}(5,T)}{\mathcal{I}(0,3) + \mathcal{I}(3,5) + \mathcal{I}(5,T)}$$

[5] We use the convention $t_0^\star = 0$.

(c) The parameters $(\lambda_1, \lambda_2, \lambda_3)$ satisfy the following set of equations:

$$\left\{ \begin{array}{c} \mathcal{s}(3) = (1 - \mathcal{R}) \times \lambda_1 \\ \mathcal{s}(5) = (1 - \mathcal{R}) \times \frac{\lambda_1 \mathcal{I}(0,3) + \lambda_2 \mathcal{I}(3,5)}{\mathcal{I}(0,3) + \mathcal{I}(3,5)} \\ \mathcal{s}(7) = (1 - \mathcal{R}) \times \frac{\lambda_1 \mathcal{I}(0,3) + \lambda_2 \mathcal{I}(3,5) + \lambda_3 \mathcal{I}(5,7)}{\mathcal{I}(0,3) + \mathcal{I}(3,5) + \mathcal{I}(5,7)} \end{array} \right. \tag{3.1}$$

From the first equation, we estimate $\hat{\lambda}_1$:

$$\hat{\lambda}_1 = \frac{\mathcal{s}(3)}{(1 - \mathcal{R})}$$

We can now solve numerically the second equation and we obtain $\hat{\lambda}_2$. Finally, we solve the nonlinear third equation to obtain $\hat{\lambda}_3$. This iterative approach of calibration is known as the bootstrapping method.

(d) When r is equal to zero and λ_m is small, the function $\mathcal{I}\left(t^\star_{m-1}, T\right)$ becomes:

$$\begin{aligned} \mathcal{I}\left(t^\star_{m-1}, T\right) &= \mathbf{S}\left(t^\star_{m-1}\right) \frac{1 - e^{-\lambda_m \left(T - t^\star_{m-1}\right)}}{\lambda_m} \\ &\approx \mathbf{S}\left(t^\star_{m-1}\right)\left(T - t^\star_{m-1}\right) \\ &\approx T - t^\star_{m-1} \end{aligned}$$

We introduce the following notation:

$$\lambda(T) = \frac{\mathcal{s}(T)}{(1 - \mathcal{R})}$$

Using Equation (3.1), we deduce that:

$$\begin{aligned} \hat{\lambda}_1 &= \lambda(3) \\ \hat{\lambda}_2 &= \left(\frac{\mathcal{I}(0,3) + \mathcal{I}(3,5)}{\mathcal{I}(3,5)}\right)\lambda(5) - \frac{\mathcal{I}(0,3)}{\mathcal{I}(3,5)}\hat{\lambda}_1 \\ &\approx \frac{5\lambda(5) - 3\lambda(3)}{2} \\ \hat{\lambda}_3 &= \left(\frac{\mathcal{I}(0,3) + \mathcal{I}(3,5) + \mathcal{I}(5,7)}{\mathcal{I}(5,7)}\right)\lambda(7) - \frac{\hat{\lambda}_1 \mathcal{I}(0,3) + \hat{\lambda}_2 \mathcal{I}(3,5)}{\mathcal{I}(5,7)} \\ &\approx \frac{7\lambda(7) - 5\lambda(5)}{2} \end{aligned}$$

We notice that:

$$\begin{aligned} \mathcal{s}(3) &= (1 - \mathcal{R}) \times \hat{\lambda}_1 \\ \mathcal{s}(5) &= (1 - \mathcal{R}) \times \left(\frac{3\hat{\lambda}_1 + 2\hat{\lambda}_2}{5}\right) \\ \mathcal{s}(7) &= (1 - \mathcal{R}) \times \left(\frac{3\hat{\lambda}_1 + 2\hat{\lambda}_2 + 2\hat{\lambda}_3}{7}\right) \end{aligned}$$

The spread is then a weighted average of the different hazard rates, whose weights are proportional to the interval time between two knots.

(e) Using a numerical solver, we obtain $\hat{\lambda}_1 = 166.7$ bps, $\hat{\lambda}_2 = 401.2$ bps and $\hat{\lambda}_3 = 322.4$ bps[6].

(f) Since $\mathbf{S}(\boldsymbol{\tau}) \sim \mathcal{U}_{[0,1]}$, the simulated default time t is $\mathbf{S}^{-1}(u)$ where u is a uniform random number. If $u > \mathbf{S}(3)$, we have $e^{-\lambda_1 t} = u$ or $t = -\lambda_1^{-1} \ln u$. If $\mathbf{S}(5) < u \leq \mathbf{S}(3)$, it follows that $\mathbf{S}(3) e^{-\lambda_2 (t-3)} = u$ or $t = 3 + \lambda_2^{-1} (\ln \mathbf{S}(3) - \ln u)$. Finally, we obtain $t = 5 + \lambda_3^{-1} (\ln \mathbf{S}(5) - \ln u)$ if $u < \mathbf{S}(5)$. Using the previous numerical values, we find that $\mathbf{S}(3) = 0.951$ and $\mathbf{S}(5) = 0.878$. The simulated default times are then:

$$t = \begin{cases} 2.449 \text{ for } u = 0.96 \\ 46.54 \text{ for } u = 0.23 \\ 4.380 \text{ for } u = 0.90 \\ 7.881 \text{ for } u = 0.80 \end{cases}$$

3.4.4 Modeling loss given default

1. The loss given default is equal to:

$$\text{LGD} = 1 - \boldsymbol{\mathcal{R}} + c$$

where c is the recovery (or litigation) cost. Consider for example a \$200 credit and suppose that the borrower defaults. If we recover \$140 and the litigation cost is \$20, we obtain $\boldsymbol{\mathcal{R}} = 70\%$ and $\text{LGD} = 40\%$, but not $\text{LGD} = 30\%$.

2. The amounts outstanding of credit is:

$$\begin{aligned} \text{EAD} &= 250\,000 \times 50\,000 \\ &= \$12.5 \text{ bn} \end{aligned}$$

The annual loss after recovery is equal to:

$$\begin{aligned} L &= \text{EAD} \times (1 - \boldsymbol{\mathcal{R}}) \times \text{PD} + C \\ &= 43.75 + 12.5 \\ &= \$56.25 \text{ mn} \end{aligned}$$

where C is the litigation cost. We deduce that:

$$\begin{aligned} \text{LGD} &= \frac{L}{\text{EAD} \times \text{PD}} \\ &= \frac{54}{12.5 \times 10^3 \times 1\%} \\ &= 45\% \end{aligned}$$

This figure is larger than 35%, which is the loss given default without taking into account the recovery cost.

3. (a) The Beta distribution allows to obtain all the forms of LGD (bell curve, inverted-U shaped curve, etc.). The uniform distribution corresponds to the case $\alpha = 1$ and $\beta = 1$. Indeed, we have:

$$\begin{aligned} f(x) &= \frac{x^{\alpha-1} (1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1} (1-u)^{\beta-1} \, \mathrm{d}u} \\ &= 1 \end{aligned}$$

[6] If we consider the approximated formulas, the solutions are $\hat{\lambda}_1 = 166.7$ bps, $\hat{\lambda}_2 = 375.0$ bps and $\hat{\lambda}_3 = 308.3$ bps.

(b) We have:

$$\begin{aligned}\boldsymbol{\ell}(\alpha,\beta) &= \sum_{i=1}^{n} \ln f(x_i) \\ &= -n \ln \mathbf{B}(\alpha,\beta) + (\alpha-1)\sum_{i=1}^{n} \ln x_i + (\beta-1)\sum_{i=1}^{n} \ln(1-x_i)\end{aligned}$$

The first-order conditions are:

$$\frac{\partial\, \boldsymbol{\ell}(\alpha,\beta)}{\partial\, \alpha} = -n\frac{\partial_\alpha \mathbf{B}(\alpha,\beta)}{\mathbf{B}(\alpha,\beta)} + \sum_{i=1}^{n} \ln x_i = 0$$

and:

$$\frac{\partial\, \boldsymbol{\ell}(\alpha,\beta)}{\partial\, \beta} = -n\frac{\partial_\beta \mathbf{B}(\alpha,\beta)}{\mathbf{B}(\alpha,\beta)} + \sum_{i=1}^{n} \ln(1-x_i) = 0$$

(c) Let μ_{LGD} and σ_{LGD} be the mean and standard deviation of the LGD parameter. The method of moments consists in estimating α and β such that:

$$\frac{\alpha}{\alpha+\beta} = \mu_{\mathrm{LGD}}$$

and:

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \sigma^2_{\mathrm{LGD}}$$

We have:

$$\beta = \alpha\frac{(1-\mu_{\mathrm{LGD}})}{\mu_{\mathrm{LGD}}}$$

and:

$$(\alpha+\beta)^2(\alpha+\beta+1)\,\sigma^2_{\mathrm{LGD}} = \alpha\beta$$

It follows that:

$$\begin{aligned}(\alpha+\beta)^2 &= \left(\alpha + \alpha\frac{(1-\mu_{\mathrm{LGD}})}{\mu_{\mathrm{LGD}}}\right)^2 \\ &= \frac{\alpha^2}{\mu^2_{\mathrm{LGD}}}\end{aligned}$$

and:

$$\alpha\beta = \frac{\alpha^2}{\mu^2_{\mathrm{LGD}}}\left(\alpha + \alpha\frac{(1-\mu_{\mathrm{LGD}})}{\mu_{\mathrm{LGD}}} + 1\right)\sigma^2_{\mathrm{LGD}} = \alpha^2\frac{(1-\mu_{\mathrm{LGD}})}{\mu_{\mathrm{LGD}}}$$

We deduce that:

$$\alpha\left(1 + \frac{(1-\mu_{\mathrm{LGD}})}{\mu_{\mathrm{LGD}}}\right) = \frac{(1-\mu_{\mathrm{LGD}})\,\mu_{\mathrm{LGD}}}{\sigma^2_{\mathrm{LGD}}} - 1$$

We finally obtain:

$$\hat{\alpha}_{\mathrm{MM}} = \frac{\mu^2_{\mathrm{LGD}}(1-\mu_{\mathrm{LGD}})}{\sigma^2_{\mathrm{LGD}}} - \mu_{\mathrm{LGD}} \tag{3.2}$$

$$\hat{\beta}_{\mathrm{MM}} = \frac{\mu_{\mathrm{LGD}}(1-\mu_{\mathrm{LGD}})^2}{\sigma^2_{\mathrm{LGD}}} - (1-\mu_{\mathrm{LGD}}) \tag{3.3}$$

4. (a) The mean of the loss given default is equal to:

$$\begin{aligned}\mu_{\text{LGD}} &= \frac{100 \times 0\% + 100 \times 25\% + 600 \times 50\% + \ldots}{1000} \\ &= 50\%\end{aligned}$$

The expression of the expected loss is:

$$\text{EL} = \sum_{i=1}^{100} \text{EAD}_i \times \mathbb{E}\left[\text{LGD}_i\right] \times \text{PD}_i$$

where PD_i is the default probability of credit i. We finally obtain:

$$\begin{aligned}\text{EL} &= \sum_{i=1}^{100} 10\,000 \times 50\% \times 1\% \\ &= \$5\,000\end{aligned}$$

(b) We have $\mu_{\text{LGD}} = 50\%$ and:

$$\begin{aligned}\sigma_{\text{LGD}} &= \sqrt{\frac{100 \times (0 - 0.5)^2 + 100 \times (0.25 - 0.5)^2 + \ldots}{1000}} \\ &= \sqrt{\frac{2 \times 0.5^2 + 2 \times 0.25^2}{10}} \\ &= \sqrt{\frac{0.625}{10}} \\ &= 25\%\end{aligned}$$

Using Equations (3.2) and (3.3), we deduce that:

$$\begin{aligned}\hat{\alpha}_{\text{MM}} &= \frac{0.5^2 \times (1 - 0.5)}{0.25^2} - 0.5 = 1.5 \\ \hat{\beta}_{\text{MM}} &= \frac{0.5 \times (1 - 0.5)^2}{0.25^2} - (1 - 0.5) = 1.5\end{aligned}$$

(c) The previous portfolio is homogeneous and infinitely fine-grained. In this case, we know that the unexpected loss depends on the mean of the loss given default and not on the entire probability distribution. Because the expected value of the calibrated Beta distribution is 50%, there is no difference with the uniform distribution, which has also a mean equal to 50%. This result holds for the Basel model with one factor, and remains true when they are more factors.

3.4.5 Modeling default times with a Markov chain

1. We have $P(4) = P(2)\,P(2)$ and $P(6) = P(4)\,P(2)$.

2. In a piecewise exponential model, the survival function has the following expression:

$$\mathbf{S}(t) = \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t - t_{m-1}^\star\right)} \qquad \text{if } t \in \left]t_{m-1}^\star, t_m^\star\right]$$

We deduce that:

$$\lambda_m = \frac{\ln \mathbf{S}\left(t_{m-1}^\star\right) - \ln \mathbf{S}\left(t_m^\star\right)}{t_m^\star - t_{m-1}^\star}$$

with $\mathbf{S}(t_0^\star) = \mathbf{S}(0) = 1$. Here, the knots of the piecewise function are $t_1^\star = 2$, $t_2^\star = 4$ and $t_3^\star = 6$. If we consider the risk class A, we deduce that:

$$\begin{aligned}
\lambda_1 &= \frac{\ln 1 - \ln(1 - 1\%)}{2 - 0} = 50.3 \text{ bps} \\
\lambda_2 &= \frac{\ln(1 - 1\%) - \ln(1 - 2.49\%)}{4 - 2} = 75.8 \text{ bps} \\
\lambda_3 &= \frac{\ln(1 - 2.49\%) - \ln(1 - 4.296\%)}{6 - 4} = 93.5 \text{ bps}
\end{aligned}$$

We finally obtain the following results:

Rating	A	B	C
λ_1	50.3	256.5	1115.7
λ_2	75.8	275.9	856.9
λ_3	93.5	277.8	650.2

3. Let $P(t)$ be the transition matrix between 0 and t. The Markov generator of $P(t)$ is the matrix $\Lambda = (\lambda_{i,j})$ defined by:

$$P(t) = \exp(t\Lambda)$$

where e^M is the matrix exponential of the matrix M. We deduce that:

$$\Lambda = \frac{\ln P(t)}{t}$$

In this example, the direct estimator is given by:

$$\hat{\Lambda} = \frac{\ln P(2)}{2}$$

We verify that $\hat{\Lambda}$ is a Markov generator because $\sum_{j=1}^{4} \lambda_{i,j} = 0$ and $\lambda_{i,j} \geq 0$ when $i \neq j$.

4. For the piecewise exponential model, we proceed as in Question 2 by adding the knots $t_m^\star = 2m$ with $m \geq 4$. In this case, we have:

$$\lambda_i(t) = \lambda_m \qquad \text{if } t \in \left]2m - 2, 2m\right]$$

with:

$$\lambda_m = \frac{\ln \mathbf{S}_i(2m - 2) - \ln \mathbf{S}_i(2m)}{2}$$

and $\mathbf{S}_i(2m) = 1 - P_{i,4}(2m)$. For the Markov generator, we have:

$$\begin{aligned}
\mathbf{S}_i(t) &= 1 - \mathbf{e}_i^\top P(t)\,\mathbf{e}_4 \\
&= 1 - \mathbf{e}_i^\top e^{t\hat{\Lambda}} \mathbf{e}_4
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\lambda_i(t) &= \frac{-\partial_t \mathbf{S}_i(t)}{\mathbf{S}_i(t)} \\
&= \frac{\mathbf{e}_i^\top \hat{\Lambda} e^{t\hat{\Lambda}} \mathbf{e}_4}{1 - \mathbf{e}_i^\top e^{t\hat{\Lambda}} \mathbf{e}_4}
\end{aligned}$$

In the long-run, the markov chain is stationary. This means that the default probability of the different risk classes is the same when t tends to ∞ and we have:

$$\lambda_A(\infty) = \lambda_B(\infty) = \lambda_C(\infty) = 147.6 \text{ bps}$$

In the short-run, the hazard rate are ranked with respect to the risk class:

$$\lambda_A(0) < \lambda_B(0) < \lambda_C(0)$$

We deduce that the function $\lambda_A(t)$ is increasing whereas the function $\lambda_C(t)$ is decreasing. For the rating B, the behavior of the hazard function is more complex. It first increases like $\lambda_A(t)$ and reaches a maximum at $t = 4.2$, because the transition probability to risk classes C and D is very high. Then, it decreases because of the stationarity property.

3.4.6 Continuous-time modeling of default risk

1. The Chapman-Kolmogorov equation is:

$$P(n) = P(n-1)P$$

We deduce that:

$$\begin{aligned} P(n) &= P(0)\prod_{t=1}^{n} P \\ &= P^n \end{aligned}$$

because $P(0) = I_4$. We have:

$$P(10) = \begin{pmatrix} 62.60\% & 13.14\% & 5.53\% & 18.73\% \\ 38.42\% & 20.74\% & 6.81\% & 34.03\% \\ 21.90\% & 12.29\% & 4.35\% & 61.46\% \\ 0.00\% & 0.00\% & 0.00\% & 100.00\% \end{pmatrix}$$

2. (a) The eigendecomposition of P is equal to $P = VDV^{-1}$, meaning that:

$$PV = VD$$

We deduce that:

$$\begin{aligned} P(2)V &= PVD \\ &= VDD \\ &= VD^2 \end{aligned}$$

By recursion, we obtain:

$$P(n)V = VD^n$$

We can then calculate $P(n)$ as follows:

$$P(n) = VD^nV^{-1}$$

The eigendecomposition of $P(n)$ is similar to the eigendecomposition of P: the eigenvectors are the same, only the eigenvalues are different.

(b) We have:

$$V = \begin{pmatrix} 0.4670 & -0.2808 & -0.0264 & 1.0000 \\ 0.3561 & 0.8486 & -0.2373 & 1.0000 \\ 0.2065 & 0.5363 & 0.8609 & 1.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 0.9717 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.8111 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.5571 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

We deduce that:

$$D^{10} = \begin{pmatrix} 0.7506 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.1233 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0029 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

We verify that:

$$\begin{aligned} VD^{10}V^{-1} &= \begin{pmatrix} 62.60\% & 13.14\% & 5.53\% & 18.73\% \\ 38.42\% & 20.74\% & 6.81\% & 34.03\% \\ 21.90\% & 12.29\% & 4.35\% & 61.46\% \\ 0.00\% & 0.00\% & 0.00\% & 100.00\% \end{pmatrix} \\ &= P(10) \end{aligned}$$

3. Let $\mathfrak{R}_i(n)$ be the rating of a firm at time n whose initial rating is the state i. We have:

$$\begin{aligned} \mathbf{S}_i(n) &= 1 - \Pr\{\mathfrak{R}_i(n) = D\} \\ &= 1 - \mathbf{e}_i^\top P(n)\,\mathbf{e}_4 \\ &= 1 - \mathbf{e}_i^\top P^n \mathbf{e}_4 \\ &= 1 - (P^n)_{i,4} \end{aligned}$$

In the piecewise exponential model, we recall that the survival function has the following expression:

$$\mathbf{S}_i(n) = \mathbf{S}_i(n-1)\, e^{-\lambda_i(n)}$$

We deduce that:

$$\begin{aligned} \lambda_i(n) &= \ln \mathbf{S}_i(n-1) - \ln \mathbf{S}_i(n) \\ &= \ln\left(\frac{1 - \mathbf{e}_i^\top P^{n-1}\mathbf{e}_4}{1 - \mathbf{e}_i^\top P^n \mathbf{e}_4}\right) \end{aligned}$$

We verify that:

$$\begin{aligned} \lambda_i(1) &= \ln\left(\frac{1}{1 - \mathbf{e}_i^\top P^n \mathbf{e}_4}\right) \\ &= -\ln\left(1 - \mathbf{e}_i^\top P^n \mathbf{e}_4\right) \end{aligned}$$

because $\mathbf{S}_i(0) = 1$. Numerical values of $\mathbf{S}_i(n)$ and $\lambda_i(n)$ are given in Table 3.1.

TABLE 3.1: Numerical values of $\mathbf{S}_i(n)$ and $\lambda_i(n)$

n	$\mathbf{S}_A(n)$	$\mathbf{S}_B(n)$	$\mathbf{S}_C(n)$	$\lambda_A(n)$	$\lambda_B(n)$	$\lambda_C(n)$
0	1.0000	1.0000	1.0000			
1	0.7914	0.6360	0.3708	0.2339	0.4526	0.9921
2	0.7704	0.6139	0.3575	0.0269	0.0354	0.0367
3	0.7498	0.5932	0.3451	0.0272	0.0343	0.0352
4	0.7295	0.5737	0.3335	0.0274	0.0334	0.0341
5	0.7096	0.5552	0.3226	0.0276	0.0327	0.0332
6	0.6901	0.5377	0.3123	0.0278	0.0320	0.0324
7	0.6711	0.5211	0.3026	0.0280	0.0315	0.0318
8	0.6525	0.5051	0.2933	0.0281	0.0310	0.0313
9	0.6344	0.4899	0.2843	0.0282	0.0307	0.0308
10	0.6167	0.4752	0.2758	0.0283	0.0303	0.0305
50	0.1962	0.1496	0.0868	0.0287	0.0287	0.0287
100	0.0468	0.0357	0.0207	0.0287	0.0287	0.0287

4. Let $P(t)$ be the transition matrix between 0 and t. The Markov generator of $P(t)$ is the matrix $\Lambda = (\lambda_{i,j})$ defined by:

$$P(t) = \exp(t\Lambda)$$

where e^M is the matrix exponential of the matrix M. We deduce that:

$$\Lambda = t^{-1} \ln P(t)$$

In particular, we have:

$$\begin{aligned} \hat{\Lambda} &= \frac{\ln P(n)}{n} \\ &= \frac{\ln P^n}{n} \\ &= \ln P \end{aligned}$$

We obtain:

$$\hat{\Lambda} = \begin{pmatrix} -6.4293 & 3.2282 & 2.4851 & 0.7160 \\ 11.3156 & -23.5006 & 9.9915 & 2.1936 \\ 5.3803 & 21.6482 & -52.3649 & 25.3364 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} \times 10^{-2}$$

We verify that $\hat{\Lambda}$ is a Markov generator because $\sum_{j=1}^{4} \lambda_{i,j} = 0$ and $\lambda_{i,j} \geq 0$ when $i \neq j$.

5. We have:

$$P(t) = \exp(t\Lambda)$$

We remind that:

$$e^M = I + M + \frac{M^2}{2!} + \frac{M^3}{3!} + \ldots$$

We deduce that:

$$P(t) = I_4 + t\Lambda + \frac{t^2}{2}\Lambda^2 + \frac{t^3}{6}\Lambda^3 + \frac{t^4}{24}\Lambda^4 + \ldots$$

The 6-month transition probability matrix is equal to:

$$\begin{aligned} P\left(\frac{1}{2}\right) &= \exp\left(\frac{\Lambda}{2}\right) \\ &= \begin{pmatrix} 96.90\% & 1.56\% & 1.11\% & 0.43\% \\ 5.32\% & 89.19\% & 4.17\% & 1.33\% \\ 2.60\% & 8.99\% & 77.20\% & 11.21\% \\ 0.00\% & 0.00\% & 0.00\% & 100.00\% \end{pmatrix} \end{aligned}$$

6. We have:

$$\begin{aligned} \mathbf{S}_i(t) &= 1 - \Pr\{\mathfrak{R}_i(t) = D\} \\ &= 1 - \mathbf{e}_i^\top P(t)\,\mathbf{e}_4 \\ &= 1 - \mathbf{e}_i^\top e^{t\Lambda}\mathbf{e}_4 \end{aligned}$$

We know that:

$$\lambda_i(t) = \frac{f_i(t)}{\mathbf{S}_i(t)} = -\frac{\partial_t\,\mathbf{S}_i(t)}{\mathbf{S}_i(t)}$$

We deduce that:

$$\lambda_i(t) = \frac{\mathbf{e}_i^\top \Lambda e^{t\Lambda}\mathbf{e}_4}{1 - \mathbf{e}_i^\top e^{t\Lambda}\mathbf{e}_4}$$

3.4.7 Derivation of the original Basel granularity adjustment

1. We deduce that:

$$\begin{aligned} \mu(x) &= \mathbb{E}[L_i \mid X = x] \\ &= \mathbb{E}[\mathrm{LGD}_i]\,p_i(x) \\ &= E_i p_i(1 + \varpi_i(x-1)) \end{aligned}$$

and:

$$\begin{aligned} \upsilon(x) &= \sigma^2(L_i \mid X = x) \\ &= A_i \\ &= \mathbb{E}^2[\mathrm{LGD}_i]\,p_i(x)(1 - p_i(x)) + \sigma^2(\mathrm{LGD}_i)\,p_i(x) \\ &= \left(\mathbb{E}^2[\mathrm{LGD}_i] + \sigma^2(\mathrm{LGD}_i)\right)p_i(x) - \mathbb{E}^2[\mathrm{LGD}_i]\,p_i^2(x) \end{aligned}$$

If we assume that $p_i^2(x) \approx 0$, it follows that:

$$\begin{aligned} \upsilon(x) &\approx \left(\mathbb{E}^2[\mathrm{LGD}_i] + \sigma^2(\mathrm{LGD}_i)\right)p_i(x) \\ &= \left(E_i^2 + \sigma^2(\mathrm{LGD}_i)\right)p_i(1 + \varpi_i(x-1)) \end{aligned}$$

We have:

$$\begin{aligned} E_i^2 + \sigma^2(\mathrm{LGD}_i) &= E_i^2 + \frac{1}{4}E_i(1 - E_i) \\ &= E_i\left(\frac{1}{4} + \frac{3}{4}E_i\right) \end{aligned}$$

We conclude that:

$$\upsilon(x) = E_i\left(\frac{1}{4} + \frac{3}{4}E_i\right)p_i(1 + \varpi_i(x-1))$$

2. The computation of the derivatives of $\mu(x)$ gives:

$$\frac{\partial\,\mu(x)}{\partial\,x} = E_i p_i \varpi_i$$

and:

$$\frac{\partial^2\,\mu(x)}{\partial\,x^2} = 0$$

For the variance, we obtain:

$$\frac{\partial\,\upsilon(x)}{\partial\,x} = E_i\left(\frac{1}{4}+\frac{3}{4}E_i\right)p_i\varpi_i$$

Since we have:

$$h(x) = \frac{\beta_g^{\alpha_g} x^{\alpha_g-1} e^{-\beta_g x}}{\Gamma(\alpha_g)}$$

and:

$$\ln h(x) = -\ln\Gamma(\alpha_g) + \alpha_g \ln\beta_g + (\alpha_g-1)\ln x - \beta_g x$$

We deduce that:

$$\partial_x \ln h(x) = \frac{(\alpha_g-1)}{x} - \frac{1}{\beta_g}$$

The granularity adjustment function is:

$$\begin{aligned}
\beta(x) &= \frac{1}{2}\upsilon(x)\frac{\partial_x^2\mu(x)}{(\partial_x\mu(x))^2} - \frac{1}{2}\frac{\partial_x\upsilon(x)}{\partial_x\mu(x)} - \frac{1}{2}\upsilon(x)\frac{\partial_x\ln h(x)}{\partial_x\mu(x)} \\
&= -\frac{1}{2}E_i\left(\frac{1}{4}+\frac{3}{4}E_i\right)p_i\varpi_i \times \frac{1}{E_i p_i \varpi_i} - \\
&\quad \frac{1}{2}E_i\left(\frac{1}{4}+\frac{3}{4}E_i\right)p_i\left(1+\varpi_i(x-1)\right)\left(\frac{(\alpha_g-1)}{x}-\frac{1}{\beta_g}\right) \times \frac{1}{E_i p_i \varpi_i} \\
&= -\frac{1}{2}\left(\frac{1}{4}+\frac{3}{4}E_i\right)\left(\left(\frac{\alpha_g-1}{x}-\frac{1}{\beta_g}\right)\left(\frac{1}{\varpi_i}+(x-1)\right)+1\right)
\end{aligned}$$

3. In order to maintain the coherency with the IRB formula, we must have:

$$\Phi\left(\frac{\Phi^{-1}(p_i)+\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) = p_i\left(1+\varpi_i(x-1)\right)$$

This implies that the factor weight ϖ_i is equal to:

$$\varpi_i = \frac{1}{(x-1)}\frac{F_i}{p_i}$$

where:

$$F_i = \Phi\left(\frac{\Phi^{-1}(p_i)+\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) - p_i$$

Finally, we obtain the following expression for $\beta(x)$:

$$\begin{aligned}
\beta(x) &= \frac{1}{2}\left(\frac{1}{4}+\frac{3}{4}E_i\right)\left(-\left(\frac{\alpha_g-1}{x}-\frac{1}{\beta_g}\right)\left(\frac{p_i}{F_i}(x-1)+(x-1)\right)-1\right) \\
&= \frac{1}{2}\left(0.25+0.75E_i\right)\left(A-1+A\frac{p_i}{F_i}\right)
\end{aligned}$$

where:

$$A = -\left(\frac{\alpha_g-1}{x}-\frac{1}{\beta_g}\right)(x-1)$$

4. Since we have $\mathbb{E}[X] = \alpha_g \beta_g$ and $\sigma(X) = \sqrt{\alpha_g}\beta_g$, the parameters of the Gamma distribution are $\alpha = 0.25$ and $\beta = 4$. Since the confidence level α of the value at risk is equal to 99.5%, the quantile of the Gamma distribution $\mathcal{G}(0.25; 4)$ is equal to[7] $x_\alpha = 12.007243$ and the value of A is 3.4393485. We deduce that:

$$\begin{aligned} \beta(x_\alpha) &= \frac{1}{2}(0.25 + 0.75E_i)\left(2.4393485 + 3.4393485\frac{p_i}{F_i}\right) \\ &= (0.4 + 1.2E_i)\left(0.76229640 + 1.0747964\frac{p_i}{F_i}\right) \end{aligned}$$

   We retrieve almost the Basel formula given in BCBS (2001a, §456):

$$\beta(x_\alpha) = (0.4 + 1.2 \times \text{LGD})\left(0.76 + 1.10\frac{p_i}{F_i}\right)$$

   In order to find exactly the Basel formula, we do not use the approximation $p_1^2(x) \approx 0$ for calculating $\upsilon(x)$. In this case, we have:

$$\upsilon(x) = E_i\left(\frac{1}{4} + \frac{3}{4}E_i\right)p_i\left(1 + \varpi_i(x-1)\right) - E_i^2 p_i^2\left(1 + \varpi_i(x-1)\right)^2$$

   and:

$$\frac{\partial\,\upsilon(x)}{\partial\,x} = E_i\left(\frac{1}{4} + \frac{3}{4}E_i\right)p_i\varpi_i - 2E_i^2 p_i^2 \varpi_i\left(1 + \varpi_i(x-1)\right)$$

   We deduce that:

$$\begin{aligned} \beta(x) &= \frac{1}{2}(0.25 + 0.75E_i)\left(-\left(\frac{\alpha_g - 1}{x} - \frac{1}{\beta_g}\right)\left(\varpi_1^{-1} + (x-1)\right) - 1\right) + \\ &\quad \underbrace{E_i p_i\left(1 + \varpi_i(x-1)\right)\left(1 + \frac{1}{2}\varpi_i^{-1}\left(\frac{\alpha_g - 1}{x} - \frac{1}{\beta_g}\right)\right)}_{\text{Correction term}} \end{aligned}$$

   When the expected LGD E_i varies from 5% to 95%, the probability of default p_i varies from 10 bps to 15% and the asset correlation ρ varies from 10% to 30%, the relative error between the exact formula and the approximation is lower than 1%.

5. We deduce that:

$$\begin{aligned} \text{GA} &= \frac{1}{8\%}\left(1.5 \times \frac{\text{EAD}^\star \times \beta(x_\alpha)}{n^\star}\right) - 0.04 \times \text{RWA}_{\text{NR}} \\ &= \frac{\text{EAD}^\star}{n^\star}(0.6 + 1.8 \times E^\star)\left(9.5 + 13.75 \times \frac{p^\star}{F^\star}\right) - 0.04 \times \text{RWA}_{\text{NR}} \\ &= \frac{\text{EAD}^\star}{n^\star} \times \text{GSF} - 0.04 \times \text{RWA}_{\text{NR}} \end{aligned}$$

   where:

$$\text{GSF} = (0.6 + 1.8 \times E^\star)\left(9.5 + 13.75 \times \frac{p^\star}{F^\star}\right)$$

[7] Contrary to the Vasicek model, the conditional probability of default $p_i(X)$ is an increasing function of X in the CreditRisk+ model. Therefore, we have $x_\alpha = \mathbf{H}^{-1}(\alpha)$ and not $x_\alpha = \mathbf{H}^{-1}(1-\alpha)$.

6. Following Wilde (2001b) and Gordy (2003), the portfolio loss is equal to:

$$L = \sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times \mathrm{LGD}_i \times D_i$$

where $\mathcal{C}_j$ is the j^{th} class of risk. The goal is to build an equivalent homogeneous portfolio $w^\star$ such that:

$$L^\star = \mathrm{EAD}^* \times \mathrm{LGD}^* \times D^*$$

First, it is obvious to impose that:

$$\mathrm{EAD}^\star = \sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i$$

Wilde and Gordy also propose to equalize the default rates weighted by exposures:

$$\mathbb{E}\left[\mathrm{EAD}^\star \times D^\star\right] = \mathbb{E}\left[\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times D_i\right]$$

This implies that:

$$\begin{aligned} p^\star &= \frac{\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times p_{\mathcal{C}_j}}{\mathrm{EAD}^*} \\ &= \sum_{j=1}^{n_{\mathcal{C}}} s_{\mathcal{C}_j} \times p_{\mathcal{C}_j} \end{aligned}$$

where $p_{\mathcal{C}_j}$ is the default probability associated to Class $\mathcal{C}_j$ and $s_{\mathcal{C}_j}$ is the corresponding relative exposure:

$$s_{\mathcal{C}_j} = \frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i}{\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i' \in \mathcal{C}_j} \mathrm{EAD}_{i'}}$$

We also have:

$$\begin{aligned} \mathbb{E}\left[\mathrm{EAD}^* \times \mathrm{LGD}^* \times D^*\right] &= \mathbb{E}\left[\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times \mathrm{LGD}_i \times D_i\right] \\ \left(\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i\right) \times E^\star \times p^\star &= \sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times E_i \times p_{\mathcal{C}_j} \\ &= \sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times E_i \end{aligned}$$

Let $E_{\mathcal{C}_j}$ be the average loss given default for Class $\mathcal{C}_j$:

$$E_{\mathcal{C}_j} = \frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i \times E_i}{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i}$$

We deduce that:

$$\left(\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i\right) \times E^\star \times p^\star = \sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} \times E_{\mathcal{C}_j} \times \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i$$

or:

$$\begin{aligned}
E^{\star} &= \sum_{j=1}^{n_{\mathcal{C}}} \frac{p_{\mathcal{C}_j}}{p^{\star}} \times E_{\mathcal{C}_j} \times \frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i}{\sum_{j'=1}^{n_{\mathcal{C}}} \sum_{i' \in \mathcal{C}_{j'}} \mathrm{EAD}_{i'}} \\
&= \sum_{j=1}^{n_{\mathcal{C}}} \frac{s_{\mathcal{C}_j} \times p_{\mathcal{C}_j}}{\sum_{j'=1}^{n_{\mathcal{C}}} s_{\mathcal{C}_{j'}} \times p_{\mathcal{C}_{j'}}} \times E_{\mathcal{C}_j}
\end{aligned}$$

We remind that the conditional variance of the portfolio loss is equal to:

$$\sigma^2(L \mid X) = \sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2 \times \left(E_i^2 p_i(X)(1 - p_i(X)) + \sigma^2(\mathrm{LGD}_i)\, p_i(X)\right)$$

It follows that the expression of the unconditional variance is:

$$\begin{aligned}
\sigma^2(L) &= \sigma^2(\mathbb{E}[L \mid X]) + \mathbb{E}\left[\sigma^2(L \mid X)\right] \\
&= \underbrace{\sigma^2\left(\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i\, E_i p_i(X)\right)}_{\text{Contribution of the systematic risk}} + \\
&\qquad \underbrace{\mathbb{E}\left[\sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2 \left(E_i^2 p_i(X)(1 - p_i(X)) + \sigma^2(\mathrm{LGD}_i)\, p_i(X)\right)\right]}_{\text{Contribution of the idiosyncratic risk}}
\end{aligned}$$

and:

$$\begin{aligned}
\sigma^2(L) &= \underbrace{\sigma^2\left(\sum_{j=1}^{n_{\mathcal{C}}} \mathrm{EAD}^{\star}\, s_{\mathcal{C}_j} E_{\mathcal{C}_j} p_{\mathcal{C}_j}(X)\right)}_{\text{Contribution of the systematic risk}} + \\
&\qquad \underbrace{\sum_{j=1}^{n_{\mathcal{C}}} \left(p_{\mathcal{C}_j}(X)\left(1 - p_{\mathcal{C}_j}(X)\right) - \sigma^2\left(p_{\mathcal{C}_j}(X)\right)\right) \left(\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2\, E_i^2\right)}_{\text{Contribution of the idiosyncratic default risk}} + \\
&\qquad \underbrace{\sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} \left(\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2\, \sigma^2(\mathrm{LGD}_i)\right)}_{\text{Contribution of the idiosyncratic LGD}}
\end{aligned}$$

For the homogenous portfolio $w^{\star}$, we have:

$$\begin{aligned}
\sigma^2(L^{\star}) &= \sigma^2\left(\mathrm{EAD}^{\star}\, E^{\star} p^{\star}(X)\right) + \frac{\left(\mathrm{EAD}^{\star}\right)^2}{n^{\star}} \times \\
&\qquad \left(\left(E^{\star}\right)^2 \left(p^{\star}(1 - p^{\star}) - \sigma^2(p^{\star}(X))\right) + \sigma^2(\mathrm{LGD}^{\star})\, p^{\star}\right)
\end{aligned}$$

In the CreditRisk+ model, we have:

$$p_i(X) = p_i(1 + \varpi_i(X - 1))$$

We have already used the property that $\mathbb{E}\left[p_i\left(X\right)\right] = p_i$, which implies that $\mathbb{E}\left[X\right] = 1$. For the variance, we have:

$$\sigma^2\left(p_i\left(X\right)\right) = p_i^2 \varpi_i^2 \sigma^2\left(X\right)$$

The calibration of the systematic risk implies that the factor weight $\varpi^\star$ is equal to:

$$\begin{aligned}
\varpi^\star &= \frac{\sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} \varpi_{\mathcal{C}_j} E_{\mathcal{C}_j} s_{\mathcal{C}_j}}{p^\star E^\star} \\
&= \frac{\sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} \varpi_{\mathcal{C}_j} E_{\mathcal{C}_j} s_{\mathcal{C}_j}}{\sum_{j=1}^{n_{\mathcal{C}}} p_{\mathcal{C}_j} E_{\mathcal{C}_j} s_{\mathcal{C}_j}}
\end{aligned}$$

For the idiosyncratic default risk, we have:

$$n^\star = \frac{{E^\star}^2 p^\star\left(1-p^\star\right) - \left(p^\star \varpi^\star \sigma\left(X\right)\right)^2}{\sum_{j=1}^{n_{\mathcal{C}}} E_{\mathcal{C}_j}^2 \left(p_{\mathcal{C}_j}\left(1-p_{\mathcal{C}_j}\right) - \left(p_{\mathcal{C}_j} \varpi_{\mathcal{C}_j} \sigma\left(X\right)\right)^2\right) \frac{\sum_{i \in \mathcal{C}_j} \left(\mathrm{EAD}_i\, E_i\right)^2}{\left(\mathrm{EAD}^\star\, E_{\mathcal{C}_j}\right)^2}}$$

We use the following equalities:

$$\begin{aligned}
\frac{\sum_{i \in \mathcal{C}_j} \left(\mathrm{EAD}_i\, E_i\right)^2}{\left(\mathrm{EAD}^\star\, E_{\mathcal{C}_j}\right)^2} &= \frac{\sum_{i \in \mathcal{C}_j} \left(\mathrm{EAD}_i\, E_i\right)^2}{\left(\mathrm{EAD}^\star \frac{\sum_{i \in \mathcal{C}_j} E_i\, \mathrm{EAD}_i}{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i}\right)^2} \\
&= \frac{\sum_{i \in \mathcal{C}_j} \left(\mathrm{EAD}_i\, E_i\right)^2}{\left(\sum_{i \in \mathcal{C}_j} E_i \times \mathrm{EAD}_i\right)^2} \left(\frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i}{\sum_{j'=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_{j'}} \mathrm{EAD}_i}\right)^2 \\
&= H_{\mathcal{C}} s_{\mathcal{C}}^2
\end{aligned}$$

where $H_{\mathcal{C}}$ is the Herfindahl defined by the following expression[8]:

$$H_{\mathcal{C}} = \frac{\sum_{i \in \mathcal{C}} \left(\mathrm{EAD}_i \times E_i\right)^2}{\left(\sum_{i \in \mathcal{C}} \mathrm{EAD}_i \times E_i\right)^2}$$

Finally, the expression of $n^\star$ is equal to:

$$n^\star = \frac{1}{\sum_{j=1}^{n_{\mathcal{C}}} \Lambda_{\mathcal{C}_j} H_{\mathcal{C}_j} s_{\mathcal{C}_j}^2}$$

where:

$$\Lambda_{\mathcal{C}_j} = \frac{E_{\mathcal{C}_j}^2 \left(p_{\mathcal{C}_j}\left(1-p_{\mathcal{C}_j}\right) - \left(p_{\mathcal{C}_j} \varpi_{\mathcal{C}_j} \sigma\left(X\right)\right)^2\right)}{\left(E^\star\right)^2 \left(p^\star\left(1-p^\star\right) - \left(p^\star \varpi^\star \sigma\left(X\right)\right)^2\right)}$$

We retrieve the expression of $n^\star$ given by the Basel Committee (BCBS, 2001a, §445).

[8] We do not obtained the same result than Gordy (2003), who finds that:

$$H_{\mathcal{C}_j} = \frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2}{\left(\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i\right)^2}$$

However, there is a difference between the analysis of Gordy (2003) and the formula $\Lambda_{\mathcal{C}_j}$ proposed by the Basel Committee. In BCBS (2001a), the calibration of $n^\star$ uses both the idiosyncratic default risk and the idiosyncratic loss given default. In this case, we have:

$$\Lambda_{\mathcal{C}_j} = \frac{E^2_{\mathcal{C}_j}\left(p_{\mathcal{C}_j}\left(1-p_{\mathcal{C}_j}\right)-\left(p_{\mathcal{C}_j}\varpi_{\mathcal{C}_j}\sigma\left(X\right)\right)^2\right)+p_{\mathcal{C}_j}\sigma^2\left(\mathrm{LGD}_{\mathcal{C}_j}\right)}{\left(E^\star\right)^2\left(p^\star\left(1-p^\star\right)-\left(p^\star\varpi^\star\sigma\left(X\right)\right)^2\right)+p^\star\sigma^2\left(\mathrm{LGD}^\star\right)}$$

Using the hypothesis of the Basel Committee – $\sigma\left(X\right)=2$, we have:

$$\left(p_i\varpi_i\sigma\left(X\right)\right)^2 = \left(\frac{\sigma\left(X\right)}{\left(x_\alpha-1\right)}F_i\right)^2 = \frac{4F_i^2}{\left(x_\alpha-1\right)^2}$$

For $X\sim\mathcal{G}\left(0.25;4\right)$, we have already shown that $x_\alpha=12.007243$. We obtain:

$$\left(p_i\varpi_i\sigma\left(X\right)\right)^2 = 0.033014360\times F_i^2$$

We remind that:

$$\sigma\left(\mathrm{LGD}_i\right)=\frac{1}{2}\sqrt{E_i\left(1-E_i\right)}$$

Finally, we obtain the expression of the Basel Committee:

$$\Lambda_{\mathcal{C}_j} = \frac{E^2_{\mathcal{C}_j}\left(p_{\mathcal{C}_j}\left(1-p_{\mathcal{C}_j}\right)-0.033\times F^2_{\mathcal{C}_j}\right)+0.25\times p_{\mathcal{C}_j}E_{\mathcal{C}_j}\left(1-E_{\mathcal{C}_j}\right)}{\left(E^\star\right)^2\left(p^\star\left(1-p^\star\right)-0.033\times\left(F^\star\right)^2\right)+0.25\times p^\star E^\star\left(1-E^\star\right)}$$

where:

$$F_{\mathcal{C}_j}=\Phi\left(\frac{\Phi^{-1}\left(p_{\mathcal{C}_j}\right)+\sqrt{\rho}\Phi^{-1}\left(\alpha\right)}{\sqrt{1-\rho}}\right)-p_{\mathcal{C}_j}$$

and $F^\star=\sum_{j=1}^{n_\mathcal{C}} s_{\mathcal{C}_j}F_{\mathcal{C}_j}$.

7. For calculating the granularity adjustment, we proceed in two steps:

   - In the first step, we transform the current portfolio into an equivalent homogenous portfolio:

$$\begin{aligned} s_{\mathcal{C}_j} &= \frac{\sum_{i\in\mathcal{C}_j}\mathrm{EAD}_i}{\sum_{j=1}^{n_\mathcal{C}}\sum_{i\in\mathcal{C}_j}\mathrm{EAD}_i} \\ \mathrm{PD}_{\mathrm{AG}} &= \sum_{j=1}^{n_\mathcal{C}} s_{\mathcal{C}_j}\times\mathrm{PD}_{\mathcal{C}_j} \\ \mathrm{LGD}_{\mathrm{AG}} &= \frac{\sum_{j=1}^{n_\mathcal{C}} s_{\mathcal{C}_j}\times\mathrm{PD}_{\mathcal{C}_j}\times\mathrm{LGD}_{\mathcal{C}_j}}{\sum_{j=1}^{n_\mathcal{C}} s_{\mathcal{C}_j}\times\mathrm{PD}_{\mathcal{C}_j}} \end{aligned}$$

     where $\mathrm{PD}_{\mathcal{C}_j}$ is the default probability of Class $\mathcal{C}_j$ and $\mathrm{LGD}_{\mathcal{C}_j}$ is the average loss given default of Class $\mathcal{C}_j$:

$$\mathrm{LGD}_{\mathcal{C}_j}=\frac{\sum_{i\in\mathcal{C}_j}\mathrm{EAD}_i\times\mathrm{LGD}_i}{\sum_{i\in\mathcal{C}_j}\mathrm{EAD}_i}$$

Then, we calculate

$$F_{\mathrm{AG}} = \sum_{j=1}^{n_{\mathcal{C}}} s_{\mathcal{C}_j} \times F_{\mathcal{C}_j}$$

where $F_{\mathcal{C}_j}$ is the unit unexpected loss of Class $\mathcal{C}_j$:

$$F_{\mathcal{C}_j} = \Phi\left(\frac{\Phi^{-1}\left(\mathrm{PD}_{\mathcal{C}_j}\right) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) - \mathrm{PD}_{\mathcal{C}_j}$$

The equivalent number of loans $n^\star$ is the inverse of the Herfindahl $H^\star$ index:

$$n^\star = \frac{1}{H^\star} = \frac{1}{\sum_{j=1}^{n_{\mathcal{C}}} A_{\mathcal{C}_j} \times H_{\mathcal{C}_j} \times s_{\mathcal{C}_j}^2}$$

where:

$$H_{\mathcal{C}_j} = \frac{\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i^2}{\left(\sum_{i \in \mathcal{C}_j} \mathrm{EAD}_i\right)^2}$$

$A_{\mathcal{C}_j}$ is calculated as follows:

$$A_{\mathcal{C}_j} = \frac{\mathrm{LGD}_{\mathcal{C}_j}^2 \left(B_{\mathcal{C}_j} - 0.033 F_{\mathcal{C}_j}^2\right) + \frac{1}{4} B_{\mathcal{C}_j} \mathrm{LGD}_{\mathcal{C}_j}}{\mathrm{LGD}_{\mathrm{AG}}^2 \left(B_{\mathrm{AG}} - 0.033 F_{\mathrm{AG}}^2\right) + \frac{1}{4} B_{\mathrm{AG}}}$$

where:

$$B_i = \mathrm{PD}_i \left(1 - \mathrm{PD}_i\right)$$

- In the second step, we calculate the granularity scale factor:

$$\mathrm{GSF} = \left(0.6 + 1.8 \times \mathrm{LGD}_{\mathrm{AG}}\right) \times \left(9.5 + 13.75 \times \frac{\mathrm{PD}_{\mathrm{AG}}}{F_{\mathrm{AG}}}\right)$$

Finally, the granularity adjustment is equal to:

$$\mathrm{GA} = \frac{\mathrm{TNRE} \times \mathrm{GSF}}{n^\star} - 0.04 \times \mathrm{RWA}_{\mathrm{NR}}$$

where TNRE is the total non-retail exposure and $\mathrm{RWA}_{\mathrm{NR}}$ is the total non-retail risk-weighted assets.

3.4.8 Variance of the conditional portfolio loss

1. $D_i(X)$ is a Bernoulli random variable with parameter $p_i(X)$. We have $\mathbb{E}[D_i(X)] = p_i(X)$. By definition, the probability distribution of $D_i^2(X)$ is the same than the probability distribution of $D_i(X)$. It follows that $D_i^2(X)$ is a Bernoulli random variable with parameter $p_i(X)$. Since $D_i(X)$ and $D_j(X)$ are independent because the default times are conditionally independent in the Basel II model, we obtain:

$$\begin{aligned} \mathbb{E}\left[D_i(X) D_j(X)\right] &= \mathbb{E}\left[D_i(X)\right] \mathbb{E}\left[D_j(X)\right] \\ &= p_i(X) p_j(X) \end{aligned}$$

2. We have:

$$L(X) = \sum_{i=1}^{n} w_i \,\mathrm{LGD}_i \, D_i(X)$$

3. We have:

$$\begin{aligned}\mathbb{E}\left[L(X)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} w_i \,\mathrm{LGD}_i \, D_i(X)\right] \\ &= \sum_{i=1}^{n} w_i \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[D_i(X)\right] \\ &= \sum_{i=1}^{n} w_i \mathbb{E}\left[\mathrm{LGD}_i\right] p_i(X)\end{aligned}$$

4. We have:

$$\left(\sum_{i=1}^{n} w_i \,\mathrm{LGD}_i \, D_i(X)\right)^2 = \sum_{i=1}^{n} w_i^2 \,\mathrm{LGD}_i^2 \, D_i^2(X) + \sum_{i \neq j} w_i w_i \,\mathrm{LGD}_i \,\mathrm{LGD}_j \, D_i(X) \, D_j(X)$$

and:

$$\begin{aligned}\mathbb{E}\left[L^2(X)\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} w_i \,\mathrm{LGD}_i \, D_i(X)\right)^2\right] \\ &= \sum_{i=1}^{n} w_i^2 \mathbb{E}\left[\mathrm{LGD}_i^2\right] \mathbb{E}\left[D_i^2(X)\right] + \sum_{i \neq j} w_i w_i \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[\mathrm{LGD}_j\right] \mathbb{E}\left[D_i(X)\right] \mathbb{E}\left[D_j(X)\right]\end{aligned}$$

We also have:

$$\begin{aligned}\mathbb{E}^2\left[L(X)\right] &= \left(\sum_{i=1}^{n} w_i \mathbb{E}\left[\mathrm{LGD}_i\right] p_i(X)\right)^2 \\ &= \sum_{i=1}^{n} w_i^2 \mathbb{E}^2\left[\mathrm{LGD}_i\right] p_i^2(X) + \sum_{i \neq j} w_i w_j \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[\mathrm{LGD}_j\right] p_i(X) \, p_j(X)\end{aligned}$$

We deduce that:

$$\begin{aligned}\sigma^2\left(L(X)\right) &= \mathbb{E}\left[L^2(X)\right] - \mathbb{E}^2\left[L(X)\right] \\ &= \sum_{i=1}^{n} w_i^2 \mathbb{E}\left[\mathrm{LGD}_i^2\right] \mathbb{E}\left[D_i^2(X)\right] - \sum_{i=1}^{n} w_i^2 \mathbb{E}^2\left[\mathrm{LGD}_i\right] p_i^2(X) + \\ &\quad \sum_{i \neq j} w_i w_i \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[\mathrm{LGD}_j\right] \mathbb{E}\left[D_i(X)\right] \mathbb{E}\left[D_j(X)\right] - \\ &\quad \sum_{i \neq j} w_i w_j \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[\mathrm{LGD}_j\right] p_i(X) \, p_j(X) \\ &= \sum_{i=1}^{n} w_i^2 \mathbb{E}\left[\mathrm{LGD}_i^2\right] \mathbb{E}\left[D_i^2(X)\right] - \sum_{i=1}^{n} w_i^2 \mathbb{E}^2\left[\mathrm{LGD}_i\right] p_i^2(X) + \\ &\quad \sum_{i \neq j} w_i w_j \mathbb{E}\left[\mathrm{LGD}_i\right] \mathbb{E}\left[\mathrm{LGD}_j\right] \mathrm{cov}\left(D_i(X), D_j(X)\right) \\ &= \sum_{i=1}^{n} w_i^2 \left(\mathbb{E}\left[\mathrm{LGD}_i^2\right] \mathbb{E}\left[D_i^2(X)\right] - \mathbb{E}^2\left[\mathrm{LGD}_i\right] p_i^2(X)\right)\end{aligned}$$

because $\operatorname{cov}(D_i(X), D_j(X)) = 0$. It follows that:

$$\sigma^2(L(X)) = \sum_{i=1}^{n} w_i^2 \left(\mathbb{E}\left[\mathrm{LGD}_i^2\right] p_i(X) - \mathbb{E}^2[\mathrm{LGD}_i] p_i^2(X)\right)$$

If we note $\mathbb{E}\left[\mathrm{LGD}_i^2\right] = \sigma^2(\mathrm{LGD}_i) + \mathbb{E}^2[\mathrm{LGD}_i]$, we obtain:

$$\mathbb{E}\left[\mathrm{LGD}_i^2\right] p_i(X) = \sigma^2(\mathrm{LGD}_i) p_i(X) + \mathbb{E}^2[\mathrm{LGD}_i] p_i(X)$$

and:

$$\sigma^2(L(X)) = \sum_{i=1}^{n} w_i^2 \left(\sigma^2(\mathrm{LGD}_i) p_i(X) + \mathbb{E}^2[\mathrm{LGD}_i] p_i(X)(1 - p_i(X))\right)$$

Another expression is:

$$\sigma^2(L(X)) = \sum_{i=1}^{n} w_i^2 \left(\mathbb{E}[D_i(X)] \sigma^2(\mathrm{LGD}_i) + \mathbb{E}^2[\mathrm{LGD}_i] \sigma^2(D_i(X))\right)$$

because $\mathbb{E}[D_i(X)] = p_i(X)$ and $\sigma^2(D_i(X)) = p_i(X)(1 - p_i(X))$.

Chapter 4

Counterparty Credit Risk and Collateral Risk

4.4.1 Impact of netting agreements in counterparty credit risk

1. (a) Let $\text{MtM}_A(\mathcal{C})$ and $\text{MTM}_B(\mathcal{C})$ be the MtM values of Bank A and Bank B for the contract $\mathcal{C}$. We must theoretically verify that:

$$\begin{aligned}\text{MtM}_{A+B}(\mathcal{C}) &= \text{MTM}_A(\mathcal{C}) + \text{MTM}_B(\mathcal{C}) \\ &= 0 \end{aligned} \tag{4.1}$$

In the case of listed products, the previous relationship is verified. In the case of OTC products, there are no market prices, forcing the bank to use pricing models for the valuation. The MTM value is then a mark-to-model price. Because the two banks do not use the same model with the same parameters, we note a discrepancy between the two mark-to-market prices:

$$\text{MTM}_A(\mathcal{C}) + \text{MTM}_B(\mathcal{C}) \neq 0$$

For instance, we obtain:

$$\begin{aligned}\text{MTM}_{A+B}(\mathcal{C}_1) &= 10 - 11 = -1 \\ \text{MTM}_{A+B}(\mathcal{C}_2) &= -5 + 6 = 1 \\ \text{MTM}_{A+B}(\mathcal{C}_3) &= 6 - 3 = 3 \\ \text{MTM}_{A+B}(\mathcal{C}_4) &= 17 - 12 = 5 \\ \text{MTM}_{A+B}(\mathcal{C}_5) &= -5 + 9 = 4 \\ \text{MTM}_{A+B}(\mathcal{C}_6) &= -5 + 5 = 0 \\ \text{MTM}_{A+B}(\mathcal{C}_7) &= 1 + 1 = 2\end{aligned}$$

Only the contract $\mathcal{C}_6$ satisfies the relationship (4.1).

(b) We have:

$$\text{EAD} = \sum_{i=1}^{7} \max(\text{MTM}(\mathcal{C}_i), 0)$$

We deduce that:

$$\begin{aligned}\text{EAD}_A &= 10 + 6 + 17 + 1 = 34 \\ \text{EAD}_B &= 6 + 9 + 5 + 1 = 21\end{aligned}$$

(c) If there is a global netting agreement, the exposure at default becomes:

$$\text{EAD} = \max\left(\sum_{i=1}^{7} \text{MTM}(\mathcal{C}_i), 0\right)$$

Using the numerical values, we obtain:

$$\begin{aligned}\text{EAD}_A &= \max(10-5+6+17-5-5+1,0)\\ &= \max(19,0)\\ &= 19\end{aligned}$$

and:

$$\begin{aligned}\text{EAD}_B &= \max(-11+6-3-12+9+5+1,0)\\ &= \max(-5,0)\\ &= 0\end{aligned}$$

(d) If the netting agreement only concerns equity contracts, we have:

$$\text{EAD} = \max\left(\sum_{i=1}^{3}\text{MTM}(\mathcal{C}_i),0\right) + \sum_{i=4}^{7}\max(\text{MTM}(\mathcal{C}_i),0)$$

It follows that:

$$\begin{aligned}\text{EAD}_A &= \max(10-5+6,0)+17+1 = 29\\ \text{EAD}_B &= \max(-11+6-3,0)+9+5+1 = 15\end{aligned}$$

2. (a) The potential future exposure $e_1(t)$ is defined as follows:

$$e_1(t) = \max(x_1 + \sigma_1 W_1(t), 0)$$

We deduce that:

$$\begin{aligned}\mathbb{E}[e_1(t)] &= \int_{-\infty}^{\infty}\max(x,0) f(x)\,\mathrm{d}x\\ &= \int_0^{\infty} x f(x)\,\mathrm{d}x\end{aligned}$$

where $f(x)$ is the density function of $\text{MtM}_1(t)$. As we have $\text{MtM}_1(t) \sim \mathcal{N}(x_1, \sigma_1^2 t)$, we deduce that:

$$\mathbb{E}[e_1(t)] = \int_0^{\infty}\frac{x}{\sigma_1\sqrt{2\pi t}}\exp\left(-\frac{1}{2}\left(\frac{x-x_1}{\sigma_1\sqrt{t}}\right)^2\right)\mathrm{d}x$$

With the change of variable $y = \sigma_1^{-1}t^{-1/2}(x-x_1)$, we obtain:

$$\begin{aligned}\mathbb{E}[e_1(t)] &= \int_{\frac{-x_1}{\sigma_1\sqrt{t}}}^{\infty}\frac{x_1+\sigma_1\sqrt{t}y}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)\mathrm{d}y\\ &= x_1\int_{\frac{-x_1}{\sigma_1\sqrt{t}}}^{\infty}\phi(y)\,\mathrm{d}y + \sigma_1\sqrt{t}\int_{\frac{-x_1}{\sigma_1\sqrt{t}}}^{\infty}y\phi(y)\,\mathrm{d}y\\ &= x_1\Phi\left(\frac{x_1}{\sigma_1\sqrt{t}}\right) + \sigma_1\sqrt{t}\Big[-\phi(y)\Big]_{\frac{-x_1}{\sigma_1\sqrt{t}}}^{\infty}\\ &= x_1\Phi\left(\frac{x_1}{\sigma_1\sqrt{t}}\right) + \sigma_1\sqrt{t}\phi\left(\frac{x_1}{\sigma_1\sqrt{t}}\right)\end{aligned}$$

because $\phi(-x) = \phi(x)$ and $\Phi(-x) = 1-\Phi(x)$.

(b) When there is no netting agreement, we have:

$$e(t) = e_1(t) + e_2(t)$$

We deduce that:

$$\begin{aligned} \mathbb{E}[e(t)] &= \mathbb{E}[e_1(t)] + \mathbb{E}[e_2(t)] \\ &= x_1 \Phi\left(\frac{x_1}{\sigma_1 \sqrt{t}}\right) + \sigma_1 \sqrt{t} \phi\left(\frac{x_1}{\sigma_1 \sqrt{t}}\right) + \\ &\quad x_2 \Phi\left(\frac{x_2}{\sigma_2 \sqrt{t}}\right) + \sigma_2 \sqrt{t} \phi\left(\frac{x_2}{\sigma_2 \sqrt{t}}\right) \end{aligned}$$

(c) In the case of a netting agreement, the potential future exposure becomes:

$$\begin{aligned} e(t) &= \max\left(\mathrm{MtM}_1(t) + \mathrm{MtM}_2(t), 0\right) \\ &= \max\left(\mathrm{MtM}_{1+2}(t), 0\right) \\ &= \max\left(x_1 + x_2 + \sigma_1 W_1(t) + \sigma_2 W_2(t), 0\right) \end{aligned}$$

We deduce that:

$$\mathrm{MtM}_{1+2}(t) \sim \mathcal{N}\left(x_1 + x_2, \left(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right) t\right)$$

Using results of Question 2(a), we finally obtain:

$$\begin{aligned} \mathbb{E}[e(t)] &= (x_1 + x_2) \Phi\left(\frac{x_1 + x_2}{\sqrt{\left(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right) t}}\right) + \\ &\quad \sqrt{\left(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right) t}\, \phi\left(\frac{x_1 + x_2}{\sqrt{\left(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right) t}}\right) \end{aligned}$$

(d) We have represented the expected exposure $\mathbb{E}[e(t)]$ in Figure 4.1 when $x_1 = x_2 = 0$ and $\sigma_1 = \sigma_2$. We note that it is an increasing function of the time t and the volatility σ. We also observe that the netting agreement may have a big impact, especially when the correlation is low or negative.

4.4.2 Calculation of the effective expected positive exposure

1. We have $e(t) = \max(\mathrm{MTM}(t), 0)$ where $\mathrm{MTM}(t)$ is the mark-to-market price of the OTC contract at the future date t. We note $\mathbf{F}_{[0,t]}$ the cumulative distribution function of the random variable $e(t)$. The peak exposure is the quantile α of $\mathbf{F}_{[0,t]}$:

$$\mathrm{PE}_\alpha(t) = \mathbf{F}_{[0,t]}^{-1}(\alpha)$$

The maximum peak exposure is the maximum value of $\mathrm{PE}_\alpha(t)$:

$$\mathrm{MPE}_\alpha(0; t) = \sup_s \mathrm{PE}_\alpha(s)$$

The expected exposure is the average of the potential future exposure:

$$\mathrm{EE}(t) = \mathbb{E}[e(t)] = \int x \,\mathrm{d}\mathbf{F}_{[0,t]}(x)$$

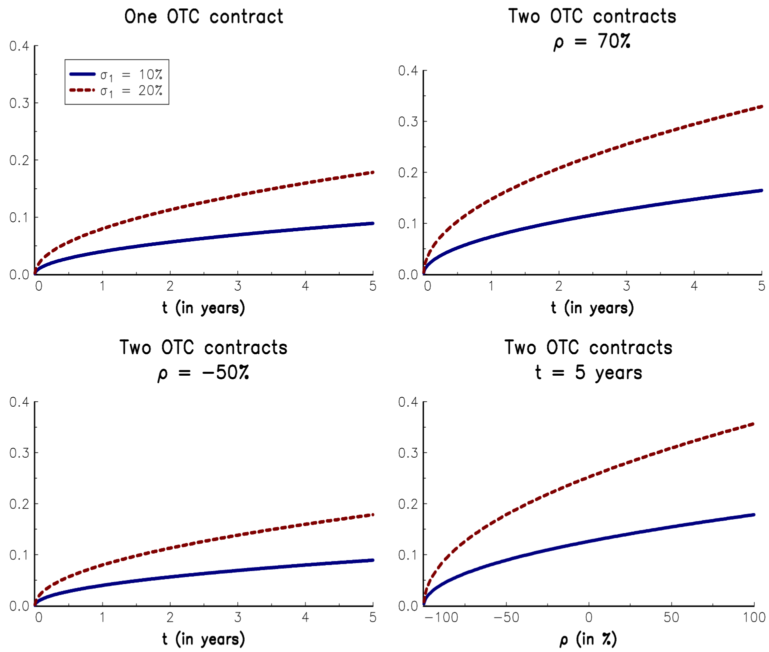

FIGURE 4.1: Expected exposure $\mathbb{E}[e(t)]$ when there is a netting agreement

We define the expected positive exposure as the weighted average over time of the expected exposure for a given holding period $[0, t]$:

$$\text{EPE}(0; t) = \frac{1}{t} \int_0^t \text{EE}(s) \, \mathrm{d}s$$

The effective expected exposure is the maximum expected exposure which occurs before the date t:

$$\begin{aligned} \text{EEE}(t) &= \sup_{s \leq t} \text{EE}(s) \\ &= \max\left(\text{EEE}\left(t^-\right), \text{EE}(t)\right) \end{aligned}$$

The effective expected positive exposure is the weighted average of effective expected exposure for a given time period $[0, t]$:

$$\text{EEPE}(0; t) = \frac{1}{t} \int_0^t \text{EEE}(s) \, \mathrm{d}s$$

2. We have:

$$\begin{aligned} \mathbf{F}_{[0,t]}(x) &= \Pr\{e(t) \leq x\} \\ &= \Pr\left\{\sigma\sqrt{t}X \leq x\right\} \\ &= \Pr\left\{X \leq \frac{x}{\sigma\sqrt{t}}\right\} \\ &= \frac{x}{\sigma\sqrt{t}} \end{aligned}$$

with $x \in \left[0, \sigma\sqrt{t}\right]$. We deduce that:

$$\begin{aligned}
\mathrm{PE}_{\alpha}(0;t) &= \alpha\sigma\sqrt{t} \\
\mathrm{MPE}_{\alpha}(0;T) &= \alpha\sigma\sqrt{T} \\
\mathrm{EE}(t) &= \int_0^{\sigma\sqrt{t}} x\frac{1}{\sigma\sqrt{t}}\,\mathrm{d}x = \frac{\sigma\sqrt{t}}{2} \\
\mathrm{EPE}(0;t) &= \frac{1}{t}\int_0^t \frac{\sigma\sqrt{s}}{2}\,\mathrm{d}s = \frac{\sigma\sqrt{t}}{3} \\
\mathrm{EEE}(t) &= \frac{\sigma\sqrt{t}}{2} \\
\mathrm{EEPE}(0;t) &= \frac{1}{t}\int_0^t \frac{\sigma\sqrt{s}}{2}\,\mathrm{d}s = \frac{\sigma\sqrt{t}}{3}
\end{aligned}$$

3. We have:

$$\begin{aligned}
\mathbf{F}_{[0,t]}(x) &= \Pr\left\{e^{\sigma\sqrt{t}X} \leq x\right\} \\
&= \Phi\left(\frac{\ln x}{\sigma\sqrt{t}}\right)
\end{aligned}$$

with $x \in [0, \infty]$. We deduce that:

$$\begin{aligned}
\mathrm{PE}_{\alpha}(t) &= \exp\left(\sigma\Phi^{-1}(\alpha)\sqrt{t}\right) \\
\mathrm{MPE}_{\alpha}(0;T) &= \exp\left(\sigma\Phi^{-1}(\alpha)\sqrt{T}\right) \\
\mathrm{EE}(t) &= \exp\left(\frac{1}{2}\sigma^2 t\right) \\
\mathrm{EPE}(0;t) &= \left(\exp\left(\frac{1}{2}\sigma^2 t\right) - 1\right) \Big/ \left(\frac{1}{2}\sigma^2 t\right) \\
\mathrm{EEE}(t) &= \exp\left(\frac{1}{2}\sigma^2 t\right) \\
\mathrm{EEPE}(0;t) &= \left(\exp\left(\frac{1}{2}\sigma^2 t\right) - 1\right) \Big/ \left(\frac{1}{2}\sigma^2 t\right)
\end{aligned}$$

4. We have:

$$\mathbf{F}_{[0,t]}(x) = \frac{x}{\sigma\left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right)}$$

with $x \in \left[0, \sigma\left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right)\right]$. We deduce that:

$$\begin{aligned}
\mathrm{PE}_{\alpha}(0) &= \alpha\sigma\left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right) \\
\mathrm{MPE}_{\alpha}(0;t) &= \mathbb{1}\{t < t^\star\} \times \mathrm{PFE}_{\alpha}(0;t) + \mathbb{1}\{t \geq t^\star\} \times \mathrm{PFE}_{\alpha}(0;t^\star) \\
\mathrm{EE}(t) &= \frac{1}{2}\sigma\left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right) \\
\mathrm{EPE}(0;t) &= \sigma\left(\frac{9t^3 - 28Tt^2 + 24T^2 t}{72}\right) \\
\mathrm{EEE}(t) &= \mathbb{1}\{t < t^\star\} \times \mathrm{EE}(t) + \mathbb{1}\{t \geq t^\star\} \times \mathrm{EE}(t^\star)
\end{aligned}$$

$$\text{EEPE}(0;t) = \frac{1}{t}\int_0^t \text{EEE}(s)\, \mathrm{d}s$$

with:

$$t^\star = \left(\frac{7-\sqrt{13}}{9}\right) T$$

This question is more difficult than the previous ones, because $e(t)$ is not a monotonically increasing function. It is increasing when $t < t_1^\star$ and then decreasing[1]. This explains that $\text{MPE}_\alpha(0;t)$ and $\text{EEE}(t)$ depends on the parameter $t^\star$.

5. The cumulative distribution function of X is:

$$\begin{aligned}
\mathbf{F}(x) &= \Pr\{X \le x\} \\
&= \int_0^x \frac{u^a}{a+1} \mathrm{d}u \\
&= x^{a+1}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\mathbf{F}_{[0,t]}(x) &= \Pr\{e(t) \le x\} \\
&= \Pr\left\{\sigma\sqrt{t}X \le x\right\} \\
&= \Pr\left\{X \le \frac{x}{\sigma\sqrt{t}}\right\} \\
&= \left(\frac{x}{\sigma\sqrt{t}}\right)^{a+1}
\end{aligned}$$

and:

$$f_{[0,t]}(x) = \frac{(a+1)\,x^a}{\left(\sigma\sqrt{t}\right)^{a+1}}$$

It follows that:

$$\text{PE}_\alpha(t) = \alpha^{1/(a+1)}\sigma\sqrt{t}$$

and:

$$\text{MPE}_\alpha(0;T) = \alpha^{1/(a+1)}\sigma\sqrt{T}$$

The expected exposure is:

$$\text{EE}(t) = \int_0^{\sigma\sqrt{t}} x\frac{(a+1)\,x^a}{\left(\sigma\sqrt{t}\right)^{a+1}}\,\mathrm{d}e = \frac{(a+1)\,\sigma\sqrt{t}}{a+2}$$

We deduce that:

$$\text{EEE}(t) = \frac{(a+1)\,\sigma\sqrt{t}}{a+2}$$

and:

$$\text{EEPE}(0;t) = \frac{1}{t}\int_0^t \frac{(a+1)\,\sigma\sqrt{s}}{a+2}\,\mathrm{d}s = \frac{2\,(a+1)\,\sigma\sqrt{t}}{3\,(a+2)}$$

[1] In fact, there is a second root:

$$t_2^\star = \left(\frac{7+\sqrt{13}}{9}\right) T$$

We observe that $e(t)$ can take a negative value when t is in the neighborhood of this solution. We ignore this problem to calculate the different measures.

6. In Figures 4.2 and 4.3, we have reported the functions $\mathrm{EE}(t)$, $\mathrm{EPE}(0;t)$, $\mathrm{EEE}(t)$ and $\mathrm{EEPE}(0;t)$ for the two exposures given in Questions 3 and 5. We notice that the second exposure has the profile of an amortizing swap where the first exposure is more like an option profile.

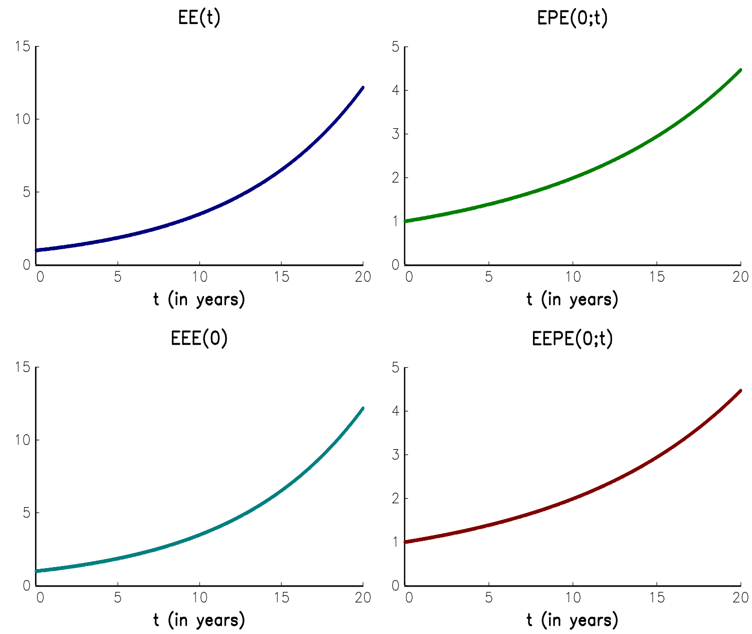

FIGURE 4.2: Credit exposure when $e(t) = \exp\left(\sigma\sqrt{t}\mathcal{N}(0,1)\right)$

4.4.3 Calculation of the required capital for counterparty credit risk

1. We have:

$$\begin{aligned} \mathbf{F}_{[0,t]}(x) &= \Pr\{e(t) \leq x\} \\ &= \Pr\left\{N\sigma\sqrt{t}U \leq x\right\} \\ &= \Pr\left\{U \leq \frac{x}{N\sigma\sqrt{t}}\right\} \\ &= \left(\frac{x}{N\sigma\sqrt{t}}\right)^{\gamma} \end{aligned}$$

with $x \in \left[0, N\sigma\sqrt{t}\right]$. We deduce that:

$$\begin{aligned} \mathrm{PE}_{\alpha}(t) &= \mathbf{F}_{[0,t]}^{-1}(\alpha) \\ &= N\sigma\sqrt{t}\alpha^{1/\gamma} \end{aligned}$$

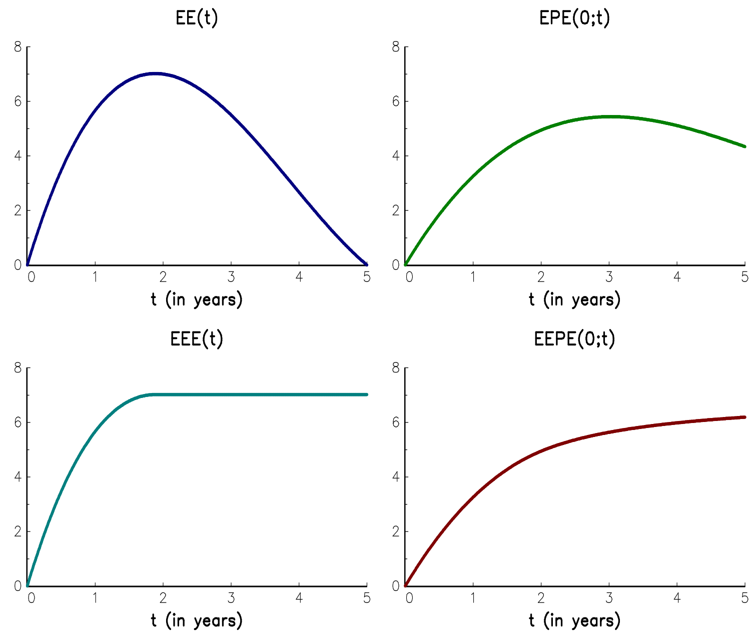

FIGURE 4.3: Credit exposure when $e(t) = \sigma\left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right)\mathcal{U}_{[0,1]}$

For the expected exposure, we obtain:

$$\begin{aligned}\mathrm{EE}(t) &= \mathbb{E}[e(t)] \\ &= \int_0^{N\sigma\sqrt{t}} x \frac{\gamma}{\left(N\sigma\sqrt{t}\right)^\gamma} x^{\gamma-1}\,\mathrm{d}x \\ &= \frac{\gamma}{\left(N\sigma\sqrt{t}\right)^\gamma}\left[\frac{x^{\gamma+1}}{\gamma+1}\right]_0^{N\sigma\sqrt{t}} \\ &= \frac{\gamma}{\gamma+1} N\sigma\sqrt{t}\end{aligned}$$

We deduce that:

$$\mathrm{EEE}(t) = \frac{\gamma}{\gamma+1} N\sigma\sqrt{t}$$

and:

$$\begin{aligned}\mathrm{EEPE}(0;t) &= \frac{1}{t}\int_0^t \mathrm{EEE}(s)\,\mathrm{d}s \\ &= \frac{1}{t}\int_0^t \frac{\gamma}{\gamma+1} N\sigma\sqrt{s}\,\mathrm{d}s \\ &= \frac{\gamma}{\gamma+1} N\sigma \frac{1}{t}\left[\frac{2}{3}s^{3/2}\right]_0^t \\ &= \frac{2\gamma}{3(\gamma+1)} N\sigma\sqrt{t}\end{aligned}$$

2. (a) When the bank uses an internal model, the regulatory exposure at default is:

$$\mathrm{EAD} = \alpha \times \mathrm{EEPE}(0;1)$$

Using the standard value $\alpha = 1.4$, we obtain:

$$\begin{aligned}\text{EAD} &= 1.4 \times \frac{4}{9} \times 3 \times 10^6 \times 0.20 \\ &= \$373\,333\end{aligned}$$

(b) While the bank uses the FIRB approach, the required capital is:

$$\mathcal{K} = \text{EAD} \times \mathbb{E}\left[\text{LGD}\right] \times \left(\Phi\left(\frac{\Phi^{-1}(\text{PD}) + \sqrt{\rho}\Phi^{-1}(99.9\%)}{\sqrt{1-\rho}}\right) - \text{PD}\right)$$

When ρ is equal to 20%, we have:

$$\begin{aligned}\frac{\Phi^{-1}(\text{PD}) + \sqrt{\rho}\Phi^{-1}(99.9\%)}{\sqrt{1-\rho}} &= \frac{-2.33 + \sqrt{0.20} \times 3.09}{\sqrt{1-0.20}} \\ &= -1.06\end{aligned}$$

By using the approximations $-1.06 \simeq 1$ and $\Phi(-1) \simeq 0.16$, we obtain:

$$\begin{aligned}\mathcal{K} &= 373\,333 \times 0.70 \times (0.16 - 0.01) \\ &= \$39\,200\end{aligned}$$

The required capital of this OTC product for counterparty credit risk is then equal to \$39 200.

4.4.4 Calculation of CVA and DVA measures

1. The positive exposure $e^+(t)$ is the maximum between zero and the mark-to-market value:

$$\begin{aligned}e^+(t) &= \max(0, \text{MtM}(t)) \\ &= \max\left(0, N\sigma\sqrt{t}X\right)\end{aligned}$$

We have:

$$\begin{aligned}\mathbf{F}_{[0,t]}(x) &= \Pr\left\{e^+(t) \le x\right\} \\ &= \Pr\left\{\max\left(0, N\sigma\sqrt{t}X\right) \le x\right\}\end{aligned}$$

We notice that:

$$\max\left(0, N\sigma\sqrt{t}X\right) = \begin{cases} 0 & \text{if } X \le 0 \\ N\sigma\sqrt{t}X & \text{otherwise}\end{cases}$$

By assuming that $x \in \left[0, N\sigma\sqrt{t}\right]$, we deduce that:

$$\begin{aligned}\mathbf{F}_{[0,t]}(x) &= \Pr\left\{e^+(t) \le x, X \le 0\right\} + \Pr\left\{e^+(t) \le x, X > 0\right\} \\ &= \Pr\left\{0 \le x, X \le 0\right\} + \Pr\left\{N\sigma\sqrt{t}X \le x, X > 0\right\} \\ &= \frac{1}{2} + \frac{1}{2}\Pr\left\{N\sigma\sqrt{t}U \le x\right\} \\ &= \frac{1}{2} + \frac{1}{2}\Pr\left\{U \le \frac{x}{N\sigma\sqrt{t}}\right\}\end{aligned}$$

where U is the standard uniform random variable. We finally obtain the following expression:

$$\mathbf{F}_{[0,t]}(x) = \frac{1}{2} + \frac{x}{2N\sigma\sqrt{t}}$$

If $x \leq 0$ or $x \geq N\sigma\sqrt{t}$, it is easy to show that $\mathbf{F}_{[0,t]}(x) = 0$ and $\mathbf{F}_{[0,t]}(x) = 1$.

2. The expected positive exposure $\text{EpE}(t)$ is defined as follows:

$$\text{EpE}(t) = \mathbb{E}\left[e^{+}(t)\right]$$

Using the expression of $\mathbf{F}_{[0,t]}(x)$, it follows that the density function of $e^{+}(t)$ is equal to:

$$\begin{aligned} f_{[0,t]}(x) &= \frac{\partial\,\mathbf{F}_{[0,t]}(x)}{\partial\,x} \\ &= \frac{1}{2N\sigma\sqrt{t}} \end{aligned}$$

We deduce that:

$$\begin{aligned} \text{EpE}(t) &= \int_0^{N\sigma\sqrt{t}} x f_{[0,t]}(x)\,\mathrm{d}x \\ &= \int_0^{N\sigma\sqrt{t}} \frac{x}{2N\sigma\sqrt{t}}\,\mathrm{d}x \\ &= \left[\frac{x^2}{4N\sigma\sqrt{t}}\right]_0^{N\sigma\sqrt{t}} \\ &= \frac{N\sigma\sqrt{t}}{4} \end{aligned}$$

3. By definition, we have:

$$\text{CVA} = (1-\boldsymbol{\mathcal{R}}_B) \times \int_0^T -B_0(t)\,\text{EpE}(t)\,\mathrm{d}\mathbf{S}_B(t)$$

4. The interest rates are equal to zero meaning that $B_0(t) = 1$. Moreover, we have $\mathbf{S}_B(t) = e^{-\lambda_B t}$. We deduce that:

$$\begin{aligned} \text{CVA} &= (1-\boldsymbol{\mathcal{R}}_B) \times \int_0^T \frac{N\sigma\sqrt{t}}{4}\lambda_B e^{-\lambda_B t}\,\mathrm{d}t \\ &= \frac{N\lambda_B(1-\boldsymbol{\mathcal{R}}_B)\,\sigma}{4}\int_0^T \sqrt{t}e^{-\lambda_B t}\,\mathrm{d}t \end{aligned}$$

The definition of the incomplete gamma function is:

$$\gamma(s,x) = \int_0^x t^{s-1}e^{-t}\,\mathrm{d}t$$

By considering the change of variable $y = \lambda_B t$, we obtain:

$$\begin{aligned} \int_0^T \sqrt{t}e^{-\lambda_B t}\,\mathrm{d}t &= \int_0^{\lambda_B T} \sqrt{\frac{y}{\lambda_B}}e^{-y}\,\frac{\mathrm{d}y}{\lambda_B} \\ &= \frac{1}{\lambda_B^{3/2}}\int_0^{\lambda_B T} y^{3/2-1}e^{-y}\,\mathrm{d}y \\ &= \frac{\gamma\left(\frac{3}{2},\lambda_B T\right)}{\lambda_B^{3/2}} \end{aligned}$$

It follows that:

$$\mathrm{CVA} = \frac{N\left(1-\boldsymbol{\mathcal{R}}_B\right)\sigma\gamma\left(\frac{3}{2},\lambda_B T\right)}{4\sqrt{\lambda_B}}$$

5. The CVA is proportional to the notional N of the OTC contract, the loss given default $(1-\boldsymbol{\mathcal{R}}_B)$ of the counterparty and the volatility σ of the underlying asset. It is an increasing function of the maturity T because we have $\gamma\left(\frac{3}{2},\lambda_B T_2\right) > \gamma\left(\frac{3}{2},\lambda_B T_1\right)$ when $T_2 > T_1$. If the maturity is not very large (less than 10 years), the CVA is an increasing function of the default intensity λ_B. The limit cases are[2]:

$$\begin{aligned}\lim_{\lambda_B\to\infty}\mathrm{CVA} &= \lim_{\lambda_B\to\infty}\frac{N\left(1-\boldsymbol{\mathcal{R}}_B\right)\sigma\gamma\left(\frac{3}{2},\lambda_B T\right)}{4\sqrt{\lambda_B}} \\ &= 0\end{aligned}$$

and:

$$\lim_{T\to\infty}\mathrm{CVA} = \frac{N\left(1-\boldsymbol{\mathcal{R}}_B\right)\sigma\Gamma\left(\frac{3}{2}\right)}{4\sqrt{\lambda_B}}$$

When the counterparty has a high default intensity, meaning that the default is imminent, the CVA is equal to zero because the mark-to-market value is close to zero. When the maturity is large, the CVA is a decreasing function of the intensity λ_B. Indeed, the probability to observe a large mark-to-market in the future increases when the default time is very far from the current date. We have illustrated these properties in Figure 4.4 with the following numerical values: $N = \$1$ mn, $\boldsymbol{\mathcal{R}}_B = 40\%$ and $\sigma = 30\%$.

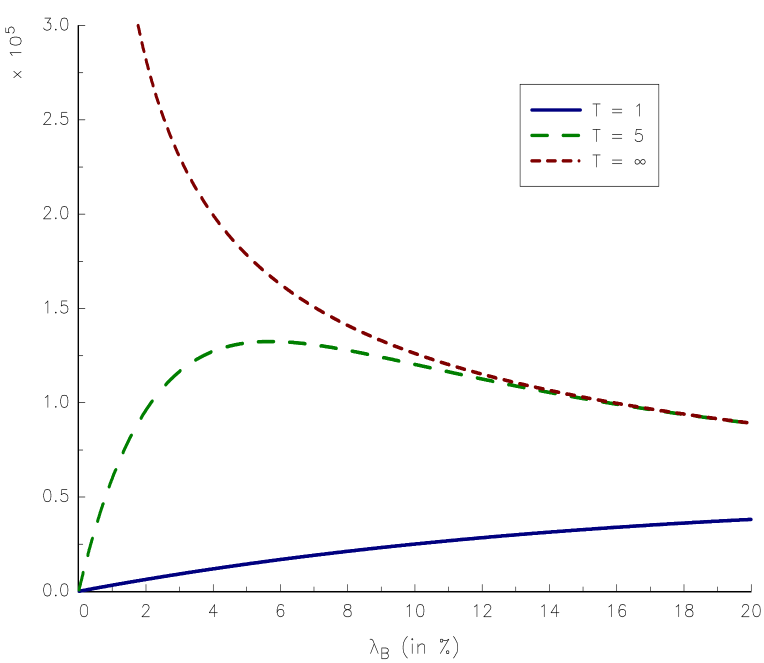

FIGURE 4.4: Evolution of the CVA with respect to maturity T and intensity λ_B

6. We notice that the mark-to-market is perfectly symmetric about 0. We deduce that

[2]We have $\lim_{x\to\infty}\gamma(s,x) = \Gamma(s)$.

the expected negative exposure $\operatorname{EnE}(t)$ is equal to the expected positive exposure $\operatorname{EpE}(t)$. It follows that the DVA is equal to:

$$\mathrm{DVA} = \frac{N\left(1-\boldsymbol{\mathcal{R}}_A\right)\sigma\gamma\left(\frac{3}{2},\lambda_A T\right)}{4\sqrt{\lambda_A}}$$

4.4.5 Approximation of the CVA for an interest rate swap

1. We have:

$$Ae^x - B \geq 0 \Leftrightarrow x \geq x^\star = \ln B - \ln A$$

It follows that:

$$\begin{aligned}
\mathbb{E}\left[\max\left(Ae^X - B, 0\right)\right] &= \int_{-\infty}^{\infty} \max\left(Ae^x - B, 0\right) \frac{1}{\sigma_X}\phi\left(\frac{x-\mu_X}{\sigma_X}\right) \mathrm{d}x \\
&= A\int_{x^\star}^{\infty} e^x \frac{1}{\sigma_X}\phi\left(\frac{x-\mu_X}{\sigma_X}\right) \mathrm{d}x - \\
&\quad B\int_{x^\star}^{\infty} \frac{1}{\sigma_X}\phi\left(\frac{x-\mu_X}{\sigma_X}\right) \mathrm{d}x
\end{aligned}$$

By considering the change of variable $y = \sigma_X^{-1}\left(x - \mu_X\right)$, we deduce that:

$$\operatorname{EpE}(t) = A\int_{y^\star}^{\infty} e^{\mu_X + \sigma_X y}\phi(y)\,\mathrm{d}y - B\int_{y^\star}^{\infty}\phi(y)\,\mathrm{d}y$$

where $y^\star = \sigma_X^{-1}\left(x^\star - \mu_X\right)$. We have:

$$\begin{aligned}
A\int_{y^\star}^{\infty} e^{\mu_X + \sigma_X y}\phi(y)\,\mathrm{d}y &= Ae^{\mu_X}\int_{y^\star}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2 + \sigma_X y}\,\mathrm{d}y \\
&= Ae^{\mu_X + \frac{1}{2}\sigma_X^2}\int_{y^\star}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\sigma_X)^2}\,\mathrm{d}y \\
&= Ae^{\mu_X + \frac{1}{2}\sigma_X^2}\int_{y^\star - \sigma_X}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,\mathrm{d}z \\
&= Ae^{\mu_X + \frac{1}{2}\sigma_X^2}\left(1 - \Phi\left(y^\star - \sigma_X\right)\right) \\
&= Ae^{\mu_X + \frac{1}{2}\sigma_X^2}\Phi\left(\frac{\mu_X + \sigma_X^2 + \ln A - \ln B}{\sigma_X}\right)
\end{aligned}$$

and:

$$\begin{aligned}
B\int_{y^\star}^{\infty}\phi(y)\,\mathrm{d}y &= B\Phi\left(-y^\star\right) \\
&= B\Phi\left(\frac{\mu_X + \ln A - \ln B}{\sigma_X}\right)
\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
\operatorname{EpE}(t) &= Ae^{\mu_X + \frac{1}{2}\sigma_X^2}\Phi\left(\frac{\mu_X + \sigma_X^2 + \ln A - \ln B}{\sigma_X}\right) - \\
&\quad B\Phi\left(\frac{\mu_X + \ln A - \ln B}{\sigma_X}\right)
\end{aligned}$$

2. The mark-to-market is an approximation of a fixed-float IRS in continuous time by assuming that the floating leg is constant, implying that the term structure of the float rate is flat (Syrkin and Shirazi, 2015). The first term of the mark-to-market is the floating leg, because the cash flows change with the time t, whereas the second term is the fixed leg[3]:

$$\operatorname{MtM}(t) = \underbrace{N \int_t^T f(t,T)\, B_t(s)\, \mathrm{d}s}_{\text{Floating leg}} - \underbrace{N \int_t^T f(0,T)\, B_t(s)\, \mathrm{d}s}_{\text{Fixed leg}}$$

Since the instantaneous forward rate follows a geometric Brownian motion, we deduce that:

$$f(0,T) = f(0,T)\, e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)}$$

and:

$$f(0,T) \sim \mathcal{LN}\left(\ln f(0,T) + \left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$$

We also have:

$$\begin{aligned} \varphi(t,T) &= \int_t^T B_t(s)\, \mathrm{d}s \\ &= \int_t^T e^{-r(s-t)}\, \mathrm{d}s \\ &= \left[-\frac{e^{-r(s-t)}}{r}\right]_t^T \\ &= \frac{1 - e^{-r(T-t)}}{r} \end{aligned}$$

It follows that:

$$\begin{aligned} \operatorname{MtM}(t) &= N\left(f(t,T) - f(0,T)\right) \int_t^T B_t(s)\, \mathrm{d}s \\ &= N f(0,T)\, \varphi(t,T) \left(e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)} - 1\right) \end{aligned}$$

The confidence interval of $\operatorname{MtM}(t)$ with confidence level α is defined by:

$$\operatorname{MtM}(t) \in \left[q_-(t;\alpha), q_-(t;\alpha)\right]$$

where:

$$q_\pm(t;\alpha) = N f(0,T)\, \varphi(t,T) \left(e^{\left(\mu - \frac{1}{2}\sigma^2\right)t \pm \sigma\sqrt{t}\Phi^{-1}\left(\frac{1-\alpha}{2}\right)} - 1\right)$$

3. For the expected mark-to-market, we have:

$$\begin{aligned} \mathbb{E}\left[\operatorname{MtM}(t)\right] &= N f(0,T)\, \varphi(t,T) \left(e^{\left(\mu - \frac{1}{2}\sigma^2\right)t} \mathbb{E}\left[e^{\sigma W(t)}\right] - 1\right) \\ &= N f(0,T)\, \varphi(t,T) \left(e^{\left(\mu - \frac{1}{2}\sigma^2\right)t} e^{\sigma^2 t} - 1\right) \\ &= N f(0,T)\, \varphi(t,T) \left(e^{\mu t} - 1\right) \end{aligned}$$

[3] $f(0,T)$ is known at time $t = 0$.

For the expected counterparty exposure, we have:

$$\begin{aligned}
\mathbb{E}\left[e\left(t\right)\right] &= \mathbb{E}\left[\max\left(f\left(0,T\right)\left(e^{\left(\mu-\frac{1}{2}\sigma^2\right)t+\sigma W(t)}-1\right),0\right)\right] \\
&= \mathbb{E}\left[\max\left(\left(f\left(0,T\right)e^{\left(\mu-\frac{1}{2}\sigma^2\right)t}e^{\sigma W(t)}-f\left(0,T\right)\right),0\right)\right] \\
&= \mathbb{E}\left[\max\left(Ae^{X}-B,0\right)\right]
\end{aligned}$$

where $A = Nf\left(0,T\right)\varphi\left(t,T\right)e^{\left(\mu-\frac{1}{2}\sigma^2\right)t}$, $B = Nf\left(0,T\right)\varphi\left(t,T\right)$ and $X \sim \mathcal{N}\left(0,\sigma^2 t\right)$. Since $\ln A - \ln B = \left(\mu - \frac{1}{2}\sigma^2\right)t$, we obtain:

$$\begin{aligned}
\mathrm{EpE}\left(t\right) &= Ae^{\frac{1}{2}\sigma^2 t}\Phi\left(\frac{\sigma^2 t + \ln A - \ln B}{\sigma\sqrt{t}}\right) - B\Phi\left(\frac{\ln A - \ln B}{\sigma\sqrt{t}}\right) \\
&= Nf\left(0,T\right)\varphi\left(t,T\right)\left(e^{\mu t}\Phi\left(\delta\left(t\right)\right) - \Phi\left(\delta\left(t\right) - \sigma\sqrt{t}\right)\right)
\end{aligned}$$

where:

$$\delta\left(t\right) = \left(\frac{\mu}{\sigma} + \frac{1}{2}\sigma\right)\sqrt{t}$$

4. We have:

$$\begin{aligned}
\mathrm{CVA}\left(t\right) &= \left(1-\boldsymbol{\mathcal{R}}\right)\times\int_t^T -B_t\left(u\right)\,\mathrm{EpE}\left(u\right)\,\mathrm{d}\mathbf{S}\left(u\right) \\
&= \left(1-\boldsymbol{\mathcal{R}}\right)\times\int_t^T \lambda e^{-(r+\lambda)(u-t)}\,\mathrm{EpE}\left(u\right)\,\mathrm{d}u \\
&= \boldsymbol{s}\times\int_t^T e^{-(r+\lambda)(u-t)}\,\mathrm{EpE}\left(u\right)\,\mathrm{d}u
\end{aligned}$$

where $\boldsymbol{s}$ is the credit spread of the counterparty.

5. Syrkin and Shirazi (2015) propose the following approximations: $e^{\mu t} - 1 \approx \mu t$,

$$\Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) \approx \Phi\left(\frac{\mu\sqrt{t}}{\sigma}\right)$$

and:

$$\Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) - \Phi\left(\left(\frac{\mu}{\sigma}-\frac{1}{2}\sigma\right)\sqrt{t}\right) \approx \sigma\sqrt{t}\phi\left(\frac{\mu\sqrt{t}}{\sigma}\right)$$

We have:

$$\begin{aligned}
(*) &= e^{\mu t}\Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) - \Phi\left(\left(\frac{\mu}{\sigma}-\frac{1}{2}\sigma\right)\sqrt{t}\right) \\
&= \left(e^{\mu t}-1\right)\Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) + \\
&\quad \Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) - \Phi\left(\left(\frac{\mu}{\sigma}-\frac{1}{2}\sigma\right)\sqrt{t}\right) \\
&\approx \mu t\Phi\left(\frac{\mu\sqrt{t}}{\sigma}\right) + \sigma\sqrt{t}\phi\left(\frac{\mu\sqrt{t}}{\sigma}\right)
\end{aligned}$$

Therefore, an approximation of the CVA is:

$$\mathrm{CVA}\left(t\right) \approx \boldsymbol{s}\times N\times f\left(0,T\right)\times\int_t^T g\left(u\right)\,\mathrm{d}u$$

where:

$$g(u) = e^{-(r+\lambda)(u-t)}\varphi(u,T)\left(\mu\Phi\left(\frac{\mu\sqrt{u}}{\sigma}\right)u + \sigma\phi\left(\frac{\mu\sqrt{u}}{\sigma}\right)\sqrt{u}\right)$$

To calculate this approximation, we use a numerical integration method. Syrkin and Shirazi (2015) provide a second approximation that does not require any integration, but it seems to be less accurate.

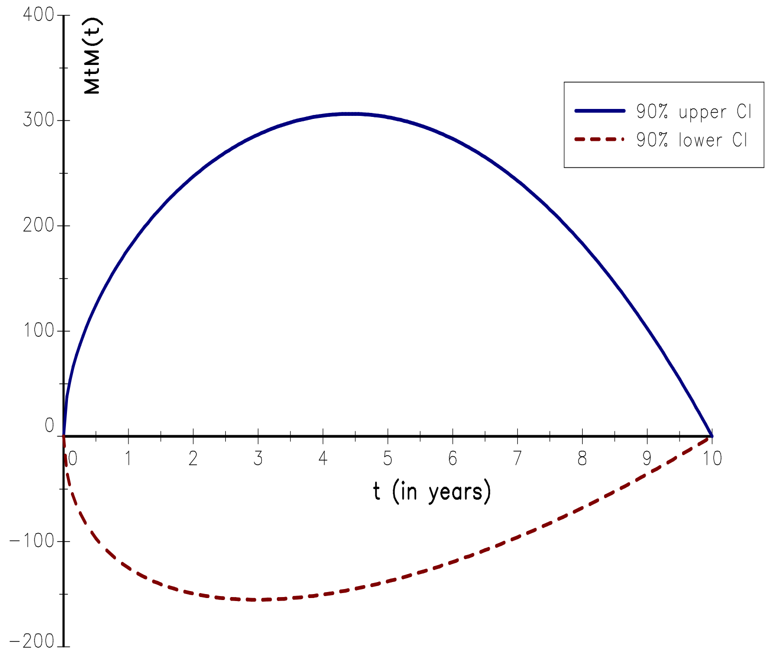

FIGURE 4.5: Confidence interval of the mark-to-market

6. All the computations are done using a Gauss-Legendre quadrature of order 128.

   (a) We have reported the 90% confidence interval of $\text{MtM}(t)$ in Figure 4.5.

   (b) The time profile of $\text{EpE}(t)$ and $\mathbb{E}[\text{MtM}(t)]$ is shown in Figure 4.6. We verify that $\text{EpE}(t) > \mathbb{E}[\text{MtM}(t)]$ and we retrieve the bell-shaped curve of IRS counterparty exposure.

   (c) In Figure 4.7, we observe that the approximation of the CVA gives good results.

   (d) When we calculate the CVA, we consider a risk-neutral probability distribution $\mathbb{Q}$. This implies that $\mu = 0\%$ is a more realistic value than $\mu = 2\%$.

4.4.6 Risk contribution of CVA with collateral

1. Since we have $\text{MtM}_i(t) = \mu_i(t) + \sigma_i(t) X_i$, we deduce that:

$$\begin{aligned}\text{MtM}(t) &= \sum_{i=1}^{n} w_i(\mu_i(t) + \sigma_i(t) X_i) \\ &= \sum_{i=1}^{n} w_i\mu_i(t) + \sum_{i=1}^{n} w_i\sigma_i(t) X_i\end{aligned}$$

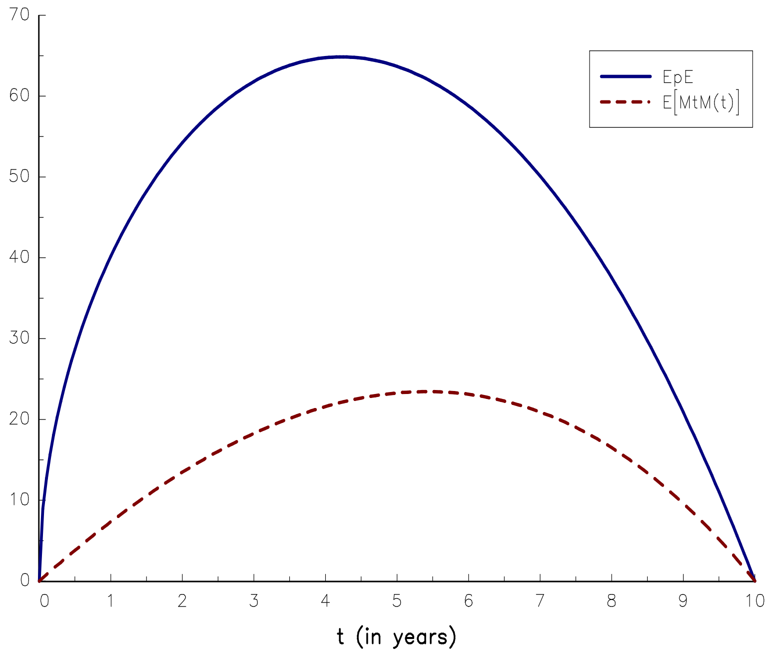

FIGURE 4.6: Comparison of $\mathrm{EpE}(t)$ and $\mathbb{E}[\mathrm{MtM}(t)]$

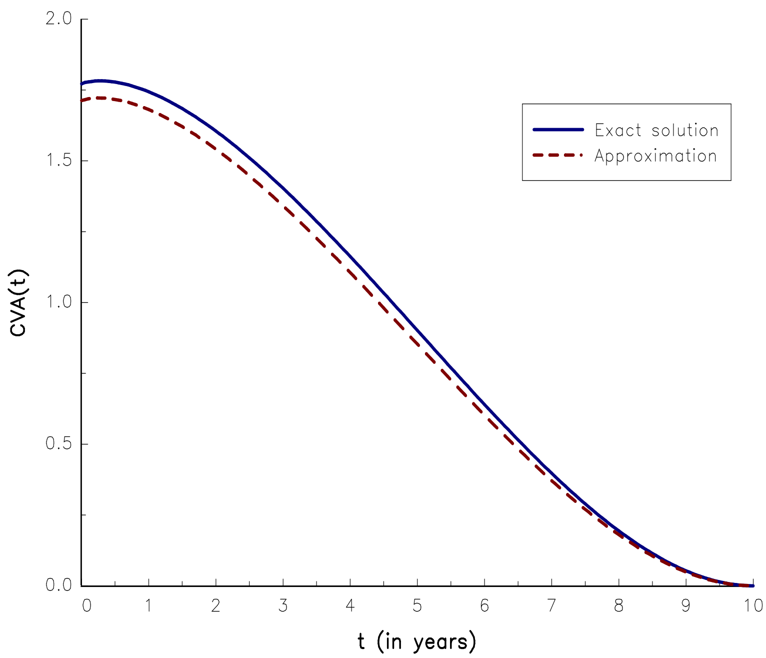

FIGURE 4.7: Approximation of the CVA

Let $\mu(t) = (\mu_1(t), \ldots, \mu_n(t))$ be the mean vector of $(\mathrm{MtM}_1(t), \ldots, \mathrm{MtM}_n(t))$. It follows that the expected value $\mu_w(t)$ of the portfolio mark-to-market is equal to:

$$\begin{aligned}\mu_w(t) &= \mathbb{E}[\mathrm{MtM}(t)] \\ &= \sum_{i=1}^{n} w_i \mu_i(t) \\ &= w^\top \mu(t)\end{aligned}$$

We define the volatility $\sigma_w(t)$ of the portfolio mark-to-market:

$$\begin{aligned}\sigma_w^2(t) &= \mathrm{var}(\mathrm{MtM}(t)) \\ &= \mathrm{var}\left(\sum_{i=1}^{n} w_i \sigma_i(t) X_i\right) \\ &= \sum_{i=1}^{n} w_i^2 \sigma_i^2(t) \mathbb{E}\left[X_i^2\right] + \sum_{j>i} w_i w_j \sigma_i(t) \sigma_j(t) \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^{n} w_i^2 \sigma_i^2(t) + \sum_{j>i} w_i w_j \sigma_i(t) \sigma_j(t) \rho_{i,j} \\ &= w^\top \Sigma(t) w\end{aligned}$$

where $\Sigma(t)$ is the covariance matrix of $(\mathrm{MtM}_1(t), \ldots, \mathrm{MtM}_n(t))$ such that:

$$\Sigma_{i,j}(t) = \rho_{i,j} \sigma_i(t) \sigma_j(t)$$

It follows that:

$$\begin{aligned}\mathrm{MtM}(t) &= \sum_{i=1}^{n} w_i \mu_i(t) + \sum_{i=1}^{n} w_i \sigma_i(t) X_i \\ &= \mu_w(t) + \sigma_w(t) X\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$. We deduce that the portfolio mark-to-market is a Gaussian random variable:

$$\mathrm{MtM}(t) \sim \mathcal{N}\left(\mu_w^2(t), \sigma_w^2(t)\right)$$

2. We have:

$$\begin{aligned}\gamma_i(t) &= \frac{\mathrm{cov}(\mathrm{MtM}_i(t), \mathrm{MtM}(t))}{\sqrt{\mathrm{var}(\mathrm{MtM}_i(t))\,\mathrm{var}(\mathrm{MtM}(t))}} \\ &= \frac{\mathbb{E}\left[\sigma_i(t) X_i \sum_{j=1}^{n} w_j \sigma_j(t) X_j\right]}{\sigma_i(t) \sigma(t)} \\ &= \frac{\mathbb{E}\left[\sum_{j=1}^{n} w_j \sigma_j(t) X_i X_j\right]}{\sigma(t)} \\ &= \sum_{j=1}^{n} \frac{w_j \sigma_j(t)}{\sigma(t)} \rho_{i,j}\end{aligned}$$

It follows that:

$$X_i = \gamma_i(t) X + \sqrt{1 - \gamma_i^2(t)}\, \varepsilon_i \tag{4.2}$$

where the idiosyncratic risks $\varepsilon_i \sim \mathcal{N}(0,1)$ are independent and satisfy $\varepsilon_i \perp X$. We verify that $\mathbb{E}[X_i] = 0$ and:

$$\begin{aligned}\sigma^2(X_i) &= \gamma_i^2(t)\,\sigma^2(X) + \left(1-\gamma_i^2(t)\right)\sigma^2(\varepsilon_i) \\ &= \gamma_i^2(t) + 1 - \gamma_i^2(t) \\ &= 1\end{aligned}$$

3. If we note $e^+(t) = \max(\mathrm{MtM}(t) - C(t), 0)$ and $C(t) = \max(\mathrm{MtM}(t) - H, 0)$, the expression of the counterparty exposure is equal to:

$$\begin{aligned}e^+(t) &= \max(\mathrm{MtM}(t) - \max(\mathrm{MtM}(t) - H, 0), 0) \\ &= \mathrm{MtM}(t)\cdot \mathbb{1}\{0 \le \mathrm{MtM}(t) < H\} + H \cdot \mathbb{1}\{\mathrm{MtM}(t) \ge H\}\end{aligned}$$

We have:

$$\begin{aligned}\mathrm{MtM}(t) \ge H &\Leftrightarrow \mu_w(t) + \sigma_w(t)\,x \ge H \\ &\Leftrightarrow x \ge x^\star(H) = \frac{H - \mu_w(t)}{\sigma_w(t)}\end{aligned}$$

and:

$$\mathrm{MtM}(t) \ge 0 \Leftrightarrow x \ge x^\star(0) = -\frac{\mu_w(t)}{\sigma_w(t)}$$

We deduce that:

$$\begin{aligned}\mathrm{EpE}(t;w) &= \int_{x^\star(0)}^{x^\star(H)} (\mu_w(t) + \sigma_w(t)\,x)\,\phi(x)\,\mathrm{d}x + \\ &\quad H\int_{x^\star(H)}^{\infty} \phi(x)\,\mathrm{d}x\end{aligned}$$

We have:

$$\begin{aligned}(*) &= \int_{x^\star(0)}^{x^\star(H)} (\mu_w(t) + \sigma_w(t)\,x)\,\phi(x)\,\mathrm{d}x \\ &= \mu_w(t)\int_{x^\star(0)}^{x^\star(H)} \phi(x)\,\mathrm{d}x + \sigma_w(t)\int_{x^\star(0)}^{x^\star(H)} x\phi(x)\,\mathrm{d}x \\ &= \mu_w(t)\left(\Phi(x^\star(H)) - \Phi(x^\star(0))\right) + \sigma_w(t)\left[-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right]_{x^\star(0)}^{x^\star(H)} \\ &= \mu_w(t)\left(\Phi(x^\star(H)) - \Phi(x^\star(0))\right) + \sigma_w(t)\left(\phi(x^\star(0)) - \phi(x^\star(H))\right)\end{aligned}$$

and:

$$\int_{x^\star(H)}^{\infty} \phi(x)\,\mathrm{d}x = 1 - \Phi(x^\star(H))$$

Using the fact that $\Phi(x) + \Phi(-x) = 1$, we finally obtain the following expression:

$$\begin{aligned}\mathrm{EpE}(t;w) &= \mu_w(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\ &\quad \sigma_w(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\ &\quad H\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right) \qquad (4.3)\end{aligned}$$

4. $C(t) = 0$ is equivalent to impose $H = +\infty$. Indeed, we verify that:

$$\begin{aligned} C(t) &= \max(\mathrm{MtM}(t) - H, 0) \\ &= \max(\mathrm{MtM}(t) - \infty, 0) \\ &= 0 \end{aligned}$$

It follows that:

$$\mathrm{EpE}(t; w) = \mu_w(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \sigma_w(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) \tag{4.4}$$

We have:

$$\frac{\partial\,\mu_w(t)}{\partial\,w_i} = \mu_i(t)$$

and:

$$\begin{aligned} \frac{\partial\,\sigma_w(t)}{\partial\,w_i} &= \frac{(\Sigma(t)\,w)_i}{\sigma(t)} \\ &= \gamma_i(t)\,\sigma_i(t) \end{aligned}$$

because we have the following relationship between $(\Sigma(t)\,w)_i$ and $\gamma_i(t)$:

$$\begin{aligned} (\Sigma(t)\,w)_i &= \sum_{j=1}^{n} \rho_{i,j}\sigma_i(t)\,\sigma_j(t)\,w_j \\ &= \sigma_i(t)\,\sigma(t) \sum_{j=1}^{n} w_j \frac{\sigma_j(t)}{\sigma(t)}\rho_{i,j} \\ &= \gamma_i(t)\,\sigma_i(t)\,\sigma(t) \end{aligned}$$

It follows that:

$$\begin{aligned} \frac{\partial}{\partial\,w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) &= \frac{\mu_i(t)}{\sigma_w(t)} - \frac{\mu_w(t)}{\sigma_w^3(t)}(\Sigma(t)\,w)_i \\ &= \frac{\mu_i(t)}{\sigma_w(t)} - \gamma_i(t)\,\sigma_i(t)\,\frac{\mu_w(t)}{\sigma_w^2(t)} \end{aligned}$$

We deduce that:

$$\frac{\partial}{\partial\,w_i}\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) = \phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\frac{\partial}{\partial\,w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)$$

and:

$$\frac{\partial}{\partial\,w_i}\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) = -\frac{\mu_w(t)}{\sigma_w(t)}\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\frac{\partial}{\partial\,w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)$$

because $\phi(x)' = -x\phi(x)$. Therefore, the expression of the marginal risk is equal to:

$$\begin{aligned} \frac{\partial\,\mathrm{EpE}(t; w)}{\partial\,w_i} = {} & \mu_i(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \\ & \mu_w(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\frac{\partial}{\partial\,w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \\ & \gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \\ & \sigma_w(t)\,\frac{\mu_w(t)}{\sigma_w(t)}\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\frac{\partial}{\partial\,w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) \end{aligned}$$

Finally, the expression of the risk contribution is given by:

$$\begin{aligned}\mathcal{RC}_i &= w_i \cdot \frac{\partial \operatorname{EpE}(t;w)}{\partial w_i} \\ &= w_i \left(\mu_i(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\right) \qquad (4.5)\end{aligned}$$

We have:

$$\begin{aligned}\sum_{i=1}^n \mathcal{RC}_i &= \sum_{i=1}^n w_i\mu_i(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \sum_{i=1}^n w_i\gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) \\ &= \Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\sum_{i=1}^n w_i\mu_i(t) + \phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\sum_{i=1}^n w_i\gamma_i(t)\,\sigma_i(t) \\ &= \Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\mu_w(t) + \phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)\sigma_w(t) \\ &= \operatorname{EpE}(t;w)\end{aligned}$$

because:

$$\begin{aligned}\sum_{i=1}^n w_i\gamma_i(t)\,\sigma_i(t) &= \sum_{i=1}^n w_i\left(\sum_{j=1}^n \frac{w_j\sigma_j(t)}{\sigma(t)}\rho_{i,j}\right)\sigma_i(t) \\ &= \frac{\sum_{i=1}^n\sum_{j=1}^n w_iw_j\rho_{i,j}\sigma_j(t)\,\sigma_i(t)}{\sigma(t)} \\ &= \frac{\sigma^2(t)}{\sigma(t)} \\ &= \sigma(t)\end{aligned}$$

We conclude that the risk measure $\operatorname{EpE}(t;w)$ satisfies the Euler allocation principle.

5. We can write:

$$\operatorname{EpE}(t;w) = E_1(t;w) - E_2(t;w) + E_3(t;w)$$

where:

$$\begin{aligned}E_1(t;w) &= \mu_w(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + \sigma_w(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) \\ E_2(t;w) &= \mu_w(t)\,\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) + \sigma_w(t)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) \\ E_3(t;w) &= H\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

We have:

$$\begin{aligned}\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) &= \frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - H\frac{\partial}{\partial w_i}\left(\frac{1}{\sigma_w(t)}\right) \\ &= \frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) + H\left(\frac{\gamma_i(t)\,\sigma_i(t)}{\sigma_w^2(t)}\right)\end{aligned}$$

We deduce that:

$$\begin{aligned}\frac{\partial}{\partial w_i}\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) &= \phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) \\ &= \phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)+ \\ & \quad H\left(\frac{\gamma_i(t)\,\sigma_i(t)}{\sigma_w^2(t)}\right)\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

and:

$$\begin{aligned}\frac{\partial}{\partial w_i}\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) &= -\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\cdot \\ & \quad \frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

We have:

$$\begin{aligned}\frac{\partial\, E_2(t;w)}{\partial w_i} &= \mu_i(t)\,\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)+ \\ & \quad \mu_w(t)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)+ \\ & \quad \gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)- \\ & \quad (\mu_w(t)-H)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

or:

$$\begin{aligned}\frac{\partial\, E_2(t;w)}{\partial w_i} &= \mu_i(t)\,\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)+\gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)+ \\ & \quad H\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

We have:

$$\frac{\partial\, E_3(t;w)}{\partial w_i}=H\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\frac{\partial}{\partial w_i}\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)$$

It follows that the marginal is equal to:

$$\begin{aligned}\frac{\partial\,\mathrm{EpE}(t;w)}{\partial w_i} &= \frac{\partial\, E_1(t;w)}{\partial w_i}-\frac{\partial\, E_2(t;w)}{\partial w_i}+\frac{\partial\, E_3(t;w)}{\partial w_i} \\ &= \mu_i(t)\,\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)+\gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)- \\ & \quad \mu_i(t)\,\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)-\gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) \\ &= \mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+ \\ & \quad \gamma_i(t)\,\sigma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)\end{aligned}$$

The expression of the risk contribution is given by:

$$\begin{aligned}\mathcal{RC}_i &= w_i\mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+\\ & w_i\gamma_i(t)\sigma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)\end{aligned} \tag{4.6}$$

We have:

$$\begin{aligned}\sum_{i=1}^n \mathcal{RC}_i &= \sum_{i=1}^n \mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+\\ & \sum_{i=1}^n w_i\gamma_i(t)\sigma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+\\ &= \mu_w(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+\\ & \sigma_w(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)\\ &= \mathrm{EpE}(t;w)-H\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\end{aligned}$$

These risk contributions do not satisfy the Euler allocation principle, meaning that it is not possible to allocate the CVA capital according to Equation (4.6).

6. Type A Euler allocation is given by:

$$\begin{aligned}\mathcal{RC}_i &= \mathbb{E}\left[w_i\,\mathrm{MtM}_i(t)\cdot\mathbb{1}\{0\leq\mathrm{MtM}(t)<H\}\right]+\\ & H\cdot\frac{\mathbb{E}\left[\mathbb{1}\{\mathrm{MtM}(t)\geq H\}\right]\cdot\mathbb{E}\left[w_i\,\mathrm{MtM}_i(t)\cdot\mathbb{1}\{\mathrm{MtM}(t)\geq H\}\right]}{\mathbb{E}\left[\mathrm{MtM}(t)\cdot\mathbb{1}\{\mathrm{MtM}(t)\geq H\}\right]}\end{aligned}$$

Using Equation (4.2), we have:

$$\begin{aligned}w_i\,\mathrm{MtM}_i(t) &= w_i\mu_i(t)+w_i\sigma_i(t)X_i\\ &= w_i\mu_i(t)+w_i\sigma_i(t)\gamma_i(t)X+w_i\sigma_i(t)\sqrt{1-\gamma_i^2(t)}\,\varepsilon_i\end{aligned}$$

Since $\varepsilon_i\perp X$, it follows that:

$$\begin{aligned}(*) &= \mathbb{E}\left[w_i\,\mathrm{MtM}_i(t)\cdot\mathbb{1}\{0\leq\mathrm{MtM}(t)<H\}\right]\\ &= \mathbb{E}\left[w_i(\mu_i(t)+\sigma_i(t)X_i)\cdot\mathbb{1}\{0\leq\mu_w(t)+\sigma_w(t)X<H\}\right]\\ &= \int_{x^\star(0)}^{x^\star(H)} w_i\mu_i(t)\phi(x)\,\mathrm{d}x+\int_{x^\star(0)}^{x^\star(H)} w_i\sigma_i(t)\gamma_i(t)x\phi(x)\,\mathrm{d}x+\\ & \int_{x^\star(0)}^{x^\star(H)}\underbrace{\mathbb{E}\left[w_i\sigma_i(t)\sqrt{1-\gamma_i^2(t)}\,\varepsilon_i\right]}_{=0}\phi(x)\,\mathrm{d}x\\ &= w_i\mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)+\\ & w_i\sigma_i(t)\gamma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right)-\phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right)\end{aligned}$$

and:

$$\begin{aligned}
(*) &= \mathbb{E}\left[\mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right] \\
&= \mathbb{E}\left[\mathbb{1}\left\{\mu_w(t) + \sigma_w(t) X \geq H\right\}\right] \\
&= \int_{x^\star(H)}^{\infty} \phi(x)\,\mathrm{d}x \\
&= \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)
\end{aligned}$$

We also have:

$$\begin{aligned}
(*) &= \mathbb{E}\left[\mathrm{MtM}(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right] \\
&= \mathbb{E}\left[\left(\mu_w(t) + \sigma_w(t) X\right) \cdot \mathbb{1}\left\{\mu_w(t) + \sigma_w(t) X \geq H\right\}\right] \\
&= \int_{x^\star(H)}^{\infty} \left(\mu_w(t) + \sigma_w(t) x\right) \phi(x)\,\mathrm{d}x \\
&= \mu_w(t)\,\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right) + \sigma_w(t)\,\phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)
\end{aligned}$$

and:

$$\begin{aligned}
(*) &= \mathbb{E}\left[w_i\,\mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right] \\
&= \mathbb{E}\left[w_i\left(\mu_i(t) + \sigma_i(t) X_i\right) \cdot \mathbb{1}\left\{\mu_w(t) + \sigma_w(t) X \geq H\right\}\right] \\
&= \int_{x^\star(H)}^{\infty} w_i \mu_i(t)\,\phi(x)\,\mathrm{d}x + \int_{x^\star(H)}^{\infty} w_i \sigma_i(t)\,\gamma_i(t)\,x\phi(x)\,\mathrm{d}x \\
&= w_i \mu_i(t)\,\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right) + w_i \sigma_i(t)\,\gamma_i(t)\,\phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)
\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
\mathcal{RC}_i &= w_i \mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\
&\quad w_i \gamma_i(t)\,\sigma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\
&\quad H\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\frac{\psi_i}{\psi_w}
\end{aligned} \tag{4.7}$$

where:

$$\psi_i = w_i \mu_i(t)\,\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right) + w_i \gamma_i(t)\,\sigma_i(t)\,\phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)$$

and[4]:

$$\psi_w = \mu_w(t)\,\Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right) + \sigma_w(t)\,\phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)$$

7. The type B Euler allocation is given by:

$$\begin{aligned}
\mathcal{RC}_i &= \mathbb{E}\left[w_i\,\mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{0 \leq \mathrm{MtM}(t) < H\right\}\right] + \\
&\quad H \cdot \mathbb{E}\left[\frac{w_i\,\mathrm{MtM}_i(t)}{\mathrm{MtM}(t)} \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]
\end{aligned}$$

[4] We notice that $\psi_w = \sum_{i=1}^{n} \psi_i$.

We have:

$$\frac{w_i \operatorname{MtM}_i(t)}{\operatorname{MtM}(t)} = \frac{w_i\mu_i(t) + w_i\sigma_i(t)\gamma_i(t)X}{\mu_w(t) + \sigma_w(t)X} + \frac{w_i\sigma_i(t)\sqrt{1-\gamma_i^2(t)}}{\mu_w(t) + \sigma_w(t)X}\varepsilon_i$$

Since $\varepsilon_i \perp X$, it follows that:

$$\begin{aligned}
(*) &= \mathbb{E}\left[\frac{w_i \operatorname{MtM}_i(t)}{\operatorname{MtM}(t)} \cdot \mathbb{1}\left\{\operatorname{MtM}(t) \geq H\right\}\right] \\
&= \mathbb{E}\left[\frac{w_i(\mu_i(t) + \gamma_i(t)\sigma_i(t)X)}{\mu_w(t) + \sigma_w(t)X} \cdot \mathbb{1}\left\{\mu_w(t) + \sigma_w(t)X \geq H\right\}\right] + \\
&\quad \mathbb{E}\left[\frac{w_i\sigma_i(t)\sqrt{1-\gamma_i^2(t)}}{\mu_w(t) + \sigma_w(t)X} \cdot \mathbb{1}\left\{\mu_w(t) + \sigma_w(t)X \geq H\right\}\right]\mathbb{E}\left[\varepsilon_i\right] \\
&= \int_{x^\star(H)}^{\infty} w_i\left(\frac{\mu_i(t) + \gamma_i(t)\sigma_i(t)x}{\mu_w(t) + \sigma_w(t)x}\right)\phi(x)\,\mathrm{d}x
\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
\mathcal{RC}_i &= w_i\mu_i(t)\left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\
&\quad w_i\gamma_i(t)\sigma_i(t)\left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right) + \\
&\quad H\int_{x^\star(H)}^{\infty} w_i\left(\frac{\mu_i(t) + \gamma_i(t)\sigma_i(t)x}{\mu_w(t) + \sigma_w(t)x}\right)\phi(x)\,\mathrm{d}x
\end{aligned} \tag{4.8}$$

8. It follows that:

$$\begin{aligned}
\operatorname{MtM}(t) &= \sum_{i=1}^{n} w_i \operatorname{MtM}_i(t) \\
&= \sum_{i=1}^{n} w_i\mu_i(t) + \sum_{i=1}^{n} w_i\sigma_i(t)X_i \\
&= \mu_w(t) + \sigma_w(t)X
\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$. The correlation between X and X_B is given by:

$$\begin{aligned}
\varrho_w(t) &= \frac{\operatorname{cov}(\operatorname{MtM}(t), X_B)}{\sqrt{\operatorname{var}(\operatorname{MtM}(t))\operatorname{var}(X_B)}} \\
&= \frac{\mathbb{E}\left[\sum_{i=1}^{n} w_i\sigma_i(t)X_iX_B\right]}{\sigma(t)} \\
&= \frac{\mathbb{E}\left[\sum_{i=1}^{n} w_i\sigma_i(t)\varrho_iX_BX_B\right]}{\sigma(t)} + \frac{\mathbb{E}\left[\sum_{i=1}^{n} w_i\sigma_i(t)\sqrt{1-\varrho_i^2}\,\eta_iX_B\right]}{\sigma(t)} \\
&= \sum_{i=1}^{n} \frac{w_i\sigma_i(t)}{\sigma(t)}\varrho_i
\end{aligned}$$

We deduce that:

$$X = \varrho_w(t)X_B + \sqrt{1-\varrho_w^2(t)}\,\eta \tag{4.9}$$

where the idiosyncratic risk $\eta \sim \mathcal{N}(0,1)$ is independent from X_B.

9. Pykhtin and Rosen (2010) notice that all previous computations involve unconditional expectations, implying that we can derive easily the expected counterparty exposure $\mathbb{E}[e(t)]$ and the corresponding risk contributions $\mathcal{RC}_i$ by replacing all unconditional expectations $\mathbb{E}[Y]$ where Y is a random variable ($\text{MtM}_i(t)$, $\text{MtM}(t)$ and $e(t)$) by conditional expectations $\mathbb{E}[Y \mid \boldsymbol{\tau} = t]$ where $\boldsymbol{\tau}$ is the default time of the counterparty. Following Redon (2006), this is equivalent to calculate the conditional expectation with respect to the random variable X_B:

$$\mathbb{E}[Y \mid \boldsymbol{\tau} = t] = \mathbb{E}[Y \mid X_B = B(t)]$$

where $B(t) = \Phi^{-1}(1 - \mathbf{S}(t))$ is the default barrier and $\mathbf{S}(t)$ is the survival function of the counterparty. For conditional means, we have:

$$\mu_i(t \mid \boldsymbol{\tau} = t) = \mu_i(t) + \varrho_i \sigma_w(t) B(t)$$

and:

$$\mu_w(t \mid \boldsymbol{\tau} = t) = \mu_w(t) + \varrho_w(t) \sigma_w(t) B(t)$$

For conditional volatilities, it follows that:

$$\sigma_i(t \mid \boldsymbol{\tau} = t) = \sqrt{1 - \varrho_i^2}\sigma_i(t)$$

and:

$$\sigma_w(t \mid \boldsymbol{\tau} = t) = \sqrt{1 - \varrho_w^2(t)}\sigma_w(t)$$

Since the unconditional correlation $\gamma_i(t)$ is equal to $\operatorname{cov}(X_i, X)$, we have:

$$\begin{aligned}
\gamma_i(t) &= \mathbb{E}[X_i, X] \\
&= \mathbb{E}\left[\left(\varrho_i X_B + \sqrt{1 - \varrho_i^2}\,\eta_i\right)\left(\varrho_w(t) X_B + \sqrt{1 - \varrho_w^2(t)}\,\eta\right)\right] \\
&= \varrho_i \varrho_w(t) + \sqrt{1 - \varrho_i^2}\sqrt{1 - \varrho_w^2(t)}\gamma_i(t \mid \boldsymbol{\tau} = t)
\end{aligned}$$

where $\gamma_i(t \mid \boldsymbol{\tau} = t)$ is the correlation between η_i and η or the conditional correlation:

$$\gamma_i(t \mid \boldsymbol{\tau} = t) = \frac{\gamma_i(t) - \varrho_i \varrho_w(t)}{\sqrt{1 - \varrho_i^2}\sqrt{1 - \varrho_w^2(t)}}$$

To compute $\text{EpE}(t; w) = \mathbb{E}[e^+(t) \mid \boldsymbol{\tau} = t]$ and $\mathcal{RC}_i$, we replace $\mu_i(t)$, $\mu_w(t)$, $\sigma_i(t)$, $\sigma_w(t)$ and $\gamma_i(t)$ by $\mu_i(t \mid \boldsymbol{\tau} = t)$, $\mu_w(t \mid \boldsymbol{\tau} = t)$, $\sigma_i(t \mid \boldsymbol{\tau} = t)$, $\sigma_w(t \mid \boldsymbol{\tau} = t)$ and $\gamma_i(t \mid \boldsymbol{\tau} = t)$ in Equations (4.3), (4.4) (4.5), (4.7) and (4.8).

Chapter 5

Operational Risk

5.4.1 Estimation of the severity distribution

1. (a) The density of the Gaussian distribution $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$ is:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

Let $X \sim \mathcal{LN}\left(\mu, \sigma^2\right)$. We have $X = \exp Y$. It follows that:

$$f(x) = g(y)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|$$

with $y = \ln x$. We deduce that:

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \times \frac{1}{x} \\ &= \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right) \end{aligned}$$

(b) For $m \geq 1$, the non-centered moment is equal to:

$$\mathbb{E}\left[X^m\right] = \int_0^\infty x^m \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$$

By considering the change of variables $y = \sigma^{-1}(\ln x - \mu)$ and $z = y - m\sigma$, we obtain:

$$\begin{aligned} \mathbb{E}\left[X^m\right] &= \int_{-\infty}^\infty e^{m\mu+m\sigma y} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \,\mathrm{d}y \\ &= e^{m\mu} \times \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2+m\sigma y} \,\mathrm{d}y \\ &= e^{m\mu} \times e^{\frac{1}{2}m^2\sigma^2} \times \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-m\sigma)^2} \,\mathrm{d}y \\ &= e^{m\mu+\frac{1}{2}m^2\sigma^2} \times \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \mathrm{d}z \\ &= e^{m\mu+\frac{1}{2}m^2\sigma^2} \end{aligned}$$

We deduce that:

$$\mathbb{E}[X] = e^{\mu+\frac{1}{2}\sigma^2}$$

and:

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}^2\left[X\right] \\ &= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)\end{aligned}$$

We can estimate the parameters μ and σ with the generalized method of moments by using the following empirical moments:

$$\begin{cases} h_{i,1}(\mu,\sigma) = x_i - e^{\mu+\frac{1}{2}\sigma^2} \\ h_{i,2}(\mu,\sigma) = \left(x_i - e^{\mu+\frac{1}{2}\sigma^2}\right)^2 - e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \end{cases}$$

(c) The log-likelihood function of the sample $\{x_1, \ldots, x_n\}$ is:

$$\begin{aligned}\boldsymbol{\ell}(\mu,\sigma) &= \sum_{i=1}^{n} \ln f(x_i) \\ &= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \sum_{i=1}^{n}\ln x_i - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln x_i - \mu}{\sigma}\right)^2\end{aligned}$$

To find the ML estimators $\hat{\mu}$ and $\hat{\sigma}$, we can proceed in two different ways:

#1 $X \sim \mathcal{LN}\left(\mu,\sigma^2\right)$ implies that $Y = \ln X \sim \mathcal{N}\left(\mu,\sigma^2\right)$. We know that the ML estimators $\hat{\mu}$ and $\hat{\sigma}$ associated to Y are:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\ \hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2}\end{aligned}$$

We deduce that the ML estimators $\hat{\mu}$ and $\hat{\sigma}$ associated to the sample $\{x_1, \ldots, x_n\}$ are:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} \ln x_i \\ \hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \hat{\mu})^2}\end{aligned}$$

#2 We maximize the log-likelihood function:

$$\{\hat{\mu},\hat{\sigma}\} = \arg\max \boldsymbol{\ell}(\mu,\sigma)$$

The first-order conditions are $\partial_\mu \boldsymbol{\ell}(\mu,\sigma) = 0$ and $\partial_\sigma \boldsymbol{\ell}(\mu,\sigma) = 0$. We deduce that:

$$\partial_\mu \boldsymbol{\ell}(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu) = 0$$

and:

$$\partial_\sigma \boldsymbol{\ell}(\mu,\sigma) = -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(\ln x_i - \mu)^2}{\sigma^3} = 0$$

We finally obtain:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i$$

and:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\ln x_i - \hat{\mu}\right)^2}$$

2. (a) The probability density function is:

$$\begin{aligned} f(x) &= \frac{\partial \Pr\{X \leq x\}}{\partial x} \\ &= \alpha \frac{x^{-(\alpha+1)}}{x_{-}^{-\alpha}} \end{aligned}$$

For $m \geq 1$, we have:

$$\begin{aligned} \mathbb{E}[X^m] &= \int_{x_{-}}^{\infty} x^m \alpha \frac{x^{-(\alpha+1)}}{x_{-}^{-\alpha}} \,\mathrm{d}x \\ &= \frac{\alpha}{x_{-}^{-\alpha}} \int_{x_{-}}^{\infty} x^{m-\alpha-1} \,\mathrm{d}x \\ &= \frac{\alpha}{x_{-}^{-\alpha}} \left[\frac{x^{m-\alpha}}{m-\alpha}\right]_{x_{-}}^{\infty} \\ &= \frac{\alpha}{\alpha - m} x_{-}^{m} \end{aligned}$$

We deduce that:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha - 1} x_{-}$$

and:

$$\begin{aligned} \operatorname{var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] \\ &= \frac{\alpha}{\alpha-2} x_{-}^2 - \left(\frac{\alpha}{\alpha-1} x_{-}\right)^2 \\ &= \frac{\alpha}{(\alpha-1)^2 (\alpha-2)} x_{-}^2 \end{aligned}$$

We can then estimate the parameter α by considering the following empirical moments:

$$\begin{aligned} h_{i,1}(\alpha) &= x_i - \frac{\alpha}{\alpha-1} x_{-} \\ h_{i,2}(\alpha) &= \left(x_i - \frac{\alpha}{\alpha-1} x_{-}\right)^2 - \frac{\alpha}{(\alpha-1)^2(\alpha-2)} x_{-}^2 \end{aligned}$$

The generalized method of moments can consider either the first moment $h_{i,1}(\alpha)$, the second moment $h_{i,2}(\alpha)$ or the joint moments $(h_{i,1}(\alpha), h_{i,2}(\alpha))$. In the first case, the estimator is:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i - n x_{-}}$$

(b) The log-likelihood function is:

$$\begin{aligned}\boldsymbol{\ell}(\alpha) &= \sum_{i=1}^{n} \ln f(x_i) \\ &= n \ln \alpha - (\alpha+1) \sum_{i=1}^{n} \ln x_i + n\alpha \ln x_-\end{aligned}$$

The first-order condition is:

$$\partial_\alpha \boldsymbol{\ell}(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} \ln x_i + \sum_{i=1}^{n} \ln x_- = 0$$

We deduce that:

$$n = \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_-}$$

The ML estimator is then:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} (\ln x_i - \ln x_-)}$$

3. The probability density function of *(iii)* is:

$$\begin{aligned}f(x) &= \frac{\partial \Pr\{X \leq x\}}{\partial x} \\ &= \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}\end{aligned}$$

It follows that the log-likelihood function is:

$$\begin{aligned}\boldsymbol{\ell}(\alpha,\beta) &= \sum_{i=1}^{n} \ln f(x_i) \\ &= -n \ln \Gamma(\alpha) + n\alpha \ln \beta + (\alpha-1) \sum_{i=1}^{n} \ln x_i - \beta \sum_{i=1}^{n} x_i\end{aligned}$$

The first-order conditions $\partial_\alpha \boldsymbol{\ell}(\alpha,\beta) = 0$ and $\partial_\beta \boldsymbol{\ell}(\alpha,\beta) = 0$ imply that:

$$n\left(\ln \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right) + \sum_{i=1}^{n} \ln x_i = 0$$

and:

$$n\frac{\alpha}{\beta} - \sum_{i=1}^{n} x_i = 0$$

4. Let $Y \sim \Gamma(\alpha,\beta)$ and $X = \exp Y$. We have:

$$f_X(x)\,|\mathrm{d}x| = f_Y(y)\,|\mathrm{d}y|$$

where f_X and f_Y are the probability density functions of X and Y. We deduce that:

$$\begin{aligned}f_X(x) &= \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} \times \frac{1}{e^y} \\ &= \frac{\beta^\alpha (\ln x)^{\alpha-1} e^{-\beta \ln x}}{x\Gamma(\alpha)} \\ &= \frac{\beta^\alpha (\ln x)^{\alpha-1}}{\Gamma(\alpha)\, x^{\beta+1}}\end{aligned}$$

The support of this probability density function is $[0, +\infty)$. The log-likelihood function associated to the sample of individual losses $\{x_1, \ldots, x_n\}$ is:

$$\begin{aligned}\boldsymbol{\ell}(\alpha, \beta) &= \sum_{i=1}^{n} \ln f(x_i) \\ &= -n \ln \Gamma(\alpha) + n\alpha \ln \beta + (\alpha - 1) \sum_{i=1}^{n} \ln(\ln x_i) - (\beta + 1) \sum_{i=1}^{n} \ln x_i\end{aligned}$$

5. (a) Using Bayes' formula, we have:

$$\begin{aligned}\Pr\{X \le x \mid X \ge H\} &= \frac{\Pr\{H \le X \le x\}}{\Pr\{X \ge H\}} \\ &= \frac{\mathbf{F}(x) - \mathbf{F}(H)}{1 - \mathbf{F}(H)}\end{aligned}$$

where $\mathbf{F}$ is the cdf of X. We deduce that the conditional probability density function is:

$$\begin{aligned}f(x \mid X \ge H) &= \partial_x \Pr\{X \le x \mid X \ge H\} \\ &= \frac{f(x)}{1 - \mathbf{F}(H)} \times \mathbb{1}\{x \ge H\}\end{aligned}$$

For the log-normal probability distribution, we obtain:

$$\begin{aligned}f(x \mid X \ge H) &= \frac{1}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \,\mathrm{d}x \\ &= \varphi \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \,\mathrm{d}x\end{aligned}$$

We note $\mathcal{M}_m(\mu, \sigma)$ the conditional moment $\mathbb{E}[X^m \mid X \ge H]$. We have:

$$\begin{aligned}\mathcal{M}_m(\mu, \sigma) &= \varphi \times \int_H^\infty \frac{x^{m-1}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \,\mathrm{d}x \\ &= \varphi \times \int_{\ln H}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2 + mx} \,\mathrm{d}x \\ &= \varphi \times e^{m\mu + \frac{1}{2}m^2\sigma^2} \times \int_{\ln H}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\left(x - \left(\mu + m\sigma^2\right)\right)^2}{\sigma^2}} \,\mathrm{d}x \\ &= \frac{1 - \Phi\left(\frac{\ln H - \mu - m\sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{m\mu + \frac{1}{2}m^2\sigma^2}\end{aligned}$$

The two first moments of $X \mid X \ge H$ are then:

$$\mathcal{M}_1(\mu, \sigma) = \mathbb{E}[X \mid X \ge H] = \frac{1 - \Phi\left(\frac{\ln H - \mu - \sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{\mu + \frac{1}{2}\sigma^2}$$

and:

$$\mathcal{M}_2(\mu, \sigma) = \mathbb{E}\left[X^2 \mid X \ge H\right] = \frac{1 - \Phi\left(\frac{\ln H - \mu - 2\sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{2\mu + 2\sigma^2}$$

We can therefore estimate μ and σ by considering the following empirical moments:

$$\left\{ \begin{array}{l} h_{i,1}\left(\mu,\sigma\right) = x_i - \mathcal{M}_1\left(\mu,\sigma\right) \\ h_{i,2}\left(\mu,\sigma\right) = \left(x_i - \mathcal{M}_1\left(\mu,\sigma\right)\right)^2 - \left(\mathcal{M}_2\left(\mu,\sigma\right) - \mathcal{M}_1^2\left(\mu,\sigma\right)\right) \end{array} \right.$$

(b) We have:

$$\begin{aligned} f\left(x \mid X \geq H\right) &= \frac{f\left(x\right)}{1-\mathbf{F}\left(H\right)} \times \mathbb{1}\left\{x \geq H\right\} \\ &= \left(\alpha \frac{x^{-(\alpha+1)}}{x_-^{-\alpha}}\right) \Big/ \left(\frac{H^{-\alpha}}{x_-^{-\alpha}}\right) \\ &= \alpha \frac{x^{-(\alpha+1)}}{H^{-\alpha}} \end{aligned}$$

The conditional probability function is then a Pareto distribution with the same parameter α but with a new threshold $x_- = H$. We can then deduce that the ML estimator $\hat{\alpha}$ is:

$$\hat{\alpha} = \frac{n}{\left(\sum_{i=1}^n \ln x_i\right) - n \ln H}$$

(c) The conditional probability density function is:

$$\begin{aligned} f\left(x \mid X \geq H\right) &= \frac{f\left(x\right)}{1-\mathbf{F}\left(H\right)} \times \mathbb{1}\left\{x \geq H\right\} \\ &= \left(\frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma\left(\alpha\right)}\right) \Big/ \int_H^\infty \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma\left(\alpha\right)} \,\mathrm{d}t \\ &= \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\int_H^\infty \beta^\alpha t^{\alpha-1} e^{-\beta t} \,\mathrm{d}t} \end{aligned}$$

We deduce that the log-likelihood function is:

$$\begin{aligned} \boldsymbol{\ell}\left(\alpha,\beta\right) &= n\alpha \ln \beta - n \ln \left(\int_H^\infty \beta^\alpha t^{\alpha-1} e^{-\beta t} \,\mathrm{d}t\right) + \\ & \quad \left(\alpha - 1\right) \sum_{i=1}^n \ln x_i - \beta \sum_{i=1}^n x_i \end{aligned}$$

5.4.2 Estimation of the frequency distribution

1. We have:

$$\Pr\left\{N = n\right\} = e^{-\lambda_Y} \frac{\lambda_Y^n}{n!}$$

We deduce that the expression of the log-likelihood function is:

$$\begin{aligned} \boldsymbol{\ell}\left(\lambda_Y\right) &= \sum_{t=1}^T \ln \Pr\left\{N = N_{Y_t}\right\} \\ &= -\lambda_Y T + \left(\sum_{t=1}^T N_{Y_t}\right) \ln \lambda_Y - \sum_{t=1}^T \ln\left(N_{Y_t}!\right) \end{aligned}$$

The first-order condition is:

$$\frac{\partial \boldsymbol{\ell}(\lambda_Y)}{\partial \lambda_Y} = -T + \frac{1}{\lambda_Y}\left(\sum_{t=1}^{T} N_{Y_t}\right) = 0$$

We deduce that the ML estimator is:

$$\hat{\lambda}_Y = \frac{1}{T}\sum_{t=1}^{T} N_{Y_t} = \frac{n}{T}$$

2. Using the same arguments, we obtain:

$$\hat{\lambda}_Q = \frac{1}{4T}\sum_{t=1}^{4T} N_{Q_t} = \frac{n}{4T} = \frac{\hat{\lambda}_Y}{4}$$

3. Considering a quarterly or annual basis has no impact on the capital charge. Indeed, the capital charge is computed with a one-year time horizon. If we use a quarterly basis, we have to find the distribution of the annual loss number. In this case, the annual loss number is the sum of the four quarterly loss numbers:

$$N_Y = N_{Q_1} + N_{Q_2} + N_{Q_3} + N_{Q_4}$$

We know that each quarterly loss number follows a Poisson distribution $\mathcal{P}\left(\hat{\lambda}_Q\right)$ and that they are independent. Because the Poisson distribution is infinitely divisible, we obtain:

$$N_{Q_1} + N_{Q_2} + N_{Q_3} + N_{Q_4} \sim \mathcal{P}\left(4\hat{\lambda}_Q\right)$$

We deduce that the annual loss number follows a Poisson distribution $\mathcal{P}\left(\hat{\lambda}_Y\right)$ in both cases.

4. This result remains valid if we consider the first moment because the MM estimator is exactly the ML estimator.

5. Since we have $\operatorname{var}(\mathcal{P}(\lambda)) = \lambda$, the MM estimator in the case of annual loss numbers is:

$$\hat{\lambda}_Y = \frac{1}{T}\sum_{t=1}^{T} N_{Y_t}^2 - \frac{n^2}{T^2}$$

If we use a quarterly basis, we obtain:

$$\begin{aligned}
\hat{\lambda}_Q &= \frac{1}{4}\left(\frac{1}{T}\sum_{t=1}^{4T} N_{Q_t}^2 - \frac{n^2}{4T^2}\right) \\
&\neq \frac{\hat{\lambda}_Y}{4}
\end{aligned}$$

There is no reason that $\hat{\lambda}_Y = 4\hat{\lambda}_Q$ meaning that the capital charge will not be the same.

5.4.3 Using the method of moments in operational risk models

1. (a) By definition, we have $\Pr\{N(t)=n\} = e^{-\lambda}\lambda^n/n!$. We deduce that:

$$\begin{aligned}
\mathbb{E}[N(t)] &= \sum_{n=0}^{\infty} n \times \Pr\{N(t)=n\} \\
&= \sum_{n=0}^{\infty} n e^{-\lambda}\frac{\lambda^n}{n!} \\
&= \lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} \\
&= \lambda
\end{aligned}$$

(b) We have:

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=0}^{m}(N(t)-i)\right] &= \sum_{n=0}^{\infty}\prod_{i=0}^{m}(n-i)\,e^{-\lambda}\frac{\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty}(n(n-1)\cdots(n-m))\,e^{-\lambda}\frac{\lambda^n}{n!}
\end{aligned}$$

The term of the sum is equal to zero when $n = 0, 1, \ldots, m$. We obtain:

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=0}^{m}(N(t)-i)\right] &= \sum_{n=m+1}^{\infty}(n(n-1)\cdots(n-m))\,e^{-\lambda}\frac{\lambda^n}{n!} \\
&= e^{-\lambda}\sum_{n=m+1}^{\infty}\frac{\lambda^n}{(n-m-1)!} \\
&= \lambda^{m+1}e^{-\lambda}\sum_{n=m+1}^{\infty}\frac{\lambda^{n-m-1}}{(n-m-1)!} \\
&= \lambda^{m+1}e^{-\lambda}\sum_{n'=0}^{\infty}\frac{\lambda^{n'}}{n'!}
\end{aligned}$$

with $n' = n-(m+1)$. It follows that:

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=0}^{m}(N(t)-i)\right] &= \lambda^{m+1}e^{-\lambda}e^{\lambda} \\
&= \lambda^{m+1} \qquad (5.1)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\operatorname{var}(N(t)) &= \mathbb{E}\left[N(t)^2\right] - \mathbb{E}^2[N(t)] \\
&= \mathbb{E}\left[N(t)^2 - N(t)\right] + \mathbb{E}[N(t)] - \mathbb{E}^2[N(t)] \\
&= \mathbb{E}[N(t)(N(t)-1)] + \mathbb{E}[N(t)] - \mathbb{E}^2[N(t)]
\end{aligned}$$

Using the formula (5.1) with $m = 1$, we finally obtain:

$$\begin{aligned}
\operatorname{var}(N(t)) &= \lambda^{1+1} + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

(c) The estimator based on the first moment is:

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} N_t$$

whereas the estimator based on the second moment is:

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\left(N_t - \frac{1}{T}\sum_{t=1}^{T} N_t\right)^2$$

2. (a) We have:

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=0}^{N(t)} X_i\right] \\ &= \mathbb{E}[N(t)]\,\mathbb{E}[X_i] \\ &= \lambda \exp\left(\mu + \frac{1}{2}\sigma^2\right)\end{aligned}$$

(b) Because $\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i\neq j} x_i x_j$, it follows that:

$$\begin{aligned}\mathbb{E}\left[S^2\right] &= \mathbb{E}\left[\sum_{i=0}^{N(t)} X_i^2 + \sum_{i\neq j}^{N(t)}\sum^{N(t)} X_i X_j\right] \\ &= \mathbb{E}[N(t)]\,\mathbb{E}\left[X_i^2\right] + \mathbb{E}[N(t)(N(t)-1)]\,\mathbb{E}[X_i X_j] \\ &= \mathbb{E}[N(t)]\,\mathbb{E}\left[X_i^2\right] + \left(\mathbb{E}\left[N(t)^2\right] - \mathbb{E}[N(t)]\right)\mathbb{E}[X_i]\,\mathbb{E}[X_j]\end{aligned}$$

We have:

$$\begin{aligned}\mathbb{E}[N(t)] &= \lambda \\ \mathbb{E}\left[N(t)^2\right] &= \operatorname{var}(N(t)) + \mathbb{E}^2[N(t)] = \lambda + \lambda^2\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[X_i^2\right] &= \operatorname{var}(X_i) + \mathbb{E}^2[X_i] \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) + \left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2 \\ &= e^{2\mu+2\sigma^2} \\ \mathbb{E}[X_i]\,\mathbb{E}[X_j] &= e^{\mu+\frac{1}{2}\sigma^2} e^{\mu+\frac{1}{2}\sigma^2} \\ &= e^{2\mu+\sigma^2}\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[S^2\right] &= \lambda\mathbb{E}\left[X_i^2\right] + \left(\lambda + \lambda^2 - \lambda\right)\mathbb{E}[X_i]\,\mathbb{E}[X_j] \\ &= \lambda\mathbb{E}\left[X_i^2\right] + \lambda^2\mathbb{E}[X_i]\,\mathbb{E}[X_j] \\ &= \lambda e^{2\mu+2\sigma^2} + \lambda^2 e^{2\mu+\sigma^2}\end{aligned}$$

and:

$$\begin{aligned}\operatorname{var}(S) &= \mathbb{E}\left[S^2\right] - \mathbb{E}^2[S] \\ &= \lambda e^{2\mu+2\sigma^2} + \lambda^2 e^{2\mu+\sigma^2} - \lambda^2\left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2 \\ &= \lambda e^{2\mu+2\sigma^2}\end{aligned}$$

(c) We have:

$$\left\{\begin{array}{l}\mathbb{E}[S] = \lambda e^{\mu+\frac{1}{2}\sigma^2} \\ \operatorname{var}(S) = \lambda e^{2\mu+2\sigma^2}\end{array}\right.$$

We deduce that:

$$\frac{\operatorname{var}(S)}{\mathbb{E}^2[S]} = \frac{\lambda e^{2\mu+2\sigma^2}}{\lambda^2 e^{2\mu+\sigma^2}} = \frac{e^{\sigma^2}}{\lambda}$$

It follows that:

$$\sigma^2 = \ln\lambda + \ln(\operatorname{var}(S)) - \ln\left(\mathbb{E}^2[S]\right)$$

and:

$$\begin{aligned}\mu &= \ln\mathbb{E}[S] - \ln\lambda - \frac{1}{2}\sigma^2 \\ &= \ln\mathbb{E}[S] + \frac{1}{2}\ln\left(\mathbb{E}^2[S]\right) - \frac{3}{2}\ln\lambda - \frac{1}{2}\ln(\operatorname{var}(S))\end{aligned}$$

Let $\hat{\lambda}$ be an estimated value of λ. We finally obtain:

$$\hat{\mu} = \ln m_S + \frac{1}{2}\ln m_S^2 - \frac{3}{2}\ln\hat{\lambda} - \frac{1}{2}\ln v_S$$

and

$$\hat{\sigma} = \sqrt{\ln\hat{\lambda} + \ln v_S - \ln m_S^2}$$

where m_S and v_S are the empirical mean and variance of aggregated losses.

3. (a) We know that the duration d between two consecutive losses that are larger than ℓ is exponentially distributed with parameter $\lambda(1-\mathbf{F}(\ell))$. We deduce that:

$$d \sim \mathcal{E}\left(\lambda\left(\frac{\ell}{x\_}\right)^{-\alpha}\right)$$

(b) We can ask experts to estimate the return time d_j for several scenarios ℓ_j and then calibrate the parameters λ and α using the method of moments and the following moment conditions:

$$\mathbb{E}[d_j] - \frac{\ell_j^\alpha}{\lambda x\_^\alpha} = 0$$

5.4.4 Calculation of the Basel II required capital

1. In order to implement the historical value-at-risk, we first calculate the daily stock returns:

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

TABLE 5.1: Stock returns $R_{A,s}$ and $R_{B,s}$ (24 first historical scenarios)

			$R_{A,s}$				
−2.01	−0.01	−0.73	−0.71	1.79	2.27	−0.15	−0.55
−0.43	1.01	0.05	0.32	2.08	−2.37	−0.55	2.57
0.29	−2.54	−0.03	0.00	−0.90	−0.03	1.96	−0.35
			$R_{B,s}$				
0.35	−0.84	0.85	1.40	1.35	1.36	−1.45	−1.95
2.17	1.51	−0.69	1.87	−0.06	−1.61	−1.25	2.20
−1.07	−2.85	−0.99	−0.06	−2.34	−1.31	3.79	−1.46

where P_t is the stock price at time t. We report the return values for stocks A and B in Table 5.1. These data are used to simulate the future P&L defined as follows:

$$\begin{aligned} \Pi_s &= 10\,000 \times 105.5 \times R_{A,s} + \\ &\quad 25\,000 \times 353.0 \times R_{B,s} \end{aligned}$$

where $R_{A,s}$ and $R_{B,s}$ are the stock returns of A and B for the s^{th} historical scenario. Table 5.2 gives the values taken by Π_s. We then calculate the order statistics $\Pi_{s:250}$ and deduce that the value-at-risk is equal to:

$$\begin{aligned} \mathrm{VaR}_{99\%}(w) &= \frac{1}{2}(323\,072 + 314\,695) \\ &= \$318\,883 \end{aligned}$$

It follows that the required capital is equal to:

$$\begin{aligned} \mathcal{K}_{\mathrm{MR}} &= (3+\xi) \times \sqrt{10} \times \mathrm{VaR}_{99\%}(w) \\ &= \$3.53 \text{ mn} \end{aligned}$$

TABLE 5.2: Daily P&L (24 first historical scenarios)

		Daily P&L Π_s			
9 972	−74 339	67 520	115 824	137 790	144 032
−129 857	−178 339	186 837	143 722	−60 767	168 780
16 234	−166 679	−116 117	221 553	−91 336	−278 402
−87 357	−5 671	−215 517	−116 172	354 813	−132 741
		Order statistic $\Pi_{s:250}$			
−340 656	−323 072	−314 695	−278 402	−277 913	−275 118
−268 632	−259 781	−255 936	−252 509	−250 117	−249 523
−243 502	−218 295	−217 514	−217 327	−215 517	−211 382
−211 018	−208 061	−192 950	−192 603	−190 993	−189 410

2. We apply the IRB formulas with the right asset class exposure. For the bank meta-credit, we have:

$$\begin{aligned} \rho(\mathrm{PD}) &= 12\% \times \frac{1-e^{-50\times 1\%}}{1-e^{-50}} + 24\% \times \frac{1-\left(1-e^{-50\times 1\%}\right)}{1-e^{-50}} \\ &= 19.28\% \end{aligned}$$

Because the maturity of the meta-credit is one year, the maturity adjustment is equal to 1. We deduce that:

$$\begin{aligned}\mathcal{K}^{\star} &= 75\% \times \Phi\left(\frac{\Phi^{-1}(1\%) + \sqrt{0.1928}\Phi^{-1}(99.9\%)}{\sqrt{1-0.1928}}\right) - 75\% \times 1\% \\ &= 9.77\%\end{aligned}$$

It follows that:

$$\mathrm{RW} = 12.5 \times 9.77\% = 122.13\%$$

and:

$$\mathrm{RWA} = 80 \times 122.13\% = \$97.70 \text{ mn}$$

We finally obtain:

$$\mathcal{K} = 8\% \times 97.70 = \$7.82 \text{ mn}$$

For the corporate meta-credit, we proceed in the same way, except that we have to incorporate the maturity adjustment. We have:

$$b(\mathrm{PD}) = (0.11852 - 0.05478 \times \ln(5\%))^2 = 7.99\%$$

and:

$$\varphi(M) = \frac{1 + (2 - 2.5) \times 0.0799}{1 - 1.5 \times 0.0799} = 1.0908$$

Using the IRB formula, we obtain $\mathcal{K}^{\star} = 15.35\%$ and $\mathcal{K} = 30.69\%$. For the SME meta-credit, we have to be careful when we calculate the correlation. Indeed, we have[1]:

$$\begin{aligned}\rho^{\mathrm{SME}}(\mathrm{PD}) &= 12\% \times \frac{1 - e^{-50 \times 2\%}}{1 - e^{-50}} + 24\% \times \frac{1 - \left(1 - e^{-50 \times 2\%}\right)}{1 - e^{-50}} - \\ &\quad 4\% \times \left(1 - \frac{(\max(30, 5) - 5)}{45}\right)\end{aligned}$$

For mortgage and retail exposures, we use a one-year maturity. The default correlation is set equal to 15% for the mortgage meta-credit whereas we consider the following formula for the retail meta-credit:

$$\begin{aligned}\rho(\mathrm{PD}) &= 3\% \times \frac{1 - e^{-35 \times 4\%}}{1 - e^{-35}} + 16\% \times \frac{1 - \left(1 - e^{-35 \times 4\%}\right)}{1 - e^{-35}} \\ &= 6.21\%\end{aligned}$$

All the results are reported in Table 5.3. At the bank level, we then obtain:

$$\begin{aligned}\mathrm{RWA} &= 97.70 + 383.65 + 55.95 + 97.87 + 122.80 \\ &= \$757.98\end{aligned}$$

and:

$$\begin{aligned}\mathcal{K}_{\mathrm{CR}} &= 7.82 + 30.69 + 4.48 + 7.83 + 9.82 \\ &= \$60.64\end{aligned}$$

[1] In order to simplify the calculation, we assume that the USD/EUR exchange rate is equal to 1.

TABLE 5.3: Calculation of capital requirements for credit exposures

Exposure	$b\,(\text{PD})$	$\varphi\,(\text{M})$	$\rho\,(\text{PD})$	$\mathcal{K}^\star$	RW	RWA	$\mathcal{K}$
Bank	0.14	1.00	19.28%	9.77%	122.13%	97.70	7.82
Corporate	0.08	1.09	12.99%	15.35%	191.83%	383.65	30.69
SME	0.11	1.46	14.64%	8.95%	111.91%	55.95	4.48
Mortgage			15.00%	15.66%	195.74%	97.87	7.83
Retail			6.21%	9.82%	122.80%	122.80	9.82

3. We calculate the capital charge for operational risk by Monte Carlo methods. The loss is equal to:

$$L = \sum_{i=1}^{N} L_i$$

where $L_i \sim \mathcal{LN}(8, 4)$ and N can take two values ($N = 5$ or $N = 10$) with:

$$\begin{aligned} \Pr\{N = 5\} &= 60\% \\ \Pr\{N = 10\} &= 40\% \end{aligned}$$

We first simulate the yearly number of operational losses N by inverting the cumulative density function:

$$\begin{aligned} \Pr\{N \leq 5\} &= 60\% \\ \Pr\{N \leq 10\} &= 100\% \end{aligned}$$

Let u_s be a uniform random variate for the s^{th} simulation. The simulated variate N_s is defined as follows:

$$N_s = \begin{cases} 5 & \text{if } u_s \leq 0.6 \\ 10 & \text{otherwise} \end{cases}$$

Then, we have to simulate the operational losses $L_i^{(s)}$ using the probability integral transform:

$$\begin{aligned} U &= \mathbf{F}(L_i) \\ &= \Phi\left(\frac{\ln L_i - 8}{2}\right) \end{aligned}$$

It follows that:

$$L_i = \exp\left(8 + 2 \times \Phi^{-1}(U)\right)$$

Let $u_i^{(s)}$ be a uniform random variate. We have:

$$L_i^{(s)} = \exp\left(8 + 2 \times \Phi^{-1}\left(u_i^{(s)}\right)\right)$$

Another way to simulate $L_i^{(s)}$ is to notice that $\Phi^{-1}(U) \sim \mathcal{N}(0, 1)$, meaning that:

$$L_i^{(s)} = \exp\left(8 + 2 \times n_i^{(s)}\right)$$

where $n_i^{(s)}$ is a normal random variate $\mathcal{N}(0, 1)$. The simulated value of the aggregated loss is then:

$$L_s = \sum_{i=1}^{N_s} L_i^{(s)}$$

Let us consider an example. We assume that $u_s = 0.2837$. It follows that $N_s = 5$. This means that we have to simulate five operational losses in the year. We obtain the following figures:

i	1	2	3	4	5
$u_i^{(s)}$	0.4351	0.0387	0.2209	0.3594	0.5902
$\Phi^{-1}\left(u_i^{(s)}\right)$	−0.1633	−1.7666	−0.7692	−0.3600	0.2282
$L_i^{(s)}$	2,150.26	87.09	640.05	1451.02	4704.84

The first loss experienced by the bank is \$2 150.26, the second loss is equal to \$87.09, etc. We deduce that the yearly total loss is equal to \$9 033.25:

$$\begin{aligned} L_s &= 2\,150.26 + 87.09 + 640.05 + 1\,451.02 + 4\,704.84 \\ &= \$9\,033.25 \end{aligned}$$

By considering n_S simulated values of L_s, the capital charge for operational risk is given by the 99.9% quantile:

$$\text{VaR}_{99.9\%} = L_{0.999 n_S : n_S}$$

For instance, if we consider one-million simulation runs, the capital charge corresponds to the 999 000$^{\text{th}}$ order statistic. In our case, we estimate the capital charge with 250 millions of simulation runs and obtain:

$$\mathcal{K}_{\text{OR}} = \$4.39 \text{ mn}$$

Because the required capital is estimated using Monte Carlo methods, there is an uncertainty on this number. For instance, we have reported the histogram of the VaR estimator with one-million simulation runs in Figure 5.1. In this case, we obtain:

$$\Pr\{4.27 \leq \text{VaR}_{99.9\%} \leq 4.50\} = 90\%$$

4. We deduce that the capital ratio of the bank is equal to:

$$\begin{aligned} \text{Cooke ratio} &= \frac{C_{\text{Bank}}}{\text{RWA} + 12.5 \times \mathcal{K}_{\text{MR}} + 12.5 \times \mathcal{K}_{\text{OR}}} \\ &= \frac{70}{757.98 + 12.5 \times 3.53 + 12.5 \times 4.39} \\ &= 8.17\% \end{aligned}$$

5.4.5 Parametric estimation of the loss severity distribution

1. We consider that the losses follow a log-logistic distribution.

   (a) By definition, the probability density function is equal to:

$$f(x;\alpha,\beta) = \frac{\partial\, \mathbf{F}(x;\alpha,\beta)}{\partial\, x}$$

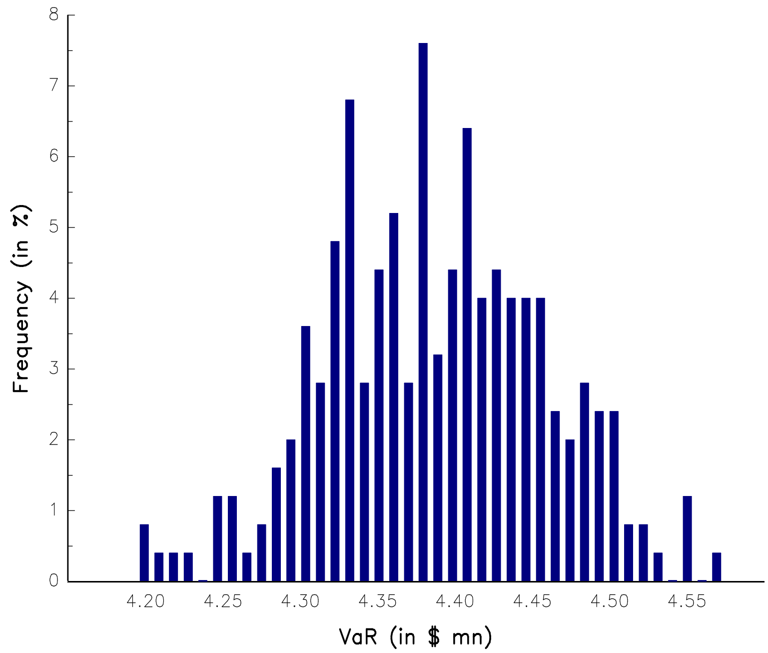

FIGURE 5.1: Histogram of the $\mathrm{VaR}_{99.9\%}$ estimator with $n_S = 10^6$

We deduce that:

$$\begin{aligned} f(x;\alpha,\beta) &= \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}\left(1+(x/\alpha)^{\beta}\right)}{\left(1+(x/\alpha)^{\beta}\right)^2} - \\ &\quad \frac{(x/\alpha)^{\beta}(\beta/\alpha)(x/\alpha)^{\beta-1}}{\left(1+(x/\alpha)^{\beta}\right)^2} \\ &= \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{\left(1+(x/\alpha)^{\beta}\right)^2} \end{aligned}$$

(b) The definition of the log-likelihood function is:

$$\boldsymbol{\ell}(\alpha,\beta) = \sum_{i=1}^{n} \ln f(x_i;\alpha,\beta)$$

We deduce that:

$$\begin{aligned} \boldsymbol{\ell}(\alpha,\beta) &= n\ln(\beta/\alpha) + (\beta-1)\sum_{i=1}^{n}\ln(x_i/\alpha) - 2\sum_{i=1}^{n}\ln\left(1+(x_i/\alpha)^{\beta}\right) \\ &= n\ln\beta - n\beta\ln\alpha + (\beta-1)\sum_{i=1}^{n}\ln x_i - \\ &\quad 2\sum_{i=1}^{n}\ln\left(1+(x_i/\alpha)^{\beta}\right) \end{aligned}$$

(c) Maximizing the log-likelihood function leads the first-order conditions:

$$\frac{\partial \boldsymbol{\ell}(\alpha,\beta)}{\partial \alpha} = -n\frac{\beta}{\alpha} + 2\frac{\beta}{\alpha}\sum_{i=1}^{n}\frac{(x_i/\alpha)^{\beta}}{1+(x_i/\alpha)^{\beta}} = 0$$

and:

$$\frac{\partial \boldsymbol{\ell}(\alpha,\beta)}{\partial \beta} = \frac{n}{\beta} - n\ln\alpha + \sum_{i=1}^{n}\ln x_i - 2\sum_{i=1}^{n}\frac{(x_i/\alpha)^{\beta}\ln(x_i/\alpha)}{1+(x_i/\alpha)^{\beta}} = 0$$

By assuming that $\beta \neq 0$, we deduce that:

$$\sum_{i=1}^{n}\mathbf{F}(x_i;\alpha,\beta) = \frac{n}{2}$$

and:

$$\frac{n}{\beta} - n\ln\alpha + \sum_{i=1}^{n}\ln x_i - 2\sum_{i=1}^{n}\frac{(x_i/\alpha)^{\beta}\ln x_i}{1+(x_i/\alpha)^{\beta}} + 2\ln\alpha\sum_{i=1}^{n}\frac{(x_i/\alpha)^{\beta}}{1+(x_i/\alpha)^{\beta}} = 0$$

We then obtain:

$$\frac{n}{\beta} + \sum_{i=1}^{n}\ln x_i - 2\sum_{i=1}^{n}\mathbf{F}(x_i;\alpha,\beta)\ln x_i = 0$$

or equivalently:

$$\sum_{i=1}^{n}\left(2\mathbf{F}(x_i;\alpha,\beta) - 1\right)\ln x_i = \frac{n}{\beta}$$

It follows that the ML estimators $\hat{\alpha}$ and $\hat{\beta}$ satisfy the following conditions:

$$\left\{\begin{array}{l} \sum_{i=1}^{n}\mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right) = n/2 \\ \sum_{i=1}^{n}\left(2\mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right) - 1\right)\ln x_i = n/\hat{\beta} \end{array}\right.$$

(d) Using the sample of loss data, we obtain:

$$2\sum_{i=1}^{10}\mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right) = 10.000$$

and:

$$\hat{\beta}\sum_{i=1}^{10}\left(2\mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right) - 1\right)\ln x_i = 9.999$$

Because the two mathematical terms are equal to $n = 10$, the first-order conditions of the ML optimization program are satisfied.

(e) We have:

$$\begin{aligned} \boldsymbol{\ell}(\alpha,\beta) &= \sum_{i=1}^{n}\ln f(x_i;\alpha,\beta) - \sum_{i=1}^{n}\ln\left(1-\mathbf{F}(H;\alpha,\beta)\right) \\ &= n\ln\beta - n\beta\ln\alpha + (\beta-1)\sum_{i=1}^{n}\ln x_i - \\ &\quad 2\sum_{i=1}^{n}\ln\left(1+(x_i/\alpha)^{\beta}\right) + n\ln\left(1+(H/\alpha)^{\beta}\right) \end{aligned}$$

5.4.6 Mixed Poisson processes

1. We recall that $\mathbb{E}[\mathcal{P}(\lambda)] = \mathrm{var}(\mathcal{P}(\lambda)) = \lambda$. We deduce that:

$$\begin{aligned} \mathbb{E}[N(t)] &= \mathbb{E}[\mathbb{E}[N(t) \mid \Lambda]] \\ &= \mathbb{E}[\Lambda] \end{aligned} \tag{5.2}$$

and:

$$\begin{aligned} \mathrm{var}(N(t)) &= \mathbb{E}\left[N(t)^2\right] - \mathbb{E}^2[N(t)] \\ &= \mathbb{E}\left[\mathbb{E}\left[N(t)^2 \mid \Lambda\right]\right] - \mathbb{E}^2[\mathbb{E}[N(t) \mid \Lambda]] \\ &= \mathbb{E}\left[\mathrm{var}(N(t) \mid \Lambda) + \mathbb{E}^2[N(t) \mid \Lambda]\right] - \mathbb{E}^2[\mathbb{E}[N(t) \mid \Lambda]] \\ &= \mathbb{E}[\mathrm{var}(N(t) \mid \Lambda)] + \mathbb{E}\left[\mathbb{E}^2[N(t) \mid \Lambda]\right] - \mathbb{E}^2[\mathbb{E}[N(t) \mid \Lambda]] \\ &= \mathbb{E}[\mathrm{var}(N(t) \mid \Lambda)] + \mathrm{var}(\mathbb{E}[N(t) \mid \Lambda]) \\ &= \mathbb{E}[\Lambda] + \mathrm{var}(\Lambda) \\ &= \mathbb{E}[N(t)] + \mathrm{var}(\Lambda) \end{aligned} \tag{5.3}$$

2. By definition, we have $\mathrm{var}(\Lambda) \geq 0$, which implies that:

$$\mathrm{var}(N(t)) \geq \mathbb{E}[N(t)]$$

The equality holds if and only if $\mathrm{var}(\Lambda) = 0$. We deduce that Λ must be constant and we obtain the Dirac distribution:

$$\Pr\{\Lambda = \lambda\} = 1$$

Since we have $N(t) \sim \mathcal{P}(\lambda)$, we deduce that:

$$\begin{aligned} p(n) &= \Pr\{N(t) = n\} \\ &= \frac{e^{-\lambda}\lambda^n}{n!} \end{aligned}$$

It follows that:

$$\begin{aligned} \varphi(n) &= \frac{(n+1) \cdot p(n+1)}{p(n)} \\ &= (n+1) \cdot \frac{e^{-\lambda}\lambda^{n+1}}{(n+1)!} \cdot \frac{n!}{e^{-\lambda}\lambda^n} \\ &= \lambda \end{aligned}$$

3. (a) We reiterate that:

$$\mathbb{E}[\mathcal{G}(\alpha, \beta)] = \frac{\alpha}{\beta}$$

and:

$$\mathrm{var}(\mathcal{G}(\alpha, \beta)) = \frac{\alpha}{\beta^2}$$

Using Equations (5.2) and (5.3), we obtain:

$$\begin{aligned} \mathbb{E}[N(t)] &= \mathbb{E}[\Lambda] \\ &= \mathbb{E}[\mathcal{G}(\alpha, \beta)] \\ &= \frac{\alpha}{\beta} \end{aligned} \tag{5.4}$$

and:

$$\begin{aligned}\operatorname{var}(N(t)) &= \mathbb{E}[\Lambda]+\operatorname{var}(\Lambda) \\ &= \mathbb{E}[\mathcal{G}(\alpha,\beta)]+\operatorname{var}(\mathcal{G}(\alpha,\beta)) \\ &= \frac{\alpha}{\beta}+\frac{\alpha}{\beta^2} \\ &= \frac{\alpha(\beta+1)}{\beta^2} \qquad (5.5)\end{aligned}$$

(b) By definition of the compound distribution, we have:

$$\begin{aligned}p(n) &= \int_0^\infty p(n \mid \mathcal{P}(\lambda)) f(\lambda \mid \mathcal{G}(\alpha,\beta))\,\mathrm{d}\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!}\frac{\beta^\alpha\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}\,\mathrm{d}\lambda \\ &= \frac{\beta^\alpha}{n!\Gamma(\alpha)}\int_0^\infty e^{-\lambda(\beta+1)}\lambda^{n+\alpha-1}\,\mathrm{d}\lambda \qquad (5.6)\end{aligned}$$

We know that:

$$\int_0^\infty t^{a-1}e^{-t}\,\mathrm{d}t = \Gamma(a)$$

We deduce that[2]:

$$\begin{aligned}\int_0^\infty t^{a-1}e^{-bt}\,\mathrm{d}t &= \int_0^\infty t^{a-1}e^{-bt}\,\mathrm{d}t \\ &= \int_0^\infty \left(\frac{x}{b}\right)^{a-1} e^{-x}\,\frac{\mathrm{d}x}{b} \\ &= \int_0^\infty x^{a-1}e^{-x}\,\mathrm{d}x \\ &= \frac{\Gamma(a)}{b^a}\end{aligned}$$

From Equation (5.6), we obtain:

$$\begin{aligned}p(n) &= \frac{\beta^\alpha}{n!\Gamma(\alpha)}\frac{\Gamma(n+\alpha)}{(\beta+1)^{n+\alpha}} \\ &= \frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)}\cdot\frac{\beta^\alpha}{(\beta+1)^{n+\alpha}} \qquad (5.7)\end{aligned}$$

We notice that:

$$\begin{aligned}\frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)} &= \frac{(n+\alpha-1)!}{n!(\alpha-1)!} \\ &= \binom{n+\alpha-1}{n}\end{aligned}$$

and:

$$\begin{aligned}\frac{\beta^\alpha}{(\beta+1)^{n+\alpha}} &= \left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^n \\ &= \left(1-\frac{1}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^n\end{aligned}$$

[2] We use the change of variable $x = bt$.

Therefore, Equation (5.7) becomes:

$$\begin{aligned} p(n) &= \binom{n+\alpha-1}{n}\left(1-\frac{1}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{n} \\ &= \binom{n+r-1}{n}(1-p)^{r}p^{n} \end{aligned}$$

This is the probability mass function of the negative binomial distribution $\mathcal{NB}(r,p)$ where[3] $r=\alpha$ and $p=1/(\beta+1)$.

(c) We have:

$$\begin{aligned} \varphi(n) &= \frac{(n+1)\cdot p(n+1)}{p(n)} \\ &= (n+1)\frac{\binom{n+r}{n+1}}{\binom{n+r-1}{n}}\frac{(1-p)^{r}p^{n+1}}{(1-p)^{r}p^{n}} \\ &= \frac{(n+r)!}{(n+r-1)!}p \\ &= pn+pr \end{aligned}$$

4. (a) Since we have $\mathbb{E}[\mathcal{E}(\lambda)]=\lambda^{-1}$ and $\operatorname{var}(\mathcal{E}(\lambda))=\lambda^{-2}$, we obtain:

$$\mathbb{E}[N(t)]=\frac{1}{\lambda}$$

and:

$$\begin{aligned} \operatorname{var}(N(t)) &= \frac{1}{\lambda}+\frac{1}{\lambda^{2}} \\ &= \frac{\lambda+1}{\lambda^{2}} \end{aligned}$$

[3] An alternative approach to find the values of r and p consists in match the first two moments. Indeed, we know that:

$$\mathbb{E}[\mathcal{NB}(r,p)]=\frac{pr}{1-p}$$

and:

$$\operatorname{var}(\mathcal{NB}(r,p))=\frac{pr}{(1-p)^{2}}$$

We deduce that:

$$\begin{aligned} \left\{\begin{array}{l} \dfrac{\alpha}{\beta}=\dfrac{pr}{1-p} \\ \dfrac{\alpha(\beta+1)}{\beta^{2}}=\dfrac{pr}{(1-p)^{2}} \end{array}\right. &\Leftrightarrow \left\{\begin{array}{l} \dfrac{\alpha}{\beta}=\dfrac{pr}{1-p} \\ \dfrac{\beta+1}{\beta}=\dfrac{1}{1-p} \end{array}\right. \\ &\Leftrightarrow \left\{\begin{array}{l} \alpha=\beta\dfrac{pr}{1-p} \\ \beta^{-1}=\dfrac{1}{1-p}-1 \end{array}\right. \\ &\Leftrightarrow \left\{\begin{array}{l} \alpha=r \\ \beta=\dfrac{1-p}{p} \end{array}\right. \\ &\Leftrightarrow \left\{\begin{array}{l} r=\alpha \\ p=\dfrac{1}{1+\beta} \end{array}\right. \end{aligned}$$

(b) We have $\mathcal{E}(\lambda) = \mathcal{G}(1, \lambda)$. We deduce that the compound Poisson distribution is the negative binomial distribution $\mathcal{NB}(r, p)$ where $r = 1$ and $p = 1/(\lambda + 1)$. In this case, the expression of the probability mass function becomes:

$$\begin{aligned} p(n) &= \binom{n+r-1}{n} (1-p)^r p^n \\ &= (1-p) p^n \end{aligned}$$

We conclude that $N(t)$ has a geometric distribution $\mathcal{G}(1/(\lambda + 1))$.

Chapter 6

Liquidity Risk

Chapter 7

Asset/Liability Management Risk

7.4.1 Constant amortization of a loan

1. We have:

$$\begin{aligned}
C_0 &= \sum_{t=1}^{n} \frac{A}{(1+i)^t} \\
&= \frac{A}{(1+i)} \sum_{t=0}^{n-1} \frac{1}{(1+i)^t} \\
&= \frac{A}{(1+i)} \times \frac{1 - \frac{1}{(1+i)^n}}{1 - \frac{1}{(1+i)}} \\
&= \left(1 - \frac{1}{(1+i)^n}\right) \frac{A}{i} \\
&= c_{(n)} A
\end{aligned}$$

where $c_{(n)}$ is the capitalization factor:

$$c_{(n)} = \frac{1 - (1+i)^{-n}}{i}$$

2. Since $C_0 = N_0$, we have:

$$\left(1 - \frac{1}{(1+i)^n}\right) \frac{A}{i} = N_0$$

We deduce that the value of the constant annuity is equal to:

$$\begin{aligned}
A &= \left(1 - \frac{1}{(1+i)^n}\right)^{-1} i N_0 \\
&= \frac{(1+i)^n}{(1+i)^n - 1} i N_0 \\
&= \frac{i}{1 - (1+i)^{-n}} N_0
\end{aligned}$$

It follows that the constant annuity rate $a_{(n)}$ is given by the following formula:

$$a_{(n)} = \frac{i}{1 - (1+i)^{-n}} = \frac{1}{c_{(n)}}$$

3. At time $t = 1$, we pay A. The interest payment is equal to $I(1) = iN_0$ while the principal payment is equal to the difference between the annuity and the interest

payment:

$$\begin{aligned}
P(1) &= A - I(1) \\
&= \left(a_{(n)} - i\right) N_0 \\
&= \left(\frac{i}{1 - (1+i)^{-n}} - i\right) N_0
\end{aligned}$$

We deduce that the amount outstanding (or remaining capital) is equal to:

$$\begin{aligned}
N(1) &= N_0 - P(1) \\
&= N_0 - \left(a_{(n)} - i\right) N_0 \\
&= \left(1 + i - \frac{i}{1 - (1+i)^{-n}}\right) N_0 \\
&= \left(\frac{1 - (1+i)^{-n+1}}{1 - (1+i)^{-n}}\right)\left(1 - \frac{1}{(1+i)^n}\right)\frac{A}{i} \\
&= \left(\frac{1 - (1+i)^{-n+1}}{1 - (1+i)^{-n}}\right)\left(\frac{(1+i)^n - 1}{(1+i)^n}\right)\frac{A}{i} \\
&= \left(\frac{(1+i)^n - 2 - i + (1+i)^{-n+1}}{(1+i)^n - 1}\right)\frac{A}{i} \\
&= \left(1 - \frac{(1+i) - (1+i)^{-n+1}}{(1+i)^n - 1}\right)\frac{A}{i}
\end{aligned}$$

We also have:

$$\begin{aligned}
C(1) &= \sum_{t=1}^{n-1} \frac{A}{(1+i)^t} \\
&= \left(1 - \frac{1}{(1+i)^{n-1}}\right)\frac{A}{i}
\end{aligned}$$

Since we have:

$$\begin{aligned}
\frac{(1+i) - (1+i)^{-n+1}}{(1+i)^n - 1} &= \frac{(1+i)^{n-1}}{(1+i)^{n-1}}\left(\frac{(1+i) - (1+i)^{-n+1}}{(1+i)^n - 1}\right) \\
&= \frac{1}{(1+i)^{n-1}}\left(\frac{(1+i)^n - (1+i)^0}{(1+i)^n - 1}\right) \\
&= \frac{1}{(1+i)^{n-1}}
\end{aligned}$$

we conclude that the amount outstanding $N(1)$ is equal to the present value of the annuity at time $t = 1$:

$$N(1) = C(1)$$

4. More generally, we have:

$$\begin{aligned}
N(t) &= C(t) \\
&= c_{(n-t)} A \\
&= \left(\frac{1 - (1+i)^{-(n-t)}}{i}\right) A
\end{aligned}$$

It follows that:

$$\begin{aligned} I(t) &= iN(t-1) \\ &= \left(1 - \frac{1}{(1+i)^{n-t+1}}\right) A \end{aligned}$$

and:

$$\begin{aligned} P(t) &= A - I(t) \\ &= \frac{1}{(1+i)^{n-t+1}} A \end{aligned}$$

7.4.2 Computation of the amortization functions $\mathbf{S}(t,u)$ and $\mathbf{S}^\star(t,u)$

1. By definition, we have:

$$\begin{aligned} \mathbf{S}(t,u) &= \mathbb{1}\{t \le u < t+m\} \\ &= \begin{cases} 1 & \text{if } u \in [t, t+m[\\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This means that the survival function is equal to one when u is between the current date t and the maturity date $T = t+m$. When u reaches T, the outstanding amount is repaid, implying that $\mathbf{S}(t,T)$ is equal to zero. It follows that:

$$\begin{aligned} \mathbf{S}^\star(t,u) &= \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,u)\,\mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s} \\ &= \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\cdot \mathbb{1}\{s \le u < s+m\}\,\mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\cdot \mathbb{1}\{s \le t < s+m\}\,\mathrm{d}s} \end{aligned}$$

For the numerator, we have:

$$\begin{aligned} \mathbb{1}\{s \le u < s+m\} = 1 &\Rightarrow u < s+m \\ &\Leftrightarrow s > u-m \end{aligned}$$

and:

$$\int_{-\infty}^{t} \mathrm{NP}(s)\cdot \mathbb{1}\{s \le u < s+m\}\,\mathrm{d}s = \int_{u-m}^{t} \mathrm{NP}(s)\,\mathrm{d}s$$

For the denominator, we have:

$$\begin{aligned} \mathbb{1}\{s \le t < s+m\} = 1 &\Rightarrow t < s+m \\ &\Leftrightarrow s > t-m \end{aligned}$$

and:

$$\int_{-\infty}^{t} \mathrm{NP}(s)\cdot \mathbb{1}\{s \le t < s+m\}\,\mathrm{d}s = \int_{t-m}^{t} \mathrm{NP}(s)\,\mathrm{d}s$$

We deduce that:

$$\mathbf{S}^\star(t,u) = \mathbb{1}\{t \le u < t+m\} \cdot \frac{\int_{u-m}^{t} \mathrm{NP}(s)\,\mathrm{d}s}{\int_{t-m}^{t} \mathrm{NP}(s)\,\mathrm{d}s}$$

In the case where the new production is a constant, we have $\text{NP}(s) = c$ and:

$$\begin{aligned}
\mathbf{S}^{\star}(t,u) &= \mathbb{1}\{t \leq u < t+m\} \cdot \frac{\int_{u-m}^{t} \mathrm{d}s}{\int_{t-m}^{t} \mathrm{d}s} \\
&= \mathbb{1}\{t \leq u < t+m\} \cdot \frac{[s]_{u-m}^{t}}{[s]_{t-m}^{t}} \\
&= \mathbb{1}\{t \leq u < t+m\} \cdot \left(\frac{t-u+m}{t-t+m}\right) \\
&= \mathbb{1}\{t \leq u < t+m\} \cdot \left(1 - \frac{u-t}{m}\right)
\end{aligned}$$

The survival function $\mathbf{S}^{\star}(t,u)$ corresponds to the case of a linear amortization.

2. If the amortization is linear, we have:

$$\mathbf{S}(t,u) = \mathbb{1}\{t \leq u < t+m\} \cdot \left(1 - \frac{u-t}{m}\right)$$

We deduce that:

$$\mathbf{S}^{\star}(t,u) = \mathbb{1}\{t \leq u < t+m\} \cdot \frac{\displaystyle\int_{u-m}^{t} \text{NP}(s)\left(1 - \frac{u-s}{m}\right) \mathrm{d}s}{\displaystyle\int_{t-m}^{t} \text{NP}(s)\left(1 - \frac{t-s}{m}\right) \mathrm{d}s}$$

In the case where the new production is a constant, we obtain:

$$\mathbf{S}^{\star}(t,u) = \mathbb{1}\{t \leq u < t+m\} \cdot \frac{\displaystyle\int_{u-m}^{t} \left(1 - \frac{u-s}{m}\right) \mathrm{d}s}{\displaystyle\int_{t-m}^{t} \left(1 - \frac{t-s}{m}\right) \mathrm{d}s}$$

For the numerator, we have:

$$\begin{aligned}
\int_{u-m}^{t} \left(1 - \frac{u-s}{m}\right) \mathrm{d}s &= \left[s - \frac{su}{m} + \frac{s^2}{2m}\right]_{u-m}^{t} \\
&= \left(t - \frac{tu}{m} + \frac{t^2}{2m}\right) - \\
&\quad \left(u - m - \frac{u^2 - mu}{m} + \frac{(u-m)^2}{2m}\right) \\
&= \left(t - \frac{tu}{m} + \frac{t^2}{2m}\right) - \left(u - \frac{m}{2} - \frac{u^2}{2m}\right) \\
&= \frac{m^2 + u^2 + t^2 + 2mt - 2mu - 2tu}{2m} \\
&= \frac{(m-u+t)^2}{2m}
\end{aligned}$$

For the denominator, we use the previous result and we set $u = t$:

$$\begin{aligned}
\int_{t-m}^{t} \left(1 - \frac{t-s}{m}\right) \mathrm{d}s &= \frac{(m-t+t)^2}{2m} \\
&= \frac{m}{2}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\mathbf{S}^{\star}(t,u) &= \mathbb{1}\{t \leq u < t+m\} \cdot \frac{\dfrac{(m-u+t)^2}{2m}}{\dfrac{m}{2}} \\
&= \mathbb{1}\{t \leq u < t+m\} \cdot \frac{(m-u+t)^2}{m^2} \\
&= \mathbb{1}\{t \leq u < t+m\} \cdot \left(1 - \frac{u-t}{m}\right)^2
\end{aligned}$$

The survival function $\mathbf{S}^{\star}(t,u)$ corresponds to the case of a parabolic amortization.

3. If the amortization is exponential, we have:

$$\mathbf{S}(t,u) = e^{-\int_t^u \lambda \,\mathrm{d}s} = e^{-\lambda(u-t)}$$

It follows that:

$$\mathbf{S}^{\star}(t,u) = \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\, e^{-\lambda(u-s)}\,\mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\, e^{-\lambda(t-s)}\,\mathrm{d}s}$$

In the case where the new production is a constant, we obtain:

$$\begin{aligned}
\mathbf{S}^{\star}(t,u) &= \frac{\int_{-\infty}^{t} e^{-\lambda(u-s)}\,\mathrm{d}s}{\int_{-\infty}^{t} e^{-\lambda(t-s)}\,\mathrm{d}s} \\
&= \frac{\left[\lambda^{-1} e^{-\lambda(u-s)}\right]_{-\infty}^{t}}{\left[\lambda^{-1} e^{-\lambda(t-s)}\right]_{-\infty}^{t}} \\
&= e^{-\lambda(u-t)} \\
&= \mathbf{S}(t,u)
\end{aligned}$$

The stock amortization function is equal to the flow amortization function.

4. We recall that the liquidity duration is equal to:

$$\mathcal{D}(t) = \int_t^{\infty} (u-t)\, f(t,u)\,\mathrm{d}u$$

where $f(t,u)$ is the density function associated to the survival function $\mathbf{S}(t,u)$. For the stock, we have:

$$\mathcal{D}^{\star}(t) = \int_t^{\infty} (u-t)\, f^{\star}(t,u)\,\mathrm{d}u$$

where $f^{\star}(t,u)$ is the density function associated to the survival function $\mathbf{S}^{\star}(t,u)$:

$$f^{\star}(t,u) = \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\, f(s,u)\,\mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s}$$

In the case where the new production is constant, we obtain:

$$\mathcal{D}^{\star}(t) = \frac{\int_t^{\infty} (u-t) \int_{-\infty}^{t} f(s,u)\,\mathrm{d}s\,\mathrm{d}u}{\int_{-\infty}^{t} \mathbf{S}(s,t)\,\mathrm{d}s}$$

Since we have $\int_{-\infty}^{t} f(s,u)\ \mathrm{d}s = \mathbf{S}(t,u)$, we deduce that:

$$\mathcal{D}^{\star}(t) = \frac{\int_t^\infty (u-t)\,\mathbf{S}(t,u)\ \mathrm{d}u}{\int_{-\infty}^{t} \mathbf{S}(s,t)\ \mathrm{d}s}$$

5. (a) In the case of the bullet repayment debt, we have:

$$\mathcal{D}(t) = m$$

and:

$$\begin{aligned}
\mathcal{D}^{\star}(t) &= \frac{\int_t^{t+m} (u-t)\ \mathrm{d}u}{\int_{t-m}^{t} \mathrm{d}s} \\
&= \frac{\left[\frac{1}{2}(u-t)^2\right]_t^{t+m}}{\left[s\right]_{t-m}^{t}} \\
&= \frac{m}{2}
\end{aligned}$$

(b) In the case of the linear amortization, we have:

$$f(t,u) = \mathbb{1}\{t \le u < t+m\} \cdot \frac{1}{m}$$

and:

$$\begin{aligned}
\mathcal{D}(t) &= \int_t^{t+m} \frac{(u-t)}{m}\ \mathrm{d}u \\
&= \frac{1}{m}\left[\frac{1}{2}(u-t)^2\right]_t^{t+m} \\
&= \frac{m}{2}
\end{aligned}$$

For the stock duration, we deduce that

$$\begin{aligned}
\mathcal{D}^{\star}(t) &= \frac{\displaystyle\int_t^{t+m} (u-t)\left(1-\frac{u-t}{m}\right)\ \mathrm{d}u}{\displaystyle\int_{t-m}^{t}\left(1-\frac{t-s}{m}\right)\ \mathrm{d}s} \\
&= \frac{\displaystyle\int_t^{t+m}\left(u-t-\frac{u^2}{m}+2\frac{tu}{m}-\frac{t^2}{m}\right)\ \mathrm{d}u}{\displaystyle\int_{t-m}^{t}\left(1-\frac{t}{m}+\frac{s}{m}\right)\ \mathrm{d}s} \\
&= \frac{\left[\frac{u^2}{2}-tu-\frac{u^3}{3m}+\frac{tu^2}{m}-\frac{t^2u}{m}\right]_t^{t+m}}{\left[s-\frac{st}{m}+\frac{s^2}{2m}\right]_{t-m}^{t}}
\end{aligned}$$

The numerator and denominator are equal to:

$$\begin{aligned}(*) &= \left[\frac{u^2}{2} - tu - \frac{u^3}{3m} + \frac{tu^2}{m} - \frac{t^2u}{m}\right]_t^{t+m} \\ &= \frac{1}{6m}\left[3mu^2 - 6mtu - 2u^3 + 6tu^2 - 6t^2u\right]_t^{t+m} \\ &= \frac{1}{6m}\left(m^3 - 3mt^2 - 2t^3\right) + \frac{1}{6m}\left(3mt^2 + 2t^3\right) \\ &= \frac{m^2}{6}\end{aligned}$$

and:

$$\begin{aligned}(*) &= \left[s - \frac{st}{m} + \frac{s^2}{2m}\right]_{t-m}^{t} \\ &= \frac{1}{2m}\left[s^2 - 2s(t-m)\right]_{t-m}^{t} \\ &= \frac{1}{2m}\left(t^2 - 2t(t-m) - (t-m)^2 + 2(t-m)^2\right) \\ &= \frac{1}{2m}\left(t^2 - 2t^2 + 2mt + t^2 - 2mt + m^2\right) \\ &= \frac{m}{2}\end{aligned}$$

We deduce that:

$$\mathcal{D}^\star(t) = \frac{m}{3}$$

(c) For the exponential amortization, we have:

$$f(t,u) = \lambda e^{-\lambda(u-t)}$$

and[1]:

$$\begin{aligned}\mathcal{D}(t) &= \int_t^\infty (u-t)\,\lambda e^{-\lambda(u-t)}\,\mathrm{d}u \\ &= \int_0^\infty v\lambda e^{-\lambda v}\,\mathrm{d}v \\ &= \frac{1}{\lambda}\end{aligned}$$

For the stock duration, we deduce that:

$$\begin{aligned}\mathcal{D}^\star(t) &= \frac{\int_t^\infty (u-t)\,e^{-\lambda(u-t)}\,\mathrm{d}u}{\int_{-\infty}^t e^{-\lambda(t-s)}\,\mathrm{d}s} \\ &= \frac{\int_0^\infty ve^{-\lambda v}\,\mathrm{d}v}{\int_0^\infty e^{-\lambda v}\,\mathrm{d}v} \\ &= \frac{1}{\lambda}\end{aligned}$$

We verify that $\mathcal{D}(t) = \mathcal{D}^\star(t)$ since we have demonstrated that $\mathbf{S}^\star(t,u) = \mathbf{S}(t,u)$.

[1] We use the change of variable $v = u - t$.

6. (a) By definition, we have:

$$\mathrm{d}N(t) = (\mathrm{NP}(t) - \mathrm{NP}(t-m))\,\mathrm{d}t$$

(b) We have:

$$f(s,t) = \frac{\mathbf{1}\{s \leq t < s+m\}}{m}$$

It follows that:

$$\begin{aligned}\int_{-\infty}^{t} \mathrm{NP}(s)\, f(s,t)\; \mathrm{d}s &= \frac{1}{m}\int_{-\infty}^{t} \mathbf{1}\{s \leq t < s+m\} \cdot \mathrm{NP}(s)\; \mathrm{d}s \\ &= \frac{1}{m}\int_{t-m}^{t} \mathrm{NP}(s)\; \mathrm{d}s\end{aligned}$$

We deduce that:

$$\mathrm{d}N(t) = \left(\mathrm{NP}(t) - \frac{1}{m}\int_{t-m}^{t} \mathrm{NP}(s)\; \mathrm{d}s\right)\; \mathrm{d}t$$

(c) We have:

$$f(s,t) = \lambda e^{-\lambda(t-s)}$$

and:

$$\begin{aligned}\int_{-\infty}^{t} \mathrm{NP}(s)\, f(s,t)\; \mathrm{d}s &= \int_{-\infty}^{t} \mathrm{NP}(s)\, \lambda e^{-\lambda(t-s)}\, \mathrm{d}s \\ &= \lambda \int_{-\infty}^{t} \mathrm{NP}(s)\, e^{-\lambda(t-s)}\, \mathrm{d}s \\ &= \lambda N(t)\end{aligned}$$

We deduce that:

$$\mathrm{d}N(t) = (\mathrm{NP}(t) - \lambda N(t))\; \mathrm{d}t$$

7.4.3 Continuous-time analysis of the constant amortization mortgage (CAM)

1. We have $\mathrm{d}N(t) = -P(t)\; \mathrm{d}t$ where $A(t) = I(t) + P(t)$ and $I(t) = iN(t)$. We deduce that:

$$\mathrm{d}N(t) = (iN(t) - A)\; \mathrm{d}t$$

We know that the solution has the following form[2]:

$$N(t) = Ce^{it} + \frac{A}{i}$$

where C is a constant. Since $N(0) = N_0$, we have:

$$C = N_0 - \frac{A}{i}$$

[2]The solution of $y'(t) = ay(t) + b$ is equal to:

$$y(t) = Ce^{at} - \frac{b}{a}$$

and:

$$N(t) = \left(N_0 - \frac{A}{i}\right) e^{it} + \frac{A}{i}$$

At the maturity m, we have $N(m) = 0$, implying that:

$$\begin{aligned}\left(N_0 - \frac{A}{i}\right) e^{im} + \frac{A}{i} = 0 \quad &\Leftrightarrow \quad \frac{A}{i} = \frac{N_0 e^{im}}{e^{im} - 1} \\ &\Leftrightarrow \quad A = \frac{iN_0}{1 - e^{-im}}\end{aligned}$$

We deduce that:

$$\begin{aligned}N(t) &= \mathbf{1}\{t < m\} \cdot \left(\left(N_0 - \frac{N_0}{1 - e^{-im}}\right) e^{it} + \frac{N_0}{1 - e^{-im}}\right) \\ &= \mathbf{1}\{t < m\} \cdot N_0 \frac{1 - e^{-i(m-t)}}{1 - e^{-im}}\end{aligned}$$

because $N(t) = 0$ when $t \geq m$.

2. More generally, we have:

$$N(t,u) = \mathbf{1}\{t \leq u < t + m\} \cdot N(t) \frac{1 - e^{-i(t+m-u)}}{1 - e^{-im}}$$

This implies that:

$$\mathbf{S}(t,u) = \mathbf{1}\{t \leq u < t + m\} \cdot \frac{1 - e^{-i(t+m-u)}}{1 - e^{-im}}$$

and:

$$\mathbf{S}^\star(t,u) = \frac{\int_{u-m}^{t} \mathrm{NP}(s)\left(1 - e^{-i(s+m-u)}\right) \mathrm{d}s}{\int_{t-m}^{t} \mathrm{NP}(s)\left(1 - e^{-i(s+m-t)}\right) \mathrm{d}s}$$

If we assume that $\mathrm{NP}(s)$ is constant, we have:

$$\begin{aligned}\mathbf{S}^\star(t,u) &= \frac{\left[s + \frac{1}{i} e^{-i(s+m-u)}\right]_{u-m}^{t}}{\left[s + \frac{1}{i} e^{-i(s+m-t)}\right]_{t-m}^{t}} \\ &= \frac{t + m - u + \dfrac{e^{-i(t+m-u)} - 1}{i}}{m + \dfrac{e^{-im} - 1}{i}} \\ &= \frac{i(t + m - u) + e^{-i(t+m-u)} - 1}{im + e^{-im} - 1}\end{aligned}$$

3. We have:

$$f(t,u) = \mathbf{1}\{t \leq u < t + m\} \cdot \frac{ie^{-i(t+m-u)}}{1 - e^{-im}}$$

It follows that:

$$\begin{aligned}
\mathcal{D}(t) &= \frac{ie^{-im}}{1-e^{-im}} \int_t^{t+m} (u-t)\, e^{i(u-t)} \,\mathrm{d}u \\
&= \frac{ie^{-im}}{1-e^{-im}} \int_0^{m} v e^{iv} \,\mathrm{d}v \\
&= \frac{ie^{-im}}{1-e^{-im}} \left(\frac{me^{im}}{i} - \frac{e^{im}-1}{i^2} \right) \\
&= \frac{1}{1-e^{-im}} \left(m - \frac{1-e^{-im}}{i} \right) \\
&= \frac{m}{1-e^{-im}} - \frac{1}{i}
\end{aligned}$$

because we have:

$$\begin{aligned}
\int_0^m v e^{iv} \,\mathrm{d}v &= \left[\frac{v e^{iv}}{i} \right]_0^m - \int_0^m \frac{e^{iv}}{i} \,\mathrm{d}v \\
&= \left[\frac{v e^{iv}}{i} \right]_0^m - \left[\frac{e^{iv}}{i^2} \right]_0^m \\
&= \frac{me^{im}}{i} - \frac{e^{im}-1}{i^2}
\end{aligned}$$

7.4.4 Valuation of non-maturity deposits

1. The current market value of liabilities is the expected discounted value of future cash flows, which are made up of interest payments $i(t)\, D(t)$ and deposit inflows $\partial_t D(t)$:

$$L_0 = \mathbb{E}\left[\int_0^\infty e^{-r(t)t} \left(i(t)\, D(t) - \partial_t D(t) \right) \,\mathrm{d}t \right] \tag{7.1}$$

2. Since we have:

$$\begin{aligned}
\int_0^\infty e^{-r(t)t} \left(i(t)\, D(t) - \partial_t D(t) \right) \,\mathrm{d}t &= \int_0^\infty e^{-r(t)t} i(t)\, D(t) \,\mathrm{d}t - \\
&\quad \int_0^\infty e^{-r(t)t} \partial_t D(t) \,\mathrm{d}t
\end{aligned}$$

and:

$$\begin{aligned}
\int_0^\infty e^{-r(t)t} \partial_t D(t) \,\mathrm{d}t &= \left[e^{-r(t)t} D(t) \right]_0^\infty + \int_0^\infty e^{-r(t)t} r(t)\, D(t) \,\mathrm{d}t \\
&= -D_0 + \int_0^\infty e^{-r(t)t} r(t)\, D(t) \,\mathrm{d}t
\end{aligned}$$

we deduce that:

$$\begin{aligned}
L_0 &= \mathbb{E}\left[\int_0^\infty e^{-r(t)t} \left(i(t)\, D(t) - \partial_t D(t) \right) \,\mathrm{d}t \right] \\
&= D_0 + \mathbb{E}\left[\int_0^\infty e^{-r(t)t} \left(i(t) - r(t) \right) D(t) \,\mathrm{d}t \right] \qquad (7.2)
\end{aligned}$$

3. The current value of deposit accounts is the difference between the current value of deposits D_0 and the current value of liabilities L_0:

$$\begin{aligned} V_0 &= D_0 - L_0 \\ &= \mathbb{E}\left[\int_0^\infty e^{-r(t)t}\left(r(t) - i(t)\right) D(t) \,\mathrm{d}t\right] \\ &= \mathbb{E}\left[\int_0^\infty e^{-r(t)t} m(t) D(t) \,\mathrm{d}t\right] \end{aligned} \tag{7.3}$$

where $m(t) = r(t) - i(t)$ is the margin of the bank. This is the equation obtained by Jarrow and Van Deventer (1998), who notice that V_0 is "*the net present value of an exotic interest rate swap paying floating at $i(t)$ and receiving floating at $r(t)$ on a random principal of $D(t)$*".

4. If the margin $m(t)$ is constant, we obtain:

$$\begin{aligned} V_0 &= \mathbb{E}\left[\int_0^\infty e^{-r(t)t} m(t) D(t) \,\mathrm{d}t\right] \\ &= m_0 \mathbb{E}\left[\int_0^\infty e^{-r(t)t} \,\mathrm{d}t\right] D_\infty \\ &= m_0 r_\infty^{-1} D_\infty \end{aligned} \tag{7.4}$$

where r_∞ can be interpreted as the average market rate[3]:

$$r_\infty = \frac{1}{\mathbb{E}\left[\int_0^\infty e^{-r(t)t} \,\mathrm{d}t\right]}$$

5. The variation of $i(t)$ is equal to a constant α plus a linear correction term $\beta(r(t) - i(t))$:

$$\begin{aligned} \frac{\mathrm{d}i(t)}{\mathrm{d}t} &= \alpha + \beta(r(t) - i(t)) \\ &= \beta r(t) + (\alpha - \beta i(t)) \end{aligned}$$

It follows that $i(t)$ is an increasing function of $r(t)$. Moreover, $i(t)$ decreases (resp. increases) if $\alpha - \beta i(t) \leq 0$ (resp. $\alpha - \beta i(t) > 0$). This implies that $i(t)$ is a mean-reverting process, where the steady state is $i_\infty = \beta^{-1}\alpha$. The variation of $D(t)$ is explained by two components:

$$\frac{\mathrm{d}D(t)}{\mathrm{d}t} = \underbrace{\gamma(D_\infty - D(t))}_{C_1(t)} - \underbrace{\delta(r(t) - i(t))}_{C_2(t)}$$

[3]In the case where $r(t)$ is constant, we notice that:

$$\begin{aligned} \mathbb{E}\left[\int_0^\infty e^{-r(t)t} \,\mathrm{d}t\right] &= \int_0^\infty e^{-rt} \,\mathrm{d}t \\ &= \left[-\frac{e^{-rt}}{r}\right]_0^\infty \\ &= \frac{1}{r} \end{aligned}$$

This justifies that r_∞ is an average interest rate.

The first component $C_1(t)$ is the traditional mean-reverting adjustment between the deposit $D(t)$ and its long-term value D_∞, whereas the second component $C_2(t)$ is the negative impact of the excess of the market rate over the savings rate. It follows that:

$$i(t) = e^{-\beta t} i_0 + \beta \int_0^t e^{-\beta(t-s)} \left(r(s) + \frac{\alpha}{\beta} \right) \mathrm{d}s \tag{7.5}$$

and:

$$D(t) = e^{-\gamma t} D_0 + \left(1 - e^{-\gamma t}\right) D_\infty - \delta \int_0^t e^{-\gamma(t-s)} \left(r(s) - i(s) \right) \mathrm{d}s \tag{7.6}$$

6. In the case where $r(t)$ is constant and equal to r_0, we obtain:

$$\begin{aligned} i(t) &= e^{-\beta t} i_0 + (\alpha + \beta r_0) \int_0^t e^{-\beta(t-s)} \mathrm{d}s \\ &= e^{-\beta t} i_0 + (\alpha + \beta r_0) \left[\frac{e^{-\beta(t-s)}}{\beta} \right]_0^t \\ &= e^{-\beta t} i_0 + \left(1 - e^{-\beta t}\right) \left(r_0 + \frac{\alpha}{\beta} \right) \\ &= i_0 + \left(1 - e^{-\beta t}\right) \left(r_0 + \frac{\alpha}{\beta} - i_0 \right) \end{aligned} \tag{7.7}$$

It follows that:

$$\begin{aligned} r(t) - i(t) &= r_0 - \left(e^{-\beta t} i_0 + \left(1 - e^{-\beta t}\right) \left(r_0 + \frac{\alpha}{\beta} \right) \right) \\ &= e^{-\beta t} (r_0 - i_0) - \frac{\alpha}{\beta} \left(1 - e^{-\beta t}\right) \end{aligned} \tag{7.8}$$

and:

$$\begin{aligned} D(t) = {} & D_\infty + e^{-\gamma t} (D_0 - D_\infty) - \delta (r_0 - i_0) \int_0^t e^{-\gamma(t-s)} e^{-\beta s} \mathrm{d}s + {} \\ & \frac{\alpha\delta}{\beta} \int_0^t e^{-\gamma(t-s)} \left(1 - e^{-\beta s}\right) \mathrm{d}s \end{aligned} \tag{7.9}$$

Since we have:

$$\int_0^t e^{-\gamma(t-s)} e^{-\beta s} \mathrm{d}s = \left[\frac{e^{-\gamma(t-s) - \beta s}}{\gamma - \beta} \right]_0^t$$

and:

$$\int_0^t e^{-\gamma(t-s)} \left(1 - e^{-\beta s}\right) \mathrm{d}s = \left[\frac{e^{-\gamma(t-s)}}{\gamma} - \frac{e^{-\gamma(t-s) - \beta s}}{\gamma - \beta} \right]_0^t$$

we deduce that:

$$\begin{aligned} D(t) = {} & D_\infty + e^{-\gamma t} (D_0 - D_\infty) - \frac{\delta \left(e^{-\beta t} - e^{-\gamma t}\right)}{\gamma - \beta} (r_0 - i_0) + {} \\ & \frac{\alpha\delta}{\beta} \left(\frac{1 - e^{-\gamma t}}{\gamma} - \frac{e^{-\beta t} - e^{-\gamma t}}{\gamma - \beta} \right) \end{aligned}$$

7. If α is equal to zero and we combine Equations (7.8) and (7.9), we obtain:

$$\begin{aligned} e^{-r_0 t}\left(r(t)-i(t)\right)D(t) &= e^{-(r_0+\beta)t}\left(r_0-i_0\right)D_\infty + \\ & e^{-(r_0+\beta+\gamma)t}\left(r_0-i_0\right)\left(D_0-D_\infty\right) - \\ & \frac{\delta\left(e^{-(r_0+2\beta)t}-e^{-(r_0+\beta+\gamma)t}\right)}{\gamma-\beta}\left(r_0-i_0\right)^2 \end{aligned}$$

It follows that:

$$\begin{aligned} V_0 &= \int_0^\infty e^{-(r_0+\beta)t}\left(r_0-i_0\right)D_\infty \,\mathrm{d}t + \\ & \int_0^\infty e^{-(r_0+\gamma+\beta)t}\left(r_0-i_0\right)\left(D_0-D_\infty\right)\mathrm{d}t - \\ & \int_0^\infty \frac{\delta e^{-(r_0+2\beta)t}}{\gamma-\beta}\left(r_0-i_0\right)^2 \mathrm{d}t + \\ & \int_0^\infty \frac{\delta e^{-(r_0+\gamma+\beta)t}}{\gamma-\beta}\left(r_0-i_0\right)^2 \mathrm{d}t \end{aligned}$$

We also have:

$$\begin{aligned} V_0 &= \left[\frac{e^{-(r_0+\beta)t}\left(r_0-i_0\right)D_\infty}{-\left(r_0+\beta\right)}\right]_0^\infty + \\ & \left[\frac{e^{-(r_0+\gamma+\beta)t}\left(r_0-i_0\right)\left(D_0-D_\infty\right)}{-\left(r_0+\gamma+\beta\right)}\right]_0^\infty - \\ & \left[\frac{\delta e^{-(r_0+2\beta)t}\left(r_0-i_0\right)^2}{-\left(r_0+2\beta\right)\left(\gamma-\beta\right)}\right]_0^\infty + \left[\frac{\delta e^{-(r_0+\gamma+\beta)t}\left(r_0-i_0\right)^2}{-\left(r_0+\gamma+\beta\right)\left(\gamma-\beta\right)}\right]_0^\infty \end{aligned}$$

Therefore, the net asset value is equal to:

$$\begin{aligned} V_0 &= \frac{\left(r_0-i_0\right)D_\infty}{\left(r_0+\beta\right)} + \frac{\left(r_0-i_0\right)\left(D_0-D_\infty\right)}{\left(r_0+\gamma+\beta\right)} + \\ & \frac{\delta\left(r_0-i_0\right)^2}{\left(r_0+\gamma+\beta\right)\left(\gamma-\beta\right)} - \frac{\delta\left(r_0-i_0\right)^2}{\left(r_0+2\beta\right)\left(\gamma-\beta\right)} \end{aligned} \tag{7.10}$$

We deduce that the sensitivity of V_0 with respect to r_0 is equal to:

$$\begin{aligned} \frac{\partial V_0}{\partial r_0} &= \frac{\left(i_0+\beta\right)D_\infty}{\left(r_0+\beta\right)^2} + \frac{\left(D_0-D_\infty\right)\left(i_0+\gamma+\beta\right)}{\left(r_0+\gamma+\beta\right)^2} + \\ & \frac{\delta\left(r_0-i_0\right)\left(r_0+i_0+2\left(\gamma+\beta\right)\right)}{\left(r_0+\gamma+\beta\right)^2\left(\gamma-\beta\right)} - \\ & \frac{\delta\left(r_0-i_0\right)\left(r_0++i_0+4\beta\right)}{\left(r_0+2\beta\right)^2\left(\gamma-\beta\right)} \end{aligned} \tag{7.11}$$

8. From Equation (7.3), we deduce that:

$$\begin{aligned} \frac{\partial V_0}{\partial r(t)} &= \mathbb{E}\left[-\int_0^\infty t e^{-r(t)t}\left(r(t)-i(t)\right)D(t)\,\mathrm{d}t\right] + \\ & \mathbb{E}\left[\int_0^\infty e^{-r(t)t}\frac{\partial\left(r(t)-i(t)\right)}{\partial r(t)}D(t)\,\mathrm{d}t\right] + \\ & \mathbb{E}\left[\int_0^\infty e^{-r(t)t}\left(r(t)-i(t)\right)\frac{\partial D(t)}{\partial r(t)}\,\mathrm{d}t\right] \end{aligned} \tag{7.12}$$

De Jong and Wielhouwer (2003) observe that the sensitivity of the net asset value is the sum of three components: the interest rate sensitivity of the expected discount margins, the margin sensitivity with respect to the market rate, and the impact of $r(t)$ on the deposit balance $D(t)$. Since we have:

$$\frac{\partial\,(r(t)-i(t))}{\partial\,r_0}=e^{-\beta t}$$

and:

$$\frac{\partial\,D(t)}{\partial\,r_0}=-\frac{\delta\left(e^{-\beta t}-e^{-\gamma t}\right)}{\gamma-\beta}$$

we deduce that:

$$\begin{aligned}\frac{\partial\,V_0}{\partial\,r_0} &= -(r_0-i_0)\int_0^\infty te^{-(r_0+\beta)t}D(t)\;\mathrm{d}t+\\ &\quad \alpha\int_0^\infty te^{-r_0t}\left(\frac{1-e^{-\beta t}}{\beta}\right)D(t)\;\mathrm{d}t+\int_0^\infty e^{-(r_0+\beta)t}D(t)\;\mathrm{d}t-\\ &\quad \frac{\delta(r_0-i_0)}{\gamma-\beta}\int_0^\infty\left(e^{-(r_0+2\beta)t}-e^{-(r_0+\beta+\gamma)t}\right)\mathrm{d}t+\\ &\quad \frac{\alpha\delta}{\gamma-\beta}\int_0^\infty\left(e^{-(r_0+\beta)t}-e^{-(r_0+\gamma)t}\right)\left(\frac{1-e^{-\beta t}}{\beta}\right)\mathrm{d}t \end{aligned} \tag{7.13}$$

This sensitivity can be computed analytically, but it is a complex formula with many terms. This is why it is better to calculate it using the Gauss-Legendre numerical integration method. The duration of deposits is then defined as:

$$\mathcal{D}_D=-\frac{1}{V_0}\frac{\partial\,V_0}{\partial\,r_0}$$

9. In Figure 7.1, we have represented the deposit rate $i(t)$ with respect to the time t. We notice that:

$$\begin{aligned}\lim_{t\to\infty}i(t) &= \lim_{t\to\infty}i_0+\left(1-e^{-\beta t}\right)\left(r_0+\frac{\alpha}{\beta}-i_0\right)\\ &= r_0+\frac{\alpha}{\beta}\end{aligned}$$

Since the margin is equal to $r(t)-i(t)$, it is natural to assume that $\alpha<0$ in order to verify the condition[4] $i(t)<r(t)$. The dynamics of $D(t)$ is given in Figure 7.2. It depends on the relative position between D_0 and D_∞. Another important parameter is the mean-reverting coefficient γ. In Figure 7.3, we have represented the mark-to-market V_0, its sensitivity with respect to r_0 and the corresponding duration. We notice that the normal case where $i_0<r_0$ corresponds to a negative duration, because the sensitivity is positive. We explain this result because $\alpha=0$ is not realistic, meaning that the margin is equal to zero on average. If we assume that α is negative or the margin is positive, we obtain a positive duration (see Figure 7.4). In particular, we verify that the duration of deposits is higher when market rates are low.

[4] This is the arbitrage condition found by Jarrow and van Deventer (1998).

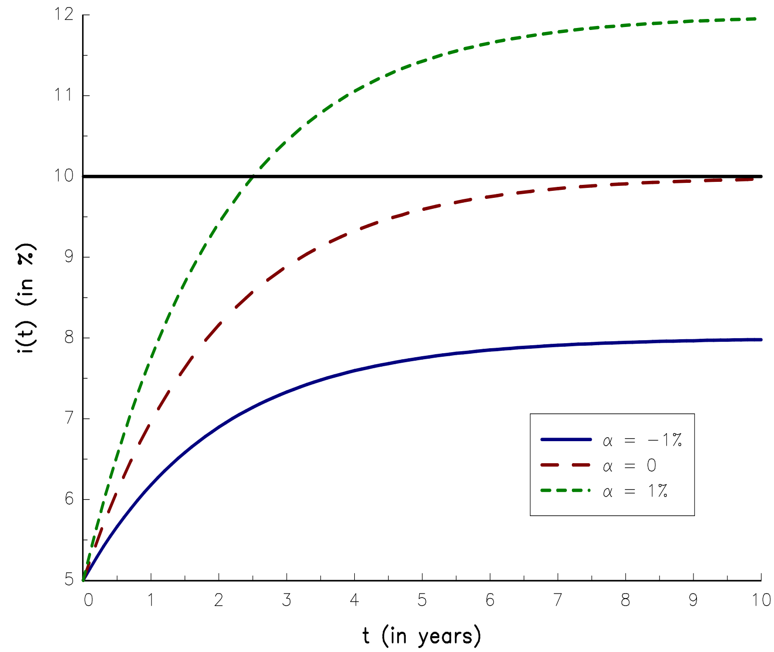

FIGURE 7.1: Dynamics of the deposit rate $i(t)$

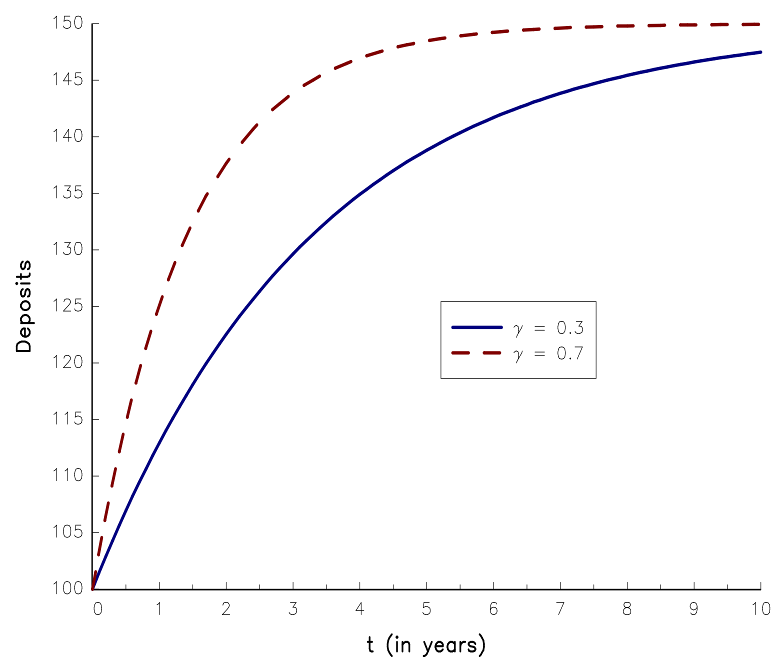

FIGURE 7.2: Dynamics of the deposit balance $D(t)$

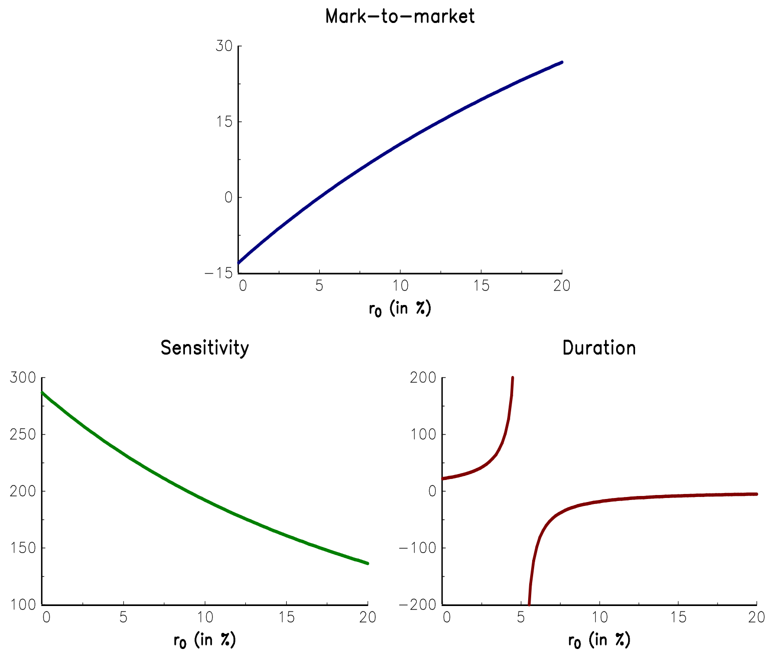

FIGURE 7.3: Duration of deposits when α is equal to zero

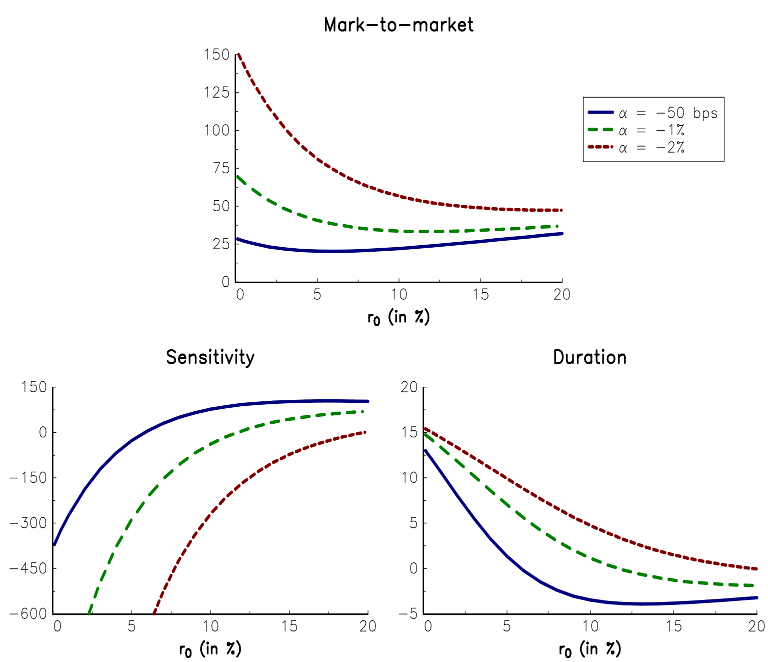

FIGURE 7.4: Duration of deposits when the margin is positive

7.4.5 Impact of prepayment on the amortization scheme of the CAM

1. We deduce that the dynamics of $N(t)$ is equal to:

$$\begin{aligned} \mathrm{d}N(t) &= \mathbb{1}\{t<m\}\cdot N_0 \frac{-ie^{-i(m-t)}}{1-e^{-im}}\,\mathrm{d}t \\ &= -ie^{-i(m-t)}\left(\mathbb{1}\{t<m\}\cdot N_0\frac{1}{1-e^{-im}}\right)\mathrm{d}t \\ &= -\frac{ie^{-i(m-t)}}{1-e^{-i(m-t)}}N(t)\,\mathrm{d}t \end{aligned}$$

2. The prepayment rate has a negative impact on $\mathrm{d}N(t)$ because it reduces the outstanding amount $N(t)$:

$$\mathrm{d}\tilde{N}(t) = -\frac{ie^{-i(m-t)}}{1-e^{-i(m-t)}}\tilde{N}(t)\,\mathrm{d}t - \lambda_p(t)\,\tilde{N}(t)\,\mathrm{d}t$$

3. It follows that:

$$\mathrm{d}\ln\tilde{N}(t) = -\left(\frac{ie^{-i(m-t)}}{1-e^{-i(m-t)}} + \lambda_p(t)\right)\mathrm{d}t$$

and:

$$\begin{aligned} \ln\tilde{N}(t) - \ln\tilde{N}(0) &= \int_0^t \frac{-ie^{-i(m-s)}}{1-e^{-i(m-s)}}\,\mathrm{d}s - \int_0^t \lambda_p(s)\,\mathrm{d}s \\ &= \left[\ln\left(1-e^{-i(m-s)}\right)\right]_0^t - \int_0^t \lambda_p(s)\,\mathrm{d}s \\ &= \ln\left(\frac{1-e^{-i(m-t)}}{1-e^{-im}}\right) - \int_0^t \lambda_p(s)\,\mathrm{d}s \end{aligned}$$

and:

$$\begin{aligned} \tilde{N}(t) &= \left(N_0\frac{1-e^{-i(m-t)}}{1-e^{-im}}\right)e^{-\int_0^t \lambda_p(s)\,\mathrm{d}s} \\ &= N(t)\,\mathbf{S}_p(t) \end{aligned}$$

where $\mathbf{S}_p(t)$ is the survival function associated to the hazard rate $\lambda_p(t)$.

4. We have:

$$\tilde{N}(t,u) = \mathbf{1}\{t \le u < t+m\}\cdot N(t)\,\frac{1-e^{-i(t+m-u)}}{1-e^{-im}}e^{-\lambda_p(u-t)}$$

this implies that:

$$\tilde{\mathbf{S}}(t,u) = \mathbf{1}\{t \le u < t+m\}\cdot\frac{e^{-\lambda_p(u-t)} - e^{-im+(i-\lambda_p)(u-t)}}{1-e^{-im}}$$

and:

$$\tilde{f}(t,u) = \mathbf{1}\{t \le u < t+m\}\cdot\frac{\lambda_p e^{-\lambda_p(u-t)} + (i-\lambda_p)\,e^{-im+(i-\lambda_p)(u-t)}}{1-e^{-im}}$$

It follows that:

$$\begin{aligned}
\tilde{\mathcal{D}}(t) &= \frac{\lambda_p}{1-e^{-im}}\int_t^{t+m}(u-t)\,e^{-\lambda_p(u-t)}\,\mathrm{d}u + \\
&\quad \frac{(i-\lambda_p)\,e^{-im}}{1-e^{-im}}\int_t^{t+m}(u-t)\,e^{(i-\lambda_p)(u-t)}\,\mathrm{d}u \\
&= \frac{\lambda_p}{1-e^{-im}}\int_0^m ve^{-\lambda_p v}\,\mathrm{d}v + \frac{(i-\lambda_p)\,e^{-im}}{1-e^{-im}}\int_0^m ve^{(i-\lambda_p)v}\,\mathrm{d}v \\
&= \frac{\lambda_p}{1-e^{-im}}\left(\frac{me^{-\lambda_p m}}{-\lambda_p} - \frac{e^{-\lambda_p m}-1}{\lambda_p^2}\right) + \\
&\quad \frac{(i-\lambda_p)\,e^{-im}}{1-e^{-im}}\left(\frac{me^{(i-\lambda_p)m}}{(i-\lambda_p)} - \frac{e^{(i-\lambda_p)m}-1}{(i-\lambda_p)^2}\right) \\
&= \frac{1}{1-e^{-im}}\left(\frac{e^{-im}-e^{-\lambda_p m}}{i-\lambda_p} + \frac{1-e^{-\lambda_p m}}{\lambda_p}\right)
\end{aligned}$$

because we have:

$$\begin{aligned}
\int_0^m ve^{\alpha v}\,\mathrm{d}v &= \left[\frac{ve^{\alpha v}}{\alpha}\right]_0^m - \int_0^m \frac{e^{\alpha v}}{\alpha}\,\mathrm{d}v \\
&= \left[\frac{ve^{\alpha v}}{\alpha}\right]_0^m - \left[\frac{e^{\alpha v}}{\alpha^2}\right]_0^m \\
&= \frac{me^{\alpha m}}{\alpha} - \frac{e^{\alpha m}-1}{\alpha^2}
\end{aligned}$$

Chapter 8

Systemic Risk and Shadow Banking System

Part II

Mathematical and Statistical Tools

Chapter 9

Model Risk of Exotic Derivatives

9.4.1 Option pricing and martingale measure

1. Since we have:
$$V(t) = \phi(t) S(t) + \psi(t) B(t)$$
we deduce that:
$$\psi(t) = \frac{V(t) - \phi(t) S(t)}{B(t)}$$
It follows that:
$$\begin{aligned} \mathrm{d}V(t) &= \phi(t)\ \mathrm{d}S(t) + \psi(t)\ \mathrm{d}B(t) \\ &= \phi(t)\ \mathrm{d}S(t) + r(V(t) - \phi(t) S(t))\ \mathrm{d}t \\ &= rV(t)\ \mathrm{d}t + \phi(t)(\mathrm{d}S(t) - rS(t)\ \mathrm{d}t) \end{aligned}$$

2. We have:
$$\begin{aligned} \mathrm{d}\tilde{S}(t) &= -re^{-rt} S(t)\ \mathrm{d}t + e^{-rt}\,\mathrm{d}S(t) \\ &= e^{-rt}(\mathrm{d}S(t) - rS(t)\ \mathrm{d}t) \end{aligned}$$
It follows that:
$$\begin{aligned} \mathrm{d}V(t) &= rV(t)\ \mathrm{d}t + \phi(t)(\mathrm{d}S(t) - rS(t)\ \mathrm{d}t) \\ &= rV(t)\ \mathrm{d}t + e^{rt}\phi(t)\ \mathrm{d}\tilde{S}(t) \end{aligned}$$
Finally, we deduce that:
$$\begin{aligned} \mathrm{d}\tilde{V}(t) &= -re^{-rt} V(t)\ \mathrm{d}t + e^{-rt}\,\mathrm{d}V(t) \\ &= -re^{-rt} V(t)\ \mathrm{d}t + e^{-rt}\left(rV(t)\ \mathrm{d}t + e^{rt}\phi(t)\ \mathrm{d}\tilde{S}(t)\right) \\ &= \phi(t)\ \mathrm{d}\tilde{S}(t) \end{aligned}$$

3. Under the probability measure $\mathbb{Q}$, we remind that:
$$\mathrm{d}S(t) = rS(t)\ \mathrm{d}t + \sigma S(t)\ \mathrm{d}W^{\mathbb{Q}}(t)$$
Then, we have:
$$\begin{aligned} \mathrm{d}\tilde{S}(t) &= e^{-rt}(\mathrm{d}S(t) - rS(t)\ \mathrm{d}t) \\ &= e^{-rt}\sigma S(t)\ \mathrm{d}W^{\mathbb{Q}}(t) \\ &= \sigma\tilde{S}(t)\ \mathrm{d}W^{\mathbb{Q}}(t) \end{aligned}$$
We conclude that $\tilde{S}(t)$ is a martingale. Since $\mathrm{d}\tilde{V}(t) = \phi(t)\ \mathrm{d}\tilde{S}(t)$, $\tilde{V}(t)$ is also a martingale. We deduce that:
$$\tilde{V}(t) = \mathbb{E}^{\mathbb{Q}}\left[\left.\tilde{V}(T)\right|\mathcal{F}_t\right]$$

and:

$$\begin{aligned} V(t) &= e^{rt}\mathbb{E}^{\mathbb{Q}}\left[\left. e^{-rT}V(T)\right| \mathcal{F}_t\right] \\ &= e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[\left. V(T)\right| \mathcal{F}_t\right] \end{aligned}$$

4. We have:

$$\mathrm{d}S(t) = \mu S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t)$$

and:

$$\begin{aligned} \mathrm{d}\tilde{S}(t) &= e^{-rt}\left(\mathrm{d}S(t) - rS(t)\,\mathrm{d}t\right) \\ &= (\mu - r)\tilde{S}(t)\,\mathrm{d}t + \sigma\tilde{S}(t)\,\mathrm{d}W(t) \end{aligned}$$

We set:

$$W^{\mathbb{Q}}(t) = W(t) + \left(\frac{\mu - r}{\sigma}\right)t$$

Using Girsanov's theorem, we know that $W^{\mathbb{Q}}(t)$ is a Brownian motion under the probability measure $\mathbb{Q}$ defined by:

$$\begin{aligned} \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} &= M(t) \\ &= \exp\left(-\frac{1}{2}\int_0^t\left(\frac{\mu - r}{\sigma}\right)^2\mathrm{d}s - \int_0^t\left(\frac{\mu - r}{\sigma}\right)\mathrm{d}W(s)\right) \end{aligned}$$

Moreover, we know that $M(t)$ is an $\mathcal{F}_t$-martingale.

5. We have:

$$V(T) = \mathbb{1}\left\{S(T) \geq K\right\}$$

and:

$$\begin{aligned} S(T) \geq K &\Leftrightarrow S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma W^{\mathbb{Q}}(T)} \geq K \\ &\Leftrightarrow W^{\mathbb{Q}}(T) \geq \frac{1}{\sigma}\left(\ln K - \ln S_0 - \left(r - \frac{1}{2}\sigma^2\right)T\right) \end{aligned}$$

We deduce that:

$$\begin{aligned} V(0) &= e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}\left\{S(T) \geq K\right\}\right] \\ &= e^{-rT}\Pr\left\{S(T) \geq K\right\} \\ &= e^{-rT}\Pr\left\{W^{\mathbb{Q}}(T) \geq \frac{1}{\sigma}\left(\ln K - \ln S_0 - \left(r - \frac{1}{2}\sigma^2\right)T\right)\right\} \\ &= e^{-rT}\Phi\left(-\frac{1}{\sigma\sqrt{T}}\left(\ln K - \ln S_0 - \left(r - \frac{1}{2}\sigma^2\right)T\right)\right) \end{aligned}$$

Therefore, the price of the binary option is:

$$V(0) = e^{-rT}\Phi\left(\frac{1}{\sigma\sqrt{T}}\left(\ln\frac{S_0}{K} + rT\right) - \frac{1}{2}\sigma\sqrt{T}\right)$$

9.4.2 The Vasicek model

1. We have:

$$\frac{1}{2}\sigma^2\frac{\partial^2 B(t,r)}{\partial r^2} + \left(a\left(b - r(t)\right) - \lambda(t)\sigma\right)\frac{\partial B(t,r)}{\partial r} + \frac{\partial B(t,r)}{\partial t} - r(t)B(t,r) = 0$$

and:

$$B(T, r(T)) = 1$$

2. We remind that the solution of the Ornstein-Uhlenbeck process is:

$$r(t) = r_0 e^{-at} + b\left(1 - e^{-at}\right) + \sigma\int_0^t e^{a(s-t)}\,\mathrm{d}W(s)$$

It follows that:

$$\begin{aligned} Z &= \int_0^T \left(r_0 e^{-at} + b\left(1 - e^{-at}\right) + \sigma\int_0^t e^{a(s-t)}\,\mathrm{d}W(s)\right)\mathrm{d}t \\ &= r_0\left[-\frac{e^{-at}}{a}\right]_0^T + b\left[t + \frac{e^{-at}}{a}\right]_0^T + \sigma\int_0^T\int_0^t e^{a(s-t)}\,\mathrm{d}W(s)\;\mathrm{d}t \\ &= bT + (r_0 - b)\left(\frac{1 - e^{-aT}}{a}\right) + \sigma\int_0^T\int_0^t e^{a(s-t)}\,\mathrm{d}W(s)\;\mathrm{d}t \end{aligned}$$

We note $I = \int_0^T\int_0^t e^{a(s-t)}\,\mathrm{d}W(s)\;\mathrm{d}t$. Using Fubini's theorem for stochastic integrals, we have:

$$\begin{aligned} I &= \int_0^T\int_t^T e^{a(t-s)}\,\mathrm{d}s\,\mathrm{d}W(t) \\ &= \int_0^T \frac{1 - e^{-a(T-t)}}{a}\,\mathrm{d}W(t) \end{aligned}$$

Since I is a sum of independent Gaussian random variables, it follows that Z is also a Gaussian random variable.

3. We have:

$$\mathbb{E}[Z] = bT + (r_0 - b)\left(\frac{1 - e^{-aT}}{a}\right)$$

and:

$$\begin{aligned} \operatorname{var}(Z) &= \mathbb{E}\left[\sigma\int_0^T \frac{1 - e^{-a(T-t)}}{a}\,\mathrm{d}W(t)\right]^2 \\ &= \frac{\sigma^2}{a^2}\int_0^T\left(1 - e^{-a(T-t)}\right)^2\mathrm{d}t \\ &= \frac{\sigma^2}{a^2}\int_0^T\left(1 - 2e^{-a(T-t)} + e^{-2a(T-t)}\right)\mathrm{d}t \\ &= \frac{\sigma^2}{a^2}\left(T - \frac{2}{a}\left(1 - e^{-aT}\right) + \frac{1}{2a}\left(1 - e^{-2aT}\right)\right) \end{aligned}$$

Another expression is:

$$\operatorname{var}(Z) = \frac{\sigma^2}{a^2}\left(T - \beta - \frac{a\beta^2}{2}\right)$$

where:

$$\beta = \frac{1 - e^{-aT}}{a}$$

4. We have:

$$\begin{aligned} B(0, r_0) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(t)\,\mathrm{d}t} \middle| \mathcal{F}_0 \right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-Z} \middle| \mathcal{F}_0 \right] \end{aligned}$$

Under the probability measure $\mathbb{Q}$, $r(t)$ is an Ornstein-Uhlenbeck process:

$$\begin{aligned} \mathrm{d}r(t) &= a(b - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \\ &= a(b - r(t))\,\mathrm{d}t + \sigma\left(\mathrm{d}W^{\mathbb{Q}}(t) - \lambda\,\mathrm{d}t\right) \\ &= (a(b - r(t)) - \lambda\sigma)\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t) \\ &= a(b' - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t) \end{aligned}$$

where:

$$b' = b - \frac{\lambda\sigma}{a}$$

It follows that:

$$B(0, r_0) = e^{-\mathbb{E}^{\mathbb{Q}}[Z] + \frac{1}{2}\operatorname{var}^{\mathbb{Q}}(Z)}$$

and:

$$\begin{aligned} -\mathbb{E}^{\mathbb{Q}}[Z] + \frac{1}{2}\operatorname{var}^{\mathbb{Q}}(Z) &= -b'T - (r_0 - b')\beta + \frac{\sigma^2}{2a^2}\left(T - \beta - \frac{a\beta^2}{2}\right) \\ &= -r_0\beta - b'(T - \beta) + \frac{\sigma^2}{2a^2}(T - \beta) - \frac{\sigma^2\beta^2}{4a} \\ &= -r_0\beta - \left(b' - \frac{\sigma^2}{2a^2}\right)(T - \beta) - \frac{\sigma^2\beta^2}{4a} \end{aligned}$$

Finally, we obtain:

$$B(0, r_0) = \exp\left(-r_0\beta - \left(b' - \frac{\sigma^2}{2a^2}\right)(T - \beta) - \frac{\sigma^2\beta^2}{4a}\right)$$

9.4.3 The Black model

1. We have:

$$\left\{ \begin{array}{l} \frac{1}{2}\sigma^2 F^2 \partial_F^2 \mathcal{C}(t, F) + \partial_t \mathcal{C}(t, F) - r\mathcal{C}(t, F) = 0 \\ \mathcal{C}(T, S(T)) = \max(F(T) - K, 0) \end{array} \right.$$

2. The Feynman-Kac formula is:

$$\mathcal{C}(0) = e^{-rT}\mathbb{E}\left[\max(F(T) - K, 0) \middle| \mathcal{F}_0\right]$$

We know that $F(T)$ is a log-normal random variable:

$$F(T) = F_0 e^{-\frac{1}{2}\sigma^2 T + \sigma(W(T) - W(0))}$$

We note $I = \mathbb{E}\left[\max\left(F\left(T\right) - K, 0\right) \middle| \mathcal{F}_0\right]$. We obtain:

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \left(F_0 e^{-\frac{1}{2}\sigma^2 T + \sigma\sqrt{T}u} - K\right)^+ \phi\left(u\right)\,\mathrm{d}u \\ &= \int_{d}^{\infty} \left(F_0 e^{-\frac{1}{2}\sigma^2 T + \sigma\sqrt{T}u} - K\right) \phi\left(u\right)\,\mathrm{d}u \\ &= F_0 e^{-\frac{1}{2}\sigma^2 T} \int_{d}^{\infty} e^{\sigma\sqrt{T}u} \phi\left(u\right)\,\mathrm{d}u - K \int_{d}^{\infty} \phi\left(u\right)\,\mathrm{d}u \end{aligned}$$

where:

$$d = -\frac{1}{\sigma\sqrt{T}} \ln \frac{F_0}{K} + \frac{1}{2}\sigma\sqrt{T}$$

We have:

$$\begin{aligned} \int_{d}^{\infty} \phi\left(u\right)\,\mathrm{d}u &= 1 - \Phi\left(d\right) \\ &= \Phi\left(-d\right) \end{aligned}$$

and:

$$\begin{aligned} \int_{d}^{\infty} e^{\sigma\sqrt{T}u} \phi\left(u\right)\,\mathrm{d}u &= \int_{d}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\sigma\sqrt{T}u - \frac{1}{2}u^2}\,\mathrm{d}u \\ &= e^{\frac{1}{2}\sigma^2 T} \int_{d}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(u - \sigma\sqrt{T}\right)^2}\,\mathrm{d}u \\ &= e^{\frac{1}{2}\sigma^2 T} \int_{d-\sigma\sqrt{T}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}\,\mathrm{d}v \\ &= e^{\frac{1}{2}\sigma^2 T} \left(1 - \Phi\left(d - \sigma\sqrt{T}\right)\right) \\ &= e^{\frac{1}{2}\sigma^2 T} \Phi\left(-d + \sigma\sqrt{T}\right) \end{aligned}$$

Finally, we deduce that:

$$\begin{aligned} \mathcal{C}\left(0\right) &= e^{-rT} \left(F_0 e^{-\frac{1}{2}\sigma^2 T} e^{\frac{1}{2}\sigma^2 T} \Phi\left(-d + \sigma\sqrt{T}\right) - K\Phi\left(-d\right)\right) \\ &= F_0 e^{-rT} \Phi\left(d_1\right) - K e^{-rT} \Phi\left(d_2\right) \end{aligned}$$

where:

$$d_1 = \frac{1}{\sigma\sqrt{T}} \ln \frac{F_0}{K} + \frac{1}{2}\sigma\sqrt{T}$$

and:

$$d_2 = \frac{1}{\sigma\sqrt{T}} \ln \frac{F_0}{K} - \frac{1}{2}\sigma\sqrt{T}$$

3. Under the risk-neutral probability measure $\mathbb{Q}$, we have:

$$\mathrm{d}S\left(t\right) = rS\left(t\right)\,\mathrm{d}t + \sigma S\left(t\right)\,\mathrm{d}W^{\mathbb{Q}}\left(t\right)$$

The price of a future contract on this stock is equal to:

$$F\left(t\right) = e^{-rt} S\left(t\right)$$

Using Ito's lemma, we deduce that:

$$\begin{aligned}\mathrm{d}F(t) &= -re^{-rt}S(t)\,\mathrm{d}t + e^{-rt}\,\mathrm{d}S(t) \\ &= \sigma e^{-rt}S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) \\ &= \sigma F(t)\,\mathrm{d}W^{\mathbb{Q}}(t)\end{aligned}$$

We can then apply the Black formula to price an European option on $F(t)$.

4. In this case, the PDE representation becomes:

$$\frac{1}{2}\sigma^2 F^2 \partial_F^2 \boldsymbol{\mathcal{C}}(t,F) + \partial_t \boldsymbol{\mathcal{C}}(t,F) - r(t)\,\boldsymbol{\mathcal{C}}(t,F) = 0$$

It follows that the Feynman-Kac formula is:

$$\boldsymbol{\mathcal{C}}(0) = \mathbb{E}\left[\left. e^{-\int_0^T r(s)\,\mathrm{d}s} \max\left(F(T) - K, 0\right)\right| \mathcal{F}_0\right]$$

Since $r(t)$ and $F(t)$ are independent, we obtain:

$$\begin{aligned}\boldsymbol{\mathcal{C}}(0) &= \mathbb{E}\left[\left. e^{-\int_0^T r(s)\,\mathrm{d}s}\right| \mathcal{F}_0\right] \cdot \mathbb{E}\left[\max\left(F(T) - K, 0\right) \middle| \mathcal{F}_0\right] \\ &= B(0,T)\cdot\left(F_0\Phi(d_1) - K\Phi(d_2)\right)\end{aligned}$$

We deduce that the discount factor e^{-rT} is replaced by the current bond price $B(0,T)$.

5. If $r(t)$ and $F(t)$ are not independent, the stochastic discount $\exp\left(-\int_0^T r(s)\,\mathrm{d}s\right)$ is not independent from the forward price $F(T)$ and we cannot separate the two terms in the mathematical expectation.

6. We remind that the price of the zero-coupon bond is given by:

$$B(t,T) = \mathbb{E}^{\mathbb{Q}}\left[\left. e^{-\int_t^T r(s)\,\mathrm{d}s}\right| \mathcal{F}_t\right]$$

The instantaneous forward rate $f(t,T)$ is defined as follows:

$$f(t,T) = -\frac{\partial \ln B(t,T)}{\partial T}$$

We consider that the numéraire is the bond price $B(t,T)$ and we note $\mathbb{Q}^\star$ the associated forward probability measure.

(a) We have:

$$\begin{aligned}\frac{\partial B(t,T)}{\partial T} &= \frac{\partial}{\partial T}\mathbb{E}^{\mathbb{Q}}\left[\left. e^{-\int_t^T r(s)\,\mathrm{d}s}\right| \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\left. \frac{\partial\, e^{-\int_t^T r(s)\,\mathrm{d}s}}{\partial T}\right| \mathcal{F}_t\right] \\ &= -\mathbb{E}^{\mathbb{Q}}\left[\left. e^{-\int_t^T r(s)\,\mathrm{d}s} r(T)\right| \mathcal{F}_t\right] \\ &= -\mathbb{E}^{\mathbb{Q}}\left[\left. \frac{M(t)}{M(T)} r(T)\right| \mathcal{F}_t\right]\end{aligned}$$

where $M(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right)$ is the spot numéraire. We consider the change of numéraire $(M(t) \longrightarrow N(t) = B(t,T))$ and we obtain:

$$\begin{aligned}\frac{\partial\, B(t,T)}{\partial\, T} &= -\mathbb{E}^{\mathbb{Q}^\star}\left[\left.\frac{N(t)}{N(T)} r(T)\right| \mathcal{F}_t\right] \\ &= -N(t)\,\mathbb{E}^{\mathbb{Q}^\star}\left[\left. r(T)\right| \mathcal{F}_t\right]\end{aligned}$$

because $N(T) = B(T,T) = 1$. Since $r(T) = f(T,T)$, we deduce that:

$$\frac{\partial\, B(t,T)}{\partial\, T} = -B(t,T)\,\mathbb{E}^{\mathbb{Q}^\star}\left[\left. f(T,T)\right| \mathcal{F}_t\right]$$

(b) We have:

$$\begin{aligned}f(t,T) &= -\frac{\partial \ln B(t,T)}{\partial\, T} \\ &= -\frac{1}{B(t,T)}\frac{\partial\, B(t,T)}{\partial\, T} \\ &= \mathbb{E}^{\mathbb{Q}^\star}\left[\left. f(T,T)\right| \mathcal{F}_t\right]\end{aligned}$$

$f(t,T)$ is then an F_t-martingale under the forward probability measure $\mathbb{Q}^\star$.

(c) We know that:

$$\begin{aligned}\mathcal{C}(0) &= \mathbb{E}^{\mathbb{Q}}\left[\left. e^{-\int_0^T r(s)\,\mathrm{d}s}\max\left(f(T,T) - K, 0\right)\right| \mathcal{F}_0\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\left.\frac{M(t)}{M(T)}\max\left(f(T,T) - K, 0\right)\right| \mathcal{F}_0\right]\end{aligned}$$

Using the change of numéraire $N(t) = B(t,T)$, we obtain:

$$\begin{aligned}\mathcal{C}(0) &= \mathbb{E}^{\mathbb{Q}^\star}\left[\left.\frac{N(t)}{N(T)}\max\left(f(T,T) - K, 0\right)\right| \mathcal{F}_0\right] \\ &= B(t,T)\,\mathbb{E}^{\mathbb{Q}^\star}\left[\left.\max\left(f(T,T) - K, 0\right)\right| \mathcal{F}_0\right]\end{aligned}$$

Using the Black model, we deduce that the price of the option is[1]:

$$\mathcal{C}(0) = B(0,T)\cdot\left(r_0\Phi(d_1) - K\Phi(d_2)\right)$$

where d_1 and d_2 are the two values defined previously.

9.4.4 Change of numéraire and the Girsanov theorem

Part one

1. Using Itô's lemma, we obtain:

$$\mathrm{d}\left(X(t)\,Y(t)\right) = X(t)\,\mathrm{d}Y(t) + Y(t)\,\mathrm{d}X(t) + \left\langle \mathrm{d}X(t), \mathrm{d}Y(t)\right\rangle$$

and:

$$\mathrm{d}\left(\frac{1}{Y(t)}\right) = -\frac{\mathrm{d}Y(t)}{Y^2(t)} + \frac{\left\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\right\rangle}{Y^3(t)}$$

[1] We have $f(0,0) = r_0$.

2. Let $Z(t)$ be the ratio of $X(t)$ and $Y(t)$:

$$Z(t) = \frac{X(t)}{Y(t)}$$

We deduce that:

$$\begin{aligned}
\mathrm{d}Z(t) &= X(t)\,\mathrm{d}\left(\frac{1}{Y(t)}\right) + \frac{1}{Y(t)}\mathrm{d}X(t) + \left\langle \mathrm{d}X(t), \mathrm{d}\left(\frac{1}{Y(t)}\right)\right\rangle \\
&= \frac{\mathrm{d}X(t)}{Y(t)} + X(t)\left(-\frac{\mathrm{d}Y(t)}{Y^2(t)} + \frac{\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\rangle}{Y^3(t)}\right) + \\
&\quad \left\langle \mathrm{d}X(t), -\frac{\mathrm{d}Y(t)}{Y^2(t)} + \frac{\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\rangle}{Y^3(t)}\right\rangle \\
&= \frac{\mathrm{d}X(t)}{Y(t)} - \frac{X(t)}{Y^2(t)}\mathrm{d}Y(t) + X(t)\frac{\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\rangle}{Y^3(t)} - \\
&\quad \frac{\langle \mathrm{d}X(t), \mathrm{d}Y(t)\rangle}{Y^2(t)}
\end{aligned}$$

and:

$$\frac{\mathrm{d}Z(t)}{Z(t)} = \frac{\mathrm{d}X(t)}{X(t)} - \frac{\mathrm{d}Y(t)}{Y(t)} + \frac{\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\rangle}{Y^2(t)} - \frac{\langle \mathrm{d}X(t), \mathrm{d}Y(t)\rangle}{X(t)\,Y(t)} \tag{9.1}$$

Part two

1. The Girsanov theorem states that the change of probability only affects the drift and not the diffusion.

2. We have:

$$\begin{aligned}
\mathrm{d}S(t) &= \mu_S^\star(t)\,S(t)\,\mathrm{d}t + \sigma_S(t)\,S(t)\,\mathrm{d}W^{\mathbb{Q}^\star}(t) \\
&= \mu_S^\star(t)\,S(t)\,\mathrm{d}t + \sigma_S(t)\,S(t)\left(\mathrm{d}W^{\mathbb{Q}}(t) - g(t)\,\mathrm{d}t\right) \\
&= \left(\mu_S^\star(t) - g(t)\,\sigma_S(t)\right)S(t)\,\mathrm{d}t + \sigma_S(t)\,S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)
\end{aligned}$$

If follows that:

$$\mu_S^\star(t) - g(t)\,\sigma_S(t) = \mu_S(t)$$

or:

$$g(t) = \frac{\mu_S^\star(t) - \mu_S(t)}{\sigma_S(t)}$$

Using Girsanov's theorem, we deduce that the Radon-Nikodym derivative is equal to:

$$\begin{aligned}
Z(t) &= \frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} \\
&= \exp\left(\int_0^t g(s)\,\mathrm{d}W^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s\right)
\end{aligned}$$

We know that $Z(t)$ is an $\mathcal{F}_t$-martingale and we have:

$$\frac{\mathrm{d}Z(t)}{Z(t)} = g(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

3. We have:

$$Z(t) = \frac{M(0)}{N(0)} \frac{N(t)}{M(t)}$$

Using Equation (9.1), we have:

$$\begin{aligned}
\frac{\mathrm{d}Z(t)}{Z(t)} &= \mu_N(t)\ \mathrm{d}t + \sigma_N(t)\ \mathrm{d}W^{\mathbb{Q}}(t) - \\
&\quad \left(\mu_M(t)\ \mathrm{d}t + \sigma_M(t)\ \mathrm{d}W^{\mathbb{Q}}(t)\right) + \\
&\quad \sigma_M^2(t)\ \mathrm{d}t - \sigma_N(t)\,\sigma_M(t)\ \mathrm{d}t \\
&= (\mu_N(t) - \mu_M(t))\ \mathrm{d}t - \sigma_M(t)\,(\sigma_N(t) - \sigma_M(t))\ \mathrm{d}t + \\
&\quad (\sigma_N(t) - \sigma_M(t))\ \mathrm{d}W^{\mathbb{Q}}(t)
\end{aligned}$$

We deduce that:

$$g(t) = \sigma_N(t) - \sigma_M(t)$$

and:

$$\mu_N(t) = \mu_M(t) + \sigma_M(t)\,(\sigma_N(t) - \sigma_M(t))$$

4. Since $g(t) = \sigma_N(t) - \sigma_M(t)$, it follows that:

$$\begin{aligned}
\mu_S^\star(t) &= \mu_S(t) + g(t)\,\sigma_S(t) \\
&= \mu_S(t) + \sigma_S(t)\,(\sigma_N(t) - \sigma_M(t))
\end{aligned} \tag{9.2}$$

5. We have:

$$\left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}N(t)}{N(t)} \right\rangle = \sigma_S(t)\,\sigma_N(t)\ \mathrm{d}t$$

and:

$$\left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}M(t)}{M(t)} \right\rangle = \sigma_S(t)\,\sigma_M(t)\ \mathrm{d}t$$

We conclude that Equation (9.2) is equivalent to:

$$\mu_S^\star(t)\ \mathrm{d}t - \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}N(t)}{N(t)} \right\rangle = \mu_S(t)\ \mathrm{d}t - \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}M(t)}{M(t)} \right\rangle$$

We also notice that:

$$\begin{aligned}
\left\langle \frac{\mathrm{d}S(t)}{S(t)}, \mathrm{d}\ln\frac{N(t)}{M(t)} \right\rangle &= \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}Z(t)}{Z(t)} \right\rangle \\
&= \sigma_S(t)\,(\sigma_N(t) - \sigma_M(t))\ \mathrm{d}t
\end{aligned}$$

and:

$$\mu_S^\star(t)\ \mathrm{d}t = \mu_S(t)\ \mathrm{d}t + \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \mathrm{d}\ln\frac{N(t)}{M(t)} \right\rangle$$

Part three

1. Using Equation (9.1), we obtain:

$$\begin{aligned}
\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} &= \frac{\mathrm{d}S(t)}{S(t)} - \frac{\mathrm{d}N(t)}{N(t)} + \frac{\langle \mathrm{d}N(t), \mathrm{d}N(t)\rangle}{N^2(t)} - \frac{\langle \mathrm{d}S(t), \mathrm{d}N(t)\rangle}{S(t)\,N(t)} \\
&= r(t)\ \mathrm{d}t + \sigma_S(t)\ \mathrm{d}W_S^{\mathbb{Q}}(t) - \left(r(t)\ \mathrm{d}t + \sigma_N(t)\ \mathrm{d}W_N^{\mathbb{Q}}(t)\right) + \\
&\quad \sigma_N^2(t)\ \mathrm{d}t - \rho\sigma_S(t)\,\sigma_N(t)\ \mathrm{d}t \\
&= \left(\sigma_N^2(t) - \rho\sigma_S(t)\,\sigma_N(t)\right)\ \mathrm{d}t + \sigma_S(t)\ \mathrm{d}W_S^{\mathbb{Q}}(t) - \\
&\quad \sigma_N(t)\ \mathrm{d}W_N^{\mathbb{Q}}(t)
\end{aligned}$$

(a) We have:

$$\mathrm{d}N(t) = r(t)\,N(t)\,\mathrm{d}t$$

and:

$$N(t) = e^{\int_0^t r(s)\,\mathrm{d}s}$$

We deduce that the discounted asset price is:

$$\begin{aligned}\tilde{S}(t) &= \frac{S(t)}{N(t)} \\ &= e^{-\int_0^t r(s)\,\mathrm{d}s} S(t)\end{aligned}$$

Since $\sigma_N(t)$ is equal to zero, it follows that:

$$\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} = \sigma_S(t)\,\mathrm{d}W_S^{\mathbb{Q}}(t)$$

We conclude that $\tilde{S}(t)$ is an $\mathcal{F}_t$-martingale under the risk-neutral probability measure $\mathbb{Q}$.

(b) We note:

$$W^{\mathbb{Q}}(t) = W_S^{\mathbb{Q}}(t) = W_N^{\mathbb{Q}}(t)$$

The Girsanov theorem gives[2]:

$$\mathrm{d}W^{\mathbb{Q}^\star}(t) = \mathrm{d}W^{\mathbb{Q}}(t) - \sigma_N(t)\,\mathrm{d}t$$

and:

$$\mathrm{d}W^{\mathbb{Q}}(t) = \mathrm{d}W^{\mathbb{Q}^\star}(t) + \sigma_N(t)\,\mathrm{d}t$$

We deduce that:

$$\begin{aligned}\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} &= \left(\sigma_N^2(t) - \sigma_S(t)\,\sigma_N(t)\right)\mathrm{d}t + \left(\sigma_S(t) - \sigma_N(t)\right)\mathrm{d}W^{\mathbb{Q}}(t) \\ &= \left(\sigma_N^2(t) - \sigma_S(t)\,\sigma_N(t)\right)\mathrm{d}t + \left(\sigma_S(t) - \sigma_N(t)\right)\sigma_N(t)\,\mathrm{d}t \\ &\quad \left(\sigma_S(t) - \sigma_N(t)\right)\mathrm{d}W^{\mathbb{Q}^\star}(t) \\ &= \tilde{\sigma}(t)\,\mathrm{d}W^{\mathbb{Q}^\star}(t)\end{aligned}$$

where:

$$\tilde{\sigma}(t) = \sigma_S(t) - \sigma_N(t)$$

(c) Let us introduce the Brownian motion $\tilde{W}^{\mathbb{Q}}(t)$ such that:

$$\tilde{\sigma}(t)\,\mathrm{d}\tilde{W}^{\mathbb{Q}}(t) = \sigma_S(t)\,\mathrm{d}W_S^{\mathbb{Q}}(t) - \sigma_N(t)\,\mathrm{d}W_N^{\mathbb{Q}}(t)$$

We have:

$$\tilde{\sigma}^2(t) = \sigma_S^2(t) - 2\rho\sigma_S(t)\,\sigma_N(t) + \sigma_N^2(t)$$

We conclude that the risk-neutral dynamics of $\tilde{S}(t)$ is given by:

$$\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} = \left(\sigma_N^2(t) - \rho\sigma_S(t)\,\sigma_N(t)\right)\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}\tilde{W}^{\mathbb{Q}}(t)$$

[2] In Part two, we have shown that $\mathrm{d}W^{\mathbb{Q}^\star}(t) = \mathrm{d}W^{\mathbb{Q}}(t) - g(t)\,\mathrm{d}t$ where $g(t) = \sigma_N(t) - \sigma_M(t)$. Here, we assume that $M(t) = 1$, implying that $\sigma_M(t) = 0$.

We now consider the following decomposition:

$$W_S^{\mathbb{Q}}(t) = \rho W_N^{\mathbb{Q}}(t) + \sqrt{1-\rho^2} W_S^{\perp}(t)$$

where $W_S^{\perp}(t) \perp W_S^{\mathbb{Q}}(t)$. We deduce that:

$$\begin{aligned} \frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} &= \left(\sigma_N^2(t) - \rho\sigma_S(t)\sigma_N(t)\right) \mathrm{d}t + \\ & \quad (\rho\sigma_S(t) - \sigma_N(t)) \,\mathrm{d}W_N^{\mathbb{Q}}(t) + \sigma_S(t)\sqrt{1-\rho^2}\,\mathrm{d}W_S^{\perp}(t) \end{aligned}$$

Since we have:

$$\mathrm{d}W^{\mathbb{Q}^\star}(t) = \mathrm{d}W_N^{\mathbb{Q}}(t) - \sigma_N(t)\,\mathrm{d}t$$

we obtain:

$$\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} = (\rho\sigma_S(t) - \sigma_N(t))\,\mathrm{d}W^{\mathbb{Q}^\star}(t) + \sigma_S(t)\sqrt{1-\rho^2}\,\mathrm{d}W_S^{\perp}(t)$$

We notice that:

$$\tilde{\sigma}(t)\,\mathrm{d}\tilde{W}^{\mathbb{Q}^\star}(t) = (\rho\sigma_S(t) - \sigma_N(t))\,\mathrm{d}W^{\mathbb{Q}^\star}(t) + \sigma_S(t)\sqrt{1-\rho^2}\,\mathrm{d}W_S^{\perp}(t)$$

We deduce that:

$$\frac{\mathrm{d}\tilde{S}(t)}{\tilde{S}(t)} = \tilde{\sigma}(t)\,\mathrm{d}\tilde{W}^{\mathbb{Q}^\star}(t)$$

9.4.5 The HJM model and the forward probability measure

1. Since we have:

$$N(t) = B(t, T_2) = e^{-\int_t^{T_2} f(t,u)\,\mathrm{d}u}$$

we deduce that the Radon-Nikodym derivative is given by:

$$\begin{aligned} \frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} &= \frac{M(0)\,N(T_2)}{N(T_2)\,N(0)} \\ &= e^{-\int_0^{T_2} r(t)\,\mathrm{d}t} \frac{N(T_2)}{N(0)} \\ &= e^{-\int_0^{T_2} (r(t) - f(0,t))\,\mathrm{d}t} \end{aligned}$$

2. We have seen that the dynamics of the instantaneous spot rate is:

$$r(t) = r(0) + \int_0^t \left(\sigma(s,t)\int_s^t \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}s + \int_0^t \sigma(s,t)\,\mathrm{d}W^{\mathbb{Q}}(s)$$

It follows that:

$$r(t) - f(0,t) = \int_0^t \left(\sigma(s,t)\int_s^t \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}s + \int_0^t \sigma(s,t)\,\mathrm{d}W^{\mathbb{Q}}(s)$$

Using Fubini's theorem, we have:

$$\begin{aligned}\int_0^{T_2} (r(t) - f(0,t))\,\mathrm{d}t &= \int_0^{T_2} \left(\int_0^t \left(\sigma(s,t)\int_s^t \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}s\right)\mathrm{d}t + \\ &\quad \int_0^{T_2}\left(\int_0^t \sigma(s,t)\,\mathrm{d}W^{\mathbb{Q}}(s)\right)\mathrm{d}t \\ &= \int_0^{T_2}\left(\int_s^{T_2}\left(\sigma(s,t)\int_v^t \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}t\right)\mathrm{d}s + \\ &\quad \int_0^{T_2}\left(\int_s^{T_2}\sigma(s,t)\,\mathrm{d}t\right)\mathrm{d}W^{\mathbb{Q}}(s)\end{aligned}$$

We remind that:

$$\begin{aligned}a(t,T_2) &= -\int_t^{T_2}\alpha(t,v)\,\mathrm{d}v \\ &= -\int_t^{T_2}\left(\sigma(t,v)\int_t^v \sigma(t,u)\,\mathrm{d}u\right)\mathrm{d}v\end{aligned}$$

and:

$$b(t,T_2) = -\int_t^{T_2}\sigma(t,v)\,\mathrm{d}v$$

We deduce that:

$$\int_0^{T_2}(r(t) - f(0,t))\,\mathrm{d}t = -\int_0^{T_2} a(t,T_2)\,\mathrm{d}t - \int_0^{T_2} b(t,T_2)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

Finally, we conclude that:

$$\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} = e^{\int_0^{T_2} a(t,T_2)\,\mathrm{d}t + \int_0^{T_2} b(t,T_2)\,\mathrm{d}W^{\mathbb{Q}}(t)}$$

3. Since the no-arbitrage condition in the HJM model is:

$$a(t,T_2) + \frac{1}{2}b^2(t,T_2) = 0$$

we obtain:

$$\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} = e^{\int_0^{T_2} g(t)\,\mathrm{d}W^{\mathbb{Q}}(t) - \frac{1}{2}\int_0^{T_2} g^2(t)\,\mathrm{d}t}$$

where:

$$g(t) = b(t,T_2)$$

The Girsanov theorem states that:

$$W^{\mathbb{Q}^\star(T_2)}(t) = W^{\mathbb{Q}}(t) - \int_0^t g(s)\,\mathrm{d}s$$

is a Brownian motion under the forward probability measure $\mathbb{Q}^\star(T_2)$. We deduce that:

$$W^{\mathbb{Q}^\star(T_2)}(t) = W^{\mathbb{Q}}(t) - \int_0^t b(s,T_2)\,\mathrm{d}s$$

4. We have:

$$\mathrm{d}W^{\mathbb{Q}^\star(T_2)}(t) = \mathrm{d}W^{\mathbb{Q}}(t) - b(t,T_2)\,\mathrm{d}t$$

It follows that:

$$\begin{aligned}
\mathrm{d}f(t,T_1) &= \alpha(t,T_1)\,\mathrm{d}t + \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}}(t) \\
&= \alpha(t,T_1)\,\mathrm{d}t + \sigma(t,T_1)\left(\mathrm{d}W^{\mathbb{Q}^\star}(t) + b(t,T_2)\,\mathrm{d}t\right) \\
&= (\alpha(t,T_1) + \sigma(t,T_1)\,b(t,T_2))\,\mathrm{d}t + \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}^\star(T_2)}(t)
\end{aligned}$$

Since we have:

$$\begin{aligned}
\alpha(t,T_1) + \sigma(t,T_1)\,b(t,T_2) &= \sigma(t,T_1)\int_t^{T_1}\sigma(t,u)\,\mathrm{d}u - \\
&\qquad \sigma(t,T_1)\int_t^{T_2}\sigma(t,u)\,\mathrm{d}u
\end{aligned}$$

We conclude that:

$$\mathrm{d}f(t,T_1) = -\left(\sigma(t,T_1)\int_{T_1}^{T_2}\sigma(t,u)\,\mathrm{d}u\right)\mathrm{d}t + \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}^\star(T_2)}(t)$$

5. When T_2 is equal to T_1, we have:

$$\int_{T_1}^{T_1}\sigma(t,u)\,\mathrm{d}u = 0$$

and:

$$\mathrm{d}f(t,T_1) = \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(t)$$

We deduce that $f(t,T_1)$ is a martingale under the forward probability measure $\mathbb{Q}^\star(T_1)$.

6. (a) Let $s \le t$. We have:

$$\frac{B(t,T)}{B(s,T)} = e^{\int_s^t\left(r(u)-\frac{1}{2}b^2(u,T)\right)\mathrm{d}u + b(u,T)\,\mathrm{d}W^{\mathbb{Q}}(u)}$$

and:

$$\begin{aligned}
\frac{B(t,T_2)}{B(t,T_1)} &= \frac{B(s,T_2)\,e^{\int_s^t\left(r(u)-\frac{1}{2}b^2(u,T_2)\right)\mathrm{d}u + b(u,T_2)\,\mathrm{d}W^{\mathbb{Q}}(u)}}{B(s,T_1)\,e^{\int_s^t\left(r(u)-\frac{1}{2}b^2(u,T_1)\right)\mathrm{d}u + b(u,T_1)\,\mathrm{d}W^{\mathbb{Q}}(u)}} \\
&= \frac{B(s,T_2)}{B(s,T_1)}e^{X(s,t)}
\end{aligned}$$

where:

$$\begin{aligned}
X(s,t) &= -\frac{1}{2}\int_s^t\left(b^2(u,T_2) - b^2(u,T_1)\right)\mathrm{d}u + \\
&\qquad \int_s^t (b(u,T_2) - b(u,T_1))\,\mathrm{d}W^{\mathbb{Q}}(u)
\end{aligned}$$

(b) We have:

$$\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(t) = \mathrm{d}W^{\mathbb{Q}}(t) - b(t,T_1)\,\mathrm{d}t$$

We deduce that:

$$\begin{aligned} X(s,t) &= -\frac{1}{2}\int_s^t \left(b^2(u,T_2) - b^2(u,T_1)\right)\mathrm{d}u + \\ &\quad \int_s^t \left(b(u,T_2) - b(u,T_1)\right)\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(u) + \\ &\quad \int_s^t \left(b(u,T_2) - b(u,T_1)\right) b(u,T_1)\,\mathrm{d}u \\ &= -\frac{1}{2}\int_s^t \left(b(u,T_2) - b(u,T_1)\right)^2 \mathrm{d}u + \\ &\quad \int_s^t \left(b(u,T_2) - b(u,T_1)\right)\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(u) \end{aligned}$$

We notice that $e^{X(s,t)}$ is an exponential martingale:

$$\mathbb{E}\left[e^{X(s,t)}\ \mathcal{F}_s\right] = e^{X(s,s)} = 1$$

We conclude that:

$$\begin{aligned} \mathbb{E}\left[\frac{B(t,T_2)}{B(t,T_1)}\ \mathcal{F}_s\right] &= \mathbb{E}\left[\frac{B(s,T_2)}{B(s,T_1)} e^{X(s,t)}\ \mathcal{F}_s\right] \\ &= \frac{B(s,T_2)}{B(s,T_1)} \end{aligned}$$

9.4.6 Equivalent martingale measure in the Libor market model

1. Since we have:

$$\begin{aligned} \frac{B(t,T_j)}{B(t,T_{j+1})} &= 1 + (T_{j+1} - T_j)\,L(t,T_j,T_{j+1}) \\ &= 1 + \delta_j L_j(t) \end{aligned}$$

we obtain:

$$\frac{B(t,T_{j+1})}{B(t,T_j)} = \frac{1}{1 + \delta_j L_j(t)}$$

It follows that:

$$\begin{aligned} \frac{B(t,T_{k+1})}{B(t,T_{i+1})} &= \frac{B(t,T_{k+1})}{B(t,T_k)} \times \frac{B(t,T_k)}{B(t,T_{k-1})} \times \cdots \times \frac{B(t,T_{i+2})}{B(t,T_{i+1})} \\ &= \prod_{j=i+1}^{k} \frac{B(t,T_{j+1})}{B(t,T_j)} \\ &= \prod_{j=i+1}^{k} \frac{1}{1 + \delta_j L_j(t)} \end{aligned}$$

2. We remind that:

$$M(t) = B(t,T_{i+1})$$

and:

$$N(t) = B(t, T_{k+1})$$

We have:

$$\begin{aligned} Z(t) &= \frac{\mathrm{d}\mathbb{Q}^\star(T_{k+1})}{\mathrm{d}\mathbb{Q}^\star(T_{i+1})} \\ &= \frac{N(t)/N(0)}{M(t)/M(0)} \end{aligned}$$

We deduce that:

$$\begin{aligned} Z(t) &= \frac{B(0, T_{i+1})\, B(t, T_{k+1})}{B(0, T_{k+1})\, B(t, T_{i+1})} \\ &= \frac{B(0, T_{i+1})}{B(0, T_{k+1})} \prod_{j=i+1}^{k} \frac{1}{1 + \delta_j L_j(t)} \end{aligned}$$

3. We have:

$$\begin{aligned} \mathrm{d} \ln Z(t) &= \mathrm{d}\left(\frac{B(0, T_{i+1})}{B(0, T_{k+1})} \prod_{j=i+1}^{k} \frac{1}{1 + \delta_j L_j(t)} \right) \\ &= -\sum_{j=i+1}^{k} \mathrm{d} \ln(1 + \delta_j L_j(t)) \\ &= -\sum_{j=i+1}^{k} \frac{\mathrm{d}(1 + \delta_j L_j(t))}{1 + \delta_j L_j(t)} \\ &= -\sum_{j=i+1}^{k} \frac{\delta_j \,\mathrm{d}L_j(t)}{1 + \delta_j L_j(t)} \\ &= -\sum_{j=i+1}^{k} \frac{\gamma_j(t)\, \delta_j L_j(t)}{1 + \delta_j L_j(t)} \,\mathrm{d}W_j^{\mathbb{Q}^\star(T_{j+1})}(t) \end{aligned}$$

4. We obtain:

$$\begin{aligned} \zeta &= \left\langle \frac{\mathrm{d}L_i(t)}{L_i(t)}, \mathrm{d} \ln Z(t) \right\rangle \\ &= \left\langle \gamma_i(t)\, \mathrm{d}W_i^{\mathbb{Q}^\star}(t), -\sum_{j=i+1}^{k} \frac{\gamma_j(t)\, \delta_j L_j(t)}{1 + \delta_j L_j(t)} \,\mathrm{d}W_j^{\mathbb{Q}^\star(T_{j+1})}(t) \right\rangle \\ &= -\gamma_i(t) \sum_{j=i+1}^{k} \frac{\gamma_j(t)\, \delta_j L_j(t)}{1 + \delta_j L_j(t)} \left\langle \mathrm{d}W_i^{\mathbb{Q}^\star(T_{i+1})}(t), \mathrm{d}W_j^{\mathbb{Q}^\star(T_{j+1})}(t) \right\rangle \\ &= -\gamma_i(t) \sum_{j=i+1}^{k} \frac{\gamma_j(t)\, \delta_j L_j(t)}{1 + \delta_j L_j(t)} \rho_{i,j} \,\mathrm{d}t \end{aligned}$$

5. Under the probability measure $\mathbb{Q}^\star(T_{i+1})$, we have:

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \mu_i(t)\, \mathrm{d}t + \gamma_i(t)\, \mathrm{d}W_i^{\mathbb{Q}^\star(T_{i+1})}(t)$$

where $\mu_i(t) = 0$. In Question 5, Part two, Exercise 9.4.4, we have shown that:

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \mu_{i,k}(t)\ \mathrm{d}t + \gamma_i(t)\ \mathrm{d}W_k^{\mathbb{Q}^\star(T_{k+1})}(t)$$

where:

$$\mu_{i,k}(t)\ \mathrm{d}t = \mu_i(t)\ \mathrm{d}t + \left\langle \frac{\mathrm{d}L_i(t)}{L_i(t)}, \mathrm{d}\ln Z(t) \right\rangle$$

We deduce that:

$$\begin{aligned}
\mu_{i,k}(t) &= \mu_i(t) - \gamma_i(t) \sum_{j=i+1}^{k} \rho_{i,j}\gamma_j(t) \frac{\delta_j L_j(t)}{1+\delta_j L_j(t)} \\
&= -\gamma_i(t) \sum_{j=i+1}^{k} \rho_{i,j}\gamma_j(t) \frac{(T_{j+1}-T_j) L(t,T_j,T_{j+1})}{1+(T_{j+1}-T_j) L(t,T_j,T_{j+1})}
\end{aligned}$$

6. If $T_{k+1} < T_{i+1}$, we have:

$$\begin{aligned}
\frac{B(t,T_{k+1})}{B(t,T_{i+1})} &= \frac{B(t,T_{k+1})}{B(t,T_{k+2})} \times \frac{B(t,T_{k+2})}{B(t,T_{k+3})} \times \cdots \times \frac{B(t,T_i)}{B(t,T_{i+1})} \\
&= \prod_{j=k+1}^{i} \frac{B(t,T_j)}{B(t,T_{j+1})} \\
&= \prod_{j=k+1}^{i} (1+\delta_j L_j(t))
\end{aligned}$$

We deduce that:

$$Z(t) = \frac{B(0,T_{i+1})}{B(0,T_{k+1})} \prod_{j=k+1}^{i} (1+\delta_j L_j(t))$$

It follows that:

$$\mathrm{d}\ln Z(t) = \sum_{j=k+1}^{i} \frac{\gamma_j(t)\,\delta_j L_j(t)}{1+\delta_j L_j(t)}\,\mathrm{d}W_j^{\mathbb{Q}^\star(T_{j+1})}(t)$$

and:

$$\zeta = \gamma_i(t) \sum_{j=k+1}^{i} \frac{\gamma_j(t)\,\delta_j L_j(t)}{1+\delta_j L_j(t)} \rho_{i,j}\,\mathrm{d}t$$

We conclude that:

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \gamma_i(t) \sum_{j=k+1}^{i} \frac{\rho_{i,j}\gamma_j(t)\,\delta_j L_j(t)}{1+\delta_j L_j(t)}\,\mathrm{d}t + \gamma_i(t)\ \mathrm{d}W_k^{\mathbb{Q}^\star(T_{k+1})}(t)$$

9.4.7 Displaced diffusion option pricing

1. We have:

$$\begin{aligned}
\mathrm{d}S(t) &= (\partial_t\alpha(t) + \partial_t\beta(t)\,X(t) + \beta(t)\,\mu(t,X(t)))\ \mathrm{d}t + \\
&\quad\ \beta(t)\,\sigma(t,X(t))\ \mathrm{d}W^{\mathbb{Q}}(t)
\end{aligned}$$

We deduce that:

$$\begin{aligned} bS(t) &= \partial_t \alpha(t) + \partial_t \beta(t) X(t) + \beta(t) \mu(t, X(t)) \\ &= \partial_t \alpha(t) + \partial_t \beta(t) \left(\frac{S(t) - \alpha(t)}{\beta(t)} \right) + \beta(t) \mu \left(t, \frac{S(t) - \alpha(t)}{\beta(t)} \right) \end{aligned}$$

2. We have:

$$\begin{aligned} bS(t) &= \partial_t \alpha(t) + \partial_t \beta(t) \left(\frac{S(t) - \alpha(t)}{\beta(t)} \right) + \mu(t) \beta(t) \left(\frac{S(t) - \alpha(t)}{\beta(t)} \right) \\ &= \partial_t \alpha(t) - \left(\frac{\partial_t \beta(t)}{\beta(t)} + \mu(t) \right) \alpha(t) + \left(\frac{\partial_t \beta(t)}{\beta(t)} + \mu(t) \right) S(t) \end{aligned}$$

meaning that:

$$\begin{cases} \partial_t \alpha(t) - \left(\beta(t)^{-1} \partial_t \beta(t) + \mu(t) \right) \alpha(t) = 0 \\ \beta(t)^{-1} \partial_t \beta(t) + \mu(t) = b \end{cases}$$

We deduce that:

$$\partial_t \alpha(t) - b\alpha(t) = 0$$

and:

$$\alpha(t) = \alpha_0 e^{bt}$$

We also have:

$$\frac{\partial_t \beta(t)}{\beta(t)} = b - \mu(t)$$

and:

$$\beta(t) = \beta_0 e^{\int_0^t (b - \mu(s)) \, \mathrm{d}s}$$

3. We deduce that:

$$\mathrm{d}S(t) = bS(t) \, \mathrm{d}t + \beta(t)^{1-\gamma} \sigma(t) (S(t) - \alpha(t))^{\gamma} \, \mathrm{d}W^{\mathbb{Q}}(t) \tag{9.3}$$

4. We have:

$$\mathrm{d}X(t) = \mu(t) X(t) \, \mathrm{d}t + \sigma(t) X(t) \, \mathrm{d}W^{\mathbb{Q}}(t)$$

and:

$$X(t) = X_0 e^{\int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s) \right) \mathrm{d}s + \int_0^t \sigma(s) \, \mathrm{d}W^{\mathbb{Q}}(s)}$$

By noticing that $S_0 = \alpha_0 + \beta_0 X_0$, it follows that:

$$\begin{aligned} S(t) &= \alpha(t) + \beta(t) X_0 e^{\int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s) \right) \mathrm{d}s + \int_0^t \sigma(s) \, \mathrm{d}W^{\mathbb{Q}}(s)} \\ &= \alpha_0 e^{bt} + (S_0 - \alpha_0) e^{\int_0^t \left(b - \frac{1}{2}\sigma^2(s) \right) \mathrm{d}s + \int_0^t \sigma(s) \, \mathrm{d}W^{\mathbb{Q}}(s)} \end{aligned}$$

5. The payoff of the European call option is:

$$\begin{aligned} f(S(T)) &= (S(T) - K)^+ \\ &= \left(\left(S(T) - \alpha_0 e^{bT} \right) - \left(K - \alpha_0 e^{bT} \right) \right)^+ \end{aligned}$$

It follows that the option price is equal to:

$$\boldsymbol{\mathcal{C}}(t_0, S_0) = C_{\mathrm{BS}} \left(S_0 - \alpha_0, K - \alpha_0 e^{bT}, \sqrt{\int_0^t \sigma^2(s) \, \mathrm{d}s}, T, b, r \right)$$

9.4.8 Dupire local volatility model

We assume that:

$$\mathrm{d}S(t) = bS(t)\,\mathrm{d}t + \sigma(t, S(t))\,S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

1. Let $\boldsymbol{\mathcal{C}}(T,K)$ be the price of the European call option, whose maturity is T and strike is K. We remind that:

$$\frac{1}{2}\sigma^2(T,K)\,K^2\partial_K^2\boldsymbol{\mathcal{C}}(T,K) - bK\partial_K\boldsymbol{\mathcal{C}}(T,K) - \partial_T\boldsymbol{\mathcal{C}}(T,K) + (b-r)\,\boldsymbol{\mathcal{C}}(T,K) = 0$$

We deduce that:

$$\sigma^2(T,K) = \frac{A'(T,K)}{B'(T,K)}$$

where:

$$A'(T,K) = 2bK\partial_K\boldsymbol{\mathcal{C}}(T,K) + 2\partial_T\boldsymbol{\mathcal{C}}(T,K) - 2(b-r)\,\boldsymbol{\mathcal{C}}(T,K)$$

and:

$$B'(T,K) = K^2\partial_K^2\boldsymbol{\mathcal{C}}(T,K)$$

2. We have:

$$\boldsymbol{\mathcal{C}}(T,K) = S_0e^{(b-r)T}\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

where:

$$d_1 = \frac{1}{\Sigma(T,K)\sqrt{T}}\left(\ln\left(\frac{S_0}{K}\right) + bT\right) + \frac{1}{2}\Sigma(T,K)\sqrt{T}$$

and $d_2 = d_1 - \Sigma(T,K)\sqrt{T}$. We note $C_{\mathrm{BS}}(T,K,\Sigma)$ the Black-Scholes formula. We have:

$$\begin{aligned}\partial_K\boldsymbol{\mathcal{C}}(T,K) &= \partial_K C_{\mathrm{BS}}(T,K,\Sigma(T,K)) + \\ &\quad \partial_K\Sigma(T,K)\,\partial_\Sigma C_{\mathrm{BS}}(T,K,\Sigma(T,K))\end{aligned}$$

and:

$$\begin{aligned}\partial_K^2\boldsymbol{\mathcal{C}}(T,K) &= \partial_K^2 C_{\mathrm{BS}}(T,K,\Sigma(T,K)) + \\ &\quad 2\partial_K\Sigma(T,K)\,\partial_{\Sigma,K}^2 C_{\mathrm{BS}}(T,K,\Sigma(T,K)) + \\ &\quad \partial_K^2\Sigma(T,K)\,\partial_\Sigma C_{\mathrm{BS}}(T,K,\Sigma(T,K)) + \\ &\quad (\partial_K\Sigma(T,K))^2\,\partial_\Sigma^2 C_{\mathrm{BS}}(T,K,\Sigma(T,K))\end{aligned}$$

The derivative of $\boldsymbol{\mathcal{C}}(T,K)$ with respect to the maturity T is equal to:

$$\begin{aligned}\partial_T\boldsymbol{\mathcal{C}}(T,K) &= \partial_T C_{\mathrm{BS}}(T,K,\Sigma(T,K)) + \\ &\quad \partial_T\Sigma(T,K)\,\partial_\Sigma C_{\mathrm{BS}}(T,K,\Sigma(T,K))\end{aligned}$$

The different Black-Scholes derivatives are[3]:

$$\begin{aligned}
\partial_K C_{\mathrm{BS}}(T,K,\Sigma) &= -e^{-rT}\Phi(d_2) \\
\partial_\Sigma C_{\mathrm{BS}}(T,K,\Sigma) &= S_0 e^{(b-r)T}\sqrt{T}\phi(d_1) = K e^{-rT}\sqrt{T}\phi(d_2) \\
\partial_K^2 C_{\mathrm{BS}}(T,K,\Sigma) &= e^{-rT}\frac{\phi(d_2)}{K\Sigma\sqrt{T}} \\
\partial_\Sigma^2 C_{\mathrm{BS}}(T,K,\Sigma) &= e^{-rT}\frac{K\sqrt{T}\phi(d_2)\, d_1 d_2}{\Sigma} \\
\partial_{\Sigma,K}^2 C_{\mathrm{BS}}(T,K,\Sigma) &= e^{-rT}\frac{d_1\phi(d_2)}{\Sigma} \\
\partial_T C_{\mathrm{BS}}(T,K,\Sigma) &= (b-r) S_0 e^{(b-r)T}\Phi(d_1) + \\
&\quad K e^{-rT}\left(r\Phi(d_2) + \frac{\Sigma\phi(d_2)}{2\sqrt{T}}\right)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
A'(T,K) &= -2bKe^{-rT}\Phi(d_2) + 2bK^2 e^{-rT}\sqrt{T}\phi(d_2)\,\partial_K\Sigma(T,K) + \\
&\quad 2(b-r) S_0 e^{(b-r)T}\Phi(d_1) + \\
&\quad 2Ke^{-rT}\left(r\Phi(d_2) + \frac{\Sigma(T,K)\phi(d_2)}{2\sqrt{T}}\right) + \\
&\quad 2Ke^{-rT}\sqrt{T}\phi(d_2)\,\partial_T\Sigma(T,K) - \\
&\quad 2(b-r)\left(S_0 e^{(b-r)T}\Phi(d_1) - Ke^{-rT}\Phi(d_2)\right) \\
&= 2bK^2 e^{-rT}\sqrt{T}\phi(d_2)\,\partial_K\Sigma(T,K) + \\
&\quad \frac{Ke^{-rT}\Sigma(T,K)\phi(d_2)}{\sqrt{T}} + 2Ke^{-rT}\sqrt{T}\phi(d_2)\,\partial_T\Sigma(T,K) \\
&= e^{-rT}\frac{K\phi(d_2)}{\Sigma(T,K)\sqrt{T}}A(T,K)
\end{aligned}$$

where:

$$\begin{aligned}
A(T,K) &= \Sigma^2(T,K) + 2bKT\Sigma(T,K)\,\partial_K\Sigma(T,K) + \\
&\quad 2T\Sigma(T,K)\,\partial_T\Sigma(T,K)
\end{aligned} \tag{9.4}$$

We also have:

$$\begin{aligned}
B'(T,K) &= e^{-rT}\frac{K\phi(d_2)}{\Sigma(T,K)\sqrt{T}} + 2e^{-rT}\frac{K^2 d_1\phi(d_2)}{\Sigma(T,K)}\partial_K\Sigma(T,K) + \\
&\quad e^{-rT}K^3\sqrt{T}\phi(d_2)\,\partial_K^2\Sigma(T,K) + \\
&\quad e^{-rT}\frac{K^3\sqrt{T}\phi(d_2)\, d_1 d_2}{\Sigma(T,K)}\left(\partial_K\Sigma(T,K)\right)^2 \\
&= e^{-rT}\frac{K\phi(d_2)}{\Sigma(T,K)\sqrt{T}}B(T,K)
\end{aligned}$$

[3]We use the fact that:

$$S_0\phi(d_1) = Ke^{-bT}\phi(d_2)$$

where:

$$\begin{aligned} B(T,K) &= 1 + 2K\sqrt{T}d_1\partial_K\Sigma(T,K) + K^2T\Sigma(T,K)\partial_K^2\Sigma(T,K) + \\ & K^2Td_1d_2(\partial_K\Sigma(T,K))^2 \end{aligned} \tag{9.5}$$

We conclude that:

$$\begin{aligned} \sigma^2(T,K) &= \frac{A'(T,K)}{B'(T,K)} \\ &= \frac{A(T,K)}{B(T,K)} \end{aligned}$$

where $A(T,K)$ and $B(T,K)$ are given by Equations (9.4) and (9.5).

3. We follow the proof given by van der Kamp (2009)[4]. Let $\tilde{f}$ denote the discounted payoff function:

$$\tilde{f}(T,S(T)) = e^{-r(T-t)}(S(T)-K)^+$$

Itô's lemma gives:

$$\begin{aligned} \mathrm{d}\tilde{f}(T,S) &= -re^{-r(T-t)}(S-K)^+\,\mathrm{d}T + bSe^{-r(T-t)}\mathbb{1}\{S>K\}\,\mathrm{d}T + \\ & \frac{1}{2}\sigma^2(T,S)S^2e^{-r(T-t)}\delta(S-K)\,\mathrm{d}T + \\ & \sigma(T,S)Se^{-r(T-t)}\mathbb{1}\{S>K\}\,\mathrm{d}W^{\mathbb{Q}}(t) \end{aligned}$$

where $\delta(x)$ is the Dirac delta function. We deduce that:

$$\begin{aligned} \partial_T\boldsymbol{\mathcal{C}}(T,K) &= \frac{\mathbb{E}\left[\left.\mathrm{d}\tilde{f}(T,S(T))\right|\mathcal{F}_t\right]}{\mathrm{d}T} \\ &= re^{-r(T-t)}K\mathbb{E}\left[\left.\mathbb{1}\{S(T)>K\}\right|\mathcal{F}_t\right] + \\ & (b-r)e^{-r(T-t)}\mathbb{E}\left[\left.S(T)\mathbb{1}\{S(T)>K\}\right|\mathcal{F}_t\right] + \\ & \frac{1}{2}e^{-r(T-t)}\mathbb{E}\left[\left.\sigma^2(T,S(T))S^2(T)\delta(S(T)-K)\right|\mathcal{F}_t\right] \end{aligned}$$

4. We have:

$$\begin{aligned} \boldsymbol{\mathcal{C}}(T,K) &= \mathbb{E}\left[\left.\tilde{f}(T,S(T))\right|\mathcal{F}_t\right] \\ &= e^{-r(T-t)}\mathbb{E}\left[\left.(S(T)-K)\mathbb{1}\{S(T)>K\}\right|\mathcal{F}_t\right] \end{aligned}$$

It follows that:

$$\begin{aligned} \partial_K\boldsymbol{\mathcal{C}}(T,K) &= -e^{-r(T-t)}\mathbb{E}\left[\left.\mathbb{1}\{S(T)>K\}\right|\mathcal{F}_t\right] - \\ & e^{-r(T-t)}\mathbb{E}\left[\left.(S(T)-K)\delta(S(T)-K)\right|\mathcal{F}_t\right] \\ &= -e^{-r(T-t)}\mathbb{E}\left[\left.\mathbb{1}\{S(T)>K\}\right|\mathcal{F}_t\right] \end{aligned}$$

and:

$$\partial_K^2\boldsymbol{\mathcal{C}}(T,K) = e^{-r(T-t)}\mathbb{E}\left[\left.\delta(S(T)-K)\right|\mathcal{F}_t\right]$$

We notice that:

$$\partial_T\boldsymbol{\mathcal{C}}(T,K) = \frac{\mathbb{E}\left[\left.\mathrm{d}\tilde{f}(T,S(T))\right|\mathcal{F}_t\right]}{\mathrm{d}T}$$

[4] van der Kamp, R. (2009), Local Volatility Modeling, *Master of Science Dissertation*, University of Twente.

Since we have:

$$\mathbb{E}\left[S(T)\,\mathbb{1}\{S(T)>K\}\right]=e^{r(T-t)}\boldsymbol{\mathcal{C}}(T,K)+K\mathbb{1}\{S(T)>K\}$$

we obtain:

$$\begin{aligned}\partial_T\boldsymbol{\mathcal{C}}(T,K) &= re^{-r(T-t)}K\mathbb{E}\left[\mathbb{1}\{S(T)>K\}\,\middle|\,\mathcal{F}_t\right]+\\ &\quad (b-r)\,\mathbb{E}\left[\boldsymbol{\mathcal{C}}(T,K)+e^{-r(T-t)}K\mathbb{1}\{S(T)>K\}\,\middle|\,\mathcal{F}_t\right]+\\ &\quad \frac{1}{2}e^{-r(T-t)}Q(T,K)\end{aligned}$$

We also have:

$$\begin{aligned}Q(T,K) &= \mathbb{E}\left[\sigma^2(T,S(T))\,S^2(T)\,\middle|\,S(T)=K\right]\cdot\mathbb{E}\left[\delta(S(T)-K)\,\middle|\,\mathcal{F}_t\right]\\ &= \mathbb{E}\left[\sigma^2(T,S(T))\,\middle|\,S(T)=K\right]K^2e^{r(T-t)}\partial^2_K\boldsymbol{\mathcal{C}}(T,K)\end{aligned}$$

We conclude that:

$$\begin{aligned}\partial_T\boldsymbol{\mathcal{C}}(T,K) &= rK\partial_K\boldsymbol{\mathcal{C}}(T,K)+(b-r)\,\boldsymbol{\mathcal{C}}(T,K)-\\ &\quad (b-r)\,K\partial_K\boldsymbol{\mathcal{C}}(T,K)+\\ &\quad \frac{1}{2}\mathbb{E}\left[\sigma^2(T,S(T))\,\middle|\,S(T)=K\right]K^2\partial^2_K C(T,K)\end{aligned}$$

and:

$$\frac{1}{2}\sigma^2(T,K)\,K^2\partial^2_K\boldsymbol{\mathcal{C}}(T,K)-bK\partial_K\boldsymbol{\mathcal{C}}(T,K)-\\ \partial_T\boldsymbol{\mathcal{C}}(T,K)+(b-r)\,\boldsymbol{\mathcal{C}}(T,K)=0$$

5. (a) Since we have:

$$x=\ln\frac{S_0}{K}+bT$$

we deduce that:

$$d_1=\frac{x}{\Sigma(T,K)\sqrt{T}}+\frac{1}{2}\Sigma(T,K)\sqrt{T}$$

and:

$$d_2=\frac{x}{\Sigma(T,K)\sqrt{T}}-\frac{1}{2}\Sigma(T,K)\sqrt{T}$$

We also notice that:

$$d_1d_2=\frac{x^2}{\Sigma^2(T,K)\,T}-\frac{1}{4}\Sigma^2(T,K)\,T$$

(b) The first derivatives of $x=\varphi(T,K)$ are equal to:

$$\partial_K\varphi(T,K)=-\frac{1}{K}$$

and:

$$\partial_T\varphi(T,K)=b$$

It follows that:

$$\begin{aligned}\partial_K \Sigma(T,K) &= \partial_K \tilde{\Sigma}(T,\varphi(T,K)) \\ &= \partial_x \tilde{\Sigma}(T,x)\,\partial_K \varphi(T,K) \\ &= -\frac{1}{K}\partial_x \tilde{\Sigma}(T,x)\end{aligned}$$

and

$$\begin{aligned}\partial_T \Sigma(T,K) &= \partial_T \tilde{\Sigma}(T,\varphi(T,K)) \\ &= \partial_T \tilde{\Sigma}(T,x) + b\partial_x \tilde{\Sigma}(T,x)\end{aligned}$$

We also have:

$$\partial_K^2 \Sigma(T,K) = \frac{1}{K^2}\partial_x \tilde{\Sigma}(T,x) + \frac{1}{K^2}\partial_x^2 \tilde{\Sigma}(T,x)$$

(c) We deduce that:

$$\tilde{\sigma}(T,x) = \sqrt{\frac{\tilde{A}(T,x)}{\tilde{B}(T,x)}}$$

where $\tilde{A}(T,x) = A(T,K)$ and $\tilde{B}(T,x) = B(T,K)$. Using Equations (9.4) and (9.5), we obtain:

$$\begin{aligned}\tilde{A}(T,x) &= \tilde{\Sigma}^2(T,x) - 2bT\tilde{\Sigma}(T,x)\,\partial_x\tilde{\Sigma}(T,x) + \\ & \quad 2T\tilde{\Sigma}(T,x)\left(\partial_T\tilde{\Sigma}(T,x) + b\partial_x\tilde{\Sigma}(T,x)\right) \\ &= \tilde{\Sigma}^2(T,x) + 2T\tilde{\Sigma}(T,x)\,\partial_T\tilde{\Sigma}(T,x)\end{aligned}$$

and:

$$\begin{aligned}\tilde{B}(T,x) &= 1 - 2\left(x\tilde{\Sigma}^{-1}(T,x) + \frac{1}{2}T\tilde{\Sigma}(T,x)\right)\partial_x\tilde{\Sigma}(T,x) + \\ & \quad T\tilde{\Sigma}(T,x)\left(\partial_x\tilde{\Sigma}(T,x) + \partial_x^2\tilde{\Sigma}(T,x)\right) + \\ & \quad \left(x^2\tilde{\Sigma}^{-2}(T,x) - \frac{1}{4}T^2\tilde{\Sigma}^2(T,x)\right)\left(\partial_x\tilde{\Sigma}(T,x)\right)^2 \\ &= \left(1 - x\tilde{\Sigma}^{-1}(T,x)\,\partial_x\tilde{\Sigma}(T,x)\right)^2 + \\ & \quad T\tilde{\Sigma}(T,x)\,\partial_x^2\tilde{\Sigma}(T,x) - \\ & \quad \frac{1}{4}\left(T\tilde{\Sigma}(T,x)\,\partial_x\tilde{\Sigma}(T,x)\right)^2\end{aligned}$$

(d) When T is equal to zero, we obtain:

$$\tilde{\sigma}^2(0,x) = \frac{\tilde{\Sigma}^2(0,x)}{\left(1 - x\tilde{\Sigma}^{-1}(0,x)\,\partial_x\tilde{\Sigma}(0,x)\right)^2}$$

and:

$$\tilde{\Sigma}(0,x) = \left(1 - \frac{x\partial_x\tilde{\Sigma}(0,x)}{\tilde{\Sigma}(0,x)}\right)\tilde{\sigma}(0,x)$$

The explicit solution of this equation is:

$$\tilde{\Sigma}(0,x) = \left(\int_0^1 \tilde{\sigma}^{-1}(0,xy)\,\mathrm{d}y\right)^{-1}$$

We deduce that:

$$\partial_x \tilde{\Sigma}(0,x) = \frac{\int_0^1 y\tilde{\sigma}^{-2}(0,xy)\,\partial_x\tilde{\sigma}(0,xy)\,\mathrm{d}y}{\left(\int_0^1 \tilde{\sigma}^{-1}(0,xy)\,\mathrm{d}y\right)^2}$$

It follows that:

$$\begin{aligned}\partial_x \tilde{\Sigma}(0,0) &= \frac{\int_0^1 y\,\mathrm{d}y}{\left(\int_0^1 \mathrm{d}y\right)^2}\partial_x\tilde{\sigma}(0,0) \\ &= \frac{1}{2}\partial_x\tilde{\sigma}(0,0)\end{aligned}$$

9.4.9 The stochastic normal model

1. We have[5]:

$$\begin{aligned}\Sigma_N(T,K) &= \Sigma_B(T,K)\sqrt{F_0K}\times \\ &\quad \frac{1+\frac{1}{24}\ln^2 F_0/K+\frac{1}{1920}\ln^4 F_0/K}{1+\frac{1}{24}\left(1-\frac{1}{120}\ln^2 F_0/K\right)\Sigma_B^2(T,K)\,T+\frac{1}{5760}\Sigma_B^4(T,K)\,T^2}\end{aligned}$$

2. We have[6]:

$$\begin{aligned}\Sigma_N(T,K) &= \alpha(F_0K)^{\beta/2}\left(\frac{z}{\chi(z)}\right)\times \\ &\quad \left(\frac{1+\frac{1}{24}\ln^2 F_0/K+\frac{1}{1920}\ln^4 F_0/K}{1+\frac{1}{24}(1-\beta)^2\ln^2 F_0/K+\frac{1}{1920}(1-\beta)^4\ln^4 F_0/K}\right)\times \\ &\quad \left(1+\left(\frac{-\beta(2-\beta)\alpha^2}{24(F_0K)^{1-\beta}}+\frac{\rho\alpha\nu\beta}{4(F_0K)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2\right)T\right)\end{aligned}$$

where:

$$z = \frac{\nu}{\alpha}(F_0K)^{(1-\beta)/2}\ln\frac{F_0}{K}$$

and

$$\chi(z) = \ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)$$

In the sequel, we introduce the notation $\varphi(z)=\sqrt{1-2\rho z+z^2}$.

3. When β is equal to 0, we obtain:

$$\Sigma_N(T,K) = \alpha\left(\frac{z}{\chi(z)}\right)\left(1+\frac{2-3\rho^2}{24}\nu^2 T\right)$$

where:

$$z = \frac{\nu}{\alpha}\sqrt{F_0K}\ln\frac{F_0}{K}$$

and:

$$\chi(z) = \ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)$$

[5] Hagan *et al.* (2002), Equation (A.64) on page 101.
[6] Hagan *et al.* (2002), Equation (A.69) on page 102.

4. Since we have:

$$\lim_{K\to F_0}\frac{z}{\chi\left(z\right)}=1$$

we deduce the expression of the ATM volatility:

$$\Sigma_N\left(T,F_0\right)=\alpha\left(1+\frac{2-3\rho^2}{24}\nu^2T\right)$$

5. We notice that z is a function of K. By introducing the notation $z=z\left(K\right)$, we have:

$$\begin{aligned}\partial_K\chi\left(z\left(K\right)\right) &= \frac{\frac{-2\rho\partial_K z(K)+2z(K)\partial_K z(K)}{2\sqrt{1-2\rho z(K)+z^2(K)}}+\partial_K z\left(K\right)}{\sqrt{1-2\rho z\left(K\right)+z^2\left(K\right)}+z\left(K\right)-\rho} \\ &= \frac{\partial_K z\left(K\right)}{\sqrt{1-2\rho z\left(K\right)+z^2\left(K\right)}}\end{aligned}$$

and:

$$\begin{aligned}\partial_K z\left(K\right) &= \frac{\nu}{\alpha}\sqrt{F_0}\partial_K\left(\sqrt{K}\left(\ln F_0-\ln K\right)\right) \\ &= \frac{\nu}{\alpha}\sqrt{F_0}\left(\frac{\ln F_0}{2\sqrt{K}}-\frac{\sqrt{K}}{K}-\frac{\ln K}{2\sqrt{K}}\right) \\ &= \frac{\nu}{\alpha}\sqrt{\frac{F_0}{K}}\left(\ln\sqrt{\frac{F_0}{K}}-1\right)\end{aligned}$$

It follows that:

$$\begin{aligned}\partial_K\left(\frac{z\left(K\right)}{\chi\left(z\left(K\right)\right)}\right) &= \frac{\chi\left(z\left(K\right)\right)\,\partial_K z\left(K\right)-z\left(K\right)\,\partial_K\chi\left(z\left(K\right)\right)}{\chi^2\left(z\left(K\right)\right)} \\ &= \left(\frac{\chi\left(z\left(K\right)\right)\varphi\left(z\left(K\right)\right)-z\left(K\right)}{\chi^2\left(z\left(K\right)\right)\varphi\left(z\left(K\right)\right)}\right)\partial_K z\left(K\right) \\ &= \left(\frac{1}{\chi\left(z\left(K\right)\right)}-\frac{z\left(K\right)}{\chi^2\left(z\right)\varphi\left(z\left(K\right)\right)}\right)\partial_K z\left(K\right)\end{aligned}$$

where:

$$\varphi\left(z\left(K\right)\right)=\sqrt{1-2\rho z\left(K\right)+z^2\left(K\right)}$$

We deduce that:

$$\begin{aligned}\partial_K\Sigma_N\left(T,K\right) &= \alpha\left(1+\frac{2-3\rho^2}{24}\nu^2T\right)\partial_K\left(\frac{z\left(K\right)}{\chi\left(z\left(K\right)\right)}\right) \\ &= \nu\left(1+\frac{2-3\rho^2}{24}\nu^2T\right)\sqrt{\frac{F_0}{K}}\left(\ln\sqrt{\frac{F_0}{K}}-1\right)\cdot \\ &\quad\left(\frac{1}{\chi\left(z\right)}-\frac{z}{\chi^2\left(z\right)\sqrt{1-2\rho z+z^2}}\right)\end{aligned} \tag{9.6}$$

6. We have:

$$\boldsymbol{\mathcal{C}}\left(T,K\right)=\left(F_0-K\right)\Phi\left(d\right)+\sigma_N\sqrt{T}\phi\left(d\right)$$

where:

$$d=\frac{F_0-K}{\sigma_N\sqrt{T}}$$

7. Using the results of Breeden and Litzenberger (1978), we have:

$$\begin{aligned}\mathbb{Q}(T,K) &= \Pr\{F(T)\leq K\}\\ &= 1+\frac{\partial\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial\,K}\end{aligned}$$

When σ_N is equal to the function $\Sigma_N(T,K)$, we deduce that:

$$\begin{aligned}\mathbb{Q}(T,K) &= 1-\Phi(d)+(F_0-K)\,\phi(d)\cdot\partial_K d+\\ &\quad \sqrt{T}\phi(d)\cdot\partial_K\Sigma_N(T,K)-\Sigma_N(T,K)\sqrt{T}d\phi(d)\cdot\partial_K d\\ &= 1-\Phi(d)+\phi(d)\sqrt{T}\cdot\partial_K\Sigma_N(T,K)\end{aligned}$$

where $\partial_K\Sigma_N(T,K)$ is given by Equation (9.6).

8. For the density function, we have:

$$\begin{aligned}q(T,K) &= -\phi(d)\cdot\partial_K d-d\phi(d)\sqrt{T}\cdot\partial_K d\cdot\partial_K\Sigma_N(T,K)+\\ &\quad \phi(d)\sqrt{T}\cdot\partial_K^2\Sigma_N(T,K)\end{aligned}$$

We notice that:

$$\begin{aligned}\partial_K d &= \frac{-\Sigma_N(T,K)\sqrt{T}-(F_0-K)\sqrt{T}\cdot\partial_K\Sigma_N(T,K)}{\Sigma_N^2(T,K)\,T}\\ &= -\frac{1+d\sqrt{T}\cdot\partial_K\Sigma_N(T,K)}{\Sigma_N(T,K)\sqrt{T}}\end{aligned}$$

It follows that:

$$\begin{aligned}q(T,K) &= \frac{\phi(d)}{\Sigma_N(T,K)\sqrt{T}}\left(1+d\sqrt{T}\cdot\partial_K\Sigma_N(T,K)\right)^2+\\ &\quad \phi(d)\sqrt{T}\cdot\partial_K^2\Sigma_N(T,K)\end{aligned}$$

To calculate the probability density function of $F(T)$, we need to calculate $\partial_K\Sigma_N(T,K)$ and $\partial_K^2\Sigma_N(T,K)$. If we use the approximation $z=\nu\alpha^{-1}(F_0-K)$, we have $\partial_K z=-\nu\alpha^{-1}$. We deduce that:

$$\partial_K\Sigma_N(T,K)=-\nu\left(1+\frac{2-3\rho^2}{24}\nu^2T\right)\left(\frac{1}{\chi(z)}-\frac{z}{\chi^2(z)\sqrt{1-2\rho z+z^2}}\right)$$

and:

$$\partial_K^2\Sigma_N(T,K)=\frac{\nu^2}{\alpha}\left(1+\frac{2-3\rho^2}{24}\nu^2T\right)D$$

where:

$$\begin{aligned}D &= \frac{2}{\chi^2(z)\left(1-2\rho z+z^2\right)}\left(\frac{z}{\chi(z)\sqrt{1-2\rho z+z^2}}-1\right)+\\ &\quad \frac{z(z-\rho)}{\chi^2(z)\left(1-2\rho z+z^2\right)^{3/2}}\end{aligned}$$

9. When $\beta=0$, the SABR model becomes:

$$\left\{\begin{array}{l}\mathrm{d}F(t)=\alpha(t)\,\mathrm{d}W_1^{\mathbb{Q}}(t)\\ \mathrm{d}\alpha(t)=\nu\alpha(t)\,\mathrm{d}W_2^{\mathbb{Q}}(t)\end{array}\right.$$

Since we have $\alpha(0) = \alpha$, we obtain:

$$\alpha(t) = \alpha e^{-\frac{1}{2}\nu^2 t + \nu W_2^{\mathbb{Q}}(t)}$$

and:

$$\mathrm{d}F(t) = \alpha e^{-\frac{1}{2}\nu^2 t + \nu W_2^{\mathbb{Q}}(t)}\,\mathrm{d}W_1^{\mathbb{Q}}(t)$$

It follows that:

$$F(t) = F_0 + \alpha \int_0^t e^{-\frac{1}{2}\nu^2 s + \nu W_2^{\mathbb{Q}}(s)}\,\mathrm{d}W_1^{\mathbb{Q}}(s)$$

Using the scaling property, we deduce that:

$$F(t) = F_0 + \frac{\alpha}{\nu}\int_0^{\nu^2 t} e^{-\frac{1}{2}s + W_2(s)}\,\mathrm{d}W_1(s)$$

where $W_1(t)$ and $W_2(t)$ have the same properties $W_1^{\mathbb{Q}}(t)$ and $W_2^{\mathbb{Q}}(t)$.

10. We note:

$$\begin{aligned}
X(t) &= \int_0^t e^{-\frac{1}{2}s + W_2(s)}\,\mathrm{d}W_1(s) \\
M^a(t) &= e^{-\frac{1}{2}at + aW_2(t)} \\
M(t) &= e^{-\frac{1}{2}t + W_2(t)}
\end{aligned}$$

We have:

$$\mathrm{d}X(t) = e^{-\frac{1}{2}t + W_2(t)}\,\mathrm{d}W_1(t) = M(t)\,\mathrm{d}W_1(t)$$

and:

$$\mathrm{d}\langle X(t)\rangle = e^{-t+2W_2(t)}\,\mathrm{d}t = M(t)^2\,\mathrm{d}t$$

Using Itô's lemma, we deduce that:

$$\begin{aligned}
\mathrm{d}X^n(t) &= nX^{n-1}(t)\,\mathrm{d}X(t) + \frac{n(n-1)}{2}X^{n-2}(t)\,\mathrm{d}\langle X(t)\rangle \\
&= nX^{n-1}(t)\,M(t)\,\mathrm{d}W_1(t) + \frac{n(n-1)}{2}X^{n-2}(t)\,M(t)^2\,\mathrm{d}t
\end{aligned}$$

Since we have:

$$\mathrm{d}M^a(t) = \frac{a(a-1)}{2}M^a(t)\,\mathrm{d}t + aM^a(t)\,\mathrm{d}W_2(t)$$

we obtain:

$$\begin{aligned}
\mathrm{d}\left(X^n(t)\,M^a(t)\right) &= X^n(t)\,\mathrm{d}M^a(t) + M^a(t)\,\mathrm{d}X^n(t) + \mathrm{d}\langle X^n(t), M^a(t)\rangle \\
&= \frac{a(a-1)}{2}X^n(t)\,M^a(t)\,\mathrm{d}t + aX^n(t)\,M^a(t)\,\mathrm{d}W_2(t) + \\
&\quad nX^{n-1}(t)\,M^a(t)\,M(t)\,\mathrm{d}W_1(t) + \\
&\quad \frac{n(n-1)}{2}X^{n-2}(t)\,M^a(t)\,M(t)^2\,\mathrm{d}t + \\
&\quad n\rho a X^{n-1}(t)\,M^a(t)\,M(t)\,\mathrm{d}t
\end{aligned}$$

It follows that:

$$\begin{aligned}\frac{\mathbb{E}\left[\mathrm{d}\left(X^n(t)\,M^a(t)\right)\right]}{\mathrm{d}t} &= \frac{a(a-1)}{2}\mathbb{E}\left[X^n(t)\,M^a(t)\right]+\\ &\quad \frac{n(n-1)}{2}\mathbb{E}\left[X^{n-2}(t)\,M^a(t)\,M(t)^2\right]+\\ &\quad n\rho a\mathbb{E}\left[X^{n-1}(t)\,M^a(t)\,M(t)\right]\end{aligned}$$

We notice that $M^a(t)\,M(t) = M^{a+1}(t)$ and $M^a(t)\,M(t)^2 = M^{a+2}(t)$. We conclude that:

$$\frac{\mathrm{d}\Psi^{n,a}(t)}{\mathrm{d}t} = \frac{a(a-1)}{2}\Psi^{n,a}(t) + \frac{n(n-1)}{2}\Psi^{n-2,a+2}(t) + n\rho a\Psi^{n-1,a+1}(t)$$

where:

$$\Psi^{n,a}(t) = \mathbb{E}\left[X^n(t)\,M^a(t)\right]$$

Therefore, the relationship between $\Psi^{n,a}(t)$ and the moments of $F(t)$ is:

$$\mathbb{E}\left[(F(t)-F_0)^n\right] = \left(\frac{\alpha}{\nu}\right)^n \Psi^{n,0}\left(\nu^2 t\right)$$

11. For $n=0$, we have:

$$\begin{aligned}\Psi^{0,a}(t) &= \mathbb{E}\left[X^0(t)\,M^a(t)\right]\\ &= \mathbb{E}\left[e^{-\frac{1}{2}at+aW_2(t)}\right]\\ &= e^{-\frac{1}{2}at}e^{\frac{1}{2}a^2t}\\ &= e^{\frac{1}{2}a(a-1)t}\end{aligned}$$

For $n=1$, we have:

$$\begin{aligned}\frac{\mathrm{d}\Psi^{1,a}(t)}{\mathrm{d}t} &= \frac{a(a-1)}{2}\Psi^{1,a}(t) + \rho a\Psi^{0,a+1}(t)\\ &= \frac{a(a-1)}{2}\Psi^{1,a}(t) + \rho a e^{\frac{1}{2}a(a+1)t}\end{aligned}$$

We deduce that[7]:

$$\begin{aligned}\Psi^{1,a}(t) &= e^{\frac{1}{2}a(a-1)t}\int_0^t e^{-\frac{1}{2}a(a-1)s}\rho a e^{\frac{1}{2}a(a+1)s}\,\mathrm{d}s\\ &= \rho e^{\frac{1}{2}a(a-1)t}\left[e^{as}\right]_0^t\\ &= \rho e^{\frac{1}{2}a(a-1)t}\left(e^{at}-1\right)\end{aligned}$$

For $n=2$, we solve the ODE:

$$\begin{aligned}\frac{\mathrm{d}\Psi^{2,a}(t)}{\mathrm{d}t} &= \frac{a(a-1)}{2}\Psi^{2,a}(t) + 2\rho a\Psi^{1,a+1}(t) + \Psi^{0,a+2}(t)\\ &= \frac{a(a-1)}{2}\Psi^{2,a}(t) + h(t)\end{aligned}$$

[7] We remind that the solution of the ODE:

$$\frac{\mathrm{d}f(t)}{\mathrm{d}t} = \alpha f(t) + \beta(t)$$

is equal to:

$$f(t) = e^{at}\int_0^t e^{-as}\beta(s)\,\mathrm{d}s$$

where:

$$h(t) = 2\rho^2 a e^{\frac{a(a+1)}{2}t}\left(e^{(a+1)t} - 1\right) + e^{\frac{(a+1)(a+2)}{2}t}$$

The solution is given by:

$$\begin{aligned}\Psi^{2,a}(t) &= e^{\frac{a(a-1)}{2}t}\int_0^t e^{-\frac{a(a-1)}{2}s} h(s)\,\mathrm{d}s \\ &= e^{\frac{a(a-1)}{2}t}\left(\left(\frac{2\rho^2 a + 1}{2a+1}\right)\left(e^{(2a+1)t} - 1\right) - 2\rho^2\left(e^{at} - 1\right)\right)\end{aligned}$$

For $n = 3$ and $a = 0$, the ODE becomes:

$$\begin{aligned}\frac{\mathrm{d}\Psi^{3,0}(t)}{\mathrm{d}t} &= 3\Psi^{1,2}(t) \\ &= 3\rho e^t\left(e^{2t} - 1\right)\end{aligned}$$

The solution is then:

$$\Psi^{3,0}(t) = \rho\left(e^{3t} - 3e^t + 2\right)$$

For $n = 4$ and $a = 0$, we obtain:

$$\begin{aligned}\frac{\mathrm{d}\Psi^{4,0}(t)}{\mathrm{d}t} &= 6\Psi^{2,2}(t) \\ &= 6e^t\left(\left(\frac{4\rho^2+1}{5}\right)\left(e^{5t} - 1\right) - 2\rho^2\left(e^{2t} - 1\right)\right)\end{aligned}$$

and:

$$\Psi^{4,0}(t) = \frac{4\rho^2+1}{5}e^{6t} - 4\rho^2 e^{3t} - \frac{6}{5}\left(4\rho^2+1\right)e^t + 12\rho^2 e^t - 4\rho^2 + 1$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[F(t) - F_0\right] &= \left(\frac{\alpha}{\nu}\right)^1 \Psi^{1,0}\left(\nu^2 t\right) \\ &= 0\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[\left(F(t) - F_0\right)^2\right] &= \left(\frac{\alpha}{\nu}\right)^2 \Psi^{2,0}\left(\nu^2 t\right) \\ &= \frac{\alpha^2}{\nu^2}\left(e^{\nu^2 t} - 1\right)\end{aligned}$$

For the third moment, we obtain:

$$\begin{aligned}\mathbb{E}\left[\left(F(t) - F_0\right)^3\right] &= \left(\frac{\alpha}{\nu}\right)^3 \Psi^{3,0}\left(\nu^2 t\right) \\ &= \rho\frac{\alpha^3}{\nu^3}\left(e^{3\nu^2 t} - 3e^{\nu^2 t} + 2\right)\end{aligned}$$

Finally, the fourth moment is equal to:

$$\begin{aligned}\mathbb{E}\left[\left(F(t) - F_0\right)^4\right] &= \left(\frac{\alpha}{\nu}\right)^4 \Psi^{4,0}\left(\nu^2 t\right) \\ &= \frac{\alpha^4}{\nu^4}\left(\frac{1}{5}\left(4\rho^2+1\right)e^{6\nu^2 t} - 4\rho^2 e^{3\nu^2 t}\right) + \\ &\quad \frac{\alpha^4}{\nu^4}\left(\frac{6}{5}\left(6\rho^2 - 1\right)e^{\nu^2 t} - 4\rho^2 + 1\right)\end{aligned}$$

12. We have:

$$\begin{aligned}\sigma^2 (F(t)) &= \frac{\alpha^2}{\nu^2}\left(e^{\nu^2 t} - 1\right) \\ &\simeq \frac{\alpha^2}{\nu^2}\left(1 + \nu^2 t + \frac{1}{2}\nu^4 t^2 + \frac{1}{6}\nu^6 t^3 - 1\right) \\ &= \alpha^2 \left(t + \frac{1}{2}\nu^2 t^2 + \frac{1}{6}\nu^4 t^3\right)\end{aligned}$$

Using the same approximation method, the skewness coefficient is:

$$\begin{aligned}\gamma_1 (F(t)) &= \frac{\rho\left(e^{3\nu^2 t} - 3e^{\nu^2 t} + 2\right)}{\left(e^{\nu^2 t} - 1\right)^{\frac{3}{2}}} \\ &\simeq 3\rho\nu\sqrt{t} + 4\rho\nu^3 t\sqrt{t}\end{aligned}$$

whereas the expression of the kurtosis is:

$$\begin{aligned}\gamma_2 (F(t)) &= \frac{\left(4\rho^2 + 1\right)e^{6\nu^2 t} - 20\rho^2\left(e^{3\nu^2 t} + 1\right) + \left(36\rho^2 - 6\right)e^{\nu^2 t} + 5}{5\left(e^{\nu^2 t} - 1\right)^2} \\ &\simeq \frac{3 + \left(7 + 11\rho^2\right)\nu^2 t}{\left(1 + \frac{1}{2}\nu^2 t + \frac{1}{6}\nu^4 t^2\right)^2}\end{aligned}$$

13. (a) Using the formula given in Question (1), we obtain the following equivalent normal volatility:

K	7%	10%	13%
$\Sigma_B (T, K)$	30%	20%	30%
$\Sigma_N (T, K)$	2.51389%	1.99667%	3.41753%

(b) The method of least squares gives $\alpha = 0.017573$, $\beta = 0$, $\nu = 1.448791$ and $\rho = 0.383867$. We verify that the fitted smile adjust perfectly the three observed volatilities.

(c) The cumulative distribution function is shown in Figure 9.1. We notice that the cdf is not an increasing function when the forward rate is close to zero. As a result, the density function takes negative value. We deduce that there is arbitrage opportunities.

(d) Using the formula calculated in Question (8), we obtain Figure 9.2. With the approximation $\sqrt{F_0 K}\ln\frac{F_0}{K} \simeq F_0 - K$, the probability density function becomes always positive.

(e) The skewness is equal to 10.43, whereas the kurtosis is equal to 1822.60. These values are very high, meaning that the stochastic normal model is far to be Gaussian. This result is surprising. However, we can show that the long-term probability distribution of $F(t)$ in the SABR model is non-degenerate contrary to Black and normal models. For instance, when ρ is equal to zero, we obtain:

$$F(\infty) \stackrel{\text{law}}{=} F_0 + \frac{\alpha}{\nu}\mathcal{N}(0,1)\sqrt{\frac{1}{2Z_{1/2}}}$$

where Z_k is a Gamma random variable with parameter k.

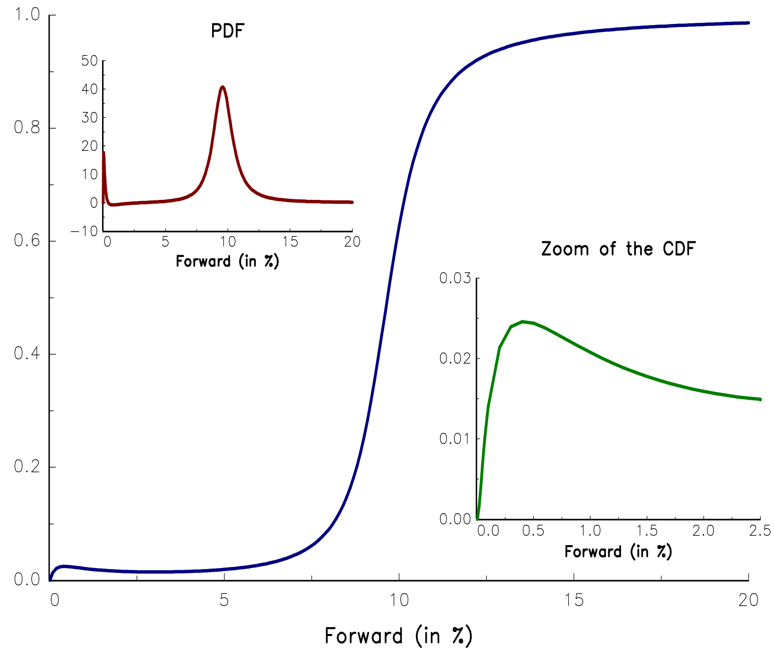

FIGURE 9.1: Cumulative distribution function of $F(1)$

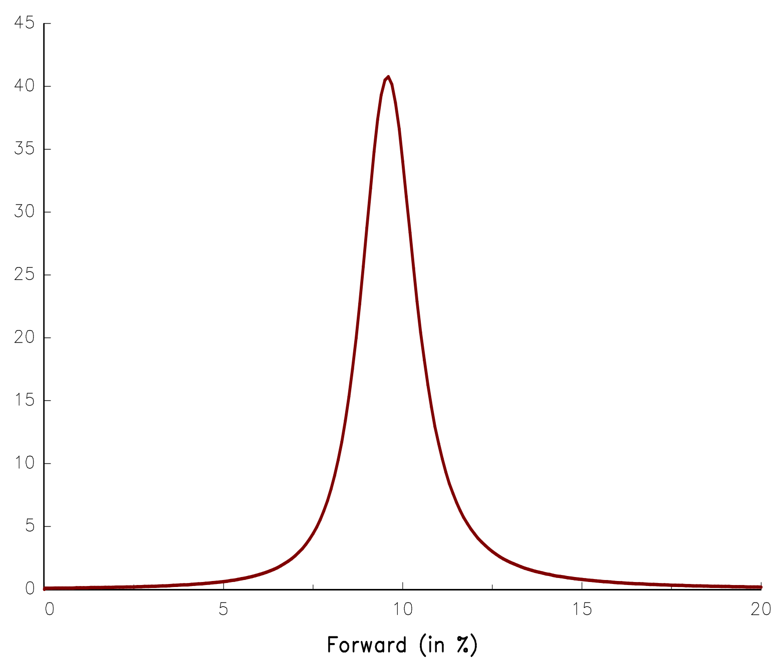

FIGURE 9.2: Probability density function of $F(1)$

Remark 1 *To find the distribution of $F(\infty)$, we use the following result*[8] *of Donati-Martin et al. (2001): Let $\xi(t) = -(ct + \sigma B(t) + N^{+}(t) - N^{-}(t))$ where $N^{+}(t)$ and $N^{-}(t)$ are two independent Poisson processes and $B(t)$ is a Brownian motion. Let $A(t) = \int_0^t \exp(\xi(s))\,\mathrm{d}s$, $X(t) = \exp(\xi(t))\int_0^t \exp(-\xi(s))\,\mathrm{d}s$ and T_α denotes an exponential variable of parameter α independent of $\xi(t)$. The law μ_α of $A(T_\alpha)$ satisfies $\mu_\alpha = \alpha^{-1}\mathcal{L}^{\star}\mu_\alpha$ where $\mathcal{L}$ denotes the infinitesimal generator of the Markov process $X(t)$. Let us consider the special case $\xi(t) = \sigma B(t) - ct$. We have:*

$$\mathrm{d}X(t) = \sigma X(t)\,\mathrm{d}B(t) + \left(\left(\frac{\sigma^2}{2} - c\right)X(t) + 1\right)\mathrm{d}t$$

and:

$$\mathcal{L} = \left(\left(\frac{\sigma^2}{2} - c\right)x + 1\right)\partial_x + \frac{1}{2}\sigma^2 x^2 \partial_x^2$$

We deduce that the density function of $A(\infty)$ is equal to:

$$f_{A(\infty)}(u) = \frac{\theta^k}{\Gamma(k)\,u^{k+1}}\exp\left(-\frac{\theta}{u}\right)$$

where $\theta = 2/\sigma^2$ and $k = 2c/\sigma^2$. Therefore, we have:

$$A(\infty) \stackrel{\text{law}}{=} \frac{\theta}{Z_k}$$

where Z_k is the Gamma random variable with parameter k. In the stochastic Gaussian model with $\rho = 0$, we have:

$$\begin{cases} \mathrm{d}F(t) = \alpha(t)\,F(t)^{\beta}\,\mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

where $W_1^{\mathbb{Q}}(t)$ and $W_2^{\mathbb{Q}}(t)$ are independent. It follows that:

$$\alpha(t) = \alpha\exp\left(-\frac{1}{2}\nu^2 t + \nu\,\mathrm{d}W_2^{\mathbb{Q}}(t)\right)$$

and:

$$F(t) = F_0 + \alpha\int_0^t \exp\left(-\frac{1}{2}\nu^2 s + \nu\,\mathrm{d}W_2^{\mathbb{Q}}(s)\right)\mathrm{d}W_1^{\mathbb{Q}}(s)$$

We deduce that $F(t) \stackrel{\text{law}}{=} F_0 + \alpha W_1^{\mathbb{Q}}(\langle F(t)\rangle)$ where:

$$\langle F(t)\rangle = \int_0^t \exp\left(2\nu W_2^{\mathbb{Q}}(s) - \nu^2 s\right)\mathrm{d}s$$

We have:

$$\langle F(\infty)\rangle \stackrel{\text{law}}{=} \frac{\theta}{Z_k}$$

where $\theta = 2/(2\nu)^2 = 1/\left(2\nu^2\right)$ and $k = 2\nu^2/(2\nu)^2 = 1/2$. Since we have $W_1^{\mathbb{Q}}(t) \stackrel{\text{law}}{=} \mathcal{N}(0,1)\sqrt{t}$, we conclude that:

$$F(\infty) \stackrel{\text{law}}{=} F_0 + \alpha\mathcal{N}(0,1)\sqrt{\frac{1}{2\nu^2 Z_{1/2}}}$$

[8] Donati-Martin, C., Ghomrasni, R., and Yor, M. (2001), On Certain Markov Processes Attached to Exponential Functionals of Brownian Motion; Applications to Asian Options, *Revista Matematica Iberoamericana*, 17(1), pp. 179-193.

9.4.10 The quadratic Gaussian model

1. We know that the bond price:

$$B(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,\mathrm{d}s}\right]$$

is the solution of the following PDE:

$$\begin{aligned}\frac{1}{2}\operatorname{tr}\left(\Sigma(t)\,\partial_X^2 B(t,T)\,\Sigma(t)^\top\right) + \\ \partial_X B(t,T)^\top\left(a(t) + B(t)\,X(t)\right) + \\ \partial_t B(t,T) - r(t)\,B(t,T) &= 0 \qquad (9.7)\end{aligned}$$

with $B(T,T) = 1$.

2. We assume that the solution of $B(t,T)$ has the following form:

$$B(t,T) = \exp\left(-\hat{\alpha}(t,T) - \hat{\beta}(t,T)^\top X(t) - X(t)^\top \hat{\Gamma}(t,T)\,X(t)\right)$$

where $\hat{\Gamma}(t,T)$ is a symmetric matrix. We obtain:

$$\frac{\partial_t B(t,T)}{B(t,T)} = -\partial_t\hat{\alpha}(t,T) - \partial_t\hat{\beta}(t,T)^\top X(t) - X(t)^\top \partial_t\hat{\Gamma}(t,T)\,X(t)$$

and:

$$\frac{\partial_X B(t,T)}{B(t,T)} = -\hat{\beta}(t,T) - 2\hat{\Gamma}(t,T)\,X(t)$$

We deduce that:

$$\begin{aligned}\frac{\partial_X^2 B(t,T)}{B(t,T)} &= \frac{\partial_X\left(\partial_X B(t,T)^\top\right)}{B(t,T)} \\ &= -2\hat{\Gamma}(t,T) + \\ &\quad \left(\hat{\beta}(t,T) + 2\hat{\Gamma}(t,T)\,X(t)\right)\left(\hat{\beta}(t,T) + 2\hat{\Gamma}(t,T)\,X(t)\right)^\top \\ &= -2\hat{\Gamma}(t,T) + \hat{\beta}(t,T)\,\hat{\beta}(t,T)^\top + \\ &\quad 2\hat{\Gamma}(t,T)\,X(t)\,\hat{\beta}(t,T)^\top + \\ &\quad 2\hat{\beta}(t,T)\,X(t)^\top\,\hat{\Gamma}(t,T) + \\ &\quad 4\hat{\Gamma}(t,T)\,X(t)\,X(t)^\top\,\hat{\Gamma}(t,T)\end{aligned}$$

By using the matrix property $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ if the product BA makes a sense, we can write Equation (9.7) as follows:

$$\begin{aligned}-\operatorname{tr}\left(\Sigma(t)\,\hat{\Gamma}(t,T)\,\Sigma(t)^\top\right) + \\ \frac{1}{2}\operatorname{tr}\left(\Sigma(t)\,\hat{\beta}(t,T)\,\hat{\beta}(t,T)^\top\,\Sigma(t)^\top\right) + \\ +2X(t)^\top\,\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\beta}(t,T) + \\ 2X(t)^\top\,\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\Gamma}(t,T)\,X(t) - \\ \left(\hat{\beta}(t,T) + 2\hat{\Gamma}(t,T)\,X(t)\right)^\top\left(a(t) + B(t)\,X(t)\right) - \\ \partial_t\hat{\alpha}(t,T) - X(t)^\top\,\partial_t\hat{\beta}(t,T) - X(t)^\top\,\partial_t\hat{\Gamma}(t,T)\,X(t) - \\ \left(\alpha(t) + X(t)^\top\,\beta(t) + X(t)^\top\,\Gamma(t)\,X(t)\right) &= 0\end{aligned}$$

We regroup the terms by the polynomial degree in X. For degree 0, we obtain:

$$\begin{aligned} -\operatorname{tr}\left(\Sigma(t)\,\hat{\Gamma}(t,T)\,\Sigma(t)^\top\right) + \\ \frac{1}{2}\operatorname{tr}\left(\Sigma(t)\,\hat{\beta}(t,T)\,\hat{\beta}(t,T)^\top\,\Sigma(t)^\top\right) - \\ \hat{\beta}(t,T)^\top a(t) - \partial_t\hat{\alpha}(t,T) - \alpha(t) &= 0 \end{aligned}$$

or:

$$\begin{aligned} \partial_t\hat{\alpha}(t,T) &= -\operatorname{tr}\left(\Sigma(t)\,\Sigma(t)^\top\,\hat{\Gamma}(t,T)\right) - \hat{\beta}(t,T)^\top a(t) + \\ &\quad \frac{1}{2}\hat{\beta}(t,T)^\top\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\beta}(t,T) - \alpha(t) \end{aligned}$$

For degree 1, we obtain:

$$\begin{aligned} 2\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\beta}(t,T) - B(t)^\top\,\hat{\beta}(t,T) - \\ 2\hat{\Gamma}(t,T)\,a(t) - \partial_t\hat{\beta}(t,T) - \beta(t) &= 0 \end{aligned}$$

or:

$$\begin{aligned} \partial_t\hat{\beta}(t,T) &= -B(t)^\top\,\hat{\beta}(t,T) + 2\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\beta}(t,T) - \\ &\quad 2\hat{\Gamma}(t,T)\,a(t) - \beta(t) \end{aligned}$$

For degree 2, we obtain:

$$\begin{aligned} 2\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\Gamma}(t,T) - 2\hat{\Gamma}(t,T)\,B(t) - \\ \partial_t\hat{\Gamma}(t,T) - \Gamma(t) &= 0 \end{aligned}$$

or:

$$\begin{aligned} \partial_t\hat{\Gamma}(t,T) &= 2\hat{\Gamma}(t,T)\,\Sigma(t)\,\Sigma(t)^\top\,\hat{\Gamma}(t,T) - \\ &\quad 2\hat{\Gamma}(t,T)\,B(t) - \Gamma(t) \end{aligned}$$

3. $B(t)$ must be a diagonal matrix in order to ensure that $\hat{\Gamma}(t,T)$ is a symmetric matrix. Indeed, if we do not consider this hypothesis, we obtain:

$$\begin{aligned} \partial_t\hat{\Gamma}(t,T) &= \frac{1}{2}\left(\hat{\Gamma}(t,T) + \hat{\Gamma}(t,T)^\top\right)\Sigma(t)\,\Sigma(t)^\top\left(\hat{\Gamma}(t,T) + \hat{\Gamma}(t,T)^\top\right) - \\ &\quad \left(\hat{\Gamma}(t,T) + \hat{\Gamma}(t,T)^\top\right)B(t) - \Gamma(t) \end{aligned}$$

It follows that the term $\left(\hat{\Gamma}(t,T) + \hat{\Gamma}(t,T)^\top\right)B(t)$ is not symmetric.

4. We recall that:

$$\mathrm{d}X(t) = \left(\tilde{a}(t) + \tilde{B}(t)\,X(t)\right)\,\mathrm{d}t + \Sigma(t)\,\mathrm{d}W^{\mathbb{Q}^\star(T)}(t)$$

where:

$$\tilde{a}(t) = a(t) - \Sigma(t)\,\Sigma(t)^\top\,\hat{\beta}(t,T)$$

and:

$$\tilde{B}(t) = B(t) - 2\Sigma(t)\,\Sigma(t)^\top\,\hat{\Gamma}(t,T)$$

It follows that:

$$\begin{aligned} X(t) &= e^{\tilde{B}(t)}X_0 + e^{\tilde{B}(t)}\int_0^t e^{-\tilde{B}(s)}\tilde{a}(s)\ \mathrm{d}s + \\ & e^{\tilde{B}(t)}\int_0^t e^{-\tilde{B}(s)}\Sigma(s)\ \mathrm{d}W^{\mathbb{Q}^\star(T)}(s) \end{aligned}$$

We conclude that $X(t)$ is Gaussian under the forward probability measure $\mathbb{Q}^\star(T)$:

$$X(t) \sim \mathcal{N}(m(0,t), V(0,t))$$

We have:

$$m(0,t) = e^{\tilde{B}(t)}X_0 + e^{\tilde{B}(t)}\int_0^t e^{-\tilde{B}(s)}\tilde{a}(s)\ \mathrm{d}s$$

that is the solution of the following EDO:

$$\begin{aligned} \partial_t m(0,t) &= \tilde{a}(t) + \tilde{B}(t)\, m(0,t) \\ &= a(t) - \Sigma(t)\Sigma(t)^\top \hat{\beta}(t,T) + \\ & B(t)\, m(0,t) - 2\Sigma(t)\Sigma(t)^\top \hat{\Gamma}(t,T)\, m(0,t) \end{aligned}$$

We also have:

$$V(0,t) = \int_0^t e^{\tilde{B}(t)}e^{-\tilde{B}(s)}\Sigma(s)\Sigma(s)^\top e^{-\tilde{B}(s)^\top}e^{\tilde{B}(t)^\top}\ \mathrm{d}s$$

or:

$$e^{-\tilde{B}(t)}V(0,t)\, e^{-\tilde{B}(t)^\top} = \int_0^t e^{-\tilde{B}(s)}\Sigma(s)\Sigma(s)^\top e^{-\tilde{B}(s)^\top}\ \mathrm{d}s$$

It follows that:

$$\begin{aligned} -\tilde{B}(t)\, e^{-\tilde{B}(t)}V(0,t)\, e^{-\tilde{B}(t)^\top} + & \\ e^{-\tilde{B}(t)}\partial_t V(0,t)\, e^{-\tilde{B}(t)^\top} - & \\ e^{-\tilde{B}(t)}V(0,t)\,\tilde{B}(t)^\top e^{-\tilde{B}(t)^\top} &= e^{-\tilde{B}(t)}\Sigma(t)\Sigma(t)^\top e^{-\tilde{B}(t)^\top} \end{aligned}$$

or:

$$\begin{aligned} \partial_t V(0,t) &= \tilde{B}(t)\, V(0,t) + V(0,t)\,\tilde{B}(t)^\top + \Sigma(t)\Sigma(t)^\top \\ &= B(t)\, V(0,t) + V(0,t)\, B(t)^\top - \\ & 4V(0,t)\,\hat{\Gamma}(t,T)^\top \Sigma(t)\Sigma(t)^\top + \Sigma(t)\Sigma(t)^\top \end{aligned}$$

In our approach, the dynamics of $m(0,t)$ and $V(0,t)$ are obtained under the forward probability measure $\mathbb{Q}^\star(T)$. In the paper of El Karoui *et al.* (1992a), the dynamics of $m(t,T)$ and $V(t,T)$ are obtained under the probability measure $\mathbb{Q}^\star(t,T)$:

$$\left\{\begin{aligned} \partial_T m(t,T) &= a(T) + B(T)\, m(t,T) - 2V(t,T)\,\Gamma(T)\, m(t,T) - \\ & V(t,T)\,\beta(T) \\ \partial_T V(t,T) &= V(t,T)\, B(T)^\top + B(T)\, V(t,T) - \\ & 2V(t,T)\,\Gamma(T)\, V(t,T) + \Sigma(T)\Sigma(T)^\top \end{aligned}\right.$$

5. The Libor rate $L(t,T_{i-1},T_i)$ at time t between the dates T_{i-1} and T_i is defined by:

$$L(t,T_{i-1},T_i) = \frac{1}{\delta_{i-1}}\left(\frac{B(t,T_{i-1})}{B(t,T_i)} - 1\right)$$

where $\delta_{i-1} = T_i - T_{i-1}$.

6. The payoff of the caplet is given by:

$$\begin{aligned} f(X) &= \delta_{i-1}\left(L(t,T_{i-1},T_i)-K\right)^+ \\ &= \left(\frac{B(t,T_{i-1})}{B(t,T_i)}-(1+\delta_{i-1}K)\right)^+ \\ &= \frac{1}{B(t,T_i)}\left(B(t,T_{i-1})-(1+\delta_{i-1}K)B(t,T_i)\right)^+ \end{aligned}$$

It follows that the price of the caplet is given by:

$$\begin{aligned} \text{Caplet} &= \mathbb{E}^{\mathbb{Q}}\left[\frac{e^{-\int_0^{T_i} r(s)\,\mathrm{d}s}}{B(t,T_i)}\left(B(t,T_{i-1})-(1+\delta_{i-1}K)B(t,T_i)\right)^+\right] \\ &= B(0,t)\,\mathbb{E}^{\mathbb{Q}^\star(t)}\left[\left(B(t,T_{i-1})-(1+\delta_{i-1}K)B(t,T_i)\right)^+\right] \\ &= B(0,t)\,\mathbb{E}^{\mathbb{Q}^\star(t)}\left[\max(0,g(X))\right] \end{aligned}$$

where:

$$\begin{aligned} g(x) &= \exp\left(-\hat{\alpha}(t,T_{i-1})-\hat{\beta}(t,T_{i-1})x-\hat{\Gamma}(t,T_{i-1})x^2\right)- \\ &\quad (1+\delta_{i-1}K)\exp\left(-\hat{\alpha}(t,T_i)-\hat{\beta}(t,T_i)x-\hat{\Gamma}(t,T_i)x^2\right) \end{aligned}$$

7. We have:

$$\begin{aligned} \text{Caplet} &= B(0,t)\int_{-\infty}^{+\infty} f(x)\,\phi(x;m(0,t),V(0,t))\,\mathrm{d}x \\ &= B(0,t)\int_{-\infty}^{+\infty}\max(0,g(x))\,\phi(x;m(0,t),V(0,t))\,\mathrm{d}x \\ &= B(0,t)\int_{\mathcal{E}} g(x)\,\phi(x;m(0,t),V(0,t))\,\mathrm{d}x \\ &= B(0,t)\int_{\mathcal{E}} h(x)\,\mathrm{d}x \end{aligned}$$

where $\mathcal{E}=\{x:g(x)\geq 0\}$ is the exercise domain of the option and:

$$h(x)=g(x)\,\phi(x;m(0,t),V(0,t))$$

We note $a_i=\hat{\Gamma}(t,T_i)$, $b_i=\hat{\beta}(t,T_i)$, $c_i=\hat{\alpha}(t,T_i)$ and $d=1+\delta_{i-1}K$. It follows that:

$$\begin{aligned} g(x)\geq 0 &\Leftrightarrow \exp\left(-a_{i-1}x^2-b_{i-1}x-c_{i-1}\right)\geq d\exp\left(-a_ix^2-b_ix-c_i\right) \\ &\Leftrightarrow a_{i-1}x^2+b_{i-1}x+c_{i-1}\leq a_ix^2+b_ix+c_i-\ln d \\ &\Leftrightarrow ax^2+bx+c\geq 0 \end{aligned}$$

where $a=a_i-a_{i-1}$, $b=b_i-b_{i-1}$, $c=c_i-c_{i-1}-\ln d$. Let $\Delta=b^2-4ac$ be the discriminant of the quadratic polynomial. If $\Delta\leq 0$ and $a>0$, $\mathcal{E}=(-\infty,+\infty)$. If $\Delta\leq 0$ and $a<0$, $\mathcal{E}=\emptyset$. If $\Delta>0$ and $a>0$, $\mathcal{E}=(-\infty,x_1]\cup[x_2,+\infty)$ where:

$$x_1=\frac{-b-\sqrt{b^2-4ac}}{2a}$$

and:

$$x_2=\frac{-b+\sqrt{b^2-4ac}}{2a}$$

If $\Delta>0$ and $a\leq 0$, $\mathcal{E}=[x_1,x_2]$.

8. We have:

$$\begin{aligned} \mathcal{J} &= \int_{x_1}^{x_2} \frac{e^{-ax^2-bx-c}}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(x-m)^2} \,\mathrm{d}x \\ &= \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi V}} e^{P(x)} \,\mathrm{d}x \end{aligned}$$

where $P(x)$ is a quadratic polynomial:

$$\begin{aligned} P(x) &= -\frac{1}{2V}(x-m)^2 - ax^2 - bx - c \\ &= -\frac{1}{2V}x^2 + \frac{m}{V}x - \frac{m^2}{2V} - ax^2 - bx - c \\ &= -\left(\frac{1}{2V} + a\right)x^2 + \left(\frac{m}{V} - b\right)x - \left(\frac{m^2}{2V} + c\right) \\ &= -\frac{1}{2}\left(\frac{1+2aV}{V}\right)x^2 + \left(\frac{m-bV}{V}\right)x - \left(\frac{m^2+2cV}{2V}\right) \end{aligned}$$

We can write $P(x)$ as follows:

$$\begin{aligned} P(x) &= -\frac{1}{2\tilde{V}}x^2 + \frac{\tilde{m}}{\tilde{V}}x - \frac{\tilde{m}^2}{2\tilde{V}} - \tilde{c} \\ &= -\frac{1}{2\tilde{V}}(x-\tilde{m})^2 - \tilde{c} \end{aligned}$$

where:

$$\begin{aligned} \tilde{V} &= \frac{V}{1+2aV} \\ \tilde{m} &= \frac{m-bV}{1+2aV} \\ \tilde{c} &= \left(\frac{m^2+2cV}{2V}\right) - \frac{\tilde{m}^2}{2\tilde{V}} \end{aligned}$$

We deduce that:

$$\begin{aligned} \mathcal{J} &= \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(x-\tilde{m})^2 - \tilde{c}} \,\mathrm{d}x \\ &= \sqrt{\frac{\tilde{V}}{V}} e^{-\tilde{c}} \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\tilde{V}}} e^{-\frac{1}{2\tilde{V}}(x-\tilde{m})^2} \,\mathrm{d}x \\ &= \sqrt{\frac{1}{1+2aV}} e^{-\tilde{c}} \left(\Phi\left(\frac{x_2-\tilde{m}}{\sqrt{\tilde{V}}}\right) - \Phi\left(\frac{x_1-\tilde{m}}{\sqrt{\tilde{V}}}\right)\right) \end{aligned}$$

We also have:

$$\frac{x-\tilde{m}}{\sqrt{\tilde{V}}} = \sqrt{\frac{1+2aV}{V}}x + \sqrt{V}\frac{b-m/V}{\sqrt{1+2aV}}$$

and:

$$\begin{aligned} \tilde{c} &= \left(\frac{m^2+2cV}{2V}\right) - \frac{\tilde{m}^2}{2\tilde{V}} \\ &= c + \frac{m^2(1+2aV)}{2V(1+2aV)} - \frac{(m-bV)^2}{2V(1+2aV)} \\ &= c + \frac{2am^2+2bm-b^2V}{2(1+2aV)} \end{aligned}$$

9. The price of the caplet is equal to:

$$\begin{aligned}\text{Caplet} &= B(0,t)\int_{\mathcal{E}} h(x)\,\mathrm{d}x \\ &= B(0,t)\,\mathcal{I}_{i-1}(\mathcal{E}) - B(0,t)(1+\delta_{i-1}K)\,\mathcal{I}_i(\mathcal{E})\end{aligned}$$

where:

$$\mathcal{I}_i(\mathcal{E}) = \int_{\mathcal{E}} e^{-\hat{\alpha}(t,T_i)-\hat{\beta}(t,T_i)x-\hat{\Gamma}(t,T_i)x^2}\phi(x;m(0,t),V(0,t))\,\mathrm{d}x$$

Since we can write $\mathcal{I}_i(\mathcal{E})$ in terms of $\mathcal{J}$, we obtain an analytical formula of the caplet price. For instance, if $\mathcal{E} = (-\infty, x_1] \cup [x_2, +\infty)$, we have:

$$\begin{aligned}\mathcal{I}_i(\mathcal{E}) &= \mathcal{J}\left(\hat{\alpha}(t,T_i), \hat{\beta}(t,T_i), \hat{\Gamma}(t,T_i), m(0,t), V(0,t), -\infty, x_1\right) + \\ &\quad \mathcal{J}\left(\hat{\alpha}(t,T_i), \hat{\beta}(t,T_i), \hat{\Gamma}(t,T_i), m(0,t), V(0,t), x_2, +\infty\right)\end{aligned}$$

If $\mathcal{E} = [x_1, x_2]$, $\mathcal{I}_i(\mathcal{E})$ becomes:

$$\mathcal{I}_i(\mathcal{E}) = \mathcal{J}\left(\hat{\alpha}(t,T_i), \hat{\beta}(t,T_i), \hat{\Gamma}(t,T_i), m(0,t), V(0,t), x_1, x_2\right)$$

9.4.11 Pricing two-asset basket options

1. Let $f(S_1(T), S_2(T)) = (\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+$ be the payoff of the option. Using Feynman-Kac representation, we know that:

$$\mathcal{C}_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r\,\mathrm{d}t} f(S_1(T), S_2(T))\right]$$

where:

$$\begin{cases} S_1(T) = S_1(0)\,e^{\left(b_1-\frac{1}{2}\sigma_1^2\right)T+\sigma_1\sqrt{T}\varepsilon_1} \\ S_2(T) = S_2(0)\,e^{\left(b_2-\frac{1}{2}\sigma_2^2\right)T+\sigma_2\sqrt{T}\varepsilon_2} \end{cases}$$

and $(\varepsilon_1, \varepsilon_2)$ is a standardized Gaussian random vector with $\rho(\varepsilon_1, \varepsilon_2) = \rho$. Since the probability density function of $(\varepsilon_1, \varepsilon_2)$ is equal to:

$$h(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(x_1^2+x_2^2-2\rho x_1 x_2\right)}$$

we have:

$$\mathcal{C}_0 = e^{-rT}\iint_{\mathbb{R}^2} g(x_1, x_2)\,h(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

where:

$$\begin{cases} g(x_1, x_2) = (g_1(x_1) + g_2(x_2) - K)^+ \\ g_1(x_1) = \alpha_1 S_1(0)\,e^{\left(b_1-\frac{1}{2}\sigma_1^2\right)T+\sigma_1\sqrt{T}x_1} \\ g_2(x_2) = \alpha_2 S_2(0)\,e^{\left(b_2-\frac{1}{2}\sigma_2^2\right)T+\sigma_2\sqrt{T}x_2} \end{cases}$$

2. (a) Since we have:

$$Ae^{b+cx} - D \geq 0 \Leftrightarrow x \geq \frac{1}{c}\ln\frac{D}{A} - \frac{b}{c}$$

we deduce that[9]:

$$
\begin{aligned}
\mathbb{E}\left[\left(Ae^{b+c\varepsilon}-D\right)^{+}\right] &= \int_{\mathbb{R}}\left(Ae^{b+cx}-D\right)^{+}\phi(x)\,\mathrm{d}x \\
&= Ae^{b}\int_{\frac{1}{c}\ln\frac{D}{A}-\frac{b}{c}}^{\infty}e^{cx}\phi(x)\,\mathrm{d}x- \\
&\quad D\int_{\frac{1}{c}\ln\frac{D}{A}-\frac{b}{c}}^{\infty}\phi(x)\,\mathrm{d}x \\
&= Ae^{b+\frac{1}{2}c^{2}}\int_{\frac{1}{c}\ln\frac{D}{A}-\frac{b}{c}-c}^{\infty}e^{cx}\phi(x)\,\mathrm{d}x- \\
&\quad D\int_{\frac{1}{c}\ln\frac{D}{A}-\frac{b}{c}}^{\infty}\phi(x)\,\mathrm{d}x \\
&= Ae^{b+\frac{1}{2}c^{2}}\Phi(d_1)-D\Phi(d_1+c)
\end{aligned}
$$

where:

$$d_1=\frac{1}{c}\left(\ln\frac{A}{D}+b\right)$$

(b) If $A<0$ and $D>0$, we have:

$$\left(Ae^{b+c\varepsilon}-D\right)^{+}=0$$

and:

$$\mathbb{E}\left[\left(Ae^{b+c\varepsilon}-D\right)^{+}\right]=0$$

If $A<0$ and $D<0$, we have:

$$\left(Ae^{b+c\varepsilon}-D\right)^{+}=\left(-D+Ae^{b+c\varepsilon}\right)^{+}$$

and:

$$\mathbb{E}\left[\left(Ae^{b+c\varepsilon}-D\right)^{+}\right]=-D\Phi(-d_1-c)+Ae^{b+\frac{1}{2}c^{2}}\Phi(-d_1)$$

If $A>0$ and $D<0$, we have:

$$\left(Ae^{b+c\varepsilon}-D\right)^{+}=Ae^{b+c\varepsilon}-D$$

and:

$$\mathbb{E}\left[\left(Ae^{b+c\varepsilon}-D\right)^{+}\right]=Ae^{b+\frac{1}{2}c^{2}}-D$$

3. Using the Cholesky decomposition, we have:

$$\varepsilon_2=\rho\varepsilon_1+\sqrt{1-\rho^{2}}\varepsilon_3$$

where $(\varepsilon_1,\varepsilon_3)$ is a standardized Gaussian random vector with $\rho(\varepsilon_1,\varepsilon_3)=0$. We deduce that the pdf of $(\varepsilon_1,\varepsilon_3)$ is given by:

$$h'(x_1,x_3)=\frac{1}{2\pi}e^{-\frac{1}{2}\left(x_1^{2}+x_3^{2}\right)}$$

[9]We recall that:

$$e^{cx}\phi(x)=e^{\frac{1}{2}c^{2}}e^{-\frac{1}{2}(x-c)^{2}}$$

Therefore, we have:

$$\mathcal{C}_0 = e^{-rT} \iint_{\mathbb{R}^2} g'(x_1, x_3)\, h'(x_1, x_3)\; \mathrm{d}x_1 \,\mathrm{d}x_3$$

where:

$$\begin{cases} g'(x_1, x_3) = (g_1'(x_1) + g_2'(x_1, x_3) - K)^+ \\ g_1'(x_1) = \alpha_1 S_1(0)\, e^{\left(b_1 - \frac{1}{2}\sigma_1^2\right)T + \sigma_1 \sqrt{T} x_1} \\ g_2'(x_1, x_3) = \alpha_2 S_2(0)\, e^{\left(b_2 - \frac{1}{2}\sigma_2^2\right)T + \rho\sigma_2\sqrt{T}x_1 + \sqrt{1-\rho^2}\sigma_2\sqrt{T}x_3} \end{cases}$$

It follows that:

$$\mathcal{C}_0 = \int_{\mathbb{R}} e^{-rT} \left(\int_{\mathbb{R}} (g_2'(x_1, x_3) - (K - g_1'(x_1)))^+ \phi(x_3)\; \mathrm{d}x_3 \right) \phi(x_1)\; \mathrm{d}x_1$$

where:

$$\begin{aligned} g_2'(x_1, x_3) \quad = \quad & \left(\alpha_2 S_2(0)\, e^{\rho\sigma_2\sqrt{T}x_1}\right) \times \\ & e^{\left(b_2 - \frac{1}{2}\rho^2\sigma_2^2 - \frac{1}{2}(1-\rho^2)\sigma_2^2\right)T + \sqrt{1-\rho^2}\sigma_2\sqrt{T}x_3} \end{aligned}$$

Since we have $S_2^\star > 0$ and $K^\star > 0$, we deduce that:

$$\mathcal{C}_0 = \int_{\mathbb{R}} \mathrm{BS}(S^\star, K^\star, \sigma^\star, T, b^\star, r)\, \phi(x_1)\; \mathrm{d}x_1$$

where:

$$\begin{cases} S^\star = \alpha_2 S_2(0)\, e^{\rho\sigma_2\sqrt{T}x_1} \\ K^\star = K - \alpha_1 S_1(0)\, e^{\left(b_1 - \frac{1}{2}\sigma_1^2\right)T + \sigma_1\sqrt{T}x_1} \\ \sigma^\star = \sigma_2\sqrt{1-\rho^2} \\ b^\star = b_2 - \frac{1}{2}\rho^2\sigma_2^2 \end{cases}$$

In this case, the Black-Scholes formula is equal to:

$$\mathrm{BS}(S^\star, K^\star, \sigma^\star, T, b^\star, r) = S^\star e^{(b^\star - r)T}\Phi(d_1) - K^\star e^{-rT}\Phi(d_2)$$

where:

$$d_1 = \frac{1}{\sigma^\star\sqrt{T}}\left(\ln\frac{S^\star}{K^\star} + b^\star T\right) + \frac{1}{2}\sigma^\star\sqrt{T}$$

and:

$$d_2 = d_1 - \sigma^\star\sqrt{T}$$

4. If $\alpha_1 > 0$, $\alpha_2 < 0$ and $K > 0$, we obtain the same formula:

$$\mathcal{C}_0 = \int_{\mathbb{R}} \mathrm{BS}(S^\star, K^\star, \sigma^\star, T, b^\star, r)\, \phi(x_1)\; \mathrm{d}x_1$$

with:

$$\mathrm{BS}(S, K, \sigma, T, b, r) = -S^\star e^{(b^\star - r)T}\Phi(-d_1) + K^\star e^{-rT}\Phi(-d_2)$$

5. In the general case, we can obtain the following options:

$$\mathbb{E}\left[(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+\right] = \begin{cases} \mathbb{E}\left[(S^\star - K^\star)^+\right] & \text{(call)} \\ \mathbb{E}\left[(K^\star - S^\star)^+\right] & \text{(put)} \\ \mathbb{E}[S^\star] + K^\star & \text{(e)} \\ 0 & \text{(0)} \end{cases}$$

where $S^\star > 0$ and $K^\star > 0$. Table 9.1 shows that we cannot always transform the two-dimensional integral into a one-dimensional integral.

TABLE 9.1: Pricing basket options with one-dimensional integration

Case	α_1	α_2	K	Type	$S^\star$	$K^\star$	1D
#1	+	+	+	(call)	$\alpha_1 S_1(T) + \alpha_2 S_2(T)$	K	
#2	+	+	−	(e)	$\alpha_1 S_1(T) + \alpha_2 S_2(T)$	$-K$	✓
#3	+	−	+	(call)	$\alpha_1 S_1(T)$	$K - \alpha_2 S_2(T)$	✓
#4	+	−	−	(call)	$\alpha_1 S_1(T) - K$	$-\alpha_2 S_2(T)$	
#5	−	+	+	(call)	$\alpha_2 S_2(T)$	$K - \alpha_1 S_1(T)$	✓
#6	−	+	−	(call)	$\alpha_2 S_2(T) - K$	$-\alpha_1 S_1(T)$	
#7	−	−	+	(0)			✓
#8	−	−	−	(put)	$\alpha_1 S_1(T) + \alpha_2 S_2(T)$	K	

Chapter 10

Statistical Inference and Model Estimation

10.3.1 Probability distribution of the t-statistic in the case of the linear regression model

1. We verify that $\mathbf{H}^\top = \left(\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)^\top = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top = \mathbf{H}$ and:

$$\begin{aligned}\mathbf{H}^2 &= \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top \\ &= \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top \\ &= \mathbf{H}\end{aligned}$$

   Since I_n is symetric, we also deduce that $\mathbf{L} = I_n - \mathbf{H}$ is symetric and idempotent:

$$\begin{aligned}\mathbf{L}^2 &= \left(I_n - \mathbf{H}\right)\left(I_n - \mathbf{H}\right) \\ &= I_n - 2\mathbf{H} + \mathbf{H}^2 \\ &= I_n - 2\mathbf{H} + \mathbf{H} \\ &= I_n - \mathbf{H}\end{aligned}$$

2. We have:

$$\mathbf{LX} = \left(I_n - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)\mathbf{X} = \mathbf{X} - \mathbf{X} = 0$$

   and:

$$\mathbf{X}^\top\mathbf{L} = \left(\mathbf{L}^\top\mathbf{X}\right)^\top = \left(\mathbf{LX}\right)^\top = 0$$

   We notice that:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y} = \mathbf{HY}$$

   and:

$$\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = \mathbf{LY}$$

   We deduce that:

$$\hat{\mathbf{U}} = \mathbf{LY} = \mathbf{L}\left(\mathbf{X}\beta + \mathbf{U}\right) = \mathbf{LX}\beta + \mathbf{LU} = \mathbf{LU}$$

3. We have:

$$\begin{aligned}\operatorname{trace}\left(\mathbf{L}\right) &= \operatorname{trace}\left(I_n - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right) \\ &= \operatorname{trace}\left(I_n\right) - \operatorname{trace}\left(\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right) \\ &= \operatorname{trace}\left(I_n\right) - \operatorname{trace}\left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{X}\right) \\ &= \operatorname{trace}\left(I_n\right) - \operatorname{trace}\left(I_K\right) \\ &= n - K\end{aligned}$$

   We know that the rank of an idempotent matrix is equal to its trace. We deduce that $\operatorname{rank}\left(\mathbf{L}\right) = \operatorname{trace}\left(\mathbf{L}\right) = n - K$.

4. We have:

$$\text{RSS}\left(\hat{\beta}\right) = \hat{\mathbf{U}}^\top \hat{\mathbf{U}} = (\mathbf{LU})^\top (\mathbf{LU}) = \mathbf{U}^\top \mathbf{L}^\top \mathbf{LU} = \mathbf{U}^\top \mathbf{LU}$$

It follows that:

$$\begin{aligned}
\mathbb{E}\left[\text{RSS}\left(\hat{\beta}\right)\right] &= \mathbb{E}\left[\mathbf{U}^\top \mathbf{LU}\right] \\
&= \mathbb{E}\left[\text{trace}\left(\mathbf{U}^\top \mathbf{LU}\right)\right] \\
&= \mathbb{E}\left[\text{trace}\left(\mathbf{LU}^\top \mathbf{U}\right)\right] \\
&= \text{trace}\left(\mathbb{E}\left[\mathbf{LU}^\top \mathbf{U}\right]\right) \\
&= \text{trace}\left(\mathbf{L}\mathbb{E}\left[\mathbf{U}^\top \mathbf{U}\right]\right) \\
&= \sigma^2 \,\text{trace}\left(\mathbf{L}\right) \\
&= (n-K)\,\sigma^2
\end{aligned}$$

and:

$$\mathbb{E}\left[\hat{\sigma}^2\right] = \mathbb{E}\left[\frac{\text{RSS}\left(\hat{\beta}\right)}{n-K}\right] = \sigma^2$$

5. We have:

$$\begin{aligned}
\mathbf{U}^\top \mathbf{LU} &= \sigma^2 \left(\left(\sigma I_n\right)^{-1} \mathbf{U}\right)^\top \mathbf{L} \left(\left(\sigma I_n\right)^{-1} \mathbf{U}\right) \\
&= \sigma^2 \mathbf{V}^\top \mathbf{LV}
\end{aligned}$$

Since $\mathbf{V}^\top \mathbf{LV}$ is a normalized Gaussian quadratic form, we have:

$$\mathbf{V}^\top \mathbf{LV} \sim \chi^2_\nu$$

because $\nu = \text{rank}\,\mathbf{L} = n-K$. We deduce that:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{\text{RSS}\left(\hat{\beta}\right)}{n-K} \\
&= \frac{\mathbf{U}^\top \mathbf{LU}}{n-K} \\
&= \frac{\sigma^2}{n-K} \mathbf{V}^\top \mathbf{LV} \\
&\sim \frac{\sigma^2}{n-K} \chi^2_{n-K}
\end{aligned}$$

6. We have:

$$\begin{aligned}
\text{cov}\left(\hat{\beta}, \hat{\mathbf{U}}\right) &= \mathbb{E}\left[\left(\hat{\beta} - \beta\right)\left(\hat{\mathbf{U}} - 0\right)^\top\right] \\
&= \mathbb{E}\left[\left(\hat{\beta} - \beta\right)\hat{\mathbf{U}}^\top\right] \\
&= \mathbb{E}\left[\hat{\beta}\hat{\mathbf{U}}^\top\right] - \beta\mathbb{E}\left[\hat{\mathbf{U}}^\top\right] \\
&= \mathbb{E}\left[\left(\beta + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{U}\right)(\mathbf{LU})^\top\right] \\
&= \beta\mathbb{E}\left[\mathbf{U}\right]^\top \mathbf{L}^\top + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbb{E}\left[\mathbf{UU}^\top\right] \mathbf{L}^\top \\
&= \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{L} \\
&= 0
\end{aligned}$$

7. We deduce that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent, because $\hat{\sigma}^2$ is a function of $\hat{\mathbf{U}}$. Moreover, we have:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 \left(\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)_{j,j}}} \sim \mathcal{N}(0,1)$$

and:

$$\frac{(n-K)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-K}$$

It follows that:

$$\begin{aligned} t\left(\hat{\beta}_j\right) &= \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\left(\hat{\beta}_j\right)} \\ &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \left(\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)_{j,j}}} \\ &= \frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 \left(\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)_{j,j}}}}{\sqrt{\frac{(n-K)\hat{\sigma}^2}{(n-K)\sigma^2}}} \\ &\sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2_{n-K}}{n-K}}} \\ &\sim \mathbf{t}_{n-K} \end{aligned}$$

10.3.2 Linear regression without a constant

1. We have:

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{1,1} & & x_{1,K} \\ & \ddots & \\ x_{n,1} & & x_{n,K} \end{pmatrix}, \\ \beta &= \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \end{aligned}$$

where $\varepsilon \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$. The sum of squared residuals $\varepsilon^\top \varepsilon$ is[1] :

$$\begin{aligned} \varepsilon^\top \varepsilon &= (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}^\top \mathbf{Y} - \beta^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\beta + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \\ &= \mathbf{Y}^\top \mathbf{Y} - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \end{aligned}$$

It follows that:

$$\begin{aligned} \hat{\beta} &= \arg\min \varepsilon^\top \varepsilon \\ &= \arg\min \frac{1}{2}\beta^\top \left(\mathbf{X}^\top \mathbf{X}\right)\beta - \beta^\top \left(\mathbf{X}^\top \mathbf{Y}\right) \end{aligned}$$

$\hat{\beta}$ is the solution of a QP problem with $Q = \mathbf{X}^\top \mathbf{X}$ and $R = \mathbf{X}^\top \mathbf{Y}$.

[1] We have $\mathbf{Y}^\top \mathbf{X}\beta = \left(\mathbf{Y}^\top \mathbf{X}\beta\right)^\top = \beta^\top \mathbf{X}^\top \mathbf{Y}$ because $\mathbf{Y}^\top \mathbf{X}\beta$ is a scalar.

2. (a) We consider that there is a constant in the explanatory variables and we note $\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{X}_\star \end{pmatrix}$ where $\mathbf{X}_\star$ is the matrix of exogenous data without the constant. We write the coefficient β as follows:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_\star \end{pmatrix}$$

The first-order condition of the previous optimization problem is $Q\hat{\beta} = R$ or $\mathbf{X}^\top \mathbf{X}\hat{\beta} = \mathbf{X}^\top \mathbf{Y}$. We deduce that:

$$\left\{ \begin{array}{rcl} \mathbf{1}^\top \mathbf{1}\hat{\beta}_0 + \mathbf{1}^\top \mathbf{X}_\star \hat{\beta}_\star & = & \mathbf{1}^\top \mathbf{Y} \\ \mathbf{X}_\star^\top \mathbf{1}\hat{\beta}_0 + \mathbf{X}_\star^\top \mathbf{X}_\star \hat{\beta}_\star & = & \mathbf{X}_\star^\top \mathbf{Y} \end{array} \right.$$

If the residuals are centered, we must verify that $\mathbf{1}^\top \hat{\varepsilon} = 0$ or $\mathbf{1}^\top \left(\mathbf{Y} - \hat{\beta}_0 \mathbf{1} - \hat{\beta}_\star \mathbf{X}_\star\right) = 0$. We have[2] :

$$\begin{array}{rcl} \mathbf{1}^\top \left(\mathbf{Y} - \hat{\beta}_0 \mathbf{1} - \hat{\beta}_\star \mathbf{X}_\star\right) & = & \mathbf{1}^\top \mathbf{Y} - \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} - \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star \\ & = & \mathbf{1}^\top \mathbf{1}\hat{\beta}_0 + \mathbf{1}^\top \mathbf{X}_\star \hat{\beta}_\star - \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} - \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star \\ & = & \operatorname{trace}\left(\mathbf{1}^\top \mathbf{1}\hat{\beta}_0\right) + \operatorname{trace}\left(\mathbf{1}^\top \mathbf{X}_\star \hat{\beta}_\star\right) - \\ & & \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} - \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star \\ & = & \operatorname{trace}\left(\mathbf{1}^\top \hat{\beta}_0 \mathbf{1}\right) + \operatorname{trace}\left(\mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star\right) - \\ & & \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} - \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star \\ & = & \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} + \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star - \mathbf{1}^\top \hat{\beta}_0 \mathbf{1} - \mathbf{1}^\top \hat{\beta}_\star \mathbf{X}_\star \\ & = & 0 \end{array}$$

Adding a constant in the explanatory variables allows to center the residuals. If there is no intercept in the linear model, there is no reason that the residuals are centered.

(b) To center the residuals, we must add the constraint $\mathbf{1}^\top \varepsilon = 0$. We have $\mathbf{1}^\top \varepsilon = \mathbf{1}^\top \mathbf{Y} - \mathbf{1}^\top \mathbf{X}\beta$, which implies that $\mathbf{1}^\top \mathbf{X}\beta = \mathbf{1}^\top \mathbf{Y}$. The QP problem becomes:

$$\begin{array}{rcl} \hat{\beta} & = & \arg\min \dfrac{1}{2}\beta^\top \left(\mathbf{X}^\top \mathbf{X}\right)\beta - \beta^\top \left(\mathbf{X}^\top \mathbf{Y}\right) \\ & & \text{s.t.} \quad \left(\mathbf{1}^\top \mathbf{X}\right)\beta = \left(\mathbf{1}^\top \mathbf{Y}\right) \end{array}$$

We obtain a new QP problem with $Q = \mathbf{X}^\top \mathbf{X}$, $R = \mathbf{X}^\top \mathbf{Y}$, $A = \mathbf{1}^\top \mathbf{X}$ et $B = \mathbf{1}^\top \mathbf{Y}$.

(c) To transform the implicit constraints, we consider the explicit parametrization:

$$\beta = C\gamma + D$$

where C is the orthonormal basis for the kernel of the matrix $A = \mathbf{1}^\top \mathbf{X}$ and D is defined as follows:

$$\begin{array}{rcl} D & = & \left(A^\top A\right)^\star A^\top B \\ & = & \left(\mathbf{X}^\top \mathbf{1}\mathbf{1}^\top \mathbf{X}\right)^\star \mathbf{X}^\top \mathbf{1}\mathbf{1}^\top \mathbf{Y} \end{array}$$

[2] We use the following properties:

- $\operatorname{trace}(a) = a$ if a is a scalar;
- $\operatorname{trace}(AB) = \operatorname{trace}(BA)$ if the matrix multiplication BA is defined.

where $\left(A^\top A\right)^\star$ is the Moore-Penrose inverse of the matrix $A^\top A$. As the dimension of β is $K \times 1$, the matrices C, γ and D have the following dimensions $K \times (K-1)$, $(K-1) \times 1$ and $K \times 1$. The objective function becomes then:

$$\begin{aligned} f(\beta) &= \frac{1}{2}\beta^\top \left(\mathbf{X}^\top \mathbf{X}\right)\beta - \beta^\top \left(\mathbf{X}^\top \mathbf{Y}\right) \\ &= \frac{1}{2}(C\gamma + D)^\top \left(\mathbf{X}^\top \mathbf{X}\right)(C\gamma + D) - (C\gamma + D)^\top \left(\mathbf{X}^\top \mathbf{Y}\right) \\ &= \frac{1}{2}\gamma^\top C^\top \mathbf{X}^\top \mathbf{X} C\gamma + \frac{1}{2}D^\top \mathbf{X}^\top \mathbf{X} C\gamma + \frac{1}{2}\gamma^\top C^\top \mathbf{X}^\top \mathbf{X} D + \\ &\quad \frac{1}{2}D^\top \mathbf{X}^\top \mathbf{X} D - \gamma^\top C^\top \mathbf{X}^\top \mathbf{Y} - D^\top \mathbf{X}^\top \mathbf{Y} \\ &= \frac{1}{2}\gamma^\top \left(C^\top \mathbf{X}^\top \mathbf{X} C\right)\gamma + \gamma^\top \left(C^\top \mathbf{X}^\top \mathbf{X} D - C^\top \mathbf{X}^\top \mathbf{Y}\right) + \\ &\quad \left(\frac{1}{2}D^\top \mathbf{X}^\top \mathbf{X} D - D^\top \mathbf{X}^\top \mathbf{Y}\right) \end{aligned}$$

We deduce that:

$$\hat{\gamma} = \left(C^\top \mathbf{X}^\top \mathbf{X} C\right)^{-1} C^\top \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} D)$$

and:

$$\hat{\beta} = C\left(C^\top \mathbf{X}^\top \mathbf{X} C\right)^{-1} C^\top \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} D) + D$$

The analytical solution consists in computing C, D and finally $\hat{\beta}$.

10.3.3 Linear regression with linear constraints

1. (a) We have:

$$\begin{aligned} \text{RSS}(\beta) &= \mathbf{U}^\top \mathbf{U} \\ &= (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \\ &= \left(\mathbf{Y}^\top - \beta^\top \mathbf{X}^\top\right)(\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}^\top \mathbf{Y} - \beta^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\beta + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \\ &= \beta^\top \mathbf{X}^\top \mathbf{X}\beta - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \mathbf{Y}^\top \mathbf{Y} \end{aligned}$$

(b) The first-order condition is:

$$\frac{\partial\, \text{RSS}(\beta)}{\partial\, \beta} = 2\mathbf{X}^\top \mathbf{X}\beta - 2^\top \mathbf{X}^\top \mathbf{Y} = 0$$

We deduce that:

$$\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y}$$

(c) We have:

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top (\mathbf{X}\beta + \mathbf{U}) \\ &= \beta + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{U} \end{aligned}$$

Since $\mathbf{X} \perp \mathbf{U}$, $\hat{\beta}$ is an unbiased estimator of β:

$$\begin{aligned} \mathbb{E}\left[\hat{\beta}\right] &= \mathbb{E}\left[\beta + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{U}\right] \\ &= \beta + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbb{E}\left[\mathbf{X}^\top \mathbf{U}\right] \\ &= \beta \end{aligned}$$

and the variance of $\hat{\beta}$ is:

$$\begin{aligned}
\operatorname{cov}\left(\hat{\beta}\right) &= \mathbb{E}\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^\top\right] \\
&= \mathbb{E}\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{U}\mathbf{U}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right] \\
&= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbb{E}\left[\mathbf{U}\mathbf{U}^\top\right]\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top I_n\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}
\end{aligned}$$

2. (a) We have:

$$\begin{aligned}
\tilde{\beta} &= \arg\min \operatorname{RSS}(\beta) \\
&\quad \text{s.t.} \left\{\begin{array}{l} A\beta = B \\ C\beta \geq D \end{array}\right.
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\tilde{\beta} &= \arg\min \frac{1}{2}\beta^\top\left(2\mathbf{X}^\top\mathbf{X}\right)\beta - \beta^\top\left(2\mathbf{X}^\top\mathbf{Y}\right) \\
&\quad \text{s.t.} \left\{\begin{array}{l} A\beta = B \\ C\beta \geq D \end{array}\right.
\end{aligned}$$

We obtain a QP program with $Q = 2\mathbf{X}^\top\mathbf{X}$ and $R = 2\mathbf{X}^\top\mathbf{Y}$.

(b) We obtain $\hat{\beta}_1 = -1.01$, $\hat{\beta}_2 = 0.95$, $\hat{\beta}_3 = 2.04$, $\hat{\beta}_4 = 3.10$ and $\hat{\beta}_5 = -0.08$.

i. If $\sum_{i=1}^5 \beta_i = 1$, we have:

$$A = \left(\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \end{array}\right) \quad \text{and} \quad B = 1$$

We obtain $\tilde{\beta}_1 = -2.40$, $\tilde{\beta}_2 = 1.08$, $\tilde{\beta}_3 = 0.49$, $\tilde{\beta}_4 = 2.43$ and $\tilde{\beta}_5 = -0.60$.

ii. If $\beta_1 = \beta_2 = \beta_5$, we have:

$$A = \left(\begin{array}{ccccc} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{array}\right) \quad \text{and} \quad B = \left(\begin{array}{c} 0 \\ 0 \end{array}\right)$$

We obtain $\tilde{\beta}_1 = \tilde{\beta}_2 = \tilde{\beta}_5 = -0.08$, $\tilde{\beta}_3 = 2.22$ and $\tilde{\beta}_4 = 3.17$.

iii. If $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4 \geq \beta_5$, we have:

$$C = \left(\begin{array}{ccccc} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right) \quad \text{and} \quad D = \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \end{array}\right)$$

We obtain $\tilde{\beta}_1 = 1.33$, $\tilde{\beta}_2 = 1.33$, $\tilde{\beta}_3 = 1.33$, $\tilde{\beta}_4 = 1.33$ and $\tilde{\beta}_5 = -0.23$.

iv. If $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4 \leq \beta_5$ and $\sum_{i=1}^5 \beta_i = 1$, we have:

$$A = \left(\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \end{array}\right), \quad B = 1,$$

and:

$$C = \left(\begin{array}{ccccc} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right) \quad \text{and} \quad D = \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \end{array}\right)$$

We obtain $\tilde{\beta}_1 = -2.63$, $\tilde{\beta}_2 = 0.91$, $\tilde{\beta}_3 = 0.91$, $\tilde{\beta}_4 = 0.91$ and $\tilde{\beta}_5 = 0.91$. The first-order condition of the QP program is:

$$Q\tilde{\beta} - R - A^\top \lambda_A + C^\top \lambda_C = 0$$

where λ_A is the Lagrange coefficient associated to the equality constraint and λ_C is a vector of dimension 4×1 corresponding to the Lagrange coefficient associated to inequality constraints. Moreover, they verify the Kuhn-Tucker conditions:

$$\min\left(\lambda_C, C\tilde{\beta} - D\right) = 0$$

Since $\lambda_A = -192.36304$, we have:

$$C^\top \lambda_C = -\left(Q\tilde{\beta} - R - A^\top \lambda_A\right) = \begin{pmatrix} 0.0000 \\ 3.7244 \\ -2.8742 \\ 24.7449 \\ -25.5951 \end{pmatrix}$$

We deduce that:

$$\lambda_C = \begin{pmatrix} 0.0000 \\ 3.7244 \\ 0.8501 \\ 25.5951 \end{pmatrix}$$

3. (a) We have:

$$f(\beta; \lambda) = \frac{1}{2}\beta^\top \left(\mathbf{X}^\top \mathbf{X}\right)\beta - \beta^\top \left(\mathbf{X}^\top \mathbf{Y}\right) - \lambda^\top (A\beta - B)$$

The first-order condition is:

$$\frac{\partial\, f(\beta; \lambda)}{\partial\, \beta} = \left(\mathbf{X}^\top \mathbf{X}\right)\beta - \left(\mathbf{X}^\top \mathbf{Y}\right) - A^\top \lambda = 0$$

We have then:

$$\tilde{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right) - \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top \lambda$$

Since $A\beta = B$, we have:

$$A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right) - A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top \lambda = B$$

or:

$$\lambda = \left(A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top\right)^{-1}\left(A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right) - B\right)$$

We deduce that:

$$\begin{aligned}\tilde{\beta} &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right) - \\ &\quad \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top \left(A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top\right)^{-1}\left(A\left(X^\top X\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right) - B\right) \\ &= \hat{\beta} - \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top \left(A\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top\right)^{-1}\left(A\hat{\beta} - B\right)\end{aligned}$$

(b) To transform the explicit constraints into implicit constraints, we consider the parametrization:

$$\beta = C\gamma + D$$

where C is the orthonormal basis associated to the kernel of A and $D = \left(A^\top A\right)^\star A^\top B$ where $\left(A^\top A\right)^\star$ is the Moore-Penrose inverse of $A^\top A$. The objective function becomes:

$$\begin{aligned}
\text{RSS}(\beta) &= \frac{1}{2}\beta^\top \left(\mathbf{X}^\top\mathbf{X}\right)\beta - \beta^\top \left(\mathbf{X}^\top\mathbf{Y}\right) \\
&= \frac{1}{2}(C\gamma + D)^\top \left(\mathbf{X}^\top\mathbf{X}\right)(C\gamma + D) - (C\gamma + D)^\top \left(\mathbf{X}^\top\mathbf{Y}\right) \\
&= \frac{1}{2}\left(\gamma^\top C^\top \mathbf{X}^\top\mathbf{X}C\gamma + D^\top\mathbf{X}^\top\mathbf{X}C\gamma\right) + \\
&\quad \frac{1}{2}\left(\gamma^\top C^\top \mathbf{X}^\top\mathbf{X}D + D^\top\mathbf{X}^\top\mathbf{X}D\right) - \\
&\quad \gamma^\top C^\top\mathbf{X}^\top\mathbf{Y} - D^\top\mathbf{X}^\top\mathbf{Y} \\
&= \frac{1}{2}\gamma^\top\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)\gamma + \gamma^\top\left(C^\top\mathbf{X}^\top\mathbf{X}D - C^\top\mathbf{X}^\top\mathbf{Y}\right) + \\
&\quad \left(\frac{1}{2}D^\top\mathbf{X}^\top\mathbf{X}D - D^\top\mathbf{X}^\top\mathbf{Y}\right)
\end{aligned}$$

Therefore we deduce that:

$$\hat{\gamma} = \left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1} C^\top\mathbf{X}^\top\left(\mathbf{Y} - \mathbf{X}D\right)$$

and:

$$\tilde{\beta} = C\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1} C^\top\mathbf{X}^\top\left(\mathbf{Y} - \mathbf{X}D\right) + D$$

(c) The expression of the estimator under explicit constraints is:

$$\begin{aligned}
\tilde{\beta} &= \hat{\beta} - \left(\mathbf{X}^\top\mathbf{X}\right)^{-1} A^\top\left(A\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\right)^{-1}\left(A\hat{\beta} - B\right) \\
&= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(I - A^\top\left(A\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\right)^{-1}A\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)\cdot \\
&\quad \left(\mathbf{X}^\top\mathbf{Y}\right) + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\left(A\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\right)^{-1}B
\end{aligned}$$

whereas the expression of the estimator under implicit constraints is:

$$\begin{aligned}
\tilde{\beta} &= C\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1}C^\top\mathbf{X}^\top\left(\mathbf{Y} - \mathbf{X}D\right) + D \\
&= C\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1}C^\top\left(\mathbf{X}^\top\mathbf{Y}\right) + \\
&\quad \left(I - C\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1}C^\top\mathbf{X}^\top\mathbf{X}\right)D
\end{aligned}$$

We also have $AC = 0$ and $D = \left(A^\top A\right)^\star A^\top B$. For any positive definite matrix M, we have:

$$M^{-1}\left(I - A^\top\left(AM^{-1}A^\top\right)^{-1}AM^{-1}\right) = C\left(C^\top MC\right)^{-1}C^\top$$

and:

$$\begin{aligned}
&\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\left(A\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}A^\top\right)^{-1} = \\
&\left(I - C\left(C^\top\mathbf{X}^\top\mathbf{X}C\right)^{-1}C^\top\mathbf{X}^\top\mathbf{X}\right)\left(A^\top A\right)^\star A^\top
\end{aligned}$$

We deduce that the two estimators are equivalent.

(d) If $\beta_1 = \beta_2$ and $\beta_1 = \beta_5 + 1$, we have:

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

We deduce that the estimator under explicit constraints is:

$$\begin{aligned} \tilde{\beta} &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \left(I - A^\top \left(A \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top\right)^{-1} A \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right) \cdot \\ & \quad \left(\mathbf{X}^\top \mathbf{Y}\right) + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top \left(A \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} A^\top\right)^{-1} B \\ &= \begin{pmatrix} 0.28040 \\ 0.28040 \\ 2.08942 \\ 3.21265 \\ -0.71960 \end{pmatrix} \end{aligned}$$

We can write the explicit constraints into implicit constraints:

$$\begin{aligned} \beta &= C\gamma + D \\ &= \begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \\ 0 \\ -\frac{2}{3} \end{pmatrix} \end{aligned}$$

We deduce that:

$$\begin{aligned} \tilde{\gamma} &= \left(C^\top \mathbf{X}^\top \mathbf{X} C\right)^{-1} C^\top \left(\mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X} D\right) \\ &= \begin{pmatrix} 2.08942 \\ 3.21265 \\ -0.09168 \end{pmatrix} \end{aligned}$$

We obtain the same solution:

$$\begin{aligned} \tilde{\beta} &= C\tilde{\gamma} + D \\ &= \begin{pmatrix} 0.28040 \\ 0.28040 \\ 2.08942 \\ 3.21265 \\ -0.71960 \end{pmatrix} \end{aligned}$$

Remark 2 *The matrices C and D of the previous $\beta = C\gamma + D$ correspond to the orthonormal matrix of A and the matrix $\left(A^\top A\right)^\star A^\top B$. However, there exist many decomposition $\beta = C\gamma + D$ because the only restriction is that C is an orthogonal matrix of A. For instance, if we choose:*

$$\begin{aligned} \beta &= C\gamma + D \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

we obtain:

$$\begin{aligned}\tilde{\gamma} &= \left(C^\top \mathbf{X}^\top \mathbf{X} C\right)^{-1} C^\top \left(\mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X} D\right) \\ &= \begin{pmatrix} 0.28040 \\ 2.08942 \\ 3.21265 \end{pmatrix}\end{aligned}$$

and:

$$\begin{aligned}\tilde{\beta} &= C\tilde{\gamma} + D \\ &= \begin{pmatrix} 0.28040 \\ 0.28040 \\ 2.08942 \\ 3.21265 \\ -0.71960 \end{pmatrix}\end{aligned}$$

10.3.4 Maximum likelihood estimation of the Poisson distribution

1. We have:

$$\begin{aligned}\boldsymbol{\ell}(\lambda) &= \sum_{i=1}^{n} \ln \Pr\{Y_i = y_i\} \\ &= \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}\right) \\ &= -n\lambda + \ln(\lambda)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \ln(y_i!)\end{aligned}$$

   It follows that:

$$\begin{aligned}\frac{\partial\,\boldsymbol{\ell}(\lambda)}{\partial\,\lambda} = 0 \quad &\Leftrightarrow \quad -n + \frac{\sum_{i=1}^n y_i}{\hat{\lambda}} = 0 \\ &\Leftrightarrow \quad \hat{\lambda} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}\end{aligned}$$

2. We have:

$$\frac{\partial^2\,\boldsymbol{\ell}(\lambda)}{\partial\,\lambda^2} = -\frac{\sum_{i=1}^n y_i}{\lambda^2}$$

   We deduce that:

$$\begin{aligned}\mathcal{I}(\lambda) &= \mathbb{E}\left[\frac{\sum_{i=1}^n Y_i}{\lambda^2}\right] \\ &= \frac{n}{\lambda}\end{aligned}$$

   The variance based on the Information matrix is then:

$$\operatorname{var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}}{n}$$

   If we use the Hessian matrix, we obtain:

$$\operatorname{var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^2}{\sum_{i=1}^n y_i} = \frac{\hat{\lambda}^2}{n\bar{y}} = \frac{\hat{\lambda}}{n}$$

   We obtain the same expression.

10.3.5 Maximum likelihood estimation of the Exponential distribution

1. We have:

$$\begin{aligned} \boldsymbol{\ell}(\lambda) &= \sum_{i=1}^{n} \ln \lambda e^{-\lambda y_i} \\ &= n \ln \lambda - \lambda \sum_{i=1}^{n} y_i \end{aligned}$$

It follows that:

$$\begin{aligned} \frac{\partial \boldsymbol{\ell}(\lambda)}{\partial \lambda} = 0 \quad &\Leftrightarrow \quad \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} y_i = 0 \\ &\Leftrightarrow \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{y}} \end{aligned}$$

2. We have:

$$\frac{\partial^2 \boldsymbol{\ell}(\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2}$$

We deduce that:

$$\mathcal{I}(\lambda) = \frac{n}{\lambda^2}$$

The variance based on the Information matrix is then:

$$\operatorname{var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^2}{n}$$

It is equal to the variance based on the Hessian matrix.

10.3.6 Relationship between the linear regression, the maximum likelihood method and the method of moments

1. We have:

$$\begin{aligned} \boldsymbol{\ell}(\theta) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - x_i^\top \beta}{\sigma} \right)^2 \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{(\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)}{2\sigma^2} \end{aligned}$$

The vector of parameters θ is:

$$\theta = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$$

2. It follows that:

$$\begin{aligned} \frac{\partial \boldsymbol{\ell}(\theta)}{\partial \beta} &= \frac{2\mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\beta)}{2\sigma^2} \\ &= \frac{\mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X}\beta}{\sigma^2} \end{aligned}$$

and:

$$\frac{\partial \boldsymbol{\ell}(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)}{2\sigma^4}$$

We deduce that:

$$\begin{aligned}\frac{\partial \boldsymbol{\ell}(\theta)}{\partial \beta}=0 &\Leftrightarrow \mathbf{X}^{\top}\mathbf{Y}-\mathbf{X}^{\top}\mathbf{X}\hat{\beta}=0 \\ &\Leftrightarrow \hat{\beta}=\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{Y}\end{aligned}$$

and:

$$\begin{aligned}\frac{\partial \boldsymbol{\ell}(\theta)}{\partial \sigma^2}=0 &\Leftrightarrow -\frac{n}{2\hat{\sigma}^2}+\frac{\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)^{\top}\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)}{2\hat{\sigma}^4}=0 \\ &\Leftrightarrow \hat{\sigma}^2=\frac{\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)^{\top}\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)}{n}\end{aligned}$$

We verify that $\hat{\beta}_{\mathrm{ML}}=\hat{\beta}_{\mathrm{OLS}}$ and $\hat{\sigma}^2_{\mathrm{ML}}<\hat{\sigma}^2_{\mathrm{OLS}}$ because:

$$\hat{\sigma}^2_{\mathrm{OLS}}=\frac{\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)^{\top}\left(\mathbf{Y}-\mathbf{X}\hat{\beta}\right)}{n-K}$$

3. We have:

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial \beta\,\partial \beta^{\top}} &= -\frac{\mathbf{X}^{\top}\mathbf{X}}{\sigma^2} \\ \frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial \beta\,\partial \sigma^2} &= -\frac{\mathbf{X}^{\top}\left(\mathbf{Y}-\mathbf{X}\beta\right)}{\sigma^4} \\ &= -\frac{\mathbf{X}^{\top}\mathbf{U}}{\sigma^4} \\ \frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial \sigma^2\,\partial \sigma^2} &= \frac{n}{2\sigma^4}-\frac{\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\top}\left(\mathbf{Y}-\mathbf{X}\beta\right)}{\sigma^6} \\ &= \frac{n}{2\sigma^4}-\frac{\mathbf{U}^{\top}\mathbf{U}}{\sigma^6}\end{aligned}$$

It follows that:

$$H(\theta)=\left(\begin{array}{cc}-\mathbf{X}^{\top}\mathbf{X}/\sigma^2 & -\mathbf{X}^{\top}\mathbf{U}/\sigma^4 \\ -\mathbf{X}^{\top}\mathbf{U}/\sigma^4 & n/\left(2\sigma^4\right)-\mathbf{U}^{\top}\mathbf{U}/\sigma^6\end{array}\right)$$

and:

$$\begin{aligned}I(\theta) &= -\mathbb{E}\left[H(\theta)\right] \\ &= \left(\begin{array}{cc}\mathbf{X}^{\top}\mathbf{X}/\sigma^2 & \mathbb{E}\left[\mathbf{X}^{\top}\mathbf{U}\right]/\sigma^4 \\ \mathbb{E}\left[\mathbf{X}^{\top}\mathbf{U}\right]/\sigma^4 & \mathbb{E}\left[\mathbf{U}^{\top}\mathbf{U}\right]/\sigma^6-\frac{n}{2}/\sigma^4\end{array}\right) \\ &= \left(\begin{array}{cc}\mathbf{X}^{\top}\mathbf{X}/\sigma^2 & 0 \\ 0 & \frac{n}{2}/\sigma^4\end{array}\right)\end{aligned}$$

because we have $\mathbb{E}\left[\mathbf{X}^{\top}\mathbf{U}\right]=0$ and $\mathbb{E}\left[\mathbf{U}^{\top}\mathbf{U}\right]=\mathbb{E}\left[\sum_{i=1}^{n}u_i^2\right]=n\sigma^2$. We deduce that:

$$\begin{aligned}\operatorname{var}(\theta) &= I(\theta)^{-1} \\ &= \left(\begin{array}{cc}\sigma^2\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} & 0 \\ 0 & 2\sigma^4/n\end{array}\right)\end{aligned}$$

Finally, we obtain:

$$\text{var}\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

and:

$$\text{var}\left(\hat{\sigma}^2\right) = \frac{2\hat{\sigma}^4}{n}$$

We notice that the expressions of $\text{var}\left(\hat{\beta}_{\text{ML}}\right)$ and $\text{var}\left(\hat{\beta}_{\text{OLS}}\right)$ are similar, but they do not use the same standard deviation of residuals $\hat{\sigma}$.

10.3.7 The Gaussian mixture model

1. We can write Y as follows:

$$Y = BY_1 + (1 - B) Y_2$$

where B is a Bernoulli random variable independent from Y_1 and Y_2, and whose parameter is π_1. We have:

$$\begin{aligned}
\mathbb{E}\left[Y^k\right] &= \mathbb{E}\left[\left(BY_1 + (1 - B) Y_2\right)^k\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{k} \binom{k}{i} \left(BY_1\right)^{k-i} \left((1 - B) Y_2\right)^i\right] \\
&= \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}\left[\left(BY_1\right)^{k-i} \left((1 - B) Y_2\right)^i\right]
\end{aligned}$$

Since Y_1 and Y_2 are independent, we have $\mathbb{E}\left[\left(BY_1\right)^{k-i} \left((1 - B) Y_2\right)^i\right]$ when $i \neq 0$ or $i \neq k$. It follows that:

$$\begin{aligned}
\mathbb{E}\left[Y^k\right] &= \mathbb{E}\left[B^k Y_1^k\right] + \mathbb{E}\left[(1 - B)^k Y_2^k\right] \\
&= \mathbb{E}\left[B^k\right] \mathbb{E}\left[Y_1^k\right] + \mathbb{E}\left[(1 - B)^k\right] \mathbb{E}\left[Y_2^k\right] \\
&= \pi_1 \mathbb{E}\left[Y_1^k\right] + \pi_2 \mathbb{E}\left[Y_2^k\right]
\end{aligned}$$

because B is independent from Y_1 and Y_2, $B^k \sim \mathcal{B}(\pi_1)$ and $(1 - B)^k \sim \mathcal{B}(\pi_2)$.

2. We deduce that:

$$\begin{aligned}
\mathbb{E}[Y] &= \pi_1 \mathbb{E}[Y_1] + \pi_2 \mathbb{E}[Y_2] \\
&= \pi_1 \mu_1 + \pi_2 \mu_2
\end{aligned}$$

and:

$$\begin{aligned}
\text{var}(Y) &= \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y] \\
&= \pi_1 \mathbb{E}\left[Y_1^2\right] + \pi_2 \mathbb{E}\left[Y_2^2\right] - \mathbb{E}^2[Y]
\end{aligned}$$

Since we know that $\mathbb{E}\left[Y_i^2\right] = \mu_i^2 + \sigma_i^2$, we obtain:

$$\begin{aligned}
\text{var}(Y) &= \pi_1 \left(\mu_1^2 + \sigma_1^2\right) + \pi_2 \left(\mu_2^2 + \sigma_2^2\right) - \left(\pi_1 \mu_1 + \pi_2 \mu_2\right)^2 \\
&= \pi_1 \sigma_1^2 + \pi_2 \sigma_2^2 + \pi_1 (1 - \pi_1) \left(\mu_1^2 + \mu_2^2\right) - 2\pi_1 \pi_2 \mu_1 \mu_2 \\
&= \pi_1 \sigma_1^2 + \pi_2 \sigma_2^2 + \pi_1 \pi_2 \left(\mu_1 - \mu_2\right)^2
\end{aligned}$$

because $\pi_2 = 1 - \pi_1$.

3. We remind that $\mathbb{E}\left[Y_i^3\right] = \mu_i^3 + 3\mu_i\sigma_i^2$. It follows that:

$$\begin{aligned}
\mathbb{E}\left[(Y - \mathbb{E}[Y])^3\right] &= \mathbb{E}\left[Y^3\right] - 3\mathbb{E}[Y]\operatorname{var}(Y) - \mathbb{E}^3[Y] \\
&= \pi_1\left(\mu_1^3 + 3\mu_1\sigma_1^2\right) + \pi_2\left(\mu_2^3 + 3\mu_2\sigma_2^2\right) - \\
&\quad 3\left(\pi_1\mu_1 + \pi_2\mu_2\right)\left(\pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2\left(\mu_1 - \mu_2\right)^2\right) - \\
&\quad \left(\pi_1\mu_1 + \pi_2\mu_2\right)^3 \\
&= \pi_1\pi_2\left(\pi_2 - \pi_1\right)\left(\mu_1 - \mu_2\right)^3 + 3\pi_1\pi_2\left(\mu_1 - \mu_2\right)\left(\sigma_1^2 - \sigma_2^2\right)
\end{aligned}$$

We deduce that the skewness coefficient is equal to:

$$\gamma_1(Y) = \frac{\pi_1\pi_2\left(\left(\pi_2 - \pi_1\right)\left(\mu_1 - \mu_2\right)^3 + 3\left(\mu_1 - \mu_2\right)\left(\sigma_1^2 - \sigma_2^2\right)\right)}{\left(\pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2\left(\mu_1 - \mu_2\right)^2\right)^{3/2}}$$

10.3.8 Parameter estimation of diffusion processes

1. The solution is:

$$X(t) = X(s)\, e^{\left(\mu - \frac{1}{2}\sigma^2\right)(t-s) + \sigma(W(t) - W(s))}$$

It follows that:

$$\ln X(t) - \ln X(s) = \left(\mu - \frac{1}{2}\sigma^2\right)(t - s) + \sigma\left(W(t) - W(s)\right)$$

Since $W(t) - W(s) \sim \mathcal{N}(0, t - s)$, we deduce that the log-likelihood function of the sample $\mathbf{X}$ is:

$$\begin{aligned}
\boldsymbol{\ell}(\mu, \sigma) &= -\frac{1}{2}\sum_{i=1}^{T}\left(\ln 2\pi + \ln\left(\sigma^2\Delta t_i\right) + \frac{\varepsilon_i^2}{\sigma^2\Delta t_i}\right) \\
&= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{T}\ln\Delta t_i - \frac{1}{2}\sum_{i=1}^{T}\frac{\varepsilon_i^2}{\sigma^2\Delta t_i}
\end{aligned}$$

where $\Delta t_i = t_i - t_{i-1}$ and ε_i is the innovation process:

$$\varepsilon_i = \ln x_i - \ln x_{i-1} - \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t_i$$

2. The solution is:

$$X(t) = X(s)\, e^{-a(t-s)} + b\left(1 - e^{-a(t-s)}\right) + \sigma\int_s^t e^{-a(t-u)}\,\mathrm{d}W(u)$$

where:

$$\int_s^t e^{-a(t-u)}\,\mathrm{d}W(u) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2a}\left(1 - e^{-2a(t-s)}\right)\right)$$

We deduce that:

$$\begin{aligned}\boldsymbol{\ell}(a,b,\sigma) &= -\frac{1}{2}\sum_{i=1}^{T}\left(\ln 2\pi + \ln\left(\frac{\sigma^2}{2a}\left(1-e^{-2a\Delta t_i}\right)\right)\right) - \\ &\quad \frac{1}{2}\sum_{i=1}^{T}\frac{2a\varepsilon_i^2}{\sigma^2\left(1-e^{-2a\Delta t_i}\right)} \\ &= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\frac{\sigma^2}{2a} - \frac{1}{2}\sum_{i=1}^{T}\ln\left(1-e^{-2a\Delta t_i}\right) - \\ &\quad \frac{a\varepsilon_i^2}{\sigma^2\left(1-e^{-2a\Delta t_i}\right)}\end{aligned}$$

where:

$$\varepsilon_i = x_i - x_{i-1}e^{-a\Delta t_i} - b\left(1-e^{-a\Delta t_i}\right)$$

3. We have:

$$X(t) - X(s) \approx \mu(s, X(s))(t-s) + \sigma(s, X(s))(W(t) - W(s))$$

We deduce that:

$$\boldsymbol{\ell}(\theta) = -\frac{1}{2}\sum_{i=1}^{T}\left(\ln 2\pi + \ln\left(\sigma^2(t_{i-1}, x_{i-1})\Delta t_i\right) + \frac{\varepsilon_i^2}{\sigma^2(t_{i-1}, x_{i-1})\Delta t_i}\right)$$

where:$\varepsilon_i = x_i - x_{i-1} - \mu(t_{i-1}, x_{i-1})\Delta t_i$. In the case of the CIR process, we obtain:

$$\begin{aligned}\boldsymbol{\ell}(a,b,\sigma) &= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{T}\ln\left(x_{i-1}\Delta t_i\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{T}\frac{\varepsilon_i^2}{\sigma^2 x_{i-1}\Delta t_i}\end{aligned} \tag{10.1}$$

where:

$$\varepsilon_i = x_i - x_{i-1} - a(b - x_{i-1})\Delta t_i$$

We assume that $X(s,t) = X(t) \mid X(s)$ is normally-distributed $\mathcal{N}(m_1(s,t), m_2(s,t))$ where $m_1(s,t) = \mathbb{E}[X(s,t)]$ and $m_2(s,t) = \mathbb{E}\left[(X(s,t) - m_1(s,t))^2\right]$.Then, we have:

$$\boldsymbol{\ell}(\theta) = -\frac{1}{2}\sum_{i=1}^{T}\left(\ln 2\pi + \ln m_2(t_{i-1}, t_i) + \frac{(x_i - m_1(t_{i-1}, t_i))^2}{m_2(t_{i-1}, t_i)}\right)$$

In the case of the CIR process, we have:

$$m_1(t_{i-1}, t_i) = x_{i-1}e^{-a\Delta t_i} + b\left(1-e^{-a\Delta t_i}\right)$$

and:

$$m_2(t_{i-1}, t_i) = \sigma^2\left(x_{i-1}\frac{\left(e^{-a\Delta t_i} - e^{-2a\Delta t_i}\right)}{a} + b\frac{\left(1-e^{-a\Delta t_i}\right)^2}{2a}\right)$$

We deduce that:

$$\begin{aligned}\boldsymbol{\ell}(a,b,\sigma) &= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \\ &\quad \frac{1}{2}\sum_{i=1}^{T}\ln\left(x_{i-1}\frac{\left(e^{-a\Delta t_i}-e^{-2a\Delta t_i}\right)}{a}+b\frac{\left(1-e^{-a\Delta t_i}\right)^2}{2a}\right) - \\ &\quad \frac{1}{2}\sum_{i=1}^{T}\frac{\varepsilon_i^2}{\sigma^2 x_{i-1}\Delta t_i} \end{aligned} \tag{10.2}$$

where:

$$\varepsilon_i = x_i - x_{i-1}e^{-a\Delta t_i} - b\left(1-e^{-a\Delta t_i}\right)$$

When $\Delta t_i \to 0$, we have $e^{-a\Delta t_i} \approx 1 - a\Delta t_i$ and $e^{-2a\Delta t_i} \approx 1 - 2a\Delta t_i$. It follows that:

$$\begin{aligned}\frac{\left(e^{-a\Delta t_i}-e^{-2a\Delta t_i}\right)}{a} &\approx \frac{(1-a\Delta t_i)-(1-2a\Delta t_i)}{a} \\ &\approx \Delta t_i\end{aligned}$$

and:

$$\begin{aligned}\frac{\left(1-e^{-a\Delta t_i}\right)^2}{2a} &\approx \frac{1-2(1-a\Delta t_i)+(1-2a\Delta t_i)}{2a} \\ &\approx 0\end{aligned}$$

We deduce that $m_1(t_{i-1},t_i) \approx x_{i-1}(1-a\Delta t_i) + ab\Delta t_i$, $m_2(t_{i-1},t_i) \approx \sigma^2 x_{i-1}\Delta t_i$ and

$$\begin{aligned}\varepsilon_i &\approx x_i - x_{i-1}(1-a\Delta t_i) - ab\Delta t_i \\ &\approx x_i - x_{i-1} - a(b-x_{i-1})\Delta t_i\end{aligned}$$

We conclude that the log-likelihood functions (10.1) and (10.2) converge to the same expression when $\Delta t_i \to 0$.

4. For the geometric Brownian motion, we have $\mathbb{E}_{t_{i-1}}[\varepsilon_i] = 0$ and $\mathbb{E}_{t_{i-1}}\left[\varepsilon_i^2 - \sigma^2\Delta t_i\right] = 0$. We deduce that:

$$\left\{\begin{array}{l} h_{i,1}(\mu,\sigma) = \ln x_i - \ln x_{i-1} - \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t_i \\ h_{i,2}(\mu,\sigma) = \left(\ln x_i - \ln x_{i-1} - \left(\mu-\frac{1}{2}\sigma^2\right)\Delta t_i\right)^2 - \sigma^2\Delta t_i \end{array}\right.$$

For the Ornstein-Uhlenbeck process, we can use the same two moment conditions and the orthogonal condition $\mathbb{E}_{t_{i-1}}[\varepsilon_i x_{i-1}] = 0$. Finally, we obtain:

$$\left\{\begin{array}{l} h_{i,1}(\theta) = x_i - x_{i-1}e^{-a\Delta t_i} - b\left(1-e^{-a\Delta t_i}\right) \\ h_{i,2}(\theta) = \left(x_t - x_{t-1}e^{-a\Delta t_i} - b\left(1-e^{-a\Delta t_i}\right)\right)^2 - \sigma^2\left(\frac{1-e^{-2a\Delta t_i}}{2a}\right) \\ h_{i,3}(\theta) = \left(x_i - x_{i-1}e^{-a\Delta t_i} - b\left(1-e^{-a\Delta t_i}\right)\right)x_{i-1} \end{array}\right.$$

For the CIR process, we proceed as for the OU process:

$$\left\{\begin{array}{l} h_{i,1}(\theta) = x_i - m_1(t_{i-1},t_i) \\ h_{i,2}(\theta) = (x_i - m_1(t_{i-1},t_i))^2 - m_2(t_{i-1},t_i) \\ h_{i,3}(\theta) = (x_i - m_1(t_{i-1},t_i))\,x_{i-1} \end{array}\right.$$

where:

$$m_1(t_{i-1},t_i) = x_{i-1}e^{-a\Delta t_i} + b\left(1-e^{-a\Delta t_i}\right)$$

and:

$$m_2(t_{i-1},t_i) = \sigma^2\left(x_{i-1}\frac{\left(e^{-a\Delta t_i}-e^{-2a\Delta t_i}\right)}{a} + b\frac{\left(1-e^{-a\Delta t_i}\right)^2}{2a}\right)$$

5. If we use the Euler-Maruyama scheme:

$$X(t) - X(s) \approx a(b - X(s))(t - s) + \sigma |X(s)|^{\gamma} (W(t) - W(s))$$

we obtain:

$$\begin{cases} h_{i,1}(\theta) = x_i - m_1(t_{i-1}, t_i) \\ h_{i,2}(\theta) = (x_i - m_1(t_{i-1}, t_i))^2 - m_2(t_{i-1}, t_i) \\ h_{i,3}(\theta) = (x_i - m_1(t_{i-1}, t_i)) x_{i-1} \\ h_{i,4}(\theta) = \left((x_i - m_1(t_{i-1}, t_i))^2 - m_2(t_{i-1}, t_i)\right) x_{i-1} \end{cases}$$

where:

$$m_1(t_{i-1}, t_i) = x_{i-1} + a(b - x_{i-1}) \Delta t_i$$

and:

$$m_2(t_{i-1}, t_i) = \sigma^2 |x_{i-1}|^{2\gamma} \Delta t_i$$

10.3.9 The Tobit model

1. We note $\tilde{X} = X \mid X \geq c$ the truncated random variable. The probability density function of $\tilde{X}$ is equal to:

$$f(x) = \frac{1}{\sigma(1 - \Phi(\alpha))} \phi\left(\frac{x - \mu}{\sigma}\right)$$

where $\alpha = \sigma^{-1}(c - \mu)$. We have[3]:

$$\begin{aligned} \mathbb{E}\left[\tilde{X}\right] &= \frac{1}{1 - \Phi(\alpha)} \int_c^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right) \mathrm{d}x \\ &= \frac{1}{1 - \Phi(\alpha)} \int_{\alpha}^{\infty} (\mu + \sigma y) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \\ &= \frac{1}{1 - \Phi(\alpha)} \left(\mu \int_{\alpha}^{\infty} \phi(y) \, \mathrm{d}y + \sigma\left[-\phi(y)\right]_{\alpha}^{\infty}\right) \\ &= \mu + \sigma\lambda(\alpha) \end{aligned} \tag{10.3}$$

where $\lambda(\alpha)$ is the inverse Mills ratio:

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} = \frac{\phi(-\alpha)}{\Phi(-\alpha)}$$

[3]We use the change of variable $y = \sigma^{-1}(x - \mu)$.

We have:

$$\begin{aligned}
\mathbb{E}\left[\tilde{X}^2\right] &= \frac{1}{1-\Phi(\alpha)}\int_c^\infty x^2 \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\,\mathrm{d}x \\
&= \frac{1}{1-\Phi(\alpha)}\int_\alpha^\infty (\mu+\sigma y)^2\,\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)\,\mathrm{d}y \\
&= \frac{1}{1-\Phi(\alpha)}\left(\mu^2\int_\alpha^\infty \phi(y)\,\mathrm{d}y + 2\mu\sigma\left[-\phi(y)\right]_\alpha^\infty + \right. \\
&\quad \left. \sigma^2\int_\alpha^\infty y^2\phi(y)\,\mathrm{d}y\right) \\
&= \mu^2 + 2\mu\sigma\lambda(\alpha) + \frac{\sigma^2}{1-\Phi(\alpha)}\left(\left[-y\phi(y)\right]_\alpha^\infty + \int_\alpha^\infty \phi(y)\,\mathrm{d}y\right) \\
&= \mu^2 + 2\mu\sigma\lambda(\alpha) + \sigma^2\left(1+\alpha\lambda(\alpha)\right)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\operatorname{var}\left(\tilde{X}\right) &= \mathbb{E}\left[\tilde{X}^2\right] - \mathbb{E}^2\left[\tilde{X}\right] \\
&= \mu^2 + 2\mu\sigma\lambda(\alpha) + \\
&\quad \sigma^2\left(1+\alpha\lambda(\alpha)\right) - \mu^2 - 2\mu\sigma\lambda(\alpha) - \sigma^2\lambda^2(\alpha) \\
&= \sigma^2\left(1-\delta(\alpha)\right)
\end{aligned}$$

where:

$$\delta(\alpha) = \lambda(\alpha)\left(\lambda(\alpha)-\alpha\right)$$

We can show that truncation reduces variance because we have $0 \leq \delta(\alpha) \leq 1$.

2. The censured random variable $\tilde{Y}$ can be written as follows:

$$\tilde{Y} = \begin{cases} X & \text{if} \quad X \geq c \\ c & \text{if} \quad X < c \end{cases}$$

We have:

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}\right] &= \Pr\{Y=c\}\,\mathbb{E}\left[\tilde{Y}\mid X<c\right] + \Pr\{Y\neq c\}\,\mathbb{E}\left[X\mid X\geq c\right] \\
&= \Pr\{X<c\}\,c + \Pr\{X\geq c\}\,\mathbb{E}\left[\tilde{X}\right] \\
&= \Phi(\alpha)\,c + \left(1-\Phi(\alpha)\right)\left(\mu+\sigma\lambda(\alpha)\right)
\end{aligned}$$

We also have:

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}^2\right] &= \Pr\{Y=c\}\,\mathbb{E}\left[\tilde{Y}^2\mid X<c\right] + \Pr\{Y\neq c\}\,\mathbb{E}\left[X^2\mid X\geq c\right] \\
&= \Phi(\alpha)\,c^2 + \left(1-\Phi(\alpha)\right)\mathbb{E}\left[\tilde{X}^2\right] \\
&= \Phi(\alpha)\,c^2 + \left(1-\Phi(\alpha)\right)\left(\mu^2 + 2\mu\sigma\lambda(\alpha) + \sigma^2\left(1+\alpha\lambda(\alpha)\right)\right)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\operatorname{var}\left(\tilde{Y}\right) &= \mathbb{E}\left[\tilde{Y}^2\right] - \mathbb{E}^2\left[\tilde{Y}\right] \\
&= \Phi(\alpha) c^2 + (1-\Phi(\alpha))\left(\mu^2 + 2\mu\sigma\lambda(\alpha) + \sigma^2(1+\alpha\lambda(\alpha))\right) - \\
&\quad \Phi^2(\alpha) c^2 - 2\Phi(\alpha)(1-\Phi(\alpha))(\mu c + \sigma c\lambda(\alpha)) - \\
&\quad (1-\Phi(\alpha))^2\left(\mu^2 + 2\mu\sigma\lambda(\alpha) + \sigma^2\lambda^2(\alpha)\right) \\
&= \Phi(\alpha)(1-\Phi(\alpha)) c^2 + \Phi(\alpha)(1-\Phi(\alpha))\mu^2 + \\
&\quad 2\Phi(\alpha)\phi(\alpha)\mu\sigma + (\phi(\alpha)(\alpha-\phi(\alpha)) + 1 - \Phi(\alpha))\sigma^2 - \\
&\quad 2\Phi(\alpha)(1-\Phi(\alpha))\mu c - 2\Phi(\alpha)\phi(\alpha)\sigma c \\
&= \Phi(\alpha)(1-\Phi(\alpha))(c-\mu)^2 - 2\Phi(\alpha)\phi(\alpha)(c-\mu)\sigma + \\
&\quad (1-\Phi(\alpha))(1+\lambda(\alpha)(\alpha-\phi(\alpha)))\sigma^2 \\
&= \Phi(\alpha)(1-\Phi(\alpha))\alpha^2\sigma^2 - 2\Phi(\alpha)\phi(\alpha)\alpha\sigma^2 + \\
&\quad (1-\Phi(\alpha))\left(1-\delta(\alpha) - \lambda(\alpha)\phi(\alpha) + \lambda^2(\alpha)\right)\sigma^2 \\
&= \sigma^2(1-\Phi(\alpha))\left(\Phi(\alpha)\alpha^2 - 2\Phi(\alpha)\lambda(\alpha)\alpha + \right. \\
&\quad \left. 1-\delta(\alpha) - \lambda(\alpha)\phi(\alpha) + \lambda^2(\alpha)\right) \\
&= \sigma^2(1-\Phi(\alpha))\Big((1-\delta(\alpha)) + (\alpha-\lambda(\alpha))^2\Phi(\alpha)\Big)
\end{aligned}$$

because we have:

$$-\Phi(\alpha)\lambda^2(\alpha) - \lambda(\alpha)\phi(\alpha) + \lambda^2(\alpha) = 0$$

3. In Figures 10.1 and 10.2, we have reported the corresponding probability density function of the truncated random variable $\tilde{X}$ and the censored random variable $\tilde{Y}$. We obtain $\mathbb{E}\left[\tilde{X}\right] = 3.7955$, $\mathbb{E}\left[\tilde{X}^2\right] = 18.3864$, $\sigma\left(\tilde{X}\right) = 1.9952$, $\mathbb{E}\left[\tilde{Y}\right] = 2.7627$, $\mathbb{E}\left[\tilde{Y}^2\right] = 11.9632$ and $\sigma\left(\tilde{Y}\right) = 2.0810$. We verify that truncation reduces variance: $\sigma\left(\tilde{X}\right) \leq \sigma(X)$. In the case of truncation, some observations are excluded, implying that we observe only a part of the probability density function. In the case od censoring, the probability density function is a mixture of continuous and discrete distributions. In particular, we observe a probability mass at the censoring point $X = c$.

4. We have:

$$\begin{aligned}
\Pr\{Y=0\} &= \Pr\{Y^\star \leq 0\} \\
&= \Pr\left\{x^\top\beta + U \leq 0\right\} \\
&= \Pr\left\{U \leq -x^\top\beta\right\} \\
&= \Phi\left(-\frac{x^\top\beta}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{x^\top\beta}{\sigma}\right)
\end{aligned}$$

We deduce that the log-likelihood function is equal to:

$$\begin{aligned}
\boldsymbol{\ell}(\theta) &= \sum_{i=1}^n (1-d_i)\ln\left(1-\Phi\left(\frac{x_i^\top\beta}{\sigma}\right)\right) - \\
&\quad \frac{1}{2}\sum_{i=1}^n d_i\left(\ln 2\pi + \ln\sigma^2 + \left(\frac{y_i - x_i^\top\beta}{\sigma}\right)^2\right)
\end{aligned}$$

where d_i is a dummy variable that is equal to 1 if $y_i > 0$.

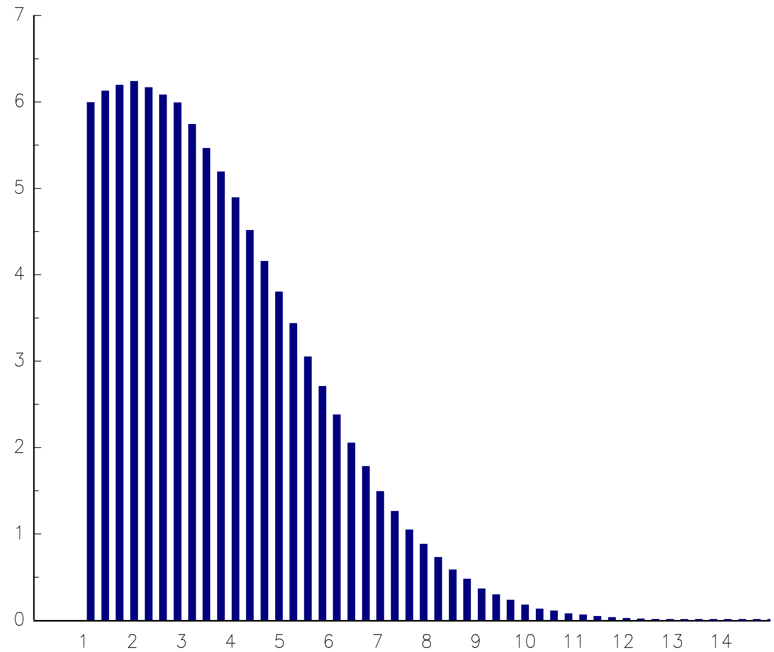

FIGURE 10.1: Frequency (in %) of the truncated random variable

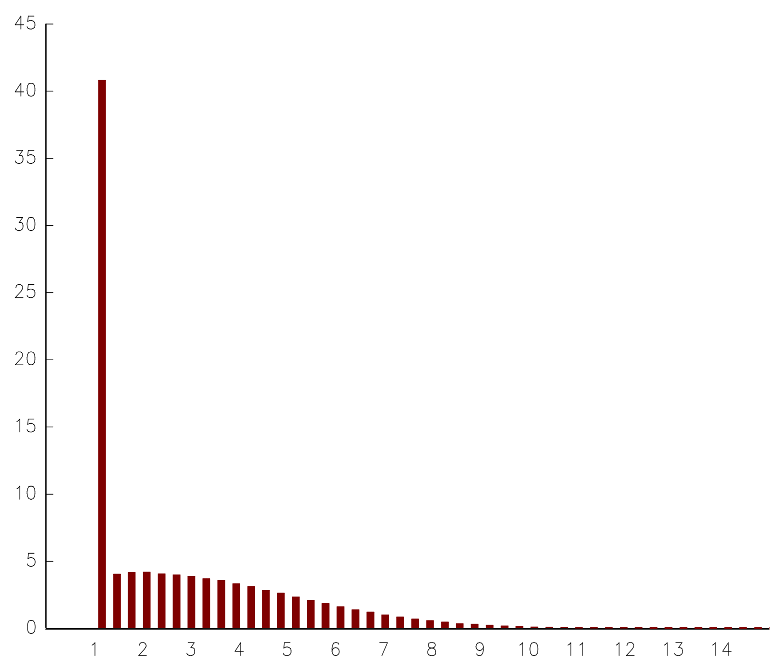

FIGURE 10.2: Frequency (in %) of the censored random variable

5. We have:

$$\frac{\partial}{\partial x}\ln\left(1-\Phi\left(f\left(x\right)\right)\right)=-\frac{\phi\left(f\left(x\right)\right)}{1-\Phi\left(f\left(x\right)\right)}f'\left(x\right)$$

and:

$$\frac{\partial}{\partial \beta}\left(\frac{x_i^\top\beta}{\sigma}\right)=\frac{x_i}{\sigma}$$

We also have:

$$\frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma}\right)=-\frac{1}{2\sigma^3}$$

We deduce that the ML estimator satisfies the following first-order conditions:

$$\frac{\partial\,\boldsymbol{\ell}\left(\theta\right)}{\partial\,\beta}=-\frac{1}{\sigma}\sum_{d_i=0}\left(\frac{\phi_i}{1-\Phi_i}\right)x_i+\frac{1}{\sigma^2}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)x_i=0 \tag{10.4}$$

and:

$$\frac{\partial\,\boldsymbol{\ell}\left(\theta\right)}{\partial\,\sigma^2}=\frac{1}{2\sigma^3}\sum_{d_i=0}\left(\frac{\phi_i}{1-\Phi_i}\right)x_i^\top\beta+\frac{1}{2\sigma^4}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)^2-\frac{n_1}{2\sigma^2}=0 \tag{10.5}$$

where $\phi_i=\phi\left(\frac{x_i^\top\beta}{\sigma}\right)$, $\Phi_i=\Phi\left(\frac{x_i^\top\beta}{\sigma}\right)$ and $n_1=\sum_{i=1}^n d_i$.

6. Since $\partial_x\phi\left(x\right)=-x\phi\left(x\right)$, we have:

$$\frac{\partial}{\partial x}\left(\frac{\phi\left(f\left(x\right)\right)}{1-\Phi\left(f\left(x\right)\right)}\right)=\frac{\phi\left(f\left(x\right)\right)\left(\phi\left(f\left(x\right)\right)-f\left(x\right)\left(1-\Phi\left(f\left(x\right)\right)\right)\right)f'\left(x\right)}{\left(1-\Phi\left(f\left(x\right)\right)\right)^2}$$

It follows that:

$$\frac{\partial}{\partial\,\beta^\top}\left(\frac{\phi_i}{1-\Phi_i}\right)=\frac{\phi_i}{\left(1-\Phi_i\right)^2}\left(\phi_i-\left(1-\Phi_i\right)\left(\frac{x_i^\top\beta}{\sigma}\right)\right)\frac{x_i}{\sigma}$$

and:

$$\frac{\partial}{\partial\,\sigma^2}\left(\frac{\phi_i}{1-\Phi_i}\right)=-\frac{1}{2\sigma^3}\frac{\phi_i}{\left(1-\Phi_i\right)^2}\left(\phi_i-\left(1-\Phi_i\right)\left(\frac{x_i^\top\beta}{\sigma}\right)\right)x_i^\top\beta$$

For the Hessian matrix, we obtain:

$$\begin{aligned}\frac{\partial^2\,\boldsymbol{\ell}\left(\theta\right)}{\partial\,\beta\,\partial\,\beta^\top} &= -\frac{1}{\sigma^2}\sum_{d_i=0}\frac{\phi_i}{\left(1-\Phi_i\right)^2}\left(\phi_i-\left(1-\Phi_i\right)\left(\frac{x_i^\top\beta}{\sigma}\right)\right)x_ix_i^\top- \\ &\quad \frac{1}{\sigma^2}\sum_{d_i=1}x_ix_i^\top\end{aligned}$$

and:

$$\begin{aligned}\frac{\partial^2\,\boldsymbol{\ell}\left(\theta\right)}{\partial\,\beta\,\partial\,\sigma^2} &= \frac{1}{2\sigma^3}\left(\sum_{d_i=0}\frac{\phi_i}{\left(1-\Phi_i\right)^2}\left(1-\Phi_i\right)+\right. \\ &\quad \left.\sum_{d_i=0}\frac{\phi_i}{\left(1-\Phi_i\right)^2}\left(\phi_i\left(\frac{x_i^\top\beta}{\sigma}\right)-\left(1-\Phi_i\right)\left(\frac{x_i^\top\beta}{\sigma}\right)^2\right)\right)x_i- \\ &\quad \frac{1}{\sigma^4}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)x_i\end{aligned}$$

We also have:

$$\frac{\partial}{\partial\,\sigma^2}\left(\frac{1}{\sigma^3}\right)=-\frac{3}{2\sigma^5}$$

and:

$$\begin{aligned}\frac{\partial^2\,\boldsymbol{\ell}(\theta)}{\partial\,\sigma^2\,\partial\,\sigma^2} &= -\frac{1}{4\sigma^4}\sum_{d_i=0}\frac{\phi_i}{(1-\Phi_i)^2}\left(3\,(1-\Phi_i)\left(\frac{x_i^\top\beta}{\sigma}\right)+\right. \\ &\quad \left.\phi_i\left(\frac{x_i^\top\beta}{\sigma}\right)^2-(1-\Phi_i)\left(\frac{x_i^\top\beta}{\sigma}\right)^3\right)- \\ &\quad \frac{1}{\sigma^6}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)^2+\frac{n_1}{2\sigma^4}\end{aligned}$$

7. By multiplying the system of equations (10.4) by $\beta^\top/\left(2\sigma^2\right)$, we obtain:

$$-\frac{1}{2\sigma^3}\sum_{d_i=0}\left(\frac{\phi_i}{1-\Phi_i}\right)x_i^\top\beta+\frac{1}{2\sigma^4}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)x_i^\top\beta=0$$

Combining this result with Equation (10.5) gives:

$$\frac{1}{2\sigma^4}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)x_i^\top\beta+\frac{1}{2\sigma^4}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)^2-\frac{n_1}{2\sigma^2}=0$$

We deduce that:

$$\begin{aligned}\sigma^2 &= \frac{1}{n_1}\sum_{d_i=1}\left(\left(y_i-x_i^\top\beta\right)x_i^\top\beta+\left(y_i-x_i^\top\beta\right)^2\right) \\ &= \frac{1}{n_1}\sum_{d_i=1}\left(y_i-x_i^\top\beta\right)y_i\end{aligned}$$

Let D_i be the Bernouilli random variable such that:

$$\begin{aligned}\Pr\{D_i=1\} &= \Pr\{Y_i^\star>0\} \\ &= \Pr\left\{U_i\geq -x_i^\top\beta\right\} \\ &= \Phi\left(\frac{x_i^\top\beta}{\sigma}\right)\end{aligned}$$

Let Ω_i be a random variable that is independent from D_i. We have:

$$\mathbb{E}\left[\sum_{d_i=0}\Omega_i\right]=\mathbb{E}\left[\sum_{i=1}^n(1-D_i)\,\Omega_i\right]=\sum_{i=1}^n(1-\Phi_i)\,\mathbb{E}\left[\Omega_i\right]$$

and:

$$\mathbb{E}\left[\sum_{d_i=1}\Omega_i\right]=\mathbb{E}\left[\sum_{i=1}^n D_i\Omega_i\right]=\sum_{i=1}^n\Phi_i\mathbb{E}\left[\Omega_i\right]$$

By introducting the notation:

$$z_i=\frac{x_i^\top\beta}{\sigma}$$

we obtain:

$$\begin{aligned}\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial\beta\,\partial\beta^\top}\right] &= -\frac{1}{\sigma^2}\sum_{i=1}^n \frac{(1-\Phi_i)\,\phi_i}{(1-\Phi_i)^2}\left(\phi_i-(1-\Phi_i)\,z_i\right)x_i x_i^\top - \\ &\quad \frac{1}{\sigma^2}\sum_{i=1}^n \Phi_i x_i x_i^\top \\ &= -\sum_{i=1}^n a_i x_i x_i^\top\end{aligned}$$

where:

$$a_i = -\frac{1}{\sigma^2}\left(\phi_i z_i - \frac{\phi_i^2}{1-\Phi_i} - \Phi_i\right)$$

Using Equation (10.3), we have:

$$\begin{aligned}\mathbb{E}\left[D_i\left(y_i - x_i^\top\beta\right)\right] &= \mathbb{E}\left[D_i U_i\right] \\ &= \mathbb{E}\left[D_i\right]\sigma\frac{\phi_i}{\Phi_i} \\ &= \sigma\phi_i\end{aligned}$$

It follows that:

$$\mathbb{E}\left[\frac{1}{\sigma^4}\sum_{d_i=1}\left(y_i - x_i^\top\beta\right)x_i\right] = \frac{1}{\sigma^3}\sum_{i=1}^n \phi_i x_i$$

and:

$$\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial\beta\,\partial\sigma^2}\right] = -\sum_{i=1}^n b_i x_i$$

where:

$$\begin{aligned}b_i &= \frac{1}{2\sigma^3}\left(-\phi_i - \frac{\phi_i^2 z_i}{1-\Phi_i} + \phi_i z_i^2\right) + \frac{1}{\sigma^4}\sigma\phi_i \\ &= \frac{1}{2\sigma^3}\left(\phi_i z_i^2 + \phi_i - \frac{\phi_i^2 z_i}{1-\Phi_i}\right)\end{aligned}$$

We have:

$$\begin{aligned}\sum_{d_i=1}\left(y_i - x_i^\top\beta\right)^2 &= \sum_{d_i=1}\left(y_i - x_i^\top\beta\right)\left(y_i - x_i^\top\beta\right) \\ &= n_1\sigma^2 - \sum_{d_i=1}\left(y_i - x_i^\top\beta\right)x_i^\top\beta\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[\frac{n_1}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{d_i=1}\left(y_i - x_i^\top\beta\right)^2\right] &= \mathbb{E}\left[-\frac{n_1}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{d_i=1}\left(y_i - x_i^\top\beta\right)x_i^\top\beta\right] \\ &= -\frac{1}{2\sigma^4}\sum_{i=1}^n \Phi_i + \frac{1}{\sigma^4}\sum_{i=1}^n \phi_i z_i\end{aligned}$$

We deduce that:

$$\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial\beta\,\partial\beta^\top}\right] = -\sum_{i=1}^n c_i$$

where:

$$\begin{aligned} c_i &= -\frac{1}{4\sigma^4}\left(-3\phi_i z_i - \frac{\phi_i^2 z_i^2}{1-\Phi_i} + \phi_i z_i^3\right) + \frac{1}{2\sigma^4}\Phi_i - \frac{1}{\sigma^4}\phi_i z_i \\ &= -\frac{1}{4\sigma^4}\left(\phi_i z_i^3 + \phi_i z_i - \frac{\phi_i^2 z_i^2}{1-\Phi_i} - 2\Phi_i\right) \end{aligned}$$

We conclude that the information matrix is equal to:

$$\begin{aligned} \mathcal{I}(\theta) &= -\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}(\theta)}{\partial\,\theta\,\partial\,\theta^\top}\right] \\ &= \begin{pmatrix} \sum_{i=1}^n a_i x_i x_i^\top & \sum_{i=1}^n b_i x_i \\ \sum_{i=1}^n b_i x_i & \sum_{i=1}^n c_i \end{pmatrix} \end{aligned}$$

We retrieve the formula obtained by Amemiya (1973).

8. We note $\mathbf{Y}_1$ the $n_1 \times 1$ vector of the explained variable and $\mathbf{X}_1$ the $n_1 \times K$ matrices of explanatory variables when the data are not censured. We also notice that:

$$\frac{\phi_i}{1-\Phi_i} = \lambda\left(x_i^\top \beta\right)$$

where λ is the inverse Mills ratio. The first-order condition (10.4) becomes:

$$-\hat{\sigma}\mathbf{X}_0^\top \mathbf{\Lambda}_0 + \mathbf{X}_1^\top\left(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}\right) = 0$$

where $\mathbf{\Lambda}_0$ is $(n-n_1)\times 1$ vector of inverse Mills ratio and $\mathbf{X}_0$ is the $(n-n_1)\times K$ matrices of explanatory variables when the data are censured. We deduce that:

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{X}_1^\top \mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top \mathbf{Y}_1 - \hat{\sigma}\left(\mathbf{X}_1^\top \mathbf{X}_1\right)^{-1}\mathbf{X}_0^\top \mathbf{\Lambda}_0 \\ &= \hat{\beta}_1 - \hat{\sigma}\left(\mathbf{X}_1^\top \mathbf{X}_1\right)^{-1}\mathbf{X}_0^\top \mathbf{\Lambda}_0 \end{aligned} \tag{10.6}$$

It follows that the OLS estimator $\hat{\beta}_1$ based on non-censured data is biased.

9. We apply results obtained in Question 1 to the random variable U with $\mu = 0$, $c = -x^\top\beta$ and $\alpha = \sigma^{-1}(c-\mu)$. We have:

$$\begin{aligned} \mathbb{E}[Y \mid Y > 0] &= \mathbb{E}\left[x^\top\beta + U \mid U > -x^\top\beta\right] \\ &= x^\top\beta + \mathbb{E}\left[U \mid U > -x^\top\beta\right] \\ &= x^\top\beta + \sigma\lambda\left(-\frac{x^\top\beta}{\sigma}\right) \end{aligned} \tag{10.7}$$

and:

$$\begin{aligned} \mathbb{E}[Y \mid Y \leq 0] &= \mathbb{E}\left[x^\top\beta + U \mid U \leq -x^\top\beta\right] \\ &= x^\top\beta + \mathbb{E}\left[U \mid U \leq -x^\top\beta\right] \\ &= x^\top\beta - \sigma\lambda\left(\frac{x^\top\beta}{\sigma}\right) \end{aligned} \tag{10.8}$$

Using Question 2, we obtain:

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\left[\max\left(x^\top\beta + U, 0\right)\right] \\
&= x^\top\beta + \mathbb{E}\left[\max\left(U, -x^\top\beta\right)\right] \\
&= x^\top\beta + \Phi\left(-\frac{x^\top\beta}{\sigma}\right)x^\top\beta + \sigma\left(1 - \Phi\left(-\frac{x^\top\beta}{\sigma}\right)\right)\lambda\left(-\frac{x^\top\beta}{\sigma}\right) \\
&= \Phi\left(\frac{x^\top\beta}{\sigma}\right)\left(x^\top\beta + \sigma\lambda\left(-\frac{x^\top\beta}{\sigma}\right)\right) \\
&= \Phi\left(\frac{x^\top\beta}{\sigma}\right)\mathbb{E}[Y \mid Y > 0]
\end{aligned}$$

From Equation (10.7), we deduce that the corresponding linear model is:

$$\mathbf{Y}_1 = \mathbf{X}_1\tilde{\beta} + \sigma\mathbf{\Lambda}_1$$

and:

$$\begin{aligned}
\tilde{\beta} &= \left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{Y}_1 - \hat{\sigma}\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{\Lambda}_1 \\
&= \hat{\beta}_1 - \hat{\sigma}\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{\Lambda}_1 \qquad (10.9)
\end{aligned}$$

The difference between the estimators (10.6) and (10.9) is the term $\mathbf{X}^\top\mathbf{\Lambda}$ which is calculated with censored data in the maximum likelihood and non-censored data in the last approach. However, the estimators (10.6) and (10.9) can not be used in practice because they depend on $\hat{\sigma}$ and on the inverse Mills ratio that is a function of $\hat{\beta}$ and $\hat{\sigma}$.

10. The ML estimates are $\hat{\beta}_0^{(\mathrm{ML})} = 2.8467$, $\hat{\beta}_1^{(\mathrm{ML})} = 1.0843$, $\hat{\beta}_2^{(\mathrm{ML})} = 0.9869$ and $\hat{\sigma}^{(\mathrm{ML})} = 5.5555$. The OLS estimates based on the non-censored data are $\hat{\beta}_0^{(\mathrm{OLS})} = 6.2002$, $\hat{\beta}_1^{(\mathrm{OLS})} = 0.6757$, $\hat{\beta}_2^{(\mathrm{OLS})} = 0.7979$. We verify that:

$$\hat{\beta}^{(\mathrm{OLS})} - \hat{\sigma}^{(\mathrm{ML})}\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_0^\top\mathbf{\Lambda}_0^{(\mathrm{ML})} = \begin{pmatrix} 2.8467 \\ 1.0843 \\ 0.9869 \end{pmatrix} = \hat{\beta}^{(\mathrm{ML})}$$

and:

$$\hat{\beta}^{(\mathrm{OLS})} - \hat{\sigma}^{(\mathrm{ML})}\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{\Lambda}_1^{(\mathrm{ML})} = \begin{pmatrix} 2.7095 \\ 1.0065 \\ 1.0522 \end{pmatrix} \neq \hat{\beta}^{(\mathrm{ML})}$$

The ML estimator combines the non-censored data – $\hat{\beta}^{(\mathrm{OLS})}$ – and the censored data – $\mathbf{X}_0^\top\mathbf{\Lambda}_0^{(\mathrm{ML})}$. This is not the case of the second estimator, which is only based on non-censored data – $\hat{\beta}^{(\mathrm{OLS})}$ and $\mathbf{X}_1^\top\mathbf{\Lambda}_1^{(\mathrm{ML})}$. The second estimator is then less efficient than the ML estimator since it does not use all the information provided by the data.

11. The conditional predicted value of $\breve{y}_i^\star$ is:

$$\breve{y}_i^\star = \begin{cases} x_i^\top\hat{\beta}^{(\mathrm{ML})} - \hat{\sigma}^{(\mathrm{ML})}\lambda\left(\dfrac{x_i^\top\hat{\beta}^{(\mathrm{ML})}}{\hat{\sigma}^{(\mathrm{ML})}}\right) & \text{if } y_i \leq 0 \\ x_i^\top\hat{\beta}^{(\mathrm{ML})} + \hat{\sigma}^{(\mathrm{ML})}\lambda\left(-\dfrac{x_i^\top\hat{\beta}^{(\mathrm{ML})}}{\hat{\sigma}^{(\mathrm{ML})}}\right) & \text{if } y_i > 0 \end{cases}$$

whereas the unconditional expectation is $\hat{y}_i^\star = x_i^\top\hat{\beta}^{(\mathrm{ML})}$. These values are reported in Table 10.1.

TABLE 10.1: Predicted values $\breve{y}_i^\star$ and $\hat{y}_i^\star$

i	1	2	3	4	5	6	7	8	9	10
y_i	4.0	0.0	0.5	0.0	0.0	17.4	18.0	0.0	0.0	9.7
$\breve{y}_i^\star$	3.5	−12.7	5.1	−5.9	−5.9	14.3	17.6	−4.2	−5.4	4.0
$\hat{y}_i^\star$	−3.0	−12.6	1.6	−3.4	−3.4	14.2	17.6	0.8	−2.4	−1.4
i	11	12	13	14	15	16	17	18	19	20
y_i	9.7	1.8	6.5	26.1	0.0	5.0	21.6	6.2	9.9	1.4
$\breve{y}_i^\star$	9.3	5.1	6.2	16.3	−15.1	3.6	19.6	5.9	9.3	6.4
$\hat{y}_i^\star$	8.6	1.6	3.9	16.3	−15.1	−2.5	19.6	3.4	8.6	4.3
i	21	22	23	24	25	26	27	28	29	30
y_i	5.0	0.0	0.0	18.1	0.0	7.7	0.0	0.0	0.0	4.0
$\breve{y}_i^\star$	4.1	−5.8	−9.4	15.7	−3.3	17.3	−6.2	−5.4	−3.4	12.2
$\hat{y}_i^\star$	−0.8	−3.2	−8.8	15.7	3.8	17.3	−3.9	−2.3	3.2	12.0

10.3.10 Derivation of Kalman filter equations

1. We have:

$$\begin{aligned}
\hat{\alpha}_{t|t-1} &= \mathbb{E}_{t-1}\left[\alpha_t\right] \\
&= \mathbb{E}_{t-1}\left[T_t \alpha_{t-1} + c_t + R_t \eta_t\right] \\
&= T_t \mathbb{E}_{t-1}\left[\alpha_{t-1}\right] + c_t \\
&= T_t \hat{\alpha}_{t-1|t-1} + c_t
\end{aligned}$$

We introduce the notation $\delta_t = \alpha_t - \hat{\alpha}_{t|t-1}$. It follows that:

$$\begin{aligned}
\delta_t &= T_t \alpha_{t-1} + c_t + R_t \eta_t - \left(T_t \hat{\alpha}_{t-1|t-1} + c_t\right) \\
&= T_t \left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right) + R_t \eta_t
\end{aligned}$$

and:

$$\begin{aligned}
\delta_t \delta_t^\top = \; & T_t \left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right)\left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right)^\top T_t^\top + \\
& 2 T_t \left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right) \eta_t^\top R_t^\top + \\
& R_t \eta_t \eta_t^\top R_t^\top
\end{aligned}$$

We deduce that:

$$\begin{aligned}
P_{t|t-1} &= \mathbb{E}_{t-1}\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)^\top\right] \\
&= T_t \mathbb{E}_{t-1}\left[\left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right)\left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right)^\top\right] T_t^\top + \\
& \quad 2 T_t \mathbb{E}_{t-1}\left[\left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}\right) \eta_t^\top\right] R_t^\top + \\
& \quad R_t \mathbb{E}_{t-1}\left[\eta_t \eta_t^\top\right] R_t^\top \\
&= T_t P_{t-1|t-1} T_t^\top + R_t Q_t R_t^\top
\end{aligned}$$

2. We have:

$$\begin{aligned}
v_t &= y_t - \mathbb{E}_{t-1}\left[y_t\right] \\
&= y_t - \mathbb{E}_{t-1}\left[Z_t \alpha_t + d_t + \epsilon_t\right] \\
&= y_t - Z_t \hat{\alpha}_{t|t-1} - d_t \\
&= Z_t \left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t
\end{aligned}$$

Since $\hat{\alpha}_{t|t-1}$ is a Gaussian vector, v_t is also Gaussian with:

$$\begin{aligned}
\mathbb{E}_{t-1}\left[v_t\right] &= \mathbb{E}_{t-1}\left[y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right] \\
&= Z_t\mathbb{E}_{t-1}\left[\alpha_t - \hat{\alpha}_{t|t-1}\right] + \mathbb{E}_{t-1}\left[\epsilon_t\right] \\
&= \mathbf{0}
\end{aligned}$$

and:

$$\begin{aligned}
F_t &= \mathbb{E}_{t-1}\left[\left(v_t - \mathbf{0}\right)\left(v_t - \mathbf{0}\right)^\top\right] \\
&= \mathbb{E}_{t-1}\left[\left(Z_t\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)\left(Z_t\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)^\top\right] \\
&= Z_t\mathbb{E}_{t-1}\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)^\top\right]Z_t^\top + \mathbb{E}_{t-1}\left[\epsilon_t\epsilon_t^\top\right] \\
&= Z_tP_{t|t-1}Z_t^\top + H_t
\end{aligned}$$

3. We have:

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\alpha_t v_t^\top\right] &= \mathbb{E}_{t-1}\left[\alpha_t\left(Z_t\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t\right)^\top\right] \\
&= \mathbb{E}_{t-1}\left[\alpha_t\left(\alpha_t^\top - \hat{\alpha}_{t|t-1}^\top\right)Z_t^\top\right] + \mathbb{E}_{t-1}\left[\alpha_t\epsilon_t^\top\right] \\
&= \mathbb{E}_{t-1}\left[\left(\alpha_t - \hat{\alpha}_{t|t-1}\right)\left(\alpha_t^\top - \hat{\alpha}_{t|t-1}^\top\right)\right]Z_t^\top + \\
&\quad \hat{\alpha}_{t|t-1}\mathbb{E}_{t-1}\left[\alpha_t^\top - \hat{\alpha}_{t|t-1}^\top\right]Z_t^\top \\
&= P_{t|t-1}Z_t^\top
\end{aligned}$$

and:

$$\begin{aligned}
\begin{pmatrix} \alpha_t \\ v_t \end{pmatrix} &= \begin{pmatrix} \alpha_t \\ Z_t\left(\alpha_t - \hat{\alpha}_{t|t-1}\right) + \epsilon_t \end{pmatrix} \\
&= \begin{pmatrix} I_m & \mathbf{0} \\ Z_t & I_n \end{pmatrix}\begin{pmatrix} \alpha_t \\ \epsilon_t \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ -Z_t\hat{\alpha}_{t|t-1} \end{pmatrix} \\
&= A_t\begin{pmatrix} \alpha_t \\ \epsilon_t \end{pmatrix} + B_t
\end{aligned}$$

Conditionally to the filtration $\mathcal{F}_{t-1}$, the random vector (α_t, v_t) is a linear combination $A_tX_t + B_t$ of the independent Gaussian random vector $X_t = (\alpha_t, \epsilon_t)$. We deduce that:

$$\begin{pmatrix} \alpha_t \\ v_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{\alpha}_{t|t-1} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & P_{t|t-1}Z_t^\top \\ Z_tP_{t|t-1} & F_t \end{pmatrix}\right)$$

4. We deduce that:

$$\begin{aligned}
\hat{\alpha}_{t|t} &= \mathbb{E}_t\left[\alpha_t\right] \\
&= \mathbb{E}\left[\alpha_t \mid v_t = y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right]
\end{aligned}$$

Using the standard results of the conditional distribution, we obtain:

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + P_{t|t-1}Z_t^\top F_t^{-1}\left(y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right)$$

and:

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}Z_t^\top F_t^{-1}Z_tP_{t|t-1}$$

5. The Kalman filter corresponds to the following recursive equations:

$$\left\{\begin{array}{l}\hat{\alpha}_{t|t-1} = T_t \hat{\alpha}_{t-1|t-1} + c_t \\ P_{t|t-1} = T_t P_{t-1|t-1} T_t^\top + R_t Q_t R_t^\top \\ v_t = y_t - Z_t \hat{\alpha}_{t|t-1} - d_t \\ F_t = Z_t P_{t|t-1} Z_t^\top + H_t \\ \hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + P_{t|t-1} Z_t^\top F_t^{-1} v_t \\ P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t^\top F_t^{-1} Z_t P_{t|t-1}\end{array}\right.$$

We have:

$$\begin{array}{rcl}\hat{\alpha}_{t+1|t} & = & T_{t+1}\hat{\alpha}_{t|t} + c_{t+1} \\ & = & T_{t+1}\left(\hat{\alpha}_{t|t-1} + P_{t|t-1} Z_t^\top F_t^{-1} v_t\right) + c_{t+1} \\ & = & T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1} + K_t v_t\end{array}$$

where K_t is the gain matrix:

$$K_t = T_{t+1} P_{t|t-1} Z_t^\top F_t^{-1}$$

Since we have $v_t = y_t - Z_t \hat{\alpha}_{t|t-1} - d_t$, we can write the state space model as follows:

$$\left\{\begin{array}{l} y_t = Z_t \hat{\alpha}_{t|t-1} + d_t + v_t \\ \hat{\alpha}_{t+1|t} = T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1} + K_t v_t\end{array}\right.$$

If $v_t = \mathbf{0}$, then $\hat{\alpha}_{t+1|t} = T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1}$. K_t indicates how the filter changes the classical estimation $T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1}$ when it takes into account innovation errors. Therefore, K_t is the correction matrix of the prediction-correction method.

6. We introduce the process $\gamma_t = \gamma_{t-1}$ with $\gamma_0 = 1$. Another representation of the state space model is:

$$\left\{\begin{array}{l} y_t = Z_t \alpha_t + d_t \gamma_t + \epsilon_t \\ \alpha_t = T_t \alpha_{t-1} + c_t \gamma_t + R_t \eta_t \\ \gamma_t = \gamma_{t-1}\end{array}\right.$$

We obtain:

$$\left\{\begin{array}{l} y_t = Z_t^\star \alpha_t^\star \\ \alpha_t^\star = T_t^\star \alpha_{t-1}^\star + R_t^\star \eta_t^\star\end{array}\right.$$

where:

$$Z_t^\star = \left(\begin{array}{ccc} Z_t & d_t & I_n\end{array}\right)$$

$$T_t^\star = \left(\begin{array}{ccc} T_t & c_t & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{array}\right)$$

$$R_t^\star = \left(\begin{array}{cc} R_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_n\end{array}\right)$$

The state vector becomes $\alpha_t^\star = (\alpha_t, \gamma_t, \epsilon_t)$ whereas the noise process $\eta_t^\star = (\eta_t, \epsilon_t)$ is a Gaussian random vector $\mathcal{N}\left(\mathbf{0}, Q_t^\star\right)$ where:

$$Q_t^\star = \left(\begin{array}{cc} Q_t & C_t^\top \\ C_t & H_t\end{array}\right)$$

7. If we apply the Kalman filter to the augmented state space model, we obtain:

$$\left\{\begin{array}{l}\hat{\alpha}^{\star}_{t|t-1} = T^{\star}_t \hat{\alpha}^{\star}_{t-1|t-1} \\ P^{\star}_{t|t-1} = T^{\star}_t P^{\star}_{t-1|t-1} T^{\star\top}_t + R^{\star}_t Q^{\star}_t R^{\star\top}_t \\ \hat{y}^{\star}_{t|t-1} = Z^{\star}_t \hat{\alpha}^{\star}_{t|t-1} \\ v^{\star}_t = y_t - \hat{y}^{\star}_{t|t-1} \\ F^{\star}_t = Z^{\star}_t P^{\star}_{t|t-1} Z^{\star\top}_t \\ \hat{\alpha}^{\star}_{t|t} = \hat{\alpha}^{\star}_{t|t-1} + P^{\star}_{t|t-1} Z^{\star\top}_t F^{\star-1}_t v^{\star}_t \\ P^{\star}_{t|t} = P^{\star}_{t|t-1} - P^{\star}_{t|t-1} Z^{\star\top}_t F^{\star-1}_t Z^{\star}_t P^{\star}_{t|t-1}\end{array}\right.$$

We have $\hat{\alpha}^{\star}_0 = (\hat{\alpha}_0, 0, \mathbf{0})$ and:

$$P^{\star}_0 = \begin{pmatrix} P_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

We assume that $P^{\star}_{t|t}$ has the following structure:

$$P^{\star}_{t|t} = \begin{pmatrix} P_{t|t} & \mathbf{0} & V_t \\ \mathbf{0} & 0 & \mathbf{0} \\ V_t^\top & \mathbf{0} & W_t \end{pmatrix}$$

We deduce that $V_t = R_t C_t^\top$ and $W_t = H_t$. Finally, we obtain:

$$\left\{\begin{array}{l}\hat{\alpha}_{t|t-1} = T_t \hat{\alpha}_{t-1|t-1} + c_t \\ P_{t|t-1} = T_t P_{t-1|t-1} T_t^\top + R_t Q_t R_t^\top \\ \hat{y}_{t|t-1} = Z_t \hat{\alpha}_{t|t-1} + d_t \\ v_t = y_t - \hat{y}_{t|t-1} \\ F_t = Z_t P_{t|t-1} Z_t^\top + 2 Z_t R_t C_t^\top + H_t \\ G_t = P_{t|t-1} Z_t^\top + R_t C_t^\top \\ \hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + G_t F_t^{-1} v_t \\ P_{t|t} = P_{t|t-1} - G_t F_t^{-1} G_t^\top\end{array}\right.$$

10.3.11 Steady state of time-invariant state space model

1. We have $\alpha_t = (y_t, \varepsilon_t)$ and:

$$\alpha_t = \begin{pmatrix} \phi_1 & 0 \\ 0 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \eta_t$$

Using the standard SSM notations, we have $c = (\mu, 0)$, $R = (1, 1)$, $Q = \sigma^2_\varepsilon$ and:

$$T = \begin{pmatrix} \phi_1 & 0 \\ 0 & 0 \end{pmatrix}$$

It follows that:

$$I_2 - T = \begin{pmatrix} 1 - \phi_1 & 0 \\ 0 & 1 \end{pmatrix}$$

and:

$$\begin{aligned}(I_2 - T)^{-1} &= \frac{1}{1-\phi_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 - \phi_1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{1-\phi_1} & 0 \\ 0 & 1 \end{pmatrix}\end{aligned}$$

The steady state $\hat{\alpha}_\infty$ is then equal to:

$$\begin{aligned}
\hat{\alpha}_\infty &= (I_2 - T)^{-1} c \\
&= \frac{1}{1-\phi_1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\phi_1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\mu}{1-\phi_1} \\ 0 \end{pmatrix}
\end{aligned}$$

We also have:

$$RQR^\top = \sigma_\varepsilon^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and:

$$T \otimes T = \begin{pmatrix} \phi_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We obtain:

$$(I_4 - T \otimes T)^{-1} = \begin{pmatrix} \frac{1}{1-\phi_1^2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and:

$$\begin{aligned}
\operatorname{vec}(P_\infty) &= (I_4 - T \otimes T)^{-1} \operatorname{vec}\left(RQR^\top\right) \\
&= \begin{pmatrix} \frac{1}{1-\phi_1^2}\sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix}
\end{aligned}$$

We finally deduce that:

$$P_\infty = \begin{pmatrix} \frac{\sigma_\varepsilon^2}{1-\phi_1^2} & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \end{pmatrix}$$

2. We have $\alpha_t = (y_t, \varepsilon_t)$ and:

$$\alpha_t = \begin{pmatrix} 0 & -\theta_1 \\ 0 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \eta_t$$

Using the standard SSM notations, we have $c = (\mu, 0)$, $R = (1, 1)$, $Q = \sigma_\varepsilon^2$ and:

$$T = \begin{pmatrix} 0 & -\theta_1 \\ 0 & 0 \end{pmatrix}$$

We obtain:

$$(I_2 - T)^{-1} = \begin{pmatrix} 1 & -\theta_1 \\ 0 & 1 \end{pmatrix}$$

and:

$$\hat{\alpha}_\infty = \begin{pmatrix} 1 & -\theta_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix}$$

We have:

$$I_4 - T \otimes T = \begin{pmatrix} 1 & 0 & 0 & -\theta_1^2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and:

$$\operatorname{vec}(P_\infty) = \begin{pmatrix} 1 & 0 & 0 & \theta_1^2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 \left(1 + \theta_1^2\right) \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix}$$

Finally, we obtain:

$$P_\infty = \sigma_\varepsilon^2 \begin{pmatrix} 1 + \theta_1^2 & 1 \\ 1 & 1 \end{pmatrix}$$

3. We have $\alpha_t = (y_t, \varepsilon_t)$ and:

$$\alpha_t = \begin{pmatrix} \phi_1 & -\theta_1 \\ 0 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \eta_t$$

Using the standard SSM notations, we have $c = (\mu, 0)$, $R = (1, 1)$, $Q = \sigma_\varepsilon^2$ and:

$$T = \begin{pmatrix} \phi_1 & -\theta_1 \\ 0 & 0 \end{pmatrix}$$

We obtain:

$$(I_2 - T)^{-1} = \frac{1}{1 - \phi_1} \begin{pmatrix} 1 & -\theta_1 \\ 0 & 1 - \phi_1 \end{pmatrix}$$

and:

$$\hat{\alpha}_\infty = \frac{1}{1 - \phi_1} \begin{pmatrix} 1 & -\theta_1 \\ 0 & 1 - \phi_1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\mu}{1-\phi_1} \\ 0 \end{pmatrix}$$

We also have:

$$T \otimes T = \begin{pmatrix} \phi_1^2 & -\phi_1\theta_1 & -\phi_1\theta_1 & \theta_1^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

It follows that:

$$\operatorname{vec}(P_\infty) = \begin{pmatrix} \frac{1}{1-\phi_1^2} & -\frac{\phi_1\theta_1}{1-\phi_1^2} & -\frac{\phi_1\theta_1}{1-\phi_1^2} & \frac{\theta_1^2}{1-\phi_1^2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix}$$

and:

$$P_\infty = \sigma_\varepsilon^2 \begin{pmatrix} \frac{1-2\phi_1\theta_1+\theta_1^2}{1-\phi_1^2} & 1 \\ 1 & 1 \end{pmatrix}$$

4. We have $\alpha_t = (y_t, u_t)$ and:

$$\alpha_t = \begin{pmatrix} 0 & \theta_1 \\ 0 & \theta_1 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \eta_t$$

Using the standard SSM notations, we have $c = (\mu, 0)$, $R = (1, 1)$, $Q = \sigma_\varepsilon^2$ and:

$$T = \begin{pmatrix} 0 & \theta_1 \\ 0 & \theta_1 \end{pmatrix}$$

We obtain:

$$\hat{\alpha}_\infty = \frac{1}{1-\theta_1} \begin{pmatrix} 1-\theta_1 & \theta_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix}$$

and:

$$\operatorname{vec}(P_\infty) = \begin{pmatrix} 1 & 0 & 0 & \frac{\theta_1^2}{1-\theta_1^2} \\ 0 & 1 & 0 & \frac{\theta_1^2}{1-\theta_1^2} \\ 0 & 0 & 1 & \frac{\theta_1^2}{1-\theta_1^2} \\ 0 & 0 & 0 & \frac{1}{1-\theta_1^2} \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\theta_1^2} \\ \frac{1}{1-\theta_1^2} \\ \frac{1}{1-\theta_1^2} \\ \frac{1}{1-\theta_1^2} \end{pmatrix} \sigma_\varepsilon^2$$

We deduce that:

$$P_\infty = \sigma_\varepsilon^2 \begin{pmatrix} \frac{1}{1-\theta_1^2} & \frac{1}{1-\theta_1^2} \\ \frac{1}{1-\theta_1^2} & \frac{1}{1-\theta_1^2} \end{pmatrix}$$

10.3.12 Kalman information filter versus Kalman covariance filter

In what follows, X^{-1} defines the inverse of the square matrix X and Y^{-1} defines the Moore-Penrose pseudo-inverse of the non-square matrix Y.

1. We have:

$$\begin{aligned} \left(I_m + AB^\top C^{-1}B\right)^{-1} A &= \left(I_m + AB^\top C^{-1}B\right)^{-1} \left(A^{-1}\right)^{-1} \\ &= \left(A^{-1}\left(I_m + AB^\top C^{-1}B\right)\right)^{-1} \\ &= \left(A^{-1} + B^\top C^{-1}B\right)^{-1} \end{aligned}$$

2. If the relationship $\left(I_m + AB^\top C^{-1}B\right)^{-1} = I_m - AB^\top\left(C + BAB^\top\right)^{-1} B$ is true, we must verify that:

$$\begin{aligned} (*) &= \left(I_m + AB^\top C^{-1}B\right)\left(I_m - AB^\top\left(C + BAB^\top\right)^{-1} B\right) \\ &= I_m \end{aligned}$$

We have:

$$\begin{aligned} (*) &= \left(I_m + AB^\top C^{-1}B\right)\left(I_m - AB^\top\left(C + BAB^\top\right)^{-1} B\right) \\ &= \left(I_m + AB^\top C^{-1}B\right) - \left(I_m + AB^\top C^{-1}B\right) AB^\top \left(C + BAB^\top\right)^{-1} B \\ &= \left(I_m + AB^\top C^{-1}B\right) - \left(I_m + AB^\top C^{-1}B\right) AB^\top \left(B^{-1}C + AB^\top\right)^{-1} \\ &= \left(I_m + AB^\top C^{-1}B\right) - \left(I_m + AB^\top C^{-1}B\right)\left(B^{-1}CB^{\top^{-1}}A^{-1} + I_m\right)^{-1} \end{aligned}$$

Since we have:

$$\begin{aligned} I_m + AB^\top C^{-1}B &= \left(AB^\top C^{-1}B\right)\left(\left(AB^\top C^{-1}B\right)^{-1} + I_m\right) \\ &= \left(AB^\top C^{-1}B\right)\left(B^{-1}CB^{\top^{-1}}A^{-1} + I_m\right) \end{aligned}$$

We deduce that:

$$\begin{aligned}
(*) &= \left(I_m + AB^\top C^{-1}B\right) - \\
&\quad \left(AB^\top C^{-1}B\right)\left(B^{-1}CB^{\top^{-1}}A^{-1} + I_m\right)\left(B^{-1}CB^{\top^{-1}}A^{-1} + I_m\right)^{-1} \\
&= \left(I_m + AB^\top C^{-1}B\right) - \left(AB^\top C^{-1}B\right) \\
&= I_m
\end{aligned}$$

3. Using Questions 1 and 2, we have:

$$\left(I_m + AB^\top C^{-1}B\right)^{-1} A = \left(A^{-1} + B^\top C^{-1}B\right)^{-1}$$

and:

$$\left(I_m + AB^\top C^{-1}B\right)^{-1} = I_m - AB^\top \left(C + BAB^\top\right)^{-1} B$$

We deduce that:

$$\begin{aligned}
\left(A^{-1} + B^\top C^{-1}B\right)^{-1} &= \left(I_m - AB^\top \left(C + BAB^\top\right)^{-1} B\right) A \\
&= A - AB^\top \left(C + BAB^\top\right)^{-1} BA
\end{aligned}$$

and:

$$\begin{aligned}
(*) &= \left(I_m + AB^\top C^{-1}B\right)^{-1} AB^\top C^{-1} \\
&= \left(A^{-1} + B^\top C^{-1}B\right)^{-1} B^\top C^{-1} \\
&= \left(A - AB^\top \left(C + BAB^\top\right)^{-1} BA\right) B^\top C^{-1} \\
&= AB^\top C^{-1} - AB^\top \left(C + BAB^\top\right)^{-1} BAB^\top C^{-1} \\
&= AB^\top C^{-1} - AB^\top \left(C + BAB^\top\right)^{-1} \left(\left(BAB^\top + C\right) C^{-1} - I\right) \\
&= AB^\top C^{-1} - AB^\top \left(C + BAB^\top\right)^{-1} \left(BAB^\top + C\right) C^{-1} + \\
&\quad AB^\top \left(C + BAB^\top\right)^{-1} \\
&= AB^\top C^{-1} - AB^\top C^{-1} + AB^\top \left(C + BAB^\top\right)^{-1}
\end{aligned}$$

Finally, we obtain the expected result:

$$\left(I_m + AB^\top C^{-1}B\right)^{-1} AB^\top C^{-1} = AB^\top \left(C + BAB^\top\right)^{-1}$$

4. We have:

$$\begin{aligned}
(*) &= \left(I_m + D^{-1}A\right)(A + D)^{-1} \\
&= (A + D)^{-1} + D^{-1}A(A + D)^{-1} \\
&= (A + D)^{-1} + D^{-1}\left(I_m + DA^{-1}\right)^{-1} \\
&= (A + D)^{-1} + D^{-1}\left(I_m - \left(I_m + DA^{-1}\right)^{-1} DA^{-1}\right) \\
&= (A + D)^{-1} + D^{-1} - D^{-1}\left(AD^{-1} + I_m\right)^{-1} \\
&= (A + D)^{-1} + D^{-1} - (A + D)^{-1} \\
&= D^{-1}
\end{aligned}$$

5. Let V be a covariance matrix. The information matrix $\mathbb{I}$ is the inverse of the covariance matrix V:

$$\mathbb{I} = V^{-1}$$

This matrix is used in the method of maximum likelihood.

6. The state $\hat{\alpha}^{\star}_{t|t}$ (resp. $\hat{\alpha}^{\star}_{t|t-1}$) is the estimator of α_t normalized by the covariance matrix given the filtration $\mathcal{F}_t$ (resp. $\mathcal{F}_{t-1}$).

7. We have:

$$\begin{aligned}
\mathbb{I}_{t|t} P_{t|t-1} &= P_{t|t}^{-1} P_{t|t-1} \\
&= P_{t|t-1}^{-1} \left(I_m - P_{t|t-1} Z_t^{\top} F_t^{-1} Z_t \right)^{-1} P_{t|t-1} \\
&= P_{t|t-1}^{-1} \left(I_m + P_{t|t-1} Z_t^{\top} H_t^{-1} Z_t \right) P_{t|t-1} \\
&= I_m + Z_t^{\top} H_t^{-1} Z_t P_{t|t-1}
\end{aligned}$$

because:

$$\begin{aligned}
(*) &= \left(I_m - P_{t|t-1} Z_t^{\top} F_t^{-1} Z_t \right)^{-1} \\
&= I_m - P_{t|t-1} Z_t^{\top} \left(-F_t + Z_t P_{t|t-1} Z_t^{\top} \right)^{-1} Z_t \\
&= I_m - P_{t|t-1} Z_t^{\top} \left(-H_t \right)^{-1} Z_t \\
&= I_m + P_{t|t-1} Z_t^{\top} H_t^{-1} Z_t
\end{aligned}$$

We also have:

$$\begin{aligned}
(*) &= \mathbb{I}_{t|t} P_{t|t-1} Z_t^{\top} \left(Z_t P_{t|t-1} Z_t^{\top} + H_t \right)^{-1} \\
&= \left(I_m + Z_t^{\top} H_t^{-1} Z_t P_{t|t-1} \right) Z_t^{\top} \left(Z_t P_{t|t-1} Z_t^{\top} + H_t \right)^{-1} \\
&= Z_t^{\top} \left(I_n + H_t^{-1} Z_t P_{t|t-1} Z_t^{\top} \right) \left(Z_t P_{t|t-1} Z_t^{\top} + H_t \right)^{-1}
\end{aligned}$$

By using Question 4 with $A = Z_t P_{t|t-1} Z_t^{\top}$ and $D = H_t$, we obtain:

$$\mathbb{I}_{t|t} P_{t|t-1} Z_t^{\top} \left(Z_t P_{t|t-1} Z_t^{\top} + H_t \right)^{-1} = Z_t^{\top} H_t^{-1}$$

8. We have:

(a)

$$\begin{aligned}
\mathbb{I}_{t|t-1} &= P_{t|t-1}^{-1} \\
&= \left(T_t P_{t-1} T_t^{\top} + R_t Q_t R_t^{\top} \right)^{-1} \\
&= \left(T_t \mathbb{I}_{t-1|t-1}^{-1} T_t^{\top} + R_t Q_t R_t^{\top} \right)^{-1}
\end{aligned}$$

(b)

$$\begin{aligned}
\hat{\alpha}^{\star}_{t|t-1} &= \mathbb{I}_{t|t-1} \hat{\alpha}_{t|t-1} \\
&= \mathbb{I}_{t|t-1} T_t \hat{\alpha}_{t-1|t-1} \\
&= \mathbb{I}_{t|t-1} T_t \mathbb{I}_{t-1|t-1}^{-1} \hat{\alpha}^{\star}_{t-1|t-1}
\end{aligned}$$

(c)

$$\begin{aligned}
\mathbb{I}_{t|t} &= P_{t|t}^{-1} \\
&= P_{t|t-1}^{-1}\left(I_m - P_{t|t-1}Z_t^\top F_t^{-1}Z_t\right)^{-1} \\
&= P_{t|t-1}^{-1}\left(I_m + P_{t|t-1}Z_t^\top H_t^{-1}Z_t\right) \\
&= \mathbb{I}_{t|t-1} + Z_t^\top H_t^{-1}Z_t
\end{aligned}$$

(d)

$$\begin{aligned}
\hat{\alpha}_{t|t}^\star &= \mathbb{I}_{t|t}\hat{\alpha}_{t|t} \\
&= \mathbb{I}_{t|t}\left(\hat{\alpha}_{t|t-1} + P_{t|t-1}Z_t^\top F_t^{-1}\left(y_t - Z_t\hat{\alpha}_{t|t-1}\right)\right) \\
&= \mathbb{I}_{t|t}\left(I_m - P_{t|t-1}Z_t^\top F_t^{-1}Z_t\right)\hat{\alpha}_{t|t-1} + \\
&\quad \mathbb{I}_{t|t}P_{t|t-1}Z_t^\top F_t^{-1}y_t \\
&= \hat{\alpha}_{t|t-1}^\star + Z_t^\top H_t^{-1}y_t
\end{aligned}$$

We deduce that the recursive equations of the Kalman information filter are:

$$\begin{cases}
\mathbb{I}_{t|t-1} = \left(T_t\mathbb{I}_{t-1|t-1}^{-1}T_t^\top + R_tQ_tR_t^\top\right)^{-1} \\
\hat{\alpha}_{t|t-1}^\star = \mathbb{I}_{t|t-1}T_t\mathbb{I}_{t-1|t-1}^{-1}\hat{\alpha}_{t-1|t-1}^\star \\
\mathbb{I}_{t|t} = \mathbb{I}_{t|t-1} + Z_t^\top H_t^{-1}Z_t \\
\hat{\alpha}_{t|t}^\star = \hat{\alpha}_{t|t-1}^\star + Z_t^\top H_t^{-1}y_t
\end{cases}$$

From a numerical point of views, the number of matrix operations is:

- 5 additions, 10 multiplications and 1 inverse for the covariance filter;
- 3 additions, 10 multiplications and 2 inverses for the information filter;

It is not obvious that the computational time is reduced when using the information filter. Its advantage may be due to the inverse of $\mathbb{I}_{t-1|t-1}$ that can be more stable than F_t^{-1} in some cases.

9. We have:

$$\boldsymbol{\ell}(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|F_t| - \frac{1}{2}\sum_{t=1}^{T}v_t^\top F_t^{-1}v_t$$

In the case of the Kalman information matrix, we have:

$$\begin{cases}
F_t = Z_t\mathbb{I}_{t|t-1}^{-1}Z_t^\top + H_t \\
v_t = y_t - Z_t\mathbb{I}_{t|t-1}^{-1}\hat{\alpha}_{t|t-1}^\star \\
\hat{\alpha}_0^\star = \mathbf{0} \\
\mathbb{I}_0 = \mathbf{0}
\end{cases}$$

In the case of the Kalman covariance matrix, we set $\alpha_0 \sim \mathcal{N}(\mathbf{0}, \kappa I_m)$ where κ is a scalar sufficiently high such that $\mathbb{I}_0 = \kappa^{-1}I_m \simeq \mathbf{0}$.

10.3.13 Granger representation theorem

1. We have:

$$y_t = \mu_t + \Phi'_1 y_{t-1} + \varepsilon_t$$

We deduce that:

$$y_t - y_{t-1} = \mu_t + \Phi'_1 y_{t-1} - y_{t-1} + \varepsilon_t$$

and:

$$\Delta y_t = \mu_t + (\Phi'_1 - I_n)\, y_{t-1} + \varepsilon_t$$

2. We have:

$$y_t = \mu_t + \Phi'_1 y_{t-1} + \Phi'_2 y_{t-2} + \varepsilon_t$$

We deduce that:

$$\begin{aligned} \Delta y_t &= \mu_t + (\Phi'_1 - I_n)\, y_{t-1} + \Phi'_2 y_{t-2} + \varepsilon_t \\ &= \mu_t + (\Phi'_1 + \Phi'_2 - I_n)\, y_{t-1} - \Phi'_2 \Delta y_{t-1} + \varepsilon_t \end{aligned}$$

3. The relationship is true for $p = 1$ and $p = 2$. We note:

$$\Pi^{(p)} = \sum_{i=1}^{p} \Phi'_i - I_n$$

and

$$\Phi_i^{(p)} = -\sum_{j=i+1}^{p} \Phi'_j$$

We notice that:

$$\begin{aligned} \Pi^{(p)} &= \sum_{i=1}^{p} \Phi'_i - I_n \\ &= \left(\sum_{i=1}^{p-1} \Phi'_i - I_n\right) + \Phi'_p \\ &= \Pi^{(p-1)} + \Phi'_p \end{aligned}$$

and:

$$\begin{aligned} \Phi_i^{(p)} &= -\sum_{j=i+1}^{p} \Phi'_j \\ &= -\sum_{j=i+1}^{p-1} \Phi'_j - \Phi'_p \\ &= \Phi_i^{(p-1)} - \Phi'_p \end{aligned}$$

We prove the relationship by induction. Let us assume that it holds for the order $p-1$. We have:

$$\begin{aligned} \Delta y_t &= y_t - y_{t-1} \\ &= \mu_t + \sum_{i=1}^{p} \Phi'_i y_{t-i} + \varepsilon_t - y_{t-1} \\ &= \mu_t + \sum_{i=1}^{p-1} \Phi'_i y_{t-i} + \varepsilon_t - y_{t-1} + \Phi'_p y_{t-p} \\ &= \mu_t + \Pi^{(p-1)} y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^{(p-1)} \Delta y_{t-i} + \varepsilon_t + \Phi'_p y_{t-p} \\ &= \mu_t + \left(\Pi^{(p)} - \Phi'_p\right) y_{t-1} + \sum_{i=1}^{p-1} \left(\Phi_i^{(p)} + \Phi'_p\right) \Delta y_{t-i} + \\ &\quad\ \varepsilon_t + \Phi'_p y_{t-p} \\ &= \mu_t + \Pi^{(p)} y_{t-1} + \sum_{i=1}^{p} \Phi_i^{(p)} \Delta y_{t-i} + \varepsilon_t + \eta_t \end{aligned}$$

Since $\Phi_p^{(p)} = \mathbf{0}$, the value of η_t is equal to:

$$\begin{aligned}
\eta_t &= -\Phi'_p y_{t-1} + \sum_{i=1}^{p-1} \Phi'_p \Delta y_{t-i} - \Phi_p^{(p)} \Delta y_{t-p} + \Phi'_p y_{t-p} \\
&= -\Phi'_p y_{t-1} + \sum_{i=1}^{p-1} \Phi'_p \left(y_{t-i} - y_{t-i-1}\right) + \Phi'_p y_{t-p} \\
&= -\Phi'_p y_{t-1} + \Phi'_p y_{t-1} - \Phi'_p y_{t-2} + \Phi'_p y_{t-2} + \\
&\quad \ldots - \Phi'_p y_{t-p} + \Phi'_p y_{t-p} \\
&= \mathbf{0}
\end{aligned}$$

It follows that the statement also holds for the order p.

10.3.14 Probability distribution of the periodogram

1. Since $a(\lambda_j)$ and $b(\lambda_j)$ are the sums of Gaussian random variables, they are also Gaussian. We have:

$$\begin{aligned}
\mathbb{E}\left[a\left(\lambda_j\right)\right] &= \mathbb{E}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \cos\left(\lambda_j t\right)\right] \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{E}\left[y_t\right] \cos\left(\lambda_j t\right) \\
&= 0
\end{aligned}$$

and $\mathbb{E}\left[b\left(\lambda_j\right)\right] = 0$. For the variance, we have:

$$\begin{aligned}
\operatorname{var}\left(a\left(\lambda_j\right)\right) &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \cos\left(\lambda_j t\right)\right)^2\right] \\
&= \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[y_t^2\right] \cos^2\left(\lambda_j t\right) + \\
&\quad \frac{1}{n} \sum_{s \neq t} \mathbb{E}\left[y_s y_t\right] \cos\left(\lambda_j s\right) \cos\left(\lambda_j t\right) \\
&= \frac{\sigma^2}{n} \sum_{t=1}^{n} \cos^2\left(\lambda_j t\right) + 0 \\
&= \frac{\sigma^2}{n} \sum_{t=1}^{n} \left(\frac{\cos\left(2\lambda_j t\right) + 1}{2}\right)
\end{aligned}$$

If $\lambda_j \neq 0$, we obtain:

$$\begin{aligned}
\lim_{n \to \infty} \operatorname{var}\left(a\left(\lambda_j\right)\right) &= \frac{\sigma^2}{2} + \frac{\sigma^2}{2} \lim_{n \to \infty} \sum_{t=1}^{n} \frac{\cos\left(2\lambda_j t\right)}{n} \\
&= \frac{\sigma^2}{2}
\end{aligned}$$

We also have:

$$\begin{aligned}
\operatorname{var}\left(b\left(\lambda_j\right)\right) &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t \sin\left(\lambda_j t\right)\right)^2\right] \\
&= \frac{\sigma^2}{n}\sum_{t=1}^{n}\sin^2\left(\lambda_j t\right) \\
&= \frac{\sigma^2}{n}\sum_{t=1}^{n}\left(\frac{1-\cos\left(2\lambda_j t\right)}{2}\right)
\end{aligned}$$

f $\lambda_j \neq 0$, we obtain:

$$\lim_{n\to\infty}\operatorname{var}\left(b\left(\lambda_j\right)\right) = \frac{\sigma^2}{2}$$

We deduce that $a\left(\lambda_j\right) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2}\right)$ and $b\left(\lambda_j\right) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2}\right)$.

2. We have:

$$\begin{aligned}
\lim_{n\to\infty}\operatorname{cov}\left(a\left(\lambda_j\right), b\left(\lambda_j\right)\right) &= \lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\sum_{s=1}^{n} y_s \cos\left(\lambda_j s\right)\sum_{t=1}^{n} y_t \sin\left(\lambda_j t\right)\right] \\
&= \lim_{n\to\infty}\frac{\sigma^2}{n}\sum_{t=1}^{n}\cos\left(\lambda_j t\right)\sin\left(\lambda_j t\right) \\
&= 0
\end{aligned}$$

It follows that $a\left(\lambda_j\right)$ and $b\left(\lambda_j\right)$ are asymptotically independent. We conclude that:

$$\frac{2}{\sigma^2}\left(a^2\left(\lambda_j\right) + b^2\left(\lambda_j\right)\right) \sim \chi_2^2$$

and:

$$\frac{4\pi}{\sigma^2} I_y\left(\lambda_j\right) \sim \chi_2^2$$

3. We verify that:

$$\mathbb{E}\left[I_y\left(\lambda_j\right)\right] = \frac{f_y\left(\lambda_j\right)}{2}\mathbb{E}\left[\chi_2^2\right] = f_y\left(\lambda_j\right)$$

and:

$$\operatorname{var}\left(I_y\left(\lambda_j\right)\right) = \frac{f_y^2\left(\lambda_j\right)}{4}\operatorname{var}\left(\chi_2^2\right) = f_y^2\left(\lambda_j\right)$$

Since we have:

$$\Pr\left\{0.0506 \leq \chi_2^2 \leq 7.3778\right\} = 95\%$$

we deduce that:

$$\Pr\left\{0.0506 \leq 2\frac{I_y\left(\lambda_j\right)}{f_y\left(\lambda_j\right)} \leq 7.3778\right\} = 95\%$$

Finally, we obtain:

$$\Pr\left\{0.27 \cdot I_y\left(\lambda_j\right) \leq f_y\left(\lambda_j\right) \leq 39.5 \cdot I_y\left(\lambda_j\right)\right\} = 95\%$$

4. If $\lambda_j = 0$, we obtain:

$$\lim_{n\to\infty} \operatorname{var}(a(0)) = \frac{\sigma^2}{n}\sum_{t=1}^{n}\left(\frac{1+1}{2}\right) = \sigma^2$$

and:

$$\operatorname{var}(b(0)) = 0$$

It follows that:

$$\frac{1}{\sigma^2}\left(a^2(0) + b^2(0)\right) \sim \chi_1^2$$

and:

$$\frac{2\pi}{\sigma^2} I_y(0) \sim \chi_1^2$$

For a white noise process, we have $f_y(0) = (2\pi)^{-1}\sigma^2$. Therefore, we can make the hypothesis that:

$$\lim_{n\to\infty} \frac{I_y(0)}{f_y(0)} \sim \chi_1^2$$

We notice that:

$$\lim_{n\to\infty} \mathbb{E}[I_y(0)] = f_y(0)\cdot\mathbb{E}\left[\chi_1^2\right] = f_y(0)$$

and:

$$\lim_{n\to\infty} \operatorname{var}(I_y(0)) = f_y^2(0)\cdot\operatorname{var}\left(\chi_1^2\right) = 2f_y^2(0)$$

10.3.15 Spectral density function of structural time series models

1. We use the canonical representation of state space models. For Model (M1), we have $Z_t = 1$, $\alpha_t = \mu_t$, $d_t = 0$, $H_t = \sigma_\varepsilon^2$, $T_t = 1$, $c_t = 0$, $R_t = 1$ and $Q_t = \sigma_\eta^2$. For model (M2), we obtain $Z_t = \begin{pmatrix} 1 & 0 \end{pmatrix}$, $\alpha_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}$, $d_t = 0$, $H_t = \sigma_\varepsilon^2$, $T_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $c_t = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $R_t = I$ and $Q_t = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}$.

2. For Model (M1), we have:

$$\begin{aligned}
(1-L)\,y_t &= (\mu_t + \varepsilon_t) - (\mu_{t-1} + \varepsilon_{t-1}) \\
&= (\mu_t - \mu_{t-1}) + (\varepsilon_t - \varepsilon_{t-1}) \\
&= \eta_t + (1-L)\,\varepsilon_t
\end{aligned}$$

Since the sum of two stationary processes is stationary, it follows that $\eta_t + (1-L)\,\varepsilon_t$ is stationary. We deduce that the stationary form is $\mathcal{S}(y_t) = (1-L)\,y_t$. The spectral density function is equal to:

$$\begin{aligned}
f_{\mathcal{S}(y)}(\lambda) &= (2\pi)^{-1}\left(\sigma_\eta^2 + \left|1 - e^{-i\lambda}\right|^2 \sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + \left|1 - (\cos(-\lambda) + i\sin(-\lambda))\right|^2 \sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + \left|1 - (\cos\lambda - i\sin\lambda)\right|^2 \sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + \left|(1-\cos\lambda) + i\sin\lambda\right|^2 \sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + \left((1-\cos\lambda)^2 + \sin^2\lambda\right)\sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + \left(1 - 2\cos\lambda + \cos^2\lambda + \sin^2\lambda\right)\sigma_\varepsilon^2\right) \\
&= (2\pi)^{-1}\left(\sigma_\eta^2 + 2(1-\cos\lambda)\,\sigma_\varepsilon^2\right)
\end{aligned}$$

For Model (M2), we have:

$$\begin{aligned}
(1-L)\, y_t &= (\mu_t + \varepsilon_t) - (\mu_{t-1} + \varepsilon_{t-1}) \\
&= (\mu_t - \mu_{t-1}) + (\varepsilon_t - \varepsilon_{t-1}) \\
&= \beta_{t-1} + \eta_t + (\varepsilon_t - \varepsilon_{t-1})
\end{aligned}$$

$(1-L)\, y_t$ is not stationary because the process β_t is integrated of order 1. We have:

$$\begin{aligned}
(1-L)^2 y_t &= (1-L)\left(\beta_{t-1} + \eta_t + (\varepsilon_t - \varepsilon_{t-1})\right) \\
&= (\beta_{t-1} - \beta_{t-2}) + (\eta_t - \eta_{t-1}) + (1-L)^2 \varepsilon_t \\
&= \zeta_{t-1} + (1-L)\,\eta_t + (1-L)^2 \varepsilon_t
\end{aligned}$$

Since this is the sum of three independent stationary processes, the stationary form of y_t is equal to $S(y_t) = (1-L)^2 y_t$. We have[4]:

$$\begin{aligned}
\left|\left(1-e^{-i\lambda}\right)^2\right|^2 &= \left|1-e^{-i\lambda}\right|^2 \left|1-e^{-i\lambda}\right|^2 \\
&= \left(2\left(1-\cos\lambda\right)\right)^2 \\
&= 4\left(1-\cos\lambda\right)^2
\end{aligned}$$

We conclude that:

$$f_{S(y)}(\lambda) = \frac{\sigma_\zeta^2 + 2\left(1-\cos\lambda\right)\sigma_\eta^2 + 4\left(1-\cos\lambda\right)^2 \sigma_\varepsilon^2}{2\pi}$$

3. In Figure 10.3, we have represented the spectral density functions of Models (M1) and (M2) when $\sigma_\varepsilon = \sigma_\eta = \sigma_\zeta = 1$. We observe that they are similar for low frequencies, and the difference between the two processes comes from the dynamics on high frequencies.

4. μ_t is the stochastic trend, β_t is an AR(1) process that can be viewed as a mean-reverting component when $\phi < 0$ and γ_t is a stochastic seasonal process. When $\sigma_\omega = 0$, we have:

$$\gamma_{t-s+1} + \ldots + \gamma_{t-1} + \gamma_t = 0$$

[4]Another way to find this result is to notice that $(1-L)^2 = 1 - 2L + L^2$. Therefore, we have:

$$\begin{aligned}
\left|\left(1-e^{-i\lambda}\right)^2\right|^2 &= \left|1-2e^{-i\lambda}+\left(e^{-i\lambda}\right)^2\right|^2 \\
&= \left|1-2e^{-i\lambda}+e^{-2i\lambda}\right|^2 \\
&= \left|(1-2\cos\lambda+\cos 2\lambda) + i\,(2\sin\lambda - \sin 2\lambda)\right|^2 \\
&= (1-2\cos\lambda+\cos 2\lambda)^2 + (2\sin\lambda-\sin 2\lambda)^2 \\
&= 1 - 4\cos\lambda + 4\cos^2\lambda + 2\cos 2\lambda - 4\cos\lambda\cos 2\lambda + \cos^2 2\lambda + \\
&\quad\; 4\sin^2\lambda - 4\sin\lambda\sin 2\lambda + \sin^2 2\lambda \\
&= 6 - 4\cos\lambda + 2\cos 2\lambda - 4\left(\cos\lambda\cos 2\lambda + \sin\lambda\sin 2\lambda\right) \\
&= 6 - 4\cos\lambda + 2\cos 2\lambda - 4\cos\left(\lambda - 2\lambda\right) \\
&= 6 - 8\cos\lambda + 2\cos 2\lambda \\
&= 4 - 8\cos\lambda + 2\left(1+\cos 2\lambda\right) \\
&= 4 - 8\cos\lambda + 2\left(1+\cos^2\lambda - \sin^2\lambda\right) \\
&= 4 - 8\cos\lambda + 4\cos^2\lambda \\
&= 4\left(1-\cos\lambda\right)^2
\end{aligned}$$

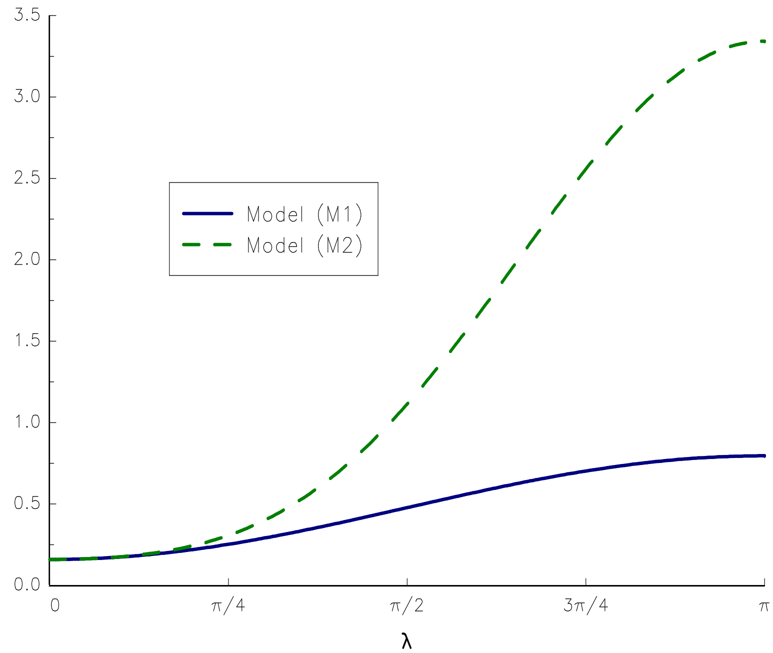

FIGURE 10.3: Spectral density function of Models (M1) and (M2)

Since we have $\gamma_{t-s} + \ldots + \gamma_{t-2} + \gamma_{t-1} = 0$, we deduce that:

$$\begin{aligned} \gamma_t &= -(\gamma_{t-s+1} + \ldots + \gamma_{t-1}) \\ &= \gamma_{t-s} \end{aligned}$$

We obtain a deterministic seasonal time series, where s represents the period length of a season. For example, if $s = 4$, we obtain:

$$\begin{cases} \gamma_t = \gamma_{t-4} = \gamma_{t-8} = \ldots \\ \gamma_{t+1} = \gamma_{t-3} = \gamma_{t-7} = \ldots \\ \gamma_{t+2} = \gamma_{t-2} = \gamma_{t-6} = \ldots \\ \gamma_{t+3} = \gamma_{t-1} = \gamma_{t-5} = \ldots \end{cases}$$

The process repeats every four time periods. If $\sigma_\omega \neq 0$, we have $\gamma_{t-s+1} + \ldots + \gamma_{t-1} + \gamma_t = \omega_t$ and $\gamma_{t-s} + \ldots + \gamma_{t-2} + \gamma_{t-1} = \omega_{t-1}$. Therefore, we have:

$$\begin{aligned} \gamma_t &= \omega_t - (\gamma_{t-s+1} + \ldots + \gamma_{t-1}) \\ &= \omega_t - (\omega_{t-1} - \gamma_{t-s}) \\ &= \gamma_{t-s} + (\omega_t - \omega_{t-1}) \end{aligned}$$

We deduce that:

$$\begin{aligned} \mathbb{E}_{t-s}[\gamma_t] &= \mathbb{E}_{t-s}[\gamma_{t-s} + (\omega_t - \omega_{t-1})] \\ &= \gamma_{t-s} \end{aligned}$$

It follows that γ_t is a stochastic seasonal process.

5. We have:

$$\begin{aligned} z_t &= (1-L)(1-L^s)\, y_t \\ &= (1-L^s)\,\eta_t + (1-L)(1-L^s)\,\beta_t + (1-L)(1-L^s)\,\varepsilon_t + \\ &\quad (1-L)(1-L^s)\,\gamma_t \end{aligned}$$

and:

$$\begin{aligned} (1-L)(1-L^s)\,\gamma_t &= (\gamma_t - \gamma_{t-1}) - (\gamma_{t-s} - \gamma_{t-s-1}) \\ &= (\gamma_t - \gamma_{t-s}) - (\gamma_{t-1} - \gamma_{t-s-1}) \\ &= (\omega_t - \omega_{t-1}) - (\omega_{t-1} - \omega_{t-2}) \\ &= \omega_t - 2\omega_{t-1} + \omega_{t-2} \\ &= (1-L)^2\,\omega_t \end{aligned}$$

We deduce that:

$$\begin{aligned} z_t &= (1-L^s)\,\eta_t + (1-L)(1-L^s)\,\beta_t + \\ &\quad (1-L)(1-L^s)\,\varepsilon_t + (1-L)^2\,\omega_t \end{aligned}$$

If we assume that $|\phi| < 1$, then β_t is stationary. Moreover, we know that η_t, ε_t and ω_t are stationary. We conclude that z_t is stationary and $\mathcal{S}(y_t) = (1-L)(1-L^s)\, y_t$ is a stationary form of y_t.

6. Another stationary form of y_t is $(1-L^s)\, y_t$. Indeed, we have:

$$(1-L^s)\, y_t = (1-L^s)\,\mu_t + (1-L^s)\,\beta_t + (1-L^s)\,\varepsilon_t + (1-L)\,\omega_t$$

and:

$$\begin{aligned} (1-L^s)\,\mu_t &= \mu_t - \mu_{t-s} \\ &= (\mu_{t-1} + \eta_t) - \mu_{t-s} \\ &= \eta_t + (\mu_{t-2} + \eta_{t-1}) - \mu_{t-s} \\ &= \eta_t + \eta_{t-1} + \ldots + \eta_{t-s-1} \end{aligned}$$

We deduce that $(1-L^s)\,\mu_t$ and $(1-L^s)\, y_t$ are stationary.

7. We note $g_\lambda(\varphi(L)) = \left|\varphi\left(e^{-i\lambda}\right)\right|^2$. We have:

$$\begin{aligned} g_\lambda(1-L^s) &= \left|1 - \left(e^{-i\lambda}\right)^s\right|^2 \\ &= \left|1 - e^{-is\lambda}\right|^2 \\ &= (1-\cos s\lambda)^2 + \sin^2 s\lambda \\ &= 2(1-\cos s\lambda) \end{aligned}$$

and:

$$\begin{aligned} g_\lambda((1-L)(1-L^s)) &= g_\lambda(1-L) \cdot g_\lambda(1-L^s) \\ &= 4(1-\cos\lambda)(1-\cos s\lambda) \end{aligned}$$

We remind that

$$\begin{aligned} g_\lambda\left((1-\phi L)^{-1}\right) &= \frac{1}{\left|1-\phi e^{-i\lambda}\right|^2} \\ &= \frac{1}{(1-\phi\cos\lambda)^2+\phi^2\sin^2\lambda} \\ &= \frac{1}{1-2\phi\cos\lambda+\phi^2} \end{aligned}$$

We deduce that:

$$\begin{aligned} 2\pi f_{\mathcal{S}(y)}(\lambda) &= g_\lambda\left(1-L^s\right)\sigma_\eta^2 + g_\lambda\left(\frac{(1-L)\left(1-L^s\right)}{1-\phi L}\right)\sigma_\zeta^2 + \\ & \quad g_\lambda\left((1-L)\left(1-L^s\right)\right)\sigma_\varepsilon^2 + g_\lambda\left((1-L)^2\right)\sigma_\omega^2 \\ &= 2\left(1-\cos s\lambda\right)\sigma_\eta^2 + \\ & \quad \left(\frac{4\left(1-\cos\lambda\right)\left(1-\cos s\lambda\right)}{1-2\phi\cos\lambda+\phi^2}\right)\sigma_\zeta^2 + \\ & \quad 4\left(1-\cos\lambda\right)\left(1-\cos s\lambda\right)\sigma_\varepsilon^2 \\ & \quad \left(4-8\cos\lambda+4\cos^2\lambda\right)\sigma_\omega^2 \end{aligned}$$

We have seen that:

$$\begin{aligned} g_\lambda\left((1-L)\left(1-L^s\right)\right) &= g_\lambda\left(1-L-L^s+L^{s+1}\right) \\ &= 4-4\cos\lambda-4\cos s\lambda + \\ & \quad 2\cos(s-1)\lambda + 2\cos(s-1)\lambda \end{aligned}$$

By using the properties of trigonometric functions, we obtain:

$$\begin{aligned} g_\lambda\left((1-L)\left(1-L^s\right)\right) &= 4-4\cos\lambda-4\cos s\lambda + \\ & \quad 2\left(\cos s\lambda\cos\lambda-\sin s\lambda\sin\lambda\right) + \\ & \quad 2\left(\cos s\lambda\cos\lambda+\sin s\lambda\sin\lambda\right) \\ &= 4-4\cos\lambda-4\cos s\lambda+4\cos s\lambda\cos\lambda \\ &= 4\left(1-\cos\lambda\right)\left(1-\cos s\lambda\right) \end{aligned}$$

The spectral density function is then defined as follows:

$$\begin{aligned} f_{\mathcal{S}(y)}(\lambda) &= \pi^{-1}\left(1-\cos s\lambda\right)\sigma_\eta^2 + \\ & \quad \pi^{-1}\left(\frac{2-2\cos\lambda+\sum_{j=-1}^{1}\left(3\left|j\right|-2\right)\cos\left(s+j\right)\lambda}{1-2\phi\cos\lambda+\phi^2}\right)\sigma_\zeta^2 + \\ & \quad 2\pi^{-1}\left(1-\cos\lambda\right)\left(1-\cos s\lambda\right)\sigma_\varepsilon^2 \\ & \quad 2\pi^{-1}\left(1-2\cos\lambda+\cos^2\lambda\right)\sigma_\omega^2 \end{aligned}$$

10.3.16 Spectral density function of some processes

We note $g_\lambda\left(\varphi(L)\right) = \left|\varphi\left(e^{-i\lambda}\right)\right|^2$.

1. We have:

$$\begin{aligned} g_\lambda (1 - L^s) &= \left|1 - \left(e^{-i\lambda}\right)^s\right|^2 \\ &= \left|1 - e^{-is\lambda}\right|^2 \\ &= (1 - \cos s\lambda)^2 + \sin^2 s\lambda \\ &= 2(1 - \cos s\lambda) \end{aligned}$$

Since we have $(1 - L^s) y_t = \varepsilon_t$, we deduce that:

$$\begin{aligned} f_y(\lambda) &= \frac{\sigma_\varepsilon^2}{2\pi g_\lambda (1 - L^s)} \\ &= \frac{\sigma_\varepsilon^2}{4\pi (1 - \cos(s\lambda))} \end{aligned}$$

2. We have:

$$\begin{aligned} f(\lambda) &= \frac{\sigma_\varepsilon^2}{2\pi \left|\left(1 - e^{-i\lambda}\right)^d\right|^2} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \left|1 - e^{-i\lambda}\right|^{-2d} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \left((1 - \cos\lambda)^2 + \sin^2\lambda\right)^{-d} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} (2(1 - \cos\lambda))^{-d} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \left(4\sin^2\frac{\lambda}{2}\right)^{-d} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \left(2\sin\frac{\lambda}{2}\right)^{-2d} \end{aligned}$$

because[5]:

$$\begin{aligned} \sin^2\frac{\lambda}{2} &= \frac{1}{2}\left(\cos\left(\frac{\lambda}{2} - \frac{\lambda}{2}\right) - \cos\left(\frac{\lambda}{2} + \frac{\lambda}{2}\right)\right) \\ &= \frac{1}{2}(1 - \cos\lambda) \end{aligned}$$

3. We have:

$$z_t = (1 - \phi L)^{-1} u_t + (1 - \theta L)^{-1} v_t$$

We deduce that:

$$f_z(\lambda) = \frac{\sigma_u^2}{2\pi (1 - 2\phi\cos\lambda + \phi^2)} + \frac{\left(1 - 2\theta\cos\lambda + \theta^2\right)\sigma_u^2}{2\pi}$$

[5]We use the following trigonometric identity:

$$\sin\alpha\sin\beta = \frac{1}{2}(\cos(\alpha - \beta) - \cos(\alpha + \beta))$$

(a) The simulated time series z is represented in the first panel Figure 10.4. In the second panel, we also give the periodogram of z:

$$I_z(\lambda_j) = \frac{|d_z(\lambda_j)|^2}{2\pi n} = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} z_t e^{-i\lambda_j t}\right|^2$$

where $\lambda_j = 2\pi(j-1)/n$ and $j \in \{1, \ldots, n\}$.

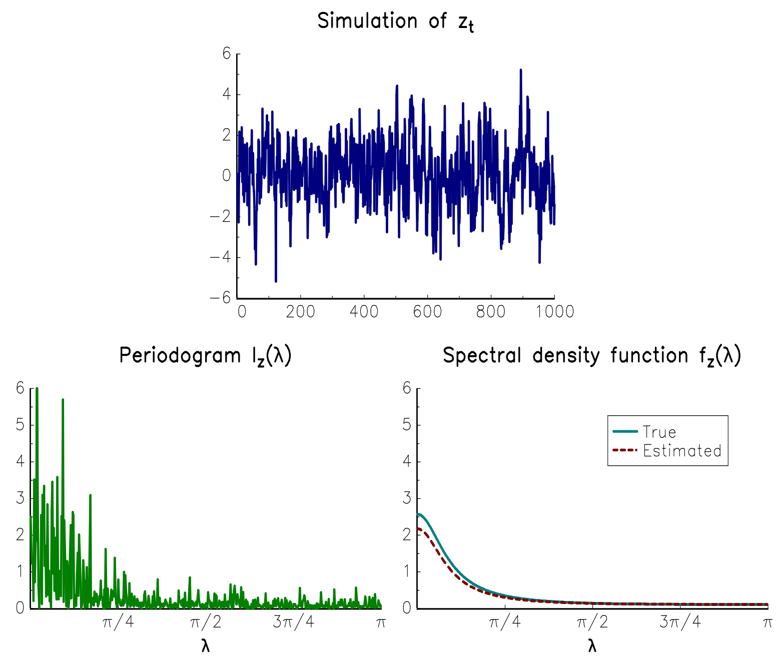

FIGURE 10.4: The AR(1) + MA(1) stochastic process

(b) The Whittle log-likelihood is equal to:

$$\ell(\phi, \sigma_u, \theta, \sigma_v) \simeq -n \ln 2\pi - \frac{1}{2}\sum_{j=1}^{n} \ln f_z(\lambda_j) - \frac{1}{2}\sum_{j=1}^{n} \frac{I_z(\lambda_j)}{f_z(\lambda_j)}$$

where $\lambda_j = 2\pi j/n$ et $j \in \{0, 1, \ldots, n-1\}$. With the simulation in Figure 10.4, we obtain the following estimates: $\hat{\phi} = 0.755$, $\hat{\sigma}_u = 0.896$, $\hat{\theta} = 0.120$ and $\hat{\sigma}_v = 0.595$. The true and estimation spectral density functions are given in the third panel in 10.4.

Chapter 11

Copulas and Dependence

11.4.1 Gumbel logistic copula

1. We recall that the expression of the Gumbel logistic copula is:

$$\mathbf{C}(u_1, u_2) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}$$

We have:

$$\begin{aligned} \partial_1 \mathbf{C}(u_1, u_2) &= \frac{u_2 (u_1 + u_2 - u_1 u_2) - u_1 u_2 (1 - u_2)}{(u_1 + u_2 - u_1 u_2)^2} \\ &= \frac{u_2^2}{(u_1 + u_2 - u_1 u_2)^2} \end{aligned}$$

We deduce that the copula density is:

$$\begin{aligned} c(u_1, u_2) &= \partial_{1,2}^2 \mathbf{C}(u_1, u_2) \\ &= \frac{2u_2 (u_1 + u_2 - u_1 u_2)^2 - 2u_2^2 (u_1 + u_2 - u_1 u_2)(1 - u_1)}{(u_1 + u_2 - u_1 u_2)^4} \\ &= \frac{2u_1 u_2}{(u_1 + u_2 - u_1 u_2)^3} \end{aligned}$$

2. We have:

$$\begin{aligned} \lambda^+(u) &= \frac{1 - 2u + \mathbf{C}(u, u)}{1 - u} \\ &= \frac{(1 - 2u)(2 - u) + u}{(1 - u)(2 - u)} \\ &= \frac{2u^2 - 4u + 2}{u^2 - 3u + 2} \end{aligned}$$

Using L'Hospital's rule, it follows that:

$$\begin{aligned} \lambda^+ &= \lim_{u \to 1} \frac{2u^2 - 4u + 2}{u^2 - 3u + 2} \\ &= \lim_{u \to 1} \frac{4u - 4}{2u - 3} \\ &= 0 \end{aligned}$$

The Gumbel logistic copula has then no upper tail dependence. For the lower tail dependence, we obtain:

$$\begin{aligned}\lambda^{+} &= \lim_{u\to 0}\frac{\mathbf{C}(u,u)}{u} \\ &= \lim_{u\to 0}\frac{u}{2u-u^2} \\ &= \lim_{u\to 0}\frac{1}{2-2u} \\ &= \frac{1}{2}\end{aligned}$$

We verify that it has a lower tail dependence.

11.4.2 Farlie-Gumbel-Morgenstern copula

1. We have:

$$\begin{aligned}\mathbf{C}(u,0) &= \mathbf{C}(0,u)=0 \\ \mathbf{C}(u,1) &= \mathbf{C}(1,u)=u \\ \frac{\partial^2\,\mathbf{C}(u_1,u_2)}{\partial\, u_1\partial\, u_2} &= 1+\theta(1-2u_1)(1-2u_2)\end{aligned}$$

As we have $-1\leq 1+\theta(1-2u_1)(1-2u_2)\leq 1$, it follows that:

$$\frac{\partial^2\mathbf{C}(u_1,u_2)}{\partial\, u_1\partial\, u_2}\geq 0$$

We deduce that $\mathbf{C}$ is a copula function.

2. We have:

$$\begin{aligned}\lambda &= \lim_{u\to 1^-}\frac{1-2u+\mathbf{C}(u,u)}{1-u} \\ &= \lim_{u\to 1^-}\frac{1-2u+u^2\left(1+\theta(1-u)^2\right)}{1-u} \\ &= \lim_{u\to 1^-}(1-u)\left(1+\theta u^2\right) \\ &= 0\end{aligned}$$

For the Kendall's tau, we obtain:

$$\tau =$$

The Spearman's rho is equal to:

$$\varrho =$$

3. We calculate the conditional copula:

We simulate (U_1,U_2) in the following way:

By applying the PIT method, we obtain:

$$\begin{aligned}x_1 &= \mu+\sigma\Phi^{-1}(u_1) \\ x_2 &= -\frac{1}{\lambda}\ln(1-u_2)\end{aligned}$$

4. We have:

$$\begin{aligned} c(u_1,u_2) &= \frac{\partial^2 \mathbf{C}(u_1,u_2)}{\partial u_1 \partial u_2} \\ &= 1+\theta(1-2u_1)(1-2u_2) \end{aligned}$$

It follows that:

$$\begin{aligned} f(x_1,x_2) &= c(\mathbf{F}_1(x_1),\mathbf{F}_2(x_2)) \times f_2(x_1) \times f_1(x_1) \\ &= \frac{\lambda}{\sigma}\left(1+\theta\left(1-2\Phi\left(\frac{x_1-\mu}{\sigma}\right)\right)\left(2e^{-\lambda x_2}-1\right)\right) \times \\ & \phi\left(\frac{x_1-\mu}{\sigma}\right) \times e^{-\lambda x_2} \end{aligned}$$

We deduce that :

$$\begin{aligned} \boldsymbol{\ell} &= n\ln\lambda - \frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi + \\ & \sum_{i=1}^{n}\ln\left(1+\theta\left(1-2\Phi\left(\frac{x_{1,i}-\mu}{\sigma}\right)\right)\left(2e^{-\lambda x_{2,i}}-1\right)\right) - \\ & \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_{1,i}-\mu}{\sigma}\right)^2 - \lambda\sum_{i=1}^{n}x_{2,i} \end{aligned}$$

11.4.3 Survival copula

1. We have $\mathbf{S}(0,0)=1$ and $\mathbf{S}(\infty,\infty)=0$. We notice that:

$$\begin{aligned} \partial^2_{1,2}\mathbf{S}(x_1,x_2) &= \\ &\leq 0 \end{aligned}$$

We conclude that $\mathbf{S}$ is a survival function.

2. We have:

$$\begin{aligned} \mathbf{S}_1(x_1) &= \mathbf{S}(x_1,0) \\ &= \exp(-x_1) \end{aligned}$$

By noting $U_1 = \mathbf{S}_1(X_1)$, we deduce the expression of the survival copula:

$$\begin{aligned} \mathbf{C}(u_1,u_2) &= \exp\left(-\left(-\ln u_1 - \ln u_2 + \theta\frac{\ln u_1 \ln u_2}{\ln u_1 + \ln u_2}\right)\right) \\ &= u_1u_2\exp\left(\theta\frac{\tilde{u}_1\tilde{u}_2}{\tilde{u}_1+\tilde{u}_2}\right) \end{aligned}$$

with $\tilde{u} = -\ln u$.

11.4.4 Method of moments

$$\mathbf{C}_{\langle X_1,X_2\rangle} = \theta\mathbf{C}^- + (1-\theta)\mathbf{C}^+$$

1. We have:

$$\mathbf{F}(x_1, x_2) = \theta \times \max(\Phi(x_1) + \Phi(x_2) - 1, 0) + (1 - \theta) \times \min(\Phi(x_1), \Phi(x_2))$$

   It follows that:

$$\begin{aligned}
\mathbb{E}[X_1 X_2] &= \iint x_1 x_2 \,\mathrm{d}\mathbf{F}(x_1, x_2) \\
&= \theta \times \iint x_1 x_2 \,\mathrm{d}\mathbf{C}^{-}(\Phi(x_1), \Phi(x_1)) + \\
&\quad (1 - \theta) \times \iint x_1 x_2 \,\mathrm{d}\mathbf{C}^{+}(\Phi(x_1), \Phi(x_1)) \\
&= \theta \times (-1) + (1 - \theta) \times (+1) \\
&= 1 - 2\theta
\end{aligned}$$

   We deduce that:

$$\rho\langle X_1, X_2\rangle = \mathbb{E}[X_1 X_2] = 1 - 2\theta$$

   The linear correlation between X_1 and X_2 is equal to zero when θ takes the value $1/2$.

2. Using the notations $N_1 \sim \mathcal{N}(0,1)$ and $N_2 \sim \mathcal{N}(0,1)$, we obtain:

$$\begin{aligned}
\rho\langle X_1, X_2\rangle &= \rho\langle \mu_1 + \sigma_1 N_1, \mu_2 + \sigma_2 N_2\rangle \\
&= \rho\langle N_1, N_2\rangle \\
&= 1 - 2\theta
\end{aligned}$$

3. We have:

$$\theta = \frac{1 - \rho\langle X_1, X_2\rangle}{2}$$

   The MM estimator $\hat{\theta}_{\mathrm{MM}}$ is then equal to:

$$\hat{\theta}_{\mathrm{MM}} = \frac{1 - \hat{\rho}}{2}$$

   where $\hat{\rho}$ is the empirical correlation between X_1 and X_2.

11.4.5 Correlated loss given default rates

1. As we have $x \in [0,1]$, the parameter γ must be positive or equal to zero in order to have $\mathbf{F}(0) = 0$, $\mathbf{F}(1) = 1$ and $f(x) = \gamma x^{\gamma - 1} \geq 0$.

2. The expression of the log-likelihood function is:

$$\begin{aligned}
\boldsymbol{\ell}(\gamma) &= \sum_{i=1}^{n} \ln f(x_i) \\
&= \sum_{i=1}^{n} \ln\left(\gamma x_i^{\gamma - 1}\right) \\
&= n \ln \gamma + (\gamma - 1) \sum_{i=1}^{n} \ln x_i
\end{aligned}$$

We deduce the first-order condition:

$$\frac{\partial \boldsymbol{\ell}(\gamma)}{\partial \gamma} = 0 \Leftrightarrow \frac{n}{\gamma} + \sum_{i=1}^{n} \ln x_i = 0$$

We finally obtain the ML estimator:

$$\hat{\gamma}_{\mathrm{ML}} = -\frac{1}{\left(n^{-1}\sum_{i=1}^{n} \ln x_i\right)}$$

3. We have:

$$\begin{aligned} \mathbb{E}[\mathrm{LGD}] &= \int_0^1 x\gamma x^{\gamma-1}\,\mathrm{d}x \\ &= \gamma \int_0^1 x^{\gamma}\,\mathrm{d}x \\ &= \gamma \left[\frac{x^{\gamma+1}}{\gamma+1}\right]_0^1 \\ &= \frac{\gamma}{\gamma+1} \end{aligned}$$

Let $\bar{x}$ be the empirical mean of the sample $\{x_1, \ldots, x_n\}$. The MM estimator $\hat{\gamma}_{\mathrm{MM}}$ satisfies the following equation:

$$\frac{\hat{\gamma}_{\mathrm{MM}}}{\hat{\gamma}_{\mathrm{MM}} + 1} = \bar{x}$$

We deduce that:

$$\begin{aligned} \hat{\gamma}_{\mathrm{MM}} &= \frac{\bar{x}}{1-\bar{x}} \\ &= \frac{\sum_{i=1}^{n} x_i}{n - \sum_{i=1}^{n} x_i} \end{aligned}$$

4. In the case $x_i = 50\%$, we obtain:

$$\hat{\gamma}_{\mathrm{ML}} = -\frac{1}{\ln 0.5} = \ln 2 = 1.44$$

and:

$$\hat{\gamma}_{\mathrm{MM}} = 0.5/(1-0.5) = 1.00$$

The numerical results are different. For example, we have reported the density function of the two probability distributions in Figure 11.1.

5. We have:

$$\frac{\partial\, \mathbf{C}(u_1, u_2)}{\partial\, u_1} = u_2 e^{-\theta \ln u_1 \ln u_2} - \theta u_2 \ln u_2 e^{-\theta \ln u_1 \ln u_2}$$

and:

$$\begin{aligned} \frac{\partial^2\, \mathbf{C}(u_1, u_2)}{\partial\, u_1 \partial\, u_2} &= e^{-\theta \ln u_1 \ln u_2} - \theta \ln u_1 e^{-\theta \ln u_1 \ln u_2} - \theta \ln u_2 e^{-\theta \ln u_1 \ln u_2} - \\ &\quad \theta e^{-\theta \ln u_1 \ln u_2} + \theta^2 \ln u_1 \ln u_2 e^{-\theta \ln u_1 \ln u_2} \\ &= \left(1 - \theta - \theta \ln(u_1 u_2) + \theta^2 \ln u_1 \ln u_2\right) e^{-\theta \ln u_1 \ln u_2} \end{aligned}$$

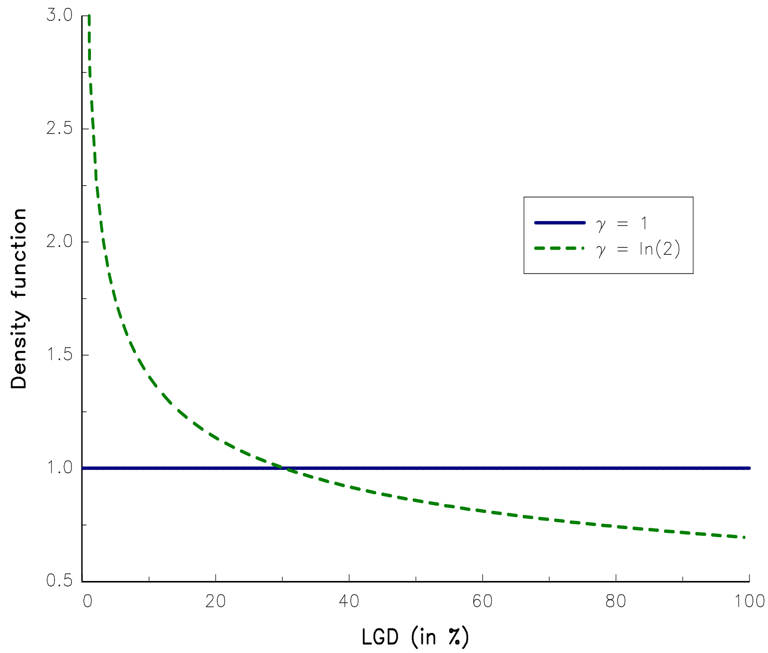

FIGURE 11.1: Density functions associated to ML and MM estimators

6. The bivariate density function is the equal to:

$$
\begin{aligned}
f(x,y) &= c(\mathbf{F}(x), \mathbf{F}(y)) \times f(x) \times f(y) \\
&= \left(1 - \theta - \theta(\gamma_1 \ln x + \gamma_2 \ln y) + \theta^2 \gamma_1 \gamma_2 \ln x \ln y\right) \times \\
&\quad e^{-\theta \gamma_1 \gamma_2 \ln x \ln y} \times \gamma_1 x^{\gamma_1 - 1} \times \gamma_2 y^{\gamma_2 - 1}
\end{aligned}
$$

where γ_1 and γ_2 are the parameters associated to the risk classes $\mathcal{C}_1$ and $\mathcal{C}_2$. It follows that the log-likelihood function is equal to:

$$
\begin{aligned}
\boldsymbol{\ell} &= n \ln \gamma_1 + n \ln \gamma_2 + (\gamma_1 - 1) \sum_{i=1}^{n} \ln x_i + (\gamma_2 - 1) \sum_{i=1}^{n} \ln y_i + \\
&\quad \sum_{i=1}^{n} \ln \left(1 - \theta - \theta(\gamma_1 \ln x_i + \gamma_2 \ln y_i) + \theta^2 \gamma_1 \gamma_2 \ln x_i \ln y_i\right) - \\
&\quad \theta \gamma_1 \gamma_2 \sum_{i=1}^{n} \ln x_i \ln y_i
\end{aligned}
$$

7. The first-order conditions are:

$$\begin{aligned}
\frac{\partial \boldsymbol{\ell}}{\partial \gamma_1} &= \frac{n}{\gamma_1} + \sum_{i=1}^{n} \ln x_i + g_1(\gamma_1, \gamma_2, \theta) \\
\frac{\partial \boldsymbol{\ell}}{\partial \gamma_2} &= \frac{n}{\gamma_2} + \sum_{i=1}^{n} \ln y_i + g_2(\gamma_1, \gamma_2, \theta) \\
\frac{\partial \boldsymbol{\ell}}{\partial \theta} &= -\sum_{i=1}^{n} \frac{1 + (\gamma_1 \ln x_i + \gamma_2 \ln y_i) - 2\theta\gamma_1\gamma_2 \ln x_i \ln y_i}{1 - \theta - \theta(\gamma_1 \ln x_i + \gamma_2 \ln y_i) + \theta^2 \gamma_1 \gamma_2 \ln x_i \ln y_i} - \\
&\quad \gamma_1\gamma_2 \sum_{i=1}^{n} \ln x_i \ln y_i
\end{aligned}$$

with:

$$\begin{aligned}
g_1(\gamma_1, \gamma_2, \theta) &= \sum_{i=1}^{n} \frac{\left(\theta^2 \gamma_2 \ln y_i - \theta\right) \ln x_i}{1 - \theta - \theta(\gamma_1 \ln x_i + \gamma_2 \ln y_i) + \theta^2 \gamma_1 \gamma_2 \ln x_i \ln y_i} - \\
&\quad \theta\gamma_2 \sum_{i=1}^{n} \ln x_i \ln y_i \\
g_2(\gamma_1, \gamma_2, \theta) &= \sum_{i=1}^{n} \frac{\left(\theta^2 \gamma_1 \ln x_i - \theta\right) \ln y_i}{1 - \theta - \theta(\gamma_1 \ln x_i + \gamma_2 \ln y_i) + \theta^2 \gamma_1 \gamma_2 \ln x_i \ln y_i} - \\
&\quad \theta\gamma_1 \sum_{i=1}^{n} \ln x_i \ln y_i
\end{aligned}$$

When $\hat{\theta}$ is equal to zero, we have $g_1(\gamma_1, \gamma_2, 0) = 0$. In this case, the estimator $\hat{\gamma}_1$ corresponds to the ML estimator $\hat{\gamma}_{\mathrm{ML}}$. When we have $\hat{\theta} \neq 0$, we obtain $g_1(\gamma_1, \gamma_2, \theta) \neq 0$ and $\hat{\gamma}_1 \neq \hat{\gamma}_{\mathrm{ML}}$. We obtain this result because more information is available in the bivariate case. The ML method can then correct the estimator $\hat{\gamma}_{\mathrm{ML}}$ by taking into account the dependence function between LGD_1 and LGD_2. For instance, if the estimated copula is equal to the Fréchet upper copula $\mathbf{C}^+$, it is obvious that the two estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are equal, even if the unidimensional ML estimators are not necessarily equal. Let us consider the following sample:

LGD_1 (in %)	50	40	60	50	80	90	70	10	40	40
LGD_2 (in %)	60	50	80	70	80	90	80	30	50	70

We obtain $\hat{\gamma}_{\mathrm{ML}} = 1.31$ for $\mathcal{C}_1$ and $\hat{\gamma}_{\mathrm{ML}} = 2.18$ for $\mathcal{C}_2$. With the bivariate ML method, we obtain $\hat{\gamma}_1 = 0.88$, $\hat{\gamma}_2 = 1.44$ and $\hat{\theta} = 1.71$.

11.4.6 Calculation of correlation bounds

1. We have:

$$\begin{aligned}
\mathbf{C}^-(u_1, u_2) &= \max(u_1 + u_2 - 1, 0) \\
\mathbf{C}^\perp(u_1, u_2) &= u_1 u_2 \\
\mathbf{C}^+(u_1, u_2) &= \min(u_1, u_2)
\end{aligned}$$

Let X_1 and X_2 be two random variables. We have:

(i) $\mathbf{C}\langle X_1, X_2\rangle = \mathbf{C}^-$ if and only if there exists a non-increasing function f such that we have $X_2 = f(X_1)$;

(ii) $\mathbf{C}\langle X_1, X_2\rangle = \mathbf{C}^\perp$ if and only if X_1 and X_2 are independent;

(iii) $\mathbf{C}\langle X_1, X_2\rangle = \mathbf{C}^+$ if and only if there exists a non-decreasing function f such that we have $X_2 = f(X_1)$.

2. We note $U_1 = 1 - \exp(-\lambda\boldsymbol{\tau})$ and $U_2 = \text{LGD}$.

(a) The dependence between $\boldsymbol{\tau}$ and LGD is maximum when we have $\mathbf{C}\langle\boldsymbol{\tau}, \text{LGD}\rangle = \mathbf{C}^+$. Since we have $U_1 = U_2$, we conclude that:

$$\text{LGD} + e^{-\lambda\boldsymbol{\tau}} - 1 = 0$$

(b) We know that:

$$\rho\langle\boldsymbol{\tau}, \text{LGD}\rangle \in [\rho_{\min}\langle\boldsymbol{\tau}, \text{LGD}\rangle, \rho_{\max}\langle\boldsymbol{\tau}, \text{LGD}\rangle]$$

where $\rho_{\min}\langle\boldsymbol{\tau}, \text{LGD}\rangle$ (resp. $\rho_{\max}\langle\boldsymbol{\tau}, \text{LGD}\rangle$) is the linear correlation corresponding to the copula $\mathbf{C}^-$ (resp. $\mathbf{C}^+$). It comes that:

$$\mathbb{E}[\boldsymbol{\tau}] = \sigma(\boldsymbol{\tau}) = \frac{1}{\lambda}$$

and:

$$\begin{aligned}\mathbb{E}[\text{LGD}] &= \frac{1}{2} \\ \sigma(\text{LGD}) &= \sqrt{\frac{1}{12}}\end{aligned}$$

In the case $\mathbf{C}\langle\boldsymbol{\tau}, \text{LGD}\rangle = \mathbf{C}^-$, we have $U_1 = 1 - U_2$. It follows that $\text{LGD} = e^{-\lambda\boldsymbol{\tau}}$. We have:

$$\begin{aligned}\mathbb{E}[\boldsymbol{\tau}\,\text{LGD}] &= \mathbb{E}\left[\boldsymbol{\tau} e^{-\lambda\boldsymbol{\tau}}\right] \\ &= \int_0^\infty t e^{-\lambda t} \lambda e^{-\lambda t}\,\mathrm{d}t \\ &= \int_0^\infty t\lambda e^{-2\lambda t}\,\mathrm{d}t \\ &= \left[-\frac{te^{-2\lambda t}}{2}\right]_0^\infty + \frac{1}{2}\int_0^\infty e^{-2\lambda t}\,\mathrm{d}t \\ &= 0 + \frac{1}{2}\left[-\frac{e^{-2\lambda t}}{2\lambda}\right]_0^\infty \\ &= \frac{1}{4\lambda}\end{aligned}$$

We deduce that:

$$\begin{aligned}\rho_{\min}\langle\boldsymbol{\tau}, \text{LGD}\rangle &= \left(\frac{1}{4\lambda} - \frac{1}{2\lambda}\right) \Big/ \left(\frac{1}{\lambda}\sqrt{\frac{1}{12}}\right) \\ &= -\frac{\sqrt{3}}{2}\end{aligned}$$

In the case $\mathbf{C}\langle\boldsymbol{\tau},\mathrm{LGD}\rangle = \mathbf{C}^{+}$, we have $\mathrm{LGD} = 1 - e^{-\lambda\boldsymbol{\tau}}$. We have:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\tau}\,\mathrm{LGD}] &= \mathbb{E}\left[\boldsymbol{\tau}\left(1-e^{-\lambda\boldsymbol{\tau}}\right)\right] \\
&= \int_0^\infty t\left(1-e^{-\lambda t}\right)\lambda e^{-\lambda t}\,\mathrm{d}t \\
&= \int_0^\infty t\lambda e^{-\lambda t}\,\mathrm{d}t - \int_0^\infty t\lambda e^{-2\lambda t}\,\mathrm{d}t \\
&= \left(\left[-te^{-\lambda t}\right]_0^\infty + \int_0^\infty e^{-\lambda t}\,\mathrm{d}t\right) - \frac{1}{4\lambda} \\
&= 0 + \left[-\frac{e^{-\lambda t}}{\lambda}\right]_0^\infty - \frac{1}{4\lambda} \\
&= \frac{3}{4\lambda}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\rho_{\max}\langle\boldsymbol{\tau},\mathrm{LGD}\rangle &= \left(\frac{3}{4\lambda}-\frac{1}{2\lambda}\right)\Big/\left(\frac{1}{\lambda}\sqrt{\frac{1}{12}}\right) \\
&= \frac{\sqrt{3}}{2}
\end{aligned}$$

We finally obtain the following result:

$$\left|\rho\langle\boldsymbol{\tau},\mathrm{LGD}\rangle\right| \leq \frac{\sqrt{3}}{2}$$

(c) We notice that $|\rho\langle\boldsymbol{\tau},\mathrm{LGD}\rangle|$ is lower than 86.6%, implying that the bounds -1 and $+1$ can not be reached.

3. (a) If the copula function of $(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ is the Fréchet upper bound copula, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are comonotone. We deduce that:

$$U_1 = U_2 \Longleftrightarrow 1 - e^{-\lambda_1\boldsymbol{\tau}_1} = 1 - e^{-\lambda_2\boldsymbol{\tau}_2}$$

and:

$$\boldsymbol{\tau}_1 = \frac{\lambda_2}{\lambda_1}\boldsymbol{\tau}_2$$

(b) We have $U_1 = 1 - U_2$. It follows that $\mathbf{S}_1(\boldsymbol{\tau}_1) = 1 - \mathbf{S}_2(\boldsymbol{\tau}_2)$. We deduce that:

$$e^{-\lambda_1\boldsymbol{\tau}_1} = 1 - e^{-\lambda_2\boldsymbol{\tau}_2}$$

and:

$$\boldsymbol{\tau}_1 = \frac{-\ln\left(1-e^{-\lambda_2\boldsymbol{\tau}_2}\right)}{\lambda_1}$$

There exists then a function f such that $\boldsymbol{\tau}_1 = f(\boldsymbol{\tau}_2)$ with:

$$f(t) = \frac{-\ln\left(1-e^{-\lambda_2 t}\right)}{\lambda_1}$$

(c) Using Question 2(b), we known that $\rho \in [\rho_{\min}, \rho_{\max}]$ where $\rho_{\min}$ and $\rho_{\max}$ are the correlations of $(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ when the copula function is respectively $\mathbf{C}^{-}$ and $\mathbf{C}^{+}$.

We also know that $\rho = 1$ (resp. $\rho = -1$) if there exists a linear and increasing (resp. decreasing) function f such that $\boldsymbol{\tau}_1 = f(\boldsymbol{\tau}_2)$. When the copula is $\mathbf{C}^+$, we have $f(t) = \frac{\lambda_2}{\lambda_1} t$ and $f'(t) = \frac{\lambda_2}{\lambda_1} > 0$. As it is a linear and increasing function, we deduce that $\rho_{\max} = 1$. When the copula is $\mathbf{C}^-$, we have:

$$f(t) = \frac{-\ln\left(1 - e^{-\lambda_2 t}\right)}{\lambda_1}$$

and:

$$f'(t) = -\frac{\lambda_2 e^{-\lambda_2 t} \ln\left(1 - e^{-\lambda_2 t}\right)}{\lambda_1 \left(1 - e^{-\lambda_2 t}\right)} < 0$$

The function $f(t)$ is decreasing, but it is not linear. We deduce that $\rho_{\min} \neq -1$ and:

$$-1 < \rho \leq 1$$

(d) When the copula is $\mathbf{C}^-$, we know that there exists a decreasing function f such that $X_2 = f(X_1)$. We also know that the linear correlation reaches the lower bound -1 if the function f is linear:

$$X_2 = a + bX_1$$

This implies that $b < 0$. When X_1 takes the value $+\infty$, we obtain:

$$X_2 = a + b \times \infty$$

As the lower bound of X_2 is equal to zero 0, we deduce that $a = +\infty$. This means that the function $f(x) = a + bx$ does not exist. We conclude that the lower bound $\rho = -1$ can not be reached.

4. (a) $X_1 + X_2$ is a Gaussian random variable because it is a linear combination of the Gaussian random vector (X_1, X_2). We have:

$$\mathbb{E}[X_1 + X_2] = \mu_1 + \mu_2$$

and:

$$\operatorname{var}(X_1 + X_2) = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2$$

We deduce that:

$$X_1 + X_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)$$

(b) We have:

$$\begin{aligned} \operatorname{cov}(Y_1, Y_2) &= \mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_2]\,\mathbb{E}[Y_2] \\ &= \mathbb{E}\left[e^{X_1 + X_2}\right] - \mathbb{E}[Y_2]\,\mathbb{E}[Y_2] \end{aligned}$$

We know that $e^{X_1 + X_2}$ is a lognormal random variable. We deduce that:

$$\begin{aligned} \mathbb{E}\left[e^{X_1 + X_2}\right] &= \exp\left(\mathbb{E}[X_1 + X_2] + \frac{1}{2}\operatorname{var}(X_1 + X_2)\right) \\ &= \exp\left(\mu_1 + \mu_2 + \frac{1}{2}\left(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)\right) \\ &= e^{\mu_1 + \frac{1}{2}\sigma_1^2} e^{\mu_2 + \frac{1}{2}\sigma_2^2} e^{\rho\sigma_1\sigma_2} \end{aligned}$$

We finally obtain:

$$\operatorname{cov}(Y_1, Y_2) = e^{\mu_1 + \frac{1}{2}\sigma_1^2} e^{\mu_2 + \frac{1}{2}\sigma_2^2}\left(e^{\rho\sigma_1\sigma_2} - 1\right)$$

(c) We have:

$$\begin{aligned}\rho\langle Y_1,Y_2\rangle &= \frac{e^{\mu_1+\frac{1}{2}\sigma_1^2}e^{\mu_2+\frac{1}{2}\sigma_2^2}\left(e^{\rho\sigma_1\sigma_2}-1\right)}{\sqrt{e^{2\mu_1+\sigma_1^2}\left(e^{\sigma_1^2}-1\right)}\sqrt{e^{2\mu_2+\sigma_2^2}\left(e^{\sigma_2^2}-1\right)}} \\ &= \frac{e^{\rho\sigma_1\sigma_2}-1}{\sqrt{e^{\sigma_1^2}-1}\sqrt{e^{\sigma_2^2}-1}}\end{aligned}$$

(d) $\rho\langle Y_1,Y_2\rangle$ is an increasing function with respect to ρ. We deduce that:

$$\rho\langle Y_1,Y_2\rangle = 1 \Longleftrightarrow \rho = 1 \text{ and } \sigma_1 = \sigma_2$$

The lower bound of $\rho\langle Y_1,Y_2\rangle$ is reached if ρ is equal to -1. In this case, we have:

$$\rho\langle Y_1,Y_2\rangle = \frac{e^{-\sigma_1\sigma_2}-1}{\sqrt{e^{\sigma_1^2}-1}\sqrt{e^{\sigma_2^2}-1}} > -1$$

It follows that $\rho\langle Y_1,Y_2\rangle \neq -1$.

(e) It is evident that:

$$\rho\langle S_1(t),S_2(t)\rangle = \frac{e^{\rho\sigma_1\sigma_2 t}-1}{\sqrt{e^{\sigma_1^2 t}-1}\sqrt{e^{\sigma_2^2 t}-1}}$$

In the case $\sigma_1=\sigma_2$ and $\rho=1$, we have $\rho\langle S_1(t),S_2(t)\rangle = 1$. Otherwise, we obtain:

$$\lim_{t\to\infty}\rho\langle S_1(t),S_2(t)\rangle = 0$$

(f) In the case of lognormal random variables, the linear correlation does not necessarily range between -1 and $+1$.

11.4.7 The bivariate Pareto copula

1. We have:

$$\begin{aligned}\mathbf{F}_1(x_1) &= \Pr\{X_1 \leq x_1\} \\ &= \Pr\{X_1 \leq x_1, X_2 \leq \infty\} \\ &= \mathbf{F}(x_1,\infty)\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbf{F}_1(x_1) &= 1-\left(\frac{\theta_1+x_1}{\theta_1}\right)^{-\alpha}-\left(\frac{\theta_2+\infty}{\theta_2}\right)^{-\alpha}+ \\ &\quad \left(\frac{\theta_1+x_1}{\theta_1}+\frac{\theta_2+\infty}{\theta_2}-1\right)^{-\alpha} \\ &= 1-\left(\frac{\theta_1+x_1}{\theta_1}\right)^{-\alpha}\end{aligned}$$

We conclude that $\mathbf{F}_1$ (and $\mathbf{F}_2$) is a Pareto distribution.

2. We have:

$$\mathbf{C}(u_1,u_2) = \mathbf{F}\left(\mathbf{F}_1^{-1}(u_1),\mathbf{F}_2^{-1}(u_2)\right)$$

It follows that:

$$\begin{aligned} & 1-\left(\frac{\theta_1+x_1}{\theta_1}\right)^{-\alpha}=u_1 \\ \Leftrightarrow \quad & \left(\frac{\theta_1+x_1}{\theta_1}\right)^{-\alpha}=1-u_1 \\ \Leftrightarrow \quad & \frac{\theta_1+x_1}{\theta_1}=\left(1-u_1\right)^{-1/\alpha} \end{aligned}$$

We deduce that:

$$\begin{aligned} \mathbf{C}\left(u_1,u_2\right) &= 1-\left(1-u_1\right)-\left(1-u_2\right)+ \\ & \quad \left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha} \\ &= u_1+u_2-1+\left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha} \end{aligned}$$

3. We have:

$$\begin{aligned} \frac{\partial\,\mathbf{C}\left(u_1,u_2\right)}{\partial\,u_1} &= 1-\alpha\left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha-1}\times \\ & \quad \left(-\frac{1}{\alpha}\right)\left(1-u_1\right)^{-1/\alpha-1}\times(-1) \\ &= 1-\left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha-1}\times \\ & \quad \left(1-u_1\right)^{-1/\alpha-1} \end{aligned}$$

We deduce that the probability density function of the copula is[1]:

$$\begin{aligned} c\left(u_1,u_2\right) &= \frac{\partial^2\,\mathbf{C}\left(u_1,u_2\right)}{\partial\,u_1\,\partial\,u_2} \\ &= -\left(-\alpha-1\right)\left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha-2}\times \\ & \quad \left(-\frac{1}{\alpha}\right)\left(1-u_2\right)^{-1/\alpha-1}\times(-1)\times\left(1-u_1\right)^{-1/\alpha-1} \\ &= \left(\frac{\alpha+1}{\alpha}\right)\left(\left(1-u_1\right)^{-1/\alpha}+\left(1-u_2\right)^{-1/\alpha}-1\right)^{-\alpha-2}\times \\ & \quad \left(1-u_1-u_2+u_1u_2\right)^{-1/\alpha-1} \end{aligned}$$

In Figure 11.2, we have reported the density of the Pareto copula when α is equal to 1 and 10.

[1] Another expression of $c\left(u_1,u_2\right)$ is:

$$\begin{aligned} c\left(u_1,u_2\right) &= \left(\frac{\alpha+1}{\alpha}\right)\left(\left(1-u_1\right)\left(1-u_2\right)\right)^{1/\alpha}\times \\ & \quad \left(\left(1-u_1\right)^{1/\alpha}+\left(1-u_2\right)^{1/\alpha}-\left(1-u_1\right)^{1/\alpha}\left(1-u_2\right)^{1/\alpha}\right)^{-\alpha-2} \end{aligned}$$

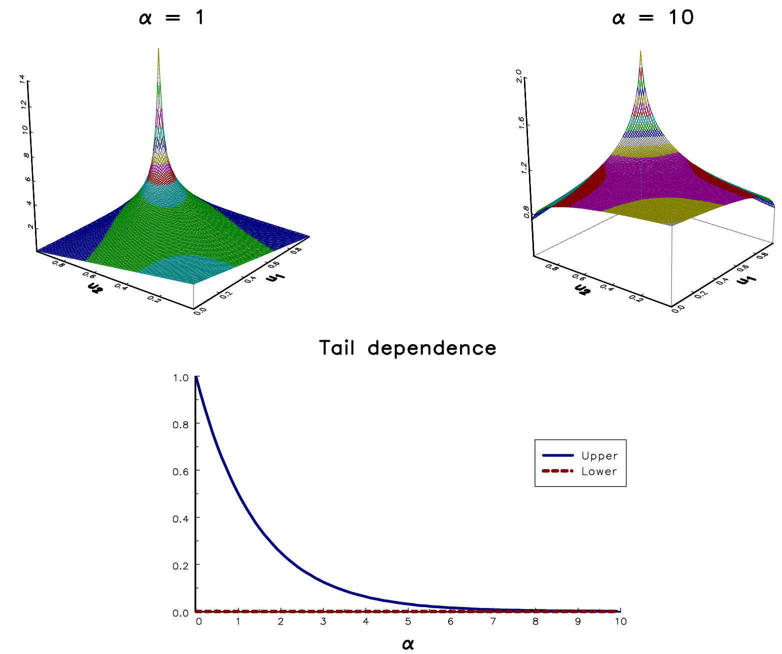

FIGURE 11.2: The Pareto copula

4. We have:

$$\begin{aligned}
\lambda^- &= \lim_{u\to 0^+} \frac{\mathbf{C}(u,u)}{u} \\
&= 2 \lim_{u\to 0^+} \frac{\partial\, \mathbf{C}(u,u)}{\partial\, u_1} \\
&= 2 \lim_{u\to 0^+} 1 - \left((1-u)^{-1/\alpha} + (1-u)^{-1/\alpha} - 1 \right)^{-\alpha-1} (1-u)^{-1/\alpha-1} \\
&= 2 \lim_{u\to 0^+} (1-1) \\
&= 0
\end{aligned}$$

and:

$$\begin{aligned}
\lambda^+ &= \lim_{u\to 1^-} \frac{1-2u+\mathbf{C}(u,u)}{1-u} \\
&= \lim_{u\to 1^-} \frac{\left((1-u)^{-1/\alpha} + (1-u)^{-1/\alpha} - 1 \right)^{-\alpha}}{1-u} \\
&= \lim_{u\to 1^-} \left(1+1-(1-u)^{1/\alpha} \right)^{-\alpha} \\
&= 2^{-\alpha}
\end{aligned}$$

The tail dependence coefficients λ^- and λ^+ are given with respect to the parameter α in Figure 11.2. We deduce that the bivariate Pareto copula function has no lower tail dependence ($\lambda^- = 0$), but an upper tail dependence ($\lambda^+ = 2^{-\alpha}$).

5. The bivariate Pareto copula family cannot reach $\mathbf{C}^-$ because λ^- is never equal to 1.

We notice that:

$$\lim_{\alpha \to \infty} \lambda^+ = 0$$

and

$$\lim_{\alpha \to 0} \lambda^+ = 1$$

This implies that the bivariate Pareto copula may reach $\mathbf{C}^\perp$ and $\mathbf{C}^+$ for these two limit cases: $\alpha \to \infty$ and $\alpha \to 0$. In fact, $\alpha \to 0$ does not correspond to the copula $\mathbf{C}^+$ because λ^- is always equal to 0.

6. (a) We note $U_1 = \mathbf{F}_1(X_1)$ and $U_2 = \mathbf{F}_2(X_2)$. X_1 and X_2 are comonotonic if and only if:

$$U_2 = U_1$$

We deduce that:

$$\begin{aligned} & 1 - \left(\frac{\theta_2 + X_2}{\theta_2}\right)^{-\alpha_2} = 1 - \left(\frac{\theta_1 + X_1}{\theta_1}\right)^{-\alpha_1} \\ \Leftrightarrow \quad & \left(\frac{\theta_2 + X_2}{\theta_2}\right)^{-\alpha_2} = \left(\frac{\theta_1 + X_1}{\theta_1}\right)^{-\alpha_1} \\ \Leftrightarrow \quad & X_2 = \theta_2 \left(\left(\frac{\theta_1 + X_1}{\theta_1}\right)^{\alpha_1/\alpha_2} - 1\right) \end{aligned}$$

We know that $\rho\langle X_1, X_2\rangle = 1$ if and only if there is an increasing linear relationship between X_1 and X_2. This implies that:

$$\frac{\alpha_1}{\alpha_2} = 1$$

(b) X_1 and X_2 are countermonotonic if and only if:

$$U_2 = 1 - U_1$$

We deduce that:

$$\begin{aligned} & \left(\frac{\theta_2 + X_2}{\theta_2}\right)^{-\alpha_2} = 1 - \left(\frac{\theta_1 + X_1}{\theta_1}\right)^{-\alpha_1} \\ \Leftrightarrow \quad & \left(\frac{\theta_2 + X_2}{\theta_2}\right)^{-\alpha_2} = 1 - \left(\frac{\theta_1 + X_1}{\theta_1}\right)^{-\alpha_1} \\ \Leftrightarrow \quad & X_2 = \theta_2 \left(\left(1 - \left(\frac{\theta_1 + X_1}{\theta_1}\right)^{-\alpha_1}\right)^{1/\alpha_2} - 1\right) \end{aligned}$$

It is not possible to obtain a decreasing linear function between X_1 and X_2. This implies that $\rho\langle X_1, X_2\rangle > -1$.

(c) We have:

$$\begin{aligned} \mathbf{F}'(x_1, x_2) &= \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \\ &= 1 - \left(\frac{\theta_1 + x_1}{\theta_1}\right)^{-\alpha_1} - \left(\frac{\theta_2 + x_2}{\theta_2}\right)^{-\alpha_2} + \\ & \quad \left(\left(\frac{\theta_1 + x_1}{\theta_1}\right)^{\alpha_1/\alpha} + \left(\frac{\theta_2 + x_2}{\theta_2}\right)^{\alpha_2/\alpha} - 1\right)^{-\alpha} \end{aligned}$$

The traditional bivariate Pareto distribution $\mathbf{F}(x_1, x_2)$ is a special case of $\mathbf{F}'(x_1, x_2)$ when:

$$\alpha_1 = \alpha_2 = \alpha$$

Using $\mathbf{F}'$ instead of $\mathbf{F}$, we can control the tail dependence, but also the univariate tail index of the two margins.

Chapter 12

Extreme Value Theory

12.4.1 Uniform order statistics

1. Since we have $f(x) = 1$ and $\mathbf{F}(x) = x$, we deduce that:

$$\begin{aligned} f_{i:n}(x) &= \frac{n!}{(i-1)!\,(n-i)!} \cdot x^{i-1} \cdot (1-x)^{n-1} \cdot 1 \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)} x^{i-1} (1-x)^{n-i} \end{aligned}$$

This is the probability density function of the Beta distribution $\mathcal{B}(\alpha, \beta)$ where $\alpha = i$ and $\beta = n - i + 1$.

2. We have:

$$\begin{aligned} \mathbb{E}[X_{i:n}] &= \mathbb{E}[\mathcal{B}(i, n-i+1)] \\ &= \frac{\alpha}{\alpha+\beta} \\ &= \frac{i}{n+1} \end{aligned}$$

3. We have:

$$\begin{aligned} \operatorname{var}(X_{i:n}) &= \operatorname{var}(\mathcal{B}(i, n-i+1)) \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{i(n-i+1)}{(n+1)^2(n+2)} \end{aligned}$$

4. We have:

Sample	$X_{i:8}$							
	1	2	3	4	5	6	7	8
1	0.04	0.14	0.24	0.34	0.45	0.55	0.72	0.94
2	0.12	0.25	0.31	0.32	0.57	0.64	0.69	0.97
3	0.11	0.17	0.17	0.26	0.50	0.50	0.69	0.85
4	0.00	0.03	0.15	0.53	0.58	0.77	0.98	0.98
5	0.15	0.25	0.46	0.62	0.65	0.74	0.85	0.89
6	0.05	0.07	0.15	0.25	0.65	0.74	0.86	0.93
7	0.12	0.16	0.33	0.34	0.55	0.61	0.63	0.95
8	0.01	0.11	0.14	0.47	0.57	0.82	0.87	0.96
9	0.27	0.55	0.57	0.68	0.73	0.78	0.83	0.85
10	0.28	0.40	0.68	0.89	0.91	0.94	0.99	0.99

The empirical and theoretical mean and standard deviation of $X_{i:8}$ are reported in Table 12.1.

TABLE 12.1: Empirical and theoretical mean and standard deviation of $X_{i:8}$

i	$\bar{X}_{i:8}$	$\mathbb{E}[X_{i:n}]$	$\hat{\sigma}(X_{i:n})$	$\sigma(X_{i:n})$
1	0.1150	0.1111	0.0981	0.0994
2	0.2130	0.2222	0.1584	0.1315
3	0.3200	0.3333	0.1918	0.1491
4	0.4700	0.4444	0.2096	0.1571
5	0.6160	0.5556	0.1302	0.1571
6	0.7090	0.6667	0.1333	0.1491
7	0.8110	0.7778	0.1241	0.1315
8	0.9310	0.8889	0.0511	0.0994

5. We reiterate that $X_{i:n} \sim \mathcal{B}(i, n-i+1)$. We deduce that the median statistic follows a symmetric Beta distribution:

$$X_{k+1:n} \sim \mathcal{B}(k+1, k+1)$$

Moreover, we have:

$$X_{i:n} \sim \mathcal{B}(i, 2k-i)$$

It follows that the density function of $X_{i:n}$ is right asymmetric if $i \leq k$, symmetric about .5 if $i = k+1$ and left asymmetric otherwise.

6. We consider the change of variable: $U = \mathbf{F}(X)$. It follows that U follows a uniform distribution. Using the previous results, we can deduce that the density function of $U_{i:n}$ is right asymmetric if $i \leq k$, symmetric about .5 if $i = k+1$ and left asymmetric otherwise. Because $\mathbf{F}(x)$ is a symmetric function about $x^{\star} = \mathbf{F}^{-1}(0.5)$, we conclude that the density function of $X_{i:n}$ is right asymmetric if $i \leq k$, symmetric about $x^{\star}$ if $i = k+1$ and left asymmetric otherwise.

12.4.2 Order statistics and return period

1. We have:

$$\begin{aligned}
\mathbf{F}_{n:n}(x) &= \Pr\{\max(X_1, \ldots, X_n) \leq x\} \\
&= \Pr\{X_1 \leq x, \ldots, X_n \leq x\} \\
&= \prod_{i=1}^{n} \Pr\{X_i \leq x\} \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)^n
\end{aligned}$$

2. The density function of $X_{n:n}$ is equal to:

$$\begin{aligned}
f_{n:n}(x) &= \partial_x \mathbf{F}_{n:n}(x) \\
&= \frac{n}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{x-\mu}{\sigma}\right)^{n-1}
\end{aligned}$$

We deduce that the log-likelihood function of a sample $(x_1, \ldots, x_m)$ of the order statistic $X_{n:n}$ is equal to:

$$\begin{aligned}\boldsymbol{\ell}_{n:n} &= m\ln n - \frac{m}{2}\ln(2\pi) - \frac{m}{2}\ln\sigma^2 - \sum_{i=1}^{m}\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \\ &\quad (n-1)\ln\Phi\left(\frac{x_i-\mu}{\sigma}\right)\end{aligned}$$

For each time period n, we calculate $\boldsymbol{\ell}_{n:n}$ and find the estimates $\hat{\mu}_{n:n}$ and $\hat{\sigma}_{n:n}$. Then we test the joint hypothesis:

$$\mathcal{H}_0 = \left\{\begin{array}{l}\hat{\mu}_{1:1} = \hat{\mu}_{2:2} = \hat{\mu}_{3:3} = \ldots = \mu \\ \hat{\sigma}_{1:1} = \hat{\sigma}_{2:2} = \hat{\sigma}_{3:3} = \ldots = \sigma\end{array}\right.$$

3. The return period is the average period between two consecutive events. It is equal to:

$$\mathcal{T} = \frac{n}{p}$$

where p is the occurrence probability of the event and n is the unit period measured in days. We have:

$$\mathcal{T}\left(\mathbf{F}_{n:n}^{-1}(\alpha)\right) = \frac{1}{1-\alpha}\times n$$

We deduce that the return periods are respectively equal to 100, 100, 500 and 2 200 days.

4. We would like to find the value α that satisfies the following equation:

$$\mathcal{T}\left(\mathbf{F}_{20:20}^{-1}(\alpha)\right) = \mathcal{T}\left(\mathbf{F}^{-1}(99.9\%)\right)$$

We have:

$$\frac{1}{1-\alpha}\times 20 = \frac{1}{1-0.999}\times 1$$

We deduce that:

$$\alpha = 1 - 20\times 0.001 = 98\%$$

12.4.3 Extreme order statistics of exponential random variables

1. Using the Bayes formula, we have:

$$\begin{aligned}\Pr\{\boldsymbol{\tau} > t \mid \boldsymbol{\tau} > s\} &= \frac{\Pr\{\tau > t \bigcap \tau > s\}}{\Pr\{\tau > s\}} \\ &= \frac{\Pr\{\tau > t\}}{\Pr\{\tau > s\}} \\ &= \frac{\mathbf{S}(t)}{\mathbf{S}(s)} \\ &= \frac{e^{-\lambda t}}{e^{-\lambda s}} \\ &= e^{-\lambda(t-s)} \\ &= \Pr\{\tau > t-s\}\end{aligned}$$

This implies that the survival function does not depend on the initial time. This Markov property is especially useful in credit models, because the default time of the counterparty does not on the past history, for instance the age of the company.

2. We have:

$$\begin{aligned}
\Pr\left\{\min\left(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right)\geq t\right\} &= \Pr\left\{\boldsymbol{\tau}_1\geq t\bigcap\ldots\bigcap\boldsymbol{\tau}_n\geq t\right\} \\
&= \prod_{i=1}^{n}\Pr\left\{\boldsymbol{\tau}_i\geq t\right\} \\
&= \exp\left(-\sum_{i=1}^{n}\lambda_i t\right)
\end{aligned}$$

and:

$$\begin{aligned}
\Pr\left\{\max\left(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right)\leq t\right\} &= \Pr\left\{\boldsymbol{\tau}_1\leq t\bigcap\ldots\bigcap\boldsymbol{\tau}_n\leq t\right\} \\
&= \prod_{i=1}^{n}\Pr\left\{\boldsymbol{\tau}_i\leq t\right\} \\
&= \prod_{i=1}^{n}\left(1-e^{-\lambda_i t}\right)
\end{aligned}$$

We deduce that:

$$\min\left(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right)\sim\mathcal{E}\left(\sum_{i=1}^{n}\lambda_i\right)$$

The distribution of $\max\left(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right)$ is not a known probability distribution. Let us consider the case $n=2$. We have:

$$\begin{aligned}
\Pr\left\{\min\left(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2\right)=\boldsymbol{\tau}_2\right\} &= \Pr\left\{\boldsymbol{\tau}_2\leq\boldsymbol{\tau}_1\right\} \\
&= \int_0^\infty\int_0^{t_1}\lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&= \int_0^\infty\lambda_1 e^{-\lambda_1 t_1}\left(\int_0^{t_1}\lambda_2 e^{-\lambda_2 t_2}\,\mathrm{d}t_2\right)\mathrm{d}t_1 \\
&= \int_0^\infty\lambda_1 e^{-\lambda_1 t_1}\left(1-e^{-\lambda_2 t_1}\right)\mathrm{d}t_1 \\
&= \int_0^\infty\lambda_1 e^{-\lambda_1 t_1}\,\mathrm{d}t_1-\lambda_1\int_0^\infty e^{-(\lambda_1+\lambda_2)t_1}\,\mathrm{d}t_1 \\
&= 1-\frac{\lambda_1}{\lambda_1+\lambda_2} \\
&= \frac{\lambda_2}{\lambda_1+\lambda_2}
\end{aligned}$$

We can generalize this result to the case $n>2$ and we finally obtain:

$$\Pr\left\{\min\left(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right)=\boldsymbol{\tau}_i\right\}=\frac{\lambda_i}{\sum_{j=1}^{n}\lambda_j}$$

3. When $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_i$ are comonotone, we have $\mathbf{S}_1\left(\tau_1\right)=\mathbf{S}_i\left(\tau_i\right)$. It follows that:

$$\boldsymbol{\tau}_i=\frac{\lambda_1}{\lambda_i}\boldsymbol{\tau}_1$$

We note $\lambda^+ = \max(\lambda_1, \ldots, \lambda_n)$ and $\lambda^- = \min(\lambda_1, \ldots, \lambda_n)$. We deduce that:

$$\begin{aligned}
\Pr\{\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) \geq t\} &= \Pr\left\{\min\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}\right)\lambda_1 \boldsymbol{\tau}_1 \geq t\right\} \\
&= \Pr\left\{\frac{\lambda_1}{\lambda^+}\boldsymbol{\tau}_1 \geq t\right\} \\
&= \Pr\left\{\boldsymbol{\tau}_1 \geq \frac{\lambda^+}{\lambda_1} t\right\} \\
&= \exp\left(-\lambda_1 \frac{\lambda^+}{\lambda_1} t\right) \\
&= \exp\left(-\lambda^+ t\right)
\end{aligned}$$

and:

$$\begin{aligned}
\Pr\{\max(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) \leq t\} &= \Pr\left\{\frac{\lambda_1}{\lambda^-}\boldsymbol{\tau}_1 \leq t\right\} \\
&= 1 - \exp\left(-\lambda_1 \frac{\lambda^-}{\lambda_1} t\right) \\
&= 1 - \exp\left(-\lambda^- t\right)
\end{aligned}$$

We finally obtain:

$$\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) \sim \mathcal{E}\left(\lambda^+\right)$$

and:

$$\max(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) \sim \mathcal{E}\left(\lambda^-\right)$$

12.4.4 Extreme value theory in the bivariate case

1. An extreme value copula $\mathbf{C}$ satisfies the following relationship:

$$\mathbf{C}\left(u_1^t, u_2^t\right) = \mathbf{C}^t(u_1, u_2)$$

for all $t > 0$.

2. The product copula $\mathbf{C}^\perp$ is an EV copula because we have:

$$\begin{aligned}
\mathbf{C}^\perp\left(u_1^t, u_2^t\right) &= u_1^t u_2^t \\
&= (u_1 u_2)^t \\
&= \left[\mathbf{C}^\perp(u_1, u_2)\right]^t
\end{aligned}$$

For the copula $\mathbf{C}^+$, we obtain:

$$\begin{aligned}
\mathbf{C}^+\left(u_1^t, u_2^t\right) &= \min\left(u_1^t, u_2^t\right) \\
&= \begin{cases} u_1^t & \text{if } u_1 \leq u_2 \\ u_2^t & \text{otherwise} \end{cases} \\
&= (\min(u_1, u_2))^t \\
&= \left[\mathbf{C}^+(u_1, u_2)\right]^t
\end{aligned}$$

However, the EV property does not hold for the Fréchet lower bound copula $\mathbf{C}^-$:

$$\mathbf{C}^-\left(u_1^t, u_2^t\right) = \max\left(u_1^t + u_2^t - 1, 0\right) \neq \max(u_1 + u_2 - 1, 0)^t$$

Indeed, we have $\mathbf{C}^{-}(0.5, 0.8) = \max(0.5 + 0.8 - 1, 0) = 0.3$ and:

$$\begin{aligned} \mathbf{C}^{-}\left(0.5^2, 0.8^2\right) &= \max(0.25 + 0.64 - 1, 0) \\ &= 0 \\ &\neq 0.3^2 \end{aligned}$$

3. We have:

$$\begin{aligned} \mathbf{C}\left(u_1^t, u_2^t\right) &= \exp\left(-\left[\left(-\ln u_1^t\right)^\theta + \left(-\ln u_2^t\right)^\theta\right]^{1/\theta}\right) \\ &= \exp\left(-\left[\left(-t\ln u_1\right)^\theta + \left(-t\ln u_2\right)^\theta\right]^{1/\theta}\right) \\ &= \exp\left(-t\left[\left(-\ln u_1\right)^\theta + \left(-\ln u_2\right)^\theta\right]^{1/\theta}\right) \\ &= \left(e^{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}}\right)^t \\ &= \mathbf{C}^t\left(u_1, u_2\right) \end{aligned}$$

4. The upper tail dependence λ is defined as follows:

$$\lambda = \lim_{u \to 1^+} \frac{1 - 2u + \mathbf{C}(u_1, u_2)}{1 - u}$$

It measures the probability to have an extreme in one direction knowing that we have already an extreme in the other direction. If λ is equal to 0, extremes are independent and the EV copula is the product copula $\mathbf{C}^{\perp}$. If λ is equal to 1, extremes are comonotonic and the EV copula is the Fréchet upper bound copula $\mathbf{C}^{+}$. Moreover, the upper tail dependence of the copula between the random variables is equal to the upper tail dependence of the copula between the extremes.

5. Using L'Hospital's rule, we have:

$$\begin{aligned} \lambda &= \lim_{u \to 1^+} \frac{1 - 2u + e^{-\left[(-\ln u)^\theta + (-\ln u)^\theta\right]^{1/\theta}}}{1 - u} \\ &= \lim_{u \to 1^+} \frac{1 - 2u + e^{-\left[2(-\ln u)^\theta\right]^{1/\theta}}}{1 - u} \\ &= \lim_{u \to 1^+} \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u} \\ &= \lim_{u \to 1^+} \frac{0 - 2 + 2^{1/\theta} u^{2^{1/\theta} - 1}}{-1} \\ &= \lim_{u \to 1^+} 2 - 2^{1/\theta} u^{2^{1/\theta} - 1} \\ &= 2 - 2^{1/\theta} \end{aligned}$$

If θ is equal to 1, we obtain $\lambda = 0$. It comes that the EV copula is the product copula. Extremes are then not correlated. This result is not surprising because the Gumbel-Houggard copula is equal to the product copula when $\theta = 1$:

$$e^{-\left[(-\ln u_1)^1 + (-\ln u_2)^1\right]^1} = u_1 u_2 = \mathbf{C}^{\perp}(u_1, u_2)$$

6. (a) We have:

$$\begin{aligned}\mathbf{C}\left(u_1^t,u_2^t\right) &= u_1^{t(1-\theta_1)}u_2^{t(1-\theta_2)}\min\left(u_1^{t\theta_1},u_2^{t\theta_2}\right)\\ &= \left(u_1^{1-\theta_1}\right)^t\left(u_2^{1-\theta_2}\right)^t\left(\min\left(u_1^{\theta_1},u_2^{\theta_2}\right)\right)^t\\ &= \left(u_1^{1-\theta_1}u_2^{1-\theta_2}\min\left(u_1^{\theta_1},u_2^{\theta_2}\right)\right)^t\\ &= \mathbf{C}^t\left(u_1,u_2\right)\end{aligned}$$

(b) If $\theta_1 > \theta_2$, we obtain:

$$\begin{aligned}\lambda &= \lim_{u\to 1^+}\frac{1-2u+u^{1-\theta_1}u^{1-\theta_2}\min\left(u^{\theta_1},u^{\theta_2}\right)}{1-u}\\ &= \lim_{u\to 1^+}\frac{1-2u+u^{1-\theta_1}u^{1-\theta_2}u^{\theta_1}}{1-u}\\ &= \lim_{u\to 1^+}\frac{1-2u+u^{2-\theta_2}}{1-u}\\ &= \lim_{u\to 1^+}\frac{0-2+(2-\theta_2)\,u^{1-\theta_2}}{-1}\\ &= \lim_{u\to 1^+}2-2u^{1-\theta_2}+\theta_2u^{1-\theta_2}\\ &= \theta_2\end{aligned}$$

If $\theta_2 > \theta_1$, we have $\lambda = \theta_1$. We deduce that the upper tail dependence of the Marshall-Olkin copula is $\min(\theta_1,\theta_2)$.

(c) If $\theta_1 = 0$ or $\theta_2 = 0$, we obtain $\lambda = 0$. It comes that the copula of the extremes is the product copula. Extremes are then not correlated.

(d) Two extremes are perfectly correlated when we have $\theta_1 = \theta_2 = 1$. In this case, we obtain:

$$\mathbf{C}\left(u_1,u_2\right) = \min\left(u_1,u_2\right) = \mathbf{C}^+\left(u_1,u_2\right)$$

12.4.5 Max-domain of attraction in the bivariate case

1. Let (X_1, X_2) be a bivariate random variable whose probability distribution is:

$$\mathbf{F}\left(x_1,x_2\right) = \mathbf{C}_{\langle X_1,X_2\rangle}\left(\mathbf{F}_1\left(x_1\right),\mathbf{F}_2\left(x_2\right)\right)$$

We know that the corresponding EV probability distribution is:

$$\mathbf{G}\left(x_1,x_2\right) = \mathbf{C}^\star_{\langle X_1,X_2\rangle}\left(\mathbf{G}_1\left(x_1\right),\mathbf{G}_2\left(x_2\right)\right)$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ are the two univariate EV probability distributions and $\mathbf{C}^\star_{\langle X_1,X_2\rangle}$ is the EV copula associated to $\mathbf{C}_{\langle X_1,X_2\rangle}$.

(a) We deduce that:

$$\begin{aligned}\mathbf{G}\left(x_1,x_2\right) &= \mathbf{C}^\perp\left(\mathbf{G}_1\left(x_1\right),\mathbf{G}_2\left(x_2\right)\right)\\ &= \mathbf{\Lambda}\left(x_1\right)\mathbf{\Psi}_1\left(x_2-1\right)\\ &= \exp\left(-e^{-x_1}+x_2-1\right)\end{aligned}$$

(b) We have:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= \mathbf{\Lambda}(x_1)\,\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right) \\ &= \exp\left(-e^{-x_1}-\left(1+\frac{x_2}{\alpha}\right)^{-\alpha}\right)\end{aligned}$$

(c) We have:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= \mathbf{\Psi}_1(x_1-1)\,\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right) \\ &= \exp\left(x_1-1-\left(1+\frac{x_2}{\alpha}\right)^{-\alpha}\right)\end{aligned}$$

2. We know that the upper tail dependence is equal to zero for the Normal copula when $\rho < 1$. We deduce that the EV copula is the product copula. We then obtain the same results as previously.

3. When the parameter ρ is equal to 1, the Normal copula is the Fréchet upper bound copula $\mathbf{C}^+$, which is an EV copula. We deduce the following results:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= \min\left(\mathbf{\Lambda}(x_1),\mathbf{\Psi}_1(x_2-1)\right) \\ &= \min\left(\exp\left(-e^{-x_1}\right),\exp(x_2-1)\right)\end{aligned} \tag{a}$$

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= \min\left(\mathbf{\Lambda}(x_1),\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right)\right) \\ &= \min\left(\exp\left(-e^{-x_1}\right),\exp\left(-\left(1+\frac{x_2}{\alpha}\right)^{-\alpha}\right)\right)\end{aligned} \tag{b}$$

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= \min\left(\mathbf{\Psi}_1(x_1-1),\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right)\right) \\ &= \min\left(\exp(x_2-1),\exp\left(-\left(1+\frac{x_2}{\alpha}\right)^{-\alpha}\right)\right)\end{aligned} \tag{c}$$

4. In the previous exercise, we have shown that the Gumbel-Houggard copula is an EV copula.

(a) We deduce that:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= e^{-\left[(-\ln\mathbf{\Lambda}(x_1))^\theta+(-\ln\mathbf{\Psi}_1(x_2-1))^\theta\right]^{1/\theta}} \\ &= \exp\left(-\left[e^{-\theta x_1}+(1-x_2)^\theta\right]^{1/\theta}\right)\end{aligned}$$

(b) We obtain:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= e^{-\left[(-\ln\mathbf{\Lambda}(x_1))^\theta+\left(-\ln\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right)\right)^\theta\right]^{1/\theta}} \\ &= \exp\left(-\left[e^{-\theta x_1}+\left(1+\frac{x_2}{\alpha}\right)^{-\alpha\theta}\right]^{1/\theta}\right)\end{aligned}$$

(c) We have:

$$\begin{aligned}\mathbf{G}(x_1,x_2) &= e^{-\left[(-\ln\mathbf{\Psi}_1(x_1-1))^\theta+\left(-\ln\mathbf{\Phi}_\alpha\left(1+\frac{x_2}{\alpha}\right)\right)^\theta\right]^{1/\theta}} \\ &= \exp\left(-\left[(1-x_1)^\theta+\left(1+\frac{x_2}{\alpha}\right)^{-\alpha\theta}\right]^{1/\theta}\right)\end{aligned}$$

Chapter 13

Monte Carlo Simulation Methods

13.4.1 Simulating random numbers using the inversion method

1. Let u_i be a uniform random variate.

   (a) We have seen that the quantile function of the distribution function $\mathcal{GEV}(\mu, \sigma, \xi)$ has the following expression:

   $$\mathbf{G}^{-1}(\alpha) = \mu - \frac{\sigma}{\xi}\left(1 - (-\ln \alpha)^{-\xi}\right)$$

   It follows that:

   $$x_i \leftarrow \mu - \frac{\sigma}{\xi}\left(1 - (-\ln u_i)^{-\xi}\right)$$

   (b) The cumulative density function of the log-normal distribution $\mathcal{LN}\left(\mu, \sigma^2\right)$ is equal to:

   $$\mathbf{F}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

   We deduce that:

   $$\mathbf{F}^{-1}(u) = \exp\left(\mu + \sigma \Phi^{-1}(u)\right)$$

   To simulate a log-normal random variate, we then use the following algorithm:

   $$x_i \leftarrow \exp\left(\mu + \sigma \Phi^{-1}(u_i)\right)$$

   (c) We have:

   $$\mathbf{F}(x) = \frac{1}{1 + (x/\alpha)^{-\beta}}$$

   and:

   $$\mathbf{F}^{-1}(u) = \alpha\left(\frac{u}{1-u}\right)^{1/\beta}$$

   To simulate a log-logistic random variate $\mathcal{LL}(\alpha, \beta)$, we use the following transformation:

   $$x_i \leftarrow \alpha\left(\frac{u_i}{1-u_i}\right)^{1/\beta}$$

2. (a) Let x_i be a random variate simulated from the probability distribution of X. A straightforward algorithm is to keep all the random variates x_i's that are higher than the threshold H:

   $$l_i \leftarrow \begin{cases} x_i \text{ if } x_i \geq H \\ \text{a missing value otherwise} \end{cases}$$

(b) We have:

$$\begin{aligned}
\mathbf{F}_L(x) &= \Pr\{X \le x \mid X \ge H\} \\
&= \frac{\Pr\{X \le x, X \ge H\}}{\Pr\{X \ge H\}} \\
&= \frac{\mathbf{F}_X(x) - \mathbf{F}_X(H)}{1 - \mathbf{F}_X(H)}
\end{aligned}$$

(c) We have:

$$\begin{aligned}
\mathbf{F}_L(x) = u &\Leftrightarrow \frac{\mathbf{F}_X(x) - \mathbf{F}_X(H)}{1 - \mathbf{F}_X(H)} = u \\
&\Leftrightarrow \mathbf{F}_X(x) = u + \mathbf{F}_X(H)(1-u) \\
&\Leftrightarrow x = \mathbf{F}_X^{-1}(u + \mathbf{F}_X(H)(1-u))
\end{aligned}$$

It follows that:

$$\mathbf{F}_L^{-1}(u) = \mathbf{F}_X^{-1}(u + \mathbf{F}_X(H)(1-u))$$

We deduce that the algorithm to simulate the random variate l_i is:

$$l_i \leftarrow \mathbf{F}_X^{-1}(u_i + \mathbf{F}_X(H)(1-u_i))$$

(d) Concerning the first algorithm, we simulate n_X values of X, but we only kept on average $n_L = n_X(1 - \mathbf{F}_X(H))$ values of L, meaning that the acceptance ratio is equal to $1 - \mathbf{F}_X(H)$. For the second algorithm, all the simulated values of u_i are kept. For instance, if $\mathbf{F}_X(H)$ is equal to 90% and we would like to simulate one million of random numbers for L, we have to simulate approximatively 10 millions of random numbers in the first algorithm, that is 10 more times than for the second algorithm. In this case, the acceptance ratio is only equal to 10%.

(e) When X follows a log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$, Algorithm (a) becomes:

$$l_i \leftarrow \exp\left(\mu + \sigma\Phi^{-1}(u_i)\right)$$

with the condition $u_i \ge \mathbf{F}_X(H) = 95.16\%$. For Algorithm (b), we have:

$$l_i \leftarrow \exp\left(\mu + \sigma\Phi^{-1}\left(u_i + (1-u_i)\,\Phi\left(\frac{\ln H - \mu}{\sigma}\right)\right)\right)$$

In Figure 13.1, we have represented the random numbers l_i generated with the two algorithms. We observe that only 4 simulated values are higher than H in the case of Algorithm (a). With Algorithm (c), all the simulated values are higher than H and it is easier to simulate a random loss located in the distribution tail.

3. (a) Let x_i be a simulated value of X_i.We have:

$$x_{1:n} = \min(x_1, \ldots, x_n)$$

and:

$$x_{n:n} = \max(x_1, \ldots, x_n)$$

(b) We have:

$$\mathbf{F}_{1:n}(x) = 1 - (1 - \mathbf{F}(x))^n$$

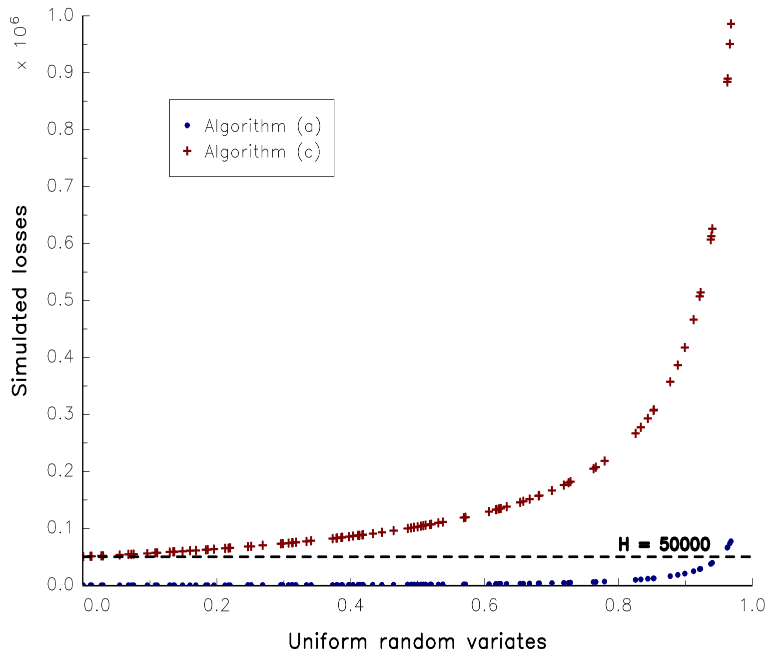

FIGURE 13.1: Simulation of conditional losses $L = X \mid X \geq H$

and:

$$\mathbf{F}_{1:n}^{-1}(u) = \mathbf{F}^{-1}\left(1 - (1-u)^{1/n}\right)$$

We deduce that a simulated value x_i^- of $X_{1:n}$ is given by:

$$x_i^- \leftarrow \mathbf{F}^{-1}\left(1 - (1-u_i)^{1/n}\right)$$

For the maximum order statistic $X_{n:n}$, we have $\mathbf{F}_{1:n}(x) = \mathbf{F}(x)^n$ and:

$$x_i^+ \leftarrow \mathbf{F}^{-1}\left(u_i^{1/n}\right)$$

(c) In Figure 13.2, we report 1 000 simulated values of $X_{1:50}$ and $X_{50:50}$ when $X_i \sim \mathcal{N}(0,1)$.

13.4.2 Simulating random numbers using the transformation method

1. The density function of $Y = h(X)$ is given by the following relationship:

$$g(y) = f(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$$

We obtain:

$$\begin{aligned} g(y) &= \frac{\beta^\alpha x^{-\alpha-1}e^{-\beta/x}}{\Gamma(\alpha)}x^2 \\ &= \frac{\beta^\alpha x^{-\alpha+1}e^{-\beta/x}}{\Gamma(\alpha)} \\ &= \frac{\beta^\alpha y^{\alpha-1}e^{-\beta y}}{\Gamma(\alpha)} \end{aligned}$$

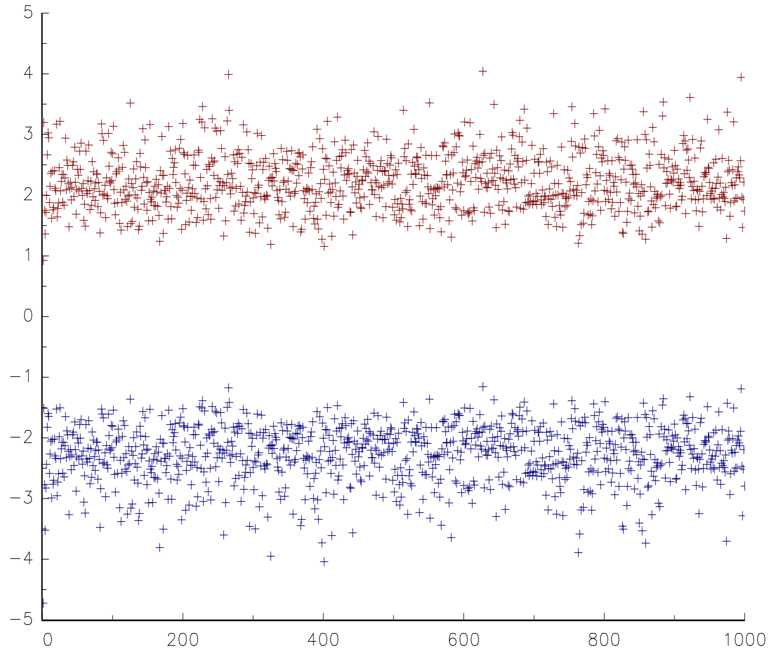

FIGURE 13.2: Simulation of $X_{1:50}$ and $X_{50:50}$ when $X_i \sim \mathcal{N}(0,1)$

It follows that $Y \sim \mathcal{G}(\alpha, \beta)$. To simulate X, we draw a gamma random variate Y and set $X = 1/Y$.

2. (a) The density function of $X \sim \mathcal{G}(\alpha, \beta)$ is equal to:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

In the case $\alpha = 1$, we obtain:

$$f(x) = \beta e^{-\beta x}$$

This is the density function of $\mathcal{E}(\beta)$. To simulate X, we apply the following transformation:

$$x \leftarrow -\frac{\ln u}{\beta}$$

where u is a uniform random number.

(b) We know that:

$$\mathcal{G}(n, \beta) = \sum_{j=1}^{n} \mathcal{G}(1, \beta)$$

We deduce that:

$$\mathcal{G}(n, \beta) = \sum_{i=1}^{n} E_i$$

where $E_i \sim \mathcal{E}(\beta)$ are *iid* exponential random variables. We deduce that the probability distribution $\mathcal{G}(n, \beta)$ can be simulated by:

$$x \leftarrow -\frac{1}{\beta} \sum_{i=1}^{n} \ln u_i$$

or:

$$x \leftarrow -\frac{1}{\beta} \ln \left(\prod_{i=1}^{n} u_i \right)$$

where $u_1, \ldots, u_n$ are *iid* uniform random variates.

3. (a) Let $Y \sim \mathcal{G}(\alpha, \delta)$ and $Z \sim \mathcal{G}(\beta, \delta)$ be two independent gamma-distributed random variables. We have:

$$f_{Y,Z}(y, z) = \frac{\delta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} z^{\beta-1} e^{-\delta(y+z)}$$

We note:

$$X = \frac{Y}{Y+Z}$$

and:

$$S = Y + Z$$

It follows that $Y = XS$ and $Z = (1 - X) S$. The Jacobian of $(y, z) = \varphi(x, s)$ is then equal to:

$$J_\varphi = \begin{pmatrix} s & x \\ -s & 1-x \end{pmatrix}$$

Since we have $\det J_\varphi = s$, we deduce that:

$$\begin{aligned} f_{X,S}(x, s) &= f_{Y,Z}(y, z) \times |s| \\ &= \frac{\delta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (xs)^{\alpha-1} ((1-x) s)^{\beta-1} e^{-\delta s} s \\ &= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \right) \times \\ &\quad \left(\frac{\delta^{\alpha+\beta}}{\Gamma(\alpha+\beta)} s^{\alpha+\beta-1} e^{-\delta s} \right) \\ &= f_X(x) f_S(s) \end{aligned}$$

It follows that the random variables X and S are independent, $X \sim \mathcal{B}(\alpha, \beta)$ and $S \sim \mathcal{G}(\alpha + \beta, \delta)$.

(b) To simulate a beta-distributed random variate, we consider the following transformation:

$$x \leftarrow \frac{y}{y+z}$$

where y and z are two independent random variates from $\mathcal{G}(\alpha, \delta)$ and $\mathcal{G}(\beta, \delta)$.

4. (a) We remind that:

$$f_{X,Y}(x, y) = f_{R,\Theta}(r, \theta) \left| \frac{1}{\det J_\varphi} \right|$$

where J_φ is the Jacobian associated to the change of variables $(x, y) = \varphi(r, \theta)$. We have:

$$J_\varphi = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and:

$$\det J_\varphi = r\cos^2\theta + r\sin^2\theta = r$$

Since R and Θ are independent, we have $f_{R,\Theta}(r,\theta) = f_R(r) f_\Theta(\theta)$. Moreover, Θ is a uniform random variable and we have:

$$f_\Theta(\theta) = \frac{1}{2\pi}$$

We deduce that:

$$f_{X,Y}(x,y) = \frac{f_R(r)}{2\pi r}$$

We also notice that:

$$\begin{aligned} X^2 + Y^2 &= R^2 \cos^2 \Theta + R^2 \sin^2 \Theta \\ &= R^2 \end{aligned}$$

Finally, we obtain the following result:

$$f_{X,Y}(x,y) = \frac{f_R\left(\sqrt{x^2+y^2}\right)}{2\pi\sqrt{x^2+y^2}}$$

Concerning the density function of X, we have:

$$f_X(x) = \int_{-\infty}^{\infty} \frac{f_R\left(\sqrt{x^2+y^2}\right)}{2\pi\sqrt{x^2+y^2}} \,\mathrm{d}y$$

(b) We assume that $R = \sqrt{2E}$ where $E \sim \mathcal{E}(1)$.

i. We have:

$$\begin{aligned} \mathbf{F}_R(r) &= \Pr\left\{\sqrt{2E} \leq r\right\} \\ &= \Pr\left\{E \leq \frac{r^2}{2}\right\} \\ &= 1 - e^{-r^2/2} \end{aligned}$$

We deduce that:

$$\begin{aligned} f_R(r) &= \partial_r \, \mathbf{F}_R(r) \\ &= re^{-r^2/2} \end{aligned}$$

ii. We have:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{f_R\left(\sqrt{x^2+y^2}\right)}{2\pi\sqrt{x^2+y^2}} \,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{e^{-\left(x^2+y^2\right)/2}}{2\pi} \,\mathrm{d}y \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \,\mathrm{d}y \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \\ &= \phi(x) \end{aligned}$$

We deduce that $X \sim \mathcal{N}(0,1)$. By symmetry, we also have $Y \sim \mathcal{N}(0,1)$. Moreover, we notice that X and Y are independent:

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{e^{-\left(x^2+y^2\right)/2}}{2\pi} \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-y^2/2}}{\sqrt{2\pi}} \\ &= f_X(x)\, f_Y(y) \end{aligned}$$

iii. We have $R = \sqrt{2E} = \sqrt{-2\ln U_1}$ and $\Theta = 2\pi U_2$ where U_1 and U_2 are two standard uniform random variables. It follows that X and Y defined by:

$$\begin{cases} X = \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ Y = \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{cases}$$

are two independent standard Gaussian random variables.

(c) We assume that:

$$\mathbf{F}_R(r) = 1 - \left(1 + \frac{r^2}{\nu}\right)^{-\nu/2}$$

i. It follows that the density function of R is equal to:

$$f_R(r) = r\left(1 + \frac{r^2}{\nu}\right)^{-\nu/2-1}$$

ii. We deduce that the joint density of (X,Y) is:

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{r}{2\pi r}\left(1 + \frac{r^2}{\nu}\right)^{-\nu/2-1} \\ &= \frac{1}{2\pi}\left(1 + \frac{x^2+y^2}{\nu}\right)^{-\nu/2-1} \end{aligned}$$

iii. We notice that $f_{X,Y}(x,y)$ is an even function of y. We deduce that:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi}\left(1 + \frac{x^2+y^2}{\nu}\right)^{-\nu/2-1} \mathrm{d}y \\ &= \int_0^{+\infty} \frac{1}{\pi}\left(\left(1+\frac{x^2}{\nu}\right)\left(1 + \frac{y^2}{\nu+x^2}\right)\right)^{-\nu/2-1} \mathrm{d}y \end{aligned}$$

We consider the following change of variable:

$$u = \left(1 + \frac{y^2}{\nu + x^2}\right)^{-1}$$

We have:

$$y = \sqrt{\left(\frac{1}{u} - 1\right)(\nu + x^2)}$$

and:

$$\begin{aligned} \mathrm{d}y &= -\frac{1}{2}\frac{(\nu + x^2)}{u^2\sqrt{(u^{-1}-1)(\nu+x^2)}}\,\mathrm{d}u \\ &= -\frac{1}{2}\frac{\sqrt{\nu+x^2}}{u^2\sqrt{u^{-1}-1}}\,\mathrm{d}u \end{aligned}$$

We obtain:

$$\begin{aligned} f_X(x) &= -\int_1^0 \frac{1}{2\pi}\left(\left(1+\frac{x^2}{\nu}\right)\frac{1}{u}\right)^{-\nu/2-1} \frac{\sqrt{\nu+x^2}}{u^2\sqrt{u^{-1}-1}}\,\mathrm{d}u \\ &= \int_0^1 \frac{1}{2\pi}\left(\left(1+\frac{x^2}{\nu}\right)\frac{1}{u}\right)^{-\nu/2-1} \frac{\sqrt{\nu+x^2}}{u^2\sqrt{u^{-1}-1}}\,\mathrm{d}u \\ &= \frac{\sqrt{\nu}}{2\pi}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}\int_0^1 u^{(\nu-1)/2}\left(1-u\right)^{-1/2}\,\mathrm{d}u \\ &= \mathcal{B}\left(\frac{\nu+1}{2},\frac{1}{2}\right)\frac{\sqrt{\nu}}{2\pi}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2} \end{aligned}$$

We have:

$$\begin{aligned} \mathcal{B}\left(\frac{\nu+1}{2},\frac{1}{2}\right)\frac{\sqrt{\nu}}{2\pi} &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\nu}{2}+1\right)}\frac{\sqrt{\nu}}{2\pi} \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi}}{\frac{\nu}{2}\Gamma\left(\frac{\nu}{2}\right)}\frac{\sqrt{\nu}}{2\pi} \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \end{aligned}$$

We finally deduce that:

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

This is the probability density function of the t_ν random variable.

iv. We have:

$$\mathbf{F}_R^{-1}(u) = \sqrt{\nu\left(\left(1-u\right)^{-2/\nu}-1\right)}$$

We deduce that the random variate r_i can be simulated using the inversion method:

$$r_i \leftarrow \sqrt{\nu\left(\left(1-u_i\right)^{-2/\nu}-1\right)}$$

where u_i is a uniform random variate.

v. It follows that:

$$\begin{cases} X = \sqrt{\nu\left(\left(1-U_1\right)^{-2/\nu}-1\right)}\cos\left(2\pi U_2\right) \\ Y = \sqrt{\nu\left(\left(1-U_1\right)^{-2/\nu}-1\right)}\sin\left(2\pi U_2\right) \end{cases}$$

where U_1 and U_2 are two independent uniform random variables.

vi. In the Box-Muller algorithm, X and Y are independent. In the Bailey algorithm, this property is not satisfied because:

$$f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$$

13.4.3 Simulating random numbers using rejection sampling

1. (a) It follows that:

$$\begin{aligned} h(x) &= \frac{f(x)}{g(x)} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathfrak{B}(\alpha,\beta)} \end{aligned}$$

We deduce that:

$$\begin{aligned} h'(x) &= \frac{(\alpha-1)x^{\alpha-2}(1-x)^{\beta-1}+(\beta-1)x^{\alpha-1}(1-x)^{\beta-2}}{\mathfrak{B}(\alpha,\beta)} \\ &= ((\alpha-1)(1-x)+(\beta-1)x)\frac{x^{\alpha-2}(1-x)^{\beta-2}}{\mathfrak{B}(\alpha,\beta)} \end{aligned}$$

and:

$$\begin{aligned} h'(x)=0 &\Leftrightarrow (\alpha-1)(1-x)+(\beta-1)x=0 \\ &\Leftrightarrow x^{\star}=\frac{\alpha-1}{\alpha+\beta-2} \end{aligned}$$

The supremum of $h(x)$ is equal to:

$$h(x^{\star})=\frac{1}{\mathfrak{B}(\alpha,\beta)}\left(\frac{\alpha-1}{\alpha+\beta-2}\right)^{\alpha-1}\left(\frac{\beta-1}{\alpha+\beta-2}\right)^{\beta-1}$$

We deduce that:

$$c=\frac{\Gamma(\alpha+\beta)(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(\alpha+\beta-2)^{\alpha+\beta-2}}$$

(b) We have reported the functions $f(x)$ and $cg(x)$ in Figure 13.3. c takes the value 1.27, 1.78, 8.00 and 2.76. The acceptance ratio is minimum in the third case when $\alpha=1$ and $\beta=8$. In fact, it corresponds to the worst situation for the acceptance-rejection algorithm. Indeed, when one parameter is equal to 1, we obtain:

$$c=\frac{\Gamma(1+\beta)(\beta-1)^{\beta-1}}{\Gamma(1)\Gamma(\beta)(\beta-1)^{\beta-1}}=\beta$$

The acceptance ratio p tends to zero when the second parameter tends to infinity:

$$\lim_{\beta\to\infty}p=\lim_{\beta\to\infty}\frac{1}{c}=0$$

(c) The acceptance-rejection algorithm becomes:

i. Generate two independent uniform random variates u_1 and u_2;

ii. Calculate v such that:

$$v=\frac{(\alpha+\beta-2)^{\alpha+\beta-2}}{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}u_1^{\alpha-1}(1-u_1)^{\beta-1}$$

iii. If $u_2\leq v$, accept u_1; otherwise, reject u_1.

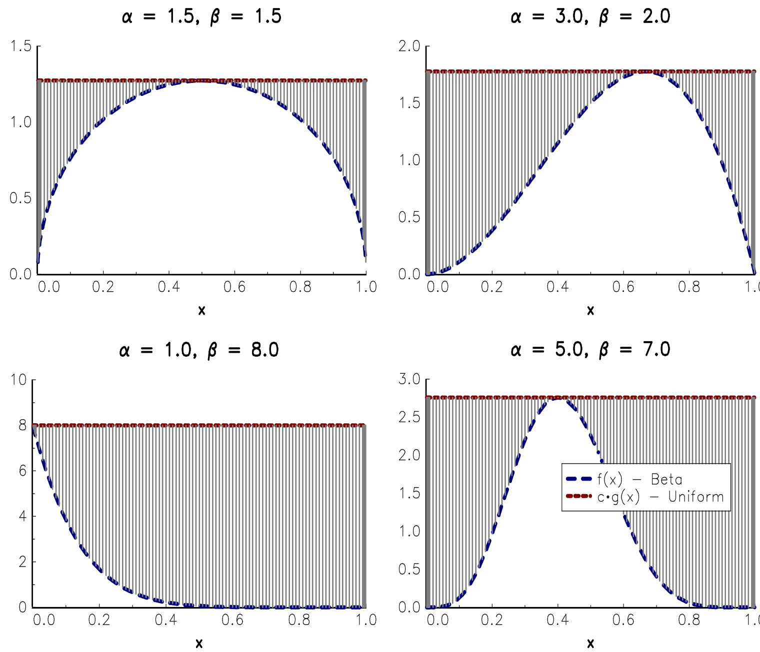

FIGURE 13.3: Rejection sampling applied to the beta distribution

2. (a) It follows that:

$$\begin{aligned} h(x) &= \frac{f(x)}{g(x)} \\ &= \frac{(1-x)^{\beta-1}}{\alpha \mathfrak{B}(\alpha,\beta)} \end{aligned}$$

Its maximum is reached at point $x^{\star} = 0$. We deduce that:

$$c = \frac{1}{\alpha \mathfrak{B}(\alpha,\beta)}$$

(b) We have $\mathbf{G}(x) = x^{\alpha}$. We use the inversion method to simulate X:

$$x \leftarrow u^{1/\alpha}$$

where u is a uniform random variate.

(c) The acceptance-rejection algorithm becomes:

i. Generate two independent uniform random variates u_1 and u_2;

ii. Calculate $x = u_1^{1/\alpha}$;

iii. Calculate v such that:

$$\begin{aligned} v &= \frac{f(x)}{cg(x)} \\ &= \left(1 - u_1^{1/\alpha}\right)^{\beta-1} \end{aligned}$$

iv. If $u_2 \leq v$, accept x; otherwise, reject x.

3. (a) We have:

$$\mathbf{G}(x) = \int_{-\infty}^{x} \frac{1}{2} e^{-|t|} \, \mathrm{d}t$$

If $x \leq 0$, we obtain:

$$\begin{aligned} \mathbf{G}(x) &= \int_{-\infty}^{x} \frac{1}{2} e^{t} \, \mathrm{d}t \\ &= \frac{1}{2} e^{x} \\ &= \frac{1}{2} - \frac{1}{2}(1 - e^{x}) \end{aligned}$$

If $x > 0$, we obtain:

$$\begin{aligned} \mathbf{G}(x) &= \frac{1}{2} e^{0} + \int_{0}^{x} \frac{1}{2} e^{-t} \, \mathrm{d}t \\ &= \frac{1}{2} + \frac{1}{2}(1 - e^{x}) \end{aligned}$$

We deduce that:

$$\mathbf{G}(x) = \frac{1}{2} + \frac{1}{2} \operatorname{sign}(x) \left(1 - e^{-x}\right)$$

and:

$$\mathbf{G}^{-1}(u) = -\operatorname{sign}\left(u - \frac{1}{2}\right) \ln\left(1 - 2\left|u - \frac{1}{2}\right|\right)$$

To simulate the Laplace distribution, we consider the following transformation:

$$x \leftarrow -\operatorname{sign}\left(u - \frac{1}{2}\right) \ln\left(1 - 2\left|u - \frac{1}{2}\right|\right)$$

where u is a uniform random variate.

(b) We have:

$$\begin{aligned} h(x) &= \frac{f(x)}{g(x)} \\ &= \sqrt{\frac{2}{\pi}} e^{-0.5x^2 + |x|} \end{aligned}$$

We have:

$$h'(x) = \begin{cases} -(x+1)\, h(x) & \text{if } x < 0 \\ -(x-1)\, h(x) & \text{if } x \geq 0 \end{cases}$$

There are two maxima: $x^{\star} = \pm 1$. We deduce that:

$$\begin{aligned} c &= \max\left(h(-1), h(1)\right) \\ &= \sqrt{\frac{2}{\pi}} e^{0.5} \\ &\approx 1.32 \end{aligned}$$

The functions $f(x)$ and $cg(x)$ are reported in Figure 13.4.

(c) The acceptance-rejection algorithm becomes:

i. Generate two independent uniform random variates u_1 and u_2;

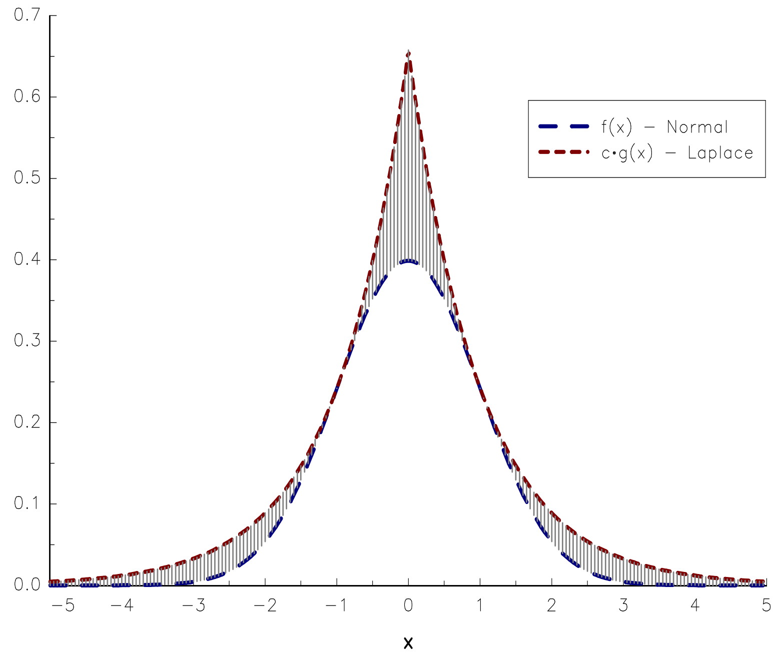

FIGURE 13.4: Rejection sampling applied to the normal distribution

ii. Calculate $x = \operatorname{sign}(u_1 - 0.5)\ln(1 - 2|u_1 - 0.5|)$;

iii. Calculate v such that:

$$\begin{aligned} v &= \frac{f(x)}{cg(x)} \\ &= e^{-0.5\left(x^2-1\right)+|x|} \end{aligned}$$

iv. If $u_2 \leq v$, accept x; otherwise, reject x.

4. (a) We have:

$$\begin{aligned} h(x) &= \frac{f(x)}{g(x)} \\ &= \frac{\pi\left(1+x^2\right)}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \\ &= \frac{\pi}{\Gamma(\alpha)}\left(x^{\alpha-1}+x^{\alpha+1}\right) e^{-x} \\ &= \frac{\pi}{\Gamma(\alpha)}\left(e^{(\alpha-1)\ln x}+e^{(\alpha+1)\ln x}\right) e^{-x} \end{aligned}$$

We deduce that:

$$\begin{aligned} h(x) &\leq \frac{\pi}{\Gamma(\alpha)}\left(e^{(\alpha+1)\ln x}+e^{(\alpha+1)\ln x}\right) e^{-x} \\ &= \frac{2\pi}{\Gamma(\alpha)} e^{(\alpha+1)\ln x} e^{-x} \\ &= \frac{2\pi}{\Gamma(\alpha)} x^{\alpha+1} e^{-x} \end{aligned}$$

We have:

$$\left(x^{\alpha+1}e^{-x}\right)' = \left(\left(\alpha+1\right)-x\right)x^{\alpha}e^{-x}$$

The maximum is reached at the point $x^{\star} = \alpha + 1$. We deduce that:

$$c = \frac{2\pi}{\Gamma\left(\alpha\right)}\left(\alpha+1\right)^{\alpha+1}e^{-(\alpha+1)}$$

(b) We have:

$$\begin{aligned} g\left(x\right) &= \frac{\Gamma\left(3/2\right)}{\Gamma\left(1\right)\sqrt{2\pi}}\left(1+\frac{x^{2}}{2}\right)^{-3/2} \\ &= \frac{1}{2\sqrt{2}}\left(1+\frac{x^{2}}{2}\right)^{-3/2} \\ &= \left(2+x^{2}\right)^{-3/2} \end{aligned}$$

and:

$$\begin{aligned} \mathbf{G}\left(x\right) &= \int_{-\infty}^{x}\left(2+t^{2}\right)^{-3/2}\,\mathrm{d}t \\ &= \left[\frac{t}{2\sqrt{2+t^{2}}}\right]_{-\infty}^{x} \\ &= \frac{1}{2}\left(1+\frac{x}{\sqrt{2+x^{2}}}\right) \end{aligned}$$

We calculate the inverse function $\mathbf{G}^{-1}\left(u\right)$:

$$\begin{aligned} \frac{1}{2}\left(1+\frac{x}{\sqrt{2+x^{2}}}\right)=u &\Leftrightarrow \frac{x^{2}}{2+x^{2}}=\left(2u-1\right)^{2} \\ &\Leftrightarrow x^{2}=2\left(2u-1\right)^{2}+x^{2}\left(2u-1\right)^{2} \\ &\Leftrightarrow x^{2}=2\left(2u-1\right)^{2}+x^{2}\left(2u-1\right)^{2} \\ &\Leftrightarrow x^{2}=\frac{2\left(2u-1\right)^{2}}{\left(4u^{2}-4u\right)} \\ &\Leftrightarrow \mathbf{G}^{-1}\left(u\right)=\frac{\sqrt{2}\left(u-0.5\right)}{\sqrt{u^{2}-u}} \end{aligned}$$

It follows that we can simulate the Student t distribution with 2 degrees of freedom by using the following transformation:

$$x \leftarrow \frac{\sqrt{2}\left(u-0.5\right)}{\sqrt{u^{2}-u}}$$

(c) In Figure 13.5, we show the acceptance ratio $p = 1/c$ for the two algorithms. It is obvious that algorithm (b) dominates algorithm (a). In particular, the acceptance ratio tends to 0 when α tends to infinity when we use the Cauchy distribution as the proposal distribution.

5. (a) We have:

$$\begin{aligned} c &= \sup\frac{p\left(k\right)}{q\left(k\right)} \\ &= K\times\max p\left(k\right) \end{aligned}$$

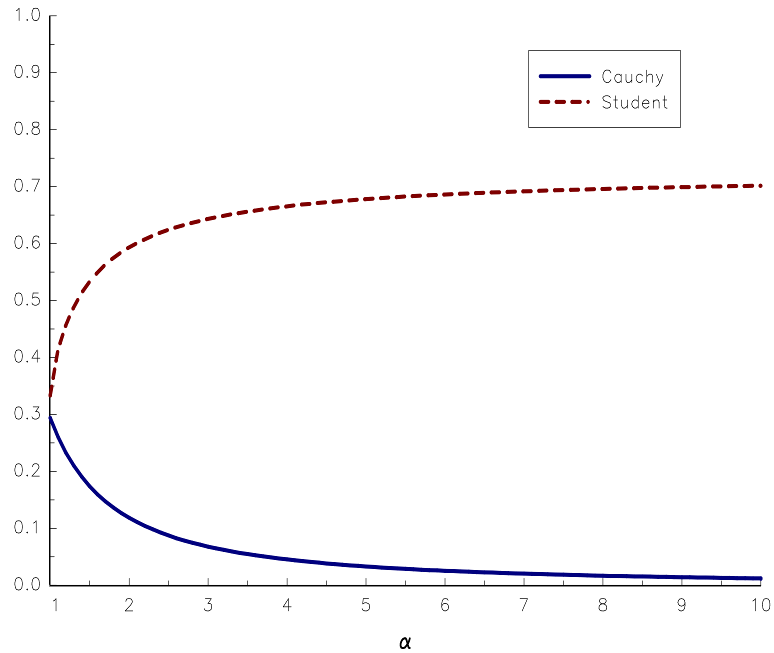

FIGURE 13.5: Acceptance ratio for the Gamma distribution

(b) We obtain $c = 5 \times 40\% = 2$. Therefore, the acceptance ratio is equal to 50% and we reject one simulation in two. This is confirmed by Figure 13.6, which shows the number of accepted and rejected values. However, the acceptance ratio is not the same for each states. For instance, it is equal to 100% for the state, which has the highest probability, but it can be low for states with small probabilities. In our experiment, we obtain the following results:

k	$f_A(k)$	$f_A^\star(k)$	$f_R(k)$	$f_R^\star(k)$
1	4.9%	9.7%	16.1%	32.5%
2	9.9%	19.6%	8.5%	17.1%
3	19.6%	38.9%	0.0%	0.0%
4	10.2%	20.2%	9.2%	18.5%
5	5.8%	11.5%	15.8%	31.9%
sum	50.4%	100.0%	49.6%	100.0%

where $f_A(k)$ and $f_R(k)$ are the frequencies of accepted and rejected values, and $f_A^\star(k)$ and $f_R^\star(k)$ are the normalized frequencies. We have rejected 49.6% of simulated values on average. Among these rejected values, 32.5% comes from the first state, 17.1% from the second state, etc. We also verify that the empirical frequencies $f_A^\star(k)$ are close to the theoretical probabilities $p(k)$.

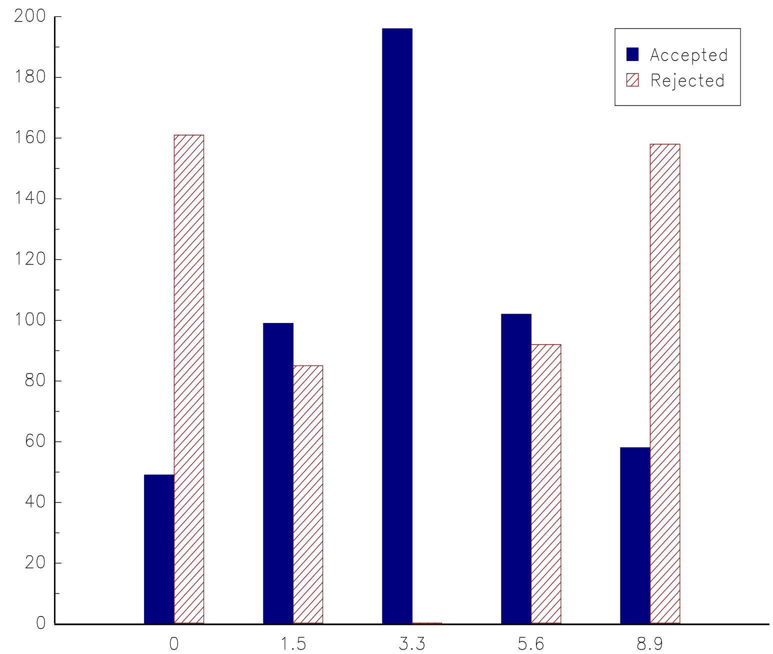

FIGURE 13.6: Histogram of accepted and rejected values

13.4.4 Simulation of Archimedean copulas

1. Let f be a function. We note $y = f(x)$. We have $\mathrm{d}y = \partial_x f(x)\,\mathrm{d}x$, $x = f^{-1}(y)$ and $\mathrm{d}x = \partial_y f^{-1}(y)\,\mathrm{d}y$. We deduce that:

$$\begin{aligned}\partial_y f^{-1}(y) &= \frac{1}{\partial_x f(x)} \\ &= \frac{1}{\partial_x f\left(f^{-1}(y)\right)}\end{aligned}$$

   We then obtain the conditional copula function:

$$\mathbf{C}_{2|1}(u_2 \mid u_1) = \frac{\varphi'(u_1)}{\varphi'\left(\varphi^{-1}\left(\varphi(u_1) + \varphi(u_2)\right)\right)}$$

   Let v_1 and v_2 be two independent uniform random variates. The simulation algorithm based on the conditional distribution is:

$$\begin{cases} u_1 = v_1 \\ \mathbf{C}_{2|1}(u_2 \mid u_1) = v_2 \end{cases}$$

   We deduce that:

$$\begin{cases} u_1 = v_1 \\ u_2 = \varphi^{-1}\left(\varphi\left(\varphi'^{-1}\left(\frac{\varphi'(v_1)}{v_2}\right)\right) - \varphi(v_1)\right) \end{cases}$$

   This is the Genest-MacKay algorithm.

2. We obtain the Gumbel-Hougaard copula:

$$\mathbf{C}(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right)$$

3. Using the Gumbel-Hougaard copula, we have $\varphi(u) = (-\ln u)^\theta$, $\varphi^{-1}(u) = \exp\left(-u^{1/\theta}\right)$ and $\varphi'(u) = -\theta u^{-1}(-\ln u)^{\theta-1}$. However, it is not possible to obtain an explicit formula for $\varphi'^{-1}(u)$. This is why we use a numerical solution $\psi(u)$ for $\varphi'^{-1}(u)$. Finally, we obtain the following simulation algorithm:

$$\begin{cases} u_1 = v_1 \\ u_2 = \exp\left(-\left[\left(-\ln\left(\psi\left(\frac{-\theta(-\ln v_1)^{\theta-1}}{v_1 v_2}\right)\right)\right)^\theta - (-\ln v_1)^\theta\right]^{1/\theta}\right) \end{cases}$$

4. We have:

$$\mathbf{C}_{2|1}(u_2 \mid u_1) = \partial_1 \mathbf{C}(u_1, u_2) = \frac{e^{-\theta u_1}\left(e^{-\theta u_2} - 1\right)}{\left(e^{-\theta} - 1\right) + \left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}$$

We deduce that:

$$\begin{aligned} & \mathbf{C}_{2|1}(u_2 \mid u_1) = v \\ \Leftrightarrow\quad & e^{-\theta u_1} e^{-\theta u_2} - e^{-\theta u_1} = v e^{-\theta} - v e^{-\theta u_1} + v\left(e^{-\theta u_1} - 1\right) e^{-\theta u_2} \\ \Leftrightarrow\quad & e^{-\theta u_2}\left((1-v) e^{-\theta u_1} + v\right) = (1-v) e^{-\theta u_1} + v e^{-\theta} \\ \Leftrightarrow\quad & u_2 = -\frac{1}{\theta} \ln\left(1 + \frac{v\left(e^{-\theta} - 1\right)}{v + (1-v) e^{-\theta u_1}}\right) \end{aligned}$$

Finally, we obtain the following simulation algorithm:

$$\begin{cases} u_1 = v_1 \\ u_2 = -\frac{1}{\theta} \ln\left(1 + \frac{v_2\left(e^{-\theta} - 1\right)}{v_2 + (1-v_2) e^{-\theta v_1}}\right) \end{cases}$$

5. We have:

$$\begin{aligned} & \varphi(u) = v \\ \Leftrightarrow\quad & \ln \frac{1 - \theta(1-u)}{u} = v \\ \Leftrightarrow\quad & 1 - \theta(1-u) = u e^v \\ \Leftrightarrow\quad & \varphi^{-1}(v) = u = \frac{1-\theta}{e^v - \theta} \end{aligned}$$

It follows that:

$$\begin{aligned} \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) &= \frac{1-\theta}{\exp\left(\ln \frac{1-\theta(1-u_1)}{u_1} + \ln \frac{1-\theta(1-u_2)}{u_2}\right) - \theta} \\ &= \frac{(1-\theta) u_1 u_2}{(1 - \theta(1-u_1))(1 - \theta(1-u_2)) - \theta u_1 u_2} \end{aligned}$$

The denominator is equal to:

$$\begin{aligned} D &= (1 - \theta(1-u_1))(1 - \theta(1-u_2)) - \theta u_1 u_2 \\ &= 1 - \theta(1-u_1) - \theta(1-u_2) + \theta^2 (1-u_1)(1-u_2) - \theta u_1 u_2 \\ &= 1 - 2\theta + \theta u_1 + \theta u_2 - \theta u_1 u_2 + \theta^2 (1-u_1)(1-u_2) \\ &= (1-\theta) - (1-\theta)\theta(1-u_1)(1-u_2) \\ &= (1-\theta)(1 - \theta(1-u_1)(1-u_2)) \end{aligned}$$

We finally obtain:

$$\varphi^{-1}\left(\varphi\left(u_1\right)+\varphi\left(u_2\right)\right)=\frac{u_1 u_2}{1-\theta\left(1-u_1\right)\left(1-u_2\right)}$$

The conditional copula is given by:

$$\begin{aligned}
\mathbf{C}_{2|1}\left(u_2 \mid u_1\right) &= \partial_1 \mathbf{C}\left(u_1, u_2\right) \\
&= \frac{u_2\left(1-\theta\left(1-u_1\right)\left(1-u_2\right)\right)-\theta u_1 u_2\left(1-u_2\right)}{\left(1-\theta\left(1-u_1\right)\left(1-u_2\right)\right)^2} \\
&= \frac{(1-\theta) u_2+\theta u_2^2}{\left(1-\theta\left(1-u_1\right)\left(1-u_2\right)\right)^2} \\
&= \frac{(1-\theta) u_2+\theta u_2^2}{\left(1-\theta+\theta u_1+\theta u_2\left(1-u_1\right)\right)^2}
\end{aligned}$$

To find the inverse conditional copula $\mathbf{C}_{2|1}^{-1}$, we have to solve the equation $\mathbf{C}_{2|1}\left(u_2 \mid u_1\right)=v$. It follows that:

$$(1-\theta) u_2+\theta u_2^2=v\left(1-\theta+\theta u_1+\theta u_2\left(1-u_1\right)\right)^2$$

or:

$$\begin{aligned}
(1-\theta) u_2+\theta u_2^2 = & \ v\left(1-\theta+\theta u_1\right)^2+ \\
& 2 \theta v u_2\left(1-\theta+\theta u_1\right)\left(1-u_1\right)+ \\
& v \theta^2 u_2^2\left(1-u_1\right)^2
\end{aligned}$$

We obtain:

$$a_\theta\left(v, u_1\right) u_2^2+b_\theta\left(v, u_1\right) u_2+c_\theta\left(v, u_1\right)=0$$

where:

$$\begin{aligned}
a_\theta\left(v, u_1\right) &= v \theta^2\left(1-u_1\right)^2-\theta \\
b_\theta\left(v, u_1\right) &= 2 \theta v\left(1-\theta+\theta u_1\right)\left(1-u_1\right)-(1-\theta) \leq 0 \\
c_\theta\left(v, u_1\right) &= v\left(1-\theta+\theta u_1\right)^2 \geq 0
\end{aligned}$$

We deduce that the solution is equal to:

$$u_2=\Psi_\theta\left(v, u_1\right)=\frac{-b_\theta\left(v, u_1\right)-\sqrt{b_\theta^2\left(v, u_1\right)-4 a_\theta\left(v, u_1\right) c_\theta\left(v, u_1\right)}}{2 a_\theta\left(v, u_1\right)}$$

Finally, we obtain the following simulation algorithm:

$$\left\{\begin{array}{l}
u_1=v_1 \\
u_2=\Psi_\theta\left(v_2, v_1\right)
\end{array}\right.$$

6. Using the previous algorithms, we obtain the following simulated random vectors:

Gumbel $\theta=1.8$		Frank $\theta=2.1$		AMH $\theta=0.6$	
u_1	u_2	u_1	u_2	u_1	u_2
0.117	0.240	0.117	0.321	0.117	0.351
0.607	0.478	0.607	0.459	0.607	0.452
0.168	0.141	0.168	0.171	0.168	0.185
0.986	0.993	0.986	0.951	0.986	0.930
0.765	0.299	0.765	0.192	0.765	0.169

13.4.5 Simulation of conditional random variables

1. $Z = (X, Y)$ is a Gaussian random vector defined as follows:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\right)$$

We have:

$$\begin{aligned} \mu_t &= \mathbb{E}[T] \\ &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x^\star - \mu_x) \end{aligned}$$

and:

$$\begin{aligned} \Sigma_{tt} &= \operatorname{cov}(T) \\ &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \end{aligned}$$

It follows that:

$$T \sim \mathcal{N}\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x^\star - \mu_x), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)$$

Let P_{tt} be the Cholesky decomposition of $\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$. We have:

$$T = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x^\star - \mu_x) + P_{tt}U$$

where $U \sim \mathcal{N}(\mathbf{0}, I)$. We deduce the following algorithm to simulate the random vector T:

(a) We simulate the vector $u = \left(u_1, \ldots, u_{n_y}\right)$ of independent Gaussian random variates $\mathcal{N}(0, 1)$;

(b) We calculate P_{tt} the Cholesky decomposition of $\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$;

(c) The simulation of the random vector T is given by:

$$t \leftarrow \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x^\star - \mu_x) + P_{tt}u$$

2. We have:

$$\begin{aligned} \mathbb{E}\left[\tilde{T}\right] &= \mathbb{E}\left[Y - \Sigma_{yx}\Sigma_{xx}^{-1}(X - x^\star)\right] \\ &= \mathbb{E}[Y] - \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbb{E}[X] - x^\star) \\ &= \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}(\mu_x - x^\star) \end{aligned}$$

We deduce that:

$$\tilde{T} - \mathbb{E}\left[\tilde{T}\right] = (Y - \mu_y) - \Sigma_{yx}\Sigma_{xx}^{-1}(X - \mu_x)$$

and:

$$\begin{aligned} \operatorname{cov}\left(\tilde{T}\right) &= \mathbb{E}\left[(Y - \mu_y)(Y - \mu_y)^\top\right] + \\ & \quad \Sigma_{yx}\Sigma_{xx}^{-1}\mathbb{E}\left[(X - \mu_x)(X - \mu_x)^\top\right]\Sigma_{xx}^{-1}\Sigma_{yx}^\top - \\ & \quad 2\mathbb{E}\left[(Y - \mu_y)(X - \mu_x)^\top\right]\Sigma_{xx}^{-1}\Sigma_{yx}^\top \\ &= \Sigma_{yy} + \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xx}\Sigma_{xx}^{-1}\Sigma_{xy} - 2\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\ &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \end{aligned}$$

As $\tilde{T}$ is a linear transformation of the Gaussian random vector Z, we obtain:

$$\tilde{T} \sim \mathcal{N}\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(x^\star - \mu_x\right), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)$$

We conclude that $\tilde{T} = T$. We deduce the following algorithm to simulate the random vector T:

(a) We simulate the vector $u = (u_1, \ldots, u_{n_z})$ of independent Gaussian random variates $\mathcal{N}(0,1)$;

(b) We calculate P_{zz} the Cholesky decomposition of Σ_{zz};

(c) We simulate the random vector Z:

$$z \leftarrow \mu_z + P_{zz}u$$

(d) We set $x = (z_1, \ldots, z_{n_x})$ and $y = \left(z_{n_x+1}, \ldots, z_{n_x+n_y}\right)$.

(e) The simulation of the random vector T is given by:

$$t \leftarrow y - \Sigma_{yx}\Sigma_{xx}^{-1}\left(x - x^\star\right)$$

3. We note $Z = (Z_1, Z_2, \ldots, Z_{n_z})$ and $Z_i(z_1, \ldots, z_{i-1}) = Z_i \mid Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}$. Let $u = (u_1, u_2, \ldots, u_{n_z})$ be a vector of independent Gaussian random variates $\mathcal{N}(0,1)$. To simulate $Z_i(z_1, \ldots, z_{i-1})$, we consider the following iteration from $i = 2$ to $i = n_z$:

$$\begin{aligned} z_i \quad \leftarrow \quad & \mu_i + \Sigma_{i,1:i-1}\Sigma_{1:i-1,1:i-1}^{-1}\left(z_{1:i-1} - \mu_{1:i-1}\right) + \\ & \sqrt{\Sigma_{i,i} - \Sigma_{i,1:i-1}\Sigma_{1:i-1,1:i-1}^{-1}\Sigma_{1:i-1,i}}\,u_i \end{aligned}$$

with:

$$z_1 \leftarrow \mu_1 + \sqrt{\Sigma_{1,1}}u_1$$

4. We obtain the following results:

z_1	−0.562	0.437	0.427	0.404	1.984
z_2	1.963	2.225	2.234	1.287	2.059
z_3	−0.808	4.013	7.643	−3.471	3.236

13.4.6 Simulation of the bivariate Normal copula

1. P is a lower triangular matrix such that we have $\Sigma = PP^\top$. We know that:

$$P = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

We verify that:

$$\begin{aligned} PP^\top &= \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}\begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \end{aligned}$$

We deduce that:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}\begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

where N_1 and N_2 are two independent standardized Gaussian random variables. Let n_1 and n_2 be two independent random variates, whose probability distribution is $\mathcal{N}(0,1)$. Using the Cholesky decomposition, we deduce that can simulate X in the following way:

$$\begin{cases} x_1 \leftarrow n_1 \\ x_2 \leftarrow \rho n_1 + \sqrt{1-\rho^2} n_2 \end{cases}$$

2. We have

$$\begin{aligned} \mathbf{C}\langle X_1, X_2 \rangle &= \mathbf{C}\langle \Phi(X_1), \Phi(X_2) \rangle \\ &= \mathbf{C}\langle U_1, U_2 \rangle \end{aligned}$$

because the function $\Phi(x)$ is non-decreasing. The copula of $U = (U_1, U_2)$ is then the copula of $X = (X_1, X_2)$.

3. We deduce that we can simulate U with the following algorithm:

$$\begin{cases} u_1 \leftarrow \Phi(x_1) = \Phi(n_1) \\ u_2 \leftarrow \Phi(x_2) = \Phi\left(\rho n_1 + \sqrt{1-\rho^2} n_2\right) \end{cases}$$

4. Let X_3 be a Gaussian random variable, which is independent from X_1 and X_2. Using the Cholesky decomposition, we know that:

$$X_2 = \rho X_1 + \sqrt{1-\rho^2} X_3$$

It follows that:

$$\begin{aligned} \Pr\{X_2 \leq x_2 | X_1 = x\} &= \Pr\left\{\rho X_1 + \sqrt{1-\rho^2} X_3 \leq x_2 \,\middle|\, X_1 = x\right\} \\ &= \Pr\left\{X_3 \leq \frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right\} \\ &= \Phi\left(\frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right) \end{aligned}$$

Then we deduce that:

$$\begin{aligned} \Phi_2(x_1, x_2; \rho) &= \Pr\{X_1 \leq x_1, X_2 \leq x_2\} \\ &= \Pr\left\{X_1 \leq x_1, X_3 \leq \frac{x_2 - \rho X_1}{\sqrt{1-\rho^2}}\right\} \\ &= \mathbb{E}\left[\Pr\left\{X_1 \leq x_1, X_3 \leq \frac{x_2 - \rho X_1}{\sqrt{1-\rho^2}} \,\middle|\, X_1\right\}\right] \\ &= \int_{-\infty}^{x_1} \Phi\left(\frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right) \phi(x) \,\mathrm{d}x \end{aligned}$$

5. Using the relationships $u_1 = \Phi(x_1)$, $u_2 = \Phi(x_2)$ and $\Phi_2(x_1, x_2; \rho) = \mathbf{C}(\Phi(x_1), \Phi(x_2); \rho)$, we obtain:

$$\begin{aligned} \mathbf{C}(u_1, u_2; \rho) &= \int_{-\infty}^{\Phi^{-1}(u_1)} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho x}{\sqrt{1-\rho^2}}\right) \phi(x) \,\mathrm{d}x \\ &= \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho \Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \mathrm{d}u \end{aligned}$$

6. We have:

$$\begin{aligned} \mathbf{C}_{2|1}(u_2 \mid u_1) &= \partial_{u_1} \mathbf{C}(u_1, u_2) \\ &= \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \end{aligned}$$

Let v_1 and v_2 be two independent uniform random variates. The simulation algorithm corresponds to the following steps:

$$\begin{cases} u_1 = v_1 \\ \mathbf{C}_{2|1}(u_1, u_2) = v_2 \end{cases}$$

We deduce that:

$$\begin{cases} u_1 \leftarrow v_1 \\ u_2 \leftarrow \Phi\left(\rho\Phi^{-1}(v_1) + \sqrt{1-\rho^2}\Phi^{-1}(v_2)\right) \end{cases}$$

7. We obtain the same algorithm, because we have the following correspondence:

$$\begin{cases} v_1 = \Phi(n_1) \\ v_2 = \Phi(n_2) \end{cases}$$

The algorithm described in Question 6 is then a special case of the Cholesky algorithm if we take $n_1 = \Phi^{-1}(v_1)$ and $n_2 = \Phi^{-1}(v_2)$. Whereas n_1 and n_2 are directly simulated in the Cholesky algorithm with a Gaussian random generator, they are simulated using the inverse transform in the conditional distribution method.

13.4.7 Computing the capital charge for operational risk

1. We obtain the following results:

α	$\mathbb{E}\left[\widehat{\mathrm{CaR}}_1(\alpha)\right]$	$\sigma\left(\widehat{\mathrm{CaR}}_1(\alpha)\right)$	$\mathrm{IC}_{95\%}\left(\widehat{\mathrm{CaR}}_1(\alpha)\right)$
90%	251 660	180	0.28%
95%	294 030	280	0.37%
99%	414 810	885	0.84%
99.9%	708 840	5 410	2.99%

where $\mathrm{IC}_{95\%}\left(\widehat{\mathrm{CaR}}_1(\alpha)\right)$ is the 95% confidence interval ratio:

$$\underset{95\%}{\mathrm{IC}}\left(\widehat{\mathrm{CaR}}_1(\alpha)\right) = 2 \times \Phi^{-1}(97.5\%) \times \frac{\sigma\left(\widehat{\mathrm{CaR}}_1(\alpha)\right)}{\mathbb{E}\left[\widehat{\mathrm{CaR}}_1(\alpha)\right]}$$

Because this ratio is lower than 5%, we conclude that one million of simulations is sufficient even if α is equal to 99.9%.

2. The results become:

α	$\mathbb{E}\left[\widehat{\mathrm{CaR}}_2(\alpha)\right]$	$\sigma\left(\widehat{\mathrm{CaR}}_2(\alpha)\right)$	$\mathrm{IC}_{95\%}\left(\widehat{\mathrm{CaR}}_2(\alpha)\right)$
90%	183 560	128	0.27%
95%	218 950	223	0.40%
99%	332 870	916	1.08%
99.9%	662 420	6 397	3.79%

We conclude that one million of simulations is sufficient to calculate the capital-at-risk.

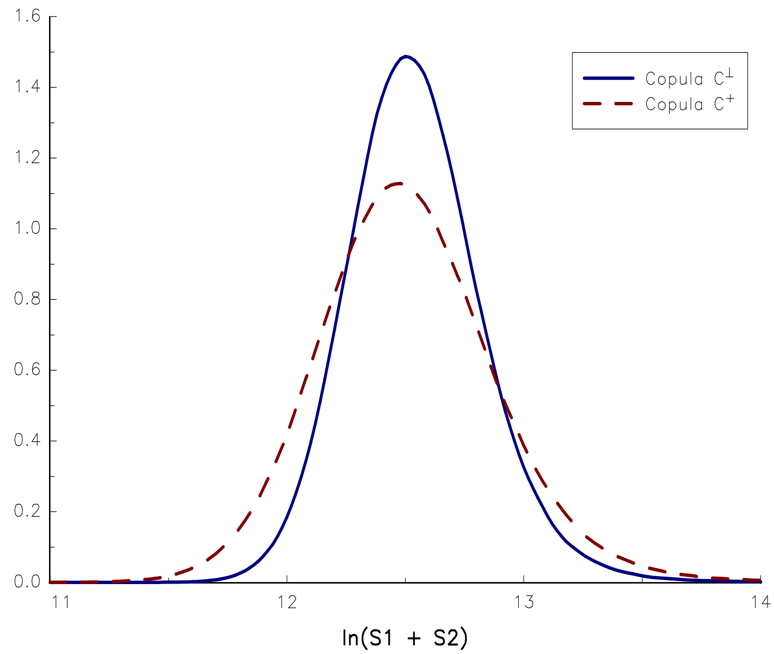

FIGURE 13.7: probability density function of $\ln(S_1 + S_2)$

3. In Figure 13.7, we have represented the probability density function of $\ln(S_1 + S_2)$ when the aggregate losses S_1 and S_2 are independent (copula $\mathbf{C}^{\perp}$) and perfectly dependent (copula $\mathbf{C}^{+}$). We obtain the following capital-at-risk:

α	$\mathbf{C}^{\perp}$	$\mathbf{C}^{+}$	$\mathrm{DR}\left(\mathbf{C}^{\perp} \mid \mathbf{C}^{+}\right)$
90%	400 240	435 220	8.04%
95%	453 864	512 980	11.52%
99%	605 927	747 680	18.96%
99.9%	993 535	1 371 260	27.55%

where $\mathrm{DR}\left(\mathbf{C}^{\perp} \mid \mathbf{C}^{+}\right)$ is the diversification ratio.

4. In Figure 13.8, we have reported the capital-at-risk calculated with the Normal copula and the Gaussian approximation defined as:

$$\begin{aligned}\widehat{\mathrm{CaR}}(\alpha) &= \bar{S}_1 + \bar{S}_2 + \sqrt{\left(\widehat{\mathrm{CaR}}_1(\alpha) - \bar{S}_1\right)^2 + \left(\widehat{\mathrm{CaR}}_2(\alpha) - \bar{S}_2\right)^2 + \ldots} \\ &\quad \overline{\ldots + 2\rho\left(\widehat{\mathrm{CaR}}_1(\alpha) - \bar{S}_1\right)\left(\widehat{\mathrm{CaR}}_2(\alpha) - \bar{S}_2\right)}\end{aligned}$$

5. Results are given in Figure 13.9.

6. Results are given in Figure 13.10.

7. For a high value of the quantile ($\alpha = 99.9\%$), the Gaussian approximation overestimates (resp. underestimates) the capital-at-risk when the dependence function is the Normal (resp. t_1) copula. We obtain this result, because the Student t_1 copula produces strong dependence when the correlation parameter ρ is equal to zero. We conclude that the Gaussian approximation is good in this example, except if the copula function highly correlates the extreme losses.

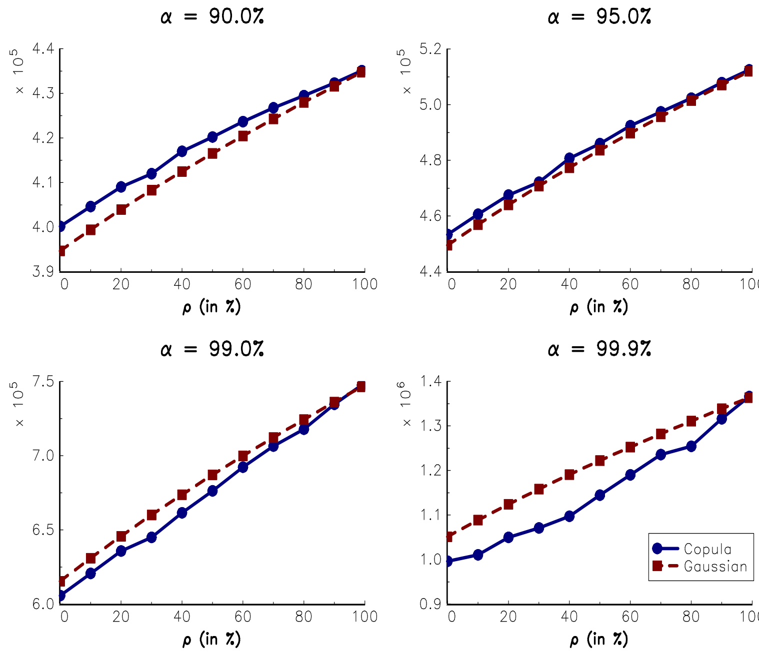

FIGURE 13.8: Capital-at-risk with the Normal copula

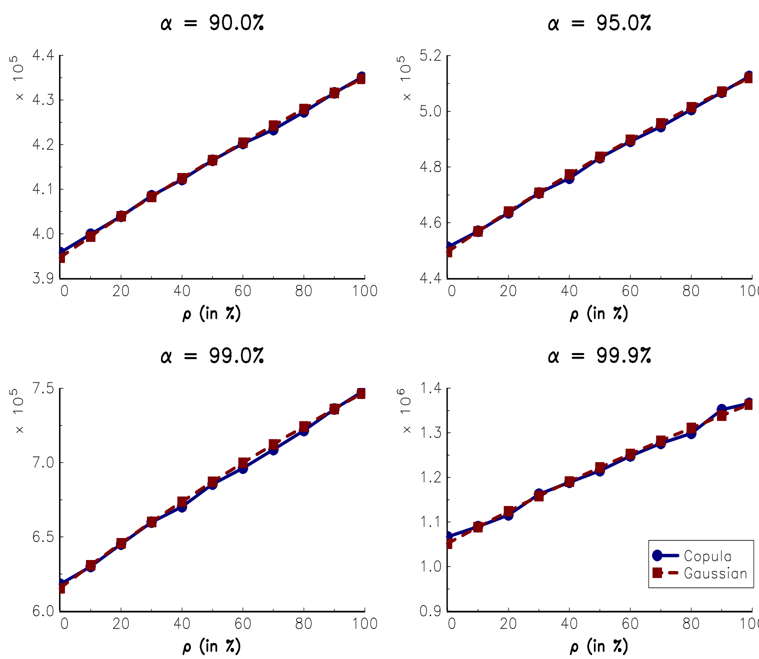

FIGURE 13.9: Capital-at-risk with the t_4 copula

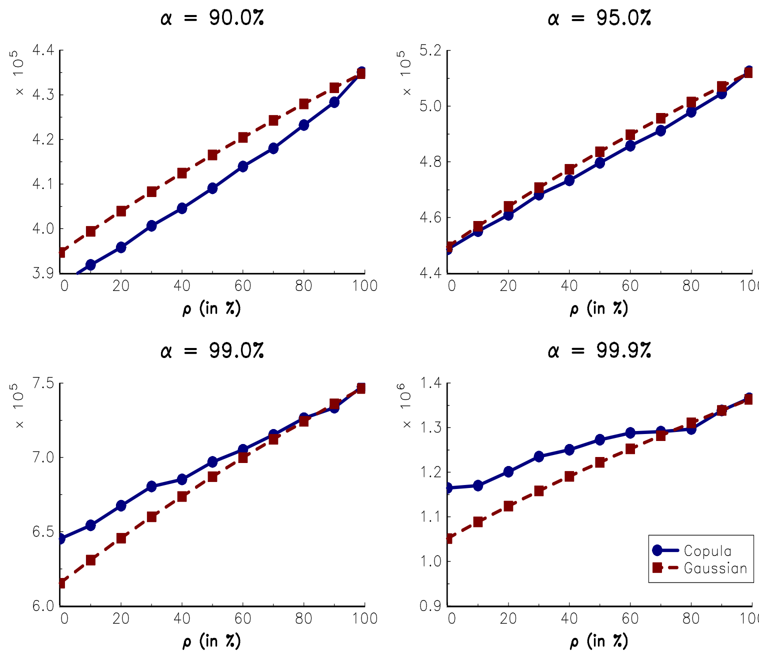

FIGURE 13.10: Capital-at-risk with the t_1 copula

13.4.8 Simulating a Brownian bridge

1. We remind that:

$$\mathbb{E}\left[W(s)\,W(t)\right] = \min(s,t)$$

We deduce that:

$$\begin{pmatrix} W(s) \\ W(t) \\ W(u) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s & s & s \\ s & t & t \\ s & t & u \end{pmatrix}\right)$$

2. We rearrange the terms of the random vector in the following way:

$$\begin{pmatrix} W(s) \\ W(u) \\ W(t) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s & s & s \\ s & u & t \\ s & t & t \end{pmatrix}\right)$$

We note:

$$B(t) = \{W(t) \mid W(s) = w_s, W(u) = w_u\}$$

We know that the conditional distribution of $W(t)$ given that $W(s) = w_s$ and $W(u) = w_u$ is Gaussian with:

$$\begin{aligned} \mathbb{E}\left[B(t)\right] &= 0 + \begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} s & s \\ s & u \end{pmatrix}^{-1} \left(\begin{pmatrix} w_s \\ w_u \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \\ &= \frac{u-t}{u-s} w_s + \frac{t-s}{u-s} w_u \end{aligned}$$

and:

$$\begin{aligned} \operatorname{var}(B(t)) &= t - \begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} s & s \\ s & u \end{pmatrix}^{-1} \begin{pmatrix} s \\ t \end{pmatrix} \\ &= \frac{(t-s)(u-t)}{u-s} \end{aligned}$$

3. We deduce that:

$$B(t) = \frac{u-t}{u-s} w_s + \frac{t-s}{u-s} w_u + \sqrt{\frac{(t-s)(u-t)}{u-s}}\varepsilon$$

where ε is a standard Gaussian random variable. To simulate $B(t)$, we then use the iterative algorithm based on filling the path and moving the starting point $(s, B(s))$ at each iteration.

13.4.9 Optimal importance sampling

1. Let X be a random variate from the distribution $\mathcal{N}(0,1)$. We have $\hat{p}_{\mathrm{MC}} = \varphi(X)$ where $\varphi = \mathbb{1}\{X \geq c\}$. We deduce that:

$$\begin{aligned} \mathbb{E}[\hat{p}_{\mathrm{MC}}] &= \mathbb{E}[\varphi(X)] \\ &= \int_{-\infty}^{\infty} \mathbb{1}\{x \geq c\} \phi(x)\,\mathrm{d}x \\ &= \int_{c}^{\infty} \phi(x)\,\mathrm{d}x \\ &= 1 - \Phi(c) \\ &= p \end{aligned}$$

Recall that $\operatorname{var}(\hat{p}_{\mathrm{MC}}) = \mathbb{E}\left[\hat{p}_{\mathrm{MC}}^2\right] - \mathbb{E}^2[\hat{p}_{\mathrm{MC}}]$. We have[1]:

$$\begin{aligned} \mathbb{E}\left[\hat{p}_{\mathrm{MC}}^2\right] &= \int_{-\infty}^{\infty} \varphi^2(x)\phi(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \mathbb{1}\{x \geq c\} \phi(x)\,\mathrm{d}x \\ &= p \end{aligned}$$

It follows that:

$$\begin{aligned} \operatorname{var}(\hat{p}_{\mathrm{MC}}) &= p - p^2 \\ &= p(1-p) \\ &= \Phi(c)(1-\Phi(c)) \end{aligned}$$

We notice that $\hat{p}_{\mathrm{MC}}$ is a Bernoulli random variable $\mathcal{B}(\Phi(c))$.

2. We note Z the random variate from the distribution $\mathcal{N}\left(\mu, \sigma^2\right)$. We have:

$$\hat{p}_{\mathrm{IS}} = \varphi(Z)\mathcal{L}(Z)$$

[1] We notice that $\varphi^2(x) = \mathbb{1}\{x \geq c\}$.

where $\mathcal{L}(Z)$ is the likelihood ratio:

$$\begin{aligned}\mathcal{L}(Z) &= \frac{f(Z)}{g(Z)} \\ &= \frac{\phi(Z)}{\frac{1}{\sigma}\phi\left(\frac{Z-\mu}{\sigma}\right)} \\ &= \sigma\exp\left(\frac{1}{2}\left(\frac{Z-\mu}{\sigma}\right)^2-\frac{1}{2}Z^2\right)\end{aligned}$$

It follows that:

$$\hat{p}_{\mathrm{IS}} = \mathbb{1}\{Z\geq c\}\,\sigma e^{\frac{1}{2}\left(\frac{Z-\mu}{\sigma}\right)^2-\frac{1}{2}Z^2}$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[\hat{p}_{\mathrm{IS}}\right] &= \mathbb{E}\left[\mathbb{1}\{Z\geq c\}\,\sigma e^{\frac{1}{2}\left(\frac{Z-\mu}{\sigma}\right)^2-\frac{1}{2}Z^2}\right] \\ &= \int_{-\infty}^{\infty}\mathbb{1}\{z\geq c\}\,e^{\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2-\frac{1}{2}z^2}\phi\left(\frac{z-\mu}{\sigma}\right)\,\mathrm{d}z \\ &= \int_{c}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,\mathrm{d}z \\ &= 1-\Phi(c) \\ &= p\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}\left[\hat{p}_{\mathrm{IS}}^2\right] &= \mathbb{E}\left[\mathbb{1}\{Z\geq c\}\,\sigma^2 e^{\left(\frac{Z-\mu}{\sigma}\right)^2-Z^2}\right] \\ &= \int_{c}^{\infty}\frac{\sigma}{\sqrt{2\pi}}e^{\left(\frac{z-\mu}{\sigma}\right)^2-z^2}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\,\mathrm{d}z \\ &= \int_{c}^{\infty}\frac{\sigma}{\sqrt{2\pi}}e^{\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2-z^2}\,\mathrm{d}z\end{aligned}$$

We have:

$$\begin{aligned}\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2-z^2 &= \frac{z^2-2\mu z+\mu^2-2\sigma^2 z^2}{2\sigma^2} \\ &= \left(\frac{1-2\sigma^2}{2\sigma^2}\right)\left(z^2-\frac{2\mu}{1-2\sigma^2}z+\frac{\mu^2}{1-2\sigma^2}\right) \\ &= \left(\frac{1-2\sigma^2}{2\sigma^2}\right)\left(\left(z-\frac{\mu}{1-2\sigma^2}\right)^2-2\frac{\mu^2\sigma^2}{\left(1-2\sigma^2\right)^2}\right) \\ &= \left(\frac{1-2\sigma^2}{2\sigma^2}\right)\left(z-\frac{\mu}{1-2\sigma^2}\right)^2-\frac{\mu^2}{1-2\sigma^2}\end{aligned}$$

We note[2]:

$$\tilde{\mu} = \frac{\mu}{1-2\sigma^2}$$

and

$$\tilde{\sigma} = \frac{\sigma}{\sqrt{2\sigma^2-1}}$$

[2]We assume that $2\sigma^2-1>0$.

It follows that:

$$\begin{aligned}
\mathbb{E}\left[\hat{p}_{\mathrm{IS}}^2\right] &= e^{-\frac{\mu^2}{1-2\sigma^2}} \int_c^\infty \frac{\sigma}{\sqrt{2\pi}} e^{\left(\frac{1-2\sigma^2}{2\sigma^2}\right)\left(z-\frac{\mu}{1-2\sigma^2}\right)^2} \,\mathrm{d}z \\
&= e^{-\frac{\mu^2}{1-2\sigma^2}} \int_c^\infty \frac{\sigma\tilde{\sigma}}{\tilde{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\tilde{\mu}}{\tilde{\sigma}}\right)^2} \,\mathrm{d}z \\
&= \frac{\sigma^2}{\sqrt{2\sigma^2-1}} e^{-\frac{\mu^2}{1-2\sigma^2}} \int_c^\infty \frac{1}{\tilde{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\tilde{\mu}}{\tilde{\sigma}}\right)^2} \,\mathrm{d}z \\
&= \frac{\sigma^2}{\sqrt{2\sigma^2-1}} e^{-\frac{\mu^2}{1-2\sigma^2}} \left(1 - \Phi\left(\frac{c-\tilde{\mu}}{\tilde{\sigma}}\right)\right)
\end{aligned}$$

We conclude that[3]:

$$\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) = \frac{\sigma^2}{\sqrt{2\sigma^2-1}} e^{-\frac{\mu^2}{1-2\sigma^2}} \left(1 - \Phi\left(\frac{c\left(2\sigma^2-1\right)+\mu}{\sigma\sqrt{2\sigma^2-1}}\right)\right) - \left(1-\Phi\left(c\right)\right)^2$$

The probability distribution of $\hat{p}_{\mathrm{IS}}$ is no longer a Bernoulli distribution.

3. We have $\operatorname{var}\left(\hat{p}_{\mathrm{MC}}\right) = 13.48 \times 10^{-4}$. In Figure 13.11, we report the relationship between μ and $\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right)$ for different values of σ. We find that the minimum value is approximately obtained for the same value of μ:

σ	$\mu^\star$	$\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) \times 10^{-4}$	$\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) / \operatorname{var}\left(\hat{p}_{\mathrm{MC}}\right)$
0.80	3.158	0.05	0.34%
1.00	3.154	0.06	0.45%
2.00	3.151	0.14	1.03%
3.00	3.151	0.22	1.60%

Therefore, we can make the hypothesis that the optimal value of μ does not highly depend on the parameter σ.

4. When σ is equal to 1, we obtain:

$$\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) = e^{\mu^2}\left(1-\Phi\left(c+\mu\right)\right) - \left(1-\Phi\left(c\right)\right)^2$$

The IS scheme is optimal if the variance $\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right)$ is minimum. The first-order condition is then:

$$\frac{\partial \operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right)}{\partial\,\mu} = 2\mu e^{\mu^2}\left(1-\Phi\left(c+\mu\right)\right) - e^{\mu^2}\phi\left(c+\mu\right) = 0$$

We deduce that the optimal value $\mu^\star$ satisfies the following nonlinear equation:

$$2\mu^\star\left(1-\Phi\left(c+\mu^\star\right)\right) = \phi\left(c+\mu^\star\right)$$

In Figure 13.12, we draw the relationship between c and $\mu^\star$. We notice that:

$$\lim_{c\to\infty} \mu^\star = c$$

We can then consider $\mu = c$. In Figure 13.12, we also report the variance ratio $\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) / \operatorname{var}\left(\hat{p}_{\mathrm{MC}}\right)$ for the two schemes $\mu = \mu^\star$ and $\mu = c$. We conclude that we obtain similar variance reduction with the heuristic scheme when $c > 1$.

[3] In the case where $\mu = 0$ and $\sigma = 1$, we retrieve the formula of the MC estimator:

$$\begin{aligned}
\operatorname{var}\left(\hat{p}_{\mathrm{IS}}\right) &= \left(1-\Phi\left(c\right)\right) - \left(1-\Phi\left(c\right)\right)^2 \\
&= \Phi\left(c\right) - \Phi^2\left(c\right)
\end{aligned}$$

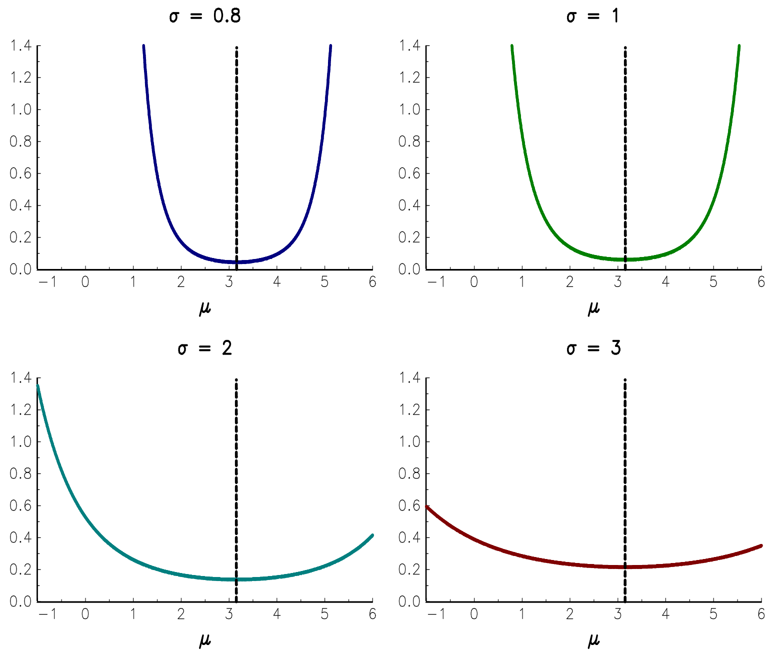

FIGURE 13.11: Variance of the IS estimator $\hat{p}_{\mathrm{IS}}$ $(\times 10^4)$

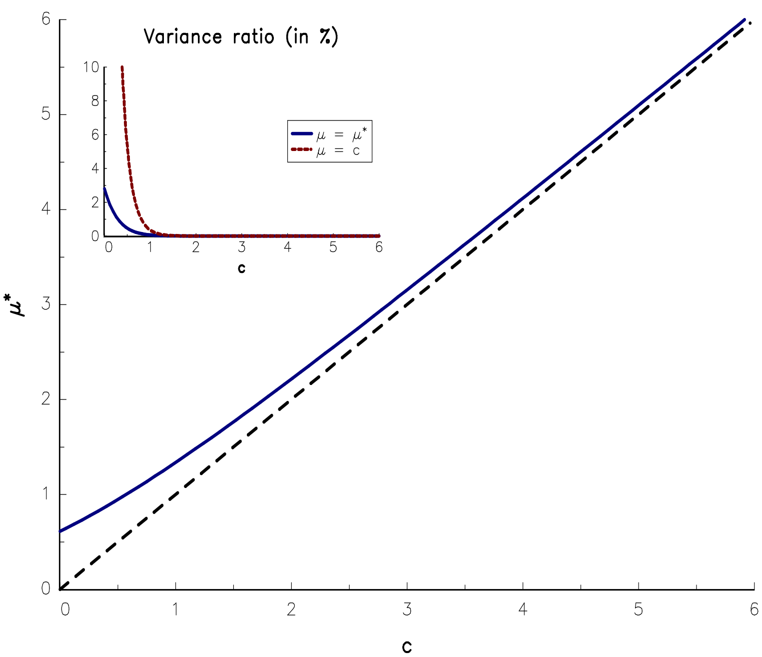

FIGURE 13.12: Optimal value $\mu^\star$ with respect to c

Chapter 14

Stress Testing and Scenario Analysis

14.3.1 Construction of a stress scenario with the GEV distribution

1. We recall that:

$$\begin{aligned}\Pr\left\{\frac{X_{n:n}-b_n}{a_n}\leq x\right\} &= \Pr\left\{X_{n:n}\leq a_n x+b_n\right\} \\ &= \mathbf{F}^n\left(a_n x+b_n\right)\end{aligned}$$

and:

$$\mathbf{G}(x)=\lim_{n\rightarrow\infty}\mathbf{F}^n\left(a_n x+b_n\right)$$

(a) We have:

$$\begin{aligned}\mathbf{F}^n\left(a_n x+b_n\right) &= \left(1-e^{-\lambda\left(\lambda^{-1}x+\lambda^{-1}\ln n\right)}\right)^n \\ &= \left(1-\frac{1}{n}e^{-x}\right)^n\end{aligned}$$

We deduce that:

$$\mathbf{G}(x)=\lim_{n\rightarrow\infty}\left(1-\frac{1}{n}e^{-x}\right)^n=e^{-e^{-x}}=\mathbf{\Lambda}(x)$$

(b) We have:

$$\begin{aligned}\mathbf{F}^n\left(a_n x+b_n\right) &= \left(n^{-1}x+1-n^{-1}\right)^n \\ &= \left(1+\frac{1}{n}(x-1)\right)^n\end{aligned}$$

We deduce that:

$$\mathbf{G}(x)=\lim_{n\rightarrow\infty}\left(1+\frac{1}{n}(x-1)\right)^n=e^{x-1}=\mathbf{\Psi}_1(x-1)$$

(c) We have:

$$\begin{aligned}\mathbf{F}^n\left(a_n x+b_n\right) &= \left(1-\left(\frac{\theta}{\theta+\theta\alpha^{-1}n^{1/\alpha}x+\theta n^{1/\alpha}-\theta}\right)^{\alpha}\right)^n \\ &= \left(1-\left(\frac{1}{\alpha^{-1}n^{1/\alpha}x+n^{1/\alpha}}\right)^{\alpha}\right)^n \\ &= \left(1-\frac{1}{n}\left(1+\frac{x}{\alpha}\right)^{-\alpha}\right)^n\end{aligned}$$

We deduce that:

$$\mathbf{G}(x)=\lim_{n\rightarrow\infty}\left(1-\frac{1}{n}\left(1+\frac{x}{\alpha}\right)^{-\alpha}\right)^n=e^{-\left(1+\frac{x}{\alpha}\right)^{-\alpha}}=\mathbf{\Phi}_{\alpha}\left(1+\frac{x}{\alpha}\right)$$

2. The GEV distribution encompasses the three EV probability distributions. This is an interesting property, because we have not to choose between the three EV distributions. We have:

$$g(x) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

We deduce that:

$$\begin{aligned} \ell &= -\frac{n}{2}\ln\sigma^2 - \left(\frac{1+\xi}{\xi}\right)\sum_{i=1}^{n}\ln\left(1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right) - \\ & \quad \sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} \end{aligned}$$

3. We notice that:

$$\lim_{\xi\to 0}(1+\xi x)^{-1/\xi} = e^{-x}$$

Then we obtain:

$$\begin{aligned} \lim_{\xi\to 0}\mathbf{G}(x) &= \lim_{\xi\to 0}\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= \exp\left\{-\lim_{\xi\to 0}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right) \end{aligned}$$

4. (a) We have:

$$\mathbf{G}^{-1}(\alpha) = \mu - \sigma\xi^{-1}\left[1-(-\ln\alpha)^{-\xi}\right]$$

When the parameter ξ is equal to 1, we obtain:

$$\mathbf{G}^{-1}(\alpha) = \mu - \sigma\left(1-(-\ln\alpha)^{-1}\right)$$

By definition, we have $\mathcal{T} = (1-\alpha)^{-1}n$. The return period $\mathcal{T}$ is then associate to the confidence level $\alpha = 1 - n/\mathcal{T}$. We deduce that:

$$\begin{aligned} R(\mathcal{T}) &\approx -\mathbf{G}^{-1}(1-n/\mathfrak{t}) \\ &= -\left(\mu - \sigma\left(1-(-\ln(1-n/\mathcal{T}))^{-1}\right)\right) \\ &= -\left(\mu + \left(\frac{\mathcal{T}}{n}-1\right)\sigma\right) \end{aligned}$$

We then replace μ and σ by their ML estimates $\hat{\mu}$ and $\hat{\sigma}$.

(b) For Portfolio #1, we obtain:

$$r(1\text{Y}) = -\left(1\% + \left(\frac{252}{21}-1\right)\times 3\%\right) = -34\%$$

For Portfolio #2, the stress scenario is equal to:

$$r(1\text{Y}) = -\left(10\% + \left(\frac{252}{21}-1\right)\times 2\%\right) = -32\%$$

We conclude that Portfolio #1 is more risky than Portfolio #2 if we consider a stress scenario analysis.

14.3.2 Conditional expectation and linearity

1. Using the conditional distribution theorem, we have:

$$\left(\frac{Y-\mu_y}{\sigma_y}\right)=\rho_{xy}\left(\frac{X-\mu_x}{\sigma_x}\right)+\sqrt{1-\rho_{xy}^2}U$$

where $U\sim\mathcal{N}(0,1)$. It follows that:

$$Y=\left(\mu_y-\rho_{xy}\frac{\sigma_y}{\sigma_x}\mu_x\right)+\rho_{xy}\frac{\sigma_y}{\sigma_x}X+\sigma_y\sqrt{1-\rho_{xy}^2}U$$

We deduce that:

$$\left\{\begin{array}{l}\beta_0=\mu_y-\frac{\rho_{xy}\sigma_y}{\sigma_x}\mu_x\\ \beta=\frac{\rho_{xy}\sigma_y}{\sigma_x}\\ \sigma=\sigma_y\sqrt{1-\rho_{xy}^2}\end{array}\right.$$

2. We have:

$$\begin{array}{rcl}m(x)&=&\mathbb{E}[Y\mid X=x]\\ &=&\mathbb{E}[\beta_0+\beta X+\sigma U\mid X=x]\\ &=&\beta_0+\beta x+\sigma\mathbb{E}[U\mid X=x]\\ &=&\beta_0+\beta x\end{array}$$

because U and X are independent.

3. Since we have $Y=\beta_0+\beta X+\sigma U$, we deduce that:

$$\begin{array}{rcl}\tilde{Y}&=&e^Y\\ &=&e^{\beta_0+\beta X+\sigma U}\\ &=&e^{\beta_0}\tilde{X}^\beta\tilde{U}^\sigma\end{array}$$

where $\tilde{U}=e^U\sim\mathcal{LN}(0,1)$. It follows that:

$$\begin{array}{rcl}\tilde{m}(x)&=&\mathbb{E}\left[\tilde{Y}\mid\tilde{X}=x\right]\\ &=&e^{\beta_0}x^\beta\mathbb{E}\left[\tilde{U}^\sigma\right]\\ &=&e^{\beta_0+\frac{1}{2}\sigma^2}x^\beta\end{array}$$

because we have $\mathbb{E}\left[\tilde{U}^\sigma\right]=\mathbb{E}\left[e^{\sigma U}\right]=e^{\frac{1}{2}\sigma^2}$. Finally, we obtain:

$$\begin{array}{rcl}\tilde{m}(x)&=&\exp\left(\beta_0+\dfrac{1}{2}\sigma^2\right)\cdot x^\beta\\ &=&\exp\left(\mu_y-\dfrac{\rho_{xy}\sigma_y}{\sigma_x}\mu_x+\dfrac{1}{2}\sigma_y^2\left(1-\rho_{xy}^2\right)\right)\cdot x^{\frac{\rho_{xy}\sigma_y}{\sigma_x}}\end{array}$$

4. In the Gaussian case, we notice that the conditional expectation is a linear function. This is not the case for the lognormal case. The use of ordinary least squares to compute a conditional stress scenario assumes that the distribution of risk factors are Gaussian.

14.3.3 Conditional quantile and linearity

1. Using the conditional distribution theorem, we know that:

$$\mathbf{F}\left(y \mid X = x\right) = \mathcal{N}\left(\mu_{y|x}, \Sigma_{yy|x}\right)$$

where:

$$\mu_{y|x} = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(x - \mu_x\right)$$

and:

$$\Sigma_{yy|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$$

We deduce that:

$$q_\alpha\left(x\right) = \mu_{y|x} + \Phi^{-1}\left(\alpha\right)\sqrt{\Sigma_{yy|x}}$$

2. We have:

$$\begin{aligned} q_\alpha\left(x\right) &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(x - \mu_x\right) + \Phi^{-1}\left(\alpha\right)\sqrt{\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}} \\ &= \beta_0\left(\alpha\right) + \beta^\top x \end{aligned}$$

where:

$$\beta_0\left(\alpha\right) = \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x + \Phi^{-1}\left(\alpha\right)\sqrt{\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}$$

and:

$$\beta = \Sigma_{xx}^{-1}\Sigma_{xy}$$

3. We reiterate that the conditional expectation is:

$$m\left(x\right) = \beta_0 + \beta^\top x$$

where:

$$\beta_0 = \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x$$

and:

$$\beta = \Sigma_{xx}^{-1}\Sigma_{xy}$$

It follows that linear regression and quantile regression produce the same estimate β, but not the same intercept. Indeed, we have:

$$\beta_0\left(\alpha\right) = \beta_0 + \Phi^{-1}\left(\alpha\right)\sqrt{\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}$$

If $\alpha > 50\%$, the intercept of the quantile regression is larger than the intercept of the linear regression:

$$\begin{cases} \beta_0\left(\alpha\right) > \beta_0 & \text{If } \alpha > 50\% \\ \beta_0\left(\alpha\right) = \beta_0 & \text{If } \alpha = 50\% \\ \beta_0\left(\alpha\right) < \beta_0 & \text{If } \alpha < 50\% \end{cases}$$

We conclude that the median regression coincides with the linear regression if (X, Y) is Gaussian.

4. We know that $Z = \Phi^{-1}\left(\mathbf{F}_\tau\left(\tau\right)\right) \sim \mathcal{N}\left(0, 1\right)$. It follows that (X, Z) is Gaussian. The expression of the conditional quantile of Z is then:

$$q_\alpha^Z\left(x\right) = \beta_0\left(\alpha\right) + \beta^\top x$$

Since we have $Z = \Phi^{-1}\left(1 - e^{-\lambda\tau}\right)$ and:

$$\tau = -\frac{\ln\left(1 - \Phi\left(Z\right)\right)}{\lambda}$$

we conclude that the conditional quantile of the default time is:

$$q_{\alpha}^{\tau}\left(x\right) = -\frac{\ln\left(1 - \Phi\left(\beta_0\left(\alpha\right) + \beta^{\top}x\right)\right)}{\lambda}$$

5. By definition, we have $\mathrm{PD} = \mathbf{F}_{\tau}\left(\tau\right)$. We deduce that:

$$q_{\alpha}^{\mathrm{PD}}\left(x\right) = \Phi\left(\beta_0\left(\alpha\right) + \beta^{\top}x\right)$$

6. By construction, ρ is the correlation between $Z = \Phi^{-1}\left(\mathbf{F}_{\tau}\left(\tau\right)\right)$ and X. We note μ_z and $\Sigma_{zz} = \sigma_z^2$ the mean and variance of Z. Since, we have $\Sigma_{xz} = \rho\sigma_x\sigma_z$, we deduce that:

$$\begin{aligned}
\beta_0\left(\alpha\right) &= \mu_z - \Sigma_{zx}\Sigma_{xx}^{-1}\mu_x + \Phi^{-1}\left(\alpha\right)\sqrt{\Sigma_{zz} - \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz}} \\
&= \mu_z - \rho\frac{\sigma_z}{\sigma_x}\mu_x + \Phi^{-1}\left(\alpha\right)\sqrt{\sigma_z^2 - \frac{\left(\rho\sigma_x\sigma_z\right)^2}{\sigma_x^2}} \\
&= \mu_z - \rho\frac{\sigma_z}{\sigma_x}\mu_x + \Phi^{-1}\left(\alpha\right)\sigma_z\sqrt{1-\rho^2}
\end{aligned}$$

and:

$$\begin{aligned}
\beta &= \Sigma_{xx}^{-1}\Sigma_{xz} \\
&= \frac{\rho\sigma_x\sigma_z}{\sigma_x^2} \\
&= \rho\frac{\sigma_z}{\sigma_x}
\end{aligned}$$

Because[1] $\mu_z = 0$ and $\sigma_z = 1$, we finally obtain:

$$\begin{aligned}
q_{\alpha}^{\mathrm{PD}}\left(x\right) &= \Phi\left(\beta_0\left(\alpha\right) + \beta^{\top}x\right) \\
&= \Phi\left(\mu_z + \rho\frac{\sigma_z}{\sigma_x}\left(x - \mu_x\right) + \Phi^{-1}\left(\alpha\right)\sigma_z\sqrt{1-\rho^2}\right) \\
&= \Phi\left(\Phi^{-1}\left(\alpha\right)\sqrt{1-\rho^2} + \rho\frac{\left(x - \mu_x\right)}{\sigma_x}\right)
\end{aligned}$$

7. We observe that the conditional quantile of the default probability is not linear with respect to the risk factor X. However, we notice that $\Phi^{-1}\left(q_{\alpha}^{\mathrm{PD}}\left(x\right)\right)$ is a linear function of X. This is why we may use the following quantile regression in order to stress the default probability:

$$\Phi^{-1}\left(\mathrm{PD}\right) = \beta_0 + \beta^{\top}X + U$$

where X is a set of risk factors.

[1] Indeed, $\mathbf{F}_{\tau}\left(\tau\right) \sim U_{[0,1]}$ and $\Phi^{-1}\left(U_{[0,1]}\right) \sim \mathcal{N}\left(0,1\right)$.

Chapter 15

Credit Scoring Models

15.4.1 Elastic net regression

1. (a) Let $f(\beta)$ be the objective function. We have:

$$\begin{aligned} f(\beta) &= \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^\top(\mathbf{Y}-\mathbf{X}\beta)+\frac{\lambda}{2}\sum_{k=1}^{K}\beta_k^2 \\ &= \frac{1}{2}\beta^\top\mathbf{X}^\top\mathbf{X}\beta-\beta^\top\mathbf{X}^\top\mathbf{Y}+\frac{1}{2}\mathbf{Y}^\top\mathbf{Y}+\frac{\lambda}{2}\beta^\top\beta \\ &= \frac{1}{2}\beta^\top\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\beta-\beta^\top\mathbf{X}^\top\mathbf{Y}+\frac{1}{2}\mathbf{Y}^\top\mathbf{Y} \end{aligned}$$

We deduce that:

$$\frac{\partial\, f(\beta)}{\partial\,\beta}=\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\beta-\mathbf{X}^\top\mathbf{Y}$$

The first order condition $\partial_\beta\, f(\beta)=\mathbf{0}$ implies that:

$$\hat\beta^{\text{ridge}}=\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{Y}$$

(b) We recall that $\hat\beta^{\text{ols}}=\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}$. We deduce that:

$$\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\hat\beta^{\text{ridge}}=\left(\mathbf{X}^\top\mathbf{X}\right)\hat\beta^{\text{ols}}=\mathbf{X}^\top\mathbf{Y}$$

and:

$$\begin{aligned} \hat\beta^{\text{ridge}} &= \left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)^{-1}\left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}\hat\beta^{\text{ols}} \\ &= \left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\right)^{-1}\hat\beta^{\text{ols}} \\ &= \left(I_K+\lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}\hat\beta^{\text{ols}} \qquad (15.1) \end{aligned}$$

(c) It follows that:

$$\begin{aligned} \mathbb{E}\left[\hat\beta^{\text{ridge}}\right] &= \mathbb{E}\left[\left(I_K+\lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}\hat\beta^{\text{ols}}\right] \\ &= \left(I_K+\lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}\beta \end{aligned}$$

where β is the true value. If $\mathbb{E}\left[\hat\beta^{\text{ridge}}\right]=\beta$, we obtain:

$$\begin{aligned} \left(I_K+\lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}=I_K &\Leftrightarrow I_K+\lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}=I_K \\ &\Leftrightarrow \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}=\mathbf{0} \\ &\Leftrightarrow \lambda=0 \end{aligned}$$

(d) From Equation (15.1), we deduce that:

$$\begin{aligned}
\operatorname{var}\left(\hat{\beta}^{\text{ridge}}\right) &= \left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1} \operatorname{var}\left(\hat{\beta}^{\text{ols}}\right)\left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1} \\
&= \sigma^2\left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)^{-1} \\
&= \sigma^2\left(\left(I_K + \lambda\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^\top\mathbf{X} + \lambda^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} + 2\lambda I_K\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^\top\mathbf{X} + Q\right)^{-1}
\end{aligned}$$

where:

$$Q = \lambda^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} + 2\lambda I_K$$

Since Q is a symmetric positive definite matrix, we have:

$$\left(\mathbf{X}^\top\mathbf{X} + Q\right) \succeq \left(\mathbf{X}^\top\mathbf{X}\right)$$

where $\succeq$ is the positive definite ordering. Finally, we obtain:

$$\operatorname{var}\left(\hat{\beta}^{\text{ols}}\right) \succeq \operatorname{var}\left(\hat{\beta}^{\text{ridge}}\right)$$

(e) We have:

$$\begin{aligned}
\hat{\mathbf{Y}} &= \mathbf{X}\hat{\beta}^{\text{ridge}} \\
&= \mathbf{X}\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{Y} \\
&= \mathbf{H}\mathbf{Y}
\end{aligned}$$

where $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top$. We deduce that:

$$\begin{aligned}
\text{df}^{(\text{model})} &= \operatorname{tr}\left(\mathbf{X}\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\right) \\
&= \operatorname{tr}\left(\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{X}\right)
\end{aligned}$$

We consider the singular value decomposition $\mathbf{X} = USV^\top$ where U and V are two orthonormal matrices, and S is a diagonal matrix that is composed of the singular values $(s_1, \ldots, s_K)$. We have $\mathbf{X}^\top\mathbf{X} = VS^2V^\top$ and:

$$\begin{aligned}
\mathbf{X}^\top\mathbf{X} + \lambda I_K &= VS^2V^\top + \lambda VV^\top \\
&= V\left(S^2 + \lambda I_K\right)V^\top
\end{aligned}$$

It follows that $\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1} = V\left(S^2 + \lambda I_K\right)^{-1}V^\top$ and:

$$\begin{aligned}
\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{X} &= V\left(S^2 + \lambda I_K\right)^{-1}V^\top VS^2V^\top \\
&= V\left(S^2 + \lambda I_K\right)^{-1}S^2V^\top
\end{aligned}$$

We finally obtain:

$$\text{df}^{(\text{model})} = \sum_{k=1}^{K}\frac{s_k^2}{s_k^2 + \lambda}$$

(f) If $\mathbf{X}$ is an orthonormal matrix, we have $\mathbf{X}^\top \mathbf{X} = I_K$ and:

$$\begin{aligned} \hat{\beta}^{\text{ridge}} &= (I_K + \lambda I_K)^{-1} \hat{\beta}^{\text{ols}} \\ &= \frac{\hat{\beta}^{\text{ols}}}{1+\lambda} \end{aligned}$$

Since we have $\text{var}\left(\hat{\beta}^{\text{ridge}}\right) = \sigma^2 (I_K + Q)^{-1}$ and $Q = \lambda^2 I_K + 2\lambda I_K$, we deduce that:

$$\begin{aligned} \text{var}\left(\hat{\beta}^{\text{ridge}}\right) &= \frac{\sigma^2}{(1 + 2\lambda + \lambda^2)} I_K \\ &= \frac{1}{(1+\lambda)^2} \text{var}\left(\hat{\beta}^{\text{ols}}\right) \end{aligned}$$

Concerning the model degree of freedom, we obtain:

$$\text{df}^{(\text{model})} = \sum_{k=1}^{K} \frac{1}{1+\lambda} = \frac{K}{1+\lambda}$$

2. (a) We have:

$$\begin{aligned} f(\beta) &= \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + \\ &\quad \frac{\lambda}{2} \left(\alpha \sum_{k=1}^{K} |\beta_k| + (1-\alpha) \sum_{k=1}^{K} \beta_k^2 \right) \\ &= \frac{1}{2} \beta^\top \left(\mathbf{X}^\top \mathbf{X} + \lambda (1-\alpha) I_K \right) \beta - \beta^\top \mathbf{X}^\top \mathbf{Y} + \frac{1}{2} \mathbf{Y}^\top \mathbf{Y} + \\ &\quad \frac{\lambda\alpha}{2} \sum_{k=1}^{K} |\beta_k| \end{aligned}$$

We note $A = \begin{pmatrix} I_K & -I_K \end{pmatrix}$. We introduce the parameter vector $\theta = (\beta^+, \beta^-)$ such that $\beta = \beta^+ - \beta^-$, $\beta^+ \geq \mathbf{0}$ and $\beta^- \geq \mathbf{0}$. We notice that:

$$\begin{aligned} \sum_{k=1}^{K} |\beta_k| &= \sum_{k=1}^{K} \left| \beta_k^+ - \beta_k^- \right| \\ &= \sum_{k=1}^{K} \beta_k^+ + \sum_{k=1}^{K} \beta_k^- \\ &= \mathbf{1}^\top \theta \end{aligned}$$

Since we have $\beta = A\theta$, it follows that:

$$\begin{aligned} f(\beta) &= \frac{1}{2} \theta^\top A^\top \left(\mathbf{X}^\top \mathbf{X} + \lambda (1-\alpha) I_K \right) A\theta - \\ &\quad \theta^\top A^\top \mathbf{X}^\top \mathbf{Y} + \frac{1}{2} \mathbf{Y}^\top \mathbf{Y} + \frac{\lambda\alpha}{2} \left(\theta^\top \mathbf{1} \right) \end{aligned}$$

The corresponding QP program is then:

$$\begin{aligned} \hat{\theta} &= \arg\min \frac{1}{2} \theta^\top Q \theta - \theta^\top R \\ &\text{u.c.} \quad \theta \geq \mathbf{0} \end{aligned}$$

where $Q = A^\top \left(\mathbf{X}^\top \mathbf{X} + \lambda (1-\alpha) I_K \right) A$ and $R = A^\top \mathbf{X}^\top \mathbf{Y} + \dfrac{\lambda\alpha}{2} \mathbf{1}$.

(b) Results are given in Figure 15.1.

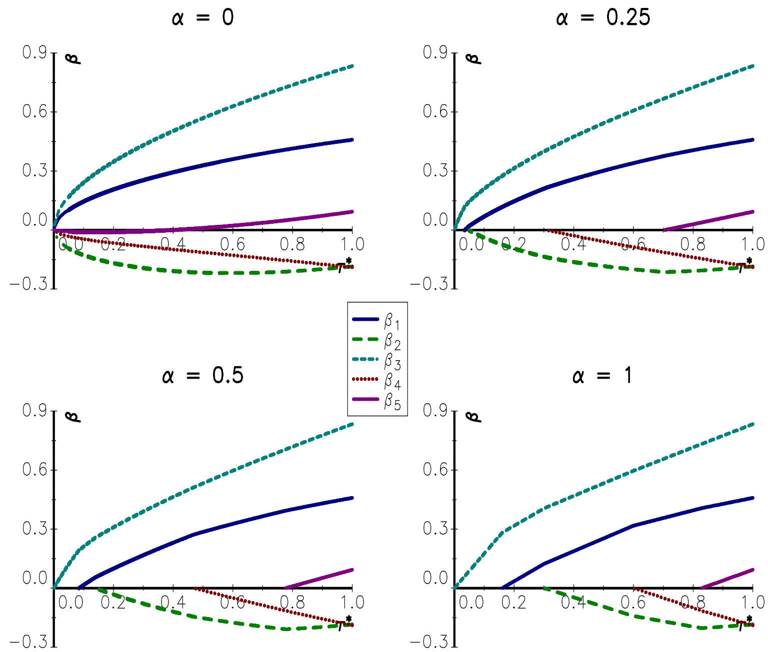

FIGURE 15.1: Comparison of lasso, ridge and elastic net estimates

15.4.2 Cross-validation of the ridge linear regression

1. The objective function is equal to:

$$\begin{aligned}\mathcal{L}(\beta;\lambda) &= \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^\top(\mathbf{Y}-\mathbf{X}\beta)+\frac{\lambda}{2}\beta^\top\beta \\ &= \frac{1}{2}\left(\mathbf{Y}^\top\mathbf{Y}-2\beta^\top\mathbf{X}^\top\mathbf{Y}+\beta^\top\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\beta\right)\end{aligned}$$

The first order condition $\partial_\beta\,\mathcal{L}(\beta;\lambda)=\mathbf{0}$ is equivalent to:

$$-\mathbf{X}^\top\mathbf{Y}+\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)\beta=\mathbf{0}$$

We deduce that:

$$\hat\beta=\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{Y}$$

2. We have:

$$\begin{aligned}\hat\beta_{-i} &= \left(\mathbf{X}_{-i}^\top\mathbf{X}_{-i}+\lambda I_K\right)^{-1}\mathbf{X}_{-i}^\top\mathbf{Y}_{-i} \\ &= \left(\mathbf{X}^\top\mathbf{X}-x_ix_i^\top+\lambda I_K\right)^{-1}\left(\mathbf{X}^\top\mathbf{Y}-x_iy_i\right) \\ &= \left(Q-x_ix_i^\top\right)^{-1}\left(\mathbf{X}^\top\mathbf{Y}-x_iy_i\right)\end{aligned}$$

where $Q = \mathbf{X}^\top \mathbf{X} + \lambda I_K$. The Sherman-Morrison-Woodbury formula[1] leads to:

$$\begin{aligned}
\left(Q - x_i x_i^\top\right)^{-1} &= Q^{-1} + \left(\frac{1}{1 - x_i^\top Q^{-1} x_i}\right) Q^{-1} x_i x_i^\top Q^{-1} \\
&= Q^{-1} + \left(\frac{1}{1 - h_i}\right) Q^{-1} x_i x_i^\top Q^{-1}
\end{aligned}$$

where $h_i = x_i^\top Q^{-1} x_i = x_i^\top \left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} x_i$. We can now obtain a formula that relates the ridge estimators $\hat{\beta}_{-i}$ and $\hat{\beta}$. Indeed, we have:

$$\begin{aligned}
\hat{\beta}_{-i} &= Q^{-1} \mathbf{X}^\top \mathbf{Y} - Q^{-1} x_i y_i + \\
&\quad \frac{Q^{-1} x_i x_i^\top Q^{-1}}{1 - h_i} \mathbf{X}^\top \mathbf{Y} - \frac{Q^{-1} x_i x_i^\top Q^{-1}}{1 - h_i} x_i y_i \\
&= \hat{\beta} - Q^{-1} x_i \left(y_i - \frac{x_i^\top}{1 - h_i} \hat{\beta} + \frac{h_i}{1 - h_i} y_i\right) \\
&= \hat{\beta} - \frac{Q^{-1} x_i}{1 - h_i} \left((1 - h_i) y_i - x_i^\top \hat{\beta} + h_i y_i\right) \\
&= \hat{\beta} - \frac{\left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} x_i \hat{u}_i}{1 - h_i}
\end{aligned}$$

where $\hat{u}_i = y_i - x_i^\top \hat{\beta}$.

3. We notice that:

$$\begin{aligned}
\hat{u}_{i,-i} &= y_i - \hat{y}_{i,-i} \\
&= y_i - x_i^\top \hat{\beta}_{-i} \\
&= y_i - x_i^\top \left(\hat{\beta} - \frac{\left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} x_i \hat{u}_i}{1 - h_i}\right) \\
&= y_i - x_i^\top \hat{\beta} + \frac{x_i^\top \left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} x_i \hat{u}_i}{1 - h_i} \\
&= \hat{u}_i + \frac{h_i \hat{u}_i}{1 - h_i} \\
&= \frac{\hat{u}_i}{1 - h_i}
\end{aligned}$$

4. We deduce that the PRESS statistic is equal to:

$$\begin{aligned}
\mathcal{P}\text{ress} &= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-i}\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{u}_i^2}{\left(1 - h_i\right)^2}
\end{aligned}$$

[1] Suppose u and v are two vectors and A is an invertible square matrix. It follows that:

$$\left(A + uv^\top\right)^{-1} = A^{-1} - \frac{1}{1 + v^\top A^{-1} u} A^{-1} u v^\top A^{-1}$$

5. Let $\mathbf{H}(\lambda) = \mathbf{X}\left(\mathbf{X}^\top \mathbf{X} + \lambda I_n\right)^{-1} \mathbf{X}^\top$ be the hat matrix[2]. By considering the singular value decomposition $\mathbf{X} = USV^\top$, we obtain:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= VSU^\top USV^\top \\ &= VS^2V^\top \end{aligned}$$

and:

$$\mathbf{X}^\top \mathbf{X} + \lambda I_n = V\left(S^2 + \lambda I_K\right) V^\top$$

It follows that:

$$\begin{aligned} \mathrm{df}^{(\mathrm{model})}(\lambda) &= \operatorname{trace} \mathbf{H}(\lambda) \\ &= \operatorname{trace}\left(\mathbf{X}\left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} \mathbf{X}^\top\right) \\ &= \operatorname{trace}\left(\left(\mathbf{X}^\top \mathbf{X} + \lambda I_K\right)^{-1} \mathbf{X}^\top \mathbf{X}\right) \end{aligned}$$

Since we have:

$$\begin{aligned} \left(\mathbf{X}^\top \mathbf{X} + \lambda I_n\right)^{-1} \mathbf{X}^\top \mathbf{X} &= \left(V\left(S^2 + \lambda I_K\right) V^\top\right)^{-1} VS^2V^\top \\ &= \left(V^\top\right)^{-1}\left(S^2 + \lambda I_K\right)^{-1} V^{-1}VS^2V^\top \\ &= V\left(S^2 + \lambda I_K\right)^{-1} S^2V^\top \end{aligned}$$

we finally obtain:

$$\begin{aligned} \mathrm{df}^{(\mathrm{model})}(\lambda) &= \operatorname{trace}\left(\left(V^\top\right)^{-1}\left(S^2 + \lambda I_K\right)^{-1} S^2V^\top\right) \\ &= \operatorname{trace}\left(\left(S^2 + \lambda I_K\right)^{-1} S^2V^\top\left(V^\top\right)^{-1}\right) \\ &= \operatorname{trace}\left(\left(S^2 + \lambda I_K\right)^{-1} S^2\right) \\ &= \sum_{k=1}^{K} \frac{s_k^2}{s_k^2 + \lambda} \end{aligned}$$

We verify the properties $\mathrm{df}^{(\mathrm{model})}(0) = K$ and $\mathrm{df}^{(\mathrm{model})}(\infty) = 0$.

6. Since we have $\hat{\mathbf{Y}} = \mathbf{H}(\lambda)\mathbf{Y}$ and $\hat{\mathbf{U}} = (I_n - \mathbf{H}(\lambda))\mathbf{Y}$, we can express the PRESS statistic as:

$$\mathcal{P}\mathrm{ress} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{((I_n - \mathbf{H}(\lambda))\mathbf{Y})_i}{(I_n - \mathbf{H}(\lambda))_{i,i}}\right)^2$$

whereas the generalized cross-validation statistic is defined by:

$$GCV = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{((I_n - \mathbf{H}(\lambda))\mathbf{Y})_i}{1 - \bar{h}}\right)^2$$

where $\bar{h} = n^{-1}\sum_{i=1}^{n}\mathbf{H}(\lambda)_{i,i}$. We verify that the generalized cross-validation statistic corresponds to the PRESS statistic where the elements $\mathbf{H}(\lambda)_{i,i}$ are replaced by their

[2] $\mathbf{H}$ transforms $\mathbf{Y}$ into $\hat{\mathbf{Y}}$ (pronounced "y-hat").

mean $\bar{h}$. We have:

$$\begin{aligned}
1-\bar{h} &= \frac{1}{n}\sum_{i=1}^{n}\left(I_n-\mathbf{H}(\lambda)\right)_{i,i} \\
&= \frac{1}{n}\operatorname{trace}\left(I_n-\mathbf{H}(\lambda)\right) \\
&= \frac{1}{n}\sum_{k=1}^{K}\left(\frac{n}{K}-\frac{s_k^2}{s_k^2+\lambda}\right) \\
&= \frac{1}{n}\sum_{k=1}^{K}\frac{n\left(s_k^2+\lambda\right)-Ks_k^2}{K\left(s_k^2+\lambda\right)} \\
&= \frac{1}{nK}\sum_{k=1}^{K}\frac{(n-K)\,s_k^2+n\lambda}{s_k^2+\lambda}
\end{aligned}$$

The GCV statistic is then equal to:

$$\begin{aligned}
GCV &= \frac{1}{n}\frac{\sum_{i=1}^{n}\left(y_i-x_i^\top\hat{\beta}\right)^2}{\left(\frac{1}{nK}\sum_{k=1}^{K}\frac{(n-K)\,s_k^2+n\lambda}{s_k^2+\lambda}\right)^2} \\
&= nK^2\left(\sum_{k=1}^{K}\frac{(n-K)\,s_k^2+n\lambda}{s_k^2+\lambda}\right)^{-2}\operatorname{RSS}\left(\hat{\beta}(\lambda)\right) \qquad (15.2)
\end{aligned}$$

The effect of λ on the GCV statistic is not obvious since we have:

$$\lambda\nearrow\Rightarrow\ \operatorname{RSS}\left(\hat{\beta}(\lambda)\right)\nearrow$$

and:

$$\lambda\nearrow\Rightarrow\ \sum_{k=1}^{K}\frac{(n-K)\,s_k^2+n\lambda}{s_k^2+\lambda}\nearrow$$

7. Since we have $1-\bar{h}=n^{-1}\operatorname{trace}\left(I_n-\mathbf{H}(\lambda)\right)$, the GCV statistic is equal to:

$$\begin{aligned}
GCV &= \frac{1}{n}\left(\frac{1}{n^{-1}\operatorname{trace}\left(I_n-\mathbf{H}(\lambda)\right)}\right)^2\operatorname{RSS}\left(\hat{\beta}(\lambda)\right) \\
&= n\left(\operatorname{trace}\left(I_n-\mathbf{H}(\lambda)\right)\right)^{-2}\operatorname{RSS}\left(\hat{\beta}(\lambda)\right)
\end{aligned}$$

From the Woodbury formula, we have[3]:

$$\begin{aligned}
I_n-\mathbf{H}(\lambda) &= I_n-\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}+\lambda I_K\right)^{-1}\mathbf{X}^\top \\
&= \left(I_n+\frac{\mathbf{X}\mathbf{X}^\top}{\lambda}\right)^{-1}
\end{aligned}$$

[3]The Woodbury matrix identity is:

$$(A+BCD)^{-1}=A^{-1}-A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$$

Let λ_i be the eigenvalues of the $(n \times n)$ symmetric real matrix $\mathbf{X}\mathbf{X}^\top$. We have:

$$\begin{aligned}
\operatorname{trace}\left(I_n - \mathbf{H}(\lambda)\right) &= \sum_{i=1}^{n} \left(1 + \frac{\lambda_i}{\lambda}\right)^{-1} \\
&= \sum_{i=1}^{n} \frac{\lambda}{\lambda + \lambda_i} \\
&= \sum_{i=1}^{K} \frac{\lambda}{\lambda + \lambda_i} + \sum_{i=K+1}^{n} \frac{\lambda}{\lambda + \lambda_i} \\
&= \sum_{i=1}^{K} \frac{\lambda}{\lambda + \lambda_i} + (n - K)
\end{aligned}$$

because the last $n - K$ eigenvalues λ_i are equal to 0. Moreover, we have $\lambda_k = s_k^2$. Finally, we obtain[4]:

$$GCV = n \left(n - K + \sum_{k=1}^{K} \frac{\lambda}{s_k^2 + \lambda}\right)^{-2} \operatorname{RSS}\left(\hat{\beta}(\lambda)\right) \tag{15.3}$$

8. The values of $\hat{\beta}_{-i}$ when λ is equal to 3 are reported in Table 15.1. In the last row, we have also given the ridge estimate $\hat{\beta}$ calculated with the full sample. Using the values of $\hat{y}_{i,-i}$, $\hat{u}_{i,-i}$, $\hat{u}_i$ and h_i (see Table 15.2), we obtain[5] $\mathcal{P}\text{ress} = 0.29104$ and $GCV = 0.28059$.

15.4.3 K-means and the Lloyd's algorithm

1. We have:

$$\begin{aligned}
\left\| x_i - x_j \right\|^2 &= \sum_{k=1}^{K} \left(x_{i,k} - x_{j,k}\right)^2 \\
&= \sum_{k=1}^{K} x_{i,k}^2 + \sum_{k=1}^{K} x_{j,k}^2 - 2 \sum_{k=1}^{K} x_{i,k} x_{j,k}
\end{aligned}$$

and:

$$\left\| x_i - \bar{x} \right\|^2 = \sum_{k=1}^{K} x_{i,k}^2 + \sum_{k=1}^{K} \bar{x}_{(k)}^2 - 2 \sum_{k=1}^{K} x_{i,k} \bar{x}_{(k)}$$

---

[4]We verify that Equations (15.2) and (15.3) are equivalent, because we have:

$$\begin{aligned}
\frac{1}{K} \sum_{k=1}^{K} \frac{(n-K) s_k^2 + n\lambda}{s_k^2 + \lambda} &= \frac{1}{K} \sum_{k=1}^{K} \frac{(n-K)\left(s_k^2 + \lambda\right) + K\lambda}{s_k^2 + \lambda} \\
&= (n-K) + \sum_{k=1}^{K} \frac{\lambda}{s_k^2 + \lambda}
\end{aligned}$$

[5]We have $\bar{h} = 0.24951$, $s_1 = 58.71914$, $s_2 = 51.42980$, $s_3 = 45.83216$, $s_4 = 37.91501$ and $s_5 = 26.61791$.

TABLE 15.1: LOOCV ridge estimates $\hat{\beta}_{-i}$

i	$\hat{\beta}_{1,-i}$	$\hat{\beta}_{2,-i}$	$\hat{\beta}_{3,-i}$	$\hat{\beta}_{4,-i}$	$\hat{\beta}_{5,-i}$
1	1.2274	−0.9805	0.1298	−0.4923	0.0398
2	1.2307	−0.9865	0.1357	−0.4946	0.0415
3	1.2335	−0.9827	0.1362	−0.4925	0.0410
4	1.2303	−0.9876	0.1355	−0.4957	0.0417
5	1.2296	−0.9849	0.1358	−0.4948	0.0420
6	1.2300	−0.9851	0.1361	−0.4941	0.0422
7	1.2335	−0.9870	0.1287	−0.4898	0.0476
8	1.2219	−0.9838	0.1357	−0.5047	0.0463
9	1.2281	−0.9844	0.1382	−0.5005	0.0445
10	1.2319	−0.9889	0.1401	−0.4912	0.0444
11	1.2299	−0.9856	0.1353	−0.4938	0.0430
12	1.2300	−0.9849	0.1355	−0.4950	0.0411
13	1.2280	−0.9817	0.1320	−0.4974	0.0407
14	1.2307	−0.9855	0.1365	−0.4965	0.0427
15	1.2314	−0.9839	0.1360	−0.4937	0.0426
16	1.2285	−0.9861	0.1390	−0.4944	0.0393
17	1.2289	−0.9843	0.1346	−0.4958	0.0390
18	1.2246	−0.9855	0.1370	−0.4892	0.0426
19	1.2267	−0.9878	0.1356	−0.4920	0.0443
20	1.2459	−0.9890	0.1386	−0.4830	0.0358
$\hat{\beta}$	1.2301	−0.9854	0.1358	−0.4941	0.0420

TABLE 15.2: Computation of $\hat{y}_{i,-i}$, $\hat{u}_{i,-i}$, $\hat{u}_i$ and h_i

i	y_i	$\hat{y}_{i,-i}$	$\hat{u}_{i,-i}$	$\hat{u}_i$	h_i
1	−23.0	−22.3270	−0.6730	−0.5130	0.2378
2	−21.0	−21.2041	0.2041	0.1796	0.1201
3	−5.0	−5.4804	0.4804	0.3950	0.1778
4	−39.6	−39.7745	0.1745	0.0857	0.5091
5	5.8	5.7076	0.0924	0.0828	0.1040
6	13.6	13.5376	0.0624	0.0525	0.1582
7	14.0	14.7404	−0.7404	−0.4168	0.4371
8	−5.2	−4.3994	−0.8006	−0.5534	0.3087
9	6.9	7.5607	−0.6607	−0.5306	0.1970
10	−5.2	−5.6244	0.4244	0.3106	0.2681
11	0.0	−0.0913	0.0913	0.0595	0.3483
12	3.0	3.2119	−0.2119	−0.1974	0.0682
13	9.2	8.9014	0.2986	0.1664	0.4428
14	26.1	26.3478	−0.2478	−0.1842	0.2568
15	−6.3	−6.4835	0.1835	0.1192	0.3506
16	11.5	10.9309	0.5691	0.4763	0.1631
17	4.8	4.3120	0.4880	0.4360	0.1065
18	35.2	34.4379	0.7621	0.5531	0.2742
19	14.0	13.2528	0.7472	0.6633	0.1123
20	−21.4	−22.5438	1.1438	0.7438	0.3497

where $\bar{x}_{(k)} = n^{-1}\sum_{i=1}^{n} x_{i,k}$. We note $S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|x_i - x_j\right\|^2$ and $S_2 = n\sum_{i=1}^{n}\left\|x_i - \bar{x}\right\|^2$. We deduce that:

$$\begin{aligned}
S_1 &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{K} x_{i,k}^2 + \sum_{k=1}^{K} x_{j,k}^2 - 2\sum_{k=1}^{K} x_{i,k}x_{j,k}\right) \\
&= \sum_{k=1}^{K}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_{i,k}^2 + x_{j,k}^2 - 2x_{i,k}x_{j,k}\right) \\
&= \sum_{k=1}^{K}\frac{1}{2}\left(n\sum_{i=1}^{n} x_{i,k}^2 + n\sum_{j=1}^{n} x_{j,k}^2 - 2\sum_{i=1}^{n} x_{i,k}\left(\sum_{j=1}^{n} x_{j,k}\right)\right) \\
&= \sum_{k=1}^{K}\left(n\sum_{i=1}^{n} x_{i,k}^2 - \sum_{i=1}^{n} x_{i,k} n\bar{x}_{(k)}\right) \\
&= \sum_{k=1}^{K}\left(n\sum_{i=1}^{n} x_{i,k}^2 - n^2\bar{x}_{(k)}^2\right)
\end{aligned}$$

and:

$$\begin{aligned}
S_2 &= n\sum_{i=1}^{n}\sum_{k=1}^{K} x_{i,k}^2 + n\sum_{i=1}^{n}\sum_{k=1}^{K}\bar{x}_{(k)}^2 - 2n\sum_{i=1}^{n}\sum_{k=1}^{K} x_{i,k}\bar{x}_{(k)} \\
&= \sum_{k=1}^{K}\left(n\sum_{i=1}^{n} x_{i,k}^2 + n\sum_{i=1}^{n}\bar{x}_{(k)}^2 - 2n\bar{x}_{(k)}\sum_{i=1}^{n} x_{i,k}\right) \\
&= \sum_{k=1}^{K}\left(n\sum_{i=1}^{n} x_{i,k}^2 + n^2\bar{x}_{(k)}^2 - 2n^2\bar{x}_{(k)}^2\right) \\
&= \sum_{k=1}^{K}\left(n\sum_{i=1}^{n} x_{i,k}^2 - n^2\bar{x}_{(k)}^2\right)
\end{aligned}$$

We conclude that $S_1 = S_2$.

2. Using the previous result, we have:

$$\frac{1}{2}\sum_{\mathcal{C}(i)=j}\sum_{\mathcal{C}(i')=j}\left\|x_i - x_{i'}\right\|^2 = n_j\sum_{\mathcal{C}(i)=j}\left\|x_i - \bar{x}_j\right\|^2$$

where $\bar{x}_j$ and n_j is the mean vector and the number of observations of Cluster $\mathcal{C}_j$.

3. The first-order conditions are:

$$\frac{\partial\, f\left(\mu_1,\ldots,\mu_{n_C}\right)}{\partial\,\mu_j} = \mathbf{0} \qquad \text{for } j = 1,\ldots,n_C$$

where:

$$f\left(\mu_1,\ldots,\mu_{n_C}\right) = \sum_{j=1}^{n_C} n_j\sum_{\mathcal{C}(i)=j}\left\|x_i - \mu_j\right\|^2$$

Since we have:

$$\frac{\partial}{\partial\,\mu_j}\left\|x_i - \mu_j\right\|^2 = -2\left(x_i - \mu_j\right)$$

it follows that:

$$\frac{\partial f(\mu_1, \ldots, \mu_{n_C})}{\partial \mu_j} = -2n_j \sum_{\mathcal{C}(i)=j} (x_i - \mu_j) = \mathbf{0}$$

We deduce that $\mu_j^\star = \left(\mu_{j,1}^\star, \ldots, \mu_{j,K}^\star\right)$ where:

$$\mu_{j,k}^\star = \frac{1}{n_j} \sum_{\mathcal{C}(i)=j} x_{i,k}$$

Finally, we verify that the optimal solution is $\mu_j^\star = \bar{x}_j$. The Llyod's algorithm exploits this result in order to find the optimal partition.

4. In Figure 15.2, we have reported the classification operated by K-means, LDA and QDA. We notice that the unsupervised K-means algorithm gives the same result as the supervised LDA algorithm.

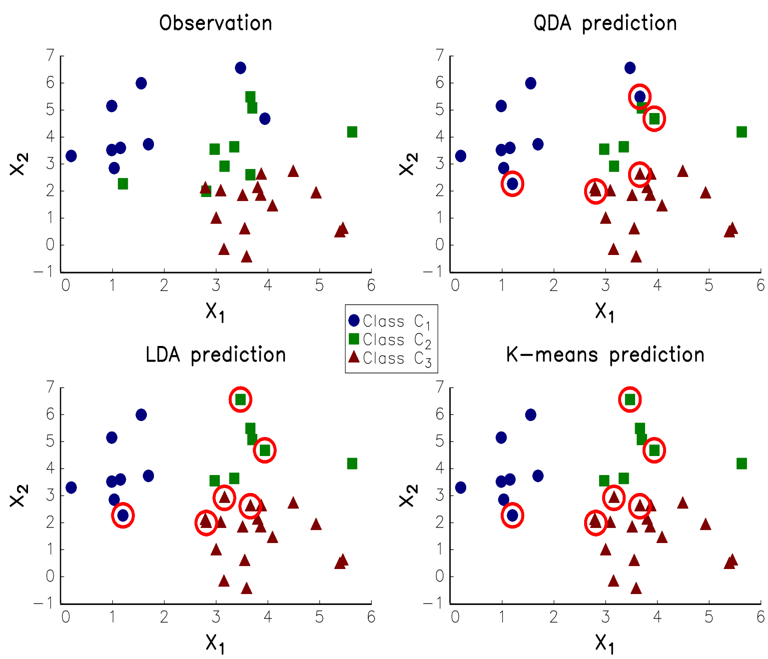

FIGURE 15.2: Comparison of LDA, QDA and K-means classification

15.4.4 Derivation of the principal component analysis[6]

1. We have:

$$\begin{aligned} \text{var}(Z_1) &= \text{var}\left(\beta_1^\top X\right) \\ &= \beta_1^\top \Sigma \beta_1 \end{aligned}$$

[6]The following exercise is taken from Chapters 1 and 2 of Jolliffe (2002).

The objective function is to maximize the variance of Z_1 under a normalization constraint[7]:

$$\begin{aligned} \beta_1 &= \arg\max \beta^\top \Sigma \beta \\ \text{s.t.} &\quad \beta^\top \beta = 1 \end{aligned}$$

The Lagrange function is then equal to:

$$\mathcal{L}(\beta; \lambda_1) = \beta^\top \Sigma \beta - \lambda_1 \left(\beta^\top \beta - 1\right)$$

Since, the first derivative $\partial_\beta \mathcal{L}(\beta; \lambda_1)$ is equal to $2\Sigma\beta - 2\lambda_1\beta$, we deduce that the first-order condition is:

$$\Sigma \beta_1 = \lambda_1 \beta_1 \tag{15.4}$$

or:

$$(\Sigma - \lambda_1 I_K)\, \beta_1 = \mathbf{0}$$

It follows that β_1 is an eigenvector of Σ and λ_1 is the associated eigenvalue. Moreover, we have:

$$\begin{aligned} \operatorname{var}(Z_1) &= \beta_1^\top \Sigma \beta_1 \\ &= \lambda_1 \end{aligned}$$

Maximizing $\operatorname{var}(Z_1)$ is then equivalent to consider the eigenvector β_1 that corresponds to the largest eigenvalue.

2. We have:

$$\begin{aligned} \operatorname{var}(Z_2) &= \operatorname{var}\left(\beta_2^\top X\right) \\ &= \beta_2^\top \Sigma \beta_2 \end{aligned}$$

Using Equation (15.4), we deduce that the covariance is:

$$\begin{aligned} \operatorname{cov}(Z_1, Z_2) &= \beta_1^\top \Sigma \beta_2 \\ &= \lambda_1 \beta_1^\top \beta_2 \end{aligned} \tag{15.5}$$

The objective function is then to maximize the variance of Z_2 under the constraints of normalization and independence between Z_1 and Z_2:

$$\begin{aligned} \beta_2 &= \arg\max \beta^\top \Sigma \beta \\ \text{s.t.} &\quad \left\{ \begin{array}{l} \beta^\top \beta = 1 \\ \beta_1^\top \beta = 0 \end{array} \right. \end{aligned}$$

The Lagrange function has the following expression:

$$\mathcal{L}(\beta; \lambda_2, \varphi) = \beta^\top \Sigma \beta - \lambda_2 \left(\beta^\top \beta - 1\right) - \varphi \beta_1^\top \beta$$

We deduce that the first-order condition is:

$$2\Sigma\beta_2 - 2\lambda_2\beta_2 - \varphi\beta_1 = \mathbf{0}$$

It follows that:

$$\beta_1^\top (2\Sigma\beta_2 - 2\lambda_2\beta_2 - \varphi\beta_1) = \beta_1^\top \mathbf{0} = 0$$

[7] Joliffe (2002) notices that the solution is $\beta_1 = \infty$ without this normalization.

or:

$$2\beta_1^\top \Sigma \beta_2 - 2\lambda_2 \beta_1^\top \beta_2 - \varphi \beta_1^\top \beta_1 = 0$$

Since the 1st and 2nd PCs are uncorrelated, Equation (15.5) implies that $\beta_1^\top \Sigma \beta_2 = 0$ and $\beta_1^\top \beta_2 = 0$. We deduce that $-\varphi \beta_1^\top \beta_1 = 0$ or $\varphi = 0$ because $\beta_1^\top \beta_1 = 1$. In this case, the first-order condition becomes:

$$\Sigma \beta_2 = \lambda_2 \beta_2$$

Again, β_2 is an eigenvector of Σ and λ_2 is the associated eigenvalue. Maximizing the variance of the second PC is then equivalent to consider the eigenvector β_2 that corresponds to the second largest eigenvalue[8].

15.4.5 Two-class separation maximization

1. The total scatter matrix is equal to:

$$\begin{aligned}
\mathbf{S} &= \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^\top \\
&= \sum_{i=1}^{n} x_i x_i^\top - 2\hat{\mu} \sum_{i=1}^{n} x_i^\top + n\hat{\mu}\hat{\mu}^\top \\
&= \sum_{i=1}^{n} x_i x_i^\top - n\hat{\mu}\hat{\mu}^\top
\end{aligned}$$

For the within-class scatter matrix, we obtain:

$$\begin{aligned}
\mathbf{S}_W &= \sum_{j=1}^{J} \mathbf{S}_j \\
&= \sum_{j=1}^{J} \sum_{i \in \mathcal{C}_j} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^\top \\
&= \sum_{j=1}^{J} \left(\sum_{i \in \mathcal{C}_j} x_i x_i^\top - n_j \hat{\mu}_j \hat{\mu}_j^\top \right) \\
&= \sum_{j=1}^{J} \sum_{i \in \mathcal{C}_j} x_i x_i^\top - \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top \\
&= \sum_{i=1}^{n} x_i x_i^\top - \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top
\end{aligned}$$

[8] We cannot use the first eigenvector because $\beta_1^\top \beta_2$ must be equal to zero.

Concerning the between-class scatter matrix, we deduce that:

$$\begin{aligned}
\mathbf{S}_B &= \sum_{j=1}^{J} n_j \left(\hat{\mu}_j - \hat{\mu}\right)\left(\hat{\mu}_j - \hat{\mu}\right)^\top \\
&= \sum_{j=1}^{J} \left(n_j \hat{\mu}_j \hat{\mu}_j^\top - 2 n_j \hat{\mu} \hat{\mu}_j^\top + n_j \hat{\mu}\hat{\mu}^\top\right) \\
&= \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top - 2\hat{\mu} \sum_{j=1}^{J} n_j \hat{\mu}_j^\top + n\hat{\mu}\hat{\mu}^\top \\
&= \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top - 2\hat{\mu}\left(n\hat{\mu}^\top\right) + n\hat{\mu}\hat{\mu}^\top \\
&= \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top - n\hat{\mu}\hat{\mu}^\top
\end{aligned}$$

because $n\hat{\mu} = \sum_{j=1}^{J} n_j \hat{\mu}_j$. It follows that:

$$\begin{aligned}
\mathbf{S}_W + \mathbf{S}_B &= \left(\sum_{i=1}^{n} x_i x_i^\top - \sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top\right) + \left(\sum_{j=1}^{J} n_j \hat{\mu}_j \hat{\mu}_j^\top - n\hat{\mu}\hat{\mu}^\top\right) \\
&= \sum_{i=1}^{n} x_i x_i^\top - n\hat{\mu}\hat{\mu}^\top \\
&= \mathbf{S}
\end{aligned}$$

2. We have:

$$\hat{\mu} = \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_2}{n_1 + n_2}$$

It follows that:

$$\begin{aligned}
\hat{\mu}_1 - \hat{\mu} &= \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_1}{n_1 + n_2} - \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_2}{n_1 + n_2} \\
&= \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_1 - n_1 \hat{\mu}_1 - n_2 \hat{\mu}_2}{n_1 + n_2} \\
&= \frac{n_2}{n_1 + n_2}\left(\hat{\mu}_1 - \hat{\mu}_2\right)
\end{aligned}$$

and:

$$\hat{\mu}_2 - \hat{\mu} = \frac{n_1}{n_1 + n_2}\left(\hat{\mu}_2 - \hat{\mu}_1\right)$$

We deduce that:

$$\begin{aligned}
\mathbf{S}_B &= n_1 \left(\hat{\mu}_1 - \hat{\mu}\right)\left(\hat{\mu}_1 - \hat{\mu}\right)^\top + n_2 \left(\hat{\mu}_2 - \hat{\mu}\right)\left(\hat{\mu}_2 - \hat{\mu}\right)^\top \\
&= n_1 \left(\frac{n_2}{n_1 + n_2}\right)^2 \left(\hat{\mu}_1 - \hat{\mu}_2\right)\left(\hat{\mu}_1 - \hat{\mu}_2\right)^\top + \\
&\quad n_2 \left(\frac{n_1}{n_1 + n_2}\right)^2 \left(\hat{\mu}_2 - \hat{\mu}_1\right)\left(\hat{\mu}_2 - \hat{\mu}_1\right)^\top \\
&= \frac{n_1 n_2}{n_1 + n_2}\left(\hat{\mu}_1 - \hat{\mu}_2\right)\left(\hat{\mu}_1 - \hat{\mu}_2\right)^\top
\end{aligned}$$

and[9]:

$$\begin{aligned}
\beta^\top \mathbf{S}_B \beta &= \frac{n_1 n_2}{n_1 + n_2} \beta^\top \left(\hat{\mu}_1 - \hat{\mu}_2\right) \left(\hat{\mu}_1 - \hat{\mu}_2\right)^\top \beta \\
&= \frac{n_1 n_2}{n_1 + n_2} \left(\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2\right) \left(\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2\right)^\top \\
&= \frac{n_1 n_2}{n_1 + n_2} \left(\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2\right)^2 \\
&= \frac{n_1 n_2}{n_1 + n_2} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2
\end{aligned}$$

where $\tilde{\mu}_j$ is defined as:

$$\tilde{\mu}_j = \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} y_i = \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} \beta^\top x_i = \beta^\top \left(\frac{1}{n_j} \sum_{i \in \mathcal{C}_j} x_i \right) = \beta^\top \hat{\mu}_j$$

3. We have:

$$\mathbf{S}_j = \sum_{i \in \mathcal{C}_j} \left(x_i - \hat{\mu}_j\right) \left(x_i - \hat{\mu}_j\right)^\top$$

and:

$$\begin{aligned}
\beta^\top \mathbf{S}_j \beta &= \beta^\top \sum_{i \in \mathcal{C}_j} \left(x_i - \hat{\mu}_j\right) \left(x_i - \hat{\mu}_j\right)^\top \beta \\
&= \sum_{i \in \mathcal{C}_j} \left(\beta^\top x_i - \beta^\top \hat{\mu}_j\right)^2 \\
&= \sum_{i \in \mathcal{C}_j} \left(y_i - \tilde{\mu}_j\right)^2 \\
&= \tilde{\mathbf{S}}_j \\
&= \tilde{s}_j^2
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\beta^\top \mathbf{S}_W \beta &= \beta^\top \left(\mathbf{S}_1 + \mathbf{S}_2\right) \beta \\
&= \beta^\top \mathbf{S}_1 \beta + \beta^\top \mathbf{S}_2 \beta \\
&= \tilde{s}_1^2 + \tilde{s}_2^2
\end{aligned}$$

4. We have:

$$\begin{aligned}
J(\beta) &= \frac{\beta^\top \mathbf{S}_B \beta}{\beta^\top \mathbf{S}_W \beta} \\
&= \frac{n_1 n_2}{n_1 + n_2} \frac{\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}
\end{aligned}$$

We finally obtain the following optimization program:

$$\beta^\star = \arg\max \frac{\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

In order to separate the class $\mathcal{C}_1$ and $\mathcal{C}_2$, we would like that $\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2$ is the largest, meaning that the projected means must be as far away as possible. At the same time, we would like that the scatters $\tilde{s}_1^2$ and $\tilde{s}_2^2$ are the smallest, meaning that the samples of each class are close to the corresponding projected mean.

[9]because $\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2$ is a scalar.

5. At the optimum, we know that:

$$\mathbf{S}_B \beta = \lambda \mathbf{S}_W \beta$$

We have:

$$\begin{aligned}
\mathbf{S}_B \beta &= \frac{n_1 n_2}{n_1 + n_2} (\hat{\mu}_1 - \hat{\mu}_2) (\hat{\mu}_1 - \hat{\mu}_2)^\top \beta \\
&= \frac{n_1 n_2}{n_1 + n_2} (\hat{\mu}_1 - \hat{\mu}_2) \left(\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2\right) \\
&= \gamma (\hat{\mu}_1 - \hat{\mu}_2)
\end{aligned}$$

where:

$$\gamma = \frac{n_1 n_2}{n_1 + n_2} \left(\beta^\top \hat{\mu}_1 - \beta^\top \hat{\mu}_2\right)$$

We deduce that:

$$\lambda \mathbf{S}_W \beta = \gamma (\hat{\mu}_1 - \hat{\mu}_2)$$

or:

$$\beta = \frac{\gamma}{\lambda} \mathbf{S}_W^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$$

It follows that the decision boundary is linear and depends on the direction $\hat{\mu}_1 - \hat{\mu}_2$. Since we have $J(\beta') = J(\beta)$ if $\beta' = c\beta$, we can choose the following optimal value:

$$\beta^\star = \mathbf{S}_W^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$$

6. We have:

$$\begin{aligned}
\mathbf{S}_W &= \mathbf{S}_1 + \mathbf{S}_2 \\
&= \begin{pmatrix} 16.86 & 8.00 \\ 8.00 & 18.00 \end{pmatrix} + \begin{pmatrix} 21.33 & 18.33 \\ 18.33 & 18.83 \end{pmatrix} \\
&= \begin{pmatrix} 38.19 & 26.33 \\ 26.33 & 36.83 \end{pmatrix}
\end{aligned}$$

and:

$$\mathbf{S}_B = \begin{pmatrix} 0.89 & -3.10 \\ -3.10 & 10.86 \end{pmatrix}$$

The equation $\mathbf{S}_B \beta = \lambda \mathbf{S}_W \beta$ is equivalent to $\mathbf{S}_W^{-1} \mathbf{S}_B \beta = \lambda \beta$. The largest eigenvalue of the matrix $\mathbf{S}_W^{-1} \mathbf{S}_B$ is equal to 0.856, and the associated eigenvector is $\beta^\star = (0.755, -0.936)$. We deduce that the scores of the 13 observations are -1.117, -2.990, -3.352, -0.544, 1.147, -1.843, -0.907, 0.755, 0.573, 2.083, 0.392, 0.784, and -0.152. We have $\bar{\mu}_1 = -1.372$ and $\bar{\mu}_1 = 0.739$. It follows that $\bar{\mu} = -0.317$. The assignment decision is then:

$$\begin{cases} s_i < -0.317 \Rightarrow i \in \mathcal{C}_1 \\ s_i > -0.317 \Rightarrow i \in \mathcal{C}_2 \end{cases}$$

We observe that the 5^{th} observation is incorrectly assigned to Class $\mathcal{C}_2$, because its score 1.147 is larger than -0.317.

15.4.6 Maximum likelihood estimation of the probit model

1. The probit model is defined by:

$$p = \Pr\{Y = 1 \mid X = x\} = \Phi\left(x^\top \beta\right)$$

We deduce that the log-likelihood function is equal to:

$$\begin{aligned}
\boldsymbol{\ell}(\beta) &= \sum_{i=1}^{n} \ln \Pr\{Y_i = y_i\} \\
&= \sum_{i=1}^{n} \ln\left((1 - p_i)^{1-y_i} p_i^{y_i}\right) \\
&= \sum_{i=1}^{n} (1 - y_i)\ln(1 - p_i) + y_i \ln p_i \\
&= \sum_{i=1}^{n} (1 - y_i)\ln\left(1 - \Phi\left(x_i^\top \beta\right)\right) + y_i \ln \Phi\left(x_i^\top \beta\right)
\end{aligned}$$

2. We have:

$$\begin{aligned}
J_{i,k}(\beta) &= \frac{\partial\, \boldsymbol{\ell}_i(\beta)}{\partial\, \beta_k} \\
&= -(1 - y_i)\frac{\phi\left(x_i^\top \beta\right) x_{i,k}}{1 - \Phi\left(x_i^\top \beta\right)} + y_i \frac{\phi\left(x_i^\top \beta\right) x_{i,k}}{\Phi\left(x_i^\top \beta\right)}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
J_{i,k}(\beta) &= \frac{-(1 - y_i)\Phi\left(x_i^\top \beta\right) + y_i\left(1 - \Phi\left(x_i^\top \beta\right)\right)}{\Phi\left(x_i^\top \beta\right)\left(1 - \Phi\left(x_i^\top \beta\right)\right)} \phi\left(x_i^\top \beta\right) x_{i,k} \\
&= \frac{\left(y_i - \Phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right)}{\Phi\left(x_i^\top \beta\right)\left(1 - \Phi\left(x_i^\top \beta\right)\right)} x_{i,k}
\end{aligned}$$

It follows that the score vector is equal to:

$$\mathcal{S}(\beta) = \sum_{i=1}^{n} \frac{\left(y_i - \Phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right)}{\Phi\left(x_i^\top \beta\right)\left(1 - \Phi\left(x_i^\top \beta\right)\right)} x_i$$

3. The (k, j) element of the Hessian matrix is:

$$H_{k,j}(\beta) = \sum_{i=1}^{n} \frac{\partial^2\, \boldsymbol{\ell}_i(\beta)}{\partial\, \beta_k \partial\, \beta_j}$$

We have $\phi'(z) = -z\phi(z)$ and:

$$\begin{aligned}
(*) &= \frac{\partial}{\partial\, \beta_j} \Phi\left(x_i^\top \beta\right)\left(1 - \Phi\left(x_i^\top \beta\right)\right) \\
&= \phi\left(x_i^\top \beta\right) x_{i,j}\left(1 - \Phi\left(x_i^\top \beta\right)\right) - \Phi\left(x_i^\top \beta\right)\phi\left(x_i^\top \beta\right) x_{i,j} \\
&= \left(1 - 2\Phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right) x_{i,j} \\
(*) &= \frac{\partial}{\partial\, \beta_j}\left(y_i - \Phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right) \\
&= -\left(y_i - \Phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right)\left(x_i^\top \beta\right) x_{i,j} - \phi^2\left(x_i^\top \beta\right) x_{i,j} \\
&= \left(-y_i x_i^\top \beta + \Phi\left(x_i^\top \beta\right) x_i^\top \beta - \phi\left(x_i^\top \beta\right)\right)\phi\left(x_i^\top \beta\right) x_{i,j}
\end{aligned}$$

We note:

$$A_i = \frac{\Phi\left(x_i^\top \beta\right)^2 \left(1 - \Phi\left(x_i^\top \beta\right)\right)^2}{\phi\left(x_i^\top \beta\right) x_{i,j} x_{i,k}} \frac{\partial^2 \boldsymbol{\ell}_i\left(\beta\right)}{\partial\, \beta_k \partial\, \beta_j}$$

and:

$$B_i = \frac{\Phi\left(x_i^\top \beta\right)^2 \left(1 - \Phi\left(x_i^\top \beta\right)\right)^2}{\phi\left(x_i^\top \beta\right)} c_i$$

We obtain:

$$\begin{aligned} A_i &= \left(-y_i x_i^\top \beta + \Phi\left(x_i^\top \beta\right) x_i^\top \beta - \phi\left(x_i^\top \beta\right)\right) \Phi\left(x_i^\top \beta\right) \left(1 - \Phi\left(x_i^\top \beta\right)\right) - \\ & \quad \left(y_i - \Phi\left(x_i^\top \beta\right)\right) \left(1 - 2\Phi\left(x_i^\top \beta\right)\right) \phi\left(x_i^\top \beta\right) \\ &= -y_i \Phi\left(x_i^\top \beta\right) x_i^\top \beta + \Phi^2\left(x_i^\top \beta\right) x_i^\top \beta - \Phi\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) + \\ & \quad y_i \Phi^2\left(x_i^\top \beta\right) x_i^\top \beta - \Phi^3\left(x_i^\top \beta\right) x_i^\top \beta + \Phi^2\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) - \\ & \quad y_i \phi\left(x_i^\top \beta\right) + \Phi\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) + 2 y_i \Phi\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) - \\ & \quad 2\Phi^2\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) \\ &= \Phi^3\left(x_i^\top \beta\right) \left(-x_i^\top \beta\right) + \Phi^2\left(x_i^\top \beta\right) \left(\left(1 + y_i\right) x_i^\top \beta - \phi\left(x_i^\top \beta\right)\right) + \\ & \quad \Phi\left(x_i^\top \beta\right) y_i \left(-x_i^\top \beta + 2\phi\left(x_i^\top \beta\right)\right) - y_i \phi\left(x_i^\top \beta\right) \end{aligned}$$

and:

$$\begin{aligned} B_i &= y_i \left(\phi\left(x_i^\top \beta\right) + x_i^\top \beta \Phi\left(x_i^\top \beta\right)\right) \left(1 - \Phi\left(x_i^\top \beta\right)\right)^2 + \\ & \quad \left(1 - y_i\right) \phi\left(x_i^\top \beta\right) - x_i^\top \beta \left(1 - \Phi\left(x_i^\top \beta\right)\right) \Phi\left(x_i^\top \beta\right)^2 \\ &= y_i \phi\left(x_i^\top \beta\right) + y_i \Phi\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) - 2 y_i \Phi\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) - \\ & \quad 2 y_i \Phi^2\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) + y_i \Phi^2\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) + \\ & \quad y_i \Phi^3\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) + \Phi^2\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) - \\ & \quad y_i \Phi^2\left(x_i^\top \beta\right) \phi\left(x_i^\top \beta\right) - \Phi^2\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) + \\ & \quad \Phi^3\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) + y_i \Phi^2\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) - y_i \Phi^3\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) \\ &= \Phi^3\left(x_i^\top \beta\right) \left(x_i^\top \beta\right) + \Phi^2\left(x_i^\top \beta\right) \left(-\left(1 + y_i\right) x_i^\top \beta + \phi\left(x_i^\top \beta\right)\right) + \\ & \quad \Phi\left(x_i^\top \beta\right) y_i \left(x_i^\top \beta - 2\phi\left(x_i^\top \beta\right)\right) + y_i \phi\left(x_i^\top \beta\right) \end{aligned}$$

Since $B_i = -A_i$, we deduce that:

$$\begin{aligned} H\left(\beta\right) &= \sum_{i=1}^n \frac{\partial^2 \boldsymbol{\ell}_i\left(\beta\right)}{\partial\, \beta \partial\, \beta^\top} \\ &= \sum_{i=1}^n A_i \frac{\phi\left(x_i^\top \beta\right)}{\Phi\left(x_i^\top \beta\right)^2 \left(1 - \Phi\left(x_i^\top \beta\right)\right)^2} \cdot \left(x_i x_i^\top\right) \\ &= -\sum_{i=1}^n B_i \frac{\phi\left(x_i^\top \beta\right)}{\Phi\left(x_i^\top \beta\right)^2 \left(1 - \Phi\left(x_i^\top \beta\right)\right)^2} \cdot \left(x_i x_i^\top\right) \\ &= -\sum_{i=1}^n H_i \cdot \left(x_i x_i^\top\right) \end{aligned}$$

4. The Newton-Raphson algorithm becomes:

$$\beta_{(s+1)} = \beta_{(s)} - H^{-1}\left(\beta_{(s)}\right) \mathcal{S}\left(\beta_{(s)}\right)$$

where $\beta_{(s)}$ is the value of β at the step s. We may initialize the algorithm with the OLS solution $\beta_{(0)} = \left(X^\top X\right)^{-1} X^\top Y$.

15.4.7 Computation of feed-forward neural networks

1. We have $u_{i,h} = \sum_{k=1}^{n_x} \beta_{h,k} x_{i,k}$ or $U = X\beta^\top$ where U is a $n \times n_z$ matrix, X is a $n \times n_x$ matrix and β is a $n_z \times n_x$ matrix. Then, we apply the non-linear transform $z_{i,h} = f_{x,z}(u_{i,h})$ or we perform the element-by-element function $Z = f_{x,z}(U)$ where Z is a $n \times n_z$ matrix. The calculation of $v_{i,j} = \sum_{h=1}^{n_z} \gamma_{j,h} z_{i,h}$ is equivalent to compute $V = Z\gamma^\top$ where V is a $n \times n_y$ matrix and γ is a $n_y \times n_z$ matrix. Finally, we have $y_j(x_i) = f_{z,y}(v_{i,j})$ or $\hat{Y} = f_{z,y}(V)$ where $\hat{Y}$ is a $n \times n_y$ matrix. It follows that:

$$\begin{aligned}
\hat{Y} &= f_{z,y}(V) \\
&= f_{z,y}\left(Z\gamma^\top\right) \\
&= f_{z,y}\left(f_{x,z}(U)\gamma^\top\right) \\
&= f_{z,y}\left(f_{x,z}\left(X\beta^\top\right)\gamma^\top\right)
\end{aligned}$$

2. If $f_{x,z}(z) = f_{z,y}(z) = z$, we deduce that:

$$\hat{Y} = X\beta^\top\gamma^\top = XA$$

where $A = (\gamma\beta)^\top$ is a $n_x \times n_y$ matrix. The least squares loss is then equal to:

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{n_y} \mathcal{L}_{i,j}(\theta) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n_y} \frac{1}{2}\left(y_j(x_i) - y_{i,j}\right)^2 \\
&= \frac{1}{2}\operatorname{trace}\left(\left(\hat{Y} - Y\right)^\intercal\left(\hat{Y} - Y\right)\right) \\
&= \frac{1}{2}\operatorname{trace}\left((XA - Y)^\intercal (XA - Y)\right)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\hat{A} &= \arg\min \frac{1}{2}\operatorname{trace}\left((XA - Y)^\intercal (XA - Y)\right) \\
&= \left(X^\intercal X\right)^{-1} X^\intercal Y
\end{aligned}$$

and:

$$\hat{\gamma}\hat{\beta} = \hat{A}^\top = Y^\top X\left(X^\intercal X\right)^{-1}$$

We conclude that it is not possible to estimate β and γ separately because it depends on the rank of the different matrices. Indeed, A has $n_x \times n_y$ parameters whereas the product $\gamma\beta$ has $n_z(n_x + n_y)$ parameters. In particular, the model is overidentified when:

$$n_z > \frac{n_x \times n_y}{n_x + n_y}$$

Otherwise, we obtain a constrained linear regression:

$$\begin{aligned}
\left(\hat{\beta}, \hat{\gamma}\right) &= \arg\min \frac{1}{2}\operatorname{trace}\left((XA - Y)^\intercal (XA - Y)\right) \\
\text{s.t. } & \gamma\beta = A^\top
\end{aligned}$$

3. Using chain rule, we have:

$$\frac{\partial\,\mathcal{L}_{i,j'}(\theta)}{\partial\,\gamma_{j,h}} = \frac{\partial\,\xi\left(y_{j'}(x_i), y_{i,j'}\right)}{\partial\,y_{j'}(x_i)} \frac{\partial\,y_{j'}(x_i)}{\partial\,v_{i,j'}} \frac{\partial\,v_{i,j'}}{\partial\,\gamma_{j,h}}$$

We notice that:

$$\frac{\partial\,v_{i,j'}}{\partial\,\gamma_{j,h}} = \begin{cases} z_{i,h} & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases}$$

It follows that:

$$\frac{\partial\,\mathcal{L}_{i,j}(\theta)}{\partial\,\gamma_{j,h}} = \xi'\left(y_j(x_i), y_{i,j}\right) f'_{z,y}(v_{i,j})\, z_{i,h}$$

and:

$$\frac{\partial\,\mathcal{L}_{i,j'}(\theta)}{\partial\,\gamma_{j,h}} = 0$$

We also have:

$$\begin{aligned} \frac{\partial\,\mathcal{L}_{i,j}(\theta)}{\partial\,\beta_{h,k}} &= \frac{\partial\,\xi\left(y_j(x_i), y_{i,j}\right)}{\partial\,y_j(x_i)} \frac{\partial\,y_j(x_i)}{\partial\,v_{i,j}} \frac{\partial\,v_{i,j}}{\partial\,z_{i,h}} \frac{\partial\,z_{i,h}}{\partial\,u_{i,h}} \frac{\partial\,u_{i,h}}{\partial\,\beta_{h,k}} \\ &= \xi'\left(y_j(x_i), y_{i,j}\right) f'_{z,y}(v_{i,j})\, \gamma_{j,h} f'_{x,z}(u_{i,h})\, x_{i,k} \end{aligned}$$

Finally, we deduce that:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\gamma} = \left(G_{y,y} \odot G_{y,v}\right)^\top Z$$

and:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\beta} = \left(\left(\left(G_{y,y} \odot G_{y,v}\right)\gamma\right) \odot G_{z,u}\right)^\top X$$

where $\partial_\gamma \mathcal{L}(\theta)$ is a $n_y \times n_z$ matrix, $G_{y,y} = \xi'\left(\hat{Y}, Y\right)$ is a $n \times n_y$ matrix, $G_{y,v} = f'_{z,y}(V)$ is a $n \times n_y$ matrix, $\partial_\beta \mathcal{L}(\theta)$ is a $n_z \times n_x$ matrix and $G_{z,u} = f'_{x,z}(U)$ is a $n \times n_z$ matrix.

4. We have:

$$f(z) = \frac{1}{1 + e^{-z}}$$

and:

$$\begin{aligned} f'(z) &= \frac{e^{-z}}{\left(1 + e^{-z}\right)^2} \\ &= \frac{1}{1 + e^{-z}} \left(1 - \frac{1}{1 + e^{-z}}\right) \\ &= f(z)\left(1 - f(z)\right) \end{aligned}$$

It follows that $f'_{z,y}(v_{i,j}) = f_{z,y}(v_{i,j})\left(1 - f_{z,y}(v_{i,j})\right) = y_j(x_i)\left(1 - y_j(x_i)\right)$ and $f'_{x,z}(u_{i,h}) = f_{x,z}(u_{i,h})\left(1 - f_{x,z}(u_{i,h})\right) = z_{i,h}\left(1 - z_{i,h}\right)$. We also have:

$$\xi(\hat{y}, y) = \frac{1}{2}\left(\hat{y} - y\right)^2$$

and:

$$\xi'(\hat{y}, y) = \hat{y} - y$$

We deduce that:

$$\frac{\partial\,\mathcal{L}_{i,j}(\theta)}{\partial\,\gamma_{j,h}} = \left(y_j(x_i) - y_{i,j}\right) y_j(x_i)\left(1 - y_j(x_i)\right) z_{i,h}$$

and:

$$\frac{\partial\,\mathcal{L}_{i,j}(\theta)}{\partial\,\beta_{h,k}} = (y_j(x_i) - y_{i,j})\, y_j(x_i)(1 - y_j(x_i))\,\gamma_{j,h} z_{i,h}(1 - z_{i,h})\, x_{i,k}$$

The matrix forms are:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\gamma} = \left(\left(\hat{Y} - Y\right) \odot G\left(\hat{Y}\right)\right)^\top Z$$

and:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\beta} = \left(\left(\left(\left(\hat{Y} - Y\right) \odot G\left(\hat{Y}\right)\right)\gamma\right) \odot G(Z)\right)^\top X$$

where $G\left(\hat{Y}\right) = \hat{Y} \odot \left(\mathbf{1}_{n\times n_y} - \hat{Y}\right)$ and $G(Z) = Z \odot (\mathbf{1}_{n\times n_z} - Z)$.

5. We have:

$$\xi(\hat{y}, y) = -\left(y \ln \hat{y} + (1 - y)\ln(1 - \hat{y})\right)$$

and:

$$\begin{aligned}
\xi'(\hat{y}, y) &= -\frac{y}{\hat{y}} + \frac{(1-y)}{(1-\hat{y})} \\
&= \frac{\hat{y}(1-y) - y(1-\hat{y})}{\hat{y}(1-\hat{y})} \\
&= \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\frac{\partial\,\mathcal{L}_i(\theta)}{\partial\,\gamma_h} &= \frac{(y(x_i) - y_i)}{y(x_i)(1 - y(x_i))} y(x_i)(1 - y(x_i))\, z_{i,h} \\
&= (y(x_i) - y_i)\, z_{i,h}
\end{aligned}$$

and:

$$\begin{aligned}
\frac{\partial\,\mathcal{L}_i(\theta)}{\partial\,\beta_{h,k}} &= \frac{(y(x_i) - y_i)}{y(x_i)(1 - y(x_i))} y(x_i)(1 - y(x_i))\,\gamma_h z_{i,h}(1 - z_{i,h})\, x_{i,k} \\
&= (y(x_i) - y_i)\,\gamma_h z_{i,h}(1 - z_{i,h})\, x_{i,k}
\end{aligned}$$

The matrix forms are:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\gamma} = \left(\hat{Y} - Y\right)^\top Z$$

and:

$$\frac{\partial\,\mathcal{L}(\theta)}{\partial\,\beta} = \left(\left(\left(\hat{Y} - Y\right)\gamma\right) \odot Z \odot (\mathbf{1} - Z)\right)^\top X$$

6. In the case of the softmax activation function, the value of $y_j(x_i)$ is equal to:

$$\begin{aligned}
y_j(x_i) &= f_{z,y}(v_{i,j}) \\
&= \frac{e^{v_{i,j}}}{\sum_{j'=1}^{n_C} e^{v_{i,j'}}}
\end{aligned}$$

This implies that $y_j(x_i)$ depends on $y_{j'}(x_i)$:

$$\sum_{j=1}^{n_C} y_j(x_i) = \sum_{j=1}^{n_C} \frac{e^{v_{i,j}}}{\sum_{j'=1}^{n_C} e^{v_{i,j'}}} = \frac{\sum_{j=1}^{n_C} e^{v_{i,j}}}{\sum_{j'=1}^{n_C} e^{v_{i,j'}}} = 1$$

The loss function is additive with respect to the index i, but not with respect to the index j.

7. We have:

$$\xi\left(\hat{y},y\right)=-y\ln\hat{y}$$

It follows that:

$$\xi'\left(\hat{y},y\right)=-\frac{y}{\hat{y}}$$

We also have:

$$f\left(z_j\right)=\frac{e^{z_j}}{\sum_{j'=1}^{n_C}e^{z_{j'}}}$$

We deduce that:

$$\begin{aligned}\frac{\partial\,f\left(z_j\right)}{\partial\,z_j} &= \frac{e^{z_j}\sum_{j'=1}^{n_C}e^{z_{j'}}-e^{2z_j}}{\left(\sum_{j'=1}^{n_C}e^{z_{j'}}\right)^2} \\ &= \frac{e^{z_j}\left(\sum_{j'=1}^{n_C}e^{z_{j'}}-e^{z_j}\right)}{\sum_{j'=1}^{n_C}e^{z_{j'}}\sum_{j'=1}^{n_C}e^{z_{j'}}} \\ &= \frac{e^{z_j}}{\sum_{j'=1}^{n_C}e^{z_{j'}}}\left(1-\frac{e^{z_j}}{\sum_{j'=1}^{n_C}e^{z_{j'}}}\right) \\ &= f\left(z_j\right)\left(1-f\left(z_j\right)\right)\end{aligned}$$

and:

$$\begin{aligned}\frac{\partial\,f\left(z_j\right)}{\partial\,z_{j'}} &= -\frac{e^{z_j+z_{j'}}}{\left(\sum_{j''=1}^{n_C}e^{z_{j''}}\right)^2} \\ &= -\frac{e^{z_j}}{\sum_{j''=1}^{n_C}e^{z_{j''}}}\frac{e^{z_{j'}}}{\sum_{j''=1}^{n_C}e^{z_{j''}}} \\ &= -f\left(z_j\right)f\left(z_{j'}\right)\end{aligned}$$

We notice that:

$$\begin{aligned}(*) &= \sum_{j'=1}^{n_C}\frac{\partial\,\xi\left(y_{j'}\left(x_i\right),y_{i,j'}\right)}{\partial\,y_{j'}\left(x_i\right)}\frac{\partial\,y_{j'}\left(x_i\right)}{\partial\,v_{i,j}} \\ &= -\left(\frac{y_{i,j}}{y_j\left(x_i\right)}\frac{\partial\,y_j\left(x_i\right)}{\partial\,v_{i,j}}+\sum_{j'\neq j}\frac{y_{i,j'}}{y_{j'}\left(x_i\right)}\frac{\partial\,y_{j'}\left(x_i\right)}{\partial\,v_{i,j}}\right) \\ &= -\left(\frac{y_{i,j}}{y_j\left(x_i\right)}f\left(v_{i,j}\right)\left(1-f\left(v_{i,j}\right)\right)-\sum_{j'\neq j}\frac{y_{i,j'}}{y_{j'}\left(x_i\right)}f\left(v_{i,j}\right)f\left(v_{i,j'}\right)\right) \\ &= -\left(\frac{y_{i,j}}{y_j\left(x_i\right)}y_j\left(x_i\right)\left(1-y_j\left(x_i\right)\right)-\sum_{j'\neq j}\frac{y_{i,j'}}{y_{j'}\left(x_i\right)}y_j\left(x_i\right)y_{j'}\left(x_i\right)\right) \\ &= -y_{i,j}\left(1-y_j\left(x_i\right)\right)+y_j\left(x_i\right)\sum_{j'\neq j}y_{i,j'} \\ &= -y_{i,j}+y_j\left(x_i\right)\sum_{j'=1}^{n_C}y_{i,j'} \\ &= y_j\left(x_i\right)-y_{i,j}\end{aligned}$$

because[10] $\sum_{j'=1}^{n_C} y_{i,j'} = 1$. We deduce that:

$$\begin{aligned}
\frac{\partial \mathcal{L}_i(\theta)}{\partial \gamma_{j,h}} &= \sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \sum_{j''=1}^{n_C} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j''}} \frac{\partial v_{i,j''}}{\partial \gamma_{j,h}} \\
&= \sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j}} z_{i,h} \\
&= \left(\sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j}} \right) z_{i,h} \\
&= (y_j(x_i) - y_{i,j}) z_{i,h}
\end{aligned}$$

and:

$$\begin{aligned}
\frac{\partial \mathcal{L}_i(\theta)}{\partial \beta_{h,k}} &= \sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \sum_{j''=1}^{n_C} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j''}} \frac{\partial v_{i,j''}}{\partial z_{i,h}} \frac{\partial z_{i,h}}{\partial u_{i,h}} \frac{\partial u_{i,h}}{\partial \beta_{h,k}} \\
&= \sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \sum_{j''=1}^{n_C} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j''}} \gamma_{j'',h} x_{i,k} \\
&= \sum_{j''=1}^{n_C} \left(\sum_{j'=1}^{n_C} \frac{\partial \xi(y_{j'}(x_i), y_{i,j'})}{\partial y_{j'}(x_i)} \frac{\partial y_{j'}(x_i)}{\partial v_{i,j''}} \right) \gamma_{j'',h} x_{i,k} \\
&= \left(\sum_{j''=1}^{n_C} (y_{j''}(x_i) - y_{i,j''}) \gamma_{j'',h} \right) x_{i,k} \\
&= \left(\sum_{j=1}^{n_C} (y_j(x_i) - y_{i,j}) \gamma_{j,h} \right) x_{i,k}
\end{aligned}$$

The matrix forms are then:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \gamma} = \left(\hat{Y} - Y \right)^{\intercal} Z$$

and:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \beta} = \left(\left(\hat{Y} - Y \right) \gamma \right)^{\top} X$$

8. We have $\partial_{\beta_{h,0}} u_{i,h} = 1$, $\partial_{\gamma_{j,0}} v_{i,j} = 1$ and $\partial_{\gamma_{j,n_{z+k}}} v_{i,j} = x_{i,k}$. We note β_0 the vector of dimension $n_z \times 1$, γ_0 the vector of dimension $n_y \times 1$ and γ_x the matrix of dimension $n_y \times n_x$. If there is a constant between the x's and the z's or between the z's and the y's, we have:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \beta_0} = (((G_{y,y} \odot G_{y,v}) \gamma) \odot G_{z,u})^{\top} \mathbf{1}_{n \times 1}$$

and:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \gamma_0} = (G_{y,y} \odot G_{y,v})^{\top} \mathbf{1}_{n \times 1}$$

In the case of direct links between the x's and the y's, we obtain:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \gamma_x} = (G_{y,y} \odot G_{y,v})^{\top} X$$

[10]All values $y_{i,j'}$ are equal to zero except one value that is equal to one.

15.4.8 Primal and dual problems of support vector machines

Hard margin classification

1. We note $\theta = (\beta_0, \beta)$ the $(1+K) \times 1$ vector of parameters. The objective function is equal to:

$$\begin{aligned} f(\theta) &= \frac{1}{2}\|\beta\|_2^2 \\ &= \frac{1}{2}\theta^\top Q\theta - \theta^\top R \end{aligned}$$

where:

$$Q = \begin{pmatrix} 0 & \mathbf{0}_K^\top \\ \mathbf{0}_K & I_K \end{pmatrix}$$

and $R = \mathbf{0}_{1+K}$. The constraints $y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1$ are equivalent to $y_i\beta_0 + y_i x_i^\top \beta \geq 1$. The matrix form $C\theta \geq D$ is defined by:

$$C = \begin{pmatrix} y_1 & y_1 x_{1,1} & \cdots & y_1 x_{1,K} \\ \vdots & \vdots & & \vdots \\ y_n & y_n x_{n,1} & \cdots & y_n x_{n,K} \end{pmatrix}$$

and $D = \mathbf{1}_n$.

2. The associated Lagrange function is:

$$\begin{aligned} \mathcal{L}(\beta_0, \beta; \alpha) &= \frac{1}{2}\|\beta\|_2^2 - \alpha^\top\left(y\beta_0 + y \odot X\beta - \mathbf{1}_n\right) \\ &= \frac{1}{2}\|\beta\|_2^2 - \left(\alpha^\top y\right)\beta_0 - \alpha^\top\left(y \odot X\beta\right) + \alpha^\top \mathbf{1}_n \\ &= \frac{1}{2}\|\beta\|_2^2 - \beta_0\left(\sum_{i=1}^n \alpha_i y_i\right) - \beta^\top\left(\sum_{i=1}^n \alpha_i y_i x_i\right) + \sum_{i=1}^n \alpha_i \end{aligned}$$

where $\alpha \geq \mathbf{0_n}$ is the vector of Lagrange multipliers. The first-order conditions are:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta_0} = -\sum_{i=1}^n \alpha_i y_i = 0$$

and:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta} = \beta - \sum_{i=1}^n \alpha_i y_i x_i = \mathbf{0}_K$$

3. At the optimum, we deduce that $\beta = \sum_{i=1}^n \alpha_i y_i x_i$ and the objective function of the dual problem is:

$$\begin{aligned} \mathcal{L}^*(\alpha) &= \frac{1}{2}\left(\sum_{i=1}^n \alpha_i y_i x_i\right)^\top\left(\sum_{i=1}^n \alpha_i y_i x_i\right) - \beta_0\left(\sum_{i=1}^n \alpha_i y_i\right) - \\ &\quad \left(\sum_{i=1}^n \alpha_i y_i x_i\right)^\top\left(\sum_{i=1}^n \alpha_i y_i x_i\right) + \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2}\left(\sum_{i=1}^n \alpha_i y_i x_i\right)^\top\left(\sum_{i=1}^n \alpha_i y_i x_i\right) \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j x_i^\top x_j \end{aligned}$$

Then, we have:

$$\begin{aligned} \hat{\alpha} &= \arg\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j \\ \text{s.t.} &\quad \alpha \geq \mathbf{0}_n \end{aligned}$$

because the Lagrange multipliers α_i are positive.

4. It follows that:

$$\hat{\alpha} = \arg\min \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j - \sum_{i=1}^{n} \alpha_i$$

Since we have $\sum_{i=1}^{n} \alpha_i = \alpha^\top \mathbf{1}_n$ and:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j = \alpha^\top \Gamma \alpha$$

where $\Gamma_{i,j} = y_i y_j x_i^\top x_j$, we deduce that:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2} \alpha^\top \Gamma \alpha - \alpha^\top \mathbf{1}_n \\ \text{s.t.} &\quad \alpha \geq \mathbf{0}_n \end{aligned}$$

5. We notice that:

$$C\theta \geq D \Leftrightarrow -C\theta \leq -D$$

By applying the direct computation, we deduce that:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2} \alpha^\top Q^* \alpha - \alpha^\top R^* \\ \text{s.t.} &\quad \alpha \geq \mathbf{0}_n \end{aligned}$$

where:

$$\begin{aligned} Q^* &= C Q^{-1} C^\top \\ &= C \begin{pmatrix} 0 & \mathbf{0}_K^\top \\ \mathbf{0}_K & I_K \end{pmatrix}^{-1} C^\top \\ &= \begin{pmatrix} y_1 & y_1 x_{1,1} & \cdots & y_1 x_{1,K} \\ \vdots & \vdots & & \vdots \\ y_n & y_n x_{n,1} & \cdots & y_n x_{n,K} \end{pmatrix} \cdot \begin{pmatrix} \infty & \mathbf{0}_K^\top \\ \mathbf{0}_K & I_K \end{pmatrix} \cdot \\ &\quad \begin{pmatrix} y_1 & y_1 x_{1,1} & \cdots & y_1 x_{1,K} \\ \vdots & \vdots & & \vdots \\ y_n & y_n x_{n,1} & \cdots & y_n x_{n,K} \end{pmatrix}^\top \\ &= \infty \left(y^\top y \right) + \Gamma \end{aligned}$$

where $\Gamma_{i,j} = y_i y_j x_i^\top x_j$ and:

$$\begin{aligned} R^* &= -C Q^{-1} R + D \\ &= -C Q^{-1} \mathbf{0} + \mathbf{1}_n \\ &= \mathbf{1}_n \end{aligned}$$

Finally, we obtain:

$$\begin{aligned}\hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \left(\infty \left(y^\top y\right) + \Gamma\right)\alpha - \alpha^\top \mathbf{1}_n \\ \text{s.t.} &\quad \alpha \geq \mathbf{0}_n\end{aligned}$$

The singularity of the matrix Q does not allow to define a proper dual problem. In particular, we observe a scaling issue of the Lagrange coefficients. This is why we reintroduce the constraint $\sum_{i=1}^n \alpha_i y_i = 0$, and replace the previous dual problem by:

$$\begin{aligned}\hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \Gamma\alpha - \alpha^\top \mathbf{1}_n \\ \text{s.t.} &\quad \left\{\begin{array}{l} y^\top \alpha = 0 \\ \alpha \geq \mathbf{0}_n \end{array}\right.\end{aligned}$$

Soft margin classification with binary hinge loss

1. We note $\theta = (\beta_0, \beta, \xi)$ the $(1 + K + n) \times 1$ vector of parameters. The objective function is equal to:

$$\begin{aligned}f(\theta) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ &= \frac{1}{2}\theta^\top Q\theta - \theta^\top R\end{aligned}$$

where:

$$Q = \left(\begin{array}{ccc} 0 & \mathbf{0}_K^\top & \mathbf{0}_n^\top \\ \mathbf{0}_K & I_K & \mathbf{0}_{K\times n} \\ \mathbf{0}_n & \mathbf{0}_{n\times K} & \mathbf{0}_{n\times n}\end{array}\right)$$

and:

$$R = \left(\begin{array}{c} \mathbf{0}_{K+1} \\ -C \cdot \mathbf{1}_n \end{array}\right)$$

The constraints $y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i$ are equivalent to $y_i\beta_0 + y_i x_i^\top \beta + \xi_i \geq 1$. The matrix form $C\theta \geq D$ is defined by:

$$C = \left(\begin{array}{cccccc} y_1 & y_1 x_{1,1} & \cdots & y_1 x_{1,K} & 1 & 0 \\ \vdots & \vdots & & \vdots & \ddots & \\ y_n & y_n x_{n,1} & \cdots & y_n x_{n,K} & 0 & 1 \end{array}\right)$$

and $D = \mathbf{1}_n$. The bounds $\xi_i \geq 0$ can be written as $\theta \geq \theta^-$ where:

$$\theta^- = \left(\begin{array}{c} -\infty \cdot \mathbf{1}_{1+K} \\ \mathbf{0}_n \end{array}\right)$$

2. The associated Lagrange function is:

$$\begin{aligned}\mathcal{L}(\beta_0, \beta, \xi; \alpha, \lambda) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i - \\ &\quad \sum_{i=1}^n \alpha_i\left(y_i\left(\beta_0 + x_i^\top \beta\right) - 1 + \xi_i\right) - \sum_{i=1}^n \lambda_i \xi_i\end{aligned}$$

where $\alpha_i \geq 0$ and $\lambda_i \geq 0$. The first-order conditions for β_0 and β are the same:

$$\frac{\partial \mathcal{L}(\beta_0, \beta; \alpha)}{\partial \beta_0} = -\sum_{i=1}^{n} \alpha_i y_i = 0$$

and:

$$\frac{\partial \mathcal{L}(\beta_0, \beta; \alpha)}{\partial \beta} = \beta - \sum_{i=1}^{n} \alpha_i y_i x_i = \mathbf{0}_K$$

The first-order condition for ξ is:

$$\frac{\partial \mathcal{L}(\beta_0, \beta; \alpha)}{\partial \xi} = C \cdot \mathbf{1}_n - \alpha - \lambda = \mathbf{0}_n$$

It follows that:

$$\begin{aligned} \mathcal{L}^*(\alpha, \lambda) &= \mathcal{L}^*(\alpha) + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \xi_i - \sum_{i=1}^{n} \lambda_i \xi_i \\ &= \mathcal{L}^*(\alpha) + \sum_{i=1}^{n} \xi_i (C - \alpha_i - \lambda_i) \\ &= \mathcal{L}^*(\alpha) \end{aligned}$$

because of the first-order condition for ξ_i. The objective function of the dual problem is then the same as previously, and does not depend on the Lagrange multipliers λ. However, $\lambda_i \geq 0$ and $C - \alpha_i - \lambda_i = 0$ implies that $C - \alpha_i \geq 0$ or $\alpha_i \leq C$. Finally, we obtain the following dual problem:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \Gamma \alpha - \alpha^\top \mathbf{1}_n \\ &\text{s.t.} \quad \begin{cases} y^\top \alpha = 0 \\ \mathbf{0}_n \leq \alpha \leq C \cdot \mathbf{1}_n \end{cases} \end{aligned}$$

3. Previously, the support vectors correspond to observations such that $\alpha_i \neq 0$. Here, the support vectors must also verify that $\xi_i = 0$, implying that $\lambda_i \neq 0$ or $\alpha_i \neq C$. Therefore, support vectors corresponds to training points such that $0 < \alpha_i < C$.

4. The Kuhn-Tucker conditions are $\min(\lambda_i, \xi_i) = 0$. We also have $\lambda_i = C - \alpha_i$. If $\alpha_i = C$, then $\lambda_i = 0$ and $\xi_i > 0$. Otherwise, we have $\lambda_i > 0$ and $\xi_i = 0$ in the case where $\alpha_i < C$. The two constraints $y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i$ and $\xi_i \geq 0$ implies that:

$$\xi_i \geq 1 - y_i\left(\beta_0 + x_i^\top \beta\right)$$

At the optimum, we deduce that:

$$\hat{\xi}_i = \max\left(0, 1 - y_i\left(\hat{\beta}_0 + x_i^\top \hat{\beta}\right)\right)$$

5. In Figure 15.3, we have represented the optimal values of β_0, β_1, β_2, $\sum_{i=1}^{n} \xi_i$ and the margin M with respect to C. We notice that the paths are not smooth. We verify that the soft margin classifier tends to the hard margin classifier when $C \to \infty$. In Figure 15.4, we show the optimal hyperplane when $C = 0.07$. We verify that the soft margin classifier has a larger margin than the hard margin classifier.

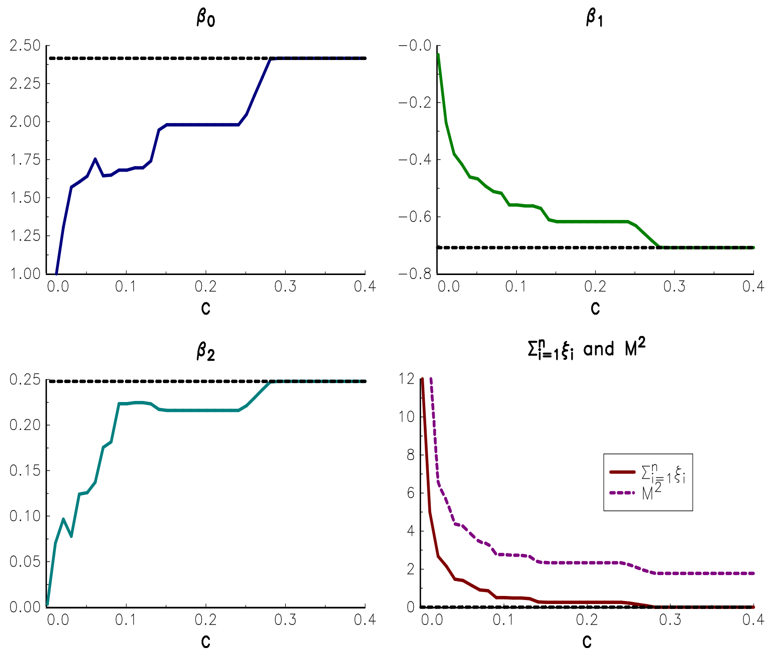

FIGURE 15.3: Optimal values of β_0, β_1, β_2, $\sum_{i=1}^n \xi_i$ and M

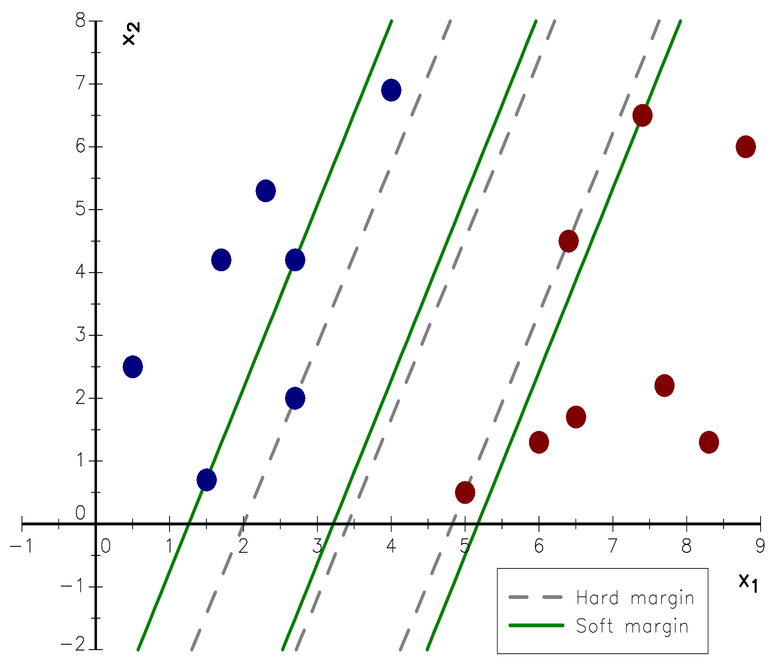

FIGURE 15.4: The hard margin classifier when $C = 0.07$

Soft margin classification with squared hinge loss

1. We note $\theta = (\beta_0, \beta, \xi)$ the $(1+K+n)\times 1$ vector of parameters. The objective function is equal to:

$$\begin{aligned} f(\theta) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 \\ &= \frac{1}{2}\theta^\top Q\theta - \theta^\top R \end{aligned}$$

where:

$$Q = \begin{pmatrix} 0 & \mathbf{0}_K^\top & \mathbf{0}_n^\top \\ \mathbf{0}_K & I_K & \mathbf{0}_{K\times n} \\ \mathbf{0}_n & \mathbf{0}_{n\times K} & 2C\cdot I_n \end{pmatrix}$$

and $R = \mathbf{0}_{1+K+n}$. The inequality and bound constraints are the same as the ones we have found for the binary hinge loss.

2. The associated Lagrange function is:

$$\begin{aligned} \mathcal{L}(\beta_0, \beta, \xi; \alpha, \lambda) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 - \\ & \quad \sum_{i=1}^n \alpha_i \left(y_i\left(\beta_0 + x_i^\top \beta\right) - 1 + \xi_i\right) - \sum_{i=1}^n \lambda_i \xi_i \end{aligned}$$

where $\alpha_i \geq 0$ and $\lambda_i \geq 0$. The first-order conditions for β_0 and β are:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta_0} = -\sum_{i=1}^n \alpha_i y_i = 0$$

and:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta} = \beta - \sum_{i=1}^n \alpha_i y_i x_i = \mathbf{0}_K$$

The first-order condition for ξ is:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\xi} = 2\cdot C\xi - \alpha - \lambda = \mathbf{0}_n$$

The Kuhn-Tucker conditions are $\min(\lambda_i, \xi_i) = 0$, implying that $\lambda_i\xi_i = 0$. This is equivalent to impose that $2C\cdot\xi - \alpha = \mathbf{0}_n$ or $\xi_i = \dfrac{\alpha_i}{2C}$. It follows that:

$$\begin{aligned} C\xi_i^2 - \alpha_i\xi_i - \lambda_i\xi_i &= C\left(\frac{\alpha_i}{2C}\right)^2 - \alpha_i\frac{\alpha_i}{2C} - 0\cdot\xi_i \\ &= \frac{\alpha_i^2}{4C} - \frac{\alpha_i^2}{2C} \\ &= -\frac{\alpha_i^2}{4C} \end{aligned}$$

and:

$$\begin{aligned} \mathcal{L}^*(\alpha, \lambda) &= \mathcal{L}^*(\alpha) - \frac{1}{4C}\sum_{i=1}^n \alpha_i^2 \\ &= \alpha^\top \mathbf{1}_n - \frac{1}{2}\alpha^\top\Gamma\alpha - \frac{1}{4C}\alpha^\top\alpha \\ &= \alpha^\top \mathbf{1}_n - \frac{1}{2}\alpha^\top\left(\Gamma + \frac{1}{2C}I_n\right)\alpha \end{aligned}$$

Finally, we obtain the following dual problem:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \left(\Gamma + \frac{1}{2C} I_n\right) \alpha - \alpha^\top \mathbf{1}_n \\ \text{s.t.} & \left\{ \begin{array}{l} y^\top \alpha = 0 \\ \alpha \geq \mathbf{0}_n \end{array} \right. \end{aligned}$$

This is a hard margin dual problem with a ridge regularization of the Γ matrix.

3. In Figure 15.5, we have represented the optimal values of β_0, β_1, β_2, $\sum_{i=1}^n \xi_i$ and the margin M with respect to C. We notice that the paths are smooth. We verify that the soft margin classifier tends to the hard margin classifier when $C \to \infty$.

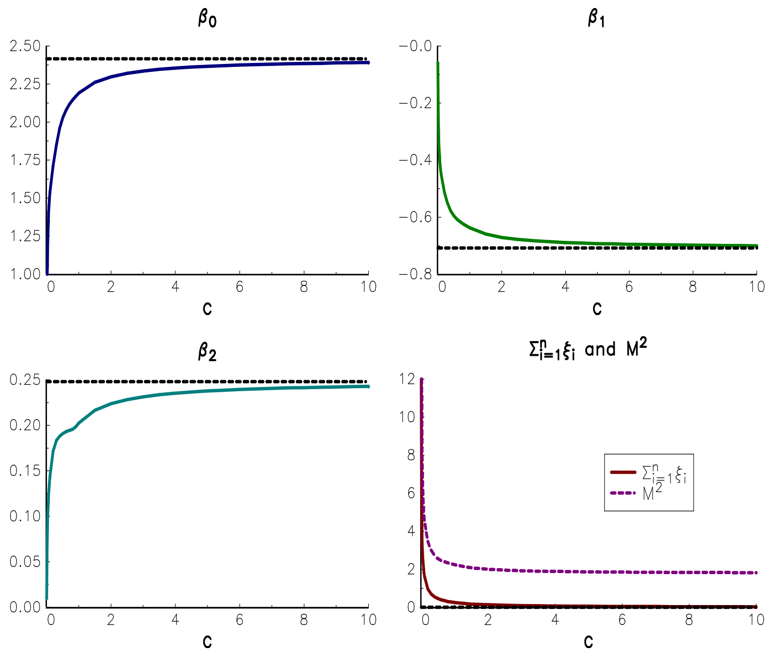

FIGURE 15.5: Convergence of soft margin classification with squared hinge loss

4. We obtain $\hat{\beta}_0 = 0.853$, $\hat{\beta}_1 = -0.371$ and $\hat{\beta}_1 = 0.226$. The optimal values of α_i and ξ_i are given in Table 15.3.

TABLE 15.3: Soft margin classification with squared hinge loss ($C = 1$)

i	1	2	3	4	5	6	7	8	
$\hat{\alpha}_i$	0.00	0.40	1.39	0.00	1.09	0.00	0.14	0.99	
$\hat{\xi}_i$	0.00	0.20	0.70	0.00	0.55	0.00	0.07	0.50	
i	9	10	11	12	13	14	15	16	17
$\hat{\alpha}_i$	0.00	0.00	1.16	0.00	0.00	0.00	0.22	2.48	3.13
$\hat{\xi}_i$	0.00	0.00	0.58	0.00	0.00	0.00	0.11	1.24	1.56

Soft margin classification with ramp loss

1. We have represented the four loss functions in Figure 15.6. The 0-1 loss function is not convex. The binary hinge loss function is convex, but not always differentiable. The

squared hinge loss function is convex and everywhere differentiable. Finally, the ramp loss function is bounded. The last three functions can be viewed as an approximation of the 0-1 loss function. Graphically, the ramp loss function is the best approximation.

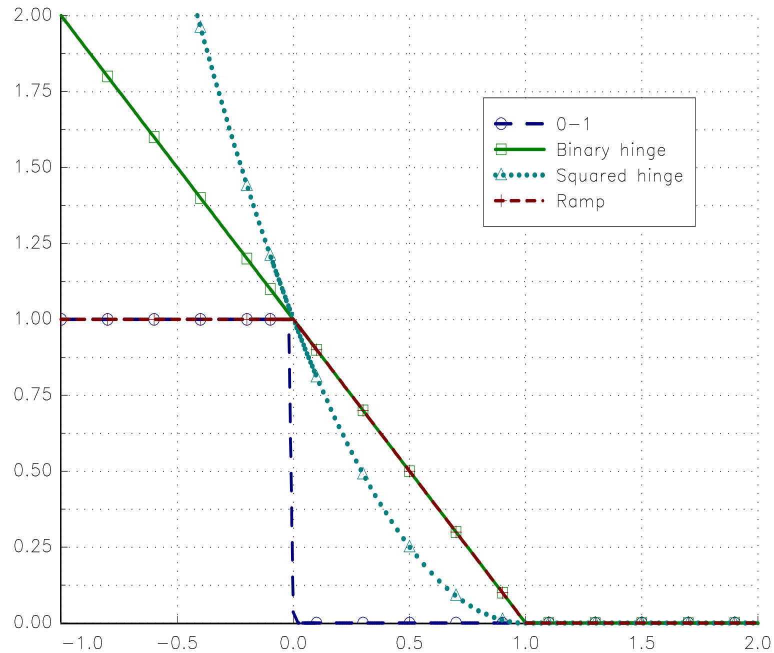

FIGURE 15.6: Comparison of SVM loss functions

2. We have:

$$\begin{aligned}
\mathcal{L}^{\text{ramp}}(x_i, y_i) &= \min\left(1, \mathcal{L}^{\text{hinge}}(x_i, y_i)\right) \\
&= \min\left(1, \max\left(0, 1 - y_i\left(\beta_0 + x_i^\top \beta\right)\right)\right) \\
&= \max\left(0, 1 - y_i\left(\beta_0 + x_i^\top \beta\right)\right) - \\
&\quad \max\left(0, -y_i\left(\beta_0 + x_i^\top \beta\right)\right) \\
&= \mathcal{L}^{\text{hinge}}(x_i, y_i) - \mathcal{L}^{\text{convex}}(x_i, y_i)
\end{aligned}$$

It follows that $\mathcal{L}^{\text{ramp}}$ is not convex, making the optimization problem tricky.

LS-SVM regression

1. We note $\theta = (\beta_0, \beta, \xi)$ the $(1 + K + n) \times 1$ vector of parameters. The objective function is equal to:

$$\begin{aligned}
f(\theta) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\xi_i^2 \\
&= \frac{1}{2}\theta^\top Q\theta - \theta^\top R
\end{aligned}$$

where:

$$Q = \begin{pmatrix} 0 & \mathbf{0}_K^\top & \mathbf{0}_n^\top \\ \mathbf{0}_K & I_K & \mathbf{0}_{K\times n} \\ \mathbf{0}_n & \mathbf{0}_{n\times K} & 2C \cdot I_n \end{pmatrix}$$

and $R = \mathbf{0}_{1+K+n}$. Let $Y = (y_i)$ and $X = (x_{i,k})$ be the output vector and the design matrix. The equality constraint is $A\theta = B$ where:

$$A = \begin{pmatrix} \mathbf{1}_n & X & I_n \end{pmatrix}$$

and $B = Y$.

2. The associated Lagrange function is:

$$\begin{aligned} \mathcal{L}(\beta_0, \beta, \xi; \alpha) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 + \\ &\quad \sum_{i=1}^n \alpha_i \left(y_i - \beta_0 - x_i^\top \beta - \xi_i\right) \end{aligned}$$

The first-order conditions for β_0 and β are:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta_0} = -\sum_{i=1}^n \alpha_i = 0$$

and:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\beta} = \beta - \sum_{i=1}^n \alpha_i x_i = \mathbf{0}_K$$

The first-order condition for ξ is:

$$\frac{\partial\,\mathcal{L}(\beta_0, \beta; \alpha)}{\partial\,\xi} = 2C \cdot \xi - \alpha = \mathbf{0}_n$$

Since we have $\alpha^\top \mathbf{1}_n = 0$, $\beta = X^\top \alpha$ and $\xi = \alpha/(2C)$, the objective function of the dual problem is equal to:

$$\begin{aligned} \mathcal{L}^*(\alpha) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 + \\ &\quad \sum_{i=1}^n \alpha_i y_i - \beta_0 \sum_{i=1}^n \alpha_i - \left(\sum_{i=1}^n \alpha_i x_i^\top\right)\beta - \sum_{i=1}^n \alpha_i \xi_i \\ &= \frac{1}{2}\beta^\top\beta + C \cdot \xi^\top \xi + \alpha^\top Y - \beta_0 \cdot \alpha^\top \mathbf{1}_n - \beta^\top \beta - \alpha^\top \xi \\ &= -\frac{1}{2}\beta^\top\beta + \frac{C}{4C^2}\alpha^\top\alpha + \alpha^\top Y - \frac{\alpha^\top\alpha}{2C} \\ &= \alpha^\top Y - \frac{1}{2}\left(\alpha^\top X X^\top \alpha\right) - \frac{\alpha^\top\alpha}{4C} \end{aligned}$$

Finally, we obtain the following dual problem:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \left(XX^\top + \frac{1}{2C}I_n\right)\alpha - \alpha^\top Y \\ &\text{s.t.} \quad \alpha^\top \mathbf{1}_n = 0 \end{aligned}$$

3. We also have $\hat{\beta} = X^\top \hat{\alpha}$. In this problem, all the training points are support vectors.

We deduce that $\hat{\beta}_0 = n^{-1} \sum_{i=1}^{n} \left(y_i - x_i^\top \hat{\beta} \right)$ and $\hat{\xi}_i = y_i - \hat{\beta}_0 - x_i^\top \hat{\beta}$. We verify that the residuals are centered:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i &= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - x_i^\top \hat{\beta} \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i^\top \hat{\beta} \right) - \hat{\beta}_0 \\
&= 0
\end{aligned}$$

ε-SVM regression

1. We note $\theta = (\beta_0, \beta, \xi^-, \xi^+)$ the $(1 + K + 2n) \times 1$ vector of parameters. The objective function is equal to:

$$\begin{aligned}
f(\theta) &= \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^{n} \left(\xi_i^- + \xi_i^+ \right) \\
&= \frac{1}{2} \theta^\top Q \theta - \theta^\top R
\end{aligned}$$

where:

$$Q = \begin{pmatrix} 0 & \mathbf{0}_K^\top & \mathbf{0}_{2n}^\top \\ \mathbf{0}_K & I_K & \mathbf{0}_{K \times 2n} \\ \mathbf{0}_{2n} & \mathbf{0}_{2n \times K} & \mathbf{0}_{2n \times 2n} \end{pmatrix}$$

and:

$$R = \begin{pmatrix} 0 \\ \mathbf{0}_K \\ -C \cdot \mathbf{1}_n \\ -C \cdot \mathbf{1}_n \end{pmatrix}$$

The constraints $\beta_0 + x_i^\top \beta - y_i \le \varepsilon + \xi_i^-$ and $y_i - \beta_0 - x_i^\top \beta \le \varepsilon + \xi_i^+$ are equivalent to $-\beta_0 - x_i^\top \beta + \xi_i^- \ge -y_i - \varepsilon$ and $\beta_0 + x_i^\top \beta + \xi_i^+ \ge y_i - \varepsilon$. The matrix form $C\theta \ge D$ is defined by:

$$C = \begin{pmatrix} -\mathbf{1}_n & -X & I_n & \mathbf{0}_{n \times n} \\ \mathbf{1}_n & X & \mathbf{0}_{n \times n} & I_n \end{pmatrix}$$

and:

$$D = \begin{pmatrix} -Y - \varepsilon \cdot \mathbf{1}_n \\ Y - \varepsilon \cdot \mathbf{1}_n \end{pmatrix}$$

The bounds $\xi_i^- \ge 0$ and $\xi_i^+ \ge 0$ can be written as $\theta \ge \theta^-$ where $\theta^- = (-\infty \cdot \mathbf{1}_{1+K}, \mathbf{0}_n, \mathbf{0}_n)$.

2. We introduce the Lagrange multipliers $\alpha_i^- \ge 0$, $\alpha_i^+ \ge 0$, $\lambda_i^- \ge 0$ and $\lambda_i^+ \ge 0$ associated

the four inequality constraints. The associated Lagrange function is:

$$\begin{aligned}
\mathcal{L}(\cdot) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \left(\xi_i^- + \xi_i^+\right) - \\
&\quad \sum_{i=1}^n \alpha_i^- \left(\varepsilon + \xi_i^- - \beta_0 - x_i^\top \beta + y_i\right) - \\
&\quad \sum_{i=1}^n \alpha_i^+ \left(\varepsilon + \xi_i^+ + \beta_0 + x_i^\top \beta - y_i\right) - \\
&\quad \sum_{i=1}^n \lambda_i^- \xi_i^- - \sum_{i=1}^n \lambda_i^+ \xi_i^+
\end{aligned}$$

The first-order conditions for β_0 and β are:

$$\frac{\partial\,\mathcal{L}(\cdot)}{\partial\,\beta_0} = \sum_{i=1}^n \alpha_i^- - \sum_{i=1}^n \alpha_i^+ = 0$$

and:

$$\frac{\partial\,\mathcal{L}(\cdot)}{\partial\,\beta} = \beta + \sum_{i=1}^n \alpha_i^- x_i - \sum_{i=1}^n \alpha_i^+ x_i = \mathbf{0}_K$$

We deduce that:

$$\mathbf{1}_n^\top \left(\alpha^- - \alpha^+\right) = 0$$

and:

$$\beta = X^\top \left(\alpha^+ - \alpha^-\right)$$

The first-order condition for ξ^- and ξ^+ are:

$$\frac{\partial\,\mathcal{L}(\cdot)}{\partial\,\xi^-} = C \cdot \mathbf{1}_n - \alpha^- - \lambda^- = \mathbf{0}_n$$

and:

$$\frac{\partial\,\mathcal{L}(\cdot)}{\partial\,\xi^+} = C \cdot \mathbf{1}_n - \alpha^+ - \lambda^+ = \mathbf{0}_n$$

This implies that:

$$C = \alpha_i^- + \lambda_i^- = \alpha_i^+ + \lambda_i^+$$

It follows that:

$$\begin{aligned}
(*) &= \sum_{i=1}^n \alpha_i^- \left(\varepsilon + \xi_i^- - \beta_0 - x_i^\top \beta + y_i\right) + \\
&\quad \sum_{i=1}^n \alpha_i^+ \left(\varepsilon + \xi_i^+ + \beta_0 + x_i^\top \beta - y_i\right) \\
&= \varepsilon\left(\alpha^- + \alpha^+\right)^\top \mathbf{1}_n - \beta_0\left(\alpha^- - \alpha^+\right)^\top \mathbf{1}_n - \\
&\quad \left(\alpha^- - \alpha^+\right)^\top X\beta + \left(\alpha^- - \alpha^+\right)^\top Y + \\
&\quad \sum_{i=1}^n \alpha_i^- \xi_i^- + \sum_{i=1}^n \alpha_i^+ \xi_i^+
\end{aligned}$$

We also have:

$$\begin{aligned} C\sum_{i=1}^{n}\left(\xi_i^- + \xi_i^+\right) &= C\sum_{i=1}^{n}\xi_i^- + C\sum_{i=1}^{n}\xi_i^+ \\ &= \sum_{i=1}^{n}\left(\alpha_i^- + \lambda_i^-\right)\xi_i^- + \sum_{i=1}^{n}\left(\alpha_i^+ + \lambda_i^+\right)\xi_i^+ \end{aligned}$$

Since $\left(\alpha^- - \alpha^+\right)^\top X = \beta^\top$, the objective function of the dual problem becomes:

$$\begin{aligned} \mathcal{L}^*(\cdot) &= \frac{1}{2}\left\|\beta\right\|_2^2 - \varepsilon\left(\alpha^- + \alpha^+\right)^\top \mathbf{1}_n + \beta^\top\beta - \left(\alpha^- - \alpha^+\right)^\top Y \\ &= -\frac{1}{2}\beta^\top\beta - \varepsilon\left(\alpha^- + \alpha^+\right)^\top \mathbf{1}_n - \left(\alpha^- - \alpha^+\right)^\top Y \end{aligned}$$

Since we have $\lambda_i^- \geq 0$, $\lambda_i^+ \geq 0$, $C\cdot\mathbf{1}_n - \alpha^- - \lambda^- = \mathbf{0}_n$ and $C\cdot\mathbf{1}_n - \alpha^+ - \lambda^+ = \mathbf{0}_n$, we deduce that $\alpha_i^- \leq C$ and $\alpha_i^+ \leq C$. Finally, we obtain the following dual problem:

$$\begin{aligned} \left\{\hat{\alpha}^-, \hat{\alpha}^+\right\} &= \arg\min \frac{1}{2}\left(\alpha^- - \alpha^+\right)^\top XX^\top\left(\alpha^- - \alpha^+\right) + \\ & \quad \varepsilon\left(\alpha^- + \alpha^+\right)^\top \mathbf{1}_n + \left(\alpha^- - \alpha^+\right)^\top Y \\ & \quad \text{s.t.} \left\{\begin{array}{l} \mathbf{1}_n^\top\left(\alpha^- - \alpha^+\right) = 0 \\ \mathbf{0}_n \leq \alpha^- \leq C\cdot\mathbf{1}_n \\ \mathbf{0}_n \leq \alpha^+ \leq C\cdot\mathbf{1}_n \end{array}\right. \end{aligned}$$

3. We note $\theta = \left(\alpha^-, \alpha^+\right)$ the $2n \times 1$ vector of parameters. The QP objective function is equal to:

$$f(\theta) = \frac{1}{2}\theta^\top Q\theta - \theta^\top R$$

where:

$$Q = \begin{pmatrix} XX^\top & -XX^\top \\ -XX^\top & XX^\top \end{pmatrix}$$

and:

$$R = \begin{pmatrix} -Y - \varepsilon\cdot\mathbf{1}_n \\ Y - \varepsilon\cdot\mathbf{1}_n \end{pmatrix}$$

The equality constraint is $A\theta = B$ where $A = \begin{pmatrix} \mathbf{1}_n^\top & -\mathbf{1}_n^\top \end{pmatrix}$ and $B = 0$. The bounds are $\theta^- = \mathbf{0}_{2n}$ and $\theta^+ = C\cdot\mathbf{1}_{2n}$.

4. We have:

$$\begin{aligned} \hat{\beta} &= X^\top\left(\hat{\alpha}^+ - \hat{\alpha}^-\right) \\ &= \sum_{i=1}^{n}\left(\hat{\alpha}_i^+ - \hat{\alpha}_i^-\right)x_i \end{aligned}$$

The Kuhn-Tucker conditions are:

$$\left\{\begin{array}{l} \min\left(\alpha_i^-, \varepsilon + \xi_i^- - \beta_0 - x_i^\top\beta + y_i\right) = 0 \\ \min\left(\alpha_i^+, \varepsilon + \xi_i^+ + \beta_0 + x_i^\top\beta - y_i\right) = 0 \\ \min\left(\lambda_i^-, \xi_i^-\right) = 0 \\ \min\left(\lambda_i^+, \xi_i^+\right) = 0 \end{array}\right.$$

We also remind that $C = \alpha_i^- + \lambda_i^- = \alpha_i^+ + \lambda_i^+$. The set $\mathcal{SV}^-$ of negative support vectors corresponds then to the observations such that $0 < \alpha_i^- < C$. In this case, we have $\varepsilon + \xi_i^- - \beta_0 - x_i^\top \beta + y_i = 0$ and $\xi_i^- = 0$. We deduce that $\hat{\beta}_0 = y_i + \varepsilon - x_i^\top \hat{\beta}$ when $i \in \mathcal{SV}^-$. The set $\mathcal{SV}^+$ of positive support vectors corresponds to the observations such that $0 < \alpha_i^+ < C$. In this case, we have $\varepsilon + \xi_i^+ + \beta_0 + x_i^\top \beta - y_i = 0$ and $\xi_i^+ = 0$. We deduce that $\hat{\beta}_0 = y_i - \varepsilon - x_i^\top \hat{\beta}$ when $i \in \mathcal{SV}^+$. Finally, we obtain:

$$\hat{\beta}_0 = \frac{\sum_{i \in \mathcal{SV}^-} \left(y_i + \varepsilon - x_i^\top \hat{\beta} \right) + \sum_{i \in \mathcal{SV}^+} \left(y_i - \varepsilon - x_i^\top \hat{\beta} \right)}{n_{\mathcal{SV}^-} + n_{\mathcal{SV}^+}}$$

where $n_{\mathcal{SV}^-}$ and $n_{\mathcal{SV}^+}$ are the number of negative and positive support vectors.

5. If $\alpha_i^- < C$, we have $\lambda_i^- > 0$ and $\xi_i^- = 0$. Otherwise, we have $\lambda_i^- = 0$ and $\xi_i^- > 0$. More precisely, we have $\varepsilon + \xi_i^- - \beta_0 - x_i^\top \beta + y_i = 0$ or:

$$\hat{\xi}_i^- = -\left(y_i + \varepsilon - \hat{\beta}_0 - x_i^\top \hat{\beta} \right)$$

If $\alpha_i^+ < C$, we have $\lambda_i^+ > 0$ and $\xi_i^+ = 0$. Otherwise, we have $\lambda_i^+ = 0$ and $\xi_i^+ > 0$. More precisely, we have $\varepsilon + \xi_i^+ + \beta_0 + x_i^\top \beta - y_i = 0$ or:

$$\hat{\xi}_i^+ = y_i - \varepsilon - \hat{\beta}_0 - x_i^\top \hat{\beta}$$

6. When ε is equal to zero, the term $\varepsilon \left(\alpha^- + \alpha^+ \right)^\top \mathbf{1}_n$ disappears in the objective function of the dual problem:

$$-\mathcal{L}^* \left(\cdot \right) = \frac{1}{2} \left(\alpha^- - \alpha^+ \right) X X^\top \left(\alpha^- - \alpha^+ \right) + \left(\alpha^- - \alpha^+ \right)^\top Y$$

By setting $\delta = \alpha^+ - \alpha^-$, we obtain the following QP problem:

$$\begin{aligned} \hat{\delta} &= \arg\min \frac{1}{2} \delta X X^\top \delta - \delta^\top Y \\ & \text{s.t.} \quad \left\{ \begin{array}{l} \mathbf{1}_n^\top \delta = 0 \\ -C \cdot \mathbf{1}_n \leq \delta \leq C \cdot \mathbf{1}_n \end{array} \right. \end{aligned}$$

The bounds are obtained by combining the inequalities $\mathbf{0}_n \leq \alpha^- \leq C \cdot \mathbf{1}_n$ and $\mathbf{0}_n \leq \alpha^+ \leq C \cdot \mathbf{1}_n$. We have $\hat{\beta} = X^\top \hat{\delta}$. The set $\mathcal{SV}$ of support vectors corresponds to the observations such that $-C < \delta_i < C$. It follows that:

$$\hat{\beta}_0 = \frac{1}{n_{\mathcal{SV}}} \sum_{i \in \mathcal{SV}} \left(y_i - x_i^\top \hat{\beta} \right)$$

where $n_{\mathcal{SV}}$ is the number of support vectors. Moreover, we have:

$$\hat{\xi}_i^- = \mathbb{1} \left\{ \hat{\delta}_i = -C \right\} \cdot \max \left(0, -\left(y_i - \hat{\beta}_0 - x_i^\top \hat{\beta} \right) \right)$$

and:

$$\hat{\xi}_i^+ = \mathbb{1} \left\{ \hat{\delta}_i = C \right\} \cdot \max \left(0, y_i - \hat{\beta}_0 - x_i^\top \hat{\beta} \right)$$

15.4.9 Derivation of the AdaBoost algorithm as the solution of the additive logit model

The following derivation comes from Section 10.4 in Hastie *et al.* (2009).

1. The objective function is equal to:

$$\begin{aligned}
\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right) &= \sum_{i=1}^{n} \mathcal{L}\left(y_i, \hat{g}_{(s-1)}\left(x_i\right) + \beta_{(s)} f_{(s)}\left(x_i\right)\right) \\
&= \sum_{i=1}^{n} e^{-y_i\left(\hat{g}_{(s-1)}(x_i)+\beta_{(s)} f_{(s)}(x_i)\right)} \\
&= \sum_{i=1}^{n} w_{i,s} e^{-y_i \beta_{(s)} f_{(s)}(x_i)}
\end{aligned}$$

where the expression of $w_{i,s}$ is equal to:

$$w_{i,s} = e^{-y_i \hat{g}_{(s-1)}(x_i)}$$

2. Since we have $y_i f_{(s)}\left(x_i\right) = 1$ if $y_i = y_{i,s}$ and $y_i f_{(s)}\left(x_i\right) = -1$ if $y_i \neq y_{i,s}$, we obtain:

$$\begin{aligned}
\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right) &= \sum_{i=1}^{n} w_{i,s} e^{-y_i \beta_{(s)} f_{(s)}(x_i)} \cdot \mathbb{1}\left\{y_i = y_{i,s}\right\} + \\
& \sum_{i=1}^{n} w_{i,s} e^{-y_i \beta_{(s)} f_{(s)}(x_i)} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\} \\
&= e^{-\beta_{(s)}} \sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i = y_{i,s}\right\} + \\
& e^{\beta_{(s)}} \sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\}
\end{aligned}$$

We notice that:

$$\sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i = y_{i,s}\right\} = \sum_{i=1}^{n} w_{i,s} - \sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\}$$

It follows that:

$$\begin{aligned}
\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right) &= e^{-\beta_{(s)}} \sum_{i=1}^{n} w_{i,s} + \left(e^{\beta_{(s)}} - e^{-\beta_{(s)}}\right) \sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\} \\
&= \left(e^{-\beta_{(s)}} + \left(e^{\beta_{(s)}} - e^{-\beta_{(s)}}\right) \frac{\sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\}}{\sum_{i=1}^{n} w_{i,s}}\right) \sum_{i=1}^{n} w_{i,s} \\
&= \left(\left(e^{\beta_{(s)}} - e^{-\beta_{(s)}}\right) \mathcal{L}_{(s)} + e^{-\beta_{(s)}}\right) \sum_{i=1}^{n} w_{i,s}
\end{aligned}$$

where $\mathcal{L}_{(s)}$ is the error rate:

$$\mathcal{L}_{(s)} = \frac{\sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\}}{\sum_{i=1}^{n} w_{i,s}}$$

3. It follows that the minimum is reached when:

$$\frac{\partial \, \mathcal{L}\left(\beta_{(s)}, f_{(s)}\right)}{\partial \, \beta} = 0$$

We have:

$$\frac{\partial \mathcal{L}\left(\beta_{(s)}, f_{(s)}\right)}{\partial \beta} = \left(\left(e^{\beta_{(s)}} + e^{-\beta_{(s)}}\right) \mathcal{L}_{(s)} - e^{-\beta_{(s)}}\right) \sum_{i=1}^{n} w_{i,s}$$

We deduce that:

$$\left(e^{\beta_{(s)}} + e^{-\beta_{(s)}}\right) \mathcal{L}_{(s)} - e^{-\beta_{(s)}} = 0$$

If we consider the change of variable $\alpha = e^{\beta_{(s)}} > 0$, we obtain:

$$\left(\alpha + \frac{1}{\alpha}\right) \mathcal{L}_{(s)} - \frac{1}{\alpha} = 0$$

or:

$$\alpha^2 = \frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}$$

The solution is:

$$\alpha^{\star} = \sqrt{\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}}$$

The optimal value of $\beta_{(s)}$ is then:

$$\begin{aligned} \hat{\beta}_{(s)} &= \ln \left(\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}\right)^{1/2} \\ &= \frac{1}{2} \ln \left(\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}\right) \end{aligned}$$

4. We have:

$$\begin{aligned} \hat{g}_{(s)}(x) &= \sum_{s'=1}^{s-1} \hat{\beta}_{(s')} \hat{f}_{(s')}(x) + \hat{\beta}_{(s)} \hat{f}_{(s)}(x) \\ &= \hat{g}_{(s-1)}(x) + \hat{\beta}_{(s)} \hat{f}_{(s)}(x) \end{aligned}$$

It follows that:

$$\begin{aligned} w_{i,s+1} &= e^{-y_i \hat{g}_{(s)}(x_i)} \\ &= e^{-y_i \hat{g}_{(s-1)}(x) - y_i \hat{\beta}_{(s)} \hat{f}_{(s)}(x_i)} \\ &= w_{i,s} e^{-y_i \hat{\beta}_{(s)} \hat{f}_{(s)}(x_i)} \end{aligned}$$

Using the fact that $-y_i \hat{f}_{(s)}(x_i) = 2 \cdot \mathbb{1}\{y_i \neq \hat{y}_{i,s}\} - 1$, the expression of $w_{i,s+1}$ becomes:

$$\begin{aligned} w_{i,s+1} &= w_{i,s} e^{2\hat{\beta}_{(s)} \mathbb{1}\{y_i \neq \hat{y}_{i,s}\}} e^{-\hat{\beta}_{(s)}} \\ &= w_{i,s} e^{w_s \cdot \mathbb{1}\{y_i \neq \hat{y}_{i,s}\}} e^{-\hat{\beta}_{(s)}} \end{aligned}$$

where:

$$w_s = 2\hat{\beta}_{(s)} = \ln \left(\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}\right)$$

The normalized weights are then:

$$\begin{aligned} w_{i,s+1} &= \frac{w_{i,s} e^{w_s \cdot \mathbb{1}\{y_i \neq \hat{y}_{i,s}\}} e^{-\hat{\beta}_{(s)}}}{\sum_{i'=1}^{n} w_{i',s} e^{w_s \cdot \mathbb{1}\left\{y_{i'} \neq \hat{y}_{i',s}\right\}} e^{-\hat{\beta}_{(s)}}} \\ &= \frac{w_{i,s} e^{w_s \cdot \mathbb{1}\{y_i \neq \hat{y}_{i,s}\}}}{\sum_{i'=1}^{n} w_{i',s} e^{w_s \cdot \mathbb{1}\left\{y_{i'} \neq \hat{y}_{i',s}\right\}}} \end{aligned}$$

5. The AdaBoost model can be viewed as an additive model, which is estimated using the forward stagewise method and the softmax loss function. However, there is a strong difference. Indeed, the solution $\hat{f}_{(s)}$ is given by:

$$\hat{f}_{(s)} = \arg\min \sum_{i=1}^{n} w_{i,s} e^{-w_s \cdot \mathbb{1}\{y_i \neq f_s(x_i)\}}$$

because $-y_i \hat{\beta}_{(s)} f_{(s)}(x_i) = 2\hat{\beta}_{(s)} \cdot \mathbb{1}\{y_i \neq y_{i,s}\} - \hat{\beta}_{(s)}$ and:

$$\mathcal{L}\left(\hat{\beta}_{(s)}, f_{(s)}\right) = e^{-\hat{\beta}_{(s)}} \sum_{i=1}^{n} w_{i,s} e^{-w_s \cdot \mathbb{1}\{y_i \neq y_{i,s}\}}$$

In the AdaBoost algorithm, the objective function for finding $\hat{f}_{(s)}$ is exogenous.

15.4.10 Weighted estimation

1. (a) We have:

$$\hat{\theta} = \arg\max \boldsymbol{\ell}_w(\theta)$$

   (b) The Jacobian matrix is:

$$J_w(\theta) = w^\top \odot J(\theta)$$

   where $J(\theta)$ is the Jacobian matrix associated to the unweighted log-likelihood function $\boldsymbol{\ell}(\theta) = \sum_{i=1}^{n} \boldsymbol{\ell}_i(\theta)$. For the Hessian matrix, we have:

$$H_w(\theta) = \sum_{i=1}^{n} w_i H_i(\theta)$$

   where $H_i(\theta)$ is the unweighted Hessian matrix:

$$H_i(\theta) = \begin{pmatrix} \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_1\,\partial\theta_1} & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_1\,\partial\theta_2 \boldsymbol{\ell}_i(\theta)} & \cdots & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_1\,\partial\theta_K} \\ \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_2\,\partial\theta_1} & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_2\,\partial\theta_2} & & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_2\,\partial\theta_K} \\ & & \ddots & \\ \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_K\,\partial\theta_1} & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_K\,\partial\theta_2} & & \dfrac{\partial^2 \boldsymbol{\ell}_i(\theta)}{\partial\theta_K\,\partial\theta_K} \end{pmatrix}$$

2. (a) The least squares loss function becomes:

$$\mathcal{L}_w(\theta) = \sum_{i=1}^{n} w_i \sum_{j=1}^{n_y} \frac{1}{2}\left(y_j(x_i) - y_{i,j}\right)^2$$

   We also have:

$$\frac{\partial\,\mathcal{L}_w(\theta)}{\partial\,\gamma} = \left(G_{y,y} \odot G_{y,v}\right)^\top (w \odot Z)$$

   and:

$$\frac{\partial\,\mathcal{L}_w(\theta)}{\partial\,\beta} = \left(\left(\left(G_{y,y} \odot G_{y,v}\right)\gamma\right) \odot G_{z,u}\right)^\top (w \odot X)$$

   where the matrices $G_{y,y}$, $G_{y,v}$, $G_{z,u}$, Z, X and γ are those defined in Exercise 15.4.7 on page 303.

(b) The cross-entropy loss function becomes:

$$\mathcal{L}_w(\theta) = -\sum_{i=1}^{n} w_i \left(y_i \ln y_j(x_i) + (1 - y_i)(1 - \ln y_j(x_i))\right)$$

We also have:

$$\frac{\partial\, \mathcal{L}_w(\theta)}{\partial\, \gamma} = \left(\hat{Y} - Y\right)^\top (w \odot Z)$$

and:

$$\frac{\partial\, \mathcal{L}_w(\theta)}{\partial\, \beta} = \left(\left(\left(\hat{Y} - Y\right)\gamma\right) \odot Z \odot (\mathbf{1} - Z)\right)^\top (w \odot X)$$

where the matrices $\hat{Y}$, Y, Z, X and γ are those defined in Exercise 15.4.7 on page 303.

3. (a) The soft margin classification problem becomes:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n} w_i \xi_i \\ &\text{s.t.} \left\{ \begin{array}{l} y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{array} \right. \quad \text{for } i = 1, \ldots, n \end{aligned}$$

(b) For the primal problem, the difference concerns the vector R:

$$R = \begin{pmatrix} \mathbf{0}_{K+1} \\ -Cw \end{pmatrix}$$

For the dual problem, the difference concerns the bounds:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \Gamma \alpha - \alpha^\top \mathbf{1}_n \\ &\text{s.t.} \left\{ \begin{array}{l} y^\top \alpha = 0 \\ \mathbf{0}_n \leq \alpha \leq Cw \end{array} \right. \end{aligned}$$

It follows that the support vectors corresponds to training points such that $0 < \alpha_i < Cw_i$.

(c) Hard margin classification assumes that the training set is linearly separable. So, there is no impact of weights on the solution. This is why it is impossible to introduce weights in the objective function of the hard margin classification problem.

Appendix A

Technical Appendix

A.4.1 Discrete-time random process

1. (a) We have $x_0 \in \{0\}$ and $\mathcal{F}_0 = \{0\}$. For $t = 1$, x_1 can take the value 0 or 1. We deduce that $x_1 \in \{0, 1\}$ and $\mathcal{F}_1 = \{\{0\}, \{1\}, \{0, 0\}, \{0, 1\}\}$. Similarly, we have $x_2 \in \{0, 1, 2\}$ and:

$$\mathcal{F}_2 = \left\{ \begin{array}{c} \{0\}, \{1\}, \{2\}, \{0,0\}, \{0,1\}, \{0,2\}, \{1,1\}, \{1,2\}, \\ \{0,0,0\}, \{0,0,1\}, \{0,1,1\}, \{0,1,2\} \end{array} \right\}$$

   (b) We have[1] $\mathbb{E}[|X(t)|] = t\sqrt{2/\pi} < \infty$. It $s < t$, we have:

$$\begin{array}{rcl} \mathbb{E}[X_t \mid \mathcal{F}_s] & = & \mathbb{E}[X_{t-1} + \varepsilon_t \mid \mathcal{F}_s] \\ & = & \mathbb{E}\left[X_s + \sum_{n=0}^{t-s+1} \varepsilon_{t-n} \mid \mathcal{F}_s\right] \\ & = & \mathbb{E}[X_s \mid \mathcal{F}_s] + \mathbb{E}\left[\sum_{n=0}^{t-s+1} \varepsilon_{t-n} \mid \mathcal{F}_s\right] \\ & = & x_s + 0 \\ & = & x_s \end{array}$$

   We deduce that X_t is a martingale.

2. We have:

$$\begin{array}{rcl} X_t & = & \phi X_{t-1} + \varepsilon_t \\ & = & \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ & = & \phi^2 X_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \\ & = & \displaystyle\sum_{n=0}^{\infty} \phi^n \varepsilon_{t-n} \end{array}$$

   because $\lim_{n\to\infty} \phi^n = 0$.

   (a) We have:

$$\begin{array}{rcl} \mathbb{E}\left[X_t^2\right] & = & \displaystyle\sum_{n=0}^{\infty} \phi^{2n} \mathbb{E}\left[\varepsilon_{t-n}^2\right] \\ & = & \sigma^2 \displaystyle\sum_{n=0}^{\infty} \phi^{2n} \\ & = & \dfrac{\sigma^2}{1-\phi^2} \\ & < & \infty \end{array}$$

[1] See Question 1(c) of Exercise A.4.2.

$\forall t \in \mathbb{Z}$, we have $\mathbb{E}[X_t] = \mathbb{E}[\sum_{n=0}^{\infty} \phi^n \varepsilon_{t-n}] = 0$. We deduce that:

$$\mathbb{E}[X_t] = \mathbb{E}[X_s]$$

$\forall (s,t) \in \mathbb{Z}$ with $t \geq s$ and $\forall u \geq 0$, we have:

$$\begin{aligned}
\mathbb{E}[X_{s+u}X_{t+u}] &= \mathbb{E}\left[\sum_{n=0}^{\infty} \phi^n \varepsilon_{s+u-n} \sum_{n=0}^{\infty} \phi^n \varepsilon_{t+u-n}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} \phi^n \varepsilon_{s+u-n} \sum_{n=t-s}^{\infty} \phi^n \varepsilon_{t+u-n}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} \phi^n \varepsilon_{s+u-n} \sum_{n=0}^{\infty} \phi^{t-s+n} \varepsilon_{s+u-n}\right] \\
&= \phi^{t-s}\mathbb{E}\left[\sum_{n=0}^{\infty} \phi^{2n} \varepsilon_{s+u-n}^2\right] \\
&= \frac{\phi^{t-s}}{1-\phi^2}\sigma^2 \\
&= \mathbb{E}[X_s X_t]
\end{aligned}$$

We deduce that X_t is a weak-sense stationary process.

(b) We have:

$$\mathbb{P}\{X_t \in \mathcal{A}\} = \int_{\mathcal{A}} \frac{\sqrt{1-\phi^2}}{\sigma^2\sqrt{2\pi}} \exp\left(-\frac{(1-\phi^2)x^2}{2\sigma^2}\right) \mathrm{d}x$$

The probability $\mathbb{P}\{X_t \in \mathcal{A}\}$ does not depend on t. We deduce that:

$$\mathbb{P}\{X_t \in \mathcal{A}\} = \mathbb{P}\{X_s \in \mathcal{A}\}$$

(c) We have:

$$\begin{aligned}
\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] &= \mathbb{E}[\phi X_{t-1} + \varepsilon_t \mid \mathcal{F}_{t-1}] \\
&= \phi x_{t-1} \\
&\neq x_{t-1}
\end{aligned}$$

It follows that X_t is a Markov process only if ϕ is equal to 0.

(d) We have:

$$\begin{aligned}
\mathbb{E}[X_t^2] &= \mathbb{E}\left[(\varepsilon_t + \theta\varepsilon_{t-1})^2\right] \\
&= (1+\theta^2)\sigma^2 \\
&< \infty
\end{aligned}$$

$\forall t \in \mathbb{Z}$, we have $\mathbb{E}[X_t] = \mathbb{E}[\varepsilon_t + \theta\varepsilon_{t-1}] = 0$. We deduce that $\mathbb{E}[X_t] = \mathbb{E}[X_s]$. $\forall (s,t) \in \mathbb{Z}$ with $t \geq s$ and $\forall u \geq 0$, we have:

$$\begin{aligned}
\mathbb{E}[X_{s+u}X_{t+u}] &= \mathbb{E}[(\varepsilon_{s+u} + \theta\varepsilon_{s+u-1})(\varepsilon_{t+u} + \theta\varepsilon_{t+u-1})] \\
&= \mathbb{E}[\varepsilon_{s+u}\varepsilon_{t+u}] + \theta\mathbb{E}[\varepsilon_{s+u-1}\varepsilon_{t+u}] + \\
&\quad \theta\mathbb{E}[\varepsilon_{s+u}\varepsilon_{t+u-1}] + \theta^2\mathbb{E}[\varepsilon_{s+u-1}\varepsilon_{t+u-1}] \\
&= \begin{cases} (1+\theta^2)\sigma^2 & \text{if } t = s \\ \theta\sigma^2 & \text{if } |t-s| = 1 \\ 0 & \text{if } |t-s| > 1 \end{cases}
\end{aligned}$$

Since we have $\mathbb{E}[X_{s+u}X_{t+u}] = \mathbb{E}[X_s X_t]$, we conclude that X_t is a weak-sense stationary process. It is easy to show that the probability $\mathbb{P}\{X_t \in \mathcal{A}\}$ does not depend on t. This implies that X_t is a strong-sense stationary process. We have:

$$\begin{aligned} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] &= \mathbb{E}[\varepsilon_t + \theta\varepsilon_{t-1} \mid \mathcal{F}_{t-1}] \\ &= \theta e_{t-1} \\ &\neq e_{t-1} + \theta e_{t-2} \\ &\neq x_{t-1} \end{aligned}$$

The MA(1) process is not a Markov process.

A.4.2 Properties of Brownian motion

1. (a) We have:

$$\begin{aligned} \mathbb{E}[W(t)] &= \mathbb{E}[W(t) - W(0)] \\ &= 0 \end{aligned}$$

(b) We assume that $s < t$. We have:

$$\begin{aligned} \operatorname{cov}(W(s)\,W(t)) &= \mathbb{E}[W(s)\,W(t)] \\ &= \mathbb{E}[W(s)(W(t) - W(s) + W(s))] \\ &= \mathbb{E}[W(s)(W(t) - W(s))] + \mathbb{E}\left[W^2(s)\right] \\ &= \mathbb{E}[(W(s) - W(0))(W(t) - W(s))] + \\ &\quad\ \mathbb{E}\left[(W(s) - W(0))^2\right] \\ &= 0 + s \\ &= s \end{aligned}$$

(c) We have:

$$\begin{aligned} \mathbb{E}[|W(t)|] &= \int_{-\infty}^{\infty} \frac{|x|}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}\,\mathrm{d}x \\ &= 2\int_0^{\infty} \frac{x}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}\,\mathrm{d}x \\ &= \sqrt{\frac{2}{\pi t}}\left[-te^{-\frac{1}{2t}x^2}\right]_0^{\infty} \\ &= \sqrt{\frac{2t}{\pi}} \\ &< \infty \end{aligned}$$

and:

$$\begin{aligned} \mathbb{E}[W(t) \mid \mathcal{F}_s] &= \mathbb{E}[W(s) + (W(t) - W(s)) \mid \mathcal{F}_s] \\ &= \mathbb{E}[W(s) \mid \mathcal{F}_s] + \mathbb{E}[W(t) - W(s) \mid \mathcal{F}_s] \\ &= w_s + 0 \\ &= w_s \end{aligned}$$

2. We only consider $h \to 0^+$ because the case $h \to 0^-$ is symmetric. We have[2]:

$$\begin{aligned}
\mathbb{P}\left\{\left|W(t+h) - W(t)\right| > \varepsilon\right\} &= 2\int_{-\infty}^{-\varepsilon} \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{1}{2h}x^2\right) \mathrm{d}x \\
&= 2\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi h}} e^{-\frac{1}{2h}\left(\sqrt{h}y+\varepsilon\right)^2} \sqrt{h}\,\mathrm{d}y \\
&= \sqrt{\frac{2}{\pi}} \exp\left(-\frac{\varepsilon^2}{2h}\right) \int_{-\infty}^{0} e^{-\frac{y^2}{2} - \frac{y\varepsilon}{\sqrt{h}}}\,\mathrm{d}y
\end{aligned}$$

We note $f(y) = -\frac{y^2}{2} - \frac{y\varepsilon}{\sqrt{h}}$. We have:

$$f'(y) = -y - \frac{\varepsilon}{\sqrt{h}}$$

Therefore, $f(y)$ is an increasing function on $\left]-\infty, -h^{-1/2}\varepsilon\right]$ and a decreasing function on $\left[-h^{-1/2}\varepsilon, 0\right]$. We deduce that:

$$0 \leq f(y) \leq \frac{\varepsilon^2}{2h}$$

and:

$$\int_{-\infty}^{0} \exp\left(-\frac{y^2}{2} - \frac{y\varepsilon}{\sqrt{h}}\right) \mathrm{d}y \leq C$$

where $C \in \mathbb{R}^+$. It follows that:

$$0 \leq \mathbb{P}\left\{\left|W(t+h) - W(t)\right| > \varepsilon\right\} \leq C\sqrt{\frac{2}{\pi}} \exp\left(-\frac{\varepsilon^2}{2h}\right)$$

and:

$$0 \leq \lim_{h \to 0^+} \mathbb{P}\left\{\left|W(t+h) - W(t)\right| > \varepsilon\right\} \leq C\sqrt{\frac{2}{\pi}} \lim_{h \to 0^+} \exp\left(-\frac{\varepsilon^2}{2h}\right)$$

Since $\lim_{h \to 0^+} \exp\left(-\frac{\varepsilon^2}{2h}\right) = 0$, we deduce that:

$$\lim_{h \to 0^+} \mathbb{P}\left\{\left|W(t+h) - W(t)\right| > \varepsilon\right\} = 0$$

3. We have:

$$\mathbb{E}\left[W^2(t)\right] = t$$

and:

$$\begin{aligned}
\mathbb{E}\left[W^2(t) \middle| \mathcal{F}_s\right] &= \mathbb{E}\left[\left(W(t) - W(s) + W(s)\right)^2 \middle| \mathcal{F}_s\right] \\
&= \mathbb{E}\left[W(s)^2 \mid \mathcal{F}_s\right] + 2\mathbb{E}\left[\left(W(t) - W(s)\right) W(s) \middle| \mathcal{F}_s\right] + \\
&\quad\ \mathbb{E}\left[\left(W(t) - W(s)\right)^2 \middle| \mathcal{F}_s\right] \\
&= 0 + 0 + (t - s) \\
&= t - s
\end{aligned}$$

[2]We use the change of variable $y = h^{-1/2}(x - \varepsilon)$.

If $X \sim \mathcal{N}(0,1)$, we know that $\mathbb{E}\left[X^3\right] = 0$ and $\mathbb{E}\left[X^4\right] = 3$. We deduce that $\mathbb{E}\left[W^3(t)\right] = 0$ and:

$$\begin{aligned} \mathbb{E}\left[W^4(t)\right] &= \mathbb{E}\left[\left(t^{1/2}\mathcal{N}(0,1)\right)^4\right] \\ &= 3t^2 \end{aligned}$$

We remind that $\mathbb{E}\left[\exp\left(\mathcal{N}\left(\mu,\sigma^2\right)\right)\right] = \exp\left(\mu + 0.5\sigma^2\right)$. We deduce that:

$$\mathbb{E}\left[e^{W(t)}\right] = e^{\frac{1}{2}t}$$

and:

$$\begin{aligned} \mathbb{E}\left[e^{W(t)} \middle| \mathcal{F}_s\right] &= \mathbb{E}\left[e^{W(s)+W(t)-W(s)} \middle| \mathcal{F}_s\right] \\ &= e^{W(s)}\mathbb{E}\left[e^{W(t)-W(s)} \middle| \mathcal{F}_s\right] \\ &= e^{W(s)}e^{\frac{1}{2}(t-s)} \\ &= e^{\frac{1}{2}(t-s)W(s)} \end{aligned}$$

4. If $X \sim \mathcal{N}(0,1)$ and $n \in \mathbb{N}^\star$, we have:

$$\begin{aligned} \mathbb{E}\left[X^{2n}\right] &= (2n-1)\,\mathbb{E}\left[X^{2(n-1)}\right] \\ &= (2n-1)(2n-3)\,\mathbb{E}\left[X^{2(n-2)}\right] \\ &= (2n-1)(2n-3)\cdots 5 \cdot 3 \cdot 1 \cdot \mathbb{E}\left[X^2\right] \\ &= (2n-1)!! \end{aligned}$$

and:

$$\mathbb{E}\left[X^{n+1}\right] = 0$$

where $n!!$ denotes the double factorial. We can also show that:

$$(2n-1)!! = \frac{(2n)!}{2^n n!}$$

It follows that:

$$\mathbb{E}\left[W^{2n}(t)\right] = \frac{(2n)!}{2^n n!}t^n$$

and:

$$\mathbb{E}\left[W^{n+1}(t)\right] = 0$$

For an even integer n, we deduce that:

$$\mathbb{E}\left[W^n(t)\right] = \frac{n!}{2^{n/2}\,(n/2)!}t^{n/2}$$

whereas for an odd integer n, $\mathbb{E}\left[W^n(t)\right]$ is equal to 0.

A.4.3 Stochastic integral for random step functions

1. We note $(*) = \int_a^b (\alpha f + \beta g)(t) \, \mathrm{d}W(t)$. We have:

$$\begin{aligned}
(*) &= \sum_{i=0}^{n-1} (\alpha f + \beta g)(t_i) \left(W(t_{i+1}) - W(t_i) \right) \\
&= \sum_{i=0}^{n-1} \left(\alpha f(t_i) + \beta g(t_i) \right) \left(W(t_{i+1}) - W(t_i) \right) \\
&= \alpha \sum_{i=0}^{n-1} f(t_i) \left(W(t_{i+1}) - W(t_i) \right) + \\
&\quad \beta \sum_{i=0}^{n-1} g(t_i) \left(W(t_{i+1}) - W(t_i) \right) \\
&= \alpha \int_a^b f(t) \, \mathrm{d}W(t) + \beta \int_a^b g(t) \, \mathrm{d}W(t)
\end{aligned}$$

We conclude that the stochastic integral verifies the linearity property. We now introduce the following partition:

$$a = t_0 < t_1 < \cdots < t_k = c < \cdots < t_n < b$$

We deduce that:

$$\begin{aligned}
\int_a^b f(t) \, \mathrm{d}W(t) &= \sum_{i=0}^{n-1} f(t_i) \left(W(t_{i+1}) - W(t_i) \right) \\
&= \sum_{i=0}^{k-1} f(t_i) \left(W(t_{i+1}) - W(t_i) \right) + \\
&\quad \sum_{i=k}^{n-1} f(t_i) \left(W(t_{i+1}) - W(t_i) \right) \\
&= \int_a^c f(t) \, \mathrm{d}W(t) + \int_c^b f(t) \, \mathrm{d}W(t)
\end{aligned}$$

It follows that the Chasles decomposition property holds.

2. We have:

$$\begin{aligned}
\mathbb{E}\left[\int_a^b f(t) \, \mathrm{d}W(t) \right] &= \mathbb{E}\left[\sum_{i=0}^{n-1} f(t_i) \left(W(t_{i+1}) - W(t_i) \right) \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[f(t_i) \left(W(t_{i+1}) - W(t_i) \right) \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[f(t_i) \right] \cdot \mathbb{E}\left[W(t_{i+1}) - W(t_i) \right] \\
&= 0
\end{aligned}$$

We note $(*) = \mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W(t)\int_a^b g(t)\,\mathrm{d}W(t)\right]$. We have:

$$\begin{aligned}
(*) &= \mathbb{E}\left[\sum_{i=0}^{n-1} f(t_i)\left(W(t_{i+1}) - W(t_i)\right)\sum_{i=0}^{n-1} g(t_i)\left(W(t_{i+1}) - W(t_i)\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} f(t_i)\,g(t_j)\left(W(t_{i+1}) - W(t_i)\right)\left(W(t_{j+1}) - W(t_j)\right)\right] \\
&= \sum_{i=0}^{n-1}\mathbb{E}\left[f(t_i)\,g(t_j)\left(W(t_{i+1}) - W(t_i)\right)^2\right] + \\
&\quad 2\sum_{i>j}\mathbb{E}\left[f(t_i)\,g(t_j)\left(W(t_{i+1}) - W(t_i)\right)\left(W(t_{j+1}) - W(t_j)\right)\right]
\end{aligned}$$

We deduce that:

$$\begin{aligned}
(*) &= \sum_{i=0}^{n-1}\mathbb{E}\left[f(t_i)\,g(t_j)\right]\cdot\mathbb{E}\left[\left(W(t_{i+1}) - W(t_i)\right)^2\right] + \\
&\quad 2\sum_{i>j}\mathbb{E}\left[f(t_i)\,g(t_j)\right]\cdot\mathbb{E}\left[\left(W(t_{i+1}) - W(t_i)\right)\left(W(t_{j+1}) - W(t_j)\right)\right] \\
&= \sum_{i=0}^{n-1}\mathbb{E}\left[f(t_i)\,g(t_j)\right](t_{i+1} - t_i) \\
&= \mathbb{E}\left[\sum_{i=0}^{n-1} f(t_i)\,g(t_j)\,(t_{i+1} - t_i)\right] \\
&= \mathbb{E}\left[\int_a^b f(t)\,g(t)\,\mathrm{d}t\right]
\end{aligned}$$

This result is known as the Itô isometry property. It is particularly useful for computing the covariance between two Itô processes $X_1(t)$ and $X_2(t)$:

$$\begin{aligned}
\operatorname{cov}\left(X_1(t), X_1(t)\right) &= \mathbb{E}\left[\int_a^b \sigma_1(t)\,\mathrm{d}W(t)\int_a^b \sigma_2(t)\,\mathrm{d}W(t)\right] \\
&= \mathbb{E}\left[\int_a^b \sigma_1(t)\,\sigma_2(t)\,\mathrm{d}t\right]
\end{aligned}$$

where $\sigma_1(t)$ and $\sigma_2(t)$ are the diffusion coefficients of $X_1(t)$ and $X_2(t)$.

3. We remind that:

$$\begin{aligned}
\operatorname{var}\left(\int_a^b f(t)\,\mathrm{d}W(t)\right) &= \mathbb{E}\left[\left(\int_a^b f(t)\,\mathrm{d}W(t)\right)^2\right] - \\
&\quad \mathbb{E}^2\left[\int_a^b f(t)\,\mathrm{d}W(t)\right]
\end{aligned}$$

It follows that:

$$\text{var}\left(\int_a^b f(t)\ \mathrm{d}W(t)\right) = \mathbb{E}\left[\int_a^b f(t)^2\ \mathrm{d}t\right] - 0^2$$

Since the mathematical expectation and the Riemann-Stieltjes are both linear, we conclude that:

$$\text{var}\left(\int_a^b f(t)\ \mathrm{d}W(t)\right) = \int_a^b \mathbb{E}\left[f^2(t)\right]\ \mathrm{d}t$$

A.4.4 Power of Brownian motion

1. We apply the Itô formula with $\mu(t,x) = 0$, $\sigma(t,x) = 1$ and $f(t,x) = x^2$. Since we have $\partial_t f(t,x) = 0$, $\partial_x f(t,x) = 2x$ and $\partial_x^2 f(t,x) = 2$, we deduce that:

$$\begin{aligned}
\mathrm{d}W^2(t) &= \mathrm{d}f(t, W(t)) \\
&= \left(0 + 2W(t) \times 0 + \frac{1}{2} \times 2 \times 1\right) \mathrm{d}t + (2W(t) \times 1)\ \mathrm{d}W(t) \\
&= \mathrm{d}t + 2W(t)\ \mathrm{d}W(t)
\end{aligned}$$

2. It follows that:

$$\int_0^t \mathrm{d}W^2(s) = \int_0^t \mathrm{d}s + 2\int_0^t W(s)\ \mathrm{d}W(s)$$

and:

$$W^2(t) = t + 2\int_0^t W(s)\ \mathrm{d}W(s)$$

Therefore, we obtain:

$$\begin{aligned}
I(t) &= \int_0^t W(s)\ \mathrm{d}W(s) \\
&= \frac{1}{2}\left(W^2(t) - t\right)
\end{aligned}$$

If follows that the expected value is equal to:

$$\begin{aligned}
\mathbb{E}\left[\int_0^t W(s)\ \mathrm{d}W(s)\right] &= \mathbb{E}\left[\frac{1}{2}\left(W^2(t) - t\right)\right] \\
&= \frac{\mathbb{E}\left[W^2(t)\right] - t}{2} \\
&= 0
\end{aligned}$$

Concerning the variance, we obtain:

$$\begin{aligned}
\text{var}\left(\int_0^t W(s)\ \mathrm{d}W(s)\right) &= \text{var}\left(\frac{1}{2}\left(W^2(t) - t\right)\right) \\
&= \frac{1}{4}\text{var}\left(W^2(t)\right) \\
&= \frac{1}{4}\left(\mathbb{E}\left[W^4(t)\right] - \mathbb{E}^2\left[W^2(t)\right]\right) \\
&= \frac{1}{4}\left(3t^2 - t^2\right) \\
&= \frac{t^2}{2}
\end{aligned}$$

We also notice that we can directly find this result by using the Itô isometry property:

$$\begin{aligned}\operatorname{var}\left(\int_0^t W(s)\,\mathrm{d}W(s)\right) &= \mathbb{E}\left[\left(\int_0^t W(s)\,\mathrm{d}W(s)\right)^2\right] \\ &= \int_0^t \mathbb{E}\left[W^2(s)\right]\,\mathrm{d}s \\ &= \int_0^t s\,\mathrm{d}s \\ &= \frac{t^2}{2}\end{aligned}$$

3. We use the function $f(t,x) = x^n$. We have $\partial_t f(t,x) = 0$, $\partial_x f(t,x) = nx^{n-1}$ and $\partial_x^2 f(t,x) = n(n-1)x^{n-2}$. The Itô formula gives:

$$\begin{aligned}\mathrm{d}W^n(t) &= \mathrm{d}f(t, W(t)) \\ &= \frac{1}{2}n(n-1)W(t)^{n-2}\,\mathrm{d}t + nW^{n-1}(t)\,\mathrm{d}W(t) \qquad \text{(A.1)}\end{aligned}$$

4. Since $I_n(t)$ is an Itô integral, we have:

$$\mathbb{E}[I_n(t)] = 0$$

and[3]:

$$\begin{aligned}\operatorname{var}(I_n(t)) &= \int_0^t \mathbb{E}\left[W^{2n}(s)\right]\,\mathrm{d}s \\ &= \int_0^t \frac{(2n)!}{2^n n!} s^n\,\mathrm{d}s \\ &= \frac{(2n)!}{2^n n!}\int_0^t s^n\,\mathrm{d}s \\ &= \frac{(2n)!}{2^n(n+1)!}t^{n+1}\end{aligned}$$

Finally, we obtain the following values of $\operatorname{var}(I_n(t))$:

n	1	2	3	4	5	6
$\operatorname{var}(I_n(t))$	$\frac{1}{2}t^2$	t^3	$\frac{15}{4}t^4$	$21t^5$	$\frac{315}{2}t^5$	$1485t^5$

5. In Question 4 of Exercise A.4.2, we have shown that:

$$\begin{aligned}\mathbb{E}[J_n(t)] &= \mathbb{E}[W^n(t)] \\ &= \begin{cases}\frac{n!}{2^{n/2}(n/2)!}t^{n/2} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd}\end{cases}\end{aligned}$$

Let us assume that n is odd. We have:

$$\begin{aligned}\operatorname{var}(J_n(t)) &= \mathbb{E}\left[W^{2n}(t)\right] \\ &= \frac{(2n)!}{2^n n!}t^n\end{aligned}$$

[3] We use the result obtained in Question 4 of Exercise A.4.2.

In the case where n is even, we deduce that:

$$\begin{aligned} \operatorname{var}\left(J_n(t)\right) &= \mathbb{E}\left[W^{2n}(t)\right] - \mathbb{E}^2\left[W^n(t)\right] \\ &= \left(\frac{(2n)!}{(n)!} - \left(\frac{n!}{(n/2)!}\right)^2\right)\frac{t^n}{2^n} \end{aligned}$$

6. We have:

$$\begin{aligned} \mathbb{E}\left[K_n(t)\right] &= \mathbb{E}\left[\int_0^t W^n(s)\,\mathrm{d}s\right] \\ &= \int_0^t \mathbb{E}\left[W^n(s)\right]\mathrm{d}s \end{aligned}$$

If n is odd, we deduce that:

$$\begin{aligned} \mathbb{E}\left[K_n(t)\right] &= \int_0^t 0\,\mathrm{d}s \\ &= 0 \end{aligned}$$

If n is even, we deduce that:

$$\begin{aligned} \mathbb{E}\left[K_n(t)\right] &= \int_0^t \frac{n!}{2^{n/2}(n/2)!}s^{n/2}\,\mathrm{d}s \\ &= \frac{n!}{2^{n/2}(n/2)!}\int_0^t s^{n/2}\,\mathrm{d}s \\ &= \frac{n!}{2^{n/2}(n/2)!}\left[\frac{s^{n/2+1}}{n/2+1}\right]_0^t \\ &= \frac{n!}{2^{n/2}(n/2+1)!}t^{n/2+1} \end{aligned}$$

7. Concerning the second non-central moment, we have:

$$\begin{aligned} \mathbb{E}\left[K_n^2(t)\right] &= \mathbb{E}\left[\left(\int_0^t W^n(s)\,\mathrm{d}s\right)^2\right] \\ &= \mathbb{E}\left[\left(\int_0^t W^n(s)\,\mathrm{d}s\right)\left(\int_0^t W^n(u)\,\mathrm{d}u\right)\right] \\ &= \mathbb{E}\left[\int_0^t\int_0^t W^n(s)\,W^n(u)\,\mathrm{d}s\,\mathrm{d}u\right] \\ &= \int_0^t\int_0^t \mathbb{E}\left[W^n(s)\,W^n(u)\right]\mathrm{d}s\,\mathrm{d}u \end{aligned}$$

by using Fubini's theorem. The challenge lies in computing the term $\mathbb{E}\left[W^n(s)\,W^n(u)\right]$. We face two difficulties. First, $W^n(s)$ and $W^n(u)$ are not independent. Therefore, the covariance involves the power series of $W(s)$. Second, we must distinguish the case $s < u$ and $u \geq u$. This is why it is long and tedious to compute the variance for high order n.

8. In the case $n = 1$, we obtain:

$$\begin{aligned}
\mathbb{E}\left[K_1^2(t)\right] &= \int_0^t \int_0^t \mathbb{E}\left[W(s)W(u)\right] \mathrm{d}s\,\mathrm{d}u \\
&= \int_0^t \left(\int_0^t \min(s,u)\ \mathrm{d}s\right) \mathrm{d}u \\
&= \int_0^t \left(\int_0^u \min(s,u)\ \mathrm{d}s + \int_u^t \min(s,u)\ \mathrm{d}s\right) \mathrm{d}u \\
&= \int_0^t \left(\int_0^u s\,\mathrm{d}s + \int_u^t u\,\mathrm{d}s\right) \mathrm{d}u \\
&= \int_0^t \left(\frac{u^2}{2} + u(t-u)\right) \mathrm{d}u \\
&= \left[\frac{u^3}{6} + \frac{u^2}{2}t - \frac{u^3}{3}\right]_0^t \\
&= \frac{t^3}{3}
\end{aligned}$$

and:

$$\begin{aligned}
\operatorname{var}(K_1(t)) &= \mathbb{E}\left[K_1^2(t)\right] - \mathbb{E}^2\left[K_1(t)\right] \\
&= \frac{1}{3}t^3
\end{aligned}$$

For $n = 2$ and $s < u$, we have:

$$\begin{aligned}
W^2(s)W^2(u) &= (W(s) - W(0))^2 (W(u) - W(0))^2 \\
&= (W(s) - W(0))^2 (W(s) - W(0) + W(u) - W(s))^2 \\
&= (W(s) - W(0))^4 + 2(W(s) - W(0))^3 (W(u) - W(s)) + \\
&\quad (W(s) - W(0))^2 (W(u) - W(s))^2
\end{aligned}$$

Since $W(s) - W(0)$ and $W(u) - W(s)$ are two independent random variables, we obtain:

$$\begin{aligned}
\mathbb{E}\left[W^2(s)W^2(u)\right] &= 3s^2 + 2 \cdot 0 + s(u-s) \\
&= 2s^2 + us
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\mathbb{E}\left[K_2^2(t)\right] &= \int_0^t \int_0^t \mathbb{E}\left[W^2(s)W^2(u)\right] \mathrm{d}s\,\mathrm{d}u \\
&= \int_0^t \left(\int_0^u \left(2s^2 + us\right) \mathrm{d}s + \int_u^t \left(2u^2 + us\right) \mathrm{d}s\right) \mathrm{d}u \\
&= \int_0^t \left(\left[\frac{2}{3}s^3 + \frac{1}{2}us^2\right]_0^u + \left[2u^2 s + \frac{1}{2}us^2\right]_u^t\right) \mathrm{d}u \\
&= \int_0^t \left(2u^2 t + \frac{1}{2}ut^2 - \frac{4}{3}u^3\right) \mathrm{d}u \\
&= \left[\frac{2}{3}u^3 t + \frac{1}{4}u^2 t^2 - \frac{1}{3}u^4\right]_0^t \\
&= \frac{7}{12}t^4
\end{aligned}$$

and:

$$\begin{aligned}\operatorname{var}(K_2(t)) &= \mathbb{E}\left[K_2^2(t)\right] - \mathbb{E}^2[K_2(t)] \\ &= \frac{7}{12}t^4 - \left(\frac{2!}{2\cdot 2!}t^2\right)^2 \\ &= \frac{1}{3}t^4\end{aligned}$$

For $n = 3$ and $s < u$, we have:

$$\begin{aligned}W^3(s)\,W^3(u) &= (W(s) - W(0))^3 (W(u) - W(0))^3 \\ &= (W(s) - W(0))^3 (W(s) - W(0) + W(u) - W(s))^3 \\ &= (W(s) - W(0))^6 + 3(W(s) - W(0))^5 (W(u) - W(s)) + \\ &\quad 3(W(s) - W(0))^4 (W(u) - W(s))^2 + \\ &\quad (W(s) - W(0))^3 (W(u) - W(s))^3\end{aligned}$$

It follows that:

$$\begin{aligned}\mathbb{E}\left[W^3(s)\,W^3(u)\right] &= \left(\frac{6!}{2^3\cdot 3!}s^3\right) + 3\cdot 0 + 3\cdot\left(\frac{4!}{2^2\cdot 2!}s^2\right)(u-s) + 0 \\ &= 6s^3 + 9us^2\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[K_3^2(t)\right] &= \int_0^t\int_0^t \mathbb{E}\left[W^3(s)\,W^3(u)\right]\,\mathrm{d}s\,\mathrm{d}u \\ &= \int_0^t\left(\int_0^u \left(6s^3 + 9us^2\right)\,\mathrm{d}s + \int_u^t \left(6u^3 + 9u^2s\right)\,\mathrm{d}s\right)\mathrm{d}u \\ &= \int_0^t\left(\left[\frac{3}{2}s^4 + 3us^3\right]_0^u + \left[6u^3s + \frac{9}{2}u^2s^2\right]_u^t\right)\mathrm{d}u \\ &= \int_0^t\left(-6u^4 + 6u^3t + \frac{9}{2}u^2t^2\right)\mathrm{d}u \\ &= \left[-\frac{6}{5}u^5 + \frac{3}{2}u^4t + \frac{3}{2}u^3t^2\right]_0^t \\ &= \frac{9}{5}t^5\end{aligned}$$

and:

$$\begin{aligned}\operatorname{var}(K_3(t)) &= \mathbb{E}\left[K_3^2(t)\right] - \mathbb{E}^2[K_3(t)] \\ &= \frac{9}{5}t^5\end{aligned}$$

9. Using Equation (A.1), we deduce that:

$$\begin{aligned}I_n(t) &= \int_0^t W^n(s)\,\mathrm{d}W(s) \\ &= \frac{1}{n+1}W^{n+1}(t) - \frac{n}{2}\int_0^t W^{n-1}(s)\,\mathrm{d}s \\ &= \frac{1}{n+1}J_{n+1}(t) - \frac{n}{2}K_{n-1}(t)\end{aligned}$$

and:

$$\begin{aligned} J_n(t) &= W^n(t) \\ &= \frac{1}{2}n(n-1)\int_0^t W^{n-2}(s)\,\mathrm{d}s + n\int_0^t W^{n-1}(s)\,\mathrm{d}W(s) \\ &= \frac{n(n-1)}{2}K_{n-2}(t) + nI_{n-1}(t) \end{aligned}$$

We also have:

$$\begin{aligned} K_n(t) &= \int_0^t W^n(s)\,\mathrm{d}s \\ &= \frac{2}{(n+2)(n+1)}J_{n+2}(t) - \frac{2}{(n+1)}I_{n+1}(t) \end{aligned}$$

A.4.5 Exponential of Brownian motion

1. We apply the Itô formula with $\mu(t,x) = 0$, $\sigma(t,x) = 1$ and $f(t,x) = e^x$. Since we have $\partial_t f(t,x) = 0$, $\partial_x f(t,x) = e^x$ and $\partial_x^2 f(t,x) = e^x$, we deduce that:

$$\begin{aligned} \mathrm{d}e^{W(t)} &= \mathrm{d}f(t, W(t)) \\ &= \frac{1}{2}e^{W(t)}\,\mathrm{d}t + e^{W(t)}\,\mathrm{d}W(t) \end{aligned}$$

and:

$$\int_0^t \mathrm{d}e^{W(s)} = \frac{1}{2}\int_0^t e^{W(s)}\,\mathrm{d}s + \int_0^t e^{W(s)}\,\mathrm{d}W(s)$$

It follows that:

$$e^{W(t)} = 1 + \frac{1}{2}\int_0^t e^{W(s)}\,\mathrm{d}s + \int_0^t e^{W(s)}\,\mathrm{d}W(s)$$

or:

$$X(t) = 1 + \frac{1}{2}Y(t) + Z(t) \tag{A.2}$$

2. $X(t) = e^{W(t)}$ is lognormal random variable with $\mathbb{E}[X(t)] = e^{\frac{1}{2}t}$ and $\operatorname{var}(X(t)) = e^{2t} - e^t$. In the case $Z(t) = \int_0^t e^{W(s)}\,\mathrm{d}W(s)$, we have:

$$\mathbb{E}[Z(t)] = 0$$

and:

$$\begin{aligned} \operatorname{var}(Z(t)) &= \int_0^t \mathbb{E}\left[e^{2W(s)}\right]\,\mathrm{d}s \\ &= \int_0^t e^{2s}\,\mathrm{d}s \\ &= \frac{1}{2}\left(e^{2t} - 1\right) \end{aligned}$$

In the case of $Y(t) = \int_0^t e^{W(s)}\,\mathrm{d}s$, we have:

$$\begin{aligned}
\mathbb{E}[Y(t)] &= \int_0^t \mathbb{E}\left[e^{W(s)}\right]\mathrm{d}s \\
&= \int_0^t e^{\frac{1}{2}s}\,\mathrm{d}s \\
&= \left[2e^{\frac{1}{2}s}\right]_0^t \\
&= 2\left(e^{\frac{1}{2}t} - 1\right)
\end{aligned}$$

and[4]:

$$\begin{aligned}
\mathbb{E}\left[Y^2(t)\right] &= \mathbb{E}\left[\left(\int_0^t e^{W(s)}\,\mathrm{d}s\right)\left(\int_0^t e^{W(u)}\,\mathrm{d}u\right)\right] \\
&= \int_0^t \int_0^t \mathbb{E}\left[e^{W(s)+W(u)}\right]\mathrm{d}s\,\mathrm{d}u \\
&= \int_0^t \int_0^t e^{\frac{1}{2}(s+u+2\min(s,u))}\,\mathrm{d}s\,\mathrm{d}u \\
&= \int_0^t \left(\int_0^u e^{\frac{1}{2}(3s+u)}\,\mathrm{d}s + \int_u^t e^{\frac{1}{2}(s+3u)}\,\mathrm{d}s\right)\mathrm{d}u \\
&= \int_0^t \left(\left[\frac{2}{3}e^{\frac{1}{2}(3s+u)}\right]_0^u + \left[2e^{\frac{1}{2}(s+3u)}\right]_u^t\right)\mathrm{d}u \\
&= \int_0^t \left(2e^{\frac{1}{2}(t+3u)} - \frac{2}{3}e^{2u} - \frac{2}{3}e^{\frac{1}{2}u}\right)\mathrm{d}u \\
&= \left[\frac{4}{3}e^{\frac{1}{2}(t+3u)} - \frac{4}{6}e^{2u} - \frac{4}{3}e^{\frac{1}{2}u}\right]_0^t \\
&= \frac{2}{3}e^{2t} - \frac{8}{3}e^{\frac{1}{2}t} + 2
\end{aligned}$$

It follows that:

$$\begin{aligned}
\operatorname{var}(Y(t)) &= \mathbb{E}\left[\left(\int_0^t e^{W(s)}\,\mathrm{d}s\right)^2\right] - \mathbb{E}^2\left[\int_0^t e^{W(s)}\,\mathrm{d}s\right] \\
&= \left(\frac{2}{3}e^{2t} - \frac{8}{3}e^{\frac{1}{2}t} + 2\right) - 4\left(e^{\frac{1}{2}t} - 1\right)^2 \\
&= \frac{2}{3}e^{2t} - 4e^t + \frac{16}{3}e^{\frac{1}{2}t} - 2
\end{aligned}$$

[4] We have $\mathbb{E}[W(s) + W(u)] = 0$ and:

$$\operatorname{var}(W(s) + W(u)) = s + u + 2\min(s,u)$$

3. From Equation (A.2), we deduce that:

$$\begin{aligned}\operatorname{cov}(Y(t), Z(t)) &= \frac{1}{4}\operatorname{var}(Y(t)) + \operatorname{var}(Z(t)) - \operatorname{var}(X(t)) \\ &= \frac{1}{4}\left(\frac{2}{3}e^{2t} - 4e^{t} + \frac{16}{3}e^{\frac{1}{2}t} - 2\right) + \frac{1}{2}\left(e^{2t} - 1\right) - \\ &\quad \left(e^{2t} - e^{t}\right) \\ &= -\frac{1}{3}e^{2t} + \frac{4}{3}e^{\frac{1}{2}t} - 1\end{aligned}$$

Finally, we obtain:

$$\rho(Y(t), Z(t)) = \frac{-1/3e^{2t} + 4/3e^{t/2} - 1}{\sqrt{\left(1/3e^{2t} - 2e^{t} + 8/3e^{t/2} - 1\right)\left(e^{2t} - 1\right)}}$$

In Figure A.1, we have reported the correlation $\rho(Y(t), Z(t))$ with respect to the time.

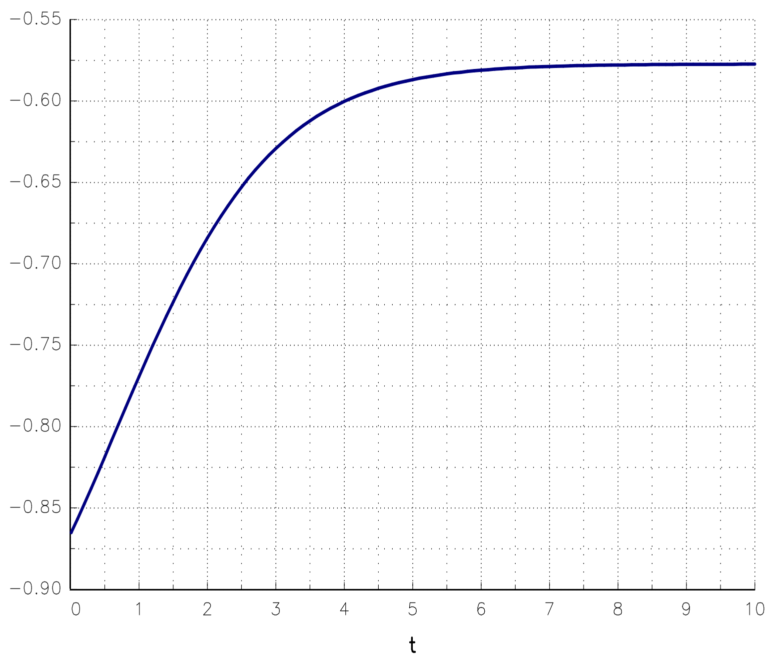

FIGURE A.1: Correlation between $\int_0^t e^{W(s)}\,\mathrm{d}s$ and $\int_0^t e^{W(s)}\,\mathrm{d}W(s)$

A.4.6 Exponential martingales

1. We have $x_s = X(s) = e^{W(s)}$ and:

$$\begin{aligned}\mathbb{E}\left[e^{W(t)} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[e^{W(s)+W(t)-W(s)} \mid \mathcal{F}_s\right] \\ &= e^{W(s)}\mathbb{E}\left[e^{W(t)-W(s)} \mid \mathcal{F}_s\right] \\ &= e^{W(s)}e^{\frac{1}{2}(t-s)} \\ &\neq x_s\end{aligned}$$

Therefore, $X(t)$ is not a martingale.

2. The previous question suggests that $m(t)$ is equal to $e^{-\frac{1}{2}t}$. We have:

$$\begin{aligned} M(t) &= e^{-\frac{1}{2}t}e^{W(t)} \\ &= e^{W(t)-\frac{1}{2}t} \end{aligned}$$

It follows that:

$$\begin{aligned} \mathbb{E}\left[M(t) \mid \mathcal{F}_s\right] &= \mathbb{E}\left[e^{W(t)-\frac{1}{2}t} \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[e^{W(s)+W(t)-W(s)-\frac{1}{2}t} \mid \mathcal{F}_s\right] \\ &= e^{W(s)}e^{\frac{1}{2}(t-s)}e^{-\frac{1}{2}t} \\ &= e^{W(s)-\frac{1}{2}s} \\ &= M(s) \end{aligned}$$

3. By applying Itô's lemma with $f(t,y) = e^y$, we obtain:

$$\begin{aligned} \mathrm{d}M(t) &= \left(-\frac{1}{2}e^{Y(t)}g^2(t) + \frac{1}{2}e^{Y(t)}g^2(t)\right)\mathrm{d}t + e^{Y(t)}g(t)\,\mathrm{d}W(t) \\ &= M(t)\,g(t)\,\mathrm{d}W(t) \end{aligned}$$

It follows that:

$$\begin{aligned} M(t) - M(0) &= \int_0^t \mathrm{d}M(s) \\ &= \int_0^t M(s)\,g(s)\,\mathrm{d}W(s) \end{aligned} \tag{A.3}$$

Since $g(t)$ is not random, we deduce that $Y(t)$ is a Gaussian process. We have:

$$\mathbb{E}\left[Y(t)\right] = -\frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s$$

and:

$$\begin{aligned} \operatorname{var}(Y(t)) &= \mathbb{E}\left[\left(\int_0^t g(s)\,\mathrm{d}W(s)\right)^2\right] \\ &= \int_0^t g^2(s)\,\mathrm{d}s \end{aligned}$$

We deduce that:

$$\begin{aligned} \mathbb{E}\left[M(t) \mid \mathcal{F}_s\right] &= \mathbb{E}\left[e^{-\frac{1}{2}\int_0^t g^2(u)\,\mathrm{d}u+\int_0^t g(u)\,\mathrm{d}W(u)} \mid \mathcal{F}_s\right] \\ &= e^{-\frac{1}{2}\int_0^s g^2(u)\,\mathrm{d}u+\int_0^s g(u)\,\mathrm{d}W(u)} \cdot \\ &\quad\; \mathbb{E}\left[e^{-\frac{1}{2}\int_s^t g^2(u)\,\mathrm{d}u+\int_s^t g(u)\,\mathrm{d}W(u)} \mid \mathcal{F}_s\right] \\ &= M(s)\,e^{-\frac{1}{2}\int_s^t g^2(u)\,\mathrm{d}u+\frac{1}{2}\int_s^t g^2(u)\,\mathrm{d}u} \\ &= M(s) \end{aligned}$$

We conclude that $M(t)$ is a martingale.

4. We notice that $M(0) = 1$. From Equation (A.3), we have:

$$M(t) = 1 + \int_0^t M(s)\, g(s)\, \mathrm{d}W(s) \tag{A.4}$$

Since $\int_0^t M(s)\, g(s)\, \mathrm{d}W(s)$ is an Itô integral, we deduce that $M(t)$ is a $\mathcal{F}_t$-martingale. We say that $M(t)$ is the exponential martingale of $X(t) = g(t)$ and we have:

$$\mathbb{E}[M(t)] = 1$$

We also notice that Equation (A.4) is related to the martingale representation theorem:

$$M(t) = \mathbb{E}[M(0)] + \int_0^t f(s)\, \mathrm{d}W(s)$$

where $f(s) = M(s)\, g(s)$.

A.4.7 Existence of solutions to stochastic differential equations

1. We have $\mu(t,x) = 1 + x$ and $\sigma(t,x) = 4$. It follows that:

$$\begin{aligned} |\mu(t,x) - \mu(t,y)| &= |1 + x - 1 - y| \\ &\leq 1 \cdot |x - y| \end{aligned}$$

and:

$$\begin{aligned} |\sigma(t,x) - \sigma(t,y)| = |4 - 4| &= 0 \\ &\leq 1 \cdot |x - y| \end{aligned}$$

We deduce that $K_1 = 1$. Using the Cauchy-Schwarz inequality, we also have:

$$\begin{aligned} |\mu(t,x)| = |1 + x| &\leq |1| + |x| \\ &\leq 4 \cdot (1 + |x|) \end{aligned}$$

and:

$$\begin{aligned} |\sigma(t,x)| &= 4 \\ &\leq 4 \cdot (1 + |x|) \end{aligned}$$

We deduce that $K_2 = 2$. We deduce that there exists a solution to the SDE and this solution is unique.

2. We have:

$$\begin{aligned} |\mu(t,x) - \mu(t,y)| &= |a(b - x) - a(b - y)| \\ &= |a| \cdot |x - y| \end{aligned}$$

and:

$$\begin{aligned} |\sigma(t,x) - \sigma(t,y)| &= |cx - cy| \\ &= |c| \cdot |x - y| \end{aligned}$$

By applying the Yamada-Watanabe theorem with $K = |a|$ and $h(u) = |c|\, u$, we conclude that the solution exists and is unique.

A.4.8 Itô calculus and stochastic integration

1. We consider the transform function:

$$\begin{aligned} Y(t) &= f(t, X(t)) \\ &= (1+t) X(t) - X(0) \end{aligned}$$

We have $\partial_t f(t,x) = x$, $\partial_x f(t,x) = 1+t$ and $\partial_x^2 f(t,x) = 0$. It follows that:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(X(t) - (1+t)\frac{X(t)}{1+t} \right) \mathrm{d}t + (1+t)\frac{1}{1+t}\,\mathrm{d}W(t) \\ &= \mathrm{d}W(t) \end{aligned}$$

and $Y(0) = X(0) - X(0) = 0$. We deduce that $Y(t) = W(t)$ and:

$$X(t) = \frac{X(0) + W(t)}{1+t}$$

2. Using $f(t,x) = x_0^{-1} - x^{-1}$, we have $\partial_t f(t,x) = 0$, $\partial_x f(t,x) = x^{-2}$ and $\partial_x^2 f(t,x) = -2x^{-3}$. It follows that:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{X(t)}{X^2(t)} - \frac{1}{2}\frac{2X^2(t)}{X^3(t)} \right) \mathrm{d}t + \frac{X^2(t)}{X^2(t)}\,\mathrm{d}W(t) \\ &= \mathrm{d}W(t) \end{aligned}$$

and $Y(0) = 0$. We deduce that $Y(t) = W(t)$ and:

$$X(t) = \frac{1}{X^{-1}(0) - W(t)}$$

3. We have:

$$\begin{aligned} \mathrm{d}X(t) &= -\left(\int_0^t \frac{1}{1-s}\,\mathrm{d}W(s) \right) \mathrm{d}t + \left(\frac{1-t}{1-t} \right) \mathrm{d}W(t) \\ &= -\frac{1}{1-t}\left(\int_0^t \frac{1-t}{1-s}\,\mathrm{d}W(s) \right) \mathrm{d}t + \mathrm{d}W(t) \\ &= -\frac{X(t)}{1-t}\,\mathrm{d}t + \mathrm{d}W(t) \end{aligned}$$

4. We have $f(t,x) = (1-t)^{-1} x$, $\partial_t f(t,x) = (1-t)^{-2} x$, $\partial_x f(t,x) = (1-t)^{-1}$ and $\partial_x^2 f(t,x) = 0$. It follows that:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{X(t)}{(1-t)^2} - \frac{X(t)}{(1-t)^2} \right) \mathrm{d}t + \frac{1}{1-t}\,\mathrm{d}W(t) \\ &= \frac{1}{1-t}\,\mathrm{d}W(t) \end{aligned}$$

We deduce that:

$$Y(t) - Y(0) = \int_0^t \frac{1}{1-s}\,\mathrm{d}W(s)$$

Since we have $X(0) = 0$, it follows that $Y(0) = 0$ and:

$$Y(t) = \int_0^t \frac{1}{1-s}\,\mathrm{d}W(s)$$

5. Using Ito's lemma, we have:

$$\mathrm{d}X(t) = \left(\partial_t f(t, W(t)) + \frac{1}{2}\partial_x^2 f(t, W(t))\right)\mathrm{d}t + \partial_x f(t, W(t))\,\mathrm{d}W(t)$$

Since $X(t)$ is a martingale, it satisfies the martingale representation theorem:

$$X(t) = \mathbb{E}[X(0)] + \int_0^t Z(s)\,\mathrm{d}W(s)$$

where $Z(t)$ is a $\mathcal{F}_t$-adapted process. We deduce that:

$$Z(t) = \partial_x f(t, W(t))$$

and:

$$\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2}\frac{\partial^2 f(t, W(t))}{\partial x^2} = 0$$

Then, $X(t)$ is a $\mathcal{F}_t$-martingale if this condition is satisfied.

6. In the case of the cubic martingale, we have $f(t,x) = x^3 - 3tx$, $\partial_t f(t,x) = -3x$, $\partial_x f(t,x) = 3x^2 - 3t$, $\partial_x^2 f(t,x) = 6x$ and:

$$\begin{aligned}\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2}\frac{\partial^2 f(t, W(t))}{\partial x^2} &= -3W(t) + \frac{1}{2}6W(t) \\ &= 0\end{aligned}$$

In the case of the quartic martingale, we have $f(t,x) = x^4 - 6tx^2 + 3t^2$, $\partial_t f(t,x) = -6x^2 + 6t$, $\partial_x f(t,x) = 4x^3 - 12tx$, $\partial_x^2 f(t,x) = 12x^2 - 12t$ and:

$$\begin{aligned}\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2}\frac{\partial^2 f(t, W(t))}{\partial x^2} &= -6W^2(t) + 6t + \frac{1}{2}\left(12W^2(t) - 12t\right) \\ &= 0\end{aligned}$$

We conclude that the necessary condition is satisfied for the cubic and quartic martingales.

7. We note $f(t,x) = e^{t/2}\cos(x)$. Since we have $\partial_t f(t,x) = \frac{1}{2}e^{t/2}\cos(x)$, $\partial_x f(t,x) = -e^{t/2}\sin(x)$, $\partial_x^2 f(t,x) = -e^{t/2}\cos(x)$, we obtain:

$$\mathrm{d}X(t) = \left(\frac{1}{2}e^{t/2}\cos W(t) - \frac{1}{2}e^{t/2}\cos W(t)\right)\mathrm{d}t - e^{t/2}\sin W(t)\,\mathrm{d}W(t)$$

It follows that:

$$X(t) = 1 - \int_0^t e^{s/2}\sin W(s)\,\mathrm{d}W(s)$$

$X(t)$ is an Itô integral. Moreover, we verify the condition:

$$\begin{aligned}\mathbb{E}\left[\int_0^t \left|e^{s/2}\sin W(s)\right|^2 \mathrm{d}s\right] &\leq \mathbb{E}\left[\int_0^t e^s\,\mathrm{d}s\right] \\ &\leq te^t \\ &< \infty\end{aligned}$$

Then, we deduce that $X(t) = e^{t/2}\cos W(t)$ is a martingale.

A.4.9 Solving a PDE with the Feynman-Kac formula

1. The function $g(t) = 1$ satisfies the Novikov condition:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s\right)\right] = e^{\frac{t}{2}} < \infty$$

We deduce that $Z(t) = W(t) - \int_0^t \mathrm{d}s$ is a Brownian motion under the probability measure $\mathbb{Q}$ defined by:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} &= \exp\left(\int_0^t \mathrm{d}W(s) - \frac{1}{2}\int_0^t \mathrm{d}s\right) \\ &= e^{W(t)-\frac{t}{2}}\end{aligned}$$

Since we have $\mathrm{d}Z(t) = \mathrm{d}W(t) - \mathrm{d}t$, we finally obtain that:

$$\begin{aligned}\mathrm{d}X(t) &= \mathrm{d}t + \mathrm{d}Z(t) + \mathrm{d}t \\ &= 2\,\mathrm{d}t + \mathrm{d}Z(t)\end{aligned}$$

2. Under the natural filtration $\mathcal{F}_t$, we have:

$$X(5) = X(t) + (5-t) + (W(5) - W(t))$$

and[5]:

$$\mathbb{E}\left[X(5) \mid \mathcal{F}_t\right] = x + (5-t)$$

If we now consider the filtration $\mathcal{G}_t$ generated by the Brownian motion $Z(t)$, we obtain:

$$X(5) = X(t) + 2(5-t) + (Z(5) - Z(t))$$

and:

$$\mathbb{E}\left[X(5) \mid \mathcal{G}_t\right] = x + 2(5-t)$$

We deduce that:

$$\begin{aligned}\mathbb{E}\left[X(5) \mid \mathcal{G}_0\right] &= x + 2(5-0) \\ &= x + 10\end{aligned}$$

3. We notice that:

$$-\partial_t V(t,x) + 10V(t,x) \neq \mathcal{A}_t V(t,x) + 4$$

where $\mathcal{A}_t$ is the infinitesimal generator of $X(t)$ with respect to the filtration $\mathcal{F}_t$. Therefore, we cannot apply the Feynman-Kac formula. However, by changing the probability measure, we have:

$$\begin{aligned}-\partial_t V(t,x) + 3V(t,x) &= \frac{1}{2}\partial_x^2 V(t,x) + 2\partial_x V(t,x) + 4 \\ &= \mathcal{A}'_t V(t,x) + 4\end{aligned}$$

where $\mathcal{A}'_t$ is the infinitesimal generator of $X(t)$ with respect to the filtration $\mathcal{G}_t$. We

[5] We note $X(t) = x$.

can then apply the Feynman-Kac formula and we have[6]:

$$\begin{aligned}
V(t,x) &= \mathbb{E}\left[X(5)\,e^{-\int_t^5 3\,\mathrm{d}s} + \int_t^5 4\left(e^{-\int_t^s 3\,\mathrm{d}u}\right)\mathrm{d}s \,\middle|\, \mathcal{G}_t\right] \\
&= e^{-3(5-t)}\cdot \mathbb{E}\left[X(5)\,|\,\mathcal{G}_t\right] + \int_t^5 4e^{-3(s-t)}\,\mathrm{d}s \\
&= e^{-3(5-t)}\cdot \mathbb{E}\left[X(5)\,|\,\mathcal{G}_t\right] + \left[-\frac{4}{3}e^{-3(s-t)}\right]_t^5 \\
&= (x+10-2t)\,e^{3t-15} + \frac{4}{3}\left(1-e^{3t-15}\right)
\end{aligned}$$

given that $X(t)=x$. We check the terminal condition:

$$\begin{aligned}
V(5,x) &= (x+10-2\times 5)\,e^{3\times 5-15} + \frac{4}{3}\left(1-e^{3\times 5-15}\right) \\
&= x
\end{aligned}$$

We also have:

$$\begin{aligned}
\partial_t V(t,x) &= 3(x+8-2t)\,e^{3t-15} \\
\partial_x V(t,x) &= e^{3t-15} \\
\partial_x^2 V(t,x) &= 0
\end{aligned}$$

It follows that $V(t,x)$ satisfies the PDE:

$$\begin{aligned}
-\partial_t V(t,x) + 3V(t,x) &= -3(x+8-2t)\,e^{3t-15} + 4\left(1-e^{3t-15}\right) + \\
&\quad\; 3(x+10-2t)\,e^{3t-15} \\
&= 2e^{3t-15} + 4 \\
&= \frac{1}{2}\partial_x^2 V(t,x) + 2\partial_x V(t,x) + 4
\end{aligned}$$

4. If the terminal value is $V(T,x)=e^x$, we obtain:

$$\begin{aligned}
V(t,x) &= \mathbb{E}\left[e^{X(5)}e^{-\int_t^5 3\,\mathrm{d}s} + \int_t^5 4\left(e^{-\int_t^s 3\,\mathrm{d}u}\right)\mathrm{d}s \,\middle|\, \mathcal{G}_t\right] \\
&= e^{-3(5-t)}\cdot \mathbb{E}\left[e^{X(5)} \,\middle|\, \mathcal{G}_t\right] + \frac{4}{3}\left(1-e^{-3(5-t)}\right)
\end{aligned}$$

We have:

$$X(5)\mid \mathcal{G}_t \sim \mathcal{N}(x+10-2t, 5-t)$$

We deduce that:

$$\begin{aligned}
\mathbb{E}\left[e^{X(5)} \,\middle|\, \mathcal{G}_t\right] &= e^{x+10-2t+\frac{1}{2}(5-t)} \\
&= e^{x+12.5-2.5t}
\end{aligned}$$

[6] At the initial date $t=0$, we have:

$$V(0,x) = (x+10)\,e^{-15} + \frac{4}{3}\left(1-e^{-15}\right)$$

We finally obtain the following solution[7]:

$$\begin{aligned} V(t,x) &= e^{-3(5-t)} \cdot e^{x+12.5-2.5t} + \frac{4}{3}\left(1-e^{-3(5-t)}\right) \\ &= e^{x-2.5+0.5t} + \frac{4}{3}\left(1-e^{3t-15}\right) \end{aligned}$$

We check the terminal condition:

$$\begin{aligned} V(5,x) &= e^{x-2.5+0.5\times 5} + \frac{4}{3}\left(1-e^{3\times 5-15}\right) \\ &= e^{x} \end{aligned}$$

We also have:

$$\begin{aligned} \partial_t V(t,x) &= 0.5e^{x-2.5+0.5t} - 4e^{3t-15} \\ \partial_x V(t,x) &= e^{x-2.5+0.5t} \\ \partial_x^2 V(t,x) &= e^{x-2.5+0.5t} \end{aligned}$$

It follows that $V(t,x)$ satisfies the PDE:

$$\begin{aligned} -\partial_t V(t,x) + 3V(t,x) &= 2.5e^{x-2.5+0.5t} + 4 \\ &= 0.5e^{x-2.5+0.5t} + 2e^{x-2.5+0.5t} + 4 \\ &= \frac{1}{2}\partial_x^2 V(t,x) + 2\partial_x V(t,x) + 4 \end{aligned}$$

A.4.10 Fokker-Planck equation

1. If we consider the following PDE:

$$\left\{ \begin{array}{l} -\partial_t V(t,x) = \frac{1}{2}\sigma^2\partial_x^2 V(t,x) + (a(b-x))\,\partial_x V(t,x) \\ V(T,x) = \mathbb{1}\{x = x_T\} \end{array} \right.$$

the solution is given by the Feynman-Kac formula:

$$\begin{aligned} V(t,x) &= \mathbb{E}\left[\mathbb{1}\{X(T) = x_T\} \mid X(t) = x\right] \\ &= \mathbb{P}\{X(T) = x_T \mid X(t) = x\} \end{aligned}$$

We have:

$$\partial_x\left[a(b-x)\,U(t,x)\right] = -aU(t,x) + a(b-x)\,\partial_x U(t,x)$$

and:

$$\partial_x^2\left[\sigma^2 U(t,x)\right] = \sigma^2\partial_x^2 U(t,x)$$

We deduce that the Fokker-Planck equation is:

$$\left\{ \begin{array}{l} \partial_t U(t,x) = aU(t,x) - a(b-x)\,\partial_x U(t,x) + \frac{1}{2}\sigma^2\partial_x^2 U(t,x) \\ U(0,x) = \mathbb{1}\{x = x_0\} \end{array} \right.$$

---

[7] At the initial date $t = 0$, we have:

$$V(0,x) = e^{x-2.5} + \frac{4}{3}\left(1-e^{-15}\right)$$

In Figure A.2, we have represented the probability density function $\mathbb{P}\{X(1) = x \mid X(0) = 0\}$ using the two approaches. For that, we solve the two PDE using finite difference methods. Let u_m^i be the numerical solution of $U(t_i, x_m)$. By construction, we have:

$$\begin{aligned} u_m^i &= \mathbb{P}\{X(t_i) = x_m \mid X(0) = 0\} \cdot \mathrm{d}x \\ &= \mathbb{P}\{X(t_i) = x_m \mid X(0) = 0\} \cdot h \end{aligned}$$

where h is the spatial mesh spacing meaning. Therefore, we have to divide the numerical solution by h in order to obtain the density.

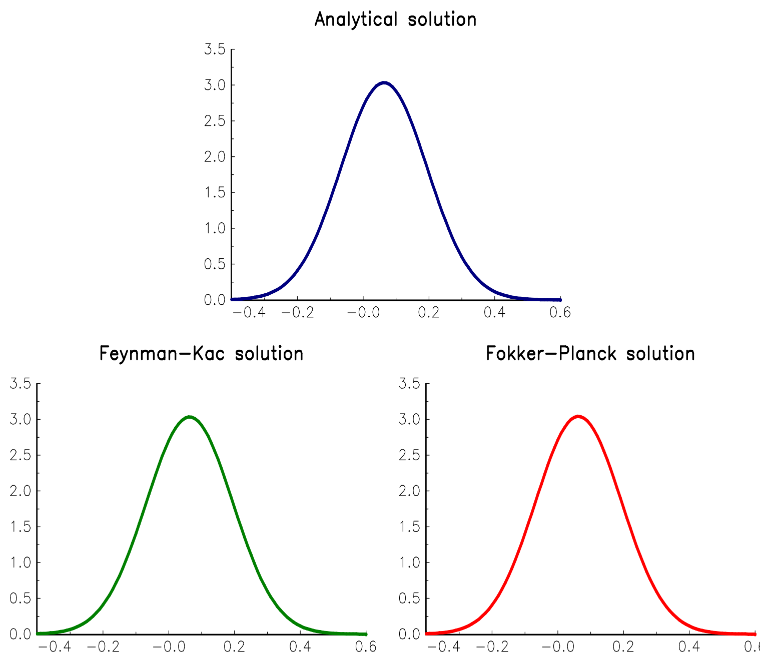

FIGURE A.2: Density function of the Ornstein-Uhlenbeck process

2. The solution is given by the Feynman-Kac PDE:

$$\begin{cases} -\partial_t V(t,x) = \frac{1}{2}\sigma^2 x^2 \partial_x^2 V(t,x) + \mu x \partial_x V(t,x) \\ V(T,x) = \mathbb{1}\{x = x_T\} \end{cases}$$

and the Fokker-Planck equation:

$$\begin{cases} \partial_t U(t,x) = \frac{1}{2}\sigma^2 x^2 \partial_x^2 U(t,x) + \left(2\sigma^2 x - \mu x\right) \partial_x U(t,x) + \left(\sigma^2 - \mu\right) U(t,x) \\ U(0,x) = \mathbb{1}\{x = x_0\} \end{cases}$$

because:

$$\partial_x [\mu x U(t,x)] = \mu U(t,x) + \mu x \partial_x U(t,x)$$

and:

$$\partial_x^2 \left[\sigma^2 x^2 U(t,x)\right] = 2\sigma^2 U(t,x) + 4\sigma^2 x \partial_x U(t,x) + \sigma^2 x^2 \partial_x^2 U(t,x)$$

In Figure A.3, we have represented the probability density function $\mathbb{P}\{X(1) = x \mid X(0) = 0\}$ using the two approaches.

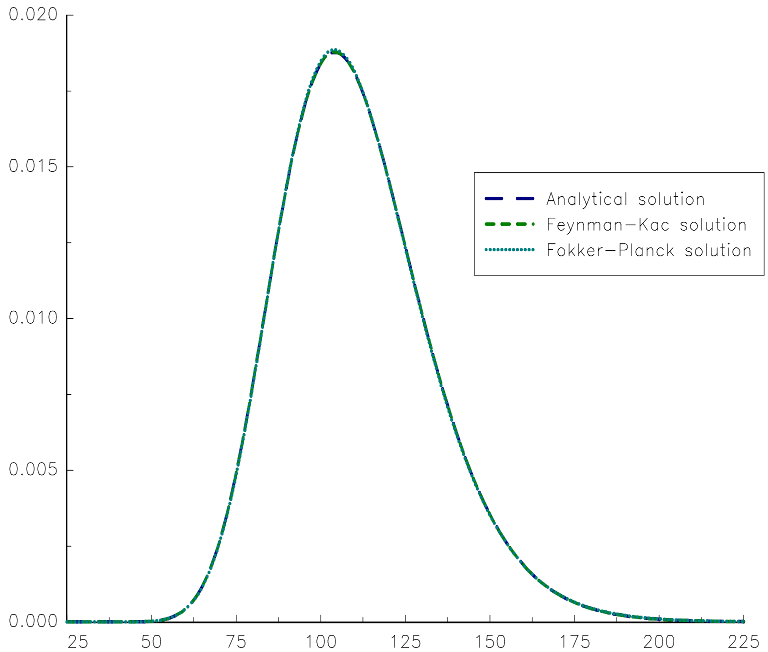

FIGURE A.3: Density function of the Geometric Brownian motion

A.4.11 Dynamic strategy based on the current asset price

1. We have:

$$\begin{aligned} \mathrm{d}V(t) &= n(t)\,\mathrm{d}S(t) \\ &= f(S(t))\,\mu(t,S(t))\,\mathrm{d}t + f(S(t))\,\sigma(t,S(t))\,\mathrm{d}W(t) \end{aligned}$$

2. We notice that $\partial_x F(x) = f(x)$ and $\partial_x^2 F(x) = f'(x)$. Using Itô's lemma, we have:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{\partial F}{\partial S}\mu(t,S(t)) + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2(t,S(t))\right)\mathrm{d}t + \\ &\quad \frac{\partial F}{\partial S}\sigma(t,S(t))\,\mathrm{d}W(t) \end{aligned}$$

We deduce that:

$$\begin{aligned} \mathrm{d}Y(t) &= f(S(t))\,\mu(t,S(t))\,\mathrm{d}t + f(S(t))\,\sigma(t,S(t))\,\mathrm{d}W(t) + \\ &\quad \frac{1}{2}f'(S(t))\,\sigma^2(t,S(t))\,\mathrm{d}t \end{aligned}$$

3. Since we have:

$$\mathrm{d}Y(t) = \mathrm{d}V(t) + \frac{1}{2}f'(S(t))\,\sigma^2(t,S(t))\,\mathrm{d}t$$

it follows that:

$$\mathrm{d}V(t) = \mathrm{d}Y(t) - \frac{1}{2}f'(S(t))\,\sigma^2(t,S(t))\,\mathrm{d}t$$

We deduce that:

$$\begin{aligned} V(T) - V(0) &= Y(T) - Y(0) - \frac{1}{2}\int_0^T f'(S(t))\,\sigma^2(t, S(t))\;\mathrm{d}t \\ &= F(S(T)) - F(S(0)) - \frac{1}{2}\int_0^T f'(S(t))\,\sigma^2(t, S(t))\;\mathrm{d}t \end{aligned}$$

Finally, we obtain:

$$\begin{aligned} V(T) &= V(0) + \int_{S(0)}^{S(T)} f(x)\;\mathrm{d}x - \frac{1}{2}\int_0^T f'(S(t))\,\sigma^2(t, S(t))\;\mathrm{d}t \\ &= G(T) + C(T) \end{aligned}$$

where:

$$G(T) = V(0) + \int_{S(0)}^{S(T)} f(x)\;\mathrm{d}x$$

and:

$$C(T) = -\frac{1}{2}\int_0^T f'(S(t))\,\sigma^2(t, S(t))\;\mathrm{d}t$$

The first term $G(T)$ can be interpreted as the option profile of the dynamic strategy at the maturity date, whereas $C(T)$ is the cost associated to the continuous trading strategy.

4. If $f(S(t)) = \mathbb{1}\{S(t) > S_\star\}$, we have:

$$\int_{S(0)}^{S(T)} \mathbb{1}\{x > S_\star\}\;\mathrm{d}x = \begin{cases} S(T) - S(0) & \text{if } S_\star \leq S(T) \\ S_\star - S(0) & \text{if } S(T) < S_\star \end{cases}$$

and:

$$f'(x) = -\delta(x - S_\star)$$

where $\delta(x)$ is the Dirac delta function. By assuming that $V(0) = S(0)$, we deduce that:

$$V(T) = S(T) + (S_\star - S(T))_+ - \frac{1}{2}\int_0^T \delta(S(t) - S_\star)\,\sigma^2(t, S(t))\;\mathrm{d}t$$

The option profile of this strategy is the underlying asset plus a put option where the strike is equal to $S^\star$. The cost of the stop-loss strategy is equal to:

$$\begin{aligned} C(T) &= -\frac{1}{2}\int_0^T \delta(S(t) - S_\star)\,\sigma^2(t, S(t))\;\mathrm{d}t \\ &< 0 \end{aligned}$$

5. If $f(S(t)) = \mathbb{1}\{S(t) < S_\star\}$, we have:

$$\int_{S(0)}^{S(T)} \mathbb{1}\{x < S_\star\}\;\mathrm{d}x = \begin{cases} S(T) - S(0) & \text{if } S(T) \leq S_\star \\ S_\star - S(0) & \text{if } S(T) > S_\star \end{cases}$$

and:

$$f'(x) = \delta(x - S_\star)$$

Use the results of the previous question, we have:

$$V(T) = S(T) - (S(T) - S_\star)_+ + \frac{1}{2}\int_0^T \delta(S(t) - S_\star)\,\sigma^2(t, S(t))\;\mathrm{d}t$$

The option profile of the stop-gain strategy is the underlying asset minus un call option where the strike is equal to $S_\star$. The cost of the strategy is positive, because if we cross the gain level $(S(t) > S_\star)$, we obtain an additional positive P&L that is equal to $S(t) - S_\star$.

6. (a) We buy the asset $(n(t) > 0)$ when the asset price $S(t)$ is below the price target $S_\star$. And we sell the asset $(n(t) < 0)$ when the asset price $S(t)$ is above the price target $S_\star$. This is a contrarian or mean-reverting strategy.

   (b) We have:

$$\begin{aligned}\int_{S(0)}^{S(T)} m\frac{S_\star - x}{x}\,\mathrm{d}x &= mS_\star\left[\ln x - \frac{x}{S_\star}\right]_{S(0)}^{S(T)} \\ &= mS_\star \ln\frac{S(T)}{S(0)} - m(S(T) - S(0))\end{aligned}$$

   and:

$$f'(x) = -m\frac{S_\star}{x^2}$$

   We deduce that:

$$\begin{aligned}V(T) - V(0) &= mS_\star(\ln S(T) - \ln S(0)) - m(S(T) - S(0)) + \\ &\quad \frac{m}{2}S_\star\int_0^T \frac{\sigma^2(t, S(t))}{S^2(t)}\,\mathrm{d}t\end{aligned}$$

   (c) When we have $\sigma(t, S(t)) = \sigma(t)\,S(t)$, we obtain:

$$V(T) - V(0) = mS_\star \ln\frac{S(T)}{S(0)} - m(S(T) - S(0)) + \frac{m}{2}S_\star\int_0^T \sigma^2(t)\;\mathrm{d}t$$

   and:

$$C(T) = \frac{m}{2}S_\star\,\mathrm{IV}(T)$$

   where $\mathrm{IV}(T)$ is the integrated variance:

$$\mathrm{IV}(T) = \int_0^T \sigma^2(t)\;\mathrm{d}t$$

   (d) When the asset volatility $\sigma(t)$ is low, the trend of the asset price is strong. It means that the asset price can continuously increase or decrease. This is the bad scenario for the strategy. The good scenario is when the asset price crosses many times the target price $S_\star$. This is why the strategy is more performing when the realized volatility is high. In Figure A.4, we have illustrated the strategy when the asset price follows a geometric Brownian motion and the target price is equal to the initial price[8]. We notice that the number of times that $S(t)$ crosses $S_\star$ increases with the volatility. Therefore, the vega of the strategy is positive.

[8] We have $S_\star = S(0) = 100$.

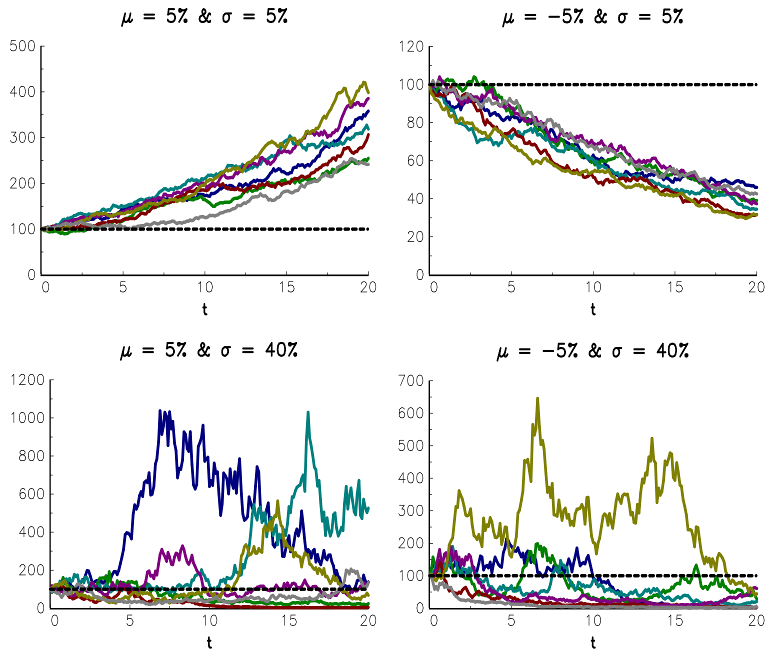

FIGURE A.4: Impact of the volatility on the mean-reverting strategy

A.4.12 Strong Markov property and maximum of Brownian motion

1. We have:

$$\Pr\{W(t) \geq x\} = \Pr\{W(t) \geq x, M(t) \geq x\} + \Pr\{W(t) \geq x, M(t) < x\}$$

Since $M(t) \geq W(t)$, we have $\Pr\{W(t) \geq x, M(t) < x\} = 0$ and:

$$\begin{aligned}\Pr\{W(t) \geq x\} &= \Pr\{W(t) \geq x, M(t) \geq x\} \\ &= \Pr\{W(t) \geq x \mid M(t) \geq x\} \cdot \Pr\{M(t) \geq x\} \\ &= \Pr\{W(t) \geq x \mid \tau_x \leq t\} \cdot \Pr\{M(t) \geq x\}\end{aligned}$$

Using the strong Markov property, we also have:

$$\begin{aligned}\Pr\{W(t) \geq x \mid \tau_x \leq t\} &= \Pr\{W(t) - W(\tau_x) \geq 0 \mid \tau_x \leq t\} \\ &= \Pr\{W(\tau_x + t - \tau_x) - W(\tau_x) \geq 0 \mid \tau_x \leq t\} \\ &= \Pr\{W(t - \tau_x) \geq 0 \mid \tau_x \leq t\} \\ &= \frac{1}{2}\end{aligned}$$

We deduce that:

$$\Pr\{W(t) \geq x\} = \frac{1}{2}\Pr\{M(t) \geq x\}$$

and:

$$\begin{aligned}\Pr\{M(t) \geq x\} &= 2\Pr\{W(t) \geq x\} \\ &= 2\left(1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right)\end{aligned}$$

2. Let $z = x - y \geq 0$. We have:

$$\begin{aligned} \Pr\{W(t) \geq 2x - y\} &= \Pr\{W(t) \geq x + z, M(t) \geq x\} + \\ & \quad \Pr\{W(t) \geq x + z, M(t) < x\} \\ &= \Pr\{W(t) \geq x + z, M(t) \geq x\} \end{aligned}$$

Indeed, we have:

$$\Pr\{W(t) \geq x + z, M(t) < x\} = 0$$

because $z \geq 0$. It follows that:

$$\begin{aligned} \Pr\{W(t) \geq 2x - y\} &= \Pr\{W(t) \geq x + z, M(t) \geq x\} \\ &= \Pr\{W(\tau_x + t - \tau_x) - W(\tau_x) \geq z \mid M(t) \geq x\} \cdot \\ & \quad \Pr\{M(t) \geq x\} \end{aligned}$$

Using the strong Markov property and the symmetry of the Brownian motion, we deduce that:

$$\begin{aligned} & \Pr\{W(\tau_x + t - \tau_x) - W(\tau_x) \geq z \mid M(t) \geq x\} \\ = \; & \Pr\{W(\tau_x + t - \tau_x) - W(\tau_x) \geq z \mid \tau_x \leq t\} \\ = \; & \Pr\{W(t - \tau_x) \geq z \mid \tau_x \leq t\} \\ = \; & \Pr\{W(t - \tau_x) \leq -z \mid \tau_x \leq t\} \\ = \; & \Pr\{W(t) \leq x - z \mid M(t) \geq x\} \end{aligned}$$

We conclude that:

$$\begin{aligned} \Pr\{W(t) \geq 2x - y\} &= \Pr\{W(t) \leq x - z \mid M(t) \geq x\} \cdot \Pr\{M(t) \geq x\} \\ &= \Pr\{W(t) \leq x - z, M(t) \geq x\} \\ &= \Pr\{W(t) \leq y, M(t) \geq x\} \end{aligned}$$

3. The joint density function of $(M(t), W(t))$ is defined as follows:

$$f(x, y) = -\frac{\partial^2 \Pr\{W(t) \leq y, M(t) \geq x\}}{\partial x \, \partial y}$$

We have:

$$\begin{aligned} \Pr\{W(t) \leq y, M(t) \geq x\} &= \Pr\{W(t) \geq 2x - y\} \\ &= 1 - \Phi\left(\frac{2x - y}{\sqrt{t}}\right) \end{aligned}$$

We deduce that:

$$\frac{\partial \Pr\{W(t) \leq y, M(t) \geq x\}}{\partial x} = -\frac{2}{\sqrt{t}} \phi\left(\frac{2x - y}{\sqrt{t}}\right)$$

It follows that:

$$\begin{aligned} f(x, y) &= -\frac{\partial}{\partial y}\left(-\frac{2}{\sqrt{t}} \phi\left(\frac{2x - y}{\sqrt{t}}\right)\right) \\ &= \frac{2}{\sqrt{2\pi t}} \frac{\partial}{\partial y}\left(\exp\left(-\frac{(2x - y)^2}{2t}\right)\right) \\ &= \frac{(2x - y)}{t^{3/2}} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{(2x - y)^2}{2t}\right) \end{aligned}$$

4. $W^{\mathbb{Q}}(t) = \mu t + W(t)$ is a standard Wiener process under the probability measure $\mathbb{Q}$ defined by[9]:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} &= \exp\left(-\mu W(t) - \frac{1}{2}\mu^2 t\right) \\ &= \exp\left(-\mu W^{\mathbb{Q}}(t) + \frac{1}{2}\mu^2 t\right)\end{aligned}$$

It follows that:

$$\begin{aligned}f_{(M_X,X)}(x,y) &= f_{\left(M_{W^{\mathbb{Q}}},W^{\mathbb{Q}}\right)}(x,y)\left|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right| \\ &= \exp\left(\mu y - \frac{1}{2}\mu^2 t\right)\frac{(2x-y)}{t^{3/2}}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{(2x-y)^2}{2t}\right) \\ &= \frac{(2x-y)}{t^{3/2}}\sqrt{\frac{2}{\pi}}\exp\left(\mu y - \frac{1}{2}\mu^2 t - \frac{(2x-y)^2}{2t}\right)\end{aligned}$$

5. To compute the density of $M_X(t)$, we use the fact that it is the marginal[10] of $f_{(M_X,X)}(x,y)$:

$$\begin{aligned}f_{M_X}(x) &= \int_{-\infty}^{x} f_{(M_X,X)}(x,y)\,\mathrm{d}y \\ &= \sqrt{\frac{2}{\pi t}}\int_{-\infty}^{x}\frac{(2x-y)}{t}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\,\mathrm{d}y \\ &= \sqrt{\frac{2}{\pi t}}\left(\left[e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\right]_{-\infty}^{x} - \mu\int_{-\infty}^{x}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\,\mathrm{d}y\right) \\ &= \sqrt{\frac{2}{\pi t}}e^{\mu x-\frac{1}{2}\mu^2 t-\frac{(2x-x)^2}{2t}} - \mu\int_{-\infty}^{x}\sqrt{\frac{2}{\pi t}}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\,\mathrm{d}y \\ &= \frac{2}{\sqrt{2\pi t}}e^{-\frac{(x-\mu t)^2}{2t}} - \mu\int_{-\infty}^{x}\sqrt{\frac{2}{\pi t}}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\,\mathrm{d}y\end{aligned}$$

Using the change of variable $z = t^{-1/2}(y - \mu t - 2x)$, we have:

$$\begin{aligned}\int_{-\infty}^{x}\sqrt{\frac{2}{\pi t}}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}\,\mathrm{d}y &= 2\int_{-\infty}^{\frac{-x-\mu t}{\sqrt{t}}}\frac{1}{\sqrt{2\pi}}e^{2\mu x-\frac{1}{2}z^2}\,\mathrm{d}z \\ &= 2e^{2\mu x}\Phi\left(\frac{-x-\mu t}{\sqrt{t}}\right)\end{aligned}$$

Finally, we obtain:

$$f_{M_X}(x) = \frac{2}{\sqrt{t}}\phi\left(\frac{x-\mu t}{\sqrt{t}}\right) - 2\mu e^{2\mu x}\Phi\left(\frac{-x-\mu t}{\sqrt{t}}\right)$$

---

[9] Using notations used to state the Girsanov theorem, we have $g(t) = \mu$.

[10] We have:

$$\frac{\mathrm{d}}{\mathrm{d}y}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}} = \mu e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}} + \frac{(2x-y)}{t}e^{\mu y-\frac{1}{2}\mu^2 t-\frac{(2x-y)^2}{2t}}$$

6. We verify that:

$$\begin{aligned}
\frac{\partial F(x)}{\partial x} &= \frac{\partial \Pr\{M_X(t) \leq x\}}{\partial x} \\
&= \frac{1}{\sqrt{t}}\phi\left(\frac{x-\mu t}{\sqrt{t}}\right) + e^{2\mu x}\frac{1}{\sqrt{t}}\phi\left(\frac{-x-\mu t}{\sqrt{t}}\right) \\
&\quad -2\mu e^{2\mu x}\Phi\left(\frac{-x-\mu t}{\sqrt{t}}\right) \\
&= \frac{2}{\sqrt{t}}\phi\left(\frac{x-\mu t}{\sqrt{t}}\right) - 2\mu e^{2\mu x}\Phi\left(\frac{-x-\mu t}{\sqrt{t}}\right) \\
&= f_{M_X}(x)
\end{aligned}$$

Moreover, we have $F(0) = 1$ and $F(\infty) = 1$. We conclude that $F(x)$ is the probability distribution of $M_X(t)$.

A.4.13 Moments of the Cox-Ingersoll-Ross process

1. We recall that:

$$\mathbb{E}[Y(\nu, \zeta)] = \nu + \zeta$$

We deduce that:

$$\begin{aligned}
\mathbb{E}[X(t)] &= \frac{1}{c}\left(\frac{4ab}{\sigma^2} + cx_0e^{-at}\right) \\
&= \frac{(1-e^{-at})\sigma^2}{4a}\left(\frac{4ab}{\sigma^2} + \frac{4a}{(1-e^{-at})\sigma^2}x_0e^{-at}\right) \\
&= x_0e^{-at} + b\left(1-e^{-at}\right)
\end{aligned}$$

2. We recall that:

$$\operatorname{var}(Y(\nu, \zeta)) = 2(\nu + 2\zeta)$$

We deduce that:

$$\begin{aligned}
\operatorname{var}(X(t)) &= \frac{2}{c^2}\left(\frac{4ab}{\sigma^2} + 2cx_0e^{-at}\right) \\
&= \frac{(1-e^{-at})^2\sigma^4}{8a^2}\left(\frac{4ab}{\sigma^2} + \frac{8a}{(1-e^{-at})\sigma^2}x_0e^{-at}\right) \\
&= \left(1-e^{-at}\right)^2\sigma^2\left(\frac{b}{2a} + \frac{1}{(1-e^{-at})a}x_0e^{-at}\right) \\
&= \frac{\sigma^2x_0}{a}\left(e^{-at} - e^{-2at}\right) + \frac{\sigma^2b}{a}\frac{(1-e^{-at})^2}{2}
\end{aligned}$$

3. We recall that:

$$\gamma_1(Y(\nu, \zeta)) = (\nu + 3\zeta)\sqrt{\frac{8}{(\nu + 2\zeta)^3}}$$

We deduce that:

$$\begin{aligned}
\gamma_1(X(t)) &= \left(\frac{4ab}{\sigma^2} + 3cx_0e^{-at}\right)\sqrt{\frac{8}{(4ab\sigma^{-2} + 2cx_0e^{-at})^3}} \\
&= \frac{\sigma(3x_0e^{-at} + b(1-e^{-at}))}{\sqrt{a}}\sqrt{\frac{2(1-e^{-at})}{(2x_0e^{-at} + b(1-e^{-at}))^3}}
\end{aligned}$$

The excess kurtosis coefficients of $Y(\nu,\zeta)$ is equal to:

$$\gamma_2(Y(\nu,\zeta)) = \frac{12(\nu+4\zeta)}{(\nu+2\zeta)^2}$$

It follows that:

$$\begin{aligned}
\gamma_2(X(t)) &= \frac{12\sigma^2\left(4ab+4c\sigma^2 x_0 e^{-at}\right)}{\left(4ab+2c\sigma^2 x_0 e^{-at}\right)^2} \\
&= \frac{3\sigma^2\left(1-e^{-at}\right)\left(4x_0 e^{-at}+b\left(1-e^{-at}\right)\right)}{a\left(x_0 e^{-at}+b\left(1-e^{-at}\right)\right)^2}
\end{aligned}$$

A.4.14 Probability density function of Heston and SABR models

1. The Fokker-Planck equation is:

$$\begin{aligned}
\partial_t U(t,x_1,x_2) &= -\partial_{x_1}\left[\mu x_1 U(t,x_1,x_2)\right] - \partial_{x_2}\left[a(b-x_2)U(t,x_1,x_2)\right] + \\
&\quad \frac{1}{2}\partial^2_{x_1}\left[x_1^2 x_2 U(t,x_1,x_2)\right] + \frac{1}{2}\partial^2_{x_2}\left[\sigma^2 x_2 U(t,x_1,x_2)\right] + \\
&\quad \rho\partial^2_{x_1,x_2}\left[\sigma x_1 x_2 U(t,x_1,x_2)\right]
\end{aligned}$$

The first-order derivatives are:

$$\begin{aligned}
\partial_{x_1}\left[\mu x_1 U(t,x_1,x_2)\right] &= \mu x_1 \partial_{x_1} U(t,x_1,x_2) + \mu U(t,x_1,x_2) \\
\partial_{x_2}\left[a(b-x_2)U(t,x_1,x_2)\right] &= a(b-x_2)\partial_{x_2} U(t,x_1,x_2) - aU(t,x_1,x_2)
\end{aligned}$$

The second-order derivatives are:

$$\begin{aligned}
\partial^2_{x_1}\left[x_1^2 x_2 U(t,x_1,x_2)\right] &= x_1^2 x_2 \partial^2_{x_1} U(t,x_1,x_2) + 4x_1 x_2 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad 2x_2 U(t,x_1,x_2) \\
\partial^2_{x_2}\left[\sigma^2 x_2 U(t,x_1,x_2)\right] &= \sigma^2 x_2 \partial^2_{x_2} U(t,x_1,x_2) + 2\sigma^2 \partial_{x_2} U(t,x_1,x_2)
\end{aligned}$$

and:

$$\begin{aligned}
\partial^2_{x_1,x_2}\left[\sigma x_1 x_2 U(t,x_1,x_2)\right] &= \sigma x_1 x_2 \partial^2_{x_1,x_2} U(t,x_1,x_2) + \\
&\quad \sigma x_1 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad \sigma x_2 \partial_{x_2} U(t,x_1,x_2) + \sigma U(t,x_1,x_2)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\partial_t U(t,x_1,x_2) &= \frac{1}{2}x_1^2 x_2 \partial^2_{x_1} U(t,x_1,x_2) + \frac{1}{2}\sigma^2 x_2 \partial^2_{x_2} U(t,x_1,x_2) + \\
&\quad \rho\sigma x_1 x_2 \partial^2_{x_1,x_2} U(t,x_1,x_2) + \\
&\quad (2x_2+\rho\sigma-\mu)\, x_1 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad \left(\sigma^2+\rho\sigma x_2 - a(b-x_2)\right)\partial_{x_2} U(t,x_1,x_2) + \\
&\quad (a+x_2+\rho\sigma-\mu)\, U(t,x_1,x_2)
\end{aligned}$$

2. The Fokker-Planck equation is:

$$\begin{aligned}
\partial_t U(t,x_1,x_2) &= \frac{1}{2}\partial^2_{x_1}\left[x_1^{2\beta} x_2^2 U(t,x_1,x_2)\right] + \frac{1}{2}\partial^2_{x_2}\left[\nu^2 x_2^2 U(t,x_1,x_2)\right] + \\
&\quad \rho\partial^2_{x_1,x_2}\left[\nu x_1^{\beta} x_2^2 U(t,x_1,x_2)\right]
\end{aligned}$$

The second-order derivatives are:

$$\begin{aligned}
\partial_{x_1}^2 \left[x_1^{2\beta} x_2^2 U(t,x_1,x_2)\right] &= x_1^{2\beta} x_2^2 \partial_{x_1}^2 U(t,x_1,x_2) + \\
&\quad 4\beta x_1^{2\beta-1} x_2^2 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad 2\beta(2\beta-1) x_1^{2\beta-2} x_2^2 U(t,x_1,x_2) \\
\partial_{x_2}^2 \left[\nu^2 x_2^2 U(t,x_1,x_2)\right] &= \nu^2 x_2^2 \partial_{x_2}^2 U(t,x_1,x_2) + \\
&\quad 4\nu^2 x_2 \partial_{x_2} U(t,x_1,x_2) + \\
&\quad 2\nu^2 U(t,x_1,x_2)
\end{aligned}$$

and:

$$\begin{aligned}
\partial_{x_1,x_2}^2 \left[\nu x_1^{\beta} x_2^2 U(t,x_1,x_2)\right] &= \nu x_1^{\beta} x_2^2 \partial_{x_1,x_2}^2 U(t,x_1,x_2) + \\
&\quad 2\nu x_1^{\beta} x_2 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad \beta\nu x_1^{\beta-1} x_2^2 \partial_{x_2} U(t,x_1,x_2) + \\
&\quad 2\beta\nu x_1^{\beta-1} x_2 U(t,x_1,x_2)
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\partial_t U(t,x_1,x_2) &= \frac{1}{2} x_1^{2\beta} x_2^2 \partial_{x_1}^2 U(t,x_1,x_2) + \frac{1}{2}\nu^2 x_2^2 \partial_{x_2}^2 U(t,x_1,x_2) + \\
&\quad \rho\nu x_1^{\beta} x_2^2 \partial_{x_1,x_2}^2 U(t,x_1,x_2) + \\
&\quad 2\left(\beta x_1^{\beta-1} x_2 + \rho\nu\right) x_1^{\beta} x_2 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad \left(2\nu^2 + \rho\beta\nu x_1^{\beta-1} x_2\right) x_2 \partial_{x_2} U(t,x_1,x_2) + \\
&\quad \left(\beta(2\beta-1) x_1^{2\beta-2} x_2^2 + \nu^2 + 2\rho\beta\nu x_1^{\beta-1} x_2\right) U(t,x_1,x_2)
\end{aligned}$$

When β is equal to 1, we obtain:

$$\begin{aligned}
\partial_t U(t,x_1,x_2) &= \frac{1}{2} x_1^2 x_2^2 \partial_{x_1}^2 U(t,x_1,x_2) + \frac{1}{2}\nu^2 x_2^2 \partial_{x_2}^2 U(t,x_1,x_2) + \\
&\quad \rho\nu x_1 x_2^2 \partial_{x_1,x_2}^2 U(t,x_1,x_2) + \\
&\quad 2(x_2 + \rho\nu) x_1 x_2 \partial_{x_1} U(t,x_1,x_2) + \\
&\quad \left(2\nu^2 + \rho\nu x_2\right) x_2 \partial_{x_2} U(t,x_1,x_2) + \\
&\quad \left(x_2^2 + \nu^2 + 2\rho\nu x_2\right) U(t,x_1,x_2)
\end{aligned}$$

3. We have reported the probability density function of Heston and SABR models in Figures A.5 and A.6.

A.4.15 Discrete dynamic programming

1. (a) We have five states $s(k) \in \{1, 1.5, 2, 2.5\}$ and eight control values $c(k) \in \{1,2,3,4,5,6,7,8\}$. We deduce that:

$$\boldsymbol{J} = \begin{pmatrix} 1.141 & 1.108 & 1.075 & 1.038 & 1.000 \\ 1.121 & 1.343 & 1.563 & 1.782 & 2.000 \\ 1.842 & 2.134 & 2.425 & 2.714 & 3.000 \\ 2.759 & 3.071 & 3.382 & 3.691 & 4.000 \end{pmatrix}$$

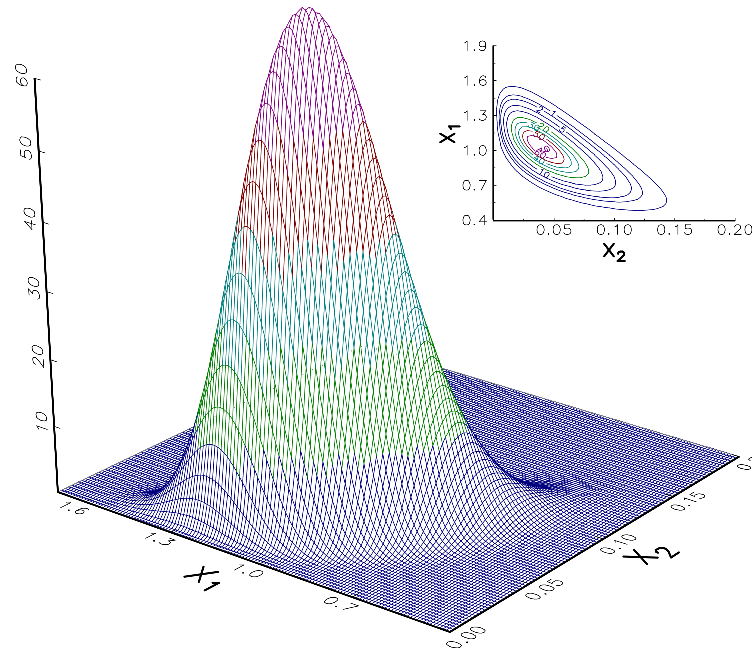

FIGURE A.5: Probability density function of the Heston model

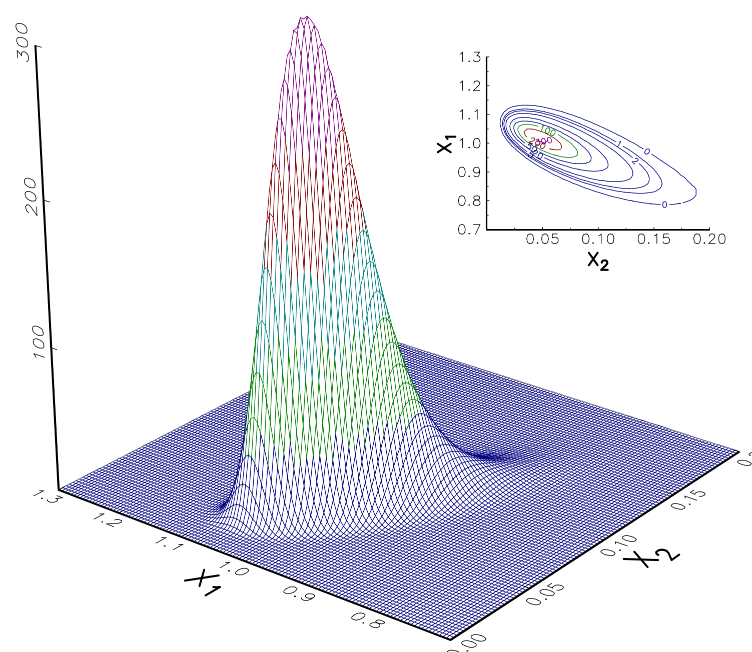

FIGURE A.6: Probability density function of the SABR model

and:

$$\boldsymbol{C}=\begin{pmatrix}1&1&2&2\\2&2&2&2\\2&2&3&3\\3&3&3&3\end{pmatrix}$$

For instance, we have $\mathcal{J}(2,2)=(\boldsymbol{J})_{3,2}=2.134$ because $s_3=2$ and $k=2$.

(b) We deduce that $\mathcal{J}(1,1)=1.141$.

(c) We notice that $c^{\star}(1)=1$ if $s(1)=1$, $c^{\star}(1)=2$ if $s(1)=1.5$ or $s(1)=2$, $c^{\star}(1)=3$ if $s(1)=2.5$, $c^{\star}(2)=1$ if $s(1)=1$, $c^{\star}(2)=2$ if $s(1)=1.5$, etc. When k is small, the objective function is mainly explained by $-\frac{\alpha}{k}(c(k)-s(k))^2$. Therefore, maximizing $f(k,s(k),c(k))$ implies that $c^{\star}(k)\approx s(k)$. This is why $c^{\star}(k)$ cannot be greater than or equal to 4.

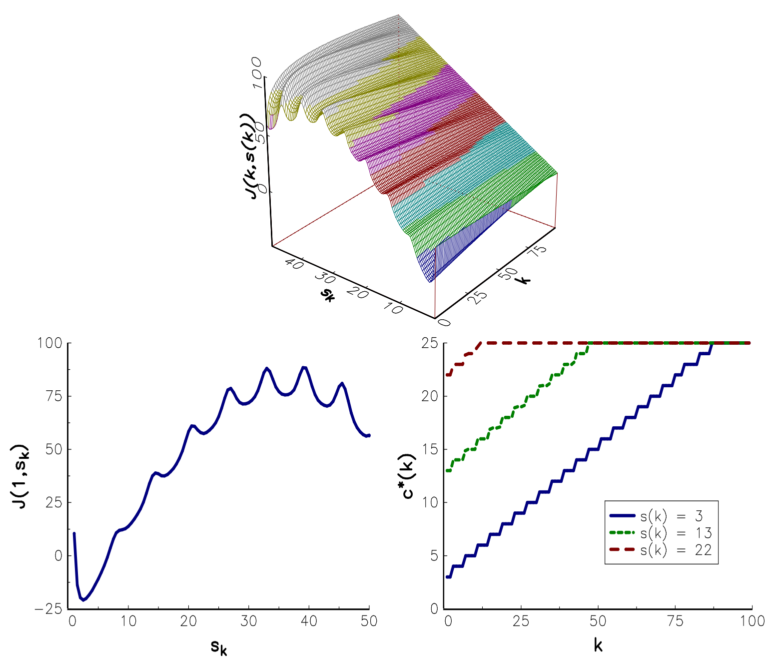

FIGURE A.7: Values taken by $\mathcal{J}(k,s(k))$ and $c^{\star}(k)$

2. (a) We have represented $\mathcal{J}(k,s(k))$ in the first panel in Figure A.7.

(b) In the second panel, we have reported $\mathcal{J}(1,s(k))$. The maximum is reached for $s(k)=39$.

(c) In the third panel, we have reported the optimal control $c^{\star}(k)$ when $s(k)$ is equal to 3, 13 and 22. We notice that $c^{\star}(k)\approx s(k)$ when $k=1$ and $c^{\star}(k)=25$ when $k=100$. The case $k=1$ has been explained in Question 1(c). In the case $k=100$, we have:

$$f(k,s(k),c(k))\approx-\frac{1}{s(k)}\ln s(k)+\beta c(k)+\gamma\sqrt{s(k)}e^{\sin s(k)}$$

Therefore, maximizing $f(k,s(k),c(k))$ implies that $c^{\star}(k)=\max c_j=25$.

A.4.16 Matrix computation

1. (a) We have $A = QTQ^*$ where:

$$Q = \begin{pmatrix} 0.737 & -0.553 & -0.390 \\ 0.526 & 0.107 & 0.844 \\ 0.425 & 0.826 & -0.370 \end{pmatrix}$$

et :

$$T = \begin{pmatrix} 1.761 & 0.000 & 0.000 \\ 0.000 & -0.143 & 0.000 \\ 0.000 & 0.000 & 0.581 \end{pmatrix}$$

(b) We obtain:

$$e^A = \begin{pmatrix} 3.693 & 1.617 & 1.682 \\ 1.617 & 2.895 & 0.820 \\ 1.682 & 0.820 & 1.887 \end{pmatrix}$$

and:

$$\ln A = \begin{pmatrix} -0.371 + 0.961i & 0.512 - 0.185i & 0.989 - 1.435i \\ 0.512 - 0.185i & -0.251 + 0.036i & 0.124 + 0.277i \\ 0.989 - 1.435i & 0.124 + 0.277i & -1.302 + 2.145i \end{pmatrix}$$

(c) We have:

$$\cos A = \frac{e^{iA} + e^{-iA}}{2}$$

and:

$$\sin A = \frac{ie^{-iA} - ie^{iA}}{2}$$

Therefore, we can calculate $\cos A$ and $\sin A$ from the matrix exponential. We obtain:

$$\cos A = \begin{pmatrix} 0.327 & -0.406 & -0.391 \\ -0.406 & 0.554 & -0.216 \\ -0.391 & -0.216 & 0.756 \end{pmatrix}$$

and:

$$\sin A = \begin{pmatrix} 0.573 & 0.209 & 0.451 \\ 0.209 & 0.661 & 0.036 \\ 0.451 & 0.036 & 0.155 \end{pmatrix}$$

We have:

$$\cos^2 A + \sin^2 A = I_3$$

(d) For transcendental functions $f(x)$, we have $f(A) = Qf(T)Q^*$. Using $f(x) = \sqrt{x}$, we obtain:

$$A^{1/2} = \begin{pmatrix} 0.836 + 0.115i & 0.264 - 0.022i & 0.525 - 0.173i \\ 0.264 - 0.022i & 0.910 + 0.004i & 0.059 + 0.033i \\ 0.525 - 0.173i & 0.059 + 0.033i & 0.344 + 0.258i \end{pmatrix}$$

2. The eigenvalues of Σ are -0.00038, 0.00866, 0.01612 and 0.05060. Σ is not a positive semi-definite matrix because one eigenvalue is negative. We have $\Sigma = (A_1 + iA_2)^2$ where:

$$A_1 = \begin{pmatrix} 0.19378 & 0.04539 & 0.01365 & -0.01440 \\ 0.04539 & 0.13513 & -0.03221 & -0.03445 \\ 0.01365 & -0.03221 & 0.02740 & -0.02833 \\ -0.01440 & -0.03445 & -0.02833 & 0.08866 \end{pmatrix}$$

and:

$$A_2 = \begin{pmatrix} 0.00024 & -0.00073 & -0.00184 & -0.00083 \\ -0.00073 & 0.00224 & 0.00563 & 0.00255 \\ -0.00184 & 0.00563 & 0.01417 & 0.00642 \\ -0.00083 & 0.00255 & 0.00642 & 0.00291 \end{pmatrix}$$

We deduce that the nearest covariance matrix is:

$$\tilde{\Sigma} = A_1^2 = \begin{pmatrix} 0.04000 & 0.01499 & 0.00196 & -0.00602 \\ 0.01499 & 0.02254 & -0.00364 & -0.00745 \\ 0.00196 & -0.00364 & 0.00278 & -0.00237 \\ -0.00602 & -0.00745 & -0.00237 & 0.01006 \end{pmatrix}$$

3. We obtain $\rho(B) = C_5(-25\%)$. This is the lower bound of constant correlation matrix. More generally, we have:

$$\rho(C_n(r)) = C_n(r^\star)$$

where $r^\star = \max(r, -1/(n-1))$. If $r < -1/(n-1)$, the nearest correlation matrix is then the lower bound.

4. We obtain:

$$\rho(C) = \begin{pmatrix} 1.0000 & & & & \\ 0.6933 & 1.0000 & & & \\ 0.6147 & 0.4571 & 1.0000 & & \\ 0.2920 & 0.7853 & 0.0636 & 1.0000 & \\ 0.7376 & 0.2025 & 0.7876 & -0.0901 & 1.0000 \end{pmatrix}$$