Course 2024–2025 in Sustainable Finance
Lecture 2. ESG Scoring

Thierry Roncalli*

*Amundi Asset Management[1]

*University of Paris-Saclay

November 2024

[1] The opinions expressed in this presentation are those of the authors and are not meant to represent the opinions or official positions of Amundi Asset Management.

Agenda

ESG data

Several issues:

- E: climate change mitigation, climate change adaptation, preservation of biodiversity, pollution prevention, circular economy
- S: inequality, inclusiveness, labor relations, investment in human capital and communities, human rights
- G: management structure, employee relations, executive remuneration

⇒ requires a lot of alternative data

Sovereign ESG data

Sovereign ESG framework

- World Bank
- Data may be download at the following webpage:

  `https://datatopics.worldbank.org/esg/framework.html`
- E: 27 variables
- S: 22 variables
- G: 18 variables

Sovereign ESG data

Table 1: The World Bank database of sovereign ESG indicators

Environmental

- Emissions & pollution (5)
- Natural capital endowment and management (6)
- Energy use & security (7)
- Environment/ climate risk & resilience (6)
- Food security (3)

Social

- Education & skills (3)
- Employment (3)
- Demography (3)
- Poverty & inequality (4)
- Health & nutrition (5)
- Access to services (4)

Governance

- Human rights (2)
- Government effectiveness (2)
- Stability & rule of law (4)
- Economic environment (3)
- Gender (4)
- Innovation (3)

Sovereign ESG data

Table 2: Indicators of the environmental pillar (World Bank database)

- **Emissions & pollution** (1) CO2 emissions (metric tons per capita); (2) GHG net emissions/removals by LUCF (Mt of CO2 equivalent); (3) Methane emissions (metric tons of CO2 equivalent per capita); (4) Nitrous oxide emissions (metric tons of CO2 equivalent per capita); (5) PM2.5 air pollution, mean annual exposure (micrograms per cubic meter);
- **Natural capital endowment & management**: (1) Adjusted savings: natural resources depletion (% of GNI); (2) Adjusted savings: net forest depletion (% of GNI); (3) Annual freshwater withdrawals, total (% of internal resources); (4) Forest area (% of land area); (5) Mammal species, threatened; (6) Terrestrial and marine protected areas (% of total territorial area);
- **Energy use & security**: (1) Electricity production from coal sources (% of total); (2) Energy imports, net (% of energy use); (3) Energy intensity level of primary energy (MJ/$2011 PPP GDP); (4) Energy use (kg of oil equivalent per capita); (5) Fossil fuel energy consumption (% of total); (6) Renewable electricity output (% of total electricity output); (7) Renewable energy consumption (% of total final energy consumption);
- **Environment/climate risk & resilience**: (1) Cooling degree days (projected change in number of degree Celsius); (2) Droughts, floods, extreme temperatures (% of population, average 1990-2009); (3) Heat Index 35 (projected change in days); (4) Maximum 5-day rainfall, 25-year return level (projected change in mm); (5) Mean drought index (projected change, unitless); (6) Population density (people per sq. km of land area)
- **Food security**: (1) Agricultural land (% of land area); (2) Agriculture, forestry, and fishing, value added (% of GDP); (3) Food production index (2004-2006 = 100);

Source: `https://datatopics.worldbank.org/esg/framework.html`.

Sovereign ESG data

Table 3: Indicators of the social pillar (World Bank database)

- **Education & skills**: (1) Government expenditure on education, total (% of government expenditure); (2) Literacy rate, adult total (% of people ages 15 and above); (3) School enrollment, primary (% gross);
- **Employment**: (1) Children in employment, total (% of children ages 7-14); (2) Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate); (3) Unemployment, total (% of total labor force) (modeled ILO estimate);
- **Demography**: (1) Fertility rate, total (births per woman); (2) Life expectancy at birth, total (years); (3) Population ages 65 and above (% of total population);
- **Poverty & inequality**: (1) Annualized average growth rate in per capita real survey mean consumption or income, total population (%); (2) Gini index (World Bank estimate); (3) Income share held by lowest 20%; (4) Poverty headcount ratio at national poverty lines (% of population);
- **Health & nutrition**: (1) Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions (% of total); (2) Hospital beds (per 1,000 people); (3) Mortality rate, under-5 (per 1,000 live births); (4) Prevalence of overweight (% of adults); (5) Prevalence of undernourishment (% of population);
- **Access to services**: (1) Access to clean fuels and technologies for cooking (% of population); (2) Access to electricity (% of population); (3) People using safely managed drinking water services (% of population); (4) People using safely managed sanitation services (% of population);

Source: `https://datatopics.worldbank.org/esg/framework.html`.

Sovereign ESG data

Table 4: Indicators of the governance pillar (World Bank database)

- **Human rights**: (1) Strength of legal rights index (0 = weak to 12 = strong); (2) Voice and accountability (estimate);
- **Government effectiveness**: (1) Government effectiveness (estimate); (2) Regulatory quality (estimate);
- **Stability & rule of law**: (1) Control of corruption (estimate); (2) Net migration; (3) Political stability and absence of violence/terrorism (estimate); (4) Rule of law (estimate)
- **Economic environment**: (1) Ease of doing business index (1 = most business-friendly regulations); (2) GDP growth (annual %); (3) Individuals using the internet (% of population);
- **Gender**: (1) Proportion of seats held by women in national parliaments (%); (2) Ratio of female to male labor force participation rate (%) (modeled ILO estimate); (3) School enrollment, primary and secondary (gross), gender parity index (GPI); (4) Unmet need for contraception (% of married women ages 15-49);
- **Innovation**: (1) Patent applications, residents; (2) Research and development expenditure (% of GDP); (3) Scientific and technical journal articles;

Source: `https://datatopics.worldbank.org/esg/framework.html`.

Where to find the data?

- National accounts statistics collected by OECD, United Nations Statistics Division (UNSD), etc.
- Internal departments and specialized databases of the World Bank: World Bank Open Data, Business Enabling Environment (BEE), Climate Change Knowledge Portal (CCKP), Global Electrification Database (GEP), etc.
- International organizations: Emission Database for Global Atmospheric Research (EDGAR), Food and Agriculture Organization FAO, International Energy Agency (IEA), International Labour Organization (ILO), World Health Organization (WHO), etc.
- NGOs: Climate Watch, etc.;
- Academic resources: International disasters database (EM-DAT) of the Centre for Research on the Epidemiology of Disasters (Université Catholique de Louvain), etc.

Other frameworks

The most known are FTSE (Beyond Ratings), Moody's (Vigeo-Eiris), MSCI, Sustainalytics, RepRisk and Verisk Mapplecroft.

⇒ The average cross-correlation between data providers is equal to 85% for the ESG score, 42% for the environmental score, 85% for the social score and 71% for the governance score

Bias towards richest countries

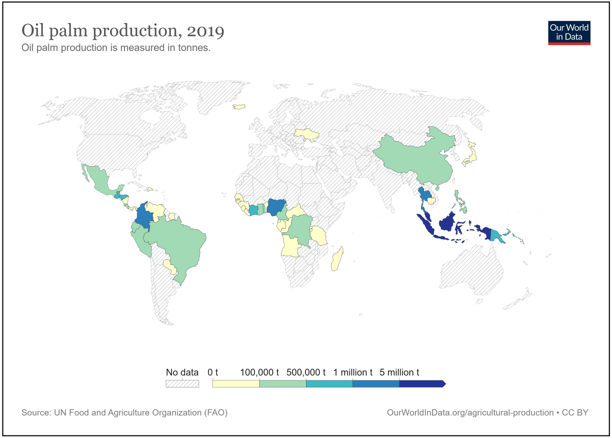
Table 5: Correlation of ESG scores with countryâĂŹs national income (GNI per capita)

Factor	ESG	E	S	G
ISS	68%	7%	86%	77%
FTSE (Beyond Ratings)	91%	74%	88%	84%
MSCI	84%	10%	90%	77%
RepRisk	78%	79%	75%	37%
RobecoSAM	89%	82%	85%	85%
Sustainalytics	95%	83%	94%	93%
V.E	60%	23%	79%	39%
Total	81%	51%	85%	70%

Source: Gratcheva *et al.* (2020).

The mushrooming growth of data

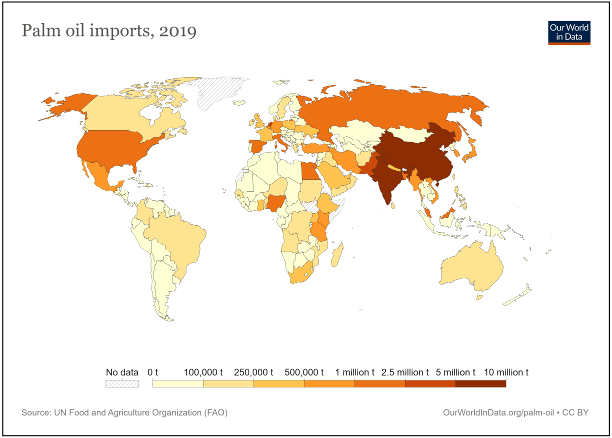
Figure 1: Palm oil production (2019)



Source: Our World in Data, `https://ourworldindata.org/palm-oil`.

The mushrooming growth of data

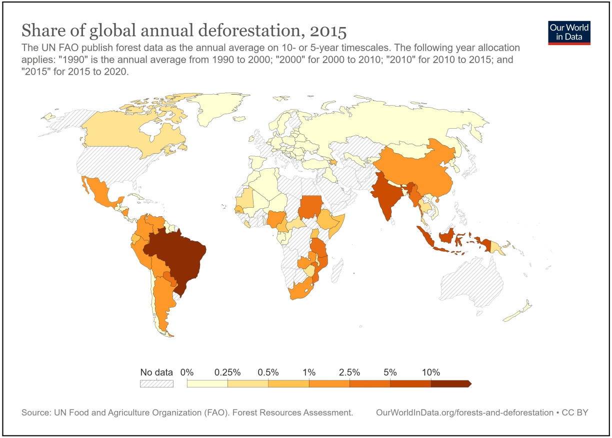
Figure 2: Palm oil imports (2019)



Source: Our World in Data, `https://ourworldindata.org/palm-oil`.

The mushrooming growth of data

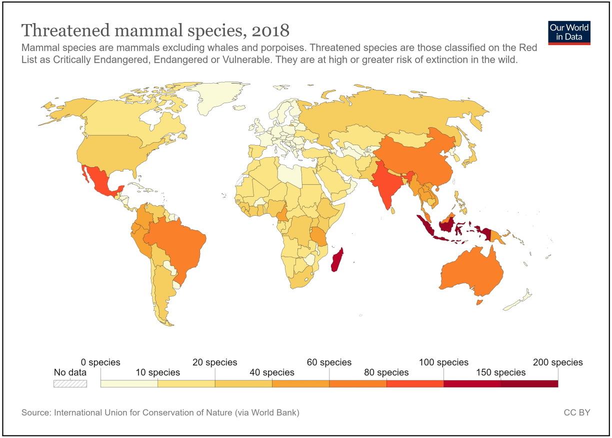
Figure 3: Share of global annual deforestation (2015)



Source: Our World in Data, `https://ourworldindata.org/deforestation`.

The mushrooming growth of data

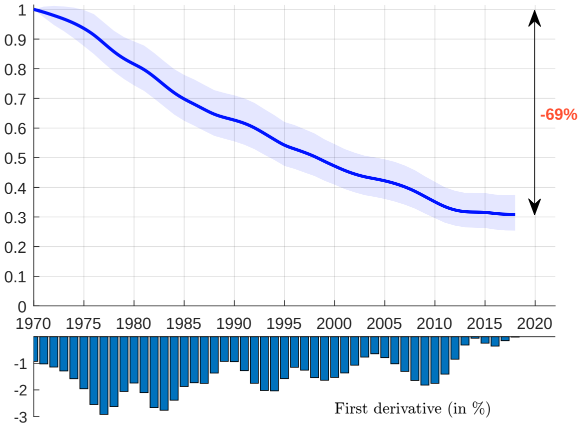
Figure 4: Threatened mammal species (2018)



Source: Our World in Data, `https://ourworldindata.org/biodiversity`.

An example with the biodiversity risk

Figure 5: Global living planet index



Source: `https://livingplanetindex.org/latest_results` & Author's calculation.

An example with the biodiversity risk

Some databases:

- the Red List Index (RLI)
- World Database on Protected Areas (WDPA)
- Integrated Biodiversity Assessment Tool (IBAT)
- Exploring Natural Capital Opportunities, Risks and Exposure (ENCORE)
- Etc.

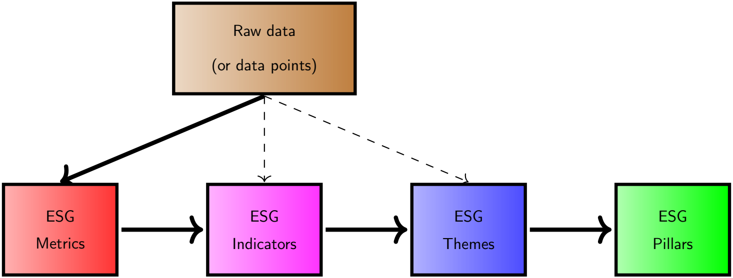
Corporate ESG data

Data sources:

1. Corporate publications (self-reporting)
    1. Annual reports
    2. Corporate sustainability reports
2. Financial and regulatory filings (standardized reporting)
    1. Mandatory reports (SFDR, CSRD, EUTR, etc.)
    2. Non-mandatory frameworks (PRI, TCFD, CDP, etc.)
3. News and other media
4. NGO reports and websites
5. Company assessment and due diligence questionnaire (DDQ)
6. Internal models

Corporate ESG data

Figure 6: From raw data to ESG pillars



Corporate ESG data

Table 6: An example of ESG criteria (corporate issuers)

Environmental	Social	Governance
• Carbon emissions	• Employment conditions	• Board independence
• Energy use	• Community involvement	• Corporate behaviour
• Pollution	• Gender equality	• Audit and control
• Waste disposal	• Diversity	• Executive compensation
• Water use	• Stakeholder opposition	• Shareholder' rights
• Renewable energy	• Access to medicine	• CSR strategy
• Green cars*		
• Green financing*		

$^{(\star)}$means a specific criterion related to one or several sectors (Green cars $\Rightarrow$ Automobiles, Green financing $\Rightarrow$ Financials)

Corporate ESG data

Some examples:

- Bloomberg rates 11 800 public companies. They use more than 120 ESG indicators and 2 000+ data points.
- ISS ESG rates about 10 000 issuers. They use more than 800 indicators and applies approximately 100 indicators per company.
- FTSE Russell rates about 7 200 securities. They use more than 300 indicators and 14 themes.
- MSCI rates 10 000 companies (14 000 issuers including subsidiaries) and 680 000 securities globally. They use 10 themes, 1000+ data points, 80 exposure metrics and 250+ management metrics.
- Refinitiv rates 12 000 public and private companies. They consider 10 themes. These themes are built using 186 metrics and 630+ data points.
- S&P Dow Jones Indices uses between 16 to 27 criteria scores, a questionnaire and 1 000 data points.
- Sustainalytics rates more than 16 300 companies. They consider 20 material ESG issues, based on 350+ indicators.

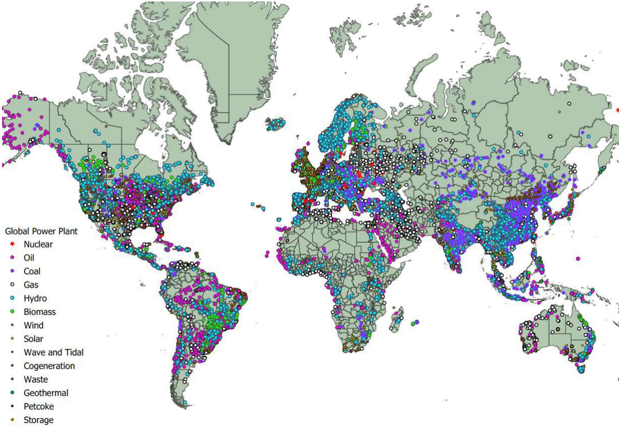
The race for alternative data

- Internet traffic, browsing activity, web scraping, product reviews, social media and sentiment data;
- Satellite imagery, geotracking data, sensor data (*e.g.*, temperature, humidity, pressure, chemical levels);
- Supply-chain data.

- Controversies $\Rightarrow$ NLP (RepRisk, daily basis: 500 000+ documents, 100 000+ sources, 23 languages)
- Geospatial data $\Rightarrow$ Physical risk

The race for alternative data

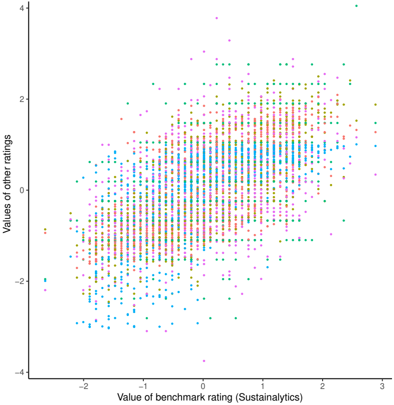
Figure 7: Geolocation of world power plants by energy source



Source: Global Power Database version 1.3 (June 2021).

The divergence of corporate ESG ratings

Figure 8: ESG rating disagreement



Source: Berg *et al.* (2022).

The divergence of corporate ESG ratings

Berg *et al.* (2022) identify three sources of divergence:

1. "**Measurement** *divergence refers to situation where rating agencies measure the same indicator using different ESG metrics* (56%)
2. **Scope** *divergence refers to situation where ratings are based on different set of ESG indicators* (38%)
3. **Weight** *divergence emerges when rating agencies take different views on the relative importance of ESG indicators*" (6%)

The divergence of corporate ESG ratings

Table 7: Rank correlation among ESG ratings

	MSCI	Refinitiv	S&P Global	
MSCI	100%			
Refinitiv	43%	100%		
S&P Global	45%	69%	100%	
Sustainalytics	53%	64%	69%	100%

Source: Billio *et al.* (2021).

One-level tree structure

- $X_1, \ldots, X_m$ are m features
- The score is linear:

$$\mathcal{S} = \sum_{j=1}^{m} \omega_j X_j$$

- ω_j is the weight of the j^{th} metric

One-level tree structure

- $X_1, \ldots, X_m$ are m features
- The score is linear:

$$\mathcal{S} = \sum_{j=1}^{m} \omega_j X_j$$

- ω_j is the weight of the j^{th} metric

One-level tree structure

The Altman Z score is equal to:

$$Z = 1.2 \cdot X_1 + 1.4 \cdot X_2 + 3.3 \cdot X_3 + 0.6 \cdot X_4 + 1.0 \cdot X_5$$

where the variables X_j represent the following financial ratios:

X_j	Ratio
X_1	Working capital / Total assets
X_2	Retained earnings / Total assets
X_3	Earnings before interest and tax / Total assets
X_4	Market value of equity / Total liabilities
X_5	Sales / Total assets

$$Z_i \Rightarrow Z_i^\star = \left(Z_i - m_z\right) / \sigma_z \Rightarrow \text{Decision rule}$$

Two-level tree structure

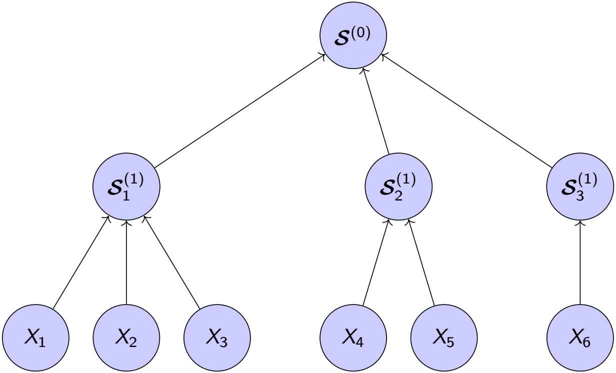
The intermediary scores are equal to:

$$\mathcal{S}_k^{(1)} = \sum_{j=1}^{m} \omega_{j,k}^{(1)} X_j$$

whereas the expression of the final score is:

$$\mathcal{S} := \mathcal{S}_1^{(0)} = \sum_{k=1}^{m_{(1)}} \omega_k^{(0)} \mathcal{S}_k^{(1)}$$

Figure 9: A two-level non-overlapping tree



- Level 1: $\omega_{1,1}^{(1)} = 50\%$; $\omega_{2,1}^{(1)} = 25\%$; $\omega_{3,1}^{(1)} = 25\%$; $\omega_{4,2}^{(1)} = 50\%$; $\omega_{5,2}^{(1)} = 50\%$; $\omega_{6,3}^{(1)} = 100\%$;
- Level 0: $\omega_1^{(0)} = \omega_2^{(0)} = \omega_3^{(0)} = 33.33\%$;

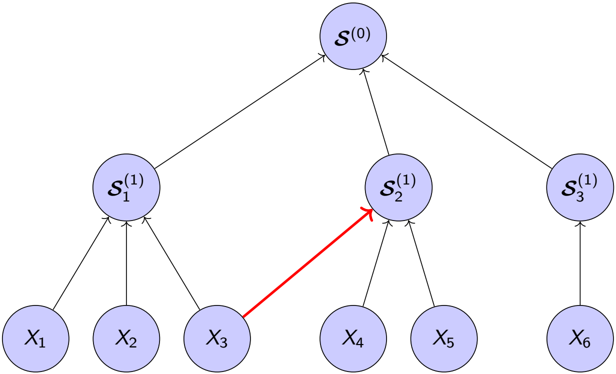
Two-level tree structure

$$\left\{\begin{array}{l}\mathcal{S}_1^{(1)} = 0.5 \cdot X_1 + 0.25 \cdot X_2 + 0.25 \cdot X_3 \\ \mathcal{S}_2^{(1)} = 0.5 \cdot X_4 + 0.5 \cdot X_5 \\ \mathcal{S}_3^{(1)} = X_6\end{array}\right.$$

$$\mathcal{S} = \frac{\mathcal{S}_1^{(1)} + \mathcal{S}_2^{(1)} + \mathcal{S}_3^{(1)}}{3}$$

Two-level tree structure

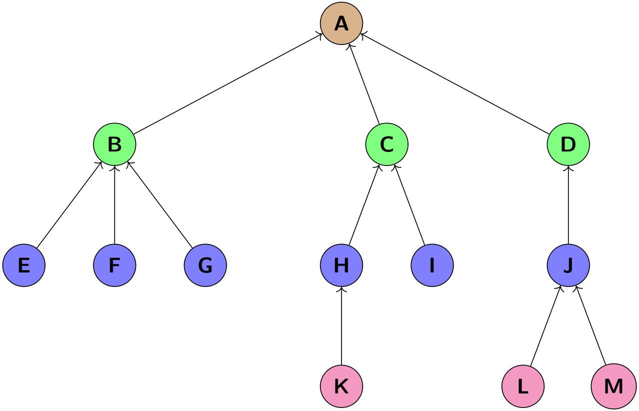
Figure 10: A two-level overlapping ~~tree~~ graph



- Level 1: $\omega_{1,1}^{(1)} = 50\%$; $\omega_{2,1}^{(1)} = 25\%$; $\omega_{3,1}^{(1)} = 25\%$; $\omega_{3,2}^{(1)} = 25\%$; $\omega_{4,2}^{(1)} = 25\%$; $\omega_{5,2}^{(1)} = 50\%$; $\omega_{6,3}^{(1)} = 100\%$;
- Level 0: $\omega_1^{(0)} = \omega_2^{(0)} = \omega_3^{(0)} = 33.33\%$;

Tree and graph theory

Figure 11: Tree data structure



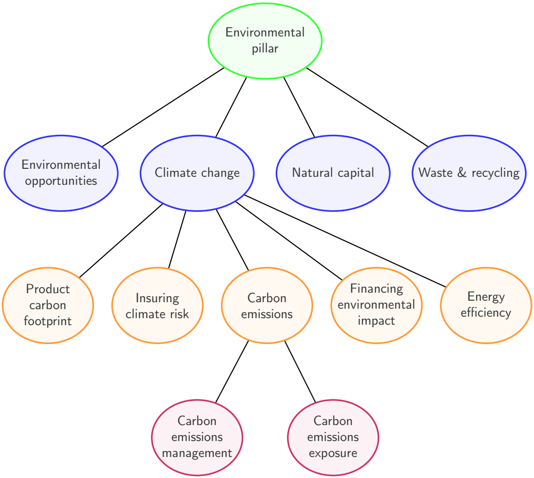
Tree and graph theory

- L is the number of levels
- We have $\mathcal{S}_j^{(L)} = X_j$
- The value of the k^{th} node at level ℓ is given by:

$$\mathcal{S}_k^{(\ell)} = \sum_{j=1}^{m_{(\ell+1)}} \omega_{j,k}^{(\ell)} \mathcal{S}_j^{(\ell+1)}$$

An example of ESG scoring tree

Figure 12: An example of ESG scoring tree (MSCI methodology)



Source: MSCI (2020)

Score normalization

Let $\omega_{(\ell)}$ be the $m_{(\ell+1)} \times m_{(\ell)}$ matrix, whose elements are $\omega_{j,k}^{(\ell)}$ for $j = 1, \ldots, m_{(\ell+1)}$ and $k = 1, \ldots, m_{(\ell)}$

The final score is equal to:

$$\mathcal{S} = \omega^\top X$$

where:

$$\omega = \omega_{(L-1)} \cdots \omega_{(1)} \omega_{(0)}$$

Score normalization

If $X \sim \mathbf{F}$, we obtain:

$$\begin{aligned}
\mathbf{G}(s) &= \Pr\{\mathcal{S} \leq s\} \\
&= \Pr\{\omega^\top X \leq s\} \\
&= \int \cdots \int \mathbb{1}\{\omega^\top x \leq s\}\, \mathrm{d}\mathbf{F}(x) \\
&= \int \cdots \int \mathbb{1}\left\{\sum_{j=1}^m \omega_j x_j \leq s\right\} \mathrm{d}\mathbf{F}(x_1, \ldots, x_m) \\
&= \int \cdots \int \mathbb{1}\left\{\sum_{j=1}^m \omega_j x_j \leq s\right\} \mathrm{d}\mathbf{C}(\mathbf{F}_1(x_1), \ldots, \mathbf{F}_m(x_m))
\end{aligned}$$

Therefore, the distribution $\mathbf{G}$ depends on the copula function $\mathbf{C}$ and the marginals $(\mathbf{F}_1, \ldots, \mathbf{F}_m)$ of $\mathbf{F}$

$$\mathbf{F}_1 \equiv \mathbf{F}_1 \equiv \ldots \equiv \mathbf{F}_m \Rightarrow \mathbf{G} \equiv \mathbf{F}_1 ?$$

Score normalization

In the independent case, we obtain a a convolution probability distribution:

$$\mathbf{G}(s) = \int \cdots \int \mathbb{1} \left\{ \sum_{j=1}^{m} \omega_j x_j \leq s \right\} \prod_{j=1}^{m} \mathrm{d}\mathbf{F}_j(x_j)$$

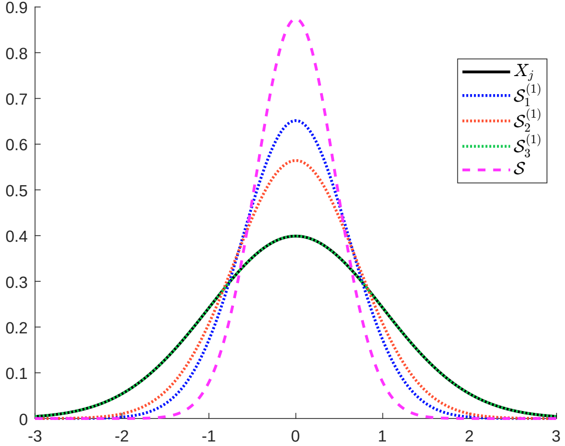
If $X_j \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right)$, we have $\omega_j X_j \sim \mathcal{N}\left(\omega_j \mu_j, \omega_j^2 \sigma_j^2\right)$. We deduce that:

$$\boldsymbol{\mathcal{S}} \sim \mathcal{N}\left(\sum_{j=1}^{m} \omega_j \mu_j, \sum_{j=1}^{m} \omega_j^2 \sigma_j^2\right) \equiv \mathcal{N}\left(\omega^\top \mu, \omega^\top \Sigma \omega\right)$$

where $\mu = (\mu_1, \ldots, \mu_m)$ and $\Sigma = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right)$.

Score normalization

Figure 13: Probability distribution of the scores based on the previous tree



Score normalization

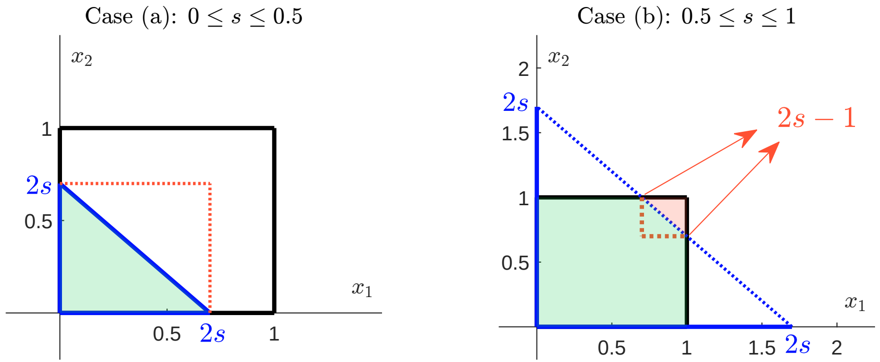
Exercise

We assume that $X_1 \sim \mathcal{U}_{[0,1]}$ and $X_2 \sim \mathcal{U}_{[0,1]}$ are two independent random variables. We consider the score $\boldsymbol{\mathcal{S}}$ defined as:

$$\boldsymbol{\mathcal{S}} = \frac{X_1 + X_2}{2}$$

Score normalization

Figure 14: Geometric interpretation of the probability mass function



Score normalization

We deduce that:

$$\Pr\left\{\mathcal{S} \leq s\right\} = \begin{cases} \dfrac{1}{2}\left(2s\right)^2 = 2s^2 & \text{if} \quad 0 \leq s \leq \dfrac{1}{2} \\ 1 - \dfrac{1}{2}\left(2 - 2s\right)^2 = -1 + 4s - 2s^2 & \text{if} \quad \dfrac{1}{2} \leq s \leq 1 \end{cases}$$

The density function is then:

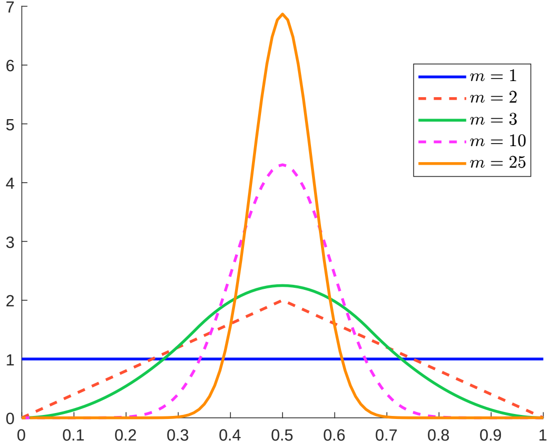
$$g\left(s\right) = \begin{cases} 4s & \text{if} \quad 0 \leq s \leq \dfrac{1}{2} \\ 4 - 4s & \text{if} \quad \dfrac{1}{2} \leq s \leq 1 \end{cases}$$

In the general case, we have:

$$\mathcal{S} = \frac{X_1 + X_2 + \cdots + X_m}{m} \sim \mathfrak{B}ates\left(m\right)$$

Score normalization

Figure 15: Probability density function of $\mathcal{S}$ (uniform distribution)



Score normalization

Exercise

We assume that $X \sim \mathcal{N}(\mu, \Sigma)$ with $\mu_j = 0$, $\sigma_j = 1$ and $\rho_{j,k} = \rho$ for $j \neq k$. Show that:

$$\mathbb{E}[\boldsymbol{\mathcal{S}}] = 0$$

and

$$\operatorname{var}(\boldsymbol{\mathcal{S}}) = \rho \mathcal{S}^2(w) + (1-\rho)\mathcal{H}(\omega)$$

where $\mathcal{S}(w) = \sum_{j=1}^m \omega_j$ is the sum index and $\mathcal{H}(\omega) = \sum_{j=1}^m \omega_j^2$ is the Herfindahl index. Deduce that:

$$\sigma_{\boldsymbol{\mathcal{S}}} = \sqrt{\rho + (1-\rho)\mathcal{H}(\omega)}$$

Score normalization

How to normalize?

$$\boldsymbol{\mathcal{S}}_k^{(\ell)} = \varphi \left(\sum_{j=1}^{m_{(\ell+1)}} \omega_{j,k}^{(\ell)} \boldsymbol{\mathcal{S}}_j^{(\ell+1)} \right)$$

Score normalization

1. m-score normalization:

$$m_i = \frac{x_i - x^-}{x^+ - x^-}$$

   where $x^- = \min x_i$ and $x^+ = \max x_i$

2. q-score normalization:

$$q_i = \mathbf{H}(x_i)$$

   where $\mathbf{H}$ is the distribution function of X

3. z-score normalization:

$$z_i = \frac{x_i - \mu}{\sigma}$$

   where μ and σ are the mathematical expectation and standard deviation of X

4. b-score normalization:

$$b_i = \mathfrak{B}^{-1}\left(\mathbf{H}(x_i); \alpha, \beta\right)$$

   where $\mathcal{B}(\alpha, \beta)$ is the beta distribution

Score normalization

Probability integral transform (PIT)

If $X \sim \mathbf{H}$ and is continuous, $Y = \mathbf{H}(X)$ is a uniform random variable.

We have $Y \in [0,1]$ and:

$$\begin{aligned}
\Pr\{Y \leq y\} &= \Pr\{\mathbf{H}(X) \leq y\} \\
&= \Pr\left\{X \leq \mathbf{H}^{-1}(y)\right\} \\
&= \mathbf{H}\left(\mathbf{H}^{-1}(y)\right) \\
&= y
\end{aligned}$$

Score normalization

Computing the empirical distribution $\hat{\mathbf{H}}$

- Let $\{x_1, x_2, \ldots, x_n\}$ be the sample
- We have:

$$q_i = \hat{\mathbf{H}}(x_i) = \Pr\{X \leq x_i\} = \frac{\#\{x_j \leq x_i\}}{n_q}$$

- $n_q = n$ or $n_q = n + 1$?

Score normalization

Exercise

What is the normalization shape of this transformation?

$$\mathcal{S} = \frac{2}{1 + e^{-z}} - 1$$

Hint: compute the density function.

Score normalization

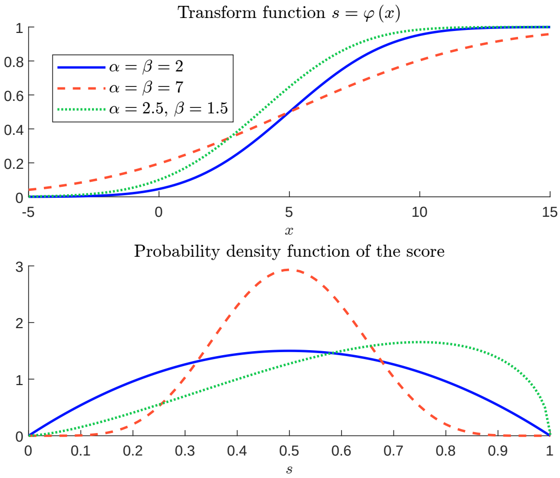
Example

The data are normally distributed with mean $\mu = 5$ and standard deviation $\sigma = 2$. To map these data into a 0/1 score, we consider the following transform:

$$s := \varphi(x) = \mathfrak{B}^{-1}\left(\Phi\left(\frac{x-5}{2}\right); \alpha, \beta\right)$$

Score normalization

Figure 16: Transforming data into $\mathcal{B}$-score



Score normalization

Example

We consider the raw data of 9 companies that belong to the same industry. The first variable measures the carbon intensity of the scope $1+2$ in 2020, while the second variable is the variation of carbon emissions between 2015 and 2020. We would like to create the score $\mathcal{S} \equiv 70\% \cdot X_1 + 30\% \cdot X_2$.

Firm	Carbon intensity in $\mathrm{tCO_2e}$/\$ mn)	Carbon momentum (in %)
1	94.0	−3.0
2	38.6	−5.5
3	30.6	5.6
4	74.4	−1.3
5	97.1	−16.8
6	57.1	−3.5
7	132.4	8.5
8	92.5	−9.1
9	64.9	−4.6

Score normalization

- q-score 0/100
- z-score
- $qz = 100 \cdot \Phi\left(z\right)$
- $zq = \Phi^{-1}\left(\frac{q}{100}\right)$
- $bz = \mathfrak{B}^{-1}\left(\Phi\left(z\right); \alpha, \beta\right)$ where $\alpha = \beta = 2$
- $bz^{\star} = \mathfrak{B}^{-1}\left(\Phi\left(z\right); \alpha, \beta\right)$ where $\alpha = 2.5$ and $\beta = 1.5$.

Score normalization

Table 8: Computation of the score $\mathcal{S} \equiv 70\% \cdot X_1 + 30\% \cdot X_2$ (q-score 0/100 normalization)

#	X_1	q_1	X_2	q_2	s	$\mathcal{S}$	$\mathfrak{R}$
1	94.00	70.00	−3.00	60.00	67.00	80.00	8
2	38.60	20.00	−5.50	30.00	23.00	10.00	1
3	30.60	10.00	5.60	80.00	31.00	20.00	2
4	74.40	50.00	−1.30	70.00	56.00	60.00	6
5	97.10	80.00	−16.80	10.00	59.00	70.00	7
6	57.10	30.00	−3.50	50.00	36.00	30.00	3
7	132.40	90.00	8.50	90.00	90.00	90.00	9
8	92.50	60.00	−9.10	20.00	48.00	50.00	5
9	64.90	40.00	−4.60	40.00	40.00	40.00	4
Mean	75.73	50.00	−3.30	50.00	50.00	50.00	
Std-dev.	31.95	27.39	7.46	27.39	20.60	27.39	

Score normalization

Table 9: Computation of the score $\mathcal{S} \equiv 70\% \cdot X_1 + 30\% \cdot X_2$ (z-score normalization)

#	X_1	z_1	X_2	z_2	s	$\mathcal{S}$	$\mathfrak{R}$
1	94.00	0.572	−3.00	0.040	0.412	0.543	8
2	38.60	−1.162	−5.50	−0.295	−0.902	−1.188	1
3	30.60	−1.413	5.60	1.193	−0.631	−0.831	2
4	74.40	−0.042	−1.30	0.268	0.051	0.067	6
5	97.10	0.669	−16.80	−1.810	−0.075	−0.099	5
6	57.10	−0.583	−3.50	−0.027	−0.416	−0.548	3
7	132.40	1.774	8.50	1.582	1.716	2.261	9
8	92.50	0.525	−9.10	−0.778	0.134	0.177	7
9	64.90	−0.339	−4.60	−0.174	−0.290	−0.382	4
Mean	75.73	0.000	−3.30	0.000	0.000	0.000	
Std-dev.	31.95	1.000	7.46	1.000	0.759	1.000	

Score normalization

Table 10: Comparison of the different scoring methods

#	q		z		qz		zq		bz		$bz^{\star}$	
	$\mathcal{S}$	$\mathfrak{R}$	$\mathcal{S}$	$\mathfrak{R}$	$\mathcal{S}$	$\mathfrak{R}$	$\mathcal{S}$	$\mathfrak{R}$	$\mathcal{S}$	$\mathfrak{R}$	$\mathcal{S}$	$\mathfrak{R}$
1	80.00	8	0.54	8	76.27	8	0.84	8	0.66	8	0.81	8
2	10.00	1	−1.19	1	9.19	1	−1.28	1	0.20	1	0.30	1
3	20.00	2	−0.83	2	21.37	2	−0.84	2	0.29	2	0.38	2
4	60.00	6	0.07	6	54.13	5	0.25	6	0.52	6	0.70	6
5	70.00	7	−0.10	5	56.65	6	0.52	7	0.51	5	0.64	5
6	30.00	3	−0.55	3	24.42	3	−0.52	3	0.34	3	0.50	3
7	90.00	9	2.26	9	98.04	9	1.28	9	0.93	9	0.96	9
8	50.00	5	0.18	7	60.39	7	0.00	5	0.56	7	0.72	7
9	40.00	4	−0.38	4	30.96	4	−0.25	4	0.39	4	0.56	4
Mean	50.00		0.00		47.94		0.00		0.49		0.62	
Std-dev.	27.39		1.00		28.79		0.82		0.22		0.21	

An example with the CEO pay ratio

The CEO pay ratio is calculated by dividing the CEO's compensation by the pay of the median employee. It is one of the key metrics for the G pillar. It has been imposed by the Dodd-Frank Act, which requires that publicly traded companies disclose:

1. the median total annual compensation of all employees other than the CEO;
2. the ratio of the CEO's annual total compensation to that of the median employee;
3. the wage ratio of the CEO to the median employee.

$\Rightarrow$ the average S&P 500 company's CEO-to-worker pay ratio was 324-to-1 in 2021 (AFL-CIO)

An example with the CEO pay ratio

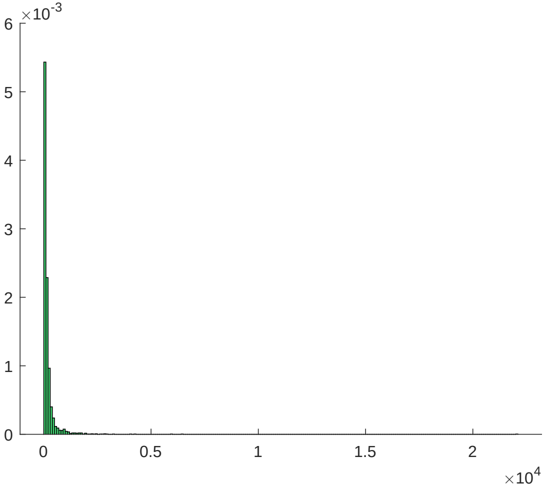
Table 11: Examples of CEO pay ratio (June 2021)

Company name	P	R	Company name	P	R
Abercrombie & Fitch	1 954	4,293	Netflix	202 931	190
McDonald's	9 291	1,939	BlackRock	133 644	182
Coca-Cola	11 285	1,657	Pfizer	98 972	181
Gap	6 177	1,558	Goldman Sachs	138 854	178
Alphabet	258 708	1,085	MSCI	55 857	165
Walmart	22 484	983	Verisk Analytics	77 055	117
Estee Lauder	30 733	697	Facebook	247 883	94
Ralph Lauren	21 358	570	Invesco	125 282	92
NIKE	25 386	550	Boeing	158 869	90
Citigroup	52 988	482	Citrix Systems	181 769	80
PepsiCo	45 896	368	Harley-Davidson	187 157	59
Microsoft	172 512	249	Amazon.com	28 848	58
Apple	57 596	201	Berkshire Hathaway	65 740	6

Source: `https://aflcio.org` (June 2021)

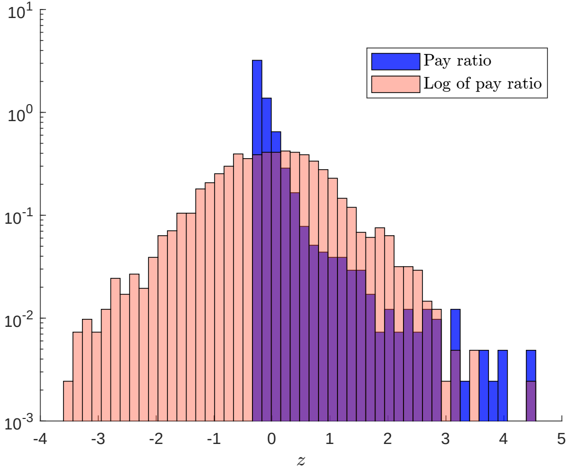
An example with the CEO pay ratio

Figure 17: Histogram of the CEO pay ratio



An example with the CEO pay ratio

Figure 18: Histogram of z-score applied to the CEO pay ratio



An example with the CEO pay ratio

What is the solution? Give the transform function $y = \varphi(x)$.

Hint: use the beta distribution.

Other statistical methods

Unsupervised learning

- Clustering (K-means, hierarchical clustering)
- Dimension reduction (PCA)

Other statistical methods

Supervised learning

- Discriminant analysis (LDA, QDA)
- Binary choice models (logistic regression, probit model)
- Regression models (OLS, lasso)

$\Rightarrow$ Advanced learning models (k-NN, neural networks and support vector machines) are not relevant in the case of ESG scoring

We need to define the response variable Y

Other statistical methods

Example with credit scoring models

- Let $\mathcal{S}_i(t)$ be the credit score of individual i at time t
- We have:

$$Y_i(t) = \mathbb{1}\{\tau_i \leq t + \delta\} = \mathbb{1}\{D_i(t + \delta) = 1\}$$

  where τ_i and D_i are the default time and the default indicator function, and δ is the time horizon (e.g., one year)
- The calibration problem of the credit scoring model is:

$$\Pr\{Y_i(t) = 0\} = f(\mathcal{S}_i(t))$$

  where f is an increasing function

Application to ESG scoring models

- Let $\mathcal{S}_i(t)$ be the ESG score of company i at time t
- Endogenous response variable:
  (a) Best-in-class oriented scoring system:
  $$Y_i(t) = \mathbb{1}\{\mathcal{S}_i(t+h) \geq s^\star\}$$
  where $s^\star$ is the best-in-class threshold
  (b) Worst-in-class oriented scoring system: $Y_i(t) = \mathbb{1}\{\mathcal{S}_i(t+h) \leq s^\star\}$ where $s^\star$ is the worst-in-class threshold
- Exogenous response variable
  (c) Binary response:
  $$Y_i(t) = \mathbf{1}\{\mathcal{C}_i(t+h) \geq 0\}$$
  where $\mathcal{C}_i(t)$ is the controversy index
  d Continuous response:
  $$Y_i(t) = \mathcal{C}_i(t+h)$$
- The calibration problem of the ESG scoring model is $\Pr\{Y_i(t) = 0\} = f(\mathcal{S}_i(t))$ or $Y_i(t) = f(\mathcal{S}_i(t))$ where the function f is increasing for case (a) and decreasing for cases (b), (c) and (d)

Performance evaluation criteria

- ESG scoring and rating
  - Shannon entropy
  - Confusion matrix
  - Binary classification ratios (TPR, FNR, TNR, FPR, PPV, ACC, F_1)
- ESG scoring
  - Performance, selection and discriminant curves
  - ROC curve
  - Gini coefficient

Shannon entropy

- Let (X, Y) be a random vector where $p_{i,j} = \Pr\{X = x_i, Y = y_j\}$, $p_i = \Pr\{X = x_i\}$ and $p_j = \Pr\{Y = y_j\}$
- The Shannon entropy of the discrete random variable X is given by:

$$\mathcal{I}(X) = -\sum_{i=1}^{n} p_i \ln p_i$$

- We have the property $0 \leq \mathcal{I}(X) \leq \ln n$
- The entropy is equal to zero if there is a state i such that $p_i = 1$ and is equal to $\ln n$ in the case of the uniform distribution $(p_i = 1/n)$
- The Shannon entropy is a measure of the average information of the system
- The lower the Shannon entropy, the more informative the system

Shannon entropy

- For a random vector (X, Y), we have:

$$\mathcal{I}(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{n} p_{i,j} \ln p_{i,j}$$

- We deduce that the conditional information of Y given X is equal to:

$$\mathcal{I}(Y \mid X) = \mathbb{E}[\mathcal{I}(Y \mid X = x)] = -\sum_{i=1}^{n} \sum_{j=1}^{n} p_{i,j} \ln \frac{p_{i,j}}{p_i} = \mathcal{I}(X, Y) - \mathcal{I}($$

- if X and Y are independent, we have $\mathcal{I}(Y \mid X) = \mathcal{I}(Y)$ and $\mathcal{I}(X, Y) = \mathcal{I}(Y) + \mathcal{I}(X)$
- if X and Y are perfectly dependent, we have $\mathcal{I}(Y \mid X) = 0$ and $\mathcal{I}(X, Y) = \mathcal{I}(X)$
- The amount of information obtained about one random variable, by the other random variable, is measured by the mutual information:

$$\mathcal{I}(X \cap Y) = \mathcal{I}(Y) + \mathcal{I}(X) - \mathcal{I}(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{i,j} \ln \frac{p_{i,j}}{p_i p_j}$$

Shannon entropy

Figure 19: Examples of Shannon entropy calculation

$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$
$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$
$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$
$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$
$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$
$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$	$^1/_{36}$

$$\mathcal{I}(X) = \mathcal{I}(Y) = 1.792$$
$$\mathcal{I}(X, Y) = 3.584$$
$$\mathcal{I}(X \cap Y) = 0$$

$^1/_6$					
	$^1/_6$				
		$^1/_6$			
			$^1/_6$		
				$^1/_6$	
					$^1/_6$

$$\mathcal{I}(X) = \mathcal{I}(Y) = 1.792$$
$$\mathcal{I}(X, Y) = 1.792$$
$$\mathcal{I}(X \cap Y) = 1.792$$

Shannon entropy

Figure 20: Examples of Shannon entropy calculation

$1/24$	$1/24$				
$1/24$	$1/24$	$1/24$	$1/48$		
	$1/24$	$1/6$	$1/24$	$1/48$	
	$1/48$	$1/24$	$1/6$	$1/24$	
		$1/48$	$1/24$	$1/24$	$1/24$
				$1/24$	$1/24$

$$\mathcal{I}(X) = \mathcal{I}(Y) = 1.683$$
$$\mathcal{I}(X, Y) = 2.774$$
$$\mathcal{I}(X \cap Y) = 0.593$$

					$1/12$
$1/8$			$1/8$		
	$1/24$				
$5/24$		$1/24$			
$3/24$				$1/24$	
$3/24$	$1/24$	$1/24$			

$$\mathcal{I}(X) = 1.658,\ \mathcal{I}(Y) = 1.328$$
$$\mathcal{I}(X, Y) = 2.236$$
$$\mathcal{I}(X \cap Y) = 0.750$$

Shannon entropy

Application to (credit/ESG) scoring

- Let $\mathcal{S}$ and Y be the score and the control variable
- For instance, Y is a binary random variable that may indicate a bad ESG risk ($Y = 0$) or a good ESG risk ($Y = 1$)
- Y may also correspond to classes defined by some quantiles
- We can measure the information of the system $(\mathcal{S}, Y)$
- We can compare two scores $\mathcal{S}_1$ and $\mathcal{S}_2$ by using the statistical measures $\mathcal{I}(\mathcal{S}_1 \cap Y)$ and $\mathcal{I}(\mathcal{S}_2 \cap Y)$

Shannon entropy

Application to (credit/ESG) scoring

$\mathcal{S}_1$: $\mathcal{I}(\mathcal{S}_1) = 1.767$, $\mathcal{I}(Y) = 1.609$, $\mathcal{I}(\mathcal{S}_1, Y) = 2.614$, $\mathcal{I}(\mathcal{S}_1 \cap Y) = 0.763$

$\mathcal{S}_2$: $\mathcal{I}(\mathcal{S}_2) = 1.771$, $\mathcal{I}(Y) = 1.609$, $\mathcal{I}(\mathcal{S}_2, Y) = 2.745$, $\mathcal{I}(\mathcal{S}_2 \cap Y) = 0.636$

$$\mathcal{S}_1 \succ \mathcal{S}_2$$

	y_1	y_2	y_3	y_4	y_5
s_1	10	9			
s_2	7	9			
s_3	3		7	2	
s_4		2	10	4	5
s_5				10	2
s_6			3	4	13

	y_1	y_2	y_3	y_4	y_5
s_1	7	10			
s_2	10	8			
s_3			5	4	3
s_4	3		10	6	4
s_5	2			5	8
s_6			5	5	5

Graphical methods

- The control variable Y can takes two values
  1. $Y = 0$ corresponds to a bad risk (or bad signal)
  2. $Y = 1$ corresponds to a good risk (or good signal)
- We assume that the probability $\Pr\{Y = 1 \mid \boldsymbol{\mathcal{S}} \geq s\}$ is increasing with respect to the level $s \in [0, 1]$, which corresponds to the rate of acceptance
- The decision rule is the following:
  - if the score of the observation is above the threshold s, the observation is selected
  - if the score of the observation is below the threshold s, the observation is not selected
- If s is equal to one, we select no observation. If s is equal to zero, we select all the observations

Performance curve

The performance curve is the parametric function $y = \mathcal{P}(x)$ defined by:

$$\begin{cases} x(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s\} \\ y(s) = \dfrac{\Pr\{Y = 0 \mid \boldsymbol{\mathcal{S}} \geq s\}}{\Pr\{Y = 0\}} \end{cases}$$

where $x(s)$ corresponds to the proportion of selected observations and $y(s)$ corresponds to the ratio between the proportion of selected bad risks and the proportion of bad risks in the population

- The score is efficient if the ratio is below one
- If $y(s) > 1$, the score selects more bad risks than those we can find in the population
- If $y(s) = 1$, the score is random and the performance is equal to zero. In this case, the selected population is representative of the total population

Selection curve

The selection curve is the parametric curve $y = \mathcal{S}(x)$ defined by:

$$\begin{cases} x(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s\} \\ y(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s \mid Y = 0\} \end{cases}$$

where $y(s)$ corresponds to the ratio of observations that are wrongly selected

- We would like that the curve $y = \mathcal{S}(x)$ is located below the bisecting line $y = x$ in order to verify that $\Pr\{\boldsymbol{\mathcal{S}} \geq s \mid Y = 0\} < \Pr\{\boldsymbol{\mathcal{S}} \geq s\}$
- The performance and selection curves are related as follows:

$$\mathcal{S}(x) = x\mathcal{P}(x)$$

Discriminant curve

The discriminant curve is the parametric curve $y = \mathcal{D}(x)$ defined by:

$$\mathcal{D}(x) = g_1\left(g_0^{-1}(x)\right)$$

where:

$$g_y(s) = \Pr\left\{\boldsymbol{\mathcal{S}} \geq s \mid Y = y\right\}$$

- It represents the proportion of good risks in the selected population with respect to the proportion of bad risks in the selected population
- The score is said to be discriminant if the curve $y = \mathcal{D}(x)$ is located above the bisecting line $y = x$

Properties

- The previous parametric curves do not depend on the probability distribution of the score $\boldsymbol{\mathcal{S}}$, but only on the ranking of the observations
- They are then invariant if we apply an increasing function to the score
- We have the following properties:
  1. The performance curve (respectively, the selection curve) is located below the line $y = 1$ (respectively, the bisecting line $y = x$) if and only if $\operatorname{cov}\left(f\left(Y\right), g\left(\boldsymbol{\mathcal{S}}\right)\right) \geq 0$ for any increasing functions f and g
  2. The performance curve is increasing if and only if:
     $$\operatorname{cov}\left(f\left(Y\right), g\left(\boldsymbol{\mathcal{S}}\right) \mid \boldsymbol{\mathcal{S}} \geq s\right) \geq 0$$
     for any increasing functions f and g, and any threshold level s
  3. The selection curve is convex if and only if $\mathbb{E}\left[f\left(Y\right) \mid \boldsymbol{\mathcal{S}} = s\right]$ is increasing with respect to the threshold level s for any increasing function f

Properties

- A score is perfect or optimal if there is a threshold level $s^\star$ such that $\Pr\{Y = 1 \mid \boldsymbol{\mathcal{S}} \geq s^\star\} = 1$ and $\Pr\{Y = 0 \mid \boldsymbol{\mathcal{S}} < s^\star\} = 1$. It separates the population between good and bad risks. Graphically, the selection curve of a perfect score is equal to:

$$y = \mathbb{1}\{x > \Pr\{Y = 1\}\} \cdot \left(1 + \frac{x - 1}{\Pr\{Y = 0\}}\right)$$

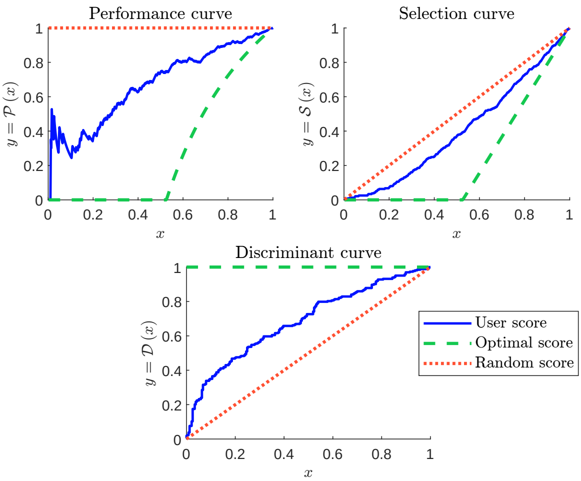
- Using the relationship $\mathcal{S}(x) = x\mathcal{P}(x)$, we deduce that the performance curve of a perfect score is given by:

$$y = \mathbb{1}\{x > \Pr\{Y = 1\}\} \cdot \left(\frac{x - \Pr\{Y = 1\}}{x \cdot \Pr\{Y = 0\}}\right)$$

- For the discriminant curve, a perfect score satisfies $\mathcal{D}(x) = 1$
- When the score is random, we have $\mathcal{S}(x) = \mathcal{D}(x) = x$ and $\mathcal{P}(x) = 1$

Properties

Figure 21: Performance, selection and discriminant curves

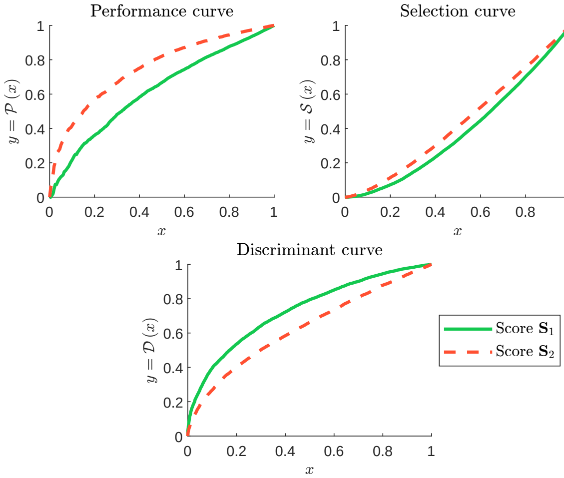


Comparing two scores $\mathcal{S}_1$ and $\mathcal{S}_2$

- The score $\mathcal{S}_1$ is more performing on the population P_1 than the score S_2 on the population P_2 if and only if the performance (or selection) curve of $(\mathcal{S}_1, P_1)$ is below the performance (or selection) curve of $(\mathcal{S}_2, P_2)$
- The score $\mathcal{S}_1$ is more discriminatory on the population P_1 than the score $\mathcal{S}_2$ on the population P_2 if and only if the discriminant curve of $(\mathcal{S}_1, P_1)$ is above the discriminant curve of $(\mathcal{S}_2, P_2)$

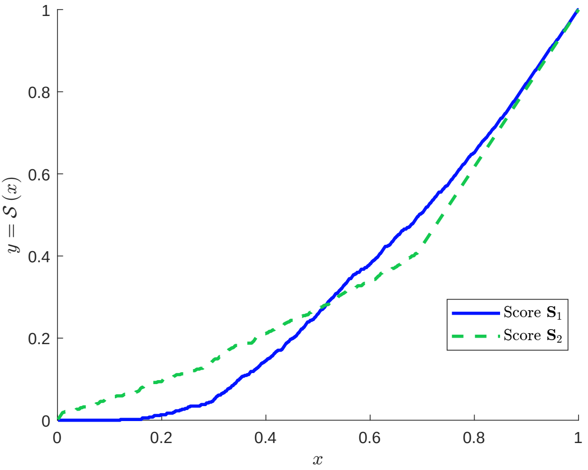
Comparing two scores $\mathcal{S}_1$ and $\mathcal{S}_2$

Figure 22: The score S_1 is better than the score S_2



Comparing two scores $\mathcal{S}_1$ and $\mathcal{S}_2$

Figure 23: Illustration of the partial ordering between two scores



Statistical methods

1. Kolmogorov-Smirnov test
2. Gini coefficient

Kolmogorov-Smirnov test

- We consider the cumulative distribution functions:

$$\mathbf{F}_0(s) = \Pr\{\boldsymbol{\mathcal{S}} \leq s \mid Y = 0\}$$

and:

$$\mathbf{F}_1(s) = \Pr\{\boldsymbol{\mathcal{S}} \leq s \mid Y = 1\}$$

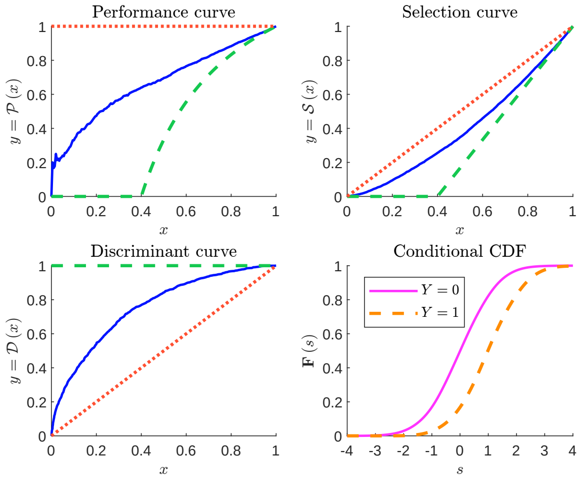
- The score $\boldsymbol{\mathcal{S}}$ is relevant if we have the stochastic dominance order $\mathbf{F}_0 \succ \mathbf{F}_1$
- In this case, the score quality is measured by the Kolmogorov-Smirnov statistic:

$$\mathrm{KS} = \max_s |\mathbf{F}_0(s) - \mathbf{F}_1(s)|$$

- It takes the value 1 if the score is perfect

Kolmogorov-Smirnov test

Figure 24: Comparison of the distributions $\mathbf{F}_0(s)$ and $\mathbf{F}_1(s)$



Gini coefficient

The Lorenz curve $y = \mathcal{L}(x)$ is the parametric curve defined by:

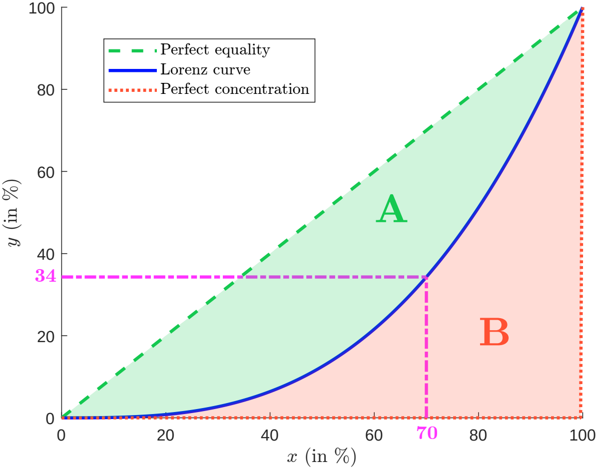
$$\begin{cases} x = \Pr\{X \leq x\} \\ y = \Pr\{Y \leq y \mid X \leq x\} \end{cases}$$

The Lorenz curve has two limit cases

1. If the wealth is perfectly concentrated, one individual holds 100% of the total wealth
2. If the wealth is perfectly allocated between all the individuals, the corresponding Lorenz curve is the bisecting line

Gini coefficient

Figure 25: An example of Lorenz curve



Gini coefficient

We define the Gini coefficient by:

$$\mathcal{G}ini\left(\mathcal{L}\right) = \frac{A}{A+B} = 1 - 2\int_0^1 \mathcal{L}\left(x\right)\,\mathrm{d}x$$

where A is the area between the Lorenz curve and the curve of perfect equality, and B is the area between the curve of perfect concentration and the Lorenz curve

- We have $0 \leq \mathcal{G}ini\left(\mathcal{L}\right) \leq 1$
- The Gini coefficient is equal to zero in the case of perfect equality and one in the case of perfect concentration

Application to scoring

- The selection curve is a Lorenz curve. We recall that $\mathbf{F}(s) = \Pr\{\boldsymbol{\mathcal{S}} \leq s\}$, $\mathbf{F}_0(s) = \Pr\{\boldsymbol{\mathcal{S}} \leq s \mid Y = 0\}$ and $\mathbf{F}_1(s) = \Pr\{\boldsymbol{\mathcal{S}} \leq s \mid Y = 1\}$. The selection curve is defined by the following parametric coordinates:

$$\left\{ \begin{array}{l} x(s) = 1 - \mathbf{F}(s) \\ y(s) = 1 - \mathbf{F}_0(s) \end{array} \right.$$

  The selection curve measures the capacity of the score for not selecting bad risks

- The precision curve is the Lorenz curve that measures the capacity of the score for selecting good risks

$$\left\{ \begin{array}{l} x(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s\} = 1 - \mathbf{F}(s) \\ y(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s \mid Y = 1\} = 1 - \mathbf{F}_1(s) \end{array} \right.$$

Application to scoring

- The receiver operating characteristic (ROC) curve is defined by:

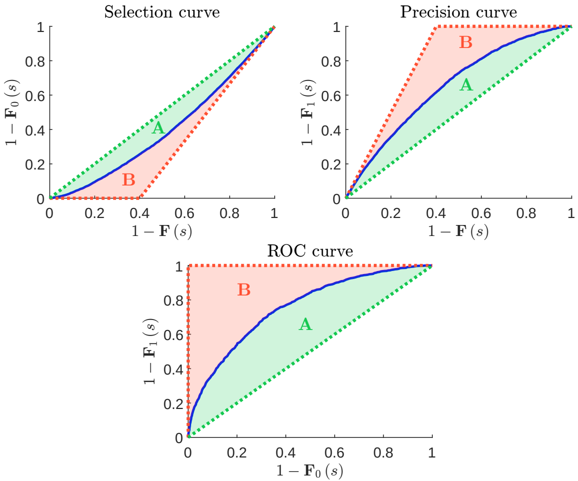
$$\left\{ \begin{array}{l} x(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s \mid Y = 0\} = 1 - \mathbf{F}_0(s) \\ y(s) = \Pr\{\boldsymbol{\mathcal{S}} \geq s \mid Y = 1\} = 1 - \mathbf{F}_1(s) \end{array} \right.$$

- The AUC measure, which corresponds to the area under the ROC curve, give the same information than the Gini coefficient since they are related by the equation:

$$\mathcal{G}ini(\mathrm{ROC}) = 2 \times \mathrm{AUC}(\mathrm{ROC}) - 1$$

Application to scoring

Figure 26: Selection, precision and ROC curves



Application to scoring

Choice of the optimal cut-off

The choice of the optimal cut-off $s^\star$ depends on the objective function. For instance, we can calibrate $s^\star$ in order to achieve a minimum universe size of ESG assets. We can also fix a given selection rate. From a statistical point of view, we must distinguish the construction of the scoring model and the decision rule. In statistical learning, we generally consider three datasets: the training set, the validation set and the test set. The training set is used for calibrating the model and its parameters whereas the validation set helps to avoid overfitting. But the decision rule is based on the test set.

Confusion matrix

- A confusion matrix is a special case of contingency matrix
- Each row of the matrix represents the frequency in a predicted class while each column represents the frequency in an actual class
- Using the test set, it takes the following form:

	$Y = 0$	$Y = 1$
$\mathcal{S} < s$	$n_{0,0}$	$n_{0,1}$
$\mathcal{S} \geq s$	$n_{1,0}$	$n_{1,1}$
	$n_0 = n_{0,0} + n_{1,0}$	$n_1 = n_{0,1} + n_{1,1}$

where $n_{i,j}$ represents the number of observations of the cell (i, j)

Confusion matrix

	$Y=0$	$Y=1$
$\mathcal{S}<s$	It is rejected and it is a bad risk (true negative)	It is rejected, but it is a good risk (false negative)
$\mathcal{S}\geq s$	It is accepted, but it is a bad risk (false positive)	It is accepted and it is a good risk (true positive)
	(negative)	(positive)

The cells $(\mathcal{S}<s, Y=0)$ and $(\mathcal{S}\geq s, Y=1)$ correspond to observations that are well-classified: true negative (TN) and true positive (TP). The cells $(\mathcal{S}\geq s, Y=0)$ and $(\mathcal{S}<s, Y=1)$ correspond to two types of errors:

1. A false positive (FP) can induce a future loss, because the risk can materialize: this is a type I error
2. A false negative (FN) potentially corresponds to a an opportunity cost: this is a type II error

Classification ratios

$$
\begin{array}{ll}
\text{True Positive Rate} & \mathrm{TPR} = \dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \\
\text{False Negative Rate} & \mathrm{FNR} = \dfrac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1-\mathrm{TPR} \\
\text{True Negative Rate} & \mathrm{TNR} = \dfrac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \\
\text{False Positive Rate} & \mathrm{FPR} = \dfrac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1-\mathrm{TNR} \\
\text{Positive Predictive Value} & \mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}
\end{array}
$$

- The true positive rate (TPR) is also known as the sensitivity or the recall. It measures the proportion of real good risks that are correctly predicted good risk
- The precision or the positive predictive value (PPV) measures the proportion of predicted good risks that are correctly real good risk

Classification ratios

1. The accuracy considers the classification of both negatives and positives:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}}$$

2. The F_1 score is the harmonic mean of precision and sensitivity:

$$F_1 = \frac{2}{1/\mathrm{precision} + 1/\mathrm{sensitivity}} = \frac{2 \cdot \mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}}$$

3. The ϕ coefficient or the Matthews correlation coefficient (MCC) is a measure of association between $\mathcal{S}$ and Y:

$$\phi = \mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \cdot (\mathrm{TP} + \mathrm{FN}) \cdot (\mathrm{TN} + \mathrm{FP}) \cdot (\mathrm{TN} + \mathrm{FN})}}$$

$\mathcal{S}$ and Y are positively associated if most of the observations fall along the diagonal cells

Backtesting of unsupervised scoring systems

Static analysis

Let $\mathcal{S}_i(t)$ be the ESG score of company i at time t. The endogenous response variable can be defined as follows:

Scoring system	Risk class	$Y_i(t)$
Best-in-class oriented	Good risk	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \geq s^\star\}$
Worst-in-class oriented	Bad risk	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \leq s^\star\}$

where $s^\star$ is the best-in-class/worst-in-class threshold to determine. $Y_i(t)$ is endogenous because it depends on the future value of the score. Here, the backtesting procedure can be seen as a stability test of the ESG scoring system. An alternative is to use an exogenous response variable based on controversies. For example, to predict bad risks, we can use the binary response $Y_i(t) = \mathbb{1}\{\mathcal{C}_i(t+\delta) \geq 0\}$ where $\mathcal{C}_i(t)$ is the controversy index.

Backtesting of unsupervised scoring systems

Dynamic analysis

- We consider the past momentum $\mathcal{M}_i(t,h) = \boldsymbol{\mathcal{S}}_i(t) - \boldsymbol{\mathcal{S}}_i(t-h)$ where h is typically the year
- The response variable is based on the future momentum $\boldsymbol{\mathcal{S}}_i(t+\delta) - \boldsymbol{\mathcal{S}}_i(t)$

Illustration using an ESG scoring system

We consider four risk classes:

Risk class	Definition	$Y_i(t)$
Worst-in-class	$\mathcal{S}_i(t) \leq \hat{\mathbf{F}}^{-1}(20\%)$	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \leq s^\star\}$
Bad risk	$\mathcal{S}_i(t) \leq \bar{\mathcal{S}}$	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \leq s^\star\}$
Good risk	$\mathcal{S}_i(t) \geq \bar{\mathcal{S}}$	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \geq s^\star\}$
Best-in-class	$\mathcal{S}_i(t) \geq \hat{\mathbf{F}}^{-1}(80\%)$	$\mathbb{1}\{\mathcal{S}_i(t+\delta) \geq s^\star\}$

where $\hat{\mathbf{F}}$ is the empirical distribution of the score and $\bar{\mathcal{S}}$ is the average of scores

Illustration using an ESG scoring system

Table 12: Optimal cut-off $s^\star$ (MSCI World)

Risk class	$\delta = 3$ months				$\delta = 12$ months			
	ACC	F_1	ϕ	$\mathcal{I}(\boldsymbol{\mathcal{S}} \cap Y)$	ACC	F_1	ϕ	$\mathcal{I}(\boldsymbol{\mathcal{S}} \cap Y)$
Worst-in-class	-0.91	-0.61	-0.68	-0.58	-0.96	-0.58	-0.67	-0.54
Bad risk	-0.01	0.18	0.02	0.02	0.01	0.24	0.04	0.05
Good risk	-0.02	-0.18	0.01	0.02	-0.01	-0.20	0.03	0.04
Best-in-class	1.05	0.79	0.85	0.77	1.08	0.76	0.83	0.72

Illustration using an ESG scoring system

Table 13: Optimal cut-off $s^{\star}$ (MSCI EM)

Risk class	$\delta = 3$ months				$\delta = 12$ months			
	ACC	F_1	ϕ	$\mathcal{I}(\boldsymbol{\mathcal{S}} \cap Y)$	ACC	F_1	ϕ	$\mathcal{I}(\boldsymbol{\mathcal{S}} \cap Y)$
Worst-in-class	-1.87	-1.19	-1.29	-1.15	-2.00	-1.17	-1.30	-1.12
Bad risk	0.13	0.23	-0.03	-0.10	0.16	0.28	-0.05	-0.14
Good risk	0.13	-0.15	-0.03	-0.14	0.16	-0.22	-0.05	-0.24
Best-in-class	0.48	0.14	0.22	0.11	0.56	0.13	0.24	0.11

Illustration using an ESG scoring system

Remark

Theoretically, the optimal cut-off is $s^\star = \Phi^{-1}(20\%) = -0.8416$ for the worst-in-class category, $s^\star = \mathbb{E}[\mathcal{N}(0,1)] = 0$ for the bad-risk and good-risk categories and $s^\star = \Phi^{-1}(80\%) = 0.8416$ for the best-in-class category, because the backtesting procedure concerns z-scores

Illustration using an ESG scoring system

Figure 27: Backtesting of ESG scores (worst-in-class & bad risk, MSCI World)

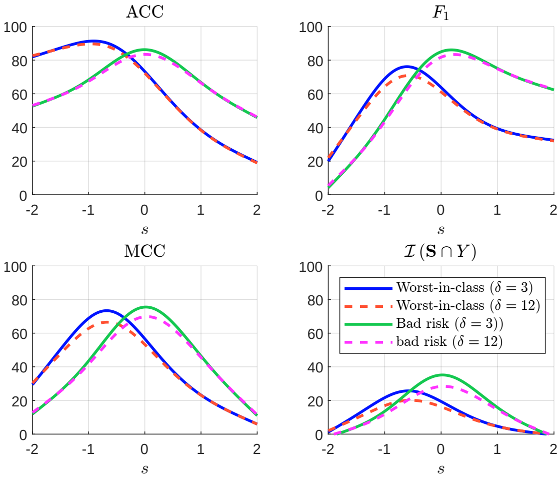


Illustration using an ESG scoring system

Figure 28: Backtesting of ESG scores (best-in-class & good risk, MSCI World)

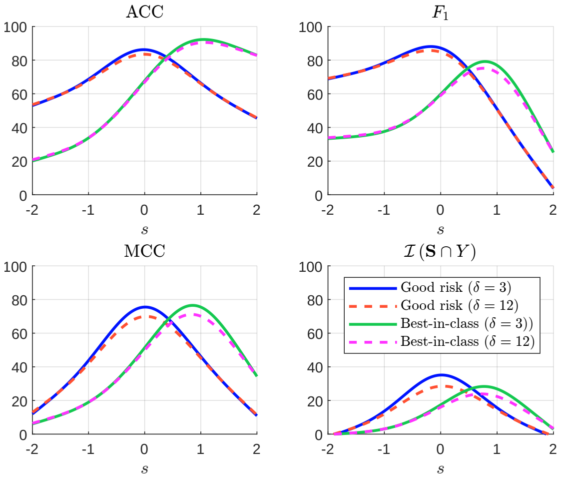


Illustration using an ESG scoring system

Figure 29: Backtesting of ESG scores (worst-in-class & bad risk, MSCI EM)

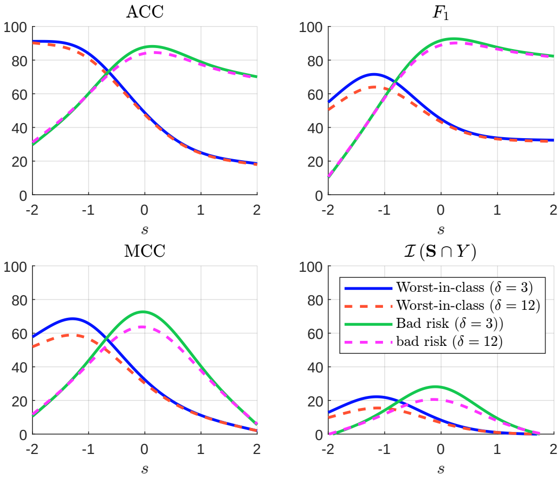
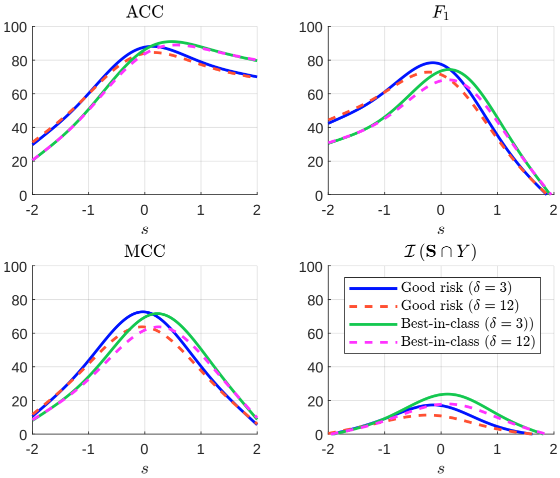


Illustration using an ESG scoring system

Figure 30: Backtesting of ESG scores (best-in-class & good risk, MSCI EM)



Definition

Table 14: Credit rating system of S&P, Moody's and Fitch

	Prime Maximum Safety	High Grade High Quality			Upper Medium Grade		
S&P/Fitch	AAA	AA+	AA	AA−	A+	A	A−
Moody's	Aaa	Aa1	Aa2	Aa3	A1	A2	A3

	Lower Medium Grade			Non Investment Grade Speculative		
S&P/Fitch	BBB+	BBB	BBB−	BB+	BB	BB−
Moody's	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3

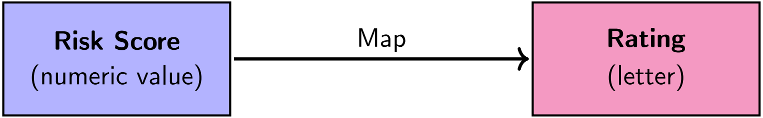
	Highly Speculative			Substantial Risk	In Poor Standing		Extremely Speculative
S&P/Fitch	B+	B	B−	CCC+	CCC	CCC−	CC
Moody's	B1	B2	B3	Caa1	Caa2	Caa3	Ca

Definition

- Amundi: A (high), B,... to G (low) — 7-grade scale
- FTSE Russell: 0 (low), 1,... to 5 (high) — 6-grade scale
- ISS ESG: 1 (high), 2,... to 10 (low) — 10-grade scale
- MSCI: AAA (high), AA,... to CCC (low) — 7-grade scale
- Refinitiv: A+ (high), A, A-, B+,... to D- (low) — 12-grade scale
- RepRisk: AAA (high), AA,... to D (low) — 8-grade scale
- Sustainanalytics: 1 (low), 2,... to 5 (high) — 5-grade scale

ESG rating process

Figure 31: From ESG score to ESG rating



Two-step approach:

1. Specification of the map function:

$$\begin{array}{rccl} \mathcal{M}ap: & \Omega_{\boldsymbol{\mathcal{S}}} & \longrightarrow & \Omega_{\mathcal{R}} \\ & \boldsymbol{\mathcal{S}} & \longmapsto & \mathcal{R} = \mathcal{M}ap\left(\boldsymbol{\mathcal{S}}\right) \end{array}$$

   where $\Omega_{\boldsymbol{\mathcal{S}}}$ is the support of ESG scores, $\Omega_{\mathcal{R}}$ is the ordered state space of ESG ratings and $\mathcal{R}$ is the ESG rating

2. Validation (and the possible *forcing*) of the rating by the analyst

ESG rating process

Example with the MSCI ESG rating system

- $\Omega_{\mathcal{S}} = [0, 10]$
- $\Omega_{\mathcal{R}} = \{\text{CCC}, \text{B}, \text{BB}, \text{BBB}, \text{A}, \text{AA}, \text{AAA}\}$
- The map function is defined as

$$\mathcal{M}ap\,(s) = \begin{cases} \text{CCC} & \text{if } \mathcal{S} \in [0, 10/7] & (0 - 1.429) \\ \text{B} & \text{if } \mathcal{S} \in [10/7, 20/7] & (1.429 - 2.857) \\ \text{BB} & \text{if } \mathcal{S} \in [20/7, 30/7] & (2.857 - 4.286) \\ \text{BBB} & \text{if } \mathcal{S} \in [30/7, 40/7] & (4.286 - 5.714) \\ \text{A} & \text{if } \mathcal{S} \in [40/7, 50/7] & (5.714 - 7.143) \\ \text{AA} & \text{if } \mathcal{S} \in [50/7, 60/7] & (7.143 - 8.571) \\ \text{AAA} & \text{if } \mathcal{S} \in [60/7, 10] & (8.571 - 10) \end{cases}$$

ESG rating process

- The map function is an increasing piecewise function
- $\boldsymbol{\mathcal{S}} \sim \mathbf{F}$ and $\boldsymbol{\mathcal{S}} \in (s^-, s^+)$
- $\left\{s_0^\star = s^-, s_1^\star, \ldots, s_{K-1}^\star, s_K^\star = s^+\right\}$ are the knots of the piecewise function
- $\Omega_{\mathcal{R}} = \{R_1, \ldots, R_K\}$ is the set of grades

$\Rightarrow$ The frequency distribution of the ratings is given by:

$$\begin{aligned} p_k &= \Pr\{\mathcal{R} = R_k\} \\ &= \Pr\left\{s_{k-1}^\star \leq \boldsymbol{\mathcal{S}} < s_k^\star\right\} \\ &= \mathbf{F}(s_k^\star) - \mathbf{F}\left(s_{k-1}^\star\right) \end{aligned}$$

ESG rating process

If we would like to build a rating system with pre-defined frequencies $(p_1, \ldots, p_K)$, we have to solve the following equation:

$$\mathbf{F}\left(s_k^\star\right) - \mathbf{F}\left(s_{k-1}^\star\right) = p_k$$

We deduce that:

$$\begin{aligned} \mathbf{F}\left(s_k^\star\right) &= p_k + \mathbf{F}\left(s_{k-1}^\star\right) \\ &= p_k + p_{k-1} + \mathbf{F}\left(s_{k-2}^\star\right) \\ &= \left(\sum_{j=1}^{k} p_j\right) + \mathbf{F}\left(s_0^\star\right) \end{aligned}$$

and:

$$s_k^\star = \mathbf{F}^{-1}\left(\sum_{j=1}^{k} p_j\right)$$

ESG rating process

Exercise

- We assume that $\mathcal{S} \sim \mathcal{U}_{[a,b]}$
- Show that $p_k = K^{-1}$ If the rating system consists in K equally-sized intervals
- Show that the knots of the map function are equal to:

$$s_k^\star = a + (b - a) \left(\sum_{j=1}^{k} p_j \right)$$

when we impose pre-defined frequencies $(p_1, \ldots, p_K)$

- If we consider a 0/100 uniform score and $\Omega_{\mathcal{R}} \times \mathbb{P} = (\mathrm{CCC}, 5\%), (\mathrm{B}, 10\%), (\mathrm{BB}, 20\%), (\mathrm{BBB}, 30\%), (\mathrm{A}, 20\%), (\mathrm{AA}, 10\%), (\mathrm{AAA}, 5\%)$, show that $s_{\mathrm{CCC}}^\star = 5$, $s_{\mathrm{B}}^\star = 15$, $s_{\mathrm{BB}}^\star = 35$, $s_{\mathrm{BBB}}^\star = 65$, $s_{\mathrm{A}}^\star = 85$ and $s_{\mathrm{AA}}^\star = 95$

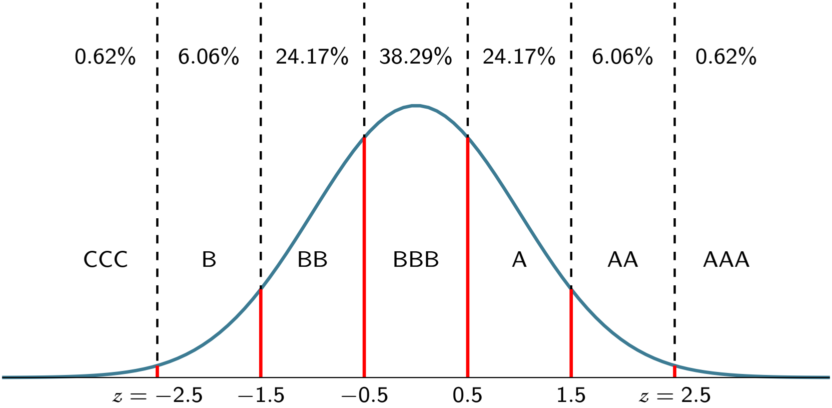
ESG rating process

For a z-score system ($\mathcal{S} \sim \mathcal{N}(0,1)$), we obtain:

$$p_k = \Phi\left(s_k^{\star}\right) - \Phi\left(s_{k-1}^{\star}\right)$$

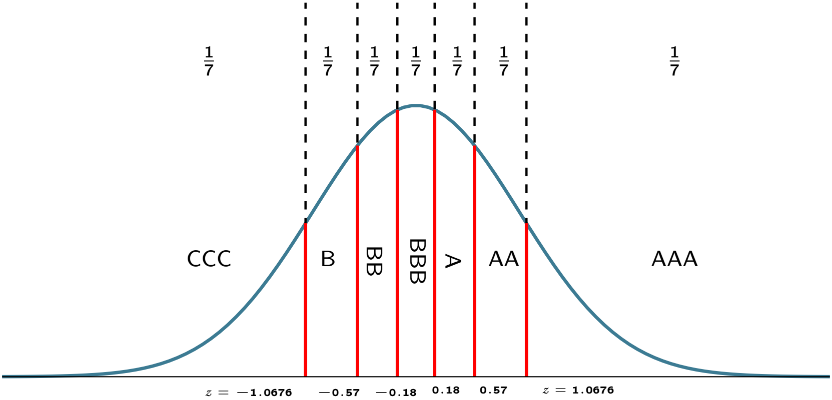
ESG rating process

Figure 32: Map function of a z-score (equal-space ratings)



ESG rating process

Figure 33: Map function of a z-score (equal-frequency ratings)



Rating migration matrix

Which rating model do you prefer? This one...

Table 15: ESG migration matrix

	AAA	AA	A	BBB	BB	B	CCC
AAA	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
AA	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
A	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
BBB	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
BB	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
B	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%
CCC	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%

$$\Rightarrow \mathcal{I}\left(\mathcal{R}\left(t\right) \mid \mathcal{R}\left(s\right)\right) = \ln 7$$

Rating migration matrix

Which rating model do you prefer? Or this one...

Table 16: ESG migration matrix

	AAA	AA	A	BBB	BB	B	CCC
AAA	100%	0%	0%	0%	0%	0%	0%
AA	0%	100%	0%	0%	0%	0%	0%
A	0%	0%	100%	0%	0%	0%	0%
BBB	0%	0%	0%	100%	0%	0%	0%
BB	0%	0%	0%	0%	100%	0%	0%
B	0%	0%	0%	0%	0%	100%	0%
CCC	0%	0%	0%	0%	0%	0%	100%

$\Rightarrow \mathcal{I}\left(\mathcal{R}\left(t\right) \mid \mathcal{R}\left(s\right)\right) = 0$

Rating migration matrix

Which rating model do you prefer? Or this one?

Table 17: ESG migration matrix

	AAA	AA	A	BBB	BB	B	CCC
AAA	96%	4%	0%	0%	0%	0%	0%
AA	2%	96%	2%	0%	0%	0%	0%
A	0%	2%	96%	2%	0%	0%	0%
BBB	0%	0%	2%	96%	2%	0%	0%
BB	0%	0%	0%	2%	96%	2%	0%
B	0%	0%	0%	0%	2%	96%	2%
CCC	0%	0%	0%	0%	0%	4%	96%

$$\Rightarrow 0 < \mathcal{I}\left(\mathcal{R}\left(t\right) \mid \mathcal{R}\left(s\right)\right) \ll \ln 7$$

Rating migration matrix

A good reference on Markov chains is:

Norris, J. R. (1997).
Markov Chains.
Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Rating migration matrix

Discrete time modeling

Definition

- $\mathcal{R}$ is a time-homogeneous Markov chain
- $\Omega_{\mathcal{R}} = \{R_1, \ldots, R_K\}$ is the state space of the chain
- $\mathbb{K} = \{1, \ldots, K\}$ is the corresponding index set
- The transition matrix is defined as $P = (p_{i,j})$
- $p_{i,j}$ is the probability that the entity migrates from rating R_i to rating R_j
- The matrix P satisfies the following properties:
  - $\forall i, j \in \mathbb{K}$, $p_{i,j} \geq 0$
  - $\forall i \in \mathbb{K}$, $\sum_{j=1}^{K} p_{i,j} = 1$

Rating migration matrix

Discrete time modeling

Table 18: ESG migration matrix #1 (one-year transition probability in %)

	AAA	AA	A	BBB	BB	B	CCC
AAA	92.76	5.66	0.90	0.45	0.23	0.00	0.00
AA	4.15	82.73	11.86	0.89	0.30	0.07	0.00
A	0.18	15.47	72.98	10.46	0.82	0.09	0.00
BBB	0.07	1.32	19.60	69.49	9.03	0.42	0.07
BB	0.04	0.19	1.55	19.36	70.88	7.75	0.23
B	0.00	0.05	0.24	1.43	21.54	74.36	2.38
CCC	0.00	0.00	0.22	0.44	2.21	13.24	83.89

Rating migration matrix

Discrete time modeling

The probability that the entity reaches the state R_j at time t given that it has reached the state R_i at time s is equal to:

$$p(s,i;t,j) = \Pr\left\{\mathcal{R}(t) = R_j \mid \mathcal{R}(s) = R_i\right\} = p_{i,j}^{(t-s)}$$

We note $p_{i,j}^{(n)}$ the n-step transition probability:

$$p_{i,j}^{(n)} = \Pr\left\{\mathcal{R}(t+n) = R_j \mid \mathcal{R}(t) = R_i\right\}$$

and the associated n-step transition matrix $P^{(n)} = \left(p_{i,j}^{(n)}\right)$

Rating migration matrix
Discrete time modeling

For $n = 2$, we obtain:

$$\begin{aligned}
p_{i,j}^{(2)} &= \Pr\left\{\mathcal{R}\left(t+2\right) = R_j \mid \mathcal{R}\left(t\right) = R_i\right\} \\
&= \sum_{k=1}^{K} \Pr\left\{\mathcal{R}\left(t+2\right) = R_j, \mathfrak{R}\left(t+1\right) = R_k \mid \mathcal{R}\left(t\right) = R_i\right\} \\
&= \sum_{k=1}^{K} \Pr\left\{\mathcal{R}\left(t+2\right) = R_j \mid \mathcal{R}\left(t+1\right) = R_k\right\} \cdot \Pr\left\{\mathcal{R}\left(t+1\right) = R_k \mid \mathcal{R}\left(t\right) = \right. \\
&= \sum_{k=1}^{K} p_{i,k} \cdot p_{k,j}
\end{aligned}$$

Rating migration matrix

Discrete time modeling

- The forward Chapman-Kolmogorov equation is :

$$p_{i,j}^{(n+m)} = \sum_{k=1}^{K} p_{i,k}^{(n)} \cdot p_{k,j}^{(m)} \qquad \forall n, m > 0$$

  or $P^{(n+m)} = P^{(n)} \cdot P^{(m)}$ with $P^{(0)} = I$

- We have:

$$\begin{aligned} P^{(n)} &= P^{(n-1)} \cdot P^{(1)} \\ &= P^{(n-2)} \cdot P^{(1)} \cdot P^{(1)} \\ &= \prod_{t=1}^{n} P^{(1)} \\ &= P^{n} \end{aligned}$$

- We deduce that:

$$p\left(t, i; t+n, j\right) = p_{i,j}^{(n)} = \mathbf{e}_i^{\top} P^{n} \mathbf{e}_j$$

Rating migration matrix

Discrete time modeling

Table 19: Two-year transition probability in % (migration matrix #1)

	AAA	AA	A	BBB	BB	B	CCC
AAA	86.28	10.08	2.25	0.92	0.44	0.02	0.00
AA	7.30	70.52	18.68	2.67	0.66	0.15	0.00
A	0.95	24.24	57.16	15.20	2.19	0.25	0.01
BBB	0.21	5.06	28.22	52.11	12.93	1.33	0.14
BB	0.09	0.79	6.07	27.45	53.68	11.37	0.55
B	0.01	0.18	0.98	6.26	31.47	57.28	3.82
CCC	0.00	0.05	0.50	1.32	6.31	21.13	70.70

Rating migration matrix

Discrete time modeling

We have:

$$
\begin{aligned}
p^{(2)}_{\mathrm{AAA},\mathrm{AAA}} &= p_{\mathrm{AAA},\mathrm{AAA}} \times p_{\mathrm{AAA},\mathrm{AAA}} + p_{\mathrm{AAA},\mathrm{AA}} \times p_{\mathrm{AA},\mathrm{AAA}} + p_{\mathrm{AAA},\mathrm{A}} \times p_{\mathrm{A},\mathrm{AAA}} + \\
&\quad p_{\mathrm{AAA},\mathrm{BBB}} \times p_{\mathrm{BBB},\mathrm{AAA}} + p_{\mathrm{AAA},\mathrm{BB}} \times p_{\mathrm{BB},\mathrm{AAA}} + \\
&\quad p_{\mathrm{AAA},\mathrm{B}} \times p_{\mathrm{B},\mathrm{AAA}} + p_{\mathrm{AAA},\mathrm{CCC}} \times p_{\mathrm{CCC},\mathrm{AAA}} \\
&= 0.9276^2 + 0.0566 \times 0.0415 + 0.0090 \times 0.0018 + \\
&\quad 0.0045 \times 0.0007 + 0.0023 \times 0.0004 \\
&= 86.28\%
\end{aligned}
$$

Rating migration matrix

Discrete time modeling

Table 20: Five-year transition probability in % (migration matrix #1)

	AAA	AA	A	BBB	BB	B	CCC
AAA	70.45	18.69	6.97	2.61	1.08	0.18	0.01
AA	13.13	50.21	26.03	7.90	2.22	0.48	0.03
A	4.35	33.20	37.78	17.99	5.52	1.08	0.09
BBB	1.50	16.49	32.49	30.90	14.61	3.63	0.38
BB	0.50	5.98	17.83	30.10	31.35	12.85	1.39
B	0.15	1.90	7.40	18.95	35.11	31.26	5.23
CCC	0.05	0.64	2.55	6.93	17.96	38.54	43.33

Rating migration matrix

Discrete time modeling

Stationary distribution

- $\pi_k^{(n)} = \Pr\{\mathcal{R}(n) = R_k\}$ is the probability of the state R_k at time n:
- $\pi^{(n)} = \left(\pi_1^{(n)}, \ldots, \pi_K^{(n)}\right)$ satisfies $\pi^{(n+1)} = P^\top \pi^{(n)}$
- The Markov chain $\mathcal{R}$ has a stationary distribution $\pi^\star$ if $\pi^\star = P^\top \pi^\star$
- $\mathcal{T}_k = \inf\{n : \mathcal{R}(n) = R_k \mid \mathcal{R}(0) = R_k\}$ is the return period of state R_k
- The average return period is then equal to:

$$\boldsymbol{\tau}_k := \mathbb{E}[\mathcal{T}_k] = \frac{1}{\pi_k^\star}$$

Rating migration matrix

Discrete time modeling

- We obtain:

$$\pi^{\star} = (17.78\%, 29.59\%, 25.12\%, 15.20\%, 8.35\%, 3.29\%, 0.67\%)$$

- The average return periods are then equal to 5.6, 3.4, 4.0, 6.6, 12.0, 30.4 and 149.0 years

$\Rightarrow$ Best-in-class (or winning-) oriented system

Rating migration matrix
Discrete time modeling

Table 21: ESG migration matrix #2 (one-month transition probability in %)

	AAA	AA	A	BBB	BB	B	CCC
AAA	93.50	5.00	0.50	0.50	0.50	0.00	0.00
AA	2.00	93.00	4.00	0.50	0.50	0.00	0.00
A	0.00	3.00	93.00	3.90	0.10	0.00	0.00
BBB	0.00	0.10	2.80	94.00	3.00	0.10	0.00
BB	0.00	0.00	0.10	3.50	94.50	1.80	0.10
B	0.00	0.00	0.00	0.10	3.70	96.00	0.20
CCC	0.00	0.00	0.00	0.40	0.50	0.60	98.50

$\Rightarrow$ The stationary distribution is $\pi^{\star} = (3.11\%, 10.10\%, 17.46\%, 27.76\%, 25.50\%, 12.68\%, 3.39\%)$ and the average return periods are equal to 32.2, 9.9, 5.7, 3.6, 3.9, 7.9 and 29.5 years

$\Rightarrow$ balanced rating system

Rating migration matrix

Discrete time modeling

Table 22: One-year probability transition in % (migration matrix #2)

	AAA	AA	A	BBB	BB	B	CCC
AAA	48.06	29.71	10.34	6.42	4.95	0.49	0.03
AA	11.65	49.25	24.10	9.60	4.87	0.49	0.03
A	2.02	17.51	49.67	24.72	5.52	0.54	0.03
BBB	0.27	3.53	17.46	55.50	20.21	2.88	0.16
BB	0.03	0.60	4.21	23.43	57.45	13.27	1.01
B	0.00	0.08	0.74	5.94	27.10	64.18	1.96
CCC	0.00	0.07	0.57	4.22	5.77	5.85	83.51

Rating migration matrix
Discrete time modeling

Table 23: One-month probability transition in % (migration matrix #1)

	AAA	AA	A	BBB	BB	B	CCC
AAA	99.36	0.53	0.05	0.04	0.02	0.00	0.00
AA	0.39	98.31	1.26	0.01	0.03	0.01	0.00
A	**−0.02**	1.65	97.14	1.21	0.02	0.01	0.00
BBB	0.01	**−0.07**	2.28	96.72	1.06	**−0.01**	0.01
BB	0.00	0.02	**−0.12**	2.29	96.92	0.88	0.01
B	0.00	0.00	0.04	**−0.15**	2.45	97.42	0.25
CCC	0.00	0.00	0.02	0.04	0.05	1.37	98.53

$\Rightarrow$ Negative probabilities

The ESG rating system is not Markovian!

Rating migration matrix

Discrete time modeling

Mean hitting time

- Let $\mathcal{A} \subset \mathbb{K}$ be a given subset. The first hitting time of $\mathcal{A}$ is given by:

$$\mathcal{T}(\mathcal{A}) = \inf \{n : \mathcal{R}(n) \in \mathcal{A}\}$$

- The mean first hitting time to target $\mathcal{A}$ from state k is defined as:

$$\boldsymbol{\tau}_k(\mathcal{A}) = \mathbb{E}[\mathcal{T}(\mathcal{A}) \mid \mathcal{R}(0) = R_k]$$

- We can show that $\boldsymbol{\tau}_k(\mathcal{A}) = 1 + \sum_{j=1}^{K} p_{k,j}\boldsymbol{\tau}_j(\mathcal{A})$
- The solution is given by the LP problem:

$$\boldsymbol{\tau}(\mathcal{A}) = \arg\min \sum_{k=1}^{K} x_k \quad \text{s.t.} \quad \begin{cases} x_k = 0 & \text{if } k \in \mathcal{A} \\ x_k = 1 + \sum_{j=1}^{K} p_{k,j}x_j & \text{if } k \notin \mathcal{A} \\ x_k \geq 0 & \end{cases}$$

Rating migration matrix
Discrete time modeling

- $\mathcal{B} = \{\text{AAA}, \text{AA}, \text{A}\}$
- $\mathcal{W} = \{\text{BB}, \text{B}, \text{CCC}\}$

Rating	$\mathcal{W}$-target				$\mathcal{B}$-target			
system	AAA	AA	A	BBB	BBB	BB	B	CCC
#1	79.21	70.04	62.34	46.54	7.50	13.28	17.58	22.68
#2	10.24	9.92	9.13	6.68	8.68	11.99	14.26	17.54

Rating migration matrix

Estimation of the transition matrix

Theoretical approach:

- Bayes theorem:

$$\begin{aligned} p_{i,j} &= \Pr\left\{\mathcal{R}(t+1) = R_j \mid \mathcal{R}(t) = R_i\right\} \\ &= \frac{\Pr\left\{\mathcal{R}(t+1) = R_j, \mathcal{R}(t) = R_i\right\}}{\Pr\left\{\mathcal{R}(t) = R_i\right\}} \end{aligned}$$

- We have seen that:

$$\Pr\left\{\mathcal{R}(t) = R_k\right\} = \mathbf{F}\left(s_k^\star\right) - \mathbf{F}\left(s_{k-1}^\star\right) = p_k$$

- We deduce that:

$$p_{i,j} = \frac{\mathbf{C}\left(\mathbf{F}\left(s_i^\star\right), \mathbf{F}\left(s_j^\star\right)\right) - \mathbf{C}\left(\mathbf{F}\left(s_{i-1}^\star\right), \mathbf{F}\left(s_j^\star\right)\right) - \mathbf{C}\left(\mathbf{F}\left(s_i^\star\right), \mathbf{F}\left(s_{j-1}^\star\right)\right) + \mathbf{C}\left(\mathbf{F}\left(s_{i-1}^\star\right), \mathbf{F}\left(s_{j-1}^\star\right)\right)}{\mathbf{F}\left(s_i^\star\right) - \mathbf{F}\left(s_{i-1}^\star\right)}$$

where $\mathbf{C}$ is the copula function of the random vector $\left(\boldsymbol{\mathcal{S}}(t), \boldsymbol{\mathcal{S}}(t+1)\right)$

Rating migration matrix

Estimation of the transition matrix

Non-parametric approach:

$$\hat{p}_{i,j}(t) = \frac{\#\left\{\mathcal{R}(t+1) = R_j, \mathcal{R}(t) = R_i\right\}}{\#\left\{\mathcal{R}(t) = R_i\right\}} = \frac{n_{i,j}(t)}{n_{i,\cdot}(t)}$$

$\Rightarrow$ cohort method vs. pooling method

Rating migration matrix

Estimation of the transition matrix

Table 24: Number of observations $n_{i,j}$ (migration matrix #1)

$n_{i,j}$	AAA	AA	A	BBB	BB	B	CCC	$n_{i,\cdot}(t)$	$\hat{p}_{i,\cdot}(t)$
AAA	2 050	125	20	10	5	0	0	2 210	3.683%
AA	280	5 580	800	60	20	5	0	6 745	11.242%
A	20	1 700	8 020	1 150	90	10	0	10 990	18.317%
BBB	10	190	2 820	10 000	1 300	60	10	14 390	23.983%
BB	5	25	200	2 500	9 150	1 000	30	12 910	21.517%
B	0	5	25	150	2 260	7 800	250	10 490	17.483%
CCC	0	0	5	10	50	300	1 900	2 265	3.775%
$n_{\cdot,j}(t)$	2 365	7 625	11 890	13 850	12 875	9 175	2 190	60 000	
$\hat{p}_{\cdot,j}(t)$	3.942%	12.708%	19.817%	23.133%	21.458%	15.292%	3.650%		100.00%

Rating migration matrix

Estimation of the transition matrix

- For the migration matrix #1, we have:

$$\pi^{\star} = (17.78\%, 29.59\%, 25.12\%, 15.20\%, 8.35\%, 3.29\%, 0.67\%)$$

- The initial empirical distribution of ratings is:

$$\hat{\pi}^{(0)} = (3.683\%, 11.242\%, 18.317\%, 23.983\%, 21.517\%, 17.483\%, 3.775\%)$$

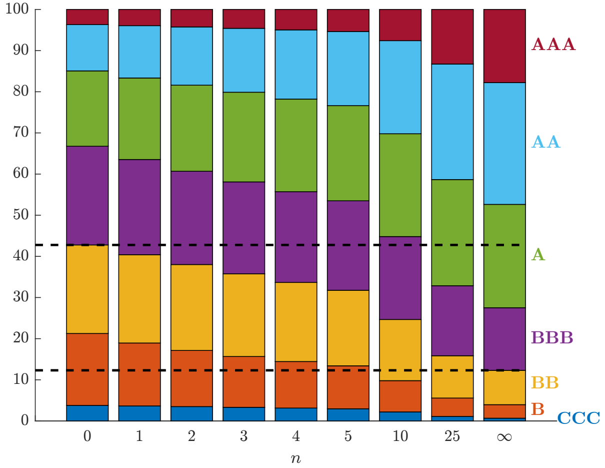
- We have:

$$\begin{aligned}\hat{\pi}^{(1)} &= \hat{P}^{\top}\hat{\pi}^{(0)} \\ &= (3.942\%, 12.708\%, 19.817\%, 23.133\%, 21.458\%, 15.290\%, 3.650\%\end{aligned}$$

Rating migration matrix

Estimation of the transition matrix

Figure 34: Dynamics of the probability distribution $\pi^{(n)}$ (migration matrix #1)



Rating migration matrix

Continuous-time modeling

Markov generator

- $t \in \mathbb{R}_+$
- The transition matrix is defined as follows:

$$P_{i,j}(s;t) = p(s,i;t,j) = \Pr\{\mathcal{R}(t) = R_j \mid \mathcal{R}(s) = R_i\}$$

- If $\mathcal{R}$ is a time-homogenous Markov, we have:

$$P(t) = P(0;t) = \exp(t\Lambda)$$

- $\Lambda = (\lambda_{i,j})$ is the Markov generator matrix $\Lambda = (\lambda_{i,j})$ where $\lambda_{i,j} \geq 0$ for all $i \neq j$ and $\lambda_{i,i} = -\sum_{j \neq i}^{K} \lambda_{i,j}$

Rating migration matrix

Continuous-time modeling

An example

- Rating system with three states: A (good rating), B (average rating) and C (bad rating)
- The Markov generator is equal to:

$$\Lambda = \begin{pmatrix} -0.30 & 0.20 & 0.10 \\ 0.15 & -0.40 & 0.25 \\ 0.10 & 0.15 & -0.25 \end{pmatrix}$$

Rating migration matrix

Continuous-time modeling

- The one-year transition probability matrix is equal to:

$$P(1) = e^{\Lambda} = \begin{pmatrix} 75.63\% & 14.84\% & 9.53\% \\ 11.63\% & 69.50\% & 18.87\% \\ 8.52\% & 11.73\% & 79.75\% \end{pmatrix}$$

- For the two-year maturity, we get:

$$P(2) = e^{2\Lambda} = \begin{pmatrix} 59.74\% & 22.65\% & 17.61\% \\ 18.49\% & 52.24\% & 29.27\% \\ 14.60\% & 18.76\% & 66.63\% \end{pmatrix}$$

- We verify that $P(2) = P(1) \cdot P(1)$ because:

$$P(t) = e^{t\Lambda} = \left(e^{\Lambda}\right)^{t} = P(1)^{t}$$

- We have:

$$P\left(\frac{1}{12}\right) = e^{\frac{1}{12}\Lambda} = \begin{pmatrix} 97.54\% & 1.62\% & 0.83 \\ 1.22\% & 96.74\% & 2.03 \\ 0.82\% & 1.22\% & 97.95 \end{pmatrix}$$

Rating migration matrix

Matrix function

Matrix function

We consider the matrix function in the space $\mathbb{M}$ of square matrices:

$$\begin{aligned} f: \quad \mathbb{M} &\longrightarrow \mathbb{M} \\ A &\longmapsto B = f(A) \end{aligned}$$

For instance, if $f(x) = \sqrt{x}$ and A is positive, we can define the matrix B such that:

$$BB^* = B^*B = A$$

B is called the square root of A and we note $B = A^{1/2}$

Rating migration matrix

Matrix function

- We consider the following Taylor expansion:

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \ldots$$

- We can show that if the series converge for $|x - x_0| < \alpha$, then the matrix $f(A)$ defined by the following expression:

$$f(A) = f(x_0) + (A - x_0 I) f'(x_0) + \frac{(A - x_0 I)^2}{2!} f''(x_0) + \ldots$$

converges to the matrix B if $|A - x_0 I| < \alpha$ and we note $B = f(A)$

Rating migration matrix
Matrix function

- In the case of the exponential function, we have:

$$f(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

- We deduce that the exponential of the matrix A is equal to:

$$B = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k"!}$$

- The logarithm of A is the matrix B such that $e^B = A$ and we note $B = \ln A$

Rating migration matrix

Matrix function

- Let A and B be two $n \times n$ square matrices. We have the properties:

$$
\begin{array}{rcl}
f\left(A^{\top}\right) & = & f(A)^{\top} \\
Af(A) & = & f(A)A \\
f\left(B^{-1}AB\right) & = & B^{-1}f(A)B
\end{array}
$$

- It follows that:

$$
\left\{
\begin{array}{ll}
e^{A^{\top}} = \left(e^{A}\right)^{\top} & \\
e^{B^{-1}AB} = B^{-1}e^{A}B & \\
Ae^{B} = e^{B}A & \text{if } AB = BA \\
e^{A+B} = e^{A}e^{B} = e^{B}e^{A} & \text{if } AB = BA
\end{array}
\right.
$$

Rating migration matrix

Matrix function

Definition

The Schur decomposition of the $n \times n$ matrix A is equal to:

$$A = QTQ^*$$

where Q is a unitary matrix and T is an upper triangular matrix

For transcendental functions, we have:

$$f(A) = Qf(T)Q^*$$

where $A = QTQ^*$ is the Schur decomposition of A

Rating migration matrix

Continuous-time modeling

Estimation of the Markov generator

We have:

$$\hat{\Lambda} = \frac{1}{t} \ln \left(\hat{P}(t) \right)$$

$\Rightarrow \hat{\Lambda}$ may not verify the Markov conditions: $\hat{\lambda}_{i,j} \geq 0$ for all $i \neq j$ and $\sum_{j=1}^{K} \lambda_{i,j} = 0$

Rating migration matrix

Continuous-time modeling

Table 25: Non-Markov generator $\Lambda' = \ln(P)$ of the migration matrix #1 (in %)

	AAA	AA	A	BBB	BB	B	CCC
AAA	−7.663	6.427	0.542	0.466	0.245	**−0.016**	**−0.00**
AA	4.770	−20.604	15.451	**−0.001**	0.318	0.066	**−0.00**
A	**−0.267**	20.259	−35.172	14.953	0.152	0.083	**−0.00**
BBB	0.102	**−1.051**	28.263	−40.366	13.100	**−0.128**	0.08
BB	0.032	0.307	**−1.762**	28.351	−37.889	10.832	0.12
B	**−0.005**	**−0.008**	0.503	**−2.240**	30.227	−31.482	3.00
CCC	0.000	**−0.024**	0.194	0.469	0.365	16.806	−17.81

Rating migration matrix

Continuous-time modeling

Israel-Rosenthal-Wei estimators

1. The first approach consists in adding the negative values back into the diagonal values:

$$\begin{cases} \bar{\lambda}_{i,j} = \max\left(\hat{\lambda}_{i,j}, 0\right) & i \neq j \\ \bar{\lambda}_{i,i} = \hat{\lambda}_{i,i} + \sum_{j \neq i} \min\left(\hat{\lambda}_{i,j}, 0\right) \end{cases}$$

2. The second estimator carries forward the negative values on the matrix entries which have the correct sign:

$$\begin{cases} G_i = \left|\hat{\lambda}_{i,i}\right| + \sum_{j \neq i} \max\left(\hat{\lambda}_{i,j}, 0\right), B_i = \sum_{j \neq i} \max\left(-\hat{\lambda}_{i,j}, 0\right) \\ \tilde{\lambda}_{i,j} = \begin{cases} 0 & \text{if } i \neq j \text{ and } \hat{\lambda}_{i,j} < 0 \\ \hat{\lambda}_{i,j} - B_i \left|\hat{\lambda}_{i,j}\right| / G_i & \text{if } G_i > 0 \\ \hat{\lambda}_{i,j} & \text{if } G_i = 0 \end{cases} \end{cases}$$

Rating migration matrix
Continuous-time modeling

Table 26: Markov generator of the migration matrix #1 (in %)

	AAA	AA	A	BBB	BB	B	CCC
AAA	−7.679	6.427	0.542	0.466	0.245	0.000	0.000
AA	4.770	−20.606	15.451	0.000	0.318	0.066	0.000
A	0.000	20.259	−35.447	14.953	0.152	0.083	0.000
BBB	0.102	0.000	28.263	−41.545	13.100	0.000	0.080
BB	0.032	0.307	0.000	38.351	−39.651	10.832	0.129
B	0.000	0.000	0.503	0.000	30.227	−33.735	3.006
CCC	0.000	0.000	0.194	0.469	0.365	16.806	−17.834

Rating migration matrix

Continuous-time modeling

Table 27: ESG migration Markov matrix #1 (one-year transition probability in %)

	AAA	AA	A	BBB	BB	B	CCC
AAA	92.75	5.66	0.90	0.45	0.23	0.01	0.00
AA	4.17	82.73	11.85	0.89	0.30	0.07	0.00
A	0.40	15.51	72.79	10.39	0.81	0.10	0.01
BBB	0.12	2.11	19.60	68.69	8.91	0.50	0.07
BB	0.04	0.43	2.79	19.25	69.65	7.61	0.23
B	0.01	0.09	0.65	2.98	21.21	72.71	2.35
CCC	0.00	0.02	0.25	0.58	2.19	13.09	83.87

Rating migration matrix

Continuous-time modeling

Table 28: **Original** migration matrix

	AAA	AA	A	BBB	BB	B	CCC
AAA	92.76	5.66	0.90	0.45	0.23	0.00	0.00
AA	4.15	82.73	11.86	0.89	0.30	0.07	0.00
A	**0.18**	15.47	72.98	10.46	0.82	0.09	0.00
BBB	0.07	**1.32**	19.60	69.49	9.03	0.42	0.07
BB	0.04	0.19	**1.55**	19.36	70.88	7.75	0.23
B	0.00	0.05	0.24	**1.43**	21.54	74.36	2.38
CCC	0.00	0.00	0.22	0.44	2.21	13.24	83.89

Table 29: **New** migration matrix

	AAA	AA	A	BBB	BB	B	CCC
AAA	92.75	5.66	0.90	0.45	0.23	0.01	0.00
AA	4.17	82.73	11.85	0.89	0.30	0.07	0.00
A	**0.40**	15.51	72.79	10.39	0.81	0.10	0.01
BBB	0.12	**2.11**	19.60	68.69	8.91	0.50	0.07
BB	0.04	0.43	**2.79**	19.25	69.65	7.61	0.23
B	0.01	0.09	0.65	**2.98**	21.21	72.71	2.35
CCC	0.00	0.02	0.25	0.58	2.19	13.09	83.87

Rating migration matrix

Continuous-time modeling

Why it is important that ESG ratings satisfy the Markov property

- Lack of memory:

$t-2$		$t-1$		t		$t+1$
AAA	$\longrightarrow$	BBB	$\longrightarrow$	BBB	$\longrightarrow$	?
BBB	$\longrightarrow$	BBB	$\longrightarrow$	BBB	$\longrightarrow$	?
BB	$\longrightarrow$	BB	$\longrightarrow$	BBB	$\longrightarrow$	?

- Non-Markov property:

$$\Pr\left\{\mathcal{R}_{c_1}(t+1)=R_j \mid \mathcal{R}_{c_1}(t)=R_i\right\} \neq \Pr\left\{\mathcal{R}_{c_2}(t+1)=R_j \mid \mathcal{R}_{c_2}(t)=R_i\right\}$$

for two different companies c_1 and c_2

Rating migration matrix

Continuous-time modeling

How to perform a dynamic analysis?

- We deduce that:

$$\pi_k(t, \mathcal{A}) = \Pr\{\mathcal{R}(t) \in \mathcal{A} \mid \mathcal{R}(0) = k\} = \sum_{j \in \mathcal{A}} \mathbf{e}_k^\top e^{t\Lambda} \mathbf{e}_j$$

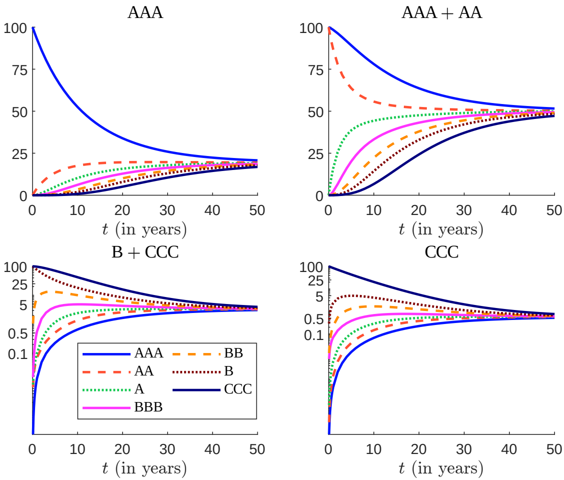
- Some properties
  - $\partial_t \exp(\Lambda t) = \Lambda \exp(\Lambda t)$
  - $\partial_t^m \exp(\Lambda t) = \Lambda^m \exp(\Lambda t)$
  - $\int_0^t e^{\Lambda s} \, \mathrm{d}s = \left(e^{\Lambda t} - I_K\right) \Lambda^{-1}$
- For example, the "*time density function*" is given by:

$$\pi_k^{(m)}(t, \mathcal{A}) := \frac{\partial \, \pi_k(t, \mathcal{A})}{\partial \, t^m} = \sum_{j \in \mathcal{A}} \mathbf{e}_k^\top \Lambda^m e^{t\Lambda} \mathbf{e}_j$$

Rating migration matrix

Continuous-time modeling

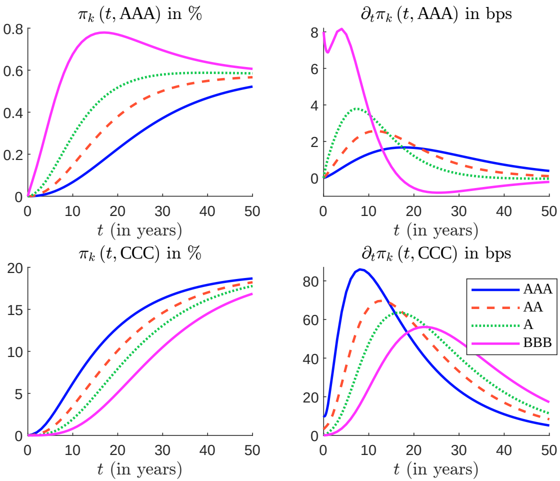
Figure 35: Probability $\pi_k(t, \mathcal{A})$ to reach $\mathcal{A}$ at time t (migration matrix #1)



Rating migration matrix

Continuous-time modeling

Figure 36: Dynamic analysis (migration matrix #1)



Rating migration matrix

Comparison with credit ratings

Table 30: Example of credit migration matrix (one-year probability transition in %)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	92.82	6.50	0.56	0.06	0.06	0.00	0.00	0.00
AA	0.63	91.87	6.64	0.65	0.06	0.11	0.04	0.00
A	0.08	2.26	91.66	5.11	0.61	0.23	0.01	0.04
BBB	0.05	0.27	5.84	87.74	4.74	0.98	0.16	0.22
BB	0.04	0.11	0.64	7.85	81.14	8.27	0.89	1.06
B	0.00	0.11	0.30	0.42	6.75	83.07	3.86	5.49
CCC	0.19	0.00	0.38	0.75	2.44	12.03	60.71	23.50
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Source: Kavvathas (2001).

Rating migration matrix
Comparison with credit ratings

The trace statistics is equal to:

$$\lambda(t) = \frac{\operatorname{trace}\left(e^{t\Lambda}\right)}{K}$$

Rating migration matrix

Comparison with credit ratings

Figure 37: Trace statistics of credit and ESG migration matrices