Optimal Decumulation: Annuitization, Drawdown, and the Flex-to-Fix Strategy*

Wajdi Maatouk
Amundi Investment Institute
wajdi.maatouk@amundi.com

Thierry Roncalli
Amundi Investment Institute
thierry.roncalli@amundi.com

Jiali Xu
Amundi Investment Institute
jiali.xu@amundi.com

September 2026

Abstract

We study the optimal decumulation of retirement wealth under longevity and market risk. In particular, we focus on the transition between drawdown, which preserves liquidity and bequest potential, and annuitization, which provides a guaranteed lifetime income. We develop a *flex-to-fix* framework in which retirees remain invested and liquid before converting part or all of their wealth into an annuity at a later date. The flex-to-fix strategy has two limiting cases: immediate annuitization (*fix-now*) and permanent drawdown (*fix-never*). A static actuarial analysis shows that the optimal fixed conversion date depends strongly on the bequest motive, the withdrawal rate, and the return of the drawdown portfolio. There is no single optimal annuitization age across retirees, although optimal dates tend to concentrate around age 65 (retirement) or ages 75–80. Then, we develop two dynamic formulations that allow the conversion decision to respond to the retiree's realized wealth-income state. An optimal-stopping formulation shows that a well-chosen fixed date already captures most of the value of delaying annuitization, while state-contingent timing adds only a small further gain. A Hamilton-Jacobi-Bellman variational inequality (HJBVI) formulation characterizes the free boundary governing a gradual transition from drawdown to annuity income. It also reveals an implicit bond effect of annuities on optimal equity allocation, and shows that spreading purchases over time substantially reduces market timing and ruin risk compared to single-date conversion rules. Finally, we propose approximating the HJBVI strategy with a simpler rule that spreads annuity purchases evenly across the flex window, once the optimal target date has been calibrated.

Keywords: Retirement planning, lifecycle investing, decumulation, withdrawal, annuity, income drawdown, fix-now, flex-to-fix, fix-never, mortality modeling, Lee-Carter model, mortality credit, bequest, dynamic programming, optimization, Hamilton-Jacobi-Bellman variational inequality.

JEL Classification: C61, G11, G50.

*The authors are very grateful to Benjamin Bruder and Laurent Caudamine for their helpful comments. The opinions expressed in this research are those of the authors and are not meant to represent the opinions or official positions of Amundi Asset Management. The authors used an AI-based language model to assist with English language polishing. The authors remain fully responsible for the content of the paper.

1 Introduction

Population aging is transforming retirement into a complex problem of financing consumption over an increasingly uncertain horizon. Rising life expectancy and sustained improvements in mortality have extended the length of retirement, putting pressure on public and occupational pension systems. As a result, a growing share of retirement risk is being transferred from collective institutions to households. This makes individual savings and investment decisions increasingly important for achieving financial security in old age. This evolution presents three related challenges. First, demographic aging generates systematic and idiosyncratic longevity risks, creating uncertainty regarding retirement duration and cost. Second, despite greater participation in financial markets, many households are still concerned about their ability to accumulate enough wealth to maintain their standard of living throughout retirement. Third, these concerns emerge in a context where traditional defined-benefit pension plans are gradually being replaced by funded and defined-contribution plans, increasing individuals' exposure to financial market risk.

What does longevity risk mean for retirement? According to United Nations (2024), the global population aged 65 and older grew from 126 million in 1950 to about 880 million in 2026. This figure is projected to surpass 1 billion by 2030, reaching roughly 1.6 billion by 2050 and 2.4 billion by the end of the century. This corresponds to an old-age dependency ratio[1] that has risen from 8 in 1950 to 16 in 2025, and is projected to reach 26 in 2050 and 40 in 2100. This means that the number of retirees has grown, and will continue to grow, putting pressure on public pension systems that were designed for a much younger population. Because this trend affects the whole population simultaneously and its future pace is uncertain, it is a source of systematic longevity risk. At the same time, the global population aged 90 and older was under 2 million in 1950, stands at around 26 million today, and is expected to reach 260 million by 2100. This concentration of very old age increases idiosyncratic longevity risk at an individual level, since it becomes difficult to predict how long anyone will need care (Vaupel *et al.*, 2021). It also drives up health and long-term care expenses for the population as a whole and for individuals. In this case, retirement planning becomes a long-term problem of financing consumption and expenditures over an uncertain lifetime.

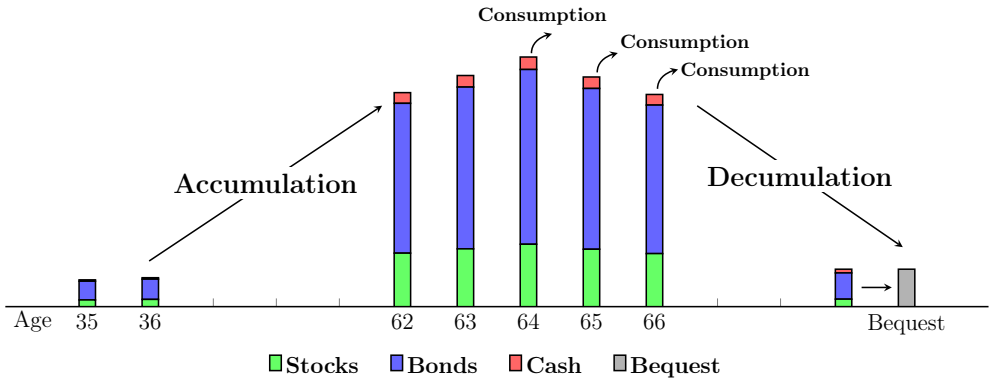
Consequently, retirement security has become a major financial objective for households. For example, the 2026 Amundi Annual Report, based on a survey of 18 000 retail investors across 26 countries, highlights "*a greater individual responsibility for long-term financial security with personal investments expected to become the largest source of retirement income*" (Amundi Investment Solutions, 2026, page 4). While wealth accumulation and income generation remain the primary motivations for investing in financial markets — cited by 53% and 37% of respondents, respectively — retirement planning has become an increasingly important objective. Globally, 36% of respondents reported investing to fund their retirement and strengthen their long-term financial security. However, this global figure masks substantial differences across age groups. The importance of retirement-related investment objectives steadily increases over the life cycle. Among individuals aged 21–30, only 27% report primarily investing to fund retirement. This proportion increases to 34% among those aged 31–40 and to 36% among individuals aged 41–50. For investors over 50, retirement funding becomes a dominant investment motive, with approximately 43–44% citing it as their main objective. These findings suggest that retirement concerns become increasingly salient as individuals approach the end of their working lives. However, recognizing retirement as an important financial objective does not necessarily imply confidence in achieving it. Survey

[1] The old-age dependency ratio is defined as the number of people aged 65 and older per 100 people of working age (15–64 years old).

results show a significant discrepancy between households' short-term optimism and their long-term fear of a financial shortfall. The 2025 GSAM survey describes this discrepancy as an *optimism gap*. While 68% of respondents consider that their retirement savings are on track, largely thanks to the strong performance of financial markets since the Covid-19 crisis, 58% of respondents also anticipate that they will run out of money during their lifetime (Goldman Sachs Asset Management, 2025, page 11). The Amundi survey points in the same direction, with confidence varying sharply by age. More than 30% of individuals under 40 are confident of achieving long-term financial security, compared to only 15% of those over 50 (Amundi Investment Solutions, 2026, pages 28-29). This fear is not unfounded for some households. For example, 13% of pensioners in Europe continue to work[2], and about 30% of them are working for financial necessity[3] (Eurostat, 2024, pages 17–22).

In this context[4], as pillar 1 and pillar 2 defined-benefit (DB) plans come under pressure, we observe strong growth in pillar 2 defined-contribution (DC) occupational plans, and, above all, in pillar 3. Three factors reinforce this trend. First, high public debt limits the fiscal ability of governments to absorb rising pension liabilities (OECD, 2025). Second, employers are increasingly unwilling to take on longevity risk in addition to business risk. Third, the nature of careers has changed. In the past, workers expected stable employment with a single employer. Today, many hold multiple jobs with different employers for shorter periods of time. This pattern is not well suited to DB plans. The shift from DB to DC is therefore no accident, but this shift transfers a growing share of longevity risk to households, making the accumulation and decumulation phases central to retirement finance.

Figure 1: Accumulation and decumulation



As shown in Figure 1, a complete pension scheme consists of two phases: pre-retirement accumulation and post-retirement decumulation. The accumulation phase is the period during which a person saves and makes long-term investments to accumulate retirement funds.

[2]There is significant heterogeneity among countries. For example, countries with a proportion greater than 25% include Cyprus, Czechia, Estonia, Iceland, Ireland, Finland, Latvia, Lithuania, Norway, Sweden, and Switzerland. Those with a proportion smaller than 5% include Croatia, Greece, Romania, and Spain.

[3]It is also worth noting that about 35% continue to work because they enjoy it.

[4]Retirement systems are generally analyzed using the World Bank's pension conceptual framework (Holzmann and Hinz, 2005, Table 1, page 10), which distinguishes five pillars: a non-contributory zero pillar that provides general social assistance; a mandatory first pillar, comprising public pension plans financed through pay-as-you-go (PAYG) arrangements; a mandatory second pillar of privately funded pensions financed by employer and employee contributions (DB/DC); a voluntary third pillar, including personal savings and voluntary individual retirement accounts; and a non-financial fourth pillar, providing non-financial sources of retirement support, such as family assistance.

In contrast, the decumulation phase is the process by which investors convert their retirement savings into income to meet their needs while continuing to invest their remaining funds (Martellini and Milhau, 2020; Bruder *et al.*, 2023). During this phase, there is a regular outflow of cash from the reserve. The two phases are closely interconnected. The adequacy of decumulation depends critically on the wealth accumulated beforehand, while accumulation choices depend on expectations about retirement consumption. Rising life expectancy complicates this balance by increasing the necessary resources for an adequate income and extending the period over which savings must be managed. Consequently, voluntary savings and supplementary pension plans are necessary to bridge the gap left by public and mandatory programs. This interdependence highlights the fact that retirement planning is an ongoing process in which saving, investing, and spending decisions are continuously linked. The accumulation phase is particularly important in this context. Without sufficient wealth at retirement, the choice of decumulation options becomes largely irrelevant. This highlights the importance of sound accumulation strategies (Bruder *et al.*, 2025). Yet accumulating enough does not settle the retirement problem. What also matters is the ability to transform that wealth into a reliable stream of lifetime income, and the two phases pose different challenges. During the accumulation phase, individuals can gauge their progress through contribution rates and investment performance. During retirement, however, they must decide how to convert wealth into consumption without exhausting it prematurely, while many key variables are uncertain: lifespan, inflation, healthcare or disability expenses, and market returns. Decumulation is therefore a major challenge.

In fact, decumulation is certainly a more complicated mathematical and financial problem than accumulation:

> "*I believe decumulation is a more difficult challenge than accumulation. My generation, the baby boomers, have begun retiring, and the ones who have managed to build up a nest egg are going to need a lot of help figuring out how to make it last. At least from my perspective, this is the sort of problem that would be fun to think about.*" (Thaler, 2019).

Similarly, Sharpe (2017) described decumulation as the "*nastiest, hardest problem in finance*", pointing to the many dimensions of the problem (actuarial mortality tables, market return scenarios, inflation pathways, etc.). Decumulation is not a new problem, but the shift from DB to DC plans has renewed it. In a DB plan, the tangible variable is the retirement income, not the wealth behind it. In France, for instance, the Agirc-Arrco statement (a mandatory complementary pension plan) reports the number of points accumulated. Multiplying this number by the current value of the point gives the annual or monthly gross pension that those points would provide today. By contrast, statements for DC and voluntary schemes usually report the current value of the accumulated wealth, and typically keep the retirement income it would finance hidden from the saver. The dynamics of accumulation also differ. In a DB plan, accrued retirement income is a nondecreasing function of time. This is a true accumulation process. In a DC plan, however, retirement wealth is not a nondecreasing function of time since it depends on the performance of the invested assets. Robert Merton emphasizes the importance of this difference:

> "*Our approach to saving is all wrong: We need to think about monthly income, not net worth.*" (Merton, 2014).

Although these three Nobel laureates in economics all stress the importance of decumulation, most of the academic literature, policy work, surveys, and industry publications still focus on accumulation. Decumulation has received less attention, even though it represents 50% of the retirement problem.

For example, the famous reports *Global Pension Assets Study* and *Pensions at a Glance* published by the Thinking Ahead Institute (TAI) and the OECD focus on the size, allocation, and performance of pension assets, but they say little about payouts (Thinking Ahead Institute, 2026; OECD, 2025). According to the TAI, the value of pension assets worldwide is around $73 trillion. This figure is comparable to the value of mutual funds ($74 trillion), but much higher than the value of insurance assets ($43 trillion) and sovereign wealth funds ($15 trillion). However, this figure represents assets under management by pension funds, not the total value of the retirement market. On the one hand, individuals also save for retirement through mutual funds, insurance products, and bank deposits, and retirees may continue to use these investment vehicles during the decumulation phase. On the other hand, pension fund assets combine accumulation assets (those of active members) and decumulation assets (those funding the pensions of retirees). Therefore, it is very difficult to measure the split between accumulation and decumulation assets, and very few study provide this information. For instance, according to The Investment Association (2026), the total value of assets in the UK private pension system is approximately £3.8 trillion, comprising £1.7 trillion in DB schemes and £2.2 trillion in DC schemes. Within the DC assets, around 70% are held in the accumulation phase, while the remaining 30% are in the decumulation phase. Within the decumulation assets, about £260 billion is invested in income drawdown products and £420 billion in annuities, representing 38% and 62% of decumulation assets, respectively.

Table 1: Net pension replacement rates from mandatory (public and private) and voluntary pension schemes for men (in %)

Country	Earnings			Country	Earnings		
	Low	Medium	High		Low	Medium	High
Australia	82.7	53.0	38.1	Switzerland	59.2	47.5	25.0
Canada	56.0	45.1	25.0	United Kingdom	76.2	54.2	39.5
Canada*	79.4	70.6	50.4	United States	62.5	51.3	40.0
France	66.1	70.0	58.9	United States*	111.4	100.7	92.2
Germany	57.7	53.3	38.8	OECD	75.4	63.2	52.9
Germany*	71.6	68.0	52.7	OECD*	81.2	69.2	59.9
Ireland	56.5	33.7	20.1	EU27	76.5	68.3	59.7
Ireland*	93.3	72.3	61.1	Argentina	103.1	78.6	66.4
Italy	70.4	79.0	81.9	Brazil	95.9	97.5	87.7
Japan	61.8	42.4	34.0	China	130.3	103.6	91.2
Netherlands	97.2	96.0	89.7	India	44.6	44.6	24.2
Norway	74.4	54.9	36.6	Indonesia	55.6	55.6	55.3
Spain	78.6	86.3	57.9	South Africa	16.8	8.9	4.9
Sweden	67.4	66.3	84.4	South Africa*	30.3	32.0	34.6

* Includes voluntary pension schemes in addition to mandatory ones.

Source: OECD (2025, Table 4.5, page 171).

One of the key questions is what level of retirement living standard can be supported by pension income. Generally, three components of financial needs in retirement can be distinguished: (1) basic needs such as housing, food, transportation, basic leisure, and clothing; (2) unexpected expenses such as healthcare and old-age dependency expenses, and (3) discretionary expenditures for a more comfortable lifestyle such as holidays, cars, leisure activities, and other non-essential consumption (OECD, 2024, Figure 5.1, page 118 and Table 5.1, page 120). Table 1 reports the net replacement rates[5] for men by country and by earnings group (low, medium, and high). For example, in the OECD countries, the

[5]The gross pension replacement rate is defined as gross pension entitlement divided by gross pre-

Table 2: Net pension wealth provided by mandatory (public and private) pension schemes for men (in multiple of annual earnings)

Country	Earnings			Country	Earnings		
	Low	Medium	High		Low	Medium	High
Australia	15.4	10.4	7.1	Spain	15.9	17.5	11.7
Canada	11.3	9.1	5.1	Sweden	11.6	11.4	14.5
France	13.5	14.3	12.0	Switzerland	13.5	10.8	5.7
Germany	12.4	11.4	8.3	United Kingdom	15.1	10.4	7.5
Greece	19.1	17.0	15.4	United States	11.5	9.4	7.4
Ireland	12.7	7.6	4.5	OECD	14.7	12.2	10.1
Italy	12.2	13.7	14.2	EU27	14.4	12.8	11.1
Japan	13.2	9.0	7.2	Argentina	19.9	15.2	12.8
Korea	11.0	7.7	4.9	Brazil	16.8	17.1	15.4
Lithuania	6.4	4.9	3.7	China	27.3	21.6	18.9
Luxembourg	25.1	22.6	20.5	India	9.0	9.0	4.9
Netherlands	17.9	17.0	15.6	Indonesia	8.0	8.0	8.0
Norway	15.6	11.5	7.6	South Africa	3.0	1.6	0.9

Source: OECD (2025, Table 4.7, page 175).

average net replacement rate is 75.4% for low earners, 63.2% for medium earners and 52.9% for high earners. When voluntary pension schemes are included, these figures rise to 81.2%, 69.2% and 59.9%, respectively. Three remarks can be made. First, most net replacement rates are below 100%. This means that retirees have a lower income in retirement than when they were working. In some countries, such as South Africa, the net replacement rate is even below 20% for all earnings groups. Exceptions above 100% are China (low and medium earners), Argentina (low earners) and the United States when voluntary pension schemes are included (low and medium earners). Second, there are significant differences in net replacement rates across earnings groups within countries. For instance, in Australia, the net replacement rate for low earners is more than 40 percentage points higher than for high earners (82.7% versus 38.1%). Third, voluntary pension saving can considerably improve retirees' income. For high earners, the net replacement rates increase from 20.1%, 40.0% and 4.9% to 61.1%, 92.2% and 34.6% in Ireland, the United States and South Africa, respectively. These results suggest that mandatory schemes alone leave many retirees with a substantial gap in their living standards, and that voluntary schemes can play a key role in closing or mitigating this issue. Another way to assess pension adequacy relative to the pre-retirement situation is to compute net pension wealth[6], which is expressed as a multiple of annual earnings. Results are reported in Table 2. For example, a net pension wealth of 10 means that the discounted value of all pension payments received during retirement is equal to ten times the annual net earnings (or ten years of net earnings). Some countries have particularly generous schemes, with net pension wealth above 20 for some earnings groups: Luxembourg (25.1, 22.6 and 20.5 for low, medium and high earners, respectively) and China (27.3 and 21.6 for low and medium earners). Other countries are much less generous. In South Africa, for instance, net pension wealth is 3.0, 1.6 and 0.9 for low, medium and high earners, respectively. These figures must be interpreted in light of the expected lifetime[7].

retirement earnings. The net pension replacement rate is defined as the individual net pension entitlement divided by net pre-retirement earnings, taking into account personal income taxes and social security contributions paid by workers and pensioners.

[6]Net pension wealth is defined as the present value, at the retirement age, of the lifetime flow of retirement incomes from mandatory pension schemes. This is calculated after the personal income taxes and social security contributions paid by pensioners. It is expressed as a multiple of annual net pre-retirement earnings. The calculation uses a uniform real discount rate of 1.5% (OECD, 2025, pages 172 and 174).

[7]The life expectancy is 79.6 years in Europe and North America, 77.8 in Argentina, 84.3 in Australia,

Retirement can also impact poverty levels. In many countries, the income poverty rate is higher among older people (aged over 65) than in the total population. On average across OECD countries, the poverty rate is 14.8% for people aged 65 and over compared to 11.5% for the total population (OECD, 2025, Table 7.2, page 211). Moreover, old-age poverty also tends to increase with age. In OECD, the rate is 17.2% for people aged 75 and over, versus 13.1% for those aged 65 to 75. In Australia, the income poverty rate reaches 22.6% for people aged 65 and over, versus 12.6% for the total population. However, these averages mask considerable heterogeneity between men and women, between singles and couples, and between homeowners and renters. In particular, housing is an important component of the retirement decumulation challenge (Callaghan *et al.*, 2020, Chart 1, page 32). In Australia, fewer than 10% of retired homeowner couples and around 15% of retired single homeowners are classified as experiencing financial stress and living in poverty. In contrast, poverty rates among retired renters are substantially higher, exceeding 20% for couples and reaching approximately 60% for singles. These figures highlight the critical role that homeownership plays in ensuring financial security in retirement.

These observations highlight that retirement risk is not just an accumulation problem. While having enough money is necessary for financial security, it is not sufficient. Retirees must also learn how to convert their savings into a steady income. However, they often make this decision with little prior experience and under considerable uncertainty. Thus, the decumulation phase presents a unique set of economic, financial, and actuarial challenges and is the natural counterpart of the accumulation phase. While the accumulation problem addresses how much to save and how to invest, the decumulation problem addresses how to spend, invest, and insure against longevity risk. Consequently, the transition from accumulating wealth to generating retirement income has become a central question in modern pension finance (Milevsky, 1998, 2001, 2006). At retirement, households face a wide range of payout options. These include fixed and inflation-linked life annuities, variable annuities, phased withdrawals through income drawdown arrangements, lump-sum distributions, and more recent forms of non-guaranteed lifetime income products. Examples of these products include collective defined-contribution schemes, target-benefit plans, group self-annuitization arrangements, and tontines (OECD, 2022b; Pensions Europe, 2025). Although these products vary in flexibility, liquidity, and risk-sharing, they all address the same fundamental challenge: providing income over an uncertain lifetime while managing investment and longevity risks.

Among these payout options, life annuities occupy a special place in standard economic theory, because they explicitly provide insurance against longevity risk and voluntary annuitization has long been regarded as the optimal way to hedge this risk. In a seminal contribution, Yaari (1965) uses a life-cycle utility framework to show that a rational consumer with an uncertain lifetime and no bequest motive should annuitize 100% of their wealth at an actuarially fair price. The intuition is straightforward. By pooling longevity risk across individuals, annuities generate a "*mortality credit*". Indeed, the accumulated capital of deceased pool members is redistributed to survivors. Thus, conditional on survival, an annuity yields an effective return well above that of conventional risk-free assets (Milevsky *et al.*, 2006). However, mortality credits are not specific to insured annuities since pooled annuity funds and tontines also deliver them, albeit without guaranteed payments (Stamos, 2008; Milevsky and Salisbury, 2015; Denuit *et al.*, 2022). Subsequent research has shown that the theoretical appeal of annuities extends well beyond Yaari's original framework. Mitchell *et al.* (1999) document that individual annuities are not actuarially fair and contain non-negligible loadings, but show that their money's worth ratios remain relatively

76.4 in Brazil, 79.0 in China, 72.7 in India, 85.1 in Japan and 66.7 in South Africa (United Nations, 2024).

high[8]. Combined with the utility value of longevity insurance, these findings suggest that annuities may remain attractive despite administrative costs. Davidoff *et al.* (2005) further relax Yaari's assumptions and show that the full annuitization result is remarkably robust and does not rely on actuarially fair pricing or restrictive assumptions about preferences. In complete markets, access to annuities improves welfare under very general conditions[9]. More recently, Milevsky (2013) emphasizes that life annuities are the most direct instrument for hedging individual longevity risk. However, observed behavior differs sharply from the predictions of standard life-cycle models. This divergence was noted by Modigliani in his Nobel lecture and his discussion of retirement saving and consumption behavior (Modigliani, 1986; Deaton, 2005). Voluntary annuity demand remains low in most countries. For example, in the U.K., only 5.9% of retired households in the 2002–04 English Longitudinal Study of Ageing (ELSA) reported income from a voluntary annuity (9.6% among stock market participants and 3.2% among stock market non-participants) (Inkmann *et al.*, 2011). Similar evidence has been documented in the United States and Australia (Brown, 2001; Benartzi *et al.*, 2011; Callaghan *et al.*, 2020). This persistent gap between normative theory and observed choices is known as the "*annuity puzzle*" (Benartzi *et al.*, 2011; Hershfield *et al.*, 2026). The academic literature has proposed numerous explanations for this discrepancy, which can be broadly grouped into two categories: rational frictions that reduce the attractiveness of annuities in real-world settings, and behavioral or psychological factors that cause individuals to depart from the predictions of standard expected-utility models. The first category remains within the rational agent framework and examines whether real-world market imperfections make annuities a "*bad deal*". The bequest motive is one of these frictions. Many retirees want to leave wealth to their heirs, but a standard single-life annuity eliminates capital at death. Lockwood (2012) shows that even a modest bequest motive can make zero annuitization optimal at market prices. However, the strength and heterogeneity of bequest motives remain debated (Bernheim, 1991; Kopczuk and Lupton, 2007). Illiquidity is a second friction. Annuitization is largely irreversible, locking up capital and exposing retirees to uninsured shocks, such as large and uncertain healthcare and long-term care expenditures. Peijnenburg *et al.* (2017) show that uninsured out-of-pocket medical expenses substantially reduce the demand for annuitization, while Reichling and Smetters (2015) show that the optimal annuity share depends on health-related and stochastic mortality risk. Adverse selection is a third friction. Since buyers have private information about their own health and longevity, insurers use annuitant mortality tables rather than population mortality tables to price, which increases the effective cost for individuals with average life expectancy. Finkelstein and Poterba (2004) document this empirically for the U.K. annuity market. Finally, a crowding-out effect is at work. Public pensions (Pillar 1) and occupational defined-benefit plans (Pillar 2) already provide a substantial amount of annuitized retirement income, and intra-household risk sharing among couples also reduces the incremental demand for voluntary annuities (Brown and Poterba, 2000; Dushi and Webb, 2004; Pashchenko, 2013). Nevertheless, quantitative studies suggest that these rational explanations, taken together, are insufficient to fully reconcile the observed annuity demand with the predictions of standard life-cycle models (Peijnenburg *et al.*, 2016). The second category of explanations draws on behavioral economics and cognitive biases that the standard capture. Annuitization exchanges liquid wealth for an irreversible income stream, a transaction that many households view as a potential loss. Loss aversion plays a central role in this perception. The prospect

[8]The money's worth ratio of an annuity is the expected present value of its payments, calculated using a given mortality table and discount curve, divided by its market price. In other words, it is the ratio between the actuarial fair price and the commercial market price. On average, the authors found that this ratio in the U.S. is between 80 and 85 cents per dollar of annuity premium.

[9]They only requires the absence of a bequest motive and a return advantage over conventional assets, namely the mortality credit.

of forfeiting one's capital to an insurer in the event of early death carries more weight than the gain from higher payouts conditional on survival. Hu and Scott (2007) formalize this argument with cumulative prospect theory and show that a loss-averse retiree may decline even with an actuarially fair annuity. Framing compounds this effect. Brown *et al.* (2008) show in a survey experiment that respondents favor annuities when they are framed as a guaranteed income, but not when they are framed as an investment, in terms of returns and principal. Agnew *et al.* (2008) find that gender, framing, and default settings significantly affect annuity choices. Other behavioral mechanisms, including complexity, choice overload, spending flexibility and decision avoidance, may also reduce annuity demand (Beshears *et al.*, 2014), as does limited financial literacy (Lusardi and Mitchell, 2014; Bavafa *et al.*, 2026). Ambiguity aversion regarding future survival prospects and the tendency to underestimate individual longevity risk also play a role. Taken together, these findings suggest that no single explanation resolves the annuity puzzle[10], which more likely reflects the interaction of rational, behavioral, and institutional factors.

The main alternative to annuitization is drawdown. With this method, retirees keep ownership of their accumulated wealth and finance their spending by making regular programmed withdrawals from an investment portfolio (Antolin, 2008; Banks and Crawford, 2022). Drawdown preserves liquidity, gives retirees control over their assets, and provides the flexibility to adapt their spending to changing circumstances. It also enables the transfer of residual wealth to heirs. These features make drawdown particularly attractive to households with uncertain expenditures or strong bequest motives (Lockwood, 2012; Peijnenburg *et al.*, 2017). However, this flexibility comes at the cost of greater financial risk. Unlike annuitants, retirees who rely on drawdown remain exposed to investment risk and bear the longevity risk associated with financing consumption from their own assets. For example, the portfolio could be depleted before death, especially if market performance is poor or longevity exceeds expectations (Gerrard *et al.*, 2004). A major concern is experiencing adverse investment returns in the early years of retirement when withdrawals have already begun, as this can permanently impair the sustainability of future income (Bengen, 1994, 1997). Inflation introduces an additional challenge as retirees must preserve the purchasing power of their withdrawals over a retirement that may span several decades (Antolin, 2008). The limiting case is the complete withdrawal of retirement wealth as a lump sum. The proceeds can then be spent, kept in liquid assets, or invested outside the pension plan. However, in this case, the withdrawn wealth no longer benefits from longevity-risk pooling. This makes the retiree responsible for managing investment, inflation, and longevity risks (Financial Conduct Authority, 2018). The growing importance of drawdown reflects broader institutional changes in pension systems. In addition to the shift from DB to DC plans, new regulations can change the retirement market. In the United Kingdom, for instance, the 2015 "*pension freedoms*" substantially liberalized access to DC pension wealth. This was followed by a sharp decline in annuity purchases and an increased use of drawdown arrangements (Financial Conduct Authority, 2018). A similar trend can be observed in Germany, where the 2026 pension reform will replace the Riester scheme. In 2027, the "*Altersvorsorgedepot*" will introduce greater flexibility during the payout phase by permitting long-term withdrawal plans as an alternative to mandatory lifetime annuitization. These developments have reinforced the practical relevance of substantial literature on retirement income sustainability and post-retirement portfolio management (Gerrard *et al.*, 2004; Milevsky and Huang, 2011; Bruder *et al.*, 2023). A central question in this literature concerns sustainable withdrawal rates.

[10] As emphasized by Brown *et al.* (2008), many proposed solutions merely shift the puzzle elsewhere. For example, if family risk-sharing substitutes for annuitization, annuity demand should increase following widowhood. However, there is little evidence that this occurs. Similarly, if delayed annuitization were the primary explanation, one would expect annuity purchases to substantially increase at advanced ages, but this is not generally observed.

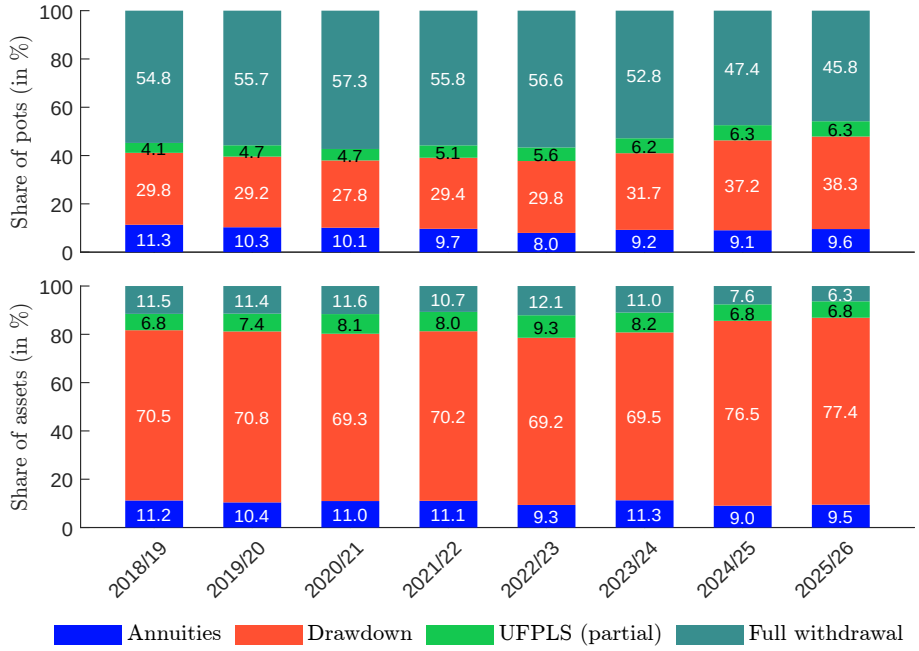
The influential *4% rule*, proposed by Bengen (1994), suggests that an initial withdrawal equal to approximately 4% of retirement wealth, adjusted for inflation, would historically avoid portfolio exhaustion over a 30-year retirement period for stock-bond portfolios. Bengen (1996) subsequently explored the impact of asset allocation, and Cooley *et al.* (1998) extended the analysis using historical simulations to estimate portfolio survival probabilities across different withdrawal rates, investment allocations, and retirement horizons. Although the 4% rule is highly influential among practitioners, it has been criticized because it does not adapt to stochastic longevity, changing market conditions, or evolving spending capacity (Scott *et al.*, 2009; Milevsky and Huang, 2011). Therefore, dynamic spending rules that adjust withdrawals based on changes in portfolio wealth, remaining planning horizons, and life expectancy have been proposed (Waring and Siegel, 2015).

Professional reports and household surveys suggest that decumulation is difficult in practice. In the United Kingdom, 77% of DC pension holders aged 40–75 who had not yet accessed their pension did not have a clear plan to do so[11] (Butt *et al.*, 2025). The difficulty persists after retirement begins. A recent U.S. survey of adults aged 45–79 with at least $100 000 in investable assets found that only 28% were comfortable with seeing their savings decline to finance retirement expenses, while 49% were uncomfortable with doing so (Corebridge Financial, 2026). Consistent with this reluctance to draw down wealth, Bavafa *et al.* (2026) use longitudinal data from the Health and Retirement Study to show that the median retiree retains about 68% of their non-housing financial wealth twelve years after age 65, while nearly 40% of retirees in the sample increased their financial wealth during the same period. The decumulation problem therefore has two dimensions. Retirees may either exhaust their wealth too quickly or preserve too much capital at the expense of current consumption. Several risks help explain this caution. In the Corebridge survey, 76% of retirees reported that concerns about persistent inflation led them to spend less than they would have liked. Other concerns include future healthcare costs (cited by 72%), long-term-care costs (69%), a major stock-market decline (68%), and the possibility of living substantially longer than expected (57%) (Corebridge Financial, 2026). Broader U.S. evidence is consistent with these concerns. The Society of Actuaries' 2024 Retirement Risk Survey found that 58% of retirees and 78% of pre-retirees were worried that their savings were not keeping pace with inflation. Additionally, 20% of retirees and 35% of pre-retirees had experienced a financial shock that reduced their assets by more than 25% (Geddes *et al.*, 2026). Long-term care is a particular concern in Europe. According to the 2024 Pension Adequacy Report, long-term care expenses for older adults with severe care needs could reach one to six times the median disposable income by 2050, and may remain unaffordable for most people, even with public support (European Commission, 2024; Pensions Europe, 2025). Some savers also underestimate their life expectancy. A 2023 U.S. study by Jackson Financial and Boston College, which surveyed 1 000 investors aged 55–84, found that more than 32% underestimated their lifetime (Pensions Europe, 2025). Holding liquid assets is a form of self-insurance against such risks. Moreover, Bavafa *et al.* (2026) show that decumulation differs sharply across asset classes. Over the twelve years following age 65, stock holdings decline by nearly 85%, IRA balances by around 50%, and checking and savings balances by only about 25%. This preference for liquidity is particularly relevant to annuitization because life annuities are usually irreversible and reduce the available capital for unexpected expenses and bequests. The OECD analysis of the payout phase highlights this trade-off: lifetime income products protect against longevity risk, while managed drawdown maintains flexibility (OECD, 2024). The OECD also finds that attitudes are more nuanced than observed annuity purchases suggest. In 15 European countries, around 80% of respon-

[11] "*This includes 56% who knew they had to make a choice but did not have a clear plan, and 21% who were not aware they had to make a choice.*" (Butt *et al.*, 2025, page 17).

dents preferred safety than performance, and 62% preferred an annuity for at least some of their retirement savings. However, preferences changed when respondents were shown their projected retirement outcomes. This illustrates how framing affects the apparent attractiveness of annuitization. In the United Kingdom, 32% of individuals aged 50–69 who were saving for retirement had never heard of a life annuity. A further 21% believed that annuities were risky products with volatile income (OECD, 2024, page 126). Therefore, low observed annuity take-up need not imply low demand for insurance against longevity risk. In fact, the demand for guaranteed income is substantial. In the 2024 DWP survey, in which respondents could select multiple objectives for their DC pension, 53% chose guaranteed lifetime income, 31% flexible lifetime income, and 12% flexible income followed by guaranteed income (Butt *et al.*, 2025, page 17). Actual choices are much more concentrated in flexible access. Among U.K. adults who accessed a DC pension in the twelve months to May 2024, 52% partially encashed a pension, 28% fully encashed a pension, and only 19% purchased an annuity (Financial Conduct Authority, 2025, page 38). Full encashment

Figure 2: Breakdown of pots and assets for the first time by method of access (U.K. data)



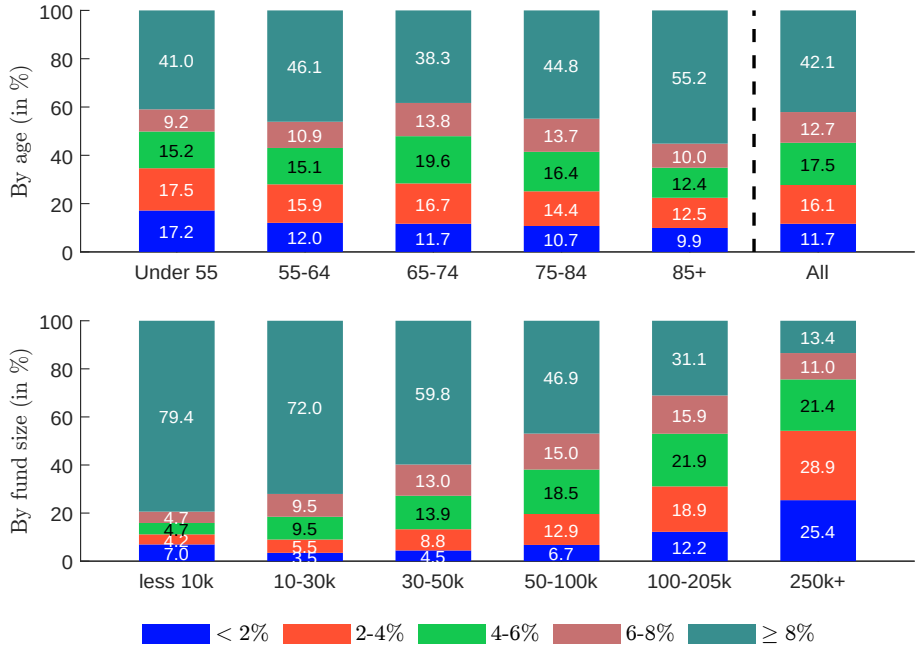
The pots accessed for the first time are split into four categories: (1) annuities purchased, (2) new drawdown policies entered into and not fully withdrawn, (3) pots from which a first partial uncrystallised funds pension lump sum (UFPLS) payment was taken and which were not fully withdrawn, and (4) full cash withdrawals from pots accessed for the first time. The upper panel measure the share of the number of pots, while the lower panel measures share of assets under administration. Fiscal years run from April to March.

Source: Financial Conduct Authority (2026), www.fca.org.uk/data/retirement-income-market-data-2025-26 & Authors' calculations.

is even more prominent when measured per pot rather than per adult. In 2025/26, it accounted for 45.8% of pots accessed for the first time, while drawdown represented 38.3% and annuities only 9.6% (Figure 2). The picture is very different when measured in assets under administration (AUA). In 2025/26, drawdown accounted for 77.4% of AUA and annuities for 9.5%, whereas full withdrawals represented only 6.3%. The gap between the

two figures is partly due to small pots, since 68% of full cash withdrawals came from pots under £10 000. This suggests that savers make short-term decisions without considering their total retirement wealth (Pensions Policy Institute, 2025). Another surprising figures is the distribution of regular withdrawal rate (Banks and Crawford, 2022). In Figure 3, we report the distribution of regular withdrawal rates by pooling the FCA data covering the period from 2017 to 2026. Although some differences are visible across age groups, these are relatively modest. By contrast, substantial heterogeneity emerges when plans are grouped by pot size. In particular, average withdrawal rates decline as fund size increases. Overall, approximately 12% of plans have a withdrawal rate below 2%, while a further 16% withdraw between 2% and 4%. At the other end of the distribution, more than 40% of plans withdraw more than 8% per year. Such withdrawal rates are difficult to reconcile with conventional

Figure 3: Annual rate of regular withdrawal by age band and fund size (U.K. data, 2017–2026)



The upper panel shows the share of the number of pots by annual withdrawal rate and age band, while the lower panel shows the share of the number of pots by annual withdrawal rate and fund size. Only pots from which the plan holders made regular partial withdrawals are considered. The exact fund-size bands are: Less than £10 000, £10 000 – £29 999, £30 000 – £49 999, £50 000 – £99 999, £100 000 – £249 999 and £250 000 and above.

Source: Financial Conduct Authority (2026), www.fca.org.uk/data/retirement-income-market-data-2025-26 & Authors' calculations.

sustainability benchmarks and may expose retirees to a significant risk of exhausting their assets prematurely. As with full encashment, however, this pattern is largely driven by the smallest pension pots, whose holders appear more likely to prioritise immediate access to wealth over long-term income planning. These choices cannot necessarily be interpreted as the result of fully informed preferences. The FCA's Retirement Outcomes Review found weak competitive pressure and low switching (Financial Conduct Authority, 2018). Many consumers followed the "*path of least resistance*" by accepting drawdown from their existing provider rather than shopping around. Approximately one-third of consumers who recently

entered drawdown were unaware of where their money was invested. Some providers defaulted consumers into cash or cash-like assets, which are generally unsuitable for long-term drawdown (Financial Conduct Authority, 2018). Thus, low annuity take-up may reflect a combination of liquidity preferences, bequest motives, imperfect information, inertia, and the complexity of comparing retirement income strategies. In other words, the institutional environment can affect observed annuitization independently of households' underlying preferences. The difficulty is compounded by limited advice and financial literacy. According to Pensions Policy Institute (2025), more than 70% of pension withdrawals occur without regulated advice or pension-wise guidance, and advice is not perceived as accessible to all. Awareness of charges is also limited. 55% of DC pension holders are unaware that fees apply, and only 12% know the exact costs. Many savers lack a structured decumulation strategy, with only 21% considering inflation risk and 21% investment risk when planning withdrawals (Pensions Policy Institute, 2025). Pensions Europe attributes these difficulties to the low level of financial literacy in Europe. According to the 2023 Eurobarometer, only 18% of EU citizens have a high level of financial literacy, 64% have a medium level and 18% have a low level (Pensions Europe, 2025). The OECD highlights that financial planning and education are particularly important during the decumulation phase when individuals must simultaneously consider longevity protection, withdrawal rates, investment risk, and sources of income (OECD, 2022a). The institutional context is also crucial. Many countries already have mandatory public and occupational pensions that provide lifetime income, which affects the marginal value of an additional private annuity. The Pension Adequacy Report documents substantial cross-country differences in adequacy (European Commission, 2024). For instance, the stronger mandatory income floor in France leaves a different role for private annuitization than in the United Kingdom, meaning that the same private pension balance can have very different implications for retirement security depending on the lifetime income supplied by the pension system.

The relative appeal of drawdown and annuitization evolves throughout retirement. Early in retirement, the flexibility of drawdown is valuable because retirees face substantial uncertainty regarding future spending needs, health care expenses, market conditions, and personal circumstances. At the same time, however, mortality credits are relatively modest at younger retirement ages, reducing the attractiveness of immediate annuitization. Nevertheless, as retirees age, the situation gradually reverses. The benefits of retaining full liquidity and investment flexibility tend to decline, while the value of longevity insurance increases as survival probabilities fall and mortality credits rise. Consequently, the choice between drawdown and annuitization should not be viewed as a binary decision made once and for all at retirement, but rather as a dynamic allocation problem whose solution may evolve over time. This observation has motivated a growing body of literature on delayed and partial annuitization, which seeks to combine the flexibility of drawdown during the early years of retirement with the income security provided by life annuities later in life. Milevsky (1998) shows that retirees with low risk aversion can benefit from delaying annuity purchases to capture the equity premium until the rising mortality credit makes annuitization attractive. Milevsky (2001) analyzes the options available to the retiree, and Milevsky and Promislow (2001) value guaranteed future annuitization options embedded in variable annuity contracts under stochastic mortality and interest rates. Milevsky and Young (2007) formalize the problem by jointly determining the optimal asset allocation and the optimal annuitization time, and Stabile (2006) treats it as a combined stochastic control and optimal stopping problem. Milevsky *et al.* (2006) study the same trade-off when the objective is to minimize the probability of financial ruin, while Horneff *et al.* (2008) compare single-date conversion with gradual annuitization and quantify the cost of switching to annuities over time. Delay can also be built into the product itself, as in the advanced life deferred annuity

of Gong and Webb (2010). From a practitioner perspective, Boyle and Webb (2022) advocate a "*flex first, fix later*" strategy, in which retirees remain flexible and switch to an annuity in their late seventies or early eighties. Clare *et al.* (2024) study this strategy for the U.K. and find, in their baseline calibration, a crossover from drawdown to annuitization at about age 73. However, this age is sensitive to market conditions and risk preferences[12]. Building on these works, we study a *flex-to-fix* strategy, in which the retiree remains invested and liquid during a flexible period before converting part or all of their wealth into an annuity at a later date. The strategy encompasses two limiting cases: immediate annuitization (*fix-now*) and permanent drawdown (*fix-never*). Our contribution is threefold. First, a static actuarial analysis shows how the optimal conversion date depends on the bequest motive, the withdrawal rate, and the return of the drawdown portfolio. We find that there is no single optimal annuitization age, although optimal dates tend to concentrate around retirement age (65) or between 75 and 80. Second, we develop two dynamic formulations: an optimal-stopping problem, which assesses the value gained by making the conversion date state-contingent, and a Hamilton-Jacobi-Bellman variational inequality (HJBVI), which characterizes the free boundary that governs the gradual transition from drawdown to annuity income. We find that a well-chosen fixed date already captures most of the value of delaying annuitization, while the HJBVI formulation reveals an implicit bond effect of annuities on the optimal equity allocation and shows that spreading purchases over time substantially reduces market timing and ruin risk compared to a single-date conversion. Third, we propose a simple rule that approximates the HJBVI strategy by spreading annuity purchases evenly across the flex window once the target date has been calibrated.

The remainder of the paper is organized as follows. Section Two introduces the mortality modeling framework. First, we examine the mortality data and emphasize the importance of data cleaning and preprocessing. Then, we present the Lee-Carter (LC) model for central death rate modeling, as well as two extensions based on Poisson and negative binomial generalized linear models. We also consider the Cairns-Blake-Dowd (CBD) model. After discussing how to calibrate these models, we show how to use them to forecast mortality rates and survival probabilities. Finally, we evaluate the models' performance using in-sample and out-of-sample backtesting exercises. Section Three focuses on annuity modeling. First, we analyze the pricing of life annuities under deterministic and stochastic mortality assumptions. Then, we investigate the sensitivity of annuity values to the underlying mortality model. Then, we examine the effects of discount rates, discuss mortality risk premia valuation, and introduce mortality credits. The second part of the section is devoted to portfolio risk analysis. After presenting a framework for modeling annuity liabilities, we develop risk measurement tools and quantify the relative contributions of systematic and idiosyncratic mortality risk. Section Four studies optimal retirement decumulation strategies by comparing three approaches: immediate annuitization (*fix-now*) which serves as the benchmark, permanent drawdown (*fix-never*), and the hybrid *flex-to-fix* strategy, which combines the first two. This section is divided into static and dynamic analyses. In the static analysis, we investigate two methods for determining the optimal annuitization date ex ante. One method uses mortality credits, and the other uses actuarial valuation principles. The dynamic analysis considers a discrete-time Bellman framework with fixed-date annuitization and a continuous-time portfolio optimization framework with withdrawals and endogenous annuitization decisions. Numerical experiments and simulation results illustrate the economic rationale behind these strategies and compare their performance with that of alternative rule-based approaches. Finally, Section Five concludes and discusses potential directions for future research.

[12]It moves to around age 78 when the lower annuity rates observed in February 2020 are used, and it ranges from around ages 67–68 for highly risk-averse retirees to about 78.5 for risk-indifferent retirees.

2 Mortality modeling

Modeling human mortality requires high-quality data and appropriate statistical models, particularly when the objective is to analyze mortality trends and forecast future mortality rates. Below, we mainly rely on the Human Mortality Database (HMD), which is widely regarded as one of the most comprehensive and reliable sources of mortality data. We then introduce two standard classes of stochastic mortality models: the Lee-Carter (LC) model and the Cairns-Blake-Dowd (CBD) model, along with their principal extensions. These models provide a statistical framework for capturing age- and time-related dynamics of mortality and are widely used in actuarial science and pension risk management. Finally, we discuss the forecasting methodologies and techniques used to project future mortality patterns and assess the sources of uncertainty in these projections.

2.1 Mortality data

The choice of mortality database is critical, since different sources vary substantially in terms of granularity, geographic coverage, historical depth, and update frequency. Below, we list some examples of country-specific and international databases:

- Continuous Mortality Investigation (CMI) is a UK research organization offering mortality projections widely used by UK pension funds and life insurers.
- The INSEE database provides official French mortality rates and life tables from 1957 to the present, calculated by the French national statistical institute.
- The Human Mortality Database (HMD) is maintained by the Max Planck Institute for Demographic Research (Germany), the University of California, Berkeley (USA), and the French Institute for Demographic Studies (France). It provides harmonized death counts, population exposure, and life tables for approximately 40 countries and regions.
- The WHO Mortality Database, maintained by the World Health Organization, is a compilation of mortality data by country, year, sex, age and cause of death from 1950 to the present.
- World Population Prospects (WPP) is produced by the United Nations Population Division. It is a global database covering more than 200 countries from 1950 to 2100.

Beyond these major international sources, many national and supranational statistical organizations (*e.g.*, OECD, Eurostat) maintain their own mortality datasets.

2.1.1 Data description

In this research paper, we use the HMD because its data files are the result of a harmonization procedure designed to create comparable mortality schedules over time and between countries (HMD, 2026). In particular, the dataset contains two important statistics[13]:

- The period deaths variable $D(x,t)$ denotes the number of deaths at age x during year t;
- The period exposure-to-risk variable $E(x,t)$ denotes exposure-to-risk at age x during year t, that is, the population at risk expressed in person-years over the calendar year.

[13] See Appendix B.1 on page 136 for more details about the HMD.

Using these two quantities, we can compute the central death rate at age x during year t:

$$m(x,t) = \frac{D(x,t)}{E(x,t)}$$

This quantity is central to mortality modeling, as it forms the basic input series on which stochastic mortality models are fitted. The table below illustrates the historical coverage available in the HMD and provides an example of how to compute $m(75, 2015)$:

	France	UK	Italy	Germany	USA	Japan
Period	1816–2023	1922–2022	1872–2023	1990–2020	1933–2024	1947–2024
$D(75, 2015)$	9 441.00	12 916.00	13 977.00	26 414.64	54 432.84	25 676.97
$E(75, 2015)$	454 188.97	458 855.33	614 300.05	971 203.48	1 824 538.24	1 343 110.93
$m(75, 2015)$	2.08%	2.81%	2.28%	2.72%	2.98%	1.91%

To interpret these figures, we consider the example of France. $D(75, 2015) = 9\,441$ means that 9 441 people, aged 75 but not yet 76 at the time of death, died in France in 2015. $E(75, 2015) = 454\,188.97$ means that the equivalent[14] of 454 188.97 people were exposed to the risk of dying at age 75 in 2015.

Remark 1 *The exposure-to-risk $E(x,t)$ is closely related to population size $P(x,t)$, which is the number of individuals aged x on January 1st of year t. In the example of France, we have $P(75, 2015) = 455\,726$ versus $E(75, 2015) = 454\,188.97$. In fact, $E(x,t)$ takes into account people entering or leaving age x during year t.*

Table 3: Central death rate $m(x,t)$ in % ($x = 75$ years)

t	France	UK	Italy	Germany	USA	Japan
1900	11.44		12.62			
1925	9.28	9.48	9.17			
1950	7.01	7.69	6.67		6.70	8.64
1980	4.64	5.71	5.26		4.73	4.84
2000	2.93	4.11	3.30	3.58	3.83	2.61
2010	2.24	2.93	2.45	3.01	3.09	2.18
2020	2.20	2.90	2.52	2.76	3.25	1.89
2021	2.15	2.71	2.41		3.30	1.83

Table 3 presents the central death rate $m(x,t)$ when x is set to 75 years. While the period covered by the HMD differs from one country to another, the age variable is standardized and can take 110 finite single-year values ($x = 0, 1, \ldots, 109$) and one open-ended value ($x = 110+$). Moreover, the data are available for three categories: female, male and total. We observe that the central death rate at age 75 has decreased dramatically since 1900. For instance, the rate was equal to 11.44% in 1900 in France, compared to 2.15% in 2021. There are also significant differences between countries. While the US death rate $m(75,t)$ was below the equivalent French, British, and Japanese rates in 1950, it is now the highest among the six countries considered. In addition, the decline in $m(75,t)$ slowed markedly since 2010 and even reversed slightly in some countries. In the US, for example, the rate rose from 3.09% to 3.30%. Table 4 presents the central death rate $m(x,t)$ for 2015. We

[14] The exposure-to-risk uses pro-rata rules. For instance, if one person turned 75 at the beginning of 2015, and lived until the end of the year, he contributes 1.0 to the exposure. If he turned 75 on July 1st, he contributes 0.5. A person who was already 75 at the start of 2015 and died on April 1st, 2015 contributes 0.25 person-years

observe a relative high value for $x = 0$, corresponding to infant mortality, followed by a sharp decrease during childhood. The death rate then increases monotonically with age, gradually at first and accelerating from around age 60.

Table 4: Central death rate $m(x,t)$ in % ($t = 2015$)

x	France	UK	Italy	Germany	USA	Japan
0	0.36	0.39	0.29	0.33	0.61	0.20
1	0.03	0.03	0.02	0.03	0.04	0.04
10	0.01	0.01	0.01	0.01	0.01	0.01
50	0.30	0.28	0.21	0.28	0.41	0.21
60	0.71	0.67	0.55	0.78	0.90	0.50
70	1.38	1.62	1.37	1.80	1.87	1.25
80	3.65	4.95	4.31	5.04	4.83	3.55
90	13.76	16.98	16.37	17.84	15.41	12.15

2.1.2 Data cleaning

The HMD uses several sources of raw data[15], combines them to produce standardized time series, and applies filtering procedures to clean the dataset. However, the harmonized HMD time series may still exhibit irregularities, in particular at older ages, for several reasons. First, age misreporting in the early twentieth century can lead to underestimated mortality rates at very high ages. As data quality improves, this may create a spurious increase in observed old-age mortality over time (Wilmoth and Lundström, 1996). Second, exposure-to-risk is smaller and death counts are sparser at these ages. Therefore, individual age-year cells are more sensitive to reporting errors and random annual fluctuations, resulting in noisier estimated rates[16]. For these reasons, we apply a cleaning procedure to detect outliers and smooth mortality rates before further analysis[17] The procedure may also attenuate temporary mortality shocks associated with exceptional events such as wars or epidemics. Since the models considered here do not include a separate mechanism for such event-specific shocks, we treat these episodes as transitory disturbances when estimating the underlying mortality dynamics. Since our objective is to model and forecast mortality, it is important to work with a dataset in which these irregularities have been addressed.

2.2 Lee-Carter model

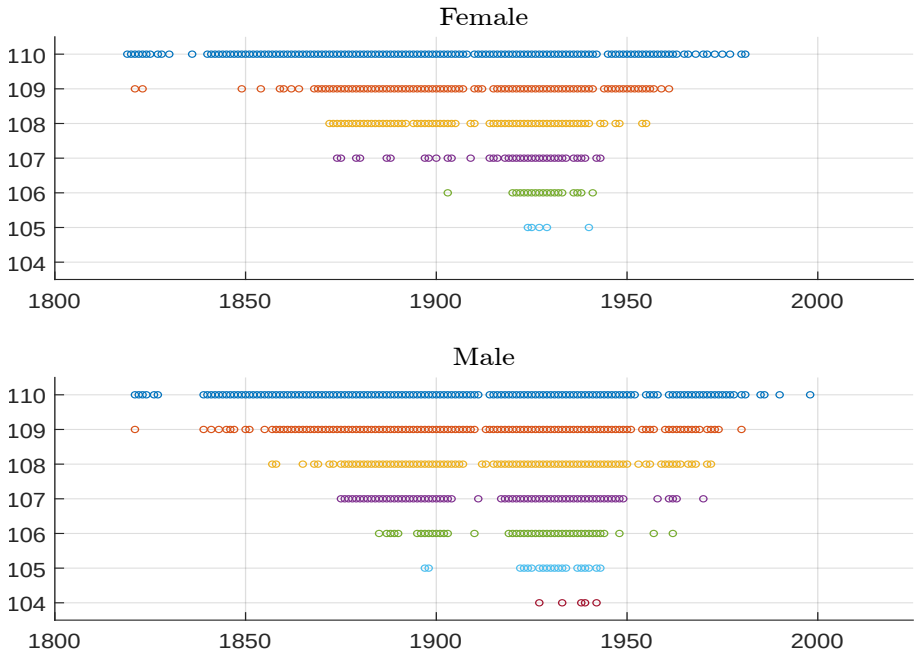
The model of Lee and Carter (1992) is certainly the most well-known stochastic model used to estimate and forecast mortality rates. Since then, the model has been expanded in several ways. Most notably, the original Gaussian error assumption has been replaced with models in which death counts follow a Poisson or negative binomial distribution. This allows for a statistically better treatment of small death counts at high ages. Below, we present these models, explain how they are calibrated, and report empirical results based on the HMD.

[15] For example, the HMD series for France uses several historical and contemporary sources, including official French statistical publications, historical mortality reconstructions, the Vallin-Meslé historical mortality tables, and modern INSEE population estimates. A crucial feature of the French HMD series is that territorial coverage is not artificially held constant over the entire historical sample. Each annual observation refers to the statistical territory covered by French national statistics at that time. For modern years, the HMD continues to use *France métropolitaine* in order to maintain continuity from 1946 onward.

[16] This is particularly true for older-age cells, since exposure-to-risk is very low. For example, Figure 4 shows the years for which exposure is exactly equal to zero for ages above 100.

[17] See Appendix B.2 on page 136 for a description of the cleaning procedure.

Figure 4: Zero exposure-to-risk (France)



2.2.1 Model specification and calibration

Lee and Carter (1992) assume that the log-mortality rate is given by:

$$\ln m(x,t) = \alpha_x + \beta_x \kappa_t + \varepsilon(x,t) \tag{1}$$

where α_x is the age-specific average log-mortality level, β_x is the age-specific loading on the period factor, κ_t is the time-varying index, and $\varepsilon(x,t) \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ is the residual error. The parameters θ of the model are[18]:

$$\theta = \{\alpha_x, \beta_x, \kappa_t : x = 0, 1, \ldots, n-1; t = 1, \ldots, T\}$$

This implies that the model summarizes nT observations into $2n + T$ parameters. For example, if $n = 110$ and $T = 208$ (the case of France), there are 22 880 observations, while the number of parameters is 428. This corresponds to a reduction in the degree of freedom by a factor of approximately 53. Once these parameters are estimated, we obtain the fitted mortality surface:

$$\hat{m}(x,t) = \exp\left(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t\right) \tag{2}$$

The time index κ_t can then be extrapolated using a time series model to forecast mortality beyond the sample period. For a forecast horizon $h > 0$, the mortality rate at age x and year $T + h$ is given by:

$$\hat{m}(x, T+h) = \exp\left(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_{T+h}\right) \tag{3}$$

[18] For notational convenience, we use the integer time scale $t = 1, \ldots, T$ rather than the calendar year scale $t = 1900, \ldots, 2023$.

Figure 5: Lee-Carter model estimation (female)

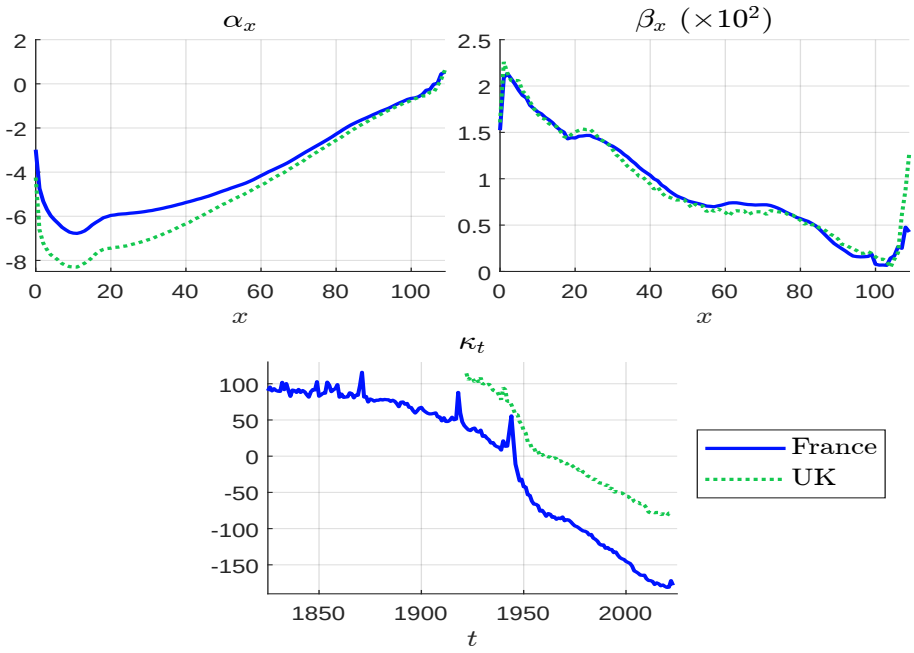
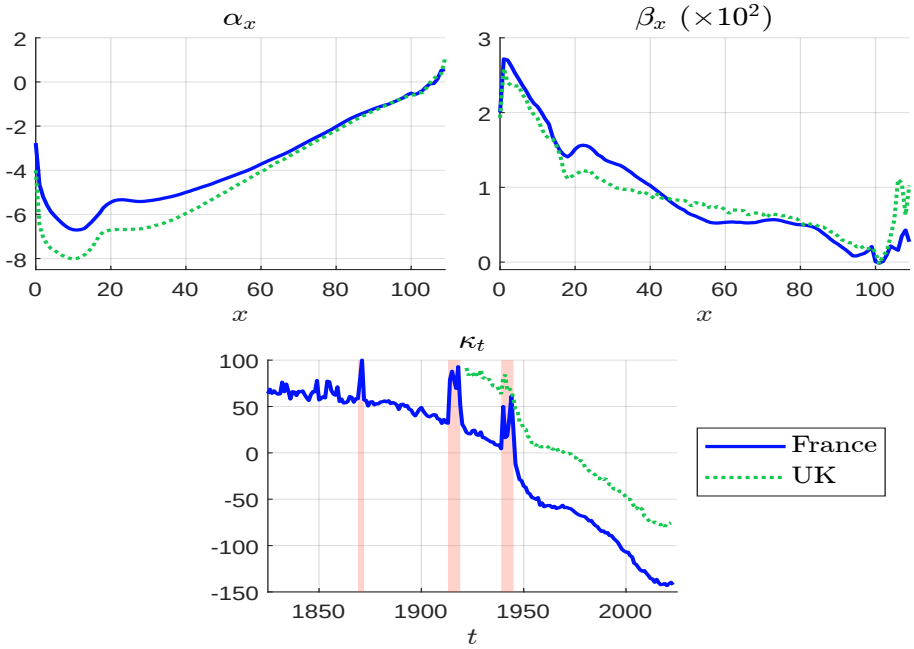


Figure 6: Lee-Carter model estimation (male)



There are two popular methods[19] for estimating the Lee-Carter model. The first, proposed by Lee and Carter (1992), uses the singular value decomposition (SVD). One drawback of this method is that the $n \times T$ matrix A, whose entries are $\ln m(x,t)$, must be fully specified. In particular, if $D(x,t)$ or $E(x,t)$ is equal to zero for some values of x and t, $\ln m(x,t)$ is not a number. This requires removing the corresponding rows (ages), even if only a few cells are affected. The second approach, proposed by Wilmoth (1993), consists in estimating the parameters using weighted least squares (WLS):

$$\hat{\theta} = \arg\min \sum_{x=0}^{n-1} \sum_{t=1}^{T} w_{x,t} \left(\ln m(x,t) - \alpha_x - \beta_x \kappa_t\right)^2$$

In this case, we can set $w_{x,t} = 0$ if $\ln m(x,t)$ is missing, and $w_{x,t} = 1$ otherwise. In the absence of missing values, the two methods produce identical results. In practice, when some values are missing, the two estimates remain very close. We verify this in the case of France (see Figures 71 and 72 on page 164). The main advantage of the WLS method is that it allows us to estimate α_x and β_x even if $\ln m(x,t)$ is missing for some values of x, without discarding entire ages from the estimation.

Figures 5 and 6 show the WLS estimates for France and the UK for the female and male groups. We note differences between females and males, and also between France and the UK. Nevertheless, there are also many common patterns. For instance, the correlation of the α_x series exceeds 99% in all cases. It is 99.9% and 99.8% between males and females in the UK and France, respectively, and 99.5% and 99.4% between France and the UK for the male and female groups, respectively. The correlation is also very high for the time index κ_t, ranging from 97% to 99.5%. The correlation is lowest for the loading coefficient β_x, ranging from 92% to 98%. The Lee-Carter model reproduces some well-known patterns in mortality. For example, mortality decreases over time, which is reflected by the downward slope of κ_t. However, we note three peaks in κ_t, corresponding to three wars: the Franco-Prussian War (1870–1871), the First World War (1914–1918), and the Second World War (1939–1945). Nevertheless, a form of plateau emerges after 2010, consistent with the slowdown in old-age mortality improvement discussed earlier. The dependence to age has two components. α_x shows that infant mortality is relatively high, and then decreases through childhood before rising again with age. α_x can thus be related to the average mortality level at each age. β_x is more difficult to interpret because it interacts with κ_t. It measures the sensitivity of each age to the general trend in mortality captured by κ_t. Globally, β_x tends to be higher at young ages ($x \leq 5$ years), indicating that these ages are more responsive to changes in the overall mortality trend[20].

Remark 2 *Comparing the fitted and observed values of $m(x,t)$ shows that the Lee-Carter model closely captures the overall shape of the mortality surface (Figures 7 and 8). However, the model fits females better than males. This is confirmed by the residual sum of squares. Indeed,* $\text{RSS}\left(\hat{\theta}\right)$ *is equal to* 816.97 *for females compared to* 1305.24 *for males. This represents a reduction of* 37.5% *in residual error for the female group. We have also computed the mean absolute errors* $\text{MAE}(t) = \frac{1}{n}\sum_x \left|\hat{m}(x,t) - m(x,t)\right|$ *and* $\text{MAE}(x) = \frac{1}{T}\sum_t \left|\hat{m}(x,t) - m(x,t)\right|$*, and the mean absolute percentage errors* $\text{MAPE}(t)$ *and* $\text{MAPE}(x)$ *where the absolute error* $\left|\hat{m}(x,t) - m(x,t)\right|$ *is replaced by the absolute percentage error*

[19] They are presented in Appendix C.1 on page 140.

[20] We also observe an increase of β_x at very old ages ($x \geq 100$ years).

Figure 7: Fitted vs. observed values of $m(x,t)$ (France, female)

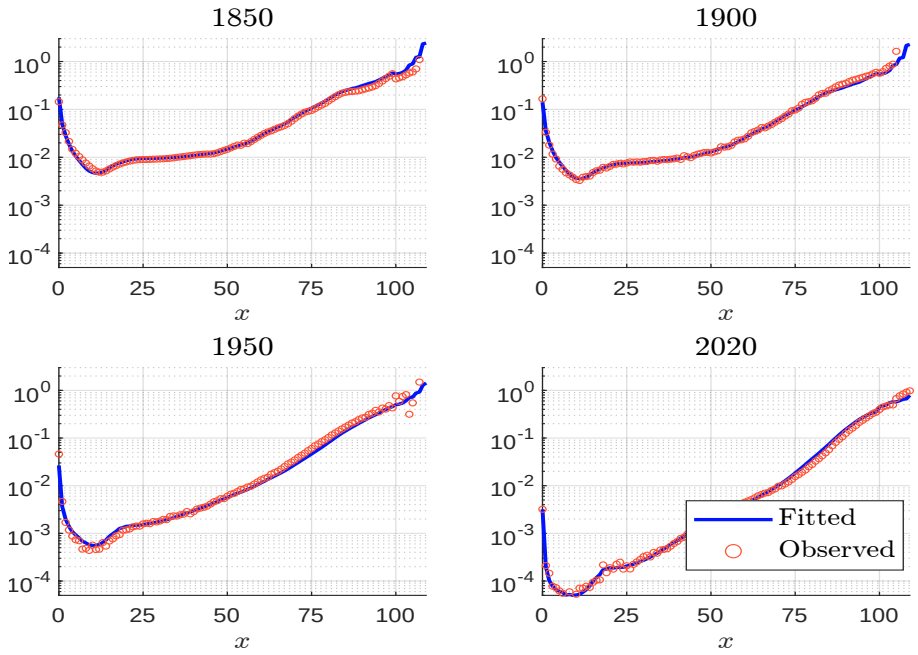


Figure 8: Fitted vs. observed values of $m(x,t)$ (France, male)

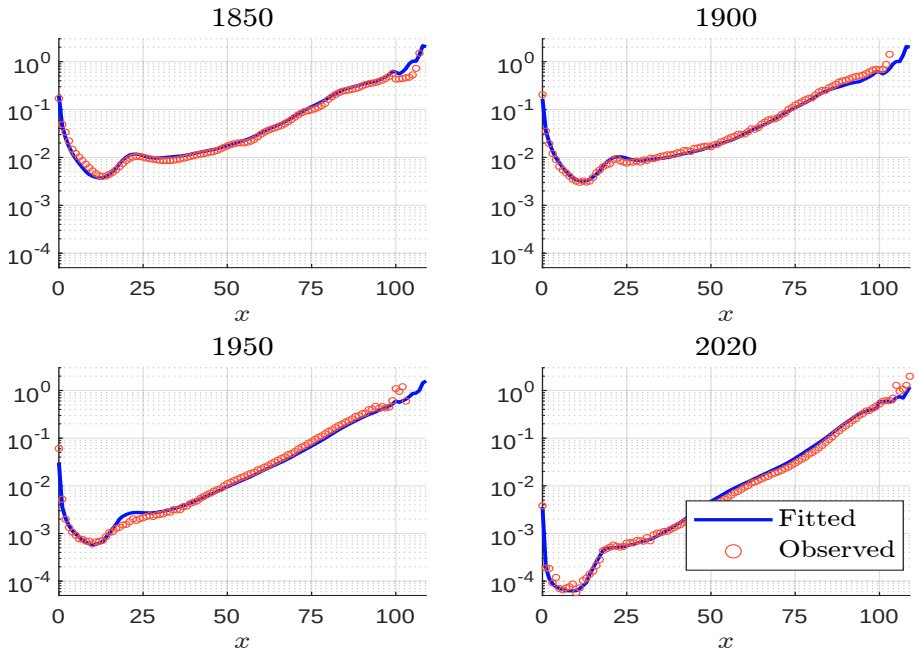


Figure 9: Mean absolute error of the central death rate (France, Lee-Carter model)

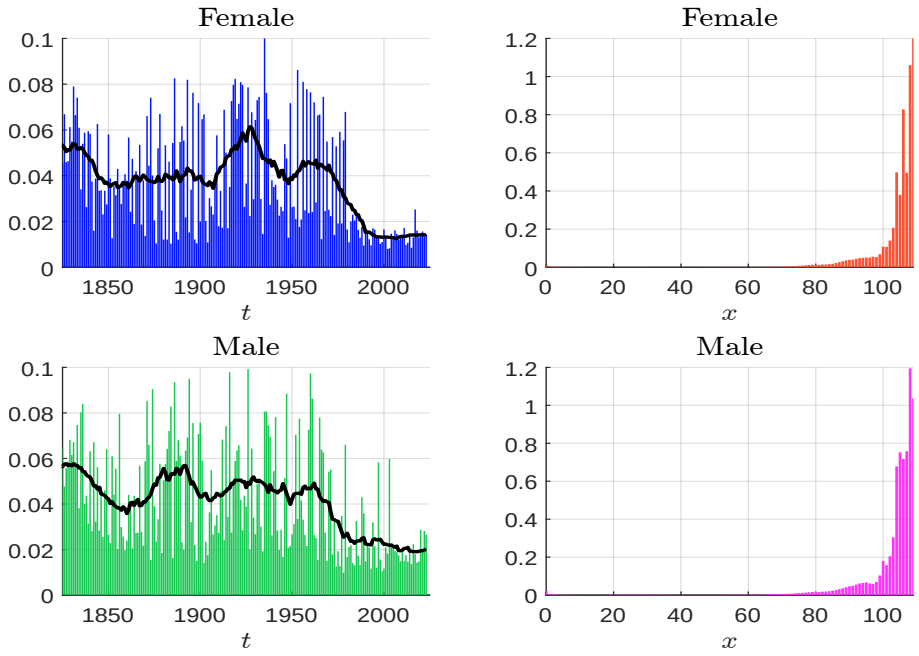
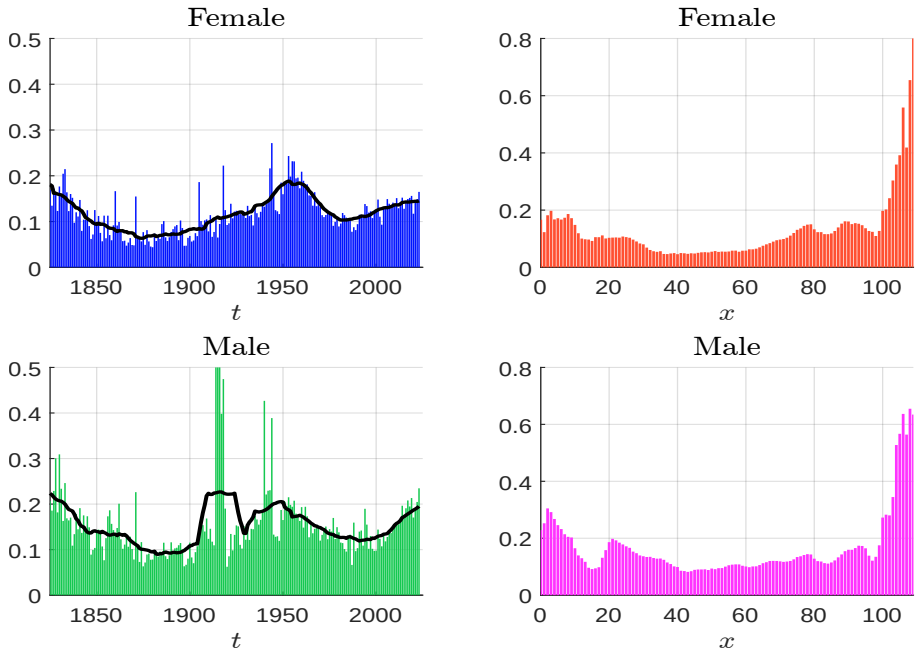


Figure 10: Mean absolute percentage error of the central death rate (France, Lee-Carter model)



$\left|\dfrac{\hat{m}(x,t)-m(x,t)}{m(x,t)}\right|$. *The results are shown in Figures 10 and 9. Since* $\mathrm{MAE}(t)$ *and* $\mathrm{MAPE}(t)$ *are relatively erratic, we also report the 15-year moving average (black line). Depending on the metric, the story differs. We observe that* $\mathrm{MAE}(t)$ *has decreased in recent years, while* $\mathrm{MAPE}(t)$ *has increased in recent years but remains, on average, at 12% for females and 13% for males. This apparent contradiction is explained by the general decline in mortality rates* $m(x,t)$ *over time. The plots of* $\mathrm{MAE}(x)$ *and* $\mathrm{MAPE}(x)$ *are more informative because they display greater variability. In particular, we can see that both errors increase dramatically at older ages (above 100 years). This is consistent with the data sparsity and reporting issues at very old ages. Further reducing in-sample error may therefore amount to fitting observation noise rather than improving the mortality signal.*

2.2.2 Extension of the Lee-Carter model

The Poisson LC model Wilmoth (1993) and Alho (2000) suggested extending the Lee-Carter model by considering a Poisson generalized linear model. Brouhns *et al.* (2002) formalized this Poisson LC model by assuming that death counts follow a Poisson distribution with parameter $\mu(x,t)$:

$$D(x,t) \sim \mathcal{P}\left(\mu(x,t)\right)$$

Since $D(x,t) = m(x,t)\,E(x,t)$, and assuming that $\mathbb{E}\left[m(x,t)\right] = e^{\alpha_x+\beta_x\kappa_t}$, we deduce that:

$$\mu(x,t) = \mathbb{E}\left[D(x,t)\right] = \mathbb{E}\left[m(x,t)\,E(x,t)\right] = E(x,t)\,\mathbb{E}\left[m(x,t)\right] = E(x,t)\,e^{\alpha_x+\beta_x\kappa_t}$$

It follows that:

$$\ln\mu(x,t) = \ln E(x,t) + \alpha_x + \beta_x\kappa_t$$

The parameters have the same interpretation as in the original Lee-Carter model. α_x is the age-specific mortality level, β_x is the age-specific loading on the period factor, and κ_t is the common time index. Thus, the Poisson specification preserves the bilinear age-period structure of the original model, while replacing the Gaussian log-mortality assumption with a count likelihood that is consistent with the discrete nature of deaths and exposures.

Remark 3 *In the original Lee-Carter model,* $\ln m(x,t) = \alpha_x+\beta_x\kappa_t+\varepsilon(x,t)$ *where* $\varepsilon(x,t) \sim \mathcal{N}\left(0,\sigma_\varepsilon^2\right)$. *Since* $m(x,t)$ *is a log-normal random variable, the mathematical expectations is* $\mathbb{E}\left[m(x,t)\right] = \exp\left(\alpha_x+\beta_x\kappa_t+\dfrac{1}{2}\sigma_\varepsilon^2\right)$. *This shows that the assumption* $\mathbb{E}\left[m(x,t)\right] = e^{\alpha_x+\beta_x\kappa_t}$ *is not derived from the Gaussian LC model but is instead a modeling choice.*

The estimation of the parameters is based on the log-likelihood function of the Poisson distribution. Let $\{k_1,\ldots,k_n\}$ be a sample of counts. We recall that the Poisson log-likelihood function takes the form:

$$\boldsymbol{\ell}(\lambda) = \sum_{i=1}^{n}\ln\frac{\lambda^{k_i}e^{-\lambda}}{k_i!} = \sum_{i=1}^{n}\left(-\lambda+k_i\ln\lambda-\ln k_i!\right)$$

Applying this formula to the death counts $D(x,t)$ with $\lambda=\mu(x,t)$ gives:

$$\begin{aligned}
\boldsymbol{\ell}(\theta) &= \sum_{x,t}\left(-\mu(x,t)+D(x,t)\ln\mu(x,t)-\ln D(x,t)!\right)\\
&= \sum_{x,t}\left(-E(x,t)\,e^{\alpha_x+\beta_x\kappa_t}+D(x,t)\left(\ln E(x,t)+\alpha_x+\beta_x\kappa_t\right)-\ln D(x,t)!\right)\\
&= \sum_{x,t}\left(D(x,t)\left(\alpha_x+\beta_x\kappa_t\right)-E(x,t)\,e^{\alpha_x+\beta_x\kappa_t}\right)+\underbrace{\sum_{x,t}\left(D(x,t)\ln E(x,t)-\ln D(x,t)!\right)}_{\text{this term does not depend on }\theta}
\end{aligned}$$

where $\theta = (\alpha_x, \beta_x, \kappa_t)$. The estimated parameters are then obtained by maximising the log-likelihood function subject to the normalization constraints:

$$\begin{aligned}\hat{\theta} &= \arg\max \sum_{x=0}^{n-1}\sum_{t=1}^{T}\left(D(x,t)(\alpha_x+\beta_x\kappa_t) - E(x,t)\,e^{\alpha_x+\beta_x\kappa_t}\right)\\ &\text{s.t.} \quad \sum_{x=0}^{n-1}\beta_x = 1 \text{ and } \sum_{t=1}^{T}\kappa_t = 0\end{aligned}$$

Empirical results and a comparison with the original Gaussian LC model are presented in the Appendix (Figures 73 and 74 on page 165). The results suggest that there are only minor differences in the age-specific mortality level parameter α_x. In contrast, significant differences are observed in the common time index κ_t, particularly at older ages, as well as in the age-specific loading parameter β_x. Comparing the mean absolute error of the central death rate reveals that the Poisson model provides a better fit at older ages (Figure 75 on page 166). However, for data since 1950, the Gaussian model generally outperforms the Poisson model.

The negative binomial LC model In the Poisson LC model, death counts $D(x,t)$ follow a Poisson distribution $\mathcal{P}(\lambda)$ with parameter $\lambda = \mu(x,t)$. This implies the equidispersion property: $\mathbb{E}\left[D(x,t)\right] = \text{var}\left(D(x,t)\right)$. In count theory, a natural competitor to the Poisson distribution is the negative binomial distribution, *i.e.*, a Poisson-gamma mixture: $\mathcal{NB}(r,p) \sim \mathcal{P}(\Lambda)$ and $\Lambda \sim \mathcal{G}(\alpha,\beta)$ where $\alpha = r$ and $\beta = (1-p)/p$. This means that the Poisson parameter becomes stochastic and follows a gamma distribution. Since $\mathbb{E}\left[\mathcal{NB}(r,p)\right] = \dfrac{pr}{(1-p)}$ and $\text{var}\left(\mathcal{NB}(r,p)\right) = \dfrac{pr}{(1-p)^2}$, we obtain the overdispersion property:

$$\text{var}\left(\mathcal{NB}(r,p)\right) = \frac{1}{1-p}\mathbb{E}\left[\mathcal{NB}(r,p)\right] \geq \mathbb{E}\left[\mathcal{NB}(r,p)\right]$$

If we set $\mu = \mathbb{E}\left[\mathcal{NB}(r,p)\right]$ and $\delta = 1/r$, we deduce that:

$$\text{var}\left(\mathcal{NB}(r,p)\right) = \frac{pr}{1-p} + \delta\frac{p^2r^2}{(1-p)^2} = \mu + \delta\mu^2$$

The probability mass function of the negative binomial distribution is:

$$p(k) = \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^r p^k$$

Using $p = \dfrac{\mu}{r+\mu} = \dfrac{\delta\mu}{1+\delta\mu}$, another expression of the pmf is:

$$p(k) = \frac{\Gamma\left(\frac{1}{\delta}+k\right)}{k!\Gamma\left(\frac{1}{\delta}\right)}\left(\frac{1}{1+\delta\mu}\right)^{1/\delta}\left(\frac{\delta\mu}{1+\delta\mu}\right)^k$$

Delwarde *et al.* (2007) propose modeling death counts with this distribution:

$$D(x,t) \sim \mathcal{NB}\left(\frac{1}{\delta}, \frac{\delta\mu(x,t)}{1+\delta\mu(x,t)}\right)$$

where $\mu(x,t) = E(x,t)\exp(\alpha_x + \beta_x\kappa_t)$. The vector of parameters is $\theta = (\alpha_x, \beta_x, \kappa_t, \delta)$. The negative binomial LC model has one more parameter than the Poisson LC model. The estimated parameters are calibrated by maximizing the log-likelihood function:

$$\begin{aligned}\hat{\theta} &= \arg\max \sum_{x=0}^{n-1}\sum_{t=1}^{T} \ell(x,t,\theta) \\ &\text{s.t.} \quad \sum_{x=0}^{n-1}\beta_x = 1 \text{ and } \sum_{t=1}^{T}\kappa_t = 0\end{aligned}$$

where[21]:

$$\begin{aligned}\ell(x,t,\theta) &= \ln \frac{\Gamma\left(\frac{1}{\delta} + D(x,t)\right)}{D(x,t)!\Gamma\left(\frac{1}{\delta}\right)} \left(\frac{1}{1+\delta\mu(x,t)}\right)^{1/\delta} \left(\frac{\delta\mu(x,t)}{1+\delta\mu(x,t)}\right)^{D(x,t)} \\ &= \ln\Gamma\left(\frac{1}{\delta} + D(x,t)\right) - \ln\Gamma\left(D(x,t)+1\right) - \ln\Gamma\left(\frac{1}{\delta}\right) - \\ &\quad \left(\frac{1}{\delta} + D(x,t)\right)\ln\left(1+\delta\mu(x,t)\right) + D(x,t)\ln\left(\delta\mu(x,t)\right) \end{aligned} \tag{4}$$

Results are reported in Figures 76, 77 and 78 on pages 166–167. Curiously, the results are similar to those of the original Lee-Carter model. However, the negative binomial specification is used because substantial overdispersion was observed in the death counts relative to the Poisson Lee-Carter model. Under Poisson equidispersion, the Pearson residual dispersion[22] ϕ should be close to one. In the fitted data, however, the estimated residual dispersion is 36.7 for females and 169.2 for males. These values are far above one and therefore indicate strong extra-Poisson variability in both sexes. After re-estimating the Lee-Carter model under the negative binomial specification, the residual dispersion falls to about 1.04 for females and 0.99 for males. Thus, the negative binomial model absorbs more than 95% of the Poisson residual overdispersion. This is precisely the type of empirical pattern for which Delwarde *et al.* (2007) proposed the negative binomial extension. In fact, we obtain $\delta = 1.63\%$ and $\delta = 3.86\%$ for the female and male groups, respectively. The larger value of δ for males implies stronger variance inflation relative to the Poisson benchmark. Defining the variance inflation factor (VIF) as the ratio of the negative binomial variance to the Poisson variance, we obtain an average VIF of about 54 for females and 139 for males, with medians of about 37 and 96, respectively. The difference between the mean and median VIFs reflects the fact that $\mu(x,t)$ increases with the fitted death count. Cells with larger populations exhibit far greater absolute overdispersion, even though δ remains constant across ages and years. In many such cells, the negative binomial variance is several times greater than the Poisson variance.

[21] We replace $D(x,t)!$ by $\Gamma\left(D(x,t)+1\right)$ because $D(x,t)$ may be a non-integer.

[22] We have:

$$\phi = \frac{1}{\text{df}}\sum_{x,t}\frac{\left(D(x,t)-\mu(x,t)\right)^2}{\text{var}\left(D(x,t)\right)}$$

where $\text{df} = nT-(2n+T)-2$ and $\text{var}\left(D(x,t)\right) = \mu(x,t)$ for the Poisson model, and $\text{df} = nT-(2n+T+1)-2$ and $\text{var}\left(D(x,t)\right) = \mu(x,t) + \delta\mu^2(x,t)$ for the negative binomial model.

2.3 Cairns-Blake-Dowd model

2.3.1 Model specification

The Cairns-Blake-Dowd (CBD) model was developed for the valuation and securitization of mortality risk (Cairns *et al.*, 2006a,b). Rather than aiming for the best possible fit across the full age range, the model was designed to focus specifically on old-age mortality, particularly for pricing financial instruments such as annuities and retirement insurance products. In what follows, we consider the parameterization proposed by Cairns *et al.* (2009). Unlike the Lee-Carter model, which models the central death rate $m(x,t)$ on the log scale, the CBD model works with the one-year death probability $q(x,t)$. More specifically, for each age x and year t, the basic CBD specification is:

$$\ln\left(\frac{q(x,t)}{1-q(x,t)}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \varepsilon(x,t)$$

where $\kappa_t^{(1)}$ controls the overall mortality level at year t, $\kappa_t^{(2)}$ controls the slope of mortality with respect to age at year t, and $\varepsilon(x,t)$ is the residual error. The age covariate is centered around its mean $\bar{x}$. In contrast with the Lee-Carter model, the CBD model includes no age-specific parameters. In fact, the age dependence is imposed through a linear function of x. This makes the model parsimonious and interpretable, but also more restrictive than the Lee-Carter model in its ability to reproduce the full shape of the observed mortality curve.

Remark 4 *The original CBD model assumes that the realized mortality rate is:*

$$q(x,t) = \frac{\exp\left(A_t^{(1)} + A_t^{(2)}(x+t)\right)}{1+\exp\left(A_t^{(1)} + A_t^{(2)}(x+t)\right)}$$

We have the following correspondence: $\kappa_t^{(1)} = A_t^{(1)} + A_t^{(2)}(\bar{x}+t)$ *and* $\kappa_t^{(2)} = A_t^{(2)}$.

Remark 5 *The LC model uses the central death rate* $m(x,t)$*, while the CBD model uses the one-year death probability* $q(x,t)$*. Cairns et al. (2009) define the force of mortality* $\lambda(x,t)$ *as the instantaneous death rate at exact time* t *for an individual aged exactly* x *at time* t*. In survival analysis literature,* $\lambda(x,t)$ *is more commonly known as the hazard function. Under the standard assumption that the force of mortality* $\lambda(x,t)$ *is constant over each unit age-time square —* $\lambda(x+s,t+s) = \lambda(x,t)$ *for* $s \in [0,1)$ *— we have the following relationship*[23]*:*

$$q(x,t) = 1 - \exp\left(-m(x,t)\right)$$

In particular, if $m(x,t)$ *is small, then* $q(x,t) \approx m(x,t)$.

2.3.2 Model calibration

To estimate the parameters of the CBD model, we use the weighted least squares (WLS):

$$\left\{\hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}\right\} = \arg\min \sum_{t\in\Omega_t}\sum_{x\in\Omega_x} w(x,t)\left(Y(x,t) - \kappa_t^{(1)} - \kappa_t^{(2)}(x-\bar{x})\right)^2$$

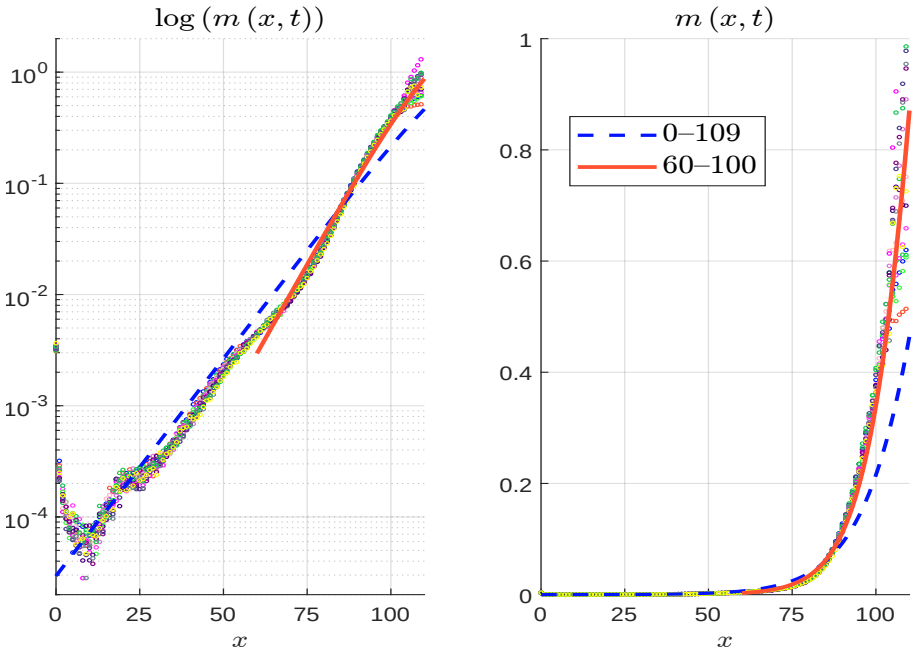
where $\bar{x}$ denotes the average age over Ω_x and the response variable is:

$$Y(x,t) = \text{logit}\left(q(x,t)\right) = \ln\left(\frac{1-\exp\left(-m(x,t)\right)}{\exp\left(-m(x,t)\right)}\right) = \ln\left(e^{m(x,t)} - 1\right)$$

[23] The proof is given in Appendix C.2 on page 141.

The condition $t \in \Omega_t$ indicates that the model is fitted on a subset of the time sample. Since $\kappa_t^{(1)}$and $\kappa_t^{(2)}$ are indexed by year, the most common approach is to estimate them year by year (Cairns *et al.*, 2009). However, other calibration strategies can also be considered. For instance, the model can be calibrated block-wise (*e.g.*, 5-year blocks) to obtain more observations per fit, or with a moving window (*e.g.*, a 10-year window) to smooth the parameter estimates over time. The condition $x \in \Omega_x$ indicates that the model is calibrated on a subset of ages. The rationale is that the linear assumption may not hold uniformly across all ages, particularly at very young or very old ages. For instance, in the context of retirement solutions, we are typically interested only in ages above the retirement date, and mortality below the age of 60 is not practically relevant. Figure 11 illustrates this by comparing a calibration over the entire age range with one restricted to ages above 60. The latter clearly provides a better fit. Moreover, we note that working on the logarithmic scale can be misleading in this context because it compresses errors at older ages[24]. This is an issue when designing a decumulation retirement solution, since precisely these older-age mortality rates are the most consequential.

Figure 11: Calibration of the CBD model (France, female, 2010–2023)



The previous results were obtained using uniform weights, corresponding to ordinary least squares. One issue with OLS is the sensitivity of the estimates to noisy observations at the oldest ages. For instance, comparing Figures 79 and 80 on page 168 shows that $\Omega_x = \{x : x \geq 60\}$ produces more volatile estimates of $\kappa_t^{(2)}$ than $\Omega_x = \{x : x \geq 60, x \leq 100\}$. This is because a low-exposure observation at age 105 receives the same weight as a much more reliable observation at age 75, despite its sampling variability being much larger. To reduce this sensitivity, the weights can be defined proportional to the inverse of the variance

[24] We have $\hat{m}(x,t) = \ln\left(1 + \exp\left(\hat{\kappa}_t^{(1)} + \hat{\kappa}_t^{(2)}(x - \bar{x})\right)\right)$.

Figure 12: WLS estimated values of κ_1 (France, female, 60–109)

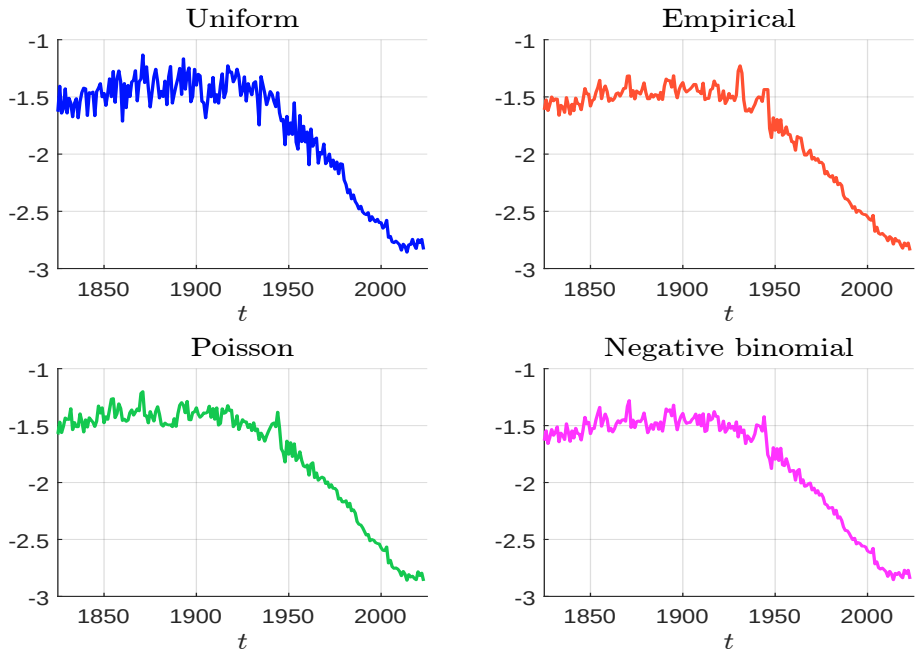
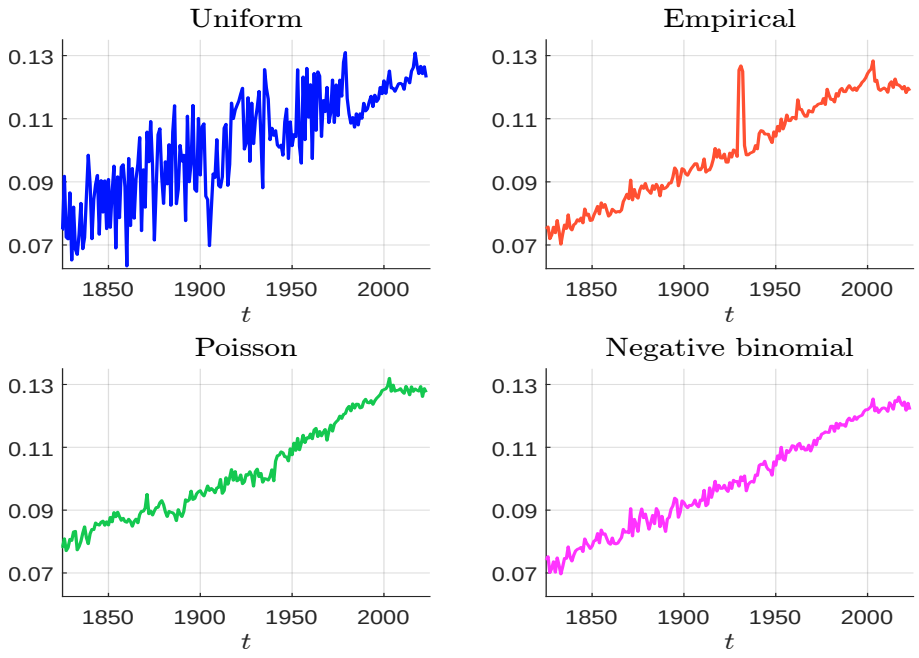


Figure 13: WLS estimated values of κ_2 (France, female, 60–109)



of the response variable $Y(x,t)$:

$$w(x,t) = \frac{1}{\text{var}\left(Y(x,t)\right)}$$

A first approach consists of replacing $\text{var}\left(Y(x,t)\right)$ by its empirical variance. The second approach is to compute the theoretical variance. In Appendix C.3 on page 142, we derive the optimal weighting schemes when death counts follow Poisson and negative binomial distributions (Equations (31) and (32)).

Figures (12) and (13) show the estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ using $\Omega_x = \{x : x \geq 60\}$ under different WLS approaches. The first panel corresponds to uniform weights, *i.e.*, the OLS method. The second panel uses the 10-year rolling-window empirical variance of $Y(x,t)$ to calculate the WLS weights. There is less noise and greater stability than with the OLS method. The third and fourth panels present the estimates obtained using Poisson and negative binomial weights[25], respectively. We observe that the Poisson specification produces the smoothest estimates.

Remark 6 *We observe some erratic behavior of κ_2, for instance in the case of the empirical weighting scheme. In Appendix C.4 on page 143, we develop an outlier detection procedure to obtain better smoothness. This procedure is particularly important for forecasting, since the projection of κ_1 and κ_2 is highly sensitive to outliers and noise in the estimated time series. Figures 81 and 82 on page 169 show the results obtained after applying the outlier correction procedure. The removal of outliers significantly reduces the noise in the estimated time series, particularly for the uniform and empirical weighting schemes. Although the Poisson and negative binomial weighting schemes also benefit from the correction, the improvement is less pronounced. This result is expected, as these methods naturally assign lower weight to the extreme observations. In other words, they already provide a partial adjustment for the influence of outliers during the estimation process.*

2.4 Model backtesting

Several considerations may motivate the choice of a statistical model. For example, a model may be chosen because it is parsimonious, computationally efficient, or easy to interpret. However, these characteristics alone are not sufficient to justify its use in practice. A model must also be evaluated based on its ability to accurately reproduce and predict the behavior of the underlying data. To assess the model performance, one must measure its predictive accuracy and the potential biases associated with its specification. This is typically achieved through a backtesting procedure in which model forecasts are compared with observed outcomes. It is important to note that forecasting performance depends on the estimation of model parameters and the forecasting methodology itself. Different forecasting procedures may lead to different results, even when based on the same underlying model. Therefore, the assessment of a model should jointly consider the quality of parameter estimation and the effectiveness of the forecasting approach.

[25] We use the ML value of δ, which was calibrated previously ($\delta = 1.63\%$). An alternative approach is to fit the CBD model with uniform weights, calculate the model-implied central death rate $\hat{m}(x,t) = \ln\left(1 + \exp\left(\hat{\kappa}_t^{(1)} + \hat{\kappa}_t^{(2)}(x - \bar{x})\right)\right)$, set $\hat{\mu}(x,t) = E(x,t)\,\hat{m}(x,t)$, and maximize the negative binomial log-likelihood function (Equation (4) on page 25) with respect to parameter δ only.

2.4.1 Forecasting

The LC and CBD models depend on two types of parameters: those that depend on age x and those that depend on year t. In what follows, we assume that the age parameters are stationary and remain fixed over time. This implies that the forecasting procedure only needs to project the time-varying parameters. Assuming that the mortality model is estimated on a window of data ending in year t, the forecasting procedure produces a projection of the variable of interest for years $t+h$ with $h \geq 1$, conditional on the information set $\mathcal{F}_t$ available at time t. For instance, in the case of the LC model, the variable of interest is the central death rate:

$$\begin{aligned}\ln \hat{m}(x,t+h) &= \mathbb{E}\left[\ln m(x,t+h) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\alpha_x + \beta_x \kappa_{t+h} + \varepsilon(x,t+h) \mid \mathcal{F}_t\right] \\ &= \alpha_x + \beta_x \mathbb{E}\left[\kappa_{t+h} \mid \mathcal{F}_t\right]\end{aligned}$$

It follows that forecasting the LC model reduces to forecast the time-varying index κ_t. In the CBD model, the variable of interest is the one-year death probability:

$$\begin{aligned}\ln\left(\frac{\hat{q}(x,t+h)}{1-\hat{q}(x,t+h)}\right) &= \mathbb{E}\left[\kappa_{t+h}^{(1)} + \kappa_{t+h}^{(2)}(x-\bar{x}) + \varepsilon(x,t+h) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\kappa_{t+h}^{(1)} \mid \mathcal{F}_t\right] + \mathbb{E}\left[\kappa_{t+h}^{(2)} \mid \mathcal{F}_t\right](x-\bar{x})\end{aligned}$$

Forecasting the CBD model is equivalent to forecast the variables $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$.

Two models are commonly used to predict the time index κ_t in the LC model: the random walk with drift and the autoregressive integrated moving average (ARIMA) model. These approaches are used, for instance, in Lee and Carter (1992), Lee (2000), Cairns *et al.* (2009), and Booth *et al.* (2006). Carter (1996) also proposes using the structural time series models (STSM) of Harvey (1990). The specifications of these models are given below.

- The random walk with drift (RWD) is defined as:

$$\kappa_{t+1} = \kappa_t + \beta + \varepsilon_t$$

  where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

- The ARIMA(p,d,q) is defined as:

$$\Phi(L)(1-L)^d \kappa_t = \Theta(L)\varepsilon_t$$

  where $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$, $\Theta(L) = 1 - \sum_{i=1}^{q} \theta_i L^i$ and $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

- The local linear trend (LLT) model is given by[26]:

$$\begin{cases} \kappa_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

  where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ and $\zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$.

[26] If $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, then $\beta_t = \beta_{t-1} = \beta$. It follows that $\mu_t = \mu_{t-1} + \beta = \mu_0 + \beta t$ and $\kappa_t = \mu_0 + \beta t + \varepsilon_t$. This means that the LLT model is a stochastic trend model.

Figure 14: Forecast of Lee-Carter time index κ_t (France, female, WLS)

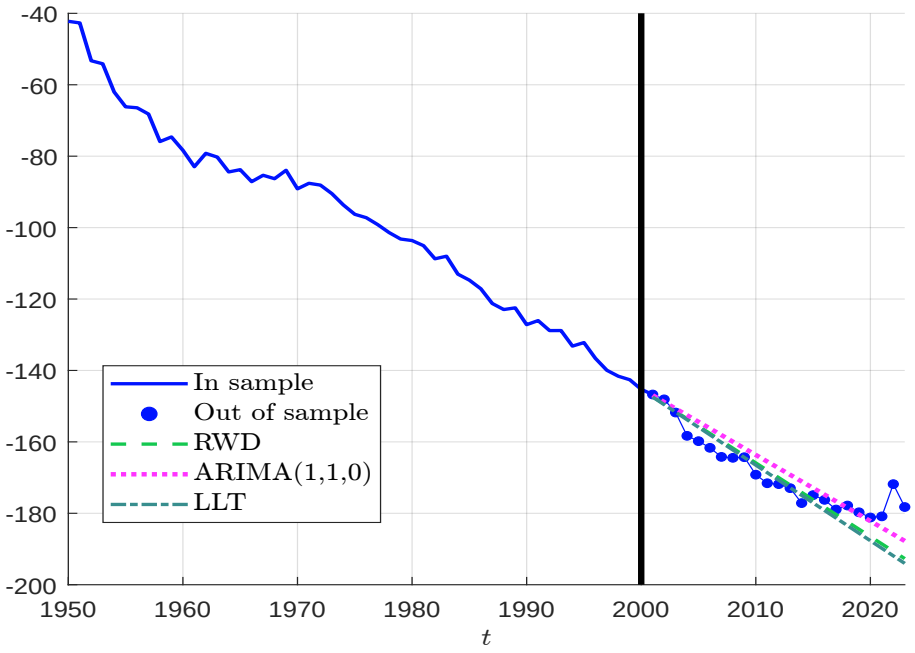
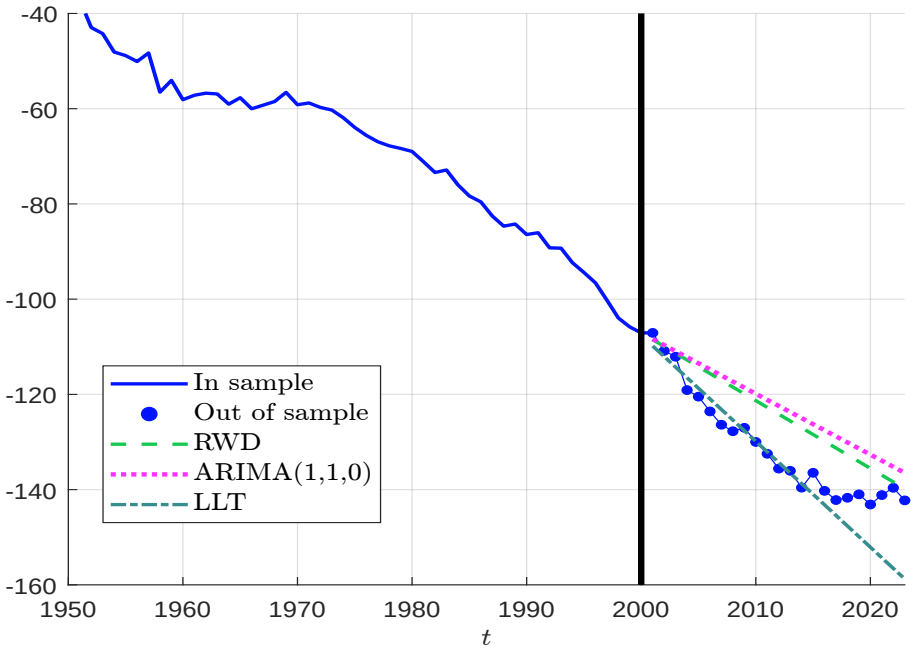


Figure 15: Forecast of Lee-Carter time index κ_t (France, male, WLS)



We consider the WLS estimates of the Lee-Carter time index κ_t, with the in-sample period set to 1950–2000. We forecast κ_t from 2001 to 2023 using three models: RWD, ARIMA(1,1,0), and LLT. Results are reported in Figures 14 and 15. We observe that the three models produce equivalent forecasts for females, while for males the LLT model performs better over the first 15 years but yields higher errors thereafter, due to a structural break in κ_t after 2015.

Remark 7 *The previous figures show the average forecast only. When we calculate the associated confidence intervals, however, we find that they can be quite wide, particularly for the RWD and LLT models. Figure 83 on page 170 illustrates this for the ARIMA(1,1,0) model, for which the 95% confidence interval is already quite wide.*

Remark 8 *We investigated several variants of the ARIMA(p,1,0) model. First, we found that the MA component adds little value and does not significantly change the projected trajectories. Second, the autoregressive coefficients ϕ_i are generally not significant for $i \geq 3$. Therefore, the choice was therefore ARIMA(1,1,0) and ARIMA(2,1,0). The optimal specification appears to depend on the projection window $(t+1, t+h)$. We ultimately selected ARIMA(1,1,0), for two reasons. First, ARIMA(2,1,0) can produce erratic forecasts for certain calibration dates. Second, $\hat{\phi}_1$ is markedly more stable than the pair $\left(\hat{\phi}_1, \hat{\phi}_2\right)$ across years. Therefore, we chose the more parsimonious specification for robustness.*

2.4.2 In-sample backtesting

For the in-sample comparison, each model is fitted to the cleaned mortality data from 1950 to 2023 for ages 50 to 95. The results are given in Table 5. They clearly separate the Lee-Carter family from the CBD family. On the log-mortality scale, the three LC variants have a female MAE close to 0.037 and a male MAE close to 0.030. By contrast, the CBD variants have a female MAE around 0.136–0.141 and a male MAE around 0.081–0.084. In this calibration window, CBD errors are therefore markedly larger than LC errors — roughly three times as large for males and nearly four times as large for females. This is consistent with the model structures. The LC model incorporates flexible age profiles $\alpha_x$and age loadings β_x, while the basic CBD specification imposes a linear old-age logit profile for each year.

Table 5: In-sample comparison on the log central death-rate scale (France, 1950–2023, age 50–95)

Sex	Model	Statistical method	$\ln m(x,t)$		$\text{logit}\, q(x,t)$		$q(x,t)$	
			MAE	RMSE	MAE	RMSE	MAE	RMSE
Female	LC	WLS	0.0363	0.0455	0.0374	0.0468	0.0018	0.0034
Female	LC	Poisson	0.0374	0.0497	0.0382	0.0505	0.0014	0.0029
Female	LC	NB	0.0362	0.0456	0.0372	0.0468	0.0017	0.0034
Female	CBD	OLS	0.1362	0.1678	0.1395	0.1715	0.0056	0.0106
Female	CBD	Poisson WLS	0.1406	0.2005	0.1428	0.2019	0.0037	0.0063
Female	CBD	NB WLS	0.1366	0.1676	0.1402	0.1717	0.0060	0.0114
Male	LC	WLS	0.0296	0.0384	0.0310	0.0404	0.0021	0.0046
Male	LC	Poisson	0.0297	0.0390	0.0311	0.0413	0.0022	0.0049
Male	LC	NB	0.0296	0.0385	0.0310	0.0406	0.0021	0.0047
Male	CBD	OLS	0.0810	0.1047	0.0844	0.1091	0.0054	0.0101
Male	CBD	Poisson WLS	0.0842	0.1090	0.0880	0.1139	0.0060	0.0114
Male	CBD	NB WLS	0.0808	0.1044	0.0844	0.1095	0.0058	0.0110

Regarding the LC model specifically, the statistical estimation method (WLS, Poisson, or NB) has a minimal impact on the accuracy of the fitted values. For females, LC NB has the lowest MAE on the log-mortality scale (0.0362), while LC WLS has the lowest RMSE (0.0455). For males, LC WLS and NB are tied for the lowest MAE (0.0296), and LC WLS again has the lowest RMSE. LC Poisson is extremely close to both alternatives in MAE, but its RMSE is noticeably higher for both sexes (0.0497 for females and 0.0390 for males, versus 0.0455/0.0456 and 0.0384/0.0385 for WLS and NB), suggesting that its fit is comparably accurate on average but somewhat less stable across the sample. The logit and $q(x,t)$ scale comparisons lead to the same broad conclusion. The LC models outperform the CBD models on the fitted sample. On the death-probability scale, LC Poisson yields the smallest q error for females, while the three LC variants are nearly tied for males. The CBD variants are less accurate on all reported scales, with q-scale MAE roughly two to four times larger than their LC counterparts. Consequently, the CBD model is useful as a parsimonious model, but the in-sample evidence favors the LC model for the French dataset.

2.4.3 Out-of-sample backtesting

The out-of-sample backtest uses an expanding calibration window. For each forecast origin t, the mortality model is re-estimated using only the observations available up to year t. The estimated time-varying parameters are projected using the selected forecasting method, and the resulting forecasts are used to reconstruct the mortality rate $\hat{m}(x,t+h)$ for a given forecast horizon h. This forecast is compared with the mortality rate $m(x,t+h)$ subsequently observed at the same age and calendar year. Then, the forecast origin advances by one year, and the entire estimation and forecasting procedure is repeated. Here, we consider forecast horizons $h=1$, $h=5$ and $h=10$. Forecast accuracy is measured on the log central death-rate scale. For each age x, forecast origin t and horizon h, the forecast error is defined as $e(x,t,h) = \ln \hat{m}(x,t+h) - \ln m(x,t+h)$. For each horizon, the mean absolute error (MAE) and root mean squared error (RMSE) are obtained by pooling the forecast errors across all forecast origins and ages. These statistics measure the out-of-sample prediction error. To illustrate the procedure, consider the WLS Lee-Carter model combined with an ARIMA(1,1,0) forecast of κ_t. The calibration sample begins in 1950, and the minimum period-factor estimation window is set to 15 years. The first forecast origin is therefore 1964. The Lee-Carter parameters α_x, β_x and κ_t are first estimated over 1950–1964. Then, the ARIMA(1,1,0) model is fitted to the 15 estimated values of κ_t over this same period and used to forecast κ_t for 1965, 1969 and 1974, corresponding to $h=1$, $h=5$ and $h=10$, respectively. The forecast mortality rate is obtained using the formula $\ln \hat{m}(x,t+h) = \alpha_x + \beta_x \hat{\kappa}_{t+h}$. These forecasts are compared with the observed mortality rates in the corresponding years. Then, the forecast origin is moved forward by one year. In the second iteration, the Lee-Carter model is re-estimated from 1950 to 1965, while the ARIMA(1,1,0) model is fitted to the most recent 15 estimated values of κ_t, *i.e.*, from 1951 to 1965. Forecasts are then produced for 1966, 1970 and 1975. This process continues with the Lee-Carter calibration window expanding by one year with each iteration, while the period-factor estimation window remains fixed at 15 observations. Since the last observation is 2023 and the longest forecast horizon is ten years, the final forecast origin is 2013. The Lee-Carter model is estimated from 1950 to 2013, the ARIMA(1,1,0) model is fitted to the values of κ_t from 1999 to 2013, and mortality forecasts are produced for 2014, 2018 and 2023. Thus, the backtest contains 50 forecast origins, ranging from 1964 to 2013. Since the forecast errors are evaluated for ages between 50 and 95, each forecast horizon contains $50 \times 46 = 2\,300$ age-origin observations. The reported statistics are therefore $\mathrm{MAE}(h) = \frac{1}{2\,300} \sum_{t=1964}^{2013} \sum_{x=50}^{95} \left| e(x,t,h) \right|$ and $\mathrm{RMSE}(h) = \sqrt{\frac{1}{2\,300} \sum_{t=1964}^{2013} \sum_{x=50}^{95} e(x,t,h)^2}$.

Table 6: Out-of-sample forecast errors of $\ln m(x,t)$ (France, 1964–2023, age 50–95)

Model	Statistical method	Forecasting method	h	Female		Male	
				MAE	RMSE	MAE	RMSE
LC	WLS	ARIMA	1	0.0383	0.0497	0.0353	0.0441
			5	0.0563	0.0711	0.0526	0.0650
			10	0.0837	0.1058	0.0762	0.0961
		RWD	1	0.0393	0.0504	0.0373	0.0461
			5	0.0571	0.0721	0.0539	0.0666
			10	0.0840	0.1061	0.0771	0.0970
		LLT	1	0.0414	0.0538	0.0378	0.0471
			5	0.0648	0.0841	0.0646	0.0818
			10	0.0970	0.1277	0.1008	0.1266
LC	Poisson	ARIMA	1	0.0420	0.0571	0.0336	0.0425
			5	0.0643	0.0855	0.0502	0.0620
			10	0.0978	0.1309	0.0728	0.0930
		RWD	1	0.0442	0.0587	0.0356	0.0446
			5	0.0659	0.0874	0.0515	0.0637
			10	0.0990	0.1320	0.0738	0.0940
		LLT	1	0.0473	0.0637	0.0363	0.0458
			5	0.0816	0.1119	0.0659	0.0827
			10	0.1298	0.1777	0.1044	0.1295
LC	NB	ARIMA	1	0.0390	0.0510	0.0344	0.0431
			5	0.0582	0.0743	0.0511	0.0633
			10	0.0868	0.1123	0.0740	0.0940
		RWD	1	0.0405	0.0521	0.0363	0.0451
			5	0.0593	0.0758	0.0524	0.0648
			10	0.0876	0.1130	0.0749	0.0949
		LLT	1	0.0448	0.0589	0.0367	0.0458
			5	0.0726	0.0957	0.0650	0.0815
			10	0.1112	0.1480	0.1021	0.1267
CBD	OLS	ARIMA	1	0.1463	0.1765	0.0824	0.1052
			5	0.1579	0.1904	0.0943	0.1193
			10	0.1717	0.2091	0.1129	0.1408
		RWD	1	0.1468	0.1769	0.0840	0.1062
			5	0.1585	0.1908	0.0960	0.1203
			10	0.1723	0.2094	0.1138	0.1414
		LLT	1	0.1474	0.1772	0.0837	0.1068
			5	0.1643	0.1977	0.1047	0.1312
			10	0.1842	0.2289	0.1316	0.1677
CBD	Poisson WLS	ARIMA	1	0.1489	0.2087	0.0835	0.1077
			5	0.1646	0.2287	0.0949	0.1210
			10	0.1849	0.2549	0.1125	0.1430
		RWD	1	0.1496	0.2088	0.0851	0.1090
			5	0.1657	0.2287	0.0964	0.1221
			10	0.1865	0.2557	0.1136	0.1438
		LLT	1	0.1509	0.2077	0.0848	0.1092
			5	0.1743	0.2303	0.1041	0.1326
			10	0.2034	0.2660	0.1316	0.1701
CBD	NB WLS	ARIMA	1	0.1466	0.1762	0.0822	0.1048
			5	0.1581	0.1903	0.0939	0.1189
			10	0.1717	0.2094	0.1124	0.1408
		RWD	1	0.1472	0.1767	0.0839	0.1059
			5	0.1587	0.1907	0.0955	0.1199
			10	0.1723	0.2096	0.1134	0.1415
		LLT	1	0.1476	0.1769	0.0838	0.1063
			5	0.1643	0.1975	0.1050	0.1309
			10	0.1838	0.2289	0.1318	0.1681

Table 6 reports the forecast accuracy of mortality models fitted to the mortality rates of French men and women aged 50-95. Forecast performance is evaluated over three projection horizons (1, 5, and 10 years). For each mortality model (Lee-Carter and Cairns-Blake-Dowd), estimation method (WLS, Poisson, and negative binomial), and forecasting approach (ARIMA, RWD, and LLT), we report the mean absolute error and root mean squared error. We obtain the following results:

- The Lee-Carter model substantially outperforms the Cairns-Blake-Dowd model by a wide margin.
  Across all specifications, the LC model produces forecast errors that are typically three to four times smaller than those of the CBD model. For example, at the one-year horizon ($h = 1$), the best-performing LC specification (LC/WLS/ARIMA) achieves an MAE of 0.0383, whereas the best CBD specification (CBD/OLS/ARIMA) achieves an MAE of 0.1463. This performance gap persists at longer forecasting horizons. The superior accuracy of the LC model likely reflects its ability to capture age-specific mortality improvement through the parameter β_x.

- The choice of mortality model is more important than the choice of estimation method.
  Within the LC family, the differences between WLS, Poisson, and NB estimation are relatively small compared to the substantial gap between the LC and CBD models. A similar pattern is observed within the CBD family, where the choice among OLS, Poisson WLS, and NB WLS has only a limited impact on forecast accuracy.

- Forecast accuracy deteriorates as the projection horizon increases.
  As expected, out-of-sample forecast errors increase with the forecasting horizon. For instance, for the LC/NB/LLT specification applied to male mortality, the RMSE rises from 0.0458 at $h = 1$ year to 0.0815 at $h = 5$ years and 0.1267 at $h = 10$ years.

- Male mortality is forecast slightly more accurately than female mortality.
  On average across all model specifications, the MAE is 0.1160 for females compared with 0.0796 for males, while the corresponding RMSE values are 0.1483 and 0.1004, respectively. Overall, forecast errors are approximately 30% lower for males than for females.

3 Annuity modeling

Life annuity pricing is the valuation of survival contingent cash flows. For an individual of age x, the value of a continuous annuity[27] paying one monetary unit per year is commonly written as follows[28]:

$$\bar{a}(x) = \int_0^\infty e^{-\varrho t} p(x,t)\, \mathrm{d}t \tag{6}$$

where ϱ is the discount rate and $p(x,t)$ is the probability that the annuitant, alive at age x, survives at age $x+t$. Thus, the basic structure is a discounted survival-weighted cash-flow valuation.

[27] Depending on the timing of payments, an annuity may be expressed as an annuity-immediate $a(x)$, paid at the end of each period, an annuity-due $\ddot{a}(x)$, paid at the beginning of each period, or a continuous annuity $\bar{a}(x)$, paid continuously. We have the following relationship under standard approximations: $\ddot{a}(x) \approx a(x)+1$ and $\bar{a}(x) \approx \ddot{a}(x) - 0.5 \approx a(x) + 0.5$.

[28] In discrete time, the analogous expression is:

$$a(x) = \sum_{t \geq 1} \frac{p(x,t)}{(1+\varrho)^t} \tag{5}$$

3.1 Literature review

The literature is distinguished by how it treats non-diversifiable longevity risk, in particular the distinction between idiosyncratic mortality risk and systematic longevity risk. Idiosyncratic mortality risk refers to deviations of individual lifetimes from their expected distribution. This component can be reduced by pooling many similar contracts, as in the standard annuity-pricing argument of Milevsky and Young (2007). By contrast, systematic longevity risk refers to changes in mortality rates that affect an entire portfolio, such as persistent improvements in mortality at retirement ages. Biffis (2005) distinguishes explicitly between such random fluctuations around expected mortality and systematic departures from it, arguing that both financial and demographic sources of uncertainty must be considered jointly when valuing life insurance contracts. How this systematic component is treated is what distinguishes one strand of this literature from another. We will first examine the classical actuarial framework, in which systematic longevity risk is incorporated into prudent assumptions rather than priced separately. Then, we will move on market consistent and risk-neutral valuation, which prices systematic longevity risk when survival contingent claims are traded or hedgeable. Finally, we will consider indifference pricing approaches, which address the residual, unhedgeable component of systematic risk, before turning to the institutional evidence on how far markets for these instruments have actually developed.

3.1.1 Insurance premium, actuarial valuation and annuity

In actuarial science, the premium is the amount paid by the policyholder to the insurer in exchange for the benefits provided by the insurance contract. Dickson *et al.* (2020) define the insurer's future loss as the difference between the present value of future outflows and the present value of future inflows associated with the policy. For a contract priced on a net premium basis, the insurer's loss random variable is given by:

$$L_0^{\text{net}} = \text{PV (benefit outflows)} - \text{PV (premium inflows)}$$

When pricing on a gross premium basis, expenses incurred by the insurer are also taken into account, and the loss random variable becomes:

$$L_0^{\text{gross}} = \text{PV (benefit outflows + expenses)} - \text{PV (gross premium inflows)}$$

Both L_0^{net} and L_0^{gross} are random variables, since the timing and amount of future cash flows depend on uncertain events, most notably the future lifetime of the insured. A fundamental principle in actuarial pricing is the equivalence principle, which states that the premium should be determined such that the expected present value of future inflows equals the expected present value of future outflows. Therefore, the expected value of the insurer's loss at policy inception should be zero:

$$\mathbb{E}[L_0] = 0$$

Dickson *et al.* (2020) use this framework to derive premiums and actuarial present values for a variety of life contingent contracts. For a life annuity, if the benefit is the annuity itself, then PV (benefit outflows) $= a\,(x)$. In the case of a single premium annuity, the equivalence principle then implies:

$$\text{Premium} = a\,(x)$$

The equivalence principle justifies then Equations (5) and (6). Milevsky (2013) reviews this actuarial framework from the perspective of retirement income economics. Milevsky traces the actuarial valuation of annuities back to Halley (1693), where survival probabilities were first combined with the time value of money. He also discusses the central role of the Gompertz annuity pricing model[29]. However, this classical treatment leaves implicit

[29] See Appendix C.5 on page 143.

whether the survival probabilities $p(x,t)$ used in Equation (6) should be interpreted under the historical (real-world) or risk-neutral measure. Carriere (1999) solves this problem by providing a no-arbitrage foundation. He shows that classical actuarial valuation formulas for life insurance and annuities are compatible with the financial no-arbitrage pricing principle when the time of death is independent of bond prices. This result formally justifies the use of real-world survival probabilities $p(x,t)$ in Equation (6) and builds a bridge between the actuarial equivalence principle and the no-arbitrage framework of financial economics.

3.1.2 Contingent claim pricing and risk-neutral valuation

The early 2000s literature on longevity-linked securities builds on this bridge by introducing traded, or potentially traded, survival contingent claims. This requires that the valuation of annuity cash flows must be consistent with the observed prices of such instruments. Blake and Burrows (2001) initiated this stream by proposing survivor bonds as instruments for transferring longevity risk. Under this view, a life annuity can be represented as a portfolio of payments of the form $b_k \mathbb{1}\{\boldsymbol{\tau} > T_k\}$, where b_k is paid at time T_k if the annuitant is alive, *i.e.*, if the survival time $\boldsymbol{\tau}$ exceeds T_k. A market-consistent value can then be written as follows:

$$V(t) = \sum_k b_k \mathbb{E}^{\mathbb{Q}} \left[\exp\left(- \int_t^{T_k} r_s \, \mathrm{d}s \right) \mathbb{1}\{\boldsymbol{\tau} > T_k\} \middle| \mathcal{F}_t \right] \tag{7}$$

where $\mathbb{Q}$ is a pricing measure consistent with traded financial assets and traded longevity-linked instruments. In the same year, Milevsky and Promislow (2001) develop this idea from a different angle, by studying mortality derivatives and the option to annuitize, treating annuitization itself as a financial option whose value depends jointly on future interest rates and future mortality rates. Dahl (2004) then embeds this logic in a full stochastic framework, specifying an equivalent martingale measure under which both financial and mortality risks enter the valuation (the financial market component through a Brownian market-price-of-risk term, and the mortality component through an adjusted mortality intensity). The resulting market reserve applies to contracts exposed to stochastic mortality generally, including, but not limited to, mortality-linked contracts.

A recurring difficulty in this literature is that mortality risk does not correspond to a single, well-defined financial risk category. Some authors treat mortality risk as analogous to interest rate risk, while others treat it as a form of credit risk. This has motivated the adaptation of established financial models of interest rate and credit risk to the mortality setting. For instance, Cairns *et al.* (2006b) developed an arbitrage-free framework for mortality-linked instruments by adapting term structure methods from interest rate modeling. They model spot mortality rates, forward mortality rates and survival probabilities in a manner similar to short rate, forward rate, and market model approaches in fixed income. Valuation examples include survivor bonds, survivor swaps, annuity futures, mortality options, and guaranteed annuity options (Blake *et al.*, 2006). Their framework allows for stochastic interest rates and stochastic mortality rates simultaneously, which is necessary when payments depend on both survival and discounting over long horizons. In this case, an annuity paying b_k at dates T_k is represented by Equation (7), where $\exp\left(-\int_t^{T_k} r_s \, \mathrm{d}s\right)$ captures stochastic interest rate discounting and $\mathbb{1}\{\boldsymbol{\tau} > T_k\}$ captures survival contingency. If interest rates and mortality are independent under $\mathbb{Q}$, the expectation decomposes into a zero-coupon bond price and a risk-neutral survival probability. Biffis (2005) takes this integration further with an affine-process formulation, modelling asset prices and mortality dynamics jointly via the analogy between death and default. Mortality intensity plays

the same role as a default intensity in credit risk, survival benefits correspond to defaultable zero-coupon claims, and annuities are built from survival contingent payments. This yields valuation expressions for life contingent claims with stochastic demographic factors, stochastic interest rates and asset-linked benefits.

3.1.3 The issue of incomplete markets

The above valuation methods are based on an implicit assumption: that the price of longevity-linked claims can be determined in the same way as the price of a European option using the Black-Scholes model, *i.e.*, by replication in a complete market. In practice, this assumption fails because longevity-linked instruments do not trade in sufficient volume to span systematic mortality risk. Incomplete market approaches address this gap directly. Young and Zariphopoulou (2002) develop a dynamic equivalent utility pricing method for insurance risks precisely because insurance cash flows often depend on risks that cannot be replicated by trading financial assets. The resulting pricing rule yields buyer and seller reservation prices that depend on preferences, risk aversion, and wealth rather than a single arbitrage-free price. Ludkovski and Young (2008) apply this indifference pricing logic directly to pure endowments and temporary life annuities under stochastic hazard rates and interest rates. They assume an insurer with exponential utility and a financial market consisting of a money-market account and default-free zero-coupon bonds. The market is driven by a stochastic hazard rate that governs conditionally independent, homogeneous lives. They derive PDEs for the indifference price of pure endowments and temporary life annuities[30]. Barrieu *et al.* (2012) takes a step back from these specific models to classify pricing methodologies by market structure. They distinguish classical static actuarial pricing under the real-world measure, risk-neutral pricing in a liquid longevity market, and indifference pricing under incomplete markets. This classification directly aligns with the focus of this review. Actuarial valuation is appropriate for pooled liabilities valued using a mortality basis and loading factors, risk-neutral valuation is appropriate when survival contingent claims are traded or hedgeable, while indifference pricing applies when systematic mortality risk remains partly unhedgeable.

3.1.4 Market development and practice

The evolution of longevity-linked products provides the institutional context for the above trade-off. Cairns *et al.* (2006b) discuss survivor bonds, short-dated mortality-linked securities, survivor swaps, annuity futures, and mortality options. They note a trade-off between standardized and customized hedges. Standardized instruments can support liquidity but leave basis risk for specific annuities or pension portfolios, while bespoke solutions fit liabilities more closely but are less liquid. Blake *et al.* (2019) document how this trade-off played out after the first wave of capital market proposals. For example, the EIB/BNP Paribas longevity bond did not attract sufficient demand. Similarly, later capital market instruments such as longevity-spread bonds, q-forwards, and tail-risk protection, reached the market only in limited volumes. By contrast, insurance-based solutions — including buy-outs, buy-ins, and longevity insurance — developed more successfully. Their explanation is comparative rather than purely structural. Once basis risk, credit risk, regulatory capital, collateral, and liquidity were taken into account, index-based capital market instruments were often less attractive than customized insurance solutions. Basis risk arises because

[30] In particular, their comparative static analyzes show that the price per contract can increase with the number of contracts when the hazard rate is stochastic, because the common hazard rate process induces dependence across lives — the opposite of the diversification benefit obtained under idiosyncratic mortality risk alone.

an index hedge references a population that may differ from the annuitant portfolio in terms of age, sex, socioeconomic composition, underwriting status, geography and mortality improvement. Credit risk reflects the long maturity of longevity transactions and the possibility that a counterparty cannot meet its obligations. Although collateralization mitigates counterparty exposure, it introduces funding costs, valuation disputes, and operational complexity. Regulatory treatment affects transaction value because a hedge that leaves material basis risk generates less capital relief. Liquidity affects both the entry price and the exit value, especially for customized contracts that do not trade on secondary markets. Blake *et al.* (2019) treat these features as central to the comparison between capital market and insurance-based risk transfer. Blake and Cairns (2021) extend this account by documenting the continued use of buy-ins, buy-outs, longevity swaps, reinsurance, and funded reinsurance. They also note the withdrawal of several investment banks from longevity risk transfer following changes in bank capital requirements, even as insurers and reinsurers expanded their role[31]. Taken together, these institutional developments explain why the incomplete market theory remains empirically relevant. Two decades after the first capital market proposals, systematic longevity risk is still predominantly transferred through customized insurance and reinsurance arrangements rather than through a liquid market of standardized instruments.

3.2 Annuity pricing

The literature reviewed above documents the limited liquidity and trading activity of longevity-linked instruments. In this context, the current empirical implementation does not draw from a sufficiently rich set of observable longevity-linked market prices from which a market price of longevity risk could be calibrated. Therefore, the empirical analysis in this paper uses a physical measure actuarial valuation, with annuity values computed as expected present values under the real-world mortality forecasts. In other words, valuation is performed using the historical probability measure $\mathbb{P}$, rather than the risk-neutral probability measure $\mathbb{Q}$.

3.2.1 Contract value

We assume that the valuation is performed at the end of calendar year t, so that the first projected mortality year is $t+1$. Here t may be the end of the sample period T or any other date before or after T. Let x denote the issue age, and x^+ the limiting age[32]. The maximum payment horizon is then $H = x^+ - x$. We consider an annuity-immediate paying an annual amount B in arrears. Let $\mathcal{F}_t$ denote the information available at the valuation date t. Conditional on $\mathcal{F}_t$, the fitted age parameters are treated as fixed, and future mortality uncertainty enters solely through the future period factors. The probability that an individual aged x at the valuation date t survives to the payment horizon h is:

$$p\left(x,t,h\right) = \prod_{j=1}^{h} \left(1 - q\left(x+j-1, t+j-1\right)\right)$$

with $p\left(x,t,0\right) = 1$. The random present value of the contract is:

$$A\left(x,t\right) = \sum_{h=1}^{H} \frac{B \cdot \mathbb{1}\left\{\boldsymbol{\tau} > t+h\right\}}{\left(1+\varrho_t\right)^h}$$

[31] In the UK pension risk transfer market, for example, funded reinsurance now allows an insurer to transfer both asset risk and longevity risk associated with pension and annuity liabilities in a single structure.

[32] x^+ can be set to ∞.

where $\boldsymbol{\tau}$ is the survival time of the annuitant. For a specified future mortality path, the actuarial value of the contract is the conditional expected present value:

$$a(x,t) = \mathbb{E}\left[A(x,t) \mid \mathcal{F}_t\right] = \sum_{h=1}^{H} \frac{B}{(1+\varrho_t)^h} \mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\tau} > t+h\right\} \mid \mathcal{F}_t\right] = B \sum_{h=1}^{H} \frac{p(x,t,h)}{(1+\varrho_t)^h} \tag{8}$$

This is the classical survival-weighted annuity valuation. As explained by Pitacco *et al.* (2009), the presence of survival probabilities does not itself make the valuation stochastic. Once the mortality path and the discount factor are fixed, the survival probabilities simply serve to determine the expected cash flows.

Remark 9 *To simplify notation, we do not distinguish in this paper between a random variable and its realization. For instance, depending on the context, $p(x,t,h)$ may be either a scalar quantity or a random variable. This is why $a(x,t)$ may deterministic or stochastic, depending on whether the underlying mortality path is fixed or random.*

3.2.2 Deterministic valuation

The deterministic valuation fixes a single (historical/projected) mortality path and evaluates the annuity along this path. This corresponds to the deterministic actuarial approach described by Pitacco *et al.* (2009), under which one age-pattern of future mortality gives one actuarial value. In the case of LC model, we have[33]:

$$\hat{\kappa}_{t+h} = \begin{cases} \kappa_{t+h} & \text{if } t+h \leq T \\ \mathbb{E}\left[\kappa_{t+h} \mid \mathcal{F}_t\right] & \text{if } t+h > T \end{cases} \tag{9}$$

where κ_{t+h} is the estimated value when $t+h \leq T$ and $\hat{\kappa}_{t+h}$ is the forecast value when $t+h > T$. The projected central death rates are then:

$$\hat{m}(x+j-1, t+j-1) = \exp\left(\alpha_{x+j-1} + \beta_{x+j-1}\hat{\kappa}_{t+j-1}\right)$$

where α_x and β_x are the estimated age-specific parameters. Applying the constant force approximation, it follows that:

$$\hat{q}(x+j-1, t+j-1) = 1 - \exp\left(-\hat{m}(x+j-1, t+j-1)\right)$$

Equation (8) becomes:

$$a(x,t) = B \sum_{h=1}^{H} \frac{1}{(1+\varrho_t)^h} \left(\prod_{j=1}^{h} \exp\left(-\exp\left(\alpha_{x+j-1} + \beta_{x+j-1}\hat{\kappa}_{t+j-1}\right)\right) \right) \tag{10}$$

In the case of the CBD model, we apply the same rule (9) to the coefficients $\hat{\kappa}^{(1)}_{t+h}$ and $\hat{\kappa}^{(2)}_{t+h}$. Finally, we obtain:

$$a(x,t) = B \sum_{h=1}^{H} \frac{1}{(1+\varrho_t)^h} \left(\prod_{j=1}^{h} \frac{1}{1+\exp\left(\hat{\kappa}^{(1)}_{t+j-1} + \hat{\kappa}^{(2)}_{t+j-1}\left(x+j-(\bar{x}+1)\right)\right)} \right) \tag{11}$$

where $\bar{x}$ denotes the mean age used to fit the CBD model. This projection-to-valuation procedure is implemented directly by Brouhns *et al.* (2002). The Lee-Carter period index is forecast using an ARIMA model, the resulting forecasts are then used to generate projected age-specific mortality rates, and the projected life table is used to calculate life-annuity net single premiums.

[33] Generally, t is greater than or equal to T.

3.2.3 Stochastic valuation

A single projected mortality path does not account for uncertainty about the future mortality trend. Olivieri (2001) analyzes this source of systematic risk in life insurance valuation, arising when a projected mortality table may not correctly represent future mortality. Pitacco *et al.* (2009) subsequently formulate this distinction in terms of mortality scenarios. A single mortality scenario produces a deterministic actuarial value, while assigning probabilities to alternative mortality scenarios permits an unconditional valuation. For instance, the deterministic valuation above corresponds to the expected mortality path, since the q-values are computed using the forecast parameters $\mathbb{E}\left[\theta_{t+h} \mid \mathcal{F}_t\right]$, where θ_t is κ_t for the LC model, and $\kappa_t^{(1)}$ or $\kappa_t^{(2)}$ for the CBD model. More generally, we can consider other parameter scenarios. Let s be a specific scenario. The annuity value conditional on this scenario is:

$$a^{(s)}(x,t) = B\sum_{h=1}^{H}\frac{p^{(s)}(x,t,h)}{(1+\varrho_t)^h} = B\sum_{h=1}^{H}\frac{\prod_{j=1}^{h}\left(1-q^{(s)}(x+j-1,t+j-1)\right)}{(1+\varrho_t)^h}$$

Here, the q-values are derived from the specific scenario s. For instance, $q^{(s)}(x,t)$ can be computed from the simulations of $\kappa_{t+h}^{(s)}$ generated by the conditional probability distribution $\kappa_{t+h} \mid \kappa_{t+h-1} \sim \mathbf{F}_{\kappa_{t+h}}(\kappa)$. Calculating the average annuity value across simulated mortality scenarios provides an estimate of the expected annuity value under mortality uncertainty. However, this average does not necessarily coincide exactly with the annuity value computed from the expected path of the period factors since the mapping from future mortality dynamics to annuity values is nonlinear. In the numerical results, however, the difference between the two valuations remains below 0.1%. In addition to the mean annuity value, the Monte Carlo approach provides the full probability distribution $\mathbf{F}_{a(x,t)}$ of annuity values resulting from uncertainty in future mortality.

Remark 10 *An alternative Monte Carlo approach consists in simulating the survival time $\boldsymbol{\tau}$ given a scenario s, rather than simulating the probability $q(x,t)$ via the time index κ_t. In this case, we have:*

$$a^{(s)}(x,t) = \mathbb{E}\left[A^{(s,r)}(x,t)\right] \approx \frac{1}{n_R}\sum_{r=1}^{n_R}A^{(s,r)}(x,t)$$

where:

$$A^{(s,r)}(x,t) = \sum_{h=1}^{H}\frac{B}{(1+\varrho_t)^h}\mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t+h\right\}$$

3.2.4 Empirical analysis

Average annuity values The empirical analysis is based on historical French mortality data for both sexes. We compute the time series of the central death rate $m(x,t)$ and use it to calculate the historical annuity value[34] via Equation (8), with a unit payment $B = 1$, a limiting age $x^+ = 95$ years, and a discount rate $\varrho_t = 3\%$ (Table 7). For an issue age of $x = 60$ years, for instance, $a(x,t)$ equals 13.88 in 1950 and 16.51 in 1980. As expected, $a(x,t)$ increases with the valuation year t and decreases with the issue age x. The former reflects the improvement in survival probabilities over time, while the latter reflects the higher mortality rates and shorter expected remaining lifetimes associated with older ages.

[34]Here, we calculate the historical value $a(x,t) = \mathbb{E}\left[A(x,t) \mid \mathcal{F}_T\right]$ conditional on information available upon T.

For some combinations of (x,t), $a(x,t)$ cannot be computed because the required central death rates are not observed historically. This occurs when $t+x^{+}-x>T$, *i.e.*, when the realized survival probability[35] $p(x,t,h)$ can not be evaluated over the full payment horizon H due to insufficient historical data. In these cases, we forecast the central death rate. To this end, we calibrate the Lee-Carter WLS model on the period $t\in[1950,2023]$ and ages $x\in[50,95]$, and use an ARIMA(1,1,0) model to forecast the period index κ_t for $t>2023$. This allows us to complete the previous table and obtain the results given in Table 8. In this case, the computation of $a(x,t)$ combines historical and forecast values of $m(x,t)$.

Table 7: Historical value of the annuity contract (France, female, $\varrho_t=3\%$, $x^{+}=95$)

x	t							
	1950	1960	1970	1980	1990	2000	2010	2020
55	16.16	17.04	17.88	18.68				
60	13.88	14.67	15.70	16.51				
65	11.41	12.16	13.09	14.02	14.86			
70	9.05	9.72	10.46	11.51	12.31			
75	6.73	7.34	7.93	8.79	9.58	10.34		
80	4.82	5.33	5.73	6.29	7.08	7.62		
85	3.27	3.60	3.88	4.22	4.75	5.08	5.50	
90	2.05	2.24	2.35	2.49	2.72	2.86	3.07	

Table 8: Value of the annuity contract (France, female, $\varrho_t=3\%$, $x^{+}=95$, historical/forecast combination)

x	t							
	1950	1960	1970	1980	1990	2000	2010	2020
55	16.16	17.04	17.88	18.68	19.26	19.47	19.70	20.13
60	13.88	14.67	15.70	16.51	17.16	17.56	17.75	18.16
65	11.41	12.16	13.09	14.02	14.86	15.44	15.63	15.95
70	9.05	9.72	10.46	11.51	12.31	12.96	13.27	13.52
75	6.73	7.34	7.93	8.79	9.58	10.34	10.79	10.91
80	4.82	5.33	5.73	6.29	7.08	7.62	8.07	8.20
85	3.27	3.60	3.88	4.22	4.75	5.08	5.50	5.61
90	2.05	2.24	2.35	2.49	2.72	2.86	3.07	3.09

Remark 11 *In practice, the annuity contract is priced using only the forecast values of $m(x,t)$, since an annuitant of age x subscribing at time t has, by definition, no mortality observations beyond the subscription date. Therefore, the most realistic approach relies exclusively on forecast values obtained in real time*[36]*. At each valuation date t, the Lee-Carter WLS model is estimated using only the mortality data available up to t, rather than the full historical sample used in Tables 7 and 8. Specifically, Table 9 reports the results obtained when the Lee-Carter WLS model is calibrated at each date t using the previous 30 years of observations and when the period index κ_t is forecasted using an ARIMA(1,1,0) model fitted to the last 15 estimated values*[37]*. As expected, the resulting annuity values differ from those obtained above. For an issue age of $x=60$ years, for example, $a(x,t)$ is now equal to* 13.28

[35] In fact, it is more a survival rate than a survival probability when using historical central death rates.

[36] This means that $a(x,t)=\mathbb{E}\left[A(x,t)\mid\mathcal{F}_t\right]$ and not $a(x,t)=\mathbb{E}\left[A(x,t)\mid\mathcal{F}_T\right]$.

[37] Two distinct estimation windows are involved: one for the mortality model and another for the ARIMA model. These windows do not necessarily coincide.

in 1950 and 15.94 *in 1980, compared to* 13.88 *and* 16.51*, respectively, using the full historical time series.*

Table 9: Value of the annuity contract (France, female, $\varrho_t = 3\%$, $x^+ = 95$, real-time rolling forecast)

x	t							
	1950	1960	1970	1980	1990	2000	2010	2020
55	15.52	16.74	16.85	18.12	19.24	19.72	19.94	20.01
60	13.28	14.47	14.64	15.94	17.12	17.66	17.99	18.14
65	10.97	12.06	12.29	13.56	14.74	15.36	15.80	16.04
70	8.68	9.62	9.88	11.03	12.15	12.84	13.37	13.68
75	6.57	7.32	7.57	8.52	9.46	10.15	10.74	11.10
80	4.75	5.30	5.50	6.20	6.89	7.47	8.03	8.36
85	3.31	3.66	3.78	4.24	4.65	5.02	5.43	5.68
90	2.06	2.25	2.31	2.51	2.68	2.84	3.02	3.14

The value of an annuity $a(x,t)$ depends on several parameters and modeling choices[38]:

1. First, it depends on the valuation date t, because mortality patterns today differ from those observed 50 years ago and will differ again in 30 years. For instance, Figure 84 on page 170 shows the time evolution of $a(x,t)$. It has increased since 1950 because life expectancy has improved.

2. Second, $a(x,t)$ depends on gender. On average, females live longer than males, so they theoretically have to pay a higher value. This pattern is again illustrated in Figure 84 on page 170.

3. Third, the discount rate ϱ_t is also a determining factor and arguably the most important factor influencing the level of the annuity contract value (Rabitti and Borgonovo, 2020). In fact, the annuity range with respect to ϱ_t is:

$$0 \leq a(x,t) \leq a^+(x,t) = B \sum_{h=1}^{H} p(x,t,h)$$

   where the lower bound is reached as $\varrho_t \to \infty$ and the upper bound is reached as $\varrho_t \to 0$. Figure 85 on page 171 illustrates the high sensitivity of $a(x,t)$ to ϱ_t. Moreover, there is no universally accepted value of ϱ_t. The appropriate choice depends heavily on the prevailing economic environment, in particular the level of interest rates[39].

4. Fourth, age x matters. As previously mentioned, the value of the annuity contract is higher for younger individuals than for older ones (see Figure 86 on page 171).

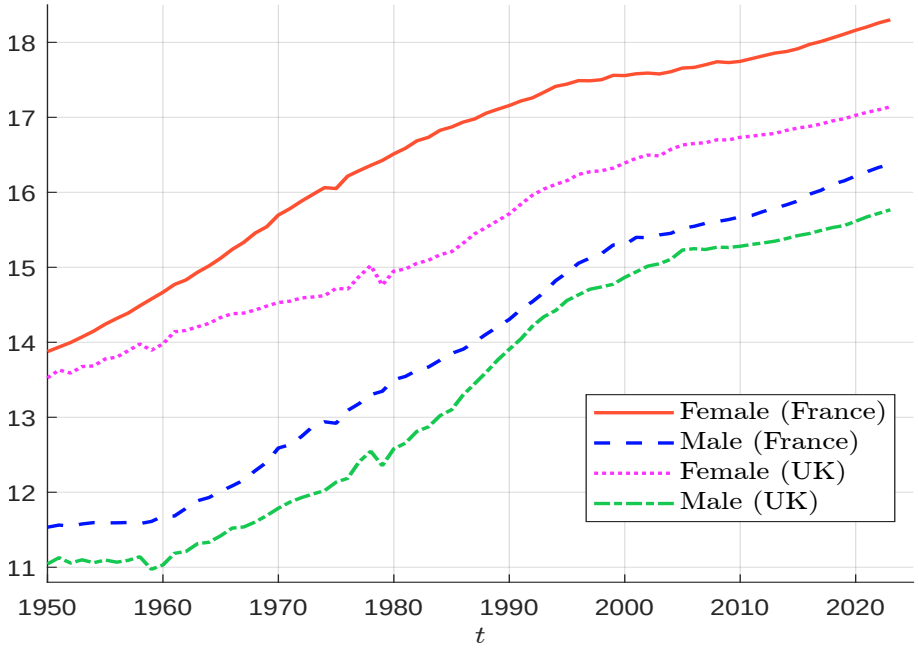
5. Fifth, the model parameterization plays a role. $a(x,t)$ depends on the calibration sample, the age range considered, the value of x^+, the choice of mortality model, and the forecasting method (see Table 36 on page 163).

6. Finally, Figure 16 below shows the impact of country on annuity value. For example, $a(x,t)$ is lower in the UK than in France, especially for females.

[38] The following empirical results use the historical/forecast combination, a limiting age of 95 years, a discount rate of 3%, the Lee-Carter WLS model calibrated on the period 1950–2023 and ages 50–95, and an ARIMA(1,1,0) model to forecast the period index.

[39] This issue will be discussed in Section 3.2.5 on page 50.

Figure 16: Value of the annuity contract (France/UK, age 60, $\varrho_t = 3\%$)



Probability distribution of annuity values The probability distribution of annuity values is estimated by computing the empirical distribution of the simulated values $a^{(s)}(x,t)$. In what follows, we consider the valuation date 2023, ensuring that the calculation is fully out-of-sample. To calibrate the Lee-Carter WLS model, we use the 30-year period from 1994 to 2023 and we obtain the estimates of α_x, β_x and κ_t. The simulated values of κ_{t+h} are generated using an ARIMA(1,1,0) model fitted on the last 15 observations of κ_t (from 2009 to 2023). Figure 87 on page 172 shows 100 simulated paths of κ_{t+h} from 2024 to 2050. For each simulated path, we compute the annuity value:

$$a^{(s)}(x,t) = B \sum_{h=1}^{H} \frac{1}{(1+\varrho_t)^h} \left(\prod_{j=1}^{h} \exp\left(-\exp\left(\alpha_{x+j-1} + \beta_{x+j-1} \kappa_{t+j-1}^{(s)} \right) \right) \right) \tag{12}$$

where $\left(\kappa_{t+1}^{(s)}, \ldots, \kappa_{t+H}^{(s)}\right)$ is the s^{th} simulated path of the time index. Equation (12) is identical to Equation (10), except that the average forecast $\hat{\kappa}_{t+j}$ is replaced by the simulated value $\kappa_{t+j}^{(s)}$. Figure 17 shows the probability density function of the annuity value[40], estimated with $n_S = 50\,000$ simulations. We note that it is close to the Gaussian approximation.

Table 10 reports several statistics about the probability distribution of the annuity value: the deterministic annuity value $a(x,t)$, the average annuity value[41], the Jensen gap[42] (in bps), the standard deviation, the quantile at the 1%, 50% (median) and 99% confidence

[40] We use $B = 1$, $\varrho_t = 3\%$ and $x^+ = 95$.

[41] It is equal to $\mu\left(a^{(s)}(x,t)\right) = \frac{1}{n_S}\sum_{s=1}^{n_S} a^{(s)}(x,t)$.

[42] It is equal to $\mathcal{J}_{a(x,t)} = \frac{\mu\left(a^{(s)}(x,t)\right)}{a(x,t)} - 1$.

Figure 17: Probability density function of the annuity value (France, valuation year 2023, issue age 65, $\varrho_t = 3\%$)

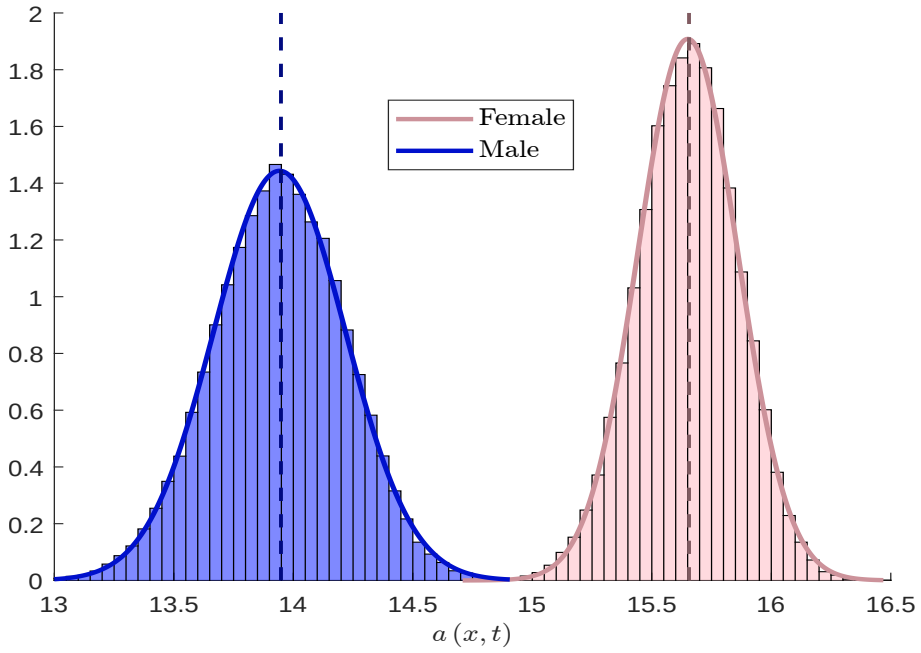


Table 10: Statistics of the stochastic annuity value (France, 2023)

	Issue age x	$a\,(x,t)$	$\mu\left(a^{(s)}\,(x,t)\right)$	$\mathcal{J}_{a(x,t)}$ (in bps)	$\sigma\left(a^{(s)}\,(x,t)\right)$	$Q_{1\%}\left(a^{(s)}\,(x,t)\right)$	$Q_{50\%}\left(a^{(s)}\,(x,t)\right)$	$Q_{99\%}\left(a^{(s)}\,(x,t)\right)$	$\text{CTE}_{99\%}\left(a^{(s)}\,(x,t)\right)$
Female	55	19.615	19.609	−3.0	0.206	19.111	19.614	20.069	20.131
	60	17.740	17.735	−3.2	0.211	17.224	17.738	18.207	18.272
	65	15.656	15.651	−3.3	0.209	15.148	15.654	16.122	16.187
	70	13.337	13.333	−3.4	0.196	12.865	13.336	13.774	13.837
	75	10.810	10.806	−3.3	0.169	10.405	10.809	11.192	11.246
	80	8.154	8.151	−2.9	0.128	7.847	8.153	8.445	8.486
	85	5.552	5.550	−2.2	0.077	5.366	5.551	5.729	5.753
	90	3.084	3.084	−1.3	0.029	3.015	3.084	3.151	3.160
Male	55	17.930	17.923	−4.4	0.300	17.194	17.928	18.595	18.688
	60	15.997	15.990	−4.4	0.293	15.286	15.993	16.652	16.744
	65	13.948	13.942	−4.3	0.276	13.286	13.946	14.571	14.658
	70	11.740	11.735	−4.0	0.244	11.159	11.736	12.291	12.369
	75	9.388	9.385	−3.5	0.196	8.925	9.386	9.834	9.897
	80	6.994	6.992	−2.8	0.138	6.671	6.993	7.309	7.353
	85	4.733	4.732	−1.8	0.076	4.555	4.732	4.906	4.932
	90	2.691	2.691	−1.0	0.028	2.627	2.691	2.754	2.763

levels, and the conditional tail expectation[43]. Several observations can be made from Table 10. First, the Monte Carlo average annuity value is very close to the deterministic annuity value: $\mu\left(a^{(s)}(x,t)\right) \approx a(x,t)$. The Jensen gap, which measures the discrepancy between these two estimates, is less than five basis points in absolute value. Second, the standard deviation of the Monte Carlo annuity value generally decreases with the issue age x. Third, the 1%–99% inter-quantile range is approximately five times the standard deviation, indicating that the Gaussian approximation does not accurately capture the tail behavior of the distribution. Given the uncertainty surrounding the annuity value, one may question whether the average annuity value constitutes an appropriate price. In actuarial practice, this quantity is referred to as the best estimate (BE). However, it does not correspond to the market price because it excludes several components, such as administrative expenses, profit margins, and the cost of holding regulatory capital. Therefore, insurers may adopt more conservative pricing rules. For example, they may use a high quantile, such as the 99th percentile, or a conditional tail expectation of the annuity value distribution to account for mortality-model risk. The choice of a conservative pricing rule depends strongly on the regulatory framework. For instance, under Solvency II in the European Union, economic capital is calibrated to a 99.5% value-at-risk over a one-year horizon. By contrast, in the United States, capital requirements for annuity products rely more heavily on conditional tail expectation measures.

Until now, the uncertainty was restricted to the time index κ_{t+j}, since the simulation of $a^{(s)}(x,t)$ assumes that all other model parameters are known with certainty and can be treated as constant. However, the Lee-Carter model is subject to additional sources of uncertainty. In particular, $a^{(s)}(x,t)$ can be computed by considering that α_x, β_x, κ_{t+h} and $\varepsilon(x,t+h)$ are stochastic variables. In this case, $a^{(s)}(x,t)$ is computed from the simulated mortality rates $q^{(s)}(x,t+h) = 1 - \exp\left(-m^{(s)}(x,t+h)\right)$, where:

$$\ln m^{(s)}(x,t+h) = \alpha_x^{(s)} + \beta_x^{(s)}\kappa_{t+h}^{(s)} + \varepsilon^{(s)}(x,t+h)$$

In this framework, the age-specific parameters $\alpha_x^{(s)}$ and $\beta_x^{(s)}$, as well as the innovations $\varepsilon^{(s)}(x,t+h)$ are treated as random. This differs from the previous approach, where only the time index $\kappa_{t+h}^{(s)}$ was assumed to be stochastic. In Appendix C.6 on page 144, we propose a methodology for simulating paths of $\ln m^{(s)}(x,t+h)$ generated by the joint simulation of $\left(\alpha_x^{(s)}, \beta_x^{(s)}, \kappa_{t+h}^{(s)}, \varepsilon^{(s)}(x,t+h)\right)$. Results are reported in Table 11. A comparison with Table 10 indicates that including uncertainty on α_x, β_x, and $\varepsilon(x,t+h)$ has a negligible impact on the results. Indeed, we can show that $\operatorname{var}\left(\varepsilon^{(s)}(x,t+h)\right) = \sigma_\varepsilon^2$, $\operatorname{var}\left(\alpha_x^{(s)}\right) = \frac{1}{T}\sigma_\varepsilon^2$ and $\operatorname{var}\left(\beta_x^{(s)}\right) = \frac{1}{\sum_{t=1}^{T}\kappa_t^2}\sigma_\varepsilon^2$. It follows that $\operatorname{var}\left(\beta_x^{(s)}\right) \ll \operatorname{var}\left(\alpha_x^{(s)}\right) \ll \operatorname{var}\left(\varepsilon^{(s)}(x,t+h)\right)$. Moreover, empirical results demonstrate that $\operatorname{var}\left(\varepsilon^{(s)}(x,t+h)\right) \ll \operatorname{var}\left(\kappa_{t+h}^{(s)}\right)$ while the magnitude of σ_ε^2 is small. These findings suggest that the dominant source of uncertainty in the Lee-Carter framework is the future evolution of the time index κ_{t+h}. By contrast, the uncertainty associated with α_x, β_x, and $\varepsilon(x,t+h)$ is of secondary importance and has only a limited effect on annuity valuations.

[43] It is defined as follows:

$$\mathrm{CTE}_\alpha\left(a^{(s)}(x,t)\right) = \mathbb{E}\left[a^{(s)}(x,t) \mid a^{(s)}(x,t) > Q_\alpha\left(a^{(s)}(x,t)\right)\right]$$

Table 11: Statistics of the stochastic annuity value (France, Monte Carlo method based on the joint simulation of $\alpha_x^{(s)}$, $\beta_x^{(s)}$, $\kappa_{t+h}^{(s)}$, and $\varepsilon^{(s)}(x, t+h)$)

	Issue age x	$a(x,t)$	$\mu\left(a^{(s)}(x,t)\right)$	$\mathcal{I}_{a(x,t)}$ (in bps)	$\sigma\left(a^{(s)}(x,t)\right)$	$Q_{1\%}\left(a^{(s)}(x,t)\right)$	$Q_{50\%}\left(a^{(s)}(x,t)\right)$	$Q_{99\%}\left(a^{(s)}(x,t)\right)$	$\text{CTE}_{99\%}\left(a^{(s)}(x,t)\right)$
Female	55	19.615	19.608	−3.8	0.206	19.106	19.612	20.062	20.124
	60	17.740	17.733	−4.0	0.211	17.223	17.737	18.204	18.266
	65	15.656	15.649	−4.0	0.209	15.148	15.653	16.120	16.181
	70	13.337	13.332	−4.0	0.197	12.865	13.334	13.777	13.835
	75	10.810	10.806	−4.0	0.170	10.404	10.807	11.196	11.247
	80	8.154	8.151	−4.0	0.130	7.844	8.151	8.449	8.492
	85	5.552	5.550	−3.7	0.081	5.359	5.550	5.737	5.765
	90	3.084	3.083	−2.9	0.035	3.001	3.084	3.164	3.176
Male	55	17.930	17.917	−7.8	0.300	17.196	17.922	18.588	18.677
	60	15.997	15.984	−8.0	0.293	15.280	15.989	16.644	16.735
	65	13.948	13.937	−8.0	0.277	13.279	13.941	14.566	14.654
	70	11.740	11.731	−7.5	0.245	11.152	11.734	12.289	12.373
	75	9.388	9.382	−6.6	0.199	8.913	9.383	9.839	9.907
	80	6.994	6.990	−5.5	0.142	6.657	6.991	7.318	7.369
	85	4.733	4.731	−4.6	0.083	4.536	4.731	4.923	4.951
	90	2.691	2.690	−3.7	0.038	2.602	2.690	2.778	2.791

Model comparison In the previous sections, the annuity values were computed using the Lee-Carter WLS model to estimate the parameters $(\alpha_x, \beta_x, \kappa_t)$ and an ARIMA(1,1,0) process to generate the time index path κ_{t+h}. Moreover, the parameters were calibrated over the last 30 years (1994–2023), while the ARIMA model was calibrated with the last 15 observations of κ_t. Here, we consider several alternative models. To obtain more stable results, we calibrate the parameters over the 74-year period 1950–2023.

Table 12: Comparison of the stochastic annuity value between mortality models (France, issue age 65, valuation date 2023)

	Model	Method	$a(x,t)$	$\mu\left(a^{(s)}(x,t)\right)$	$\mathcal{I}_{a(x,t)}$ (in bps)	$\sigma\left(a^{(s)}(x,t)\right)$	$Q_{1\%}\left(a^{(s)}(x,t)\right)$	$Q_{50\%}\left(a^{(s)}(x,t)\right)$	$Q_{99\%}\left(a^{(s)}(x,t)\right)$	$\text{CTE}_{99\%}$
Female	LC	WLS	15.791	15.786	−3.1	0.172	15.376	15.791	16.174	16.230
		Poisson	15.932	15.926	−3.8	0.206	15.430	15.930	16.380	16.442
		NB	15.785	15.780	−3.0	0.179	15.349	15.782	16.182	16.237
	CBD	OLS	15.446	15.444	−1.8	0.178	15.019	15.447	15.848	15.901
		Poisson	15.454	15.450	−2.1	0.175	15.035	15.452	15.849	15.902
		NB WLS	15.480	15.475	−3.6	0.177	15.054	15.478	15.875	15.928
Male	LC	WLS	14.336	14.331	−3.8	0.255	13.719	14.333	14.902	14.978
		Poisson	14.292	14.283	−6.3	0.281	13.610	14.290	14.916	15.002
		NB	14.312	14.306	−4.3	0.261	13.689	14.309	14.893	14.976
	CBD	OLS	13.742	13.739	−2.4	0.270	13.096	13.740	14.357	14.446
		Poisson	13.836	13.832	−2.9	0.270	13.199	13.832	14.449	14.541
		NB WLS	13.781	13.775	−3.9	0.272	13.137	13.776	14.397	14.495

Results are reported in Table 12. At age 65, the three LC specifications imply female annuity values ranging from 15.785 to 15.932, while the three CBD specifications imply values between 15.444 and 15.480. For males, the corresponding ranges are 14.283–14.336 under the LC model and 13.739–13.836 under the CBD model. Thus, the LC model yields higher annuity values than the CBD model. It is also interesting to note the sensitivity of the results to the calibration sample. Wen the age-specific parameters were calibrated over the shorter period 1994–2023, we obtained values of 15.656, 15.651, 13.948, and 13.942. Using the longer 1950–2023 sample, these values become 15.791, 15.786, 14.336, and 14.331, respectively. The dispersion ranking is approximately in line with the level ranking. LC Poisson exhibits the largest female/male standard deviation. Overall, these differences between models are substantially larger than the deterministic versus stochastic mean difference observed within any individual specification.

Remark 12 *To relate differences in projected mortality to annuity values and to understand the discrepancies between models, we rewrite Equation (8) for a fixed future mortality path as:*

$$a\left(x,t\right)=B\sum_{h=1}^{H}\frac{1}{\left(1+\varrho_{t}\right)^{h}}\prod_{j=1}^{h}\left(1-q_{j}\right)$$

where $q_{j}=q\left(x+j-1,t+j-1\right)$*. For any mortality horizon* u*, we have:*

$$\frac{\partial\, a\left(x,t\right)}{\partial\, q_{u}}=-B\sum_{h=u}^{H}\frac{1}{\left(1+\varrho_{t}\right)^{h}}\prod_{j=1,j\neq u}^{h}\left(1-q_{j}\right)=-\frac{B}{\left(1-q_{u}\right)}\sum_{h=u}^{H}\frac{p\left(x,t,h\right)}{\left(1+\varrho_{t}\right)^{h}}<0$$

A higher one-year death probability at horizon u *reduces both the expected payment at that horizon and every subsequent survival-contingent payment. Mortality differences occurring earlier in the cohort's projected path therefore carry greater weight in the valuation.*

For model $\mathcal{M}$, we define the average one-year death probability and the average survival probability as follows:

$$q_{\mathcal{M}}\left(x+h-1,t+h-1\right)=\frac{1}{n_{S}}\sum_{s=1}^{n_{S}}q_{\mathcal{M}}^{(s)}\left(x+h-1,t+h-1\right)$$

and:

$$p_{\mathcal{M}}\left(x,t,h\right)=\frac{1}{n_{S}}\sum_{s=1}^{n_{S}}p_{\mathcal{M}}^{(s)}\left(x,t,h\right)\neq\prod_{j=1}^{h}\left(1-q_{\mathcal{M}}\left(x+j-1,t+j-1\right)\right)$$

Note that $p_{\mathcal{M}}\left(x,t,h\right)$ is not the product of the $q_{\mathcal{M}}$ terms, since averaging is performed at the level of each simulated path rather than on the average death probabilities themselves. The contribution of horizon h to the annuity value is then:

$$c_{\mathcal{M}}\left(x,t,h\right)=\frac{B}{\left(1+\varrho_{t}\right)^{h}}p_{\mathcal{M}}\left(x,t,h\right)$$

so that $a\left(x,t\right)\approx\sum_{h=1}^{H}c_{\mathcal{M}}\left(x,t,h\right)$. To compare two models $\mathcal{M}_{1}$ and $\mathcal{M}_{2}$, we define the valuation gap at horizon h as follows:

$$\mathcal{G}ap_{\mathcal{M}_{2}|\mathcal{M}_{1}}\left(x,t,h\right)=c_{\mathcal{M}_{1}}\left(x,t,h\right)-c_{\mathcal{M}_{2}}\left(x,t,h\right)$$

The scenario-mean curve $q_{\mathcal{M}}$ serves as a mortality diagnostic, whereas $p_{\mathcal{M}}\left(x,t,h\right)$ and $c_{\mathcal{M}}\left(x,t,h\right)$ provide an exact decomposition of the stochastic annuity value.

Figure 18: Valuation gap analysis (France, female, age 60, baseline model = LC WLS)

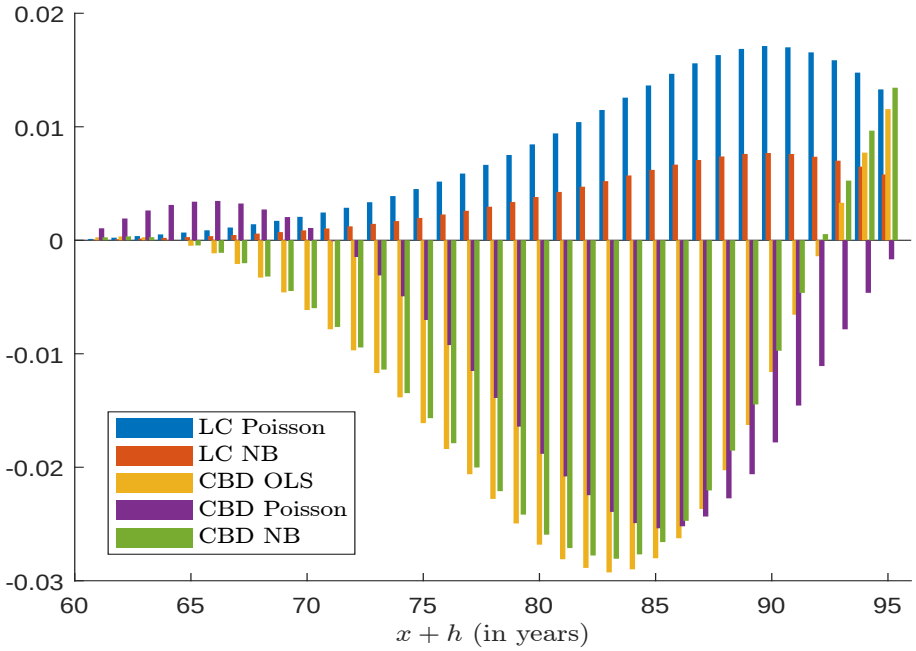
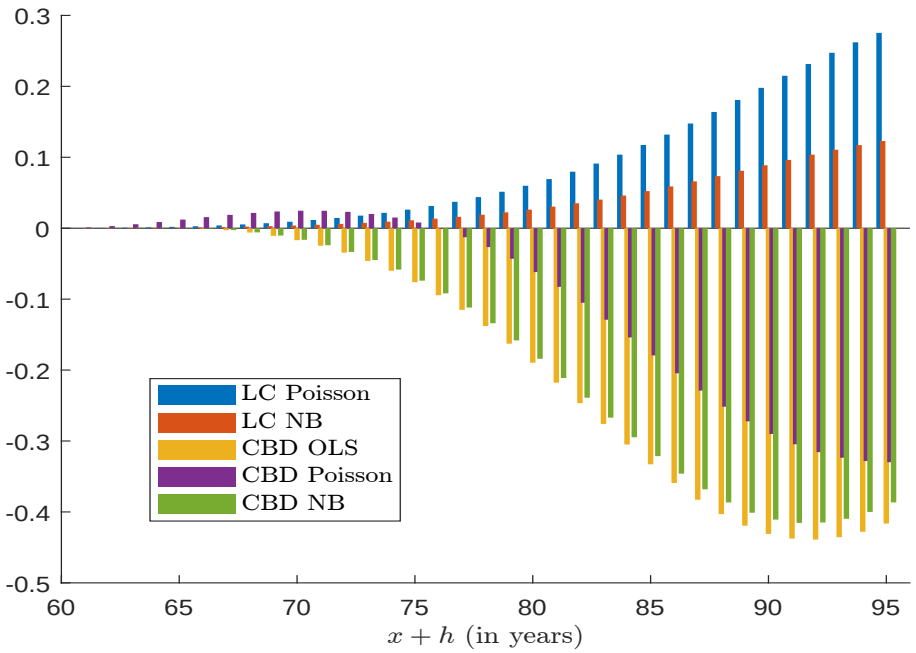


Figure 19: Cumulative gap analysis (France, female, age 60, baseline model = LC WLS)

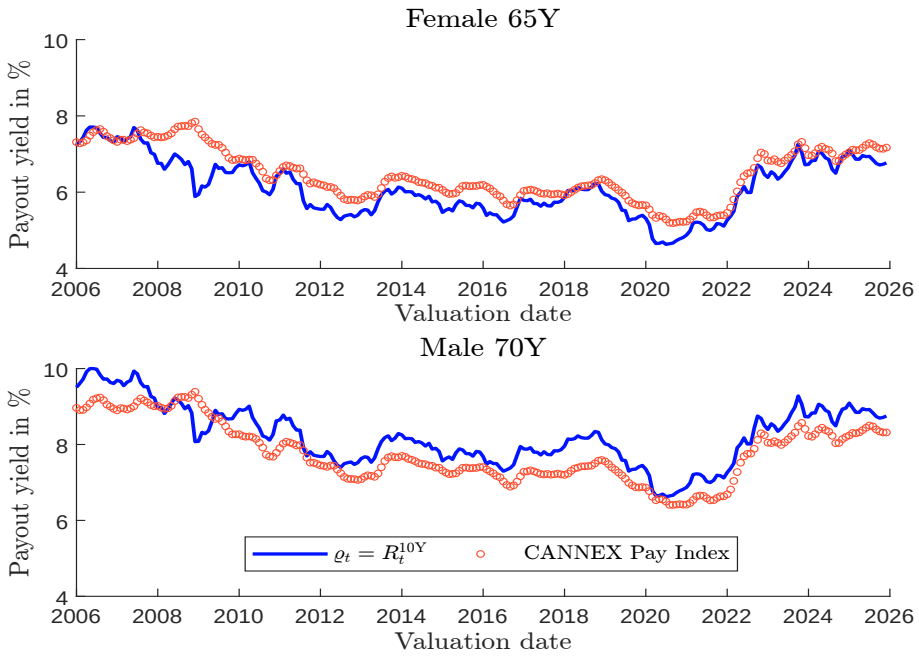


Remark 13 *In Figures 18 and 19, we report the valuation gap* $\mathcal{G}ap_{\mathcal{M}_2|\mathcal{M}_1}(x,t,h)$ *and the cumulative gap* $\sum_{j=1}^{h}\mathcal{G}ap_{\mathcal{M}_2|\mathcal{M}_1}(x,t,j)$ *for the five models (LC Poisson, LC NB, CBD OLS, CBD Poisson WLS, and CBD NB WLS) relative to the LC WLS model, which serves as the baseline. We observe that the gap is maximum for ages 85–95 for the alternative LC models, and for ages 75–90 for the CBD models. However, these observations depend heavily on the historical sample used. For example, for the male group, the results for the alternative LC models differ substantially from those obtained for the female group, whereas the CBD models remain consistent (Figures 89 and 90 on page 173).*

3.2.5 Discount rate, risk premium and mortality credit

Evidence of time-varying discount rate Until now, we have assumed that the discount rate is constant at 3%. However, under this assumption, it is impossible to replicate the temporal dynamics of market annuity premiums as observed in practice. For instance, Figure 20 shows the evolution of the CANNEX payout annuity yield index[44]. It is highly time-varying, but this volatility cannot be explained by changes in survival probabilities $p(x,t,h)$, since they vary little from month to month. Considering the formula for the

Figure 20: CANNEX payout annuity yield index in % (USA, 2005–2025)



Source: CANNEX (2026) & Authors' calculations.

annuity premium, the only explanation for the high volatility of $a(x,t)$ is that the discount rate ϱ_t depends on the valuation date. Economic theory tells us that the discount rate should track the prevailing macroeconomic interest rate environment: the discount rate should be high when interest rates are high, and low when interest rates are low. In fact, most actuarial research papers replace the discount rate ϱ_t with an interest rate r_t in the annuity valuation formula. This is the case, for instance, in several handbooks[45]. However,

[44] The annualized payout yield (or rate) of an annuity contract is defined as $y_{x,t} = B/a(x,t)$ and is expressed in %.

[45] See for instance Pitacco *et al.* (2009, Section 1.3.2, pages 9–12) and Milevsky (2013, Q17, pages 49–51).

most research papers do not clearly distinguish between the two terms, and it is often unclear which interest rate is used (a short-term or a long-term rate). To formalize the connection with financial markets, we generalize the annuity pricing formula by introducing a term structure of discount rates:

$$a(x,t) = B \sum_{h=1}^{H} \frac{p(x,t,h)}{\left(1+\varrho_{t,h}\right)^h}$$

where $\varrho_{t,h}$ is the annualized discount rate for time horizon h at valuation date t. A standard approach is to set the discount rate $\varrho_{t,h}$ to a constant interest rate depending only on the valuation date. A typical benchmark is the 10-year interest rate: $\varrho_{t,h} = R_t^{10Y}$. With this assumption, we can reproduce the volatility of the CANNEX payout annuity yield index, as shown in Figure 20. While the levels are not exactly the same[46], the dynamics are very close. The empirical correlation between the CANNEX index and our model with $\varrho_{t,h} = R_t^{10Y}$ is equal to 92.3% and 92.8% for females and males, respectively. A second approach[47] is to set $\varrho_{t,h} = R_t^h$, where R_t^h denotes the interest rate observed at time t for maturity $t+h$. Using $\varrho_{t,h} = R_t^h$ instead of $\varrho_{t,h} = R_t^{10Y}$ has only a minor quantitative impact on the calculated annuity values, as shown by the scatter plot in Figure 92 on page 174. Finally, a third approach is to consider a time horizon equal to the duration of the cash flows, defined as the survival-probability-weighted average maturity of the annuity payments:

$$\bar{h} = \left(\sum_{h=1}^{H} p(x,t,h) \cdot h\right) / \left(\sum_{h=1}^{H} p(x,t,h)\right)$$

Then, we set $\varrho_{t,h} = R_t^{\bar{h}}$, *i.e.*, a single interest rate observed at the maturity $\bar{h}$ corresponding to this duration. Again, the impact on the annuity premium is relatively minor in our example since $\bar{h}$ ranges from 9.72 to 10.20 years for males and from 12.92 to 13.44 years for females over the period 2006–2025. Since $\bar{h}$ is close to ten years throughout the sample, replacing $\varrho_{t,h} = R_t^{10Y}$ with $\varrho_{t,h} = R_t^{\bar{h}}$ does not materially change the annuity premium $a(x,t)$ and the payout yield $y_{x,t}$.

Computation of the risk premium Figure 21 compares the evolution of the annuity value under two assumptions about the discount rate: a constant rate $\varrho_{t,h} = 3\%$ and a time-varying rate $\varrho_{t,h} = R_t^{10Y}$. The left panel is based on the constant discount rate and isolates the effect of improving survival probabilities over time, since the discount rate is held fixed and cannot contribute to the observed trend. The right panel is based on $\varrho_t = R_t^{10Y}$ and combines the survival improvement effect with the fluctuations of the 10-year interest rate. Its year-to-year movements are mainly explained by the latter. In this context, the absolute value of the annuity premium $a(x,t)$ is difficult to interpret since it results from the combination of two distinct factors:

1. A demographic factor driven by the level and trend of survival probabilities $p(x,t,h)$, which reflects mortality improvements and is slow-moving and highly persistent over time;

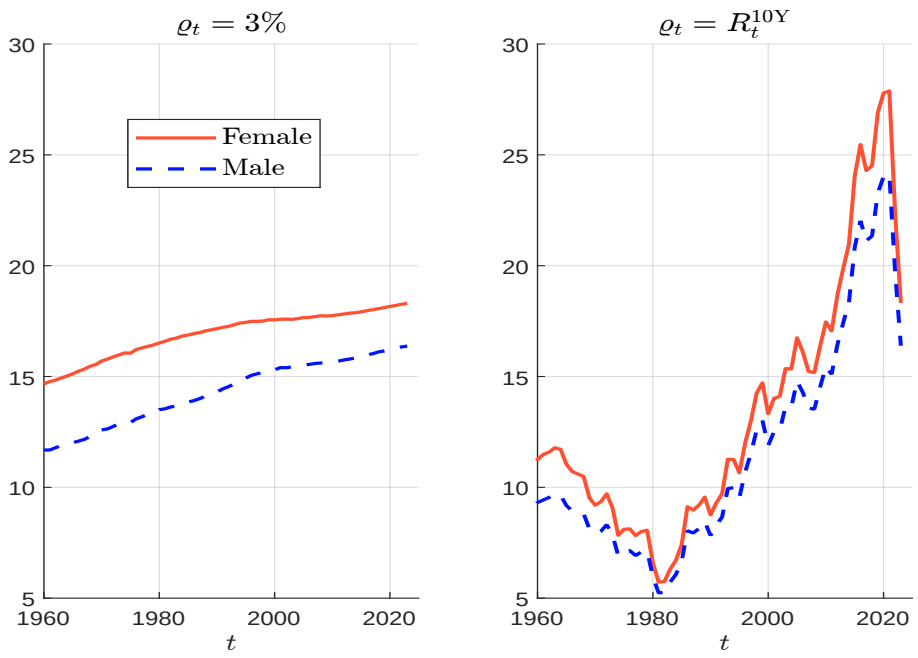
2. A financial factor driven by the level of the discount rate $\varrho_{t,h}$, which reflects the prevailing interest rate environment and can vary substantially and relatively quickly from one valuation date to another.

As these two factors operate on different timescales and have opposite implications for insurers and annuitants, it is essential to disentangle them when interpreting changes in

[46] We assume that $x^+ = 95$ years, which may explain some of the observed differences.

[47] The first approach is generally known as the benchmark method, and the second approach is called the yield curve method.

Figure 21: Impact of the discount rate on the annuity premium (France, age 60)



$a(x,t)$ over time. For that, we generally calculate the internal rate of return $i_{x,t}$, which is defined as the solution of the following equation:

$$B \sum_{h=1}^{H} \frac{p(x,t,h)}{(1+i_{x,t})^h} = a(x,t) \tag{13}$$

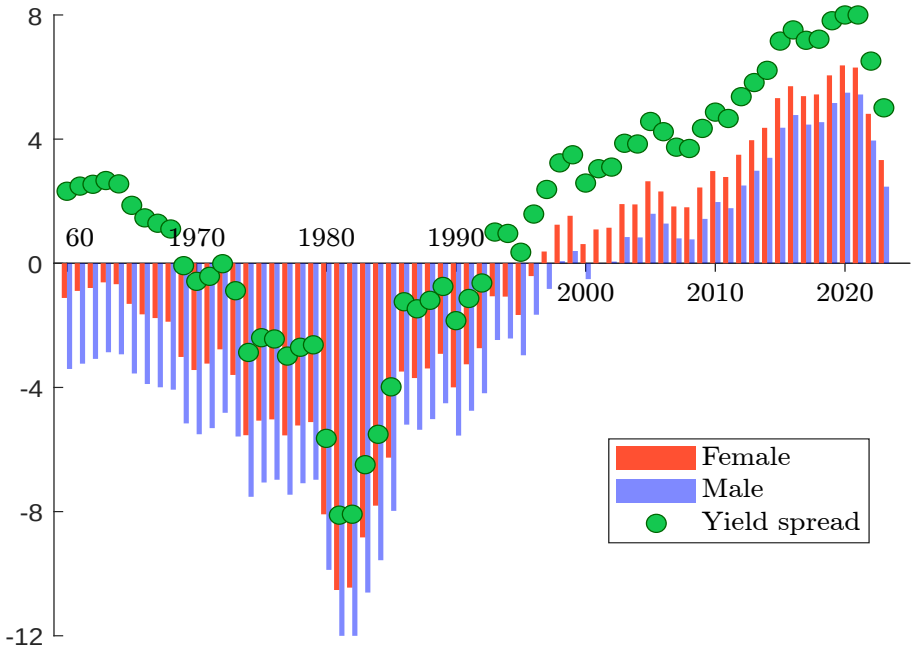
where $a(x,t)$ is the market value of the annuity premium. By construction, $i_{x,t}$ is the single flat rate that reproduces the market premium after accounting for the demographic factor. Therefore, it can be interpreted as the purely financial component embedded in the annuity premium. To assess whether this financial component is consistent with prevailing market conditions, we compare $i_{x,t}$ to the interest rate observed for a maturity matching the duration of the annuity cash flows:

$$\pi(x,t) = i_{x,t} - R_t^h$$

The risk premium $\pi(x,t)$ measures the extent to which the implied return of the annuity contract differs from what a comparable investment would offer at the same valuation date. It can be interpreted as compensation for mortality uncertainty, liquidity, and model risks borne by the insurer.

Figure 22 shows how the risk premium $\pi(x,t)$ evolves when we assume that the annuity premium is equal to 12.5 for both females and males at issue age 60. We use the 10-year interest rate as the benchmark rate ($R_t^h = R_t^{10Y}$). Overall, the risk premium is negative before 2000 and positive after 2000. From the viewpoint of the annuitant, the risk premium measures the expected excess return of buying the annuity rather than investing at the duration-matched interest rate R_t^h. Generally, we would expect $\pi(x,t)$ to be negative, since insurer typically pass on their expenses and profit margins to annuitants through pricing. Therefore, the fact that $\pi(x,t)$ turns positive after 2000 appears counterintuitive at first sight, but it coincides with a period of low and even negative long-term interest rates.

Figure 22: Annuity risk premium (France, age 60)



One issue concerning $\pi(x,t)$ is that the internal rate of return $i_{x,t}$ is not always well-defined. Indeed, a solution to Equation (13) exists if and only if $a(x,t) \leq a^{+}(x,t)$, where $a^{+}(x,t) = B\sum_{h=1}^{H} p(x,t,h)$ is the upper bound of the annuity premium reached when the discount rate is zero. When the annuity premium exceeds this bound, *i.e.*, $a(x,t) > a^{+}(x,t)$, no solution $i_{x,t}$ can reproduce the observed price, and the risk premium $\pi(x,t)$ cannot be computed. In this case, we can use an alternative measure:

$$\pi'(x,t) = y_{x,t} - R_t^h$$

where $y_{x,t} = B/a(x,t)$ is the annuity payout yield. However, unlike $\pi(x,t)$, $\pi'(x,t)$ is not a pure financial risk premium since it does not isolate the pricing loading net of mortality. For this reason, we refer to $\pi'(x,t)$ as the (gross) annuity yield spread, as opposed to the (net) risk premium $\pi(x,t)$:

- $\pi(x,t) = i_{x,t} - R_t^h$ isolates the pricing loading net of mortality.
- $\pi'(x,t) = y_{x,t} - R_t^h$ bundles the pricing loading together with the mortality credit.

Generally, we have $\pi'(x,t) \geq \pi(x,t)$. This is confirmed in Figure 22, where $\pi'(x,t)$ is represented by green circle symbols and consistently lies above $\pi(x,t)$.

Link to the mortality credit In fact, the difference between the yield spread and the risk premium isolates the mortality credit embedded in the annuity:

$$\pi'(x,t) - \pi(x,t) = y_{x,t} - i_{x,t}$$

To understand this relationship, let us calculate the present value of a perpetual life annuity with a constant discount rate ϱ_t and a constant mortality hazard rate λ_t, evaluated at age x and time t:

$$a(x,t) = B\int_x^{\infty} e^{-(\varrho_t+\lambda_t)(u-x)}\,\mathrm{d}u = \frac{B}{\varrho_t+\lambda_t}$$

The corresponding payout yield is equal to:

$$y_{x,t} = \frac{B}{a\left(x,t\right)} = \varrho_t + \lambda_t$$

It follows that the mortality credit is equal to:

$$\lambda_t = y_{x,t} - \varrho_t = y_{x,t} - i_{x,t}$$

because $i_{x,t} = \varrho_t$. We deduce that the excess yield spread over the risk premium equals the mortality hazard rate:

$$\pi'\left(x,t\right) - \pi\left(x,t\right) = \lambda_t$$

Later, in Section 4.2.1 on page 66, we will see that this is exactly the definition of the mortality credit.

3.3 Portfolio risk analysis

In this section, we extend the single-contract valuation framework to a collective portfolio of annuities. The focus shifts from individual expected cash flows to the random present value of aggregate benefit payments, which encapsulates non-systematic variations in individual lifespans and macro-level uncertainty in future mortality rates. Using this framework, we show that the mortality risk of an aggregate portfolio can be formally decomposed into two components: idiosyncratic risk (independent longevity risk arising from individual fluctuations in survival times) and systematic risk (non-diversifiable uncertainty governing population-wide longevity trends). This risk decomposition plays a fundamental role in evaluating the insurer's liability distribution. In particular, idiosyncratic risk dominates the central tendency of small portfolios, while systematic longevity risk dictates the heavy-tail dynamics of large portfolios.

3.3.1 Portfolio liability modeling

For a given demographic category (*e.g.*, gender), we consider a portfolio consisting of n_G issue-age groups. Each age group $g \in \{1,\ldots,n_G\}$ contains N_g annuitants aged x_g at the valuation date t. The total number of contracts in the portfolio is given by $N = \sum_{g=1}^{n_G} N_g$. All contracts follow the annuity-immediate specification introduced in Equation (8) on page 40. For group g, the maximum payment horizon is $H_g = x^+ - x_g$, where x^+ is the limiting age. Let $N_{g,h}$ be the random variable representing the number of annuitants in group g who survive to payment horizon h. Following Olivieri (2001), the portfolio liability $L\left(t\right)$ is represented as the aggregate random present value of the constituent annuities:

$$L\left(t\right) = \sum_{g=1}^{n_G}\sum_{h=1}^{H_g} \frac{B}{\left(1+\varrho_t\right)^h} N_{g,h} \tag{14}$$

Given the information set $\mathcal{F}_t$, the best estimate of the portfolio liability is equal to the sum of the individual actuarial present values weighted by group size:

$$\bar{L}\left(t\right) = \mathbb{E}\left[L\left(t\right) \mid \mathcal{F}_t\right] = \sum_{g=1}^{n_G} N_g a\left(x_g,t\right) \tag{15}$$

The proof is given in Appendix C.7 on page 146.

Portfolio liability risk has two fundamental sources:

1. Idiosyncratic mortality risk arises from variations in individual survival times relative to the conditional expected lifespan given a future mortality path. Following Olivieri (2001), individual lifetimes are assumed to be independent conditional on this path, so idiosyncratic risk can be reduced by pooling ($N \to \infty$).

2. Systematic mortality risk arises from uncertainty in the future mortality path itself, and affects all annuitants exposed to that path. Therefore, it cannot be diversified away within the portfolio (Cairns *et al.*, 2006b).

Let $\mathcal{P}$ denote the random future mortality path. Olivieri (2001) applies the law of total variance to separate the two sources of portfolio mortality risk:

$$\underbrace{\operatorname{var}\left(L\left(t\right) \mid \mathcal{F}_t\right)}_{\sigma^2_{\text{total}}(t)} = \underbrace{\mathbb{E}\left[\operatorname{var}\left(L\left(t\right) \mid \mathcal{F}_t, \mathcal{P}\right)\right]}_{\sigma^2_{\text{idiosyncratic}}(t)} + \underbrace{\operatorname{var}\left(\mathbb{E}\left[L\left(t\right) \mid \mathcal{F}_t, \mathcal{P}\right]\right)}_{\sigma^2_{\text{systematic}}(t)} \tag{16}$$

where $\sigma^2_{\text{idiosyncratic}}\left(t\right)$ and $\sigma^2_{\text{systematic}}\left(t\right)$ define the idiosyncratic variance (pooling/sampling risk) and systematic variance (trend/longevity risk) components, respectively.

3.3.2 Risk measurement

The systematic and idiosyncratic variance components are estimated from the conditional mean and variance of the portfolio liability under each simulated mortality path. Individual deaths do not need to be simulated for this decomposition. They are simulated separately to obtain the distribution of realized portfolio liabilities.

Estimation of the variance components Let $\mathcal{P}^{(s)}$ denote the simulated future mortality path under scenario s, for $s = 1, \ldots, n_S$. The probability distribution of future mortality is approximated by the same simulated mortality paths used in the stochastic annuity valuation. For mortality scenario s, the portfolio best estimate liability is given by:

$$\bar{L}^{(s)}\left(t\right) = \sum_{g=1}^{n_G} N_g a^{(s)}\left(x_g, t\right)$$

The systematic variance component is estimated from the dispersion of these conditional portfolio liabilities:

$$\hat{\sigma}^2_{\text{systematic}}\left(t\right) = \frac{1}{n_S - 1} \sum_{s=1}^{n_S} \left(\bar{L}^{(s)}\left(t\right) - \frac{1}{n_S} \sum_{s=1}^{n_S} \bar{L}^{(s)}\left(t\right) \right)^2$$

To compute the idiosyncratic component, we denote by $\operatorname{var}^{(s)}\left(x_g, t\right)$ the conditional variance of a single contract in group g under scenario s. The expression of $\operatorname{var}^{(s)}\left(x_g, t\right)$ is given in Appendix C.8 on page 146. It follows that the estimator of the idiosyncratic variance is:

$$\hat{\sigma}^2_{\text{idiosyncratic}}\left(t\right) = \frac{1}{n_S} \sum_{s=1}^{n_S} \sum_{g=1}^{n_G} N_g \operatorname{var}^{(s)}\left(x_g, t\right)$$

Therefore, the estimated total variance is:

$$\hat{\sigma}^2_{\text{total}}\left(t\right) = \hat{\sigma}^2_{\text{idiosyncratic}}\left(t\right) + \hat{\sigma}^2_{\text{systematic}}\left(t\right)$$

while the relative variance shares are $w_{\text{idiosyncratic}} = \hat{\sigma}^2_{\text{idiosyncratic}}\left(t\right) / \hat{\sigma}^2_{\text{total}}\left(t\right)$ and $w_{\text{systematic}} = 1 - w_{\text{idiosyncratic}}$.

Estimation of the probability distribution The full probability distribution of realized portfolio liabilities includes systematic uncertainty in mortality trends and idiosyncratic variations in individual lifespans. For each simulated mortality scenario s ($s = 1, \ldots, n_S$), the sequence of scenario-specific one-year death probabilities $q^{(s)}\left(x_g + h - 1, t + h\right)$ for time horizon $h = 1, \ldots, H_g$ is fixed for each group g. Given a mortality path $\mathcal{P}^{(s)}$, the simulation algorithm generates n_R independent realizations of portfolio deaths. To optimize computational efficiency without loss of mathematical precision, individual lifespans are not simulated separately. Instead, at each time horizon, the number of deaths in each group is sampled directly from the conditional binomial distribution. Let $N_{g,h}^{(s,r)}$ denote the number of survivors in group g at time horizon h in realization r ($r = 1, \ldots, n_R$) conditional on scenario s, with the initial condition $N_{g,0}^{(s,r)} = N_g$. Let $D_{g,h}^{(s,r)}$ denote the number of deaths occurring in group g during the period $(t + h - 1, t + h]$. Given the survivor count from the preceding horizon and the mortality scenario $\mathcal{P}^{(s)}$, the number of deaths follows a Binomial distribution:

$$D_{g,h}^{(s,r)} \mid N_{g,h-1}^{(s,r)}, \mathcal{P}^{(s)} \sim \text{Binomial}\left(N_{g,h-1}^{(s,r)}, q^{(s)}\left(x_g + h - 1, t + h - 1\right)\right)$$

The survivor count is then recursively updated as:

$$N_{g,h}^{(s,r)} = N_{g,h-1}^{(s,r)} - D_{g,h}^{(s,r)}$$

Each simulation run (s, r) yields a realization of the aggregate portfolio liability:

$$L^{(s,r)}(t) = \sum_{g=1}^{n_G} \sum_{h=1}^{H_g} \frac{B}{(1 + \varrho_t)^h} N_{g,h}^{(s,r)}$$

From these $n_S \times n_R$ simulated portfolio liability values, we estimate two probability distributions:

- The conditional probability distribution given a mortality scenario s is estimated via the empirical distribution function:

$$\hat{\mathbf{G}}^{(s)}(\ell) = \Pr\left\{L^{(s)}(t) \leq \ell\right\} \approx \frac{1}{n_R} \sum_{r=1}^{n_R} \mathbb{1}\left\{L^{(s,r)}(t) \leq \ell\right\}$$

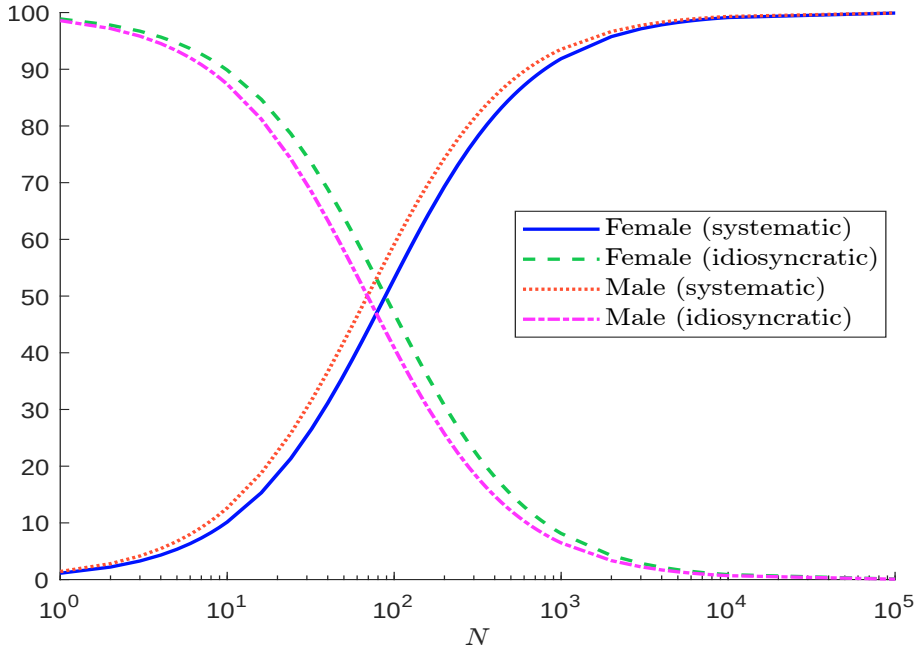
- The unconditional probability distribution is estimated by averaging these conditional empirical distributions across all n_S mortality scenarios:

$$\hat{\mathbf{G}}(\ell) = \Pr\left\{L(t) \leq \ell\right\} \approx \frac{1}{n_S} \sum_{s=1}^{n_S} \hat{\mathbf{G}}^{(s)}(\ell) = \frac{1}{n_S \times n_R} \sum_{s=1}^{n_S} \sum_{r=1}^{n_R} \mathbb{1}\left\{L^{(s,r)}(t) \leq \ell\right\}$$

3.3.3 Empirical results

Idiosyncratic vs. systematic risk We consider the same parameterization as in Section 3.2.4 on page 41. The valuation year is 2023. We use the LC WLS model, calibrated on the last 30 years. The paths are simulated using an ARIMA(1,1,0) process fitted to the last 15 observations of κ_t. The number of simulations is set to 50 000. The annuity pays $B = 1$ annually in arrears, the discount rate is $\varrho_t = 3\%$, and the limiting age is $x^+ = 95$. These are the same valuation and forecasting assumptions used for the single contract analysis. The portfolio contains eight issue age groups: $x_g \in \{55, 60, 65, 70, 75, 80, 85, 90\}$. All the

Figure 23: Variance decomposition of the portfolio liability (France, 2023, LC WLS)



results below are presented for the equally-weighted portfolio. The size of each group is then $N_g = N/n_G$.

Figure 23 shows the evolution of the variance shares $w_{\text{idiosyncratic}}$ and $w_{\text{systematic}}$ with respect to N. We observe that the idiosyncratic share is maximal when N equals one, and then decreases continuously. The systematic share exhibits the opposite pattern. To understand this, let us return to the variance formula[48]:

$$\begin{aligned}\hat{\sigma}^2_{\text{total}}(N) &= \hat{\sigma}^2_{\text{idiosyncratic}}(N) + \hat{\sigma}^2_{\text{systematic}}(N) \\ &= \frac{1}{n_S}\sum_{s=1}^{n_S}\sum_{g=1}^{n_G} N_g \operatorname{var}^{(s)}(x_g,t) + \operatorname{var}\left(\sum_{g=1}^{G} N_g a^{(s)}(x_g,t)\right)\end{aligned}$$

If we multiply the size of the portfolio by $\lambda > 1$, it follows that:

$$\begin{aligned}\hat{\sigma}^2_{\text{total}}(\lambda N) &= \frac{1}{n_S}\sum_{s=1}^{n_S}\sum_{g=1}^{n_G} \lambda N_g \operatorname{var}^{(s)}(x_g,t) + \operatorname{var}\left(\sum_{g=1}^{G} \lambda N_g a^{(s)}(x_g,t)\right) \\ &= \lambda\hat{\sigma}^2_{\text{idiosyncratic}}(N) + \lambda^2\hat{\sigma}^2_{\text{systematic}}(N)\end{aligned}$$

and:

$$\frac{1}{\lambda}\left(\frac{1}{N}\hat{\sigma}^2_{\text{total}}(\lambda N)\right) = \left(\frac{1}{N}\hat{\sigma}^2_{\text{idiosyncratic}}(N)\right) + \lambda\left(\frac{1}{N}\hat{\sigma}^2_{\text{systematic}}(N)\right)$$

We deduce that the normalized variance retains the same amount of normalized idiosyncratic risk, but an increasing amount of systematic risk as λ grows. This property of the empirical estimators can also be established from a theoretical viewpoint. Indeed, for a homogeneous

[48] We introduce the notation $\sigma^2(N)$ to link the variance to the portfolio size N.

portfolio of N annuitants aged x, the variance decomposition in Equation (16) implies:

$$\frac{\text{var}\left(L\left(t\right)\mid\mathcal{F}_t\right)}{\mathbb{E}^2\left[L\left(t\right)\mid\mathcal{F}_t\right]}=\frac{1}{N}\frac{\mathbb{E}\left[\text{var}\left(A\left(x,t\right)\mid\mathcal{F}_t,\mathcal{P}\right)\right]}{\mathbb{E}^2\left[A\left(x,t\right)\mid\mathcal{F}_t\right]}+\frac{\text{var}\left(\mathbb{E}\left[A\left(x,t\right)\mid\mathcal{F}_t,\mathcal{P}\right]\right)}{\mathbb{E}^2\left[A\left(x,t\right)\mid\mathcal{F}_t\right]}$$

The first term decreases proportionally to $1/N$, whereas the second is unaffected by portfolio size. Pitacco *et al.* (2009, Equation 7.18, page 285) obtain this result for an annuity portfolio and identify the two terms as pooling and non-pooling risk, respectively. The same result extends directly to a portfolio with a fixed age composition. Let π_g denote the proportion of annuitants in group g, with $N_g=N\pi_g$ and $\sum_{g=1}^{n_G}\pi_g=1$. Conditional independence of individual lifetimes gives:

$$\begin{aligned}\text{var}\left(L\left(t\right)\mid\mathcal{F}_t\right) &= \mathbb{E}\left[\text{var}\left(L\left(t\right)\mid\mathcal{F}_t,\mathcal{P}\right)\right]+\text{var}\left(\mathbb{E}\left[L\left(t\right)\mid\mathcal{F}_t,\mathcal{P}\right]\right)\\ &= N\sum_{g=1}^{G}\pi_g\mathbb{E}\left[\text{var}\left(A\left(x_g,t\right)\mid\mathcal{F}_t,\mathcal{P}\right)\right]+N^2\,\text{Var}\left(\sum_{g=1}^{G}\pi_g\mathbb{E}\left[A\left(x_g,t\right)\mid\mathcal{F}_t,\mathcal{P}\right]\right)\end{aligned}$$

Thus, for a fixed portfolio composition, idiosyncratic variance grows proportionally to N, whereas systematic variance grows proportionally to N^2. Whenever systematic mortality variance is nonzero, its share of total variance therefore approaches one as the portfolio becomes large. These properties are confirmed by the numerical results given in Table 13. The normalized idiosyncratic risk remains constant, while the normalized systematic risk increases with portfolio size. At $N=1$, idiosyncratic variance accounts for 98.88% and 98.57% of total variance for the female and male groups, respectively. This is normal, because for a single contract, risk is almost entirely idiosyncratic. This balance shifts quickly, however, as systematic risk comes to dominate the portfolio, reaching 90% of total variance once the portfolio holds 1 000 annuity contracts.

Table 13: Variance decomposition of the portfolio liability (France, 2023, LC WLS)

	N	$\frac{\hat{\sigma}^2_{\text{idiosyncratic}}}{N}$	$\frac{\hat{\sigma}^2_{\text{systematic}}(N)}{N}$	$\frac{\hat{\sigma}^2_{\text{total}}}{N}$	$w_{\text{idiosyncratic}}$	$w_{\text{systematic}}$
Female	1	1.87	0.02	1.89	98.88%	1.12%
	10	1.87	0.21	2.08	89.85%	10.15%
	100	1.87	2.11	3.97	46.96%	53.04%
	1 000	1.87	21.08	22.95	8.13%	91.87%
	10 000	1.87	210.81	212.67	0.88%	99.12%
	100 000	1.87	2 108.07	2 109.94	0.09%	99.91%
Male	1	2.38	0.03	2.41	98.57%	1.43%
	10	2.38	0.34	2.72	87.37%	12.63%
	100	2.38	3.44	5.82	40.89%	59.11%
	1 000	2.38	34.39	36.77	6.47%	93.53%
	10 000	2.38	343.93	346.31	0.69%	99.31%
	100 000	2.38	3 439.33	3 441.71	0.07%	99.93%

Tail risk measures Figure 24 shows the probability density function of the portfolio liability under the LC WLS benchmark. It is estimated using $n_S=50\,000$ mortality scenarios and $n_R=50$ death-process replications per scenario. According to the Kolmogorov-Smirnov statistic, the Gaussian approximation is markedly less accurate for the full stochastic liability $L^{(s,r)}\left(t\right)$ than for the best-estimate liability $\bar{L}^{(s)}\left(t\right)$. This is because $\bar{L}^{(s)}\left(t\right)$ already averages

Figure 24: Probability density function of the portfolio liability (France, 2023, LC WLS, $N = 200$)

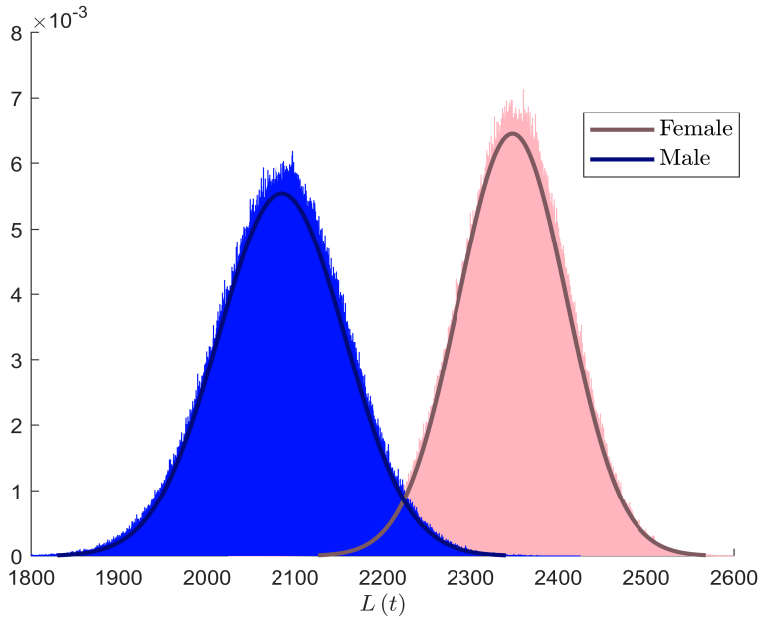
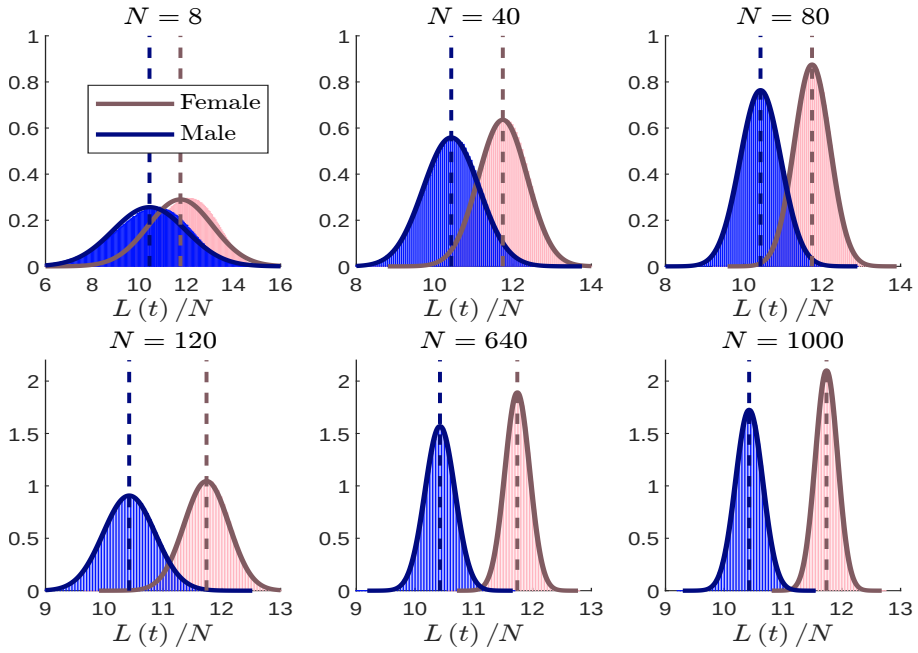


Figure 25: Convergence of the normalized portfolio liability (France, 2023, LC WLS)

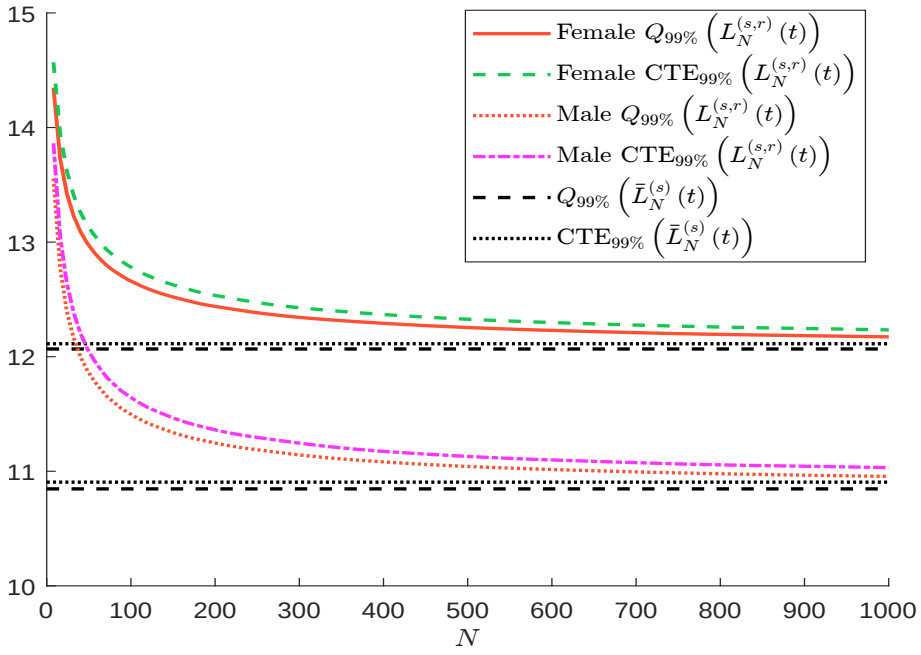


out idiosyncratic mortality risk within each scenario, inheriting the near-Gaussian behavior of the simulated κ_t paths. Conversely, $L^{(s,r)}(t)$ retains the discreteness and skewness of the underlying binomial death process, resulting in a heavier-tailed, less symmetric empirical distribution. To illustrate how this idiosyncratic component diversifies as the portfolio grows, Figure 25 reports the distribution of the normalized liability:

$$L_N^{(s,r)}(t) = \frac{L^{(s,r)}(t)}{N}$$

for several values of N. This distribution concentrates around the best-estimate liability as N increases. Idiosyncratic risk vanishes according to the law of large numbers, while the systematic risk persists. Formally, we have $\lim_{N\to\infty} L_N^{(s,r)}(t) = \sum_{g=1}^{n_G} \pi_g a^{(s)}(x_g, t)$ where $\pi_g = N_g/N$. This is consistent with the variance decomposition of Equation (16), where the idiosyncratic variance vanishes as $N \to \infty$ while the systematic variance remains. Thus, the limiting distribution of $L_N(t)$ retains only the mortality-scenario uncertainty captured by $\bar{L}^{(s)}(t)$.

Figure 26: Convergence of the tail risk measures (France, 2023, LC WLS)



Remark 14 *Figure 26 shows the evolution of* $Q_{99\%}\left(L_N^{(s,r)}(t)\right)$ *and* $\mathrm{CTE}_{99\%}\left(L_N^{(s,r)}(t)\right)$ *with respect to* N*, and illustrates that these tail risk measures converge to* $Q_{99\%}\left(\bar{L}_N^{(s)}(t)\right)$ *and* $\mathrm{CTE}_{99\%}\left(\bar{L}_N^{(s)}(t)\right)$ *as* N *increases where* $\bar{L}_N^{(s)}(t) = \bar{L}^{(s)}(t)/N$*. This convergence is a direct consequence of the vanishing idiosyncratic risk established above. Thus, the quantile and conditional tail expectation of the mixture distribution over* (s,r) *coincide with those of the distribution over* s *alone, which is driven purely by the systematic mortality scenario. We conclude that a portfolio size of* $N > 1\,000$ *is already sufficient to eliminate most of the idiosyncratic risk. Beyond this threshold, the tail risk of the portfolio is essentially determined by systematic mortality risk.*

Model risk To isolate sensitivity to the model specification, the portfolio size is fixed at $N = 10\,000$ and the portfolio composition and all other valuation and simulation settings are held unchanged. Results are reported in Table 14. For females, the best estimate liability ranges from 117 046 under the CBD Poisson WLS model to 119 288 under the LC Poisson model, while for males it ranges from 104 489 under the CBD OLS model to 107 117 under the LC WLS model. The tail risk measures follow the same broad pattern, although their ranking is not entirely determined by the best estimate liability. In particular, for males, the LC WLS model produces the highest best estimate among the LC specifications, while the LC Poisson model yields a slightly higher $Q_{99\%}\left(L\left(t\right)\right)$ and $\mathrm{CTE}_{99\%}\left(L\left(t\right)\right)$. This reversal reflects the greater relative dispersion of the LC Poisson liability distribution. Its coefficient of variation CV reaches 1.82%, compared to 1.65% for the LC WLS model. Finally, systematic mortality risk accounts for over 90% of total variance under every model. Therefore, although the choice of mortality model affects both the liability level and its tail risk, the conclusion that a large annuity portfolio is primarily exposed to systematic mortality risk remains consistent across different specifications.

Table 14: Comparison of the portfolio liability risk between mortality models (France, 2023, $N = 10\,000$, calibration period = 1950–2023)

	Model	Method	$\bar{L}\left(t\right)$	CV	$Q_{99\%}\left(L\left(t\right)\right)$	$\mathrm{CTE}_{99\%}\left(L\left(t\right)\right)$	$w_{\text{systematic}}$
Female	LC	WLS	118 192.64	1.04%	121 006.45	121 398.91	90.38%
	LC	Poisson	119 287.79	1.22%	122 549.37	122 989.86	93.20%
	LC	NB	118 156.18	1.08%	121 056.96	121 453.16	91.05%
	CBD	OLS	117 844.98	1.11%	120 831.87	121 251.65	90.18%
	CBD	Poisson WLS	117 046.44	1.10%	119 989.42	120 387.60	90.64%
	CBD	NB WLS	118 160.06	1.10%	121 139.69	121 557.82	90.13%
Male	LC	WLS	107 116.85	1.65%	111 138.85	111 699.79	93.98%
	LC	Poisson	106 745.31	1.82%	111 144.40	111 748.80	95.00%
	LC	NB	106 921.36	1.69%	111 052.76	111 615.35	94.21%
	CBD	OLS	104 488.71	1.86%	108 944.03	109 589.29	94.69%
	CBD	Poisson WLS	104 994.89	1.84%	109 443.29	110 054.20	94.83%
	CBD	NB WLS	104 860.54	1.86%	109 336.15	109 952.55	94.74%

Table 15: Maximum difference of the portfolio liability risk between mortality models (France, 2023, $N = 10\,000$)

	$\bar{L}\left(t\right)$	$Q_{99\%}\left(L\left(t\right)\right)$	$\mathrm{CTE}_{99\%}\left(L\left(t\right)\right)$
Female	1.91%	2.13%	2.16%
Male	2.52%	2.02%	1.97%

Table 15 reports the maximum relative gap across the six specifications for each risk measure. This gap is relatively modest: around 2% and peaking at 2.52% for $\bar{L}\left(t\right)$ among males, between the LC WLS and CBD OLS models. It is tempting to conclude from this that model risk is low. However, this conclusion would be premature. All six specifications are calibrated on the same historical data, over the same estimation period, and projected using the same forecasting horizon. Out of sample, differences of more than 20% can arise between models, particularly when the forecasting period includes major mortality shocks, such as World War II or the COVID-19 pandemic.

4 Optimal decumulation strategy

This section examines what may constitute an optimal decumulation strategy during retirement. We focus on the two dominant approaches observed in retirement markets: drawdown strategies and life annuities. Alternative arrangements, such as tontines and other pooled-risk products, are not considered because they remain relatively uncommon in practice and play a limited role in most retirement systems. The central question is how retirees should balance the flexibility and growth potential offered by drawdown strategies with the income security and longevity protection provided by annuities. Therefore, determining the appropriate transition between these two approaches is at the heart of the decumulation problem. We begin with a static analysis of the annuitization decision. Our goal is to identify the circumstances under which a drawdown strategy is preferable to an annuity strategy, and vice versa. Two commonly used heuristic approaches are considered. The first approach is based on mortality credits and compares the returns generated by annuities and drawdown portfolios. The second approach adopts an actuarial perspective and compares the lifetime outcomes produced by the two strategies. Although intuitive and easy to implement, both approaches share an important limitation. The annuitization age is determined *ex ante* and remains fixed throughout retirement. Consequently, the decision does not respond to realized investment performance or changes in wealth. For example, a retiree cannot accelerate or decelerate annuitization after a prolonged market downturn, nor can they postpone or anticipate it following a period of strong returns. This rigidity motivates the dynamic analysis developed in the second part of the section. Rather than prescribing a fixed conversion date at retirement, this approach allows the annuitization decision to adapt to the evolution of the retiree's financial situation. We consider two dynamic formulations. The first formulation determines the optimal annuitization date within a stochastic investment environment. The second formulation extends the analysis to a dynamic asset-allocation framework in which the retiree simultaneously chooses portfolio allocations and annuity purchases over time. Together, these approaches provide a more realistic characterization of the flex-to-fix strategy.

4.1 Paradigm shift in decumulation and the flex-to-fix strategy

The financial needs of retirees vary greatly depending on their health, wealth, and family circumstances. Consequently, the importance they place on competing decumulation objectives also varies. These objectives fall into four categories: income certainty, growth and inflation protection, flexibility and bequests[49], and longevity protection. Since these objectives often conflict with one another, products that excel in one area typically underperform in the others. Therefore, most retirement solutions focus on one or two dimensions while falling short in the rest. For instance, Table 16 compares a drawdown strategy with a single-premium immediate annuity. Along these four dimensions, the two are close to mirror images of each other. A drawdown strategy is effective for growth, inflation protection, flexibility, and bequests. An annuity purchase, on the other hand, effectively mitigates longevity risk and guarantees income certainty.

These objectives vary among retirees and change in importance over the course of retirement (Figure 27). In the early years, it is important that capital is both productive and accessible. The investment horizon is long enough to allow for meaningful portfolio growth. Spending needs are often discretionary and variable. Over time, inflation can substantially

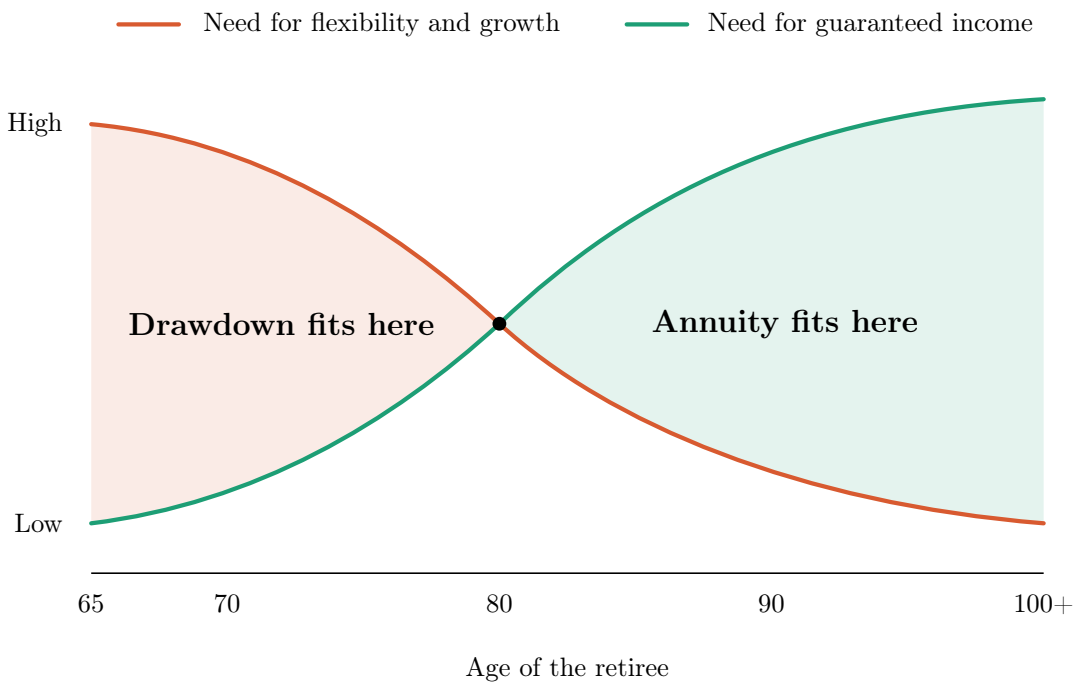
[49] We combine the concepts of bequests and flexibility into a single dimension, since bequests effectively constitute a form of liquidity. Any unspent wealth remains accessible to the retiree, who can choose whether to spend it or pass it on to their heirs.

Table 16: Comparison between drawdown strategy and single-premium immediate annuity

	Drawdown ("*flex*")	Annuity ("*fix*")
Income certainty	**Not guaranteed** The retiree bears market risk	**Guaranteed for life** The insurer bears market risk
Growth and Inflation protection	**Full upside retained** But never guaranteed	**None**
Longevity risk	**Retained by the retiree**	**Transferred to the insurer**
Flexibility	**High**	**Very low**
Bequest	**Full residual** But none if savings are exhausted	**None** Unless a guarantee period is purchased

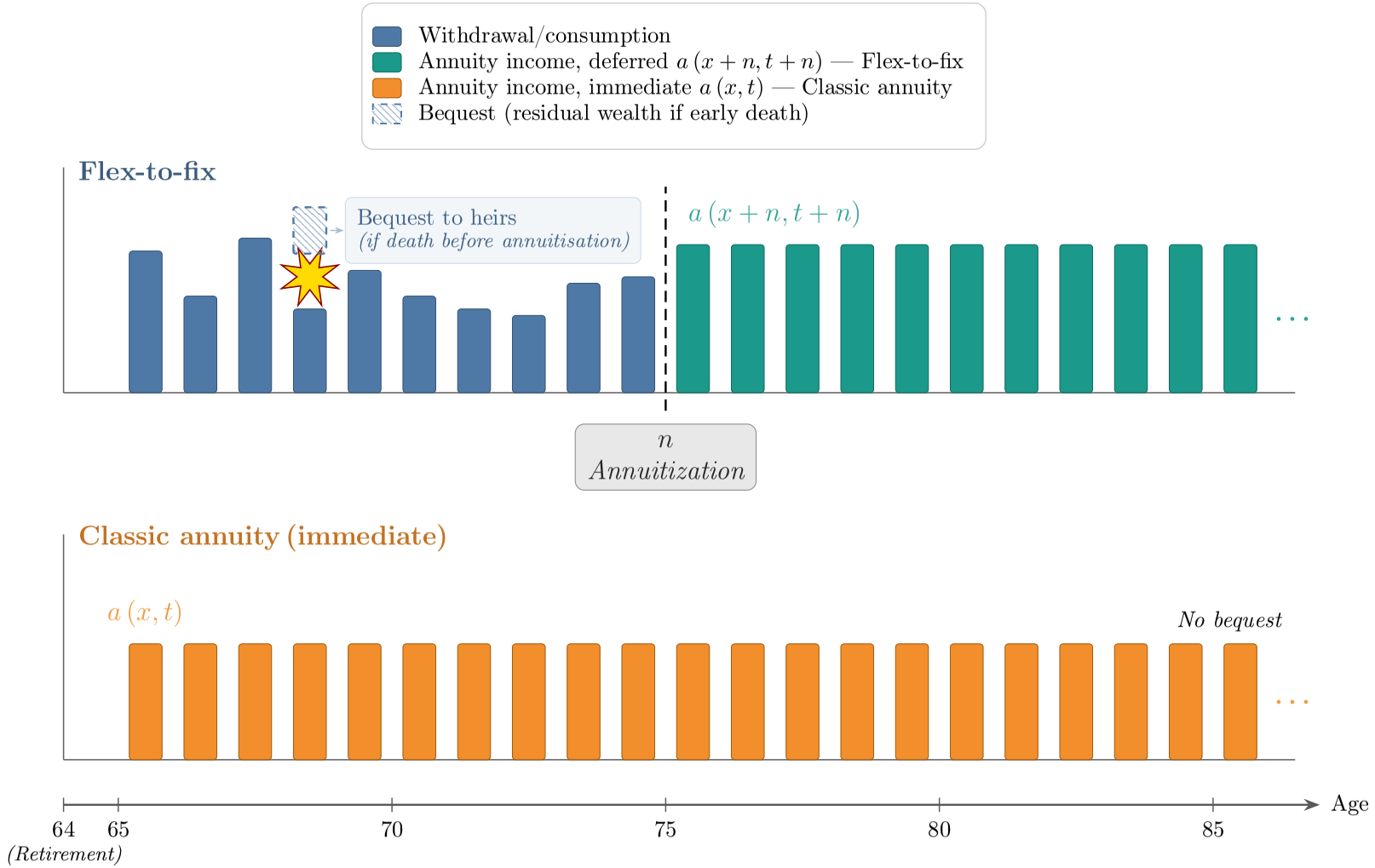
erode the purchasing power of a fixed income stream. However, these priorities gradually change as retirement progresses. A shorter time horizon reduces the benefits of taking investment risks. Essential living expenses become the primary financial obligation, making income certainty more valuable than the prospect of higher returns. Therefore, a retirement solution selected at the outset is unlikely to remain equally well-suited throughout all stages of retirement. As financial needs evolve from growth and flexibility toward security and certainty, the decumulation strategy must evolve accordingly. This observation forms the foundation of the *flex-to-fix* strategy.

Figure 27: Illustration of the evolving financial needs throughout retirement



In a flex-to-fix strategy, retirement is divided into two distinct phases. During the initial "*flex*" phase, which typically lasts the first 10 to 15 years of retirement, wealth is invested in financial markets and withdrawals are governed by a flexible spending rule. This preserves liquidity during the early retirement years, when spending needs are often uncertain and discretionary. It also maintains bequest potential, since any remaining assets are transferred

Figure 28: Comparison of annual income between a flex-to-fix strategy and an annuity



to heirs in the event of premature death. By postponing annuitization, retirees can enter retirement without making an immediate and irreversible commitment of capital. The subsequent "*fix*" phase begins at a predetermined age, such as 75 or 80. At that point, part or all of the remaining wealth is converted into a life annuity, thereby transferring longevity risk to an insurer. The annuity provides a guaranteed lifetime income that can cover essential expenditures regardless of realized lifespan or future market performance. Figure 28 compares the annual income profile generated by a flex-to-fix strategy with that of immediate annuitization at retirement. By combining these two phases, the flex-to-fix strategy seeks to align the retirement solution with the evolving needs of retirees. It preserves market exposure, liquidity, and bequest flexibility in the early years of retirement while providing income security at advanced ages. This approach has attracted increasing policy interest. For instance, Boyle and Webb (2022) and Clare *et al.* (2024) propose a "*Flex First, Fix Later*" product as a potential default decumulation pathway for U.K. defined contribution savers who do not receive financial advice. The flex-to-fix strategy also addresses several commonly cited barriers to annuitization. In particular, it preserves early retirement liquidity and bequest opportunities while avoiding the need for an all-or-nothing annuity purchase at retirement. However, it does not eliminate the fundamental frictions that limit annuity demand. Pricing loads and adverse selection remain present and could become more pronounced when annuitization is deferred because individuals who purchase annuities at older ages tend to be healthier than average retirees (Finkelstein and Poterba, 2004). Moreover, as with any voluntary annuity arrangement, demand remains constrained by the level of income already provided by public pension systems.

4.2 Static analysis

Before formulating the annuitization decision as a formal optimization problem, it is helpful to consider insights from simpler and more intuitive approaches. This section presents two commonly used methodologies for determining the optimal time for a retiree to transition from a drawdown strategy to a life annuity. Both aim to identify the optimal conversion age. However, they differ in their criteria and the evidence on which they rely. The first methodology is based on mortality credits and compares the returns available under the two strategies. At a given age, an annuity provides a return equivalent to the risk-free rate plus a mortality credit. In contrast, a drawdown portfolio offers the risk-free rate plus an uncertain risk premium. Since mortality credits increase monotonically with age and the expected return on a drawdown portfolio remains constant, the two return profiles eventually intersect. The age at which this occurs is considered the point at which annuitization becomes preferable. This approach is parsimonious, requiring only a mortality table and a few financial market assumptions. The second methodology takes a more actuarial approach, comparing lifetime outcomes rather than instantaneous returns. Instead of focusing on the return available at a specific age, this approach evaluates the conditions under which a drawdown strategy yields better retirement outcomes than an annuity strategy, and vice versa. Thus, the resulting annuitization age is derived from a comparison of realized lifetime income and wealth trajectories rather than expected rates of return.

4.2.1 The mortality credit approach

In annuity pools, survivors inherit the premiums of those who die early. This subsidy is commonly referred to as the *mortality credit*. Mathematically, we can define it by considering a discrete-time version of the tontine model[50], whose expected mortality credit is identical

[50] A tontine is a risk-pooling arrangement in which members contribute to a common fund, and the shares of those who die are redistributed among the survivors (Denuit *et al.*, 2022).

to the one embedded in an annuity's payout. Suppose that, at age x, an annuity pool is formed by N_x individuals of the same age. Each individual contributes one unit of currency, so the total principal in the pool is $V_x = N_x$. Assume that these pooled funds are invested exclusively in a risk-free asset that yields an annual rate r. According to the mortality table, the probability that a member of this cohort will die between ages x and $x+1$ is $q(x,t)$. Therefore, the survival probability is $p(x,t,1) = 1 - q(x,t)$. At the end of the year, the accumulated value of the fund is:

$$V_{x+1} = N_x (1+r)$$

while the expected number of surviving members is:

$$\mathbb{E}[N_{x+1}] = N_x p(x,t,1) = N_x \big(1 - q(x,t)\big)$$

The payout distributed to each survivor is therefore[51]:

$$\text{Payout} = \frac{N_x (1+r)}{N_x \big(1 - q(x,t)\big)} = \frac{1+r}{1-q(x,t)}$$

Factoring this expression isolates the market return from the pooling benefit:

$$\frac{1+r}{1-q(x,t)} = (1+r)\left(1 + \frac{q(x,t)}{1-q(x,t)}\right) = 1 + r + \underbrace{\frac{q(x,t)(1+r)}{1-q(x,t)}}_{\text{pooling benefit}}$$

Aside from the standard investment growth $(1+r)$, survivors receive a distinct premium simply for outliving their peers. The pure mortality credit $\text{mc}(x,t)$ per unit of capital is equal to:

$$\text{mc}(x,t) = \frac{q(x,t)(1+r)}{1-q(x,t)}$$

Since human mortality roughly follows Gompertz's law, $q(x,t)$ grows exponentially rather than linearly with age. As a result, mortality credits remain relatively marginal during early retirement but rise sharply at advanced ages.

Remark 15 *As discussed on page 53, the mortality credit is related to the mortality hazard rate. In continuous time, the hazard rate $\lambda(x,t)$ is the ratio of the density function $f(x,t)$ of the mortality time to the survival function $\mathbf{S}(x,t)$. In discrete time, the expression of the hazard rate is:*

$$\lambda(x,t) = \frac{\mathbf{S}(x,t) - \mathbf{S}(x,t+1)}{\mathbf{S}(x,t)} = \frac{q(x,t)}{1-q(x,t)}$$

We deduce that the mortality credit is the product of the hazard rate $\lambda(x,t)$ and the investment growth factor $(1+r)$:

$$\text{mc}(x,t) = \lambda(x,t)(1+r) \approx \lambda(x,t)$$

Figure 29 illustrates this dynamic using French mortality data. The exponential acceleration of $\text{mc}(x,t)$ is evident in both panels. For males in 2023, the mortality credit is only 1.32% at age 65, a subsidy sufficiently modest that it can be exceeded by a well-diversified portfolio. However, by age 75, it doubles to 2.82%, and by age 80, it reaches 4.87%. At this level, the pooling benefit becomes economically significant and increasingly difficult for a conventional investment strategy to replicate. The female panel tells a qualitatively identical

[51] We assume that $N_x \to \infty$.

Figure 29: Mortality credit in % (France, 2023, LC WLS, $r = 0$)

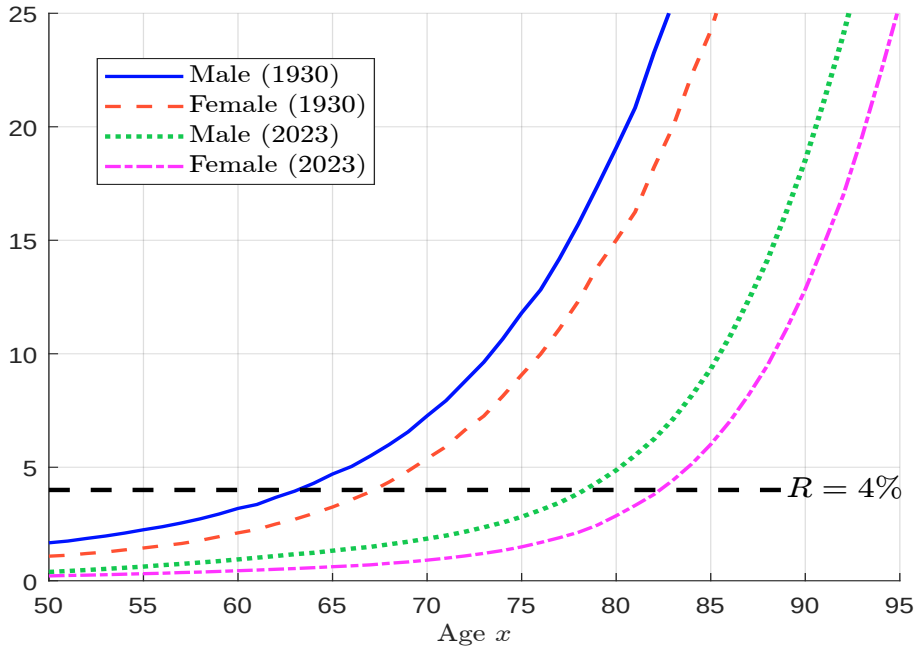
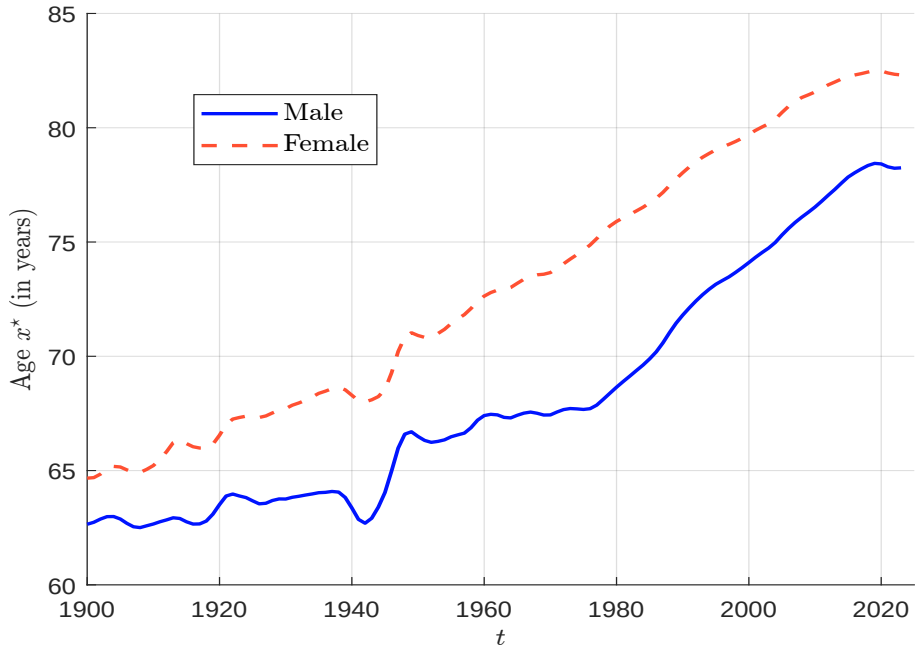


Figure 30: Evolution of the conversion age $x^\star$ in years (France, LC WLS, $r = 0$, $R = 4\%$)



story, yet it shows systematically lower levels throughout: 0.62% at age 65, 1.49% at age 75, and 2.86% at age 80. This gender gap reflects the well-documented longevity advantage of French women. Because female mortality rates $q(x,t)$ are lower, the mortality credit is compressed. This implies that women typically must survive to older ages before the pooling benefit of an annuity becomes compelling. Figure 29 also reports the mortality credits in 1930, one century ago. At that time, mortality credits were substantially higher than they are today: 4.70% at age 65, 11.80% at age 75, and 19.07% at age 80 for males, and 3.25% at age 65, 9.07% at age 75, and 15.01% at age 80 for females. More broadly, mortality credits decreased steadily throughout much of the 20th century (see Tables 37 and 38 on page 163), reflecting the changing shape of the mortality curve. Recently, however, this downward trend appears to be weaker, suggesting that mortality credits may have reached a plateau.

These empirical magnitudes have direct implications for the optimal timing of annuitization. During the early stages of retirement, when $\mathrm{mc}(x,t)$ is small for both sexes, the opportunity cost of locking wealth into an irreversible annuity contract is substantial. At these ages, the expected return generated by a flexible investment portfolio likely exceeds the mortality credit received through risk pooling. By contrast, at advanced ages, mortality credits rise sharply to levels that no risk-free financial instrument can replicate. As a result, the annuitization becomes more attractive with age. This asymmetry provides the economic rationale for the flex-to-fix strategy. With this approach, retirees initially rely on a flexible withdrawal portfolio and postpone annuitization until the mortality credits are large enough to compensate for the loss of liquidity, flexibility, and bequest potential associated with an irreversible annuity purchase. Within the above framework, it is possible to define an optimal conversion age $x^\star$ as the earliest age at which the mortality credit equals or exceeds the expected return R available from the decumulation portfolio. Formally, we have:

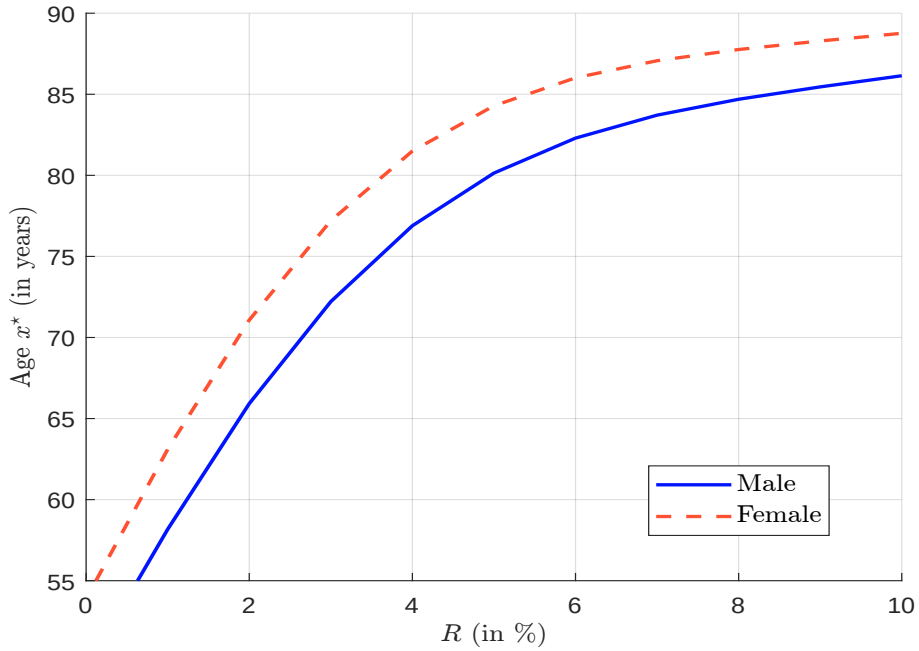
$$x^\star = \inf\left\{x : r + \mathrm{mc}(x,t) \geq R\right\}$$

The intuition is straightforward. As long as the expected portfolio return exceeds the mortality credit, it is better to remain invested in financial markets than to immediately annuitize. However, once the mortality credit surpasses the attainable portfolio return, the pooling advantage becomes sufficiently valuable to justify conversion into an annuity. Figure 30 reports the evolution of $x^\star$ between 1900 and 2023 under the assumptions $r = 0\%$ and $R = 4\%$. Consistent with the historical patterns documented above, the optimal annuitization age has increased substantially over the past century. Improvements in longevity have reduced mortality credits at any given age, thereby delaying the point at which annuitization becomes economically advantageous. Recently, however, the upward trend in $x^\star$ appears to have stabilized, mirroring the plateau observed in mortality credit dynamics. Nevertheless, this calculation should be interpreted with caution. The threshold approach is highly sensitive to the assumed portfolio return R. Small changes in expected investment performance can generate large shifts in the implied conversion age, particularly when mortality credits increase steeply with age. Figure 31 illustrates this sensitivity by reporting the values of $x^\star$ obtained under alternative assumptions for R.

4.2.2 The actuarial approach

Fixed-date annuitization problem We consider an individual aged x at retirement date t_0 who chooses a fixed annuitization date $t_a = t_0 + n$. Although the conversion date is chosen at t_0, the annuity is purchased at the prevailing price at $t_0 + n$. The special case $n = 0$ corresponds to immediate annuitization (fix-now), while $n = \infty$ denotes the strategy of remaining liquid throughout the projection horizon. Let x^+ denote the limiting age and let $H = x^+ - x$ denote the maximum projection horizon. Candidate annuitization dates are

Figure 31: Conversion age in years with respect to the return R (France, 2023, LC WLS, $r = 0$)



indexed by:

$$n \in \{0, 1, \ldots, H-1\} \cup \{\infty\}$$

We use $t \in \{t_0, t_0+1, t_0+2, \ldots\}$ to denote the time index. The timing follows an arrears convention. No withdrawal is made at $t = t_0$, while the investment return is realized before the scheduled withdrawal for $t > t_0$. We consider a simple drawdown strategy. Let W_t be the wealth at time t. We assume that the target withdrawal is proportional to the initial wealth: $w_0 = \alpha W_0$ where $W_0 = W_{t_0}$ is the initial wealth, and α is the (constant) withdrawal rate. The wealth is invested in a drawdown fund, whose return for the period $[t-1, t[$ is R_t. We deduce that the wealth process is $W_{t^-} = (1+R_t)\, W_{t-1}$. Consumption is then $\boldsymbol{c}_t = \min(w_0, W_{t^-})$. It follows that the post-withdrawal or post-consumption wealth is $W_t = W_{t^-} - \boldsymbol{c}_t$. In this model, the bequestable wealth is W_{t^-} if death occurs during period $[t-1, t[$. If the individual survives until time t, the amount available for annuitization is W_t. At the annuitization date $t_a = t_0 + n$, the individual purchases an annuity. Let $a(x+n, t_0+n)$ be the unloaded annuity price for an annual payment of one[52]. To account for administrative expenses, profit margins, and the cost of holding capital that are not included in the actuarial best estimate, we apply a proportional loading factor $\mathcal{L}$. The loaded price of one unit of annual annuity income at t_0+n is therefore $a^*(x+n, t_0+n) = (1+\mathcal{L})a(x+n, t_0+n)$. Given a terminal wealth $W_{t_a} = W_{t_0+n}$, the loaded annuity pays an annual income:

$$Y = \frac{W_{t_0+n}}{(1+\mathcal{L})a(x+n, t_0+n)} := \boldsymbol{c}_t \qquad \text{for } t > t_a$$

In the following, we assume that the annual consumption is equal to this annual income.

[52] This corresponds to $B = 1$ in Equation (8) on page 40.

There are several sources of uncertainty in the flex-to-fix strategy. The first one is the annual return of the investment. For illustrative purposes, we can assume that R_t is constant. However, it is more realistic to treat R_t as stochastic. A common approach is to assume that R_t follows a Gaussian distribution: $R_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$. A second source of uncertainty is the value of the future annuity at time t_0+n. One option is to introduce a stochastic component to simulate the survival probabilities $p\left(x+n, t_0+n, h\right)$ that are used to price the annuity $a\left(x+n, t_0+n\right)$. Alternatively, we can simplify the problem by using the current annuity price $a\left(x+n, t_0\right)$ as a proxy for the future annuity price $a\left(x+n, t_0+n\right)$. A third source of uncertainty is the loading factor $\mathcal{L}$. This is related to the money's worth ratio (MWR), which was introduced by Mitchell *et al.* (1999) and is defined as the ratio of the expected present value of an annuity's payout to its purchase premium[53]:

$$\text{MWR} = \frac{a\left(x,t\right)}{a^*\left(x,t\right)} = \frac{1}{1+\mathcal{L}}$$

Mitchell *et al.* (1999, page 1300) find that "*the prices charged for a single-premium immediate life annuity vary widely.*" For instance, the difference in annuity payout between the top 10 and bottom 10 insurance companies is roughly 20%. Using UK data, Cannon and Tonks (2004, 2008) calculate money's worth ratios and report loading figures of around 10%. However, most of these results apply to individuals aged 65 who are purchasing an immediate annuity. The evidence suggests that loadings are substantially higher for deferred annuities:

> "*This paper estimates the expected present discounted value (EPDV) of the payments from immediate and deferred annuities that were available in the US retail insurance market in May 2024. The central estimates suggest that for immediate annuities purchased at age 65, the value of payouts per premium dollar is about 87 cents for a buyer whose prospective mortality aligns with that of the US population, and about one dollar for one whose mortality aligns with that of a typical annuity purchaser. For a deferred annuity purchased at age 65 that begins payouts at age 75, the EPDVs are 73 cents and close to one dollar, respectively. The private information that contributes to the lower mortality rates of annuitants than those of the population appears to decay over time, and age-specific mortality rates of recent annuitants are somewhat lower than those of annuitants who bought policies further in the past.*" (Poterba and Solomon, 2025, page 1).

Illustration of the flex-to-fix strategy Table 17 shows the consumption calculations for a 65-year-old French male in 2023 under four different parameter combinations. We assume that $W_0 = 100$, $\alpha = 4\%$ and $t_a = 75$ years. In the fix-now strategy, consumption only depends on the loading factor. When $\mathcal{L} = 0\%$, $\boldsymbol{c}_t$ is equal to 7.40. Increasing the loading factor by 50% reduces the consumption by 33%. For the flex-to-fix strategy, pre-annuitization consumption is constant by construction and does not depend on the investment return R_t. On the contrary, investment returns have a significant impact on post-annuitization consumption. If we assume that $\min\left(w_0, W_{t^-}\right) = w_0$, we obtain the following recurrence for the wealth: $W_t = \left(1+R_t\right) W_{t-1} - \alpha W_0$. We deduce that:

$$W_{t_a} = \left(\prod_{t=t_0+1}^{t_a} \left(1+R_t\right) - \alpha \sum_{t=t_0+1}^{t_a} \prod_{s=t+1}^{t_a} \left(1+R_s\right) \right) W_0$$

[53] The loading factor is equal to 11%, 25%, 43%, 67% and 100% when the MWR is 90%, 80%, 70%, 60% and 50%, respectively.

Table 17: Value of the consumption $\boldsymbol{c}_t$ in the flex-to-fix strategy (France, male, 2023, LC WLS)

$x+n$	$\mathcal{L}$	α	Age	Fix-now	$R=0\%$	$R=1\%$	$R=2\%$	$R=3\%$	$R=4\%$	$R=5\%$
75	0%	4%	65–75	7.401	4.000	4.000	4.000	4.000	4.000	4.000
			75–95	7.401	6.560	7.501	8.538	9.679	10.933	12.308
85	0%	4%	65–85	7.401	4.000	4.000	4.000	4.000	4.000	4.000
			85–95	7.401	4.286	7.274	11.016	15.672	21.430	28.516
75	50%	4%	65–75	4.934	4.000	4.000	4.000	4.000	4.000	4.000
			75–95	4.934	4.373	5.001	5.692	6.453	7.288	8.205
75	0%	10%	65–75	7.401	10.000	10.000	10.000	10.000	10.000	10.000
			75–95	7.401	0.000	0.638	1.356	2.160	3.057	4.057

Table 18: Value of the consumption $\boldsymbol{c}_t$ in the flex-to-fix strategy (France, female, 2023, LC WLS)

$x+n$	$\mathcal{L}$	α	Age	Fix-now	$R=0\%$	$R=1\%$	$R=2\%$	$R=3\%$	$R=4\%$	$R=5\%$
75	0%	4%	65–75	6.476	4.000	4.000	4.000	4.000	4.000	4.000
			75–95	6.476	5.621	6.428	7.317	8.295	9.369	10.547
85	0%	4%	65–85	6.476	4.000	4.000	4.000	4.000	4.000	4.000
			85–95	6.476	3.632	6.164	9.336	13.281	18.161	24.166
75	50%	4%	65–75	4.317	4.000	4.000	4.000	4.000	4.000	4.000
			75–95	4.317	3.748	4.286	4.878	5.530	6.246	7.031
75	0%	10%	65–75	6.476	10.000	10.000	10.000	10.000	10.000	10.000
			75–95	6.476	0.000	0.547	1.162	1.851	2.620	3.477

If investment returns are constant, the wealth available at annuitization is given by:

$$W_{t_a}(\alpha, R) = \left((1+R)^{t_a - t_0} - \alpha \sum_{t=t_0+1}^{t_a} (1+R)^{t_a - t} \right) W_0$$

Let $n = t_a - t_0$. It follows that:

$$\begin{aligned} W_{t_a}(\alpha, R) &= \left((1+R)^n - \alpha \frac{(1+R)^n - 1}{R} \right) W_0 \\ &= \left(\frac{\alpha + (R-\alpha)(1+R)^n}{R} \right) W_0 \end{aligned}$$

We can show that $\partial_R W_{t_a}(\alpha, R) > 0$, $\lim_{R \to 0} W_{t_a}(\alpha, R) = (1 - \alpha n) W_0$, and $W_{t_a}(\alpha, \alpha) = W_{t_a}(R, R) = W_0$. Finally, we obtain:

$$\frac{\boldsymbol{c}_t}{W_0} = \begin{cases} \alpha & \text{if } t_0 < t \leq t_a \\ \dfrac{\alpha + (R-\alpha)(1+R)^n}{R(1+\mathcal{L})\, a(x+n, t_0+n)} & \text{if } t > t_a \end{cases}$$

This formula highlights the importance of the relative position of R with respect to α. In particular, if α is too high relative to R, post-annuitization consumption can fall to zero, because the wealth W_{t_a} is exhausted before annuitization. This is exactly what happens in the last case in the table ($t_a = 75$, $\mathcal{L} = 0\%$, $\alpha = 10\% \gg R = 0\%$).

While the effects of the parameters $\mathcal{L}$, α, and R_t are relatively easy to understand, the impact of the annuitization date t_a is more difficult. Indeed, we have:

$$\frac{\partial \boldsymbol{c}_t}{\partial n} = \left(\underbrace{\frac{(R-\alpha)(1+R)^n \ln(1+R)}{\alpha + (R-\alpha)(1+R)^n}}_{\text{Wealth accumulation effect}} + \underbrace{-\frac{\partial_n a(x+n, t_0+n)}{a(x+n, t_0+n)}}_{\text{Annuity price effect}} \right) \boldsymbol{c}_t$$

The post-annuitization consumption is then determined by two competing forces:

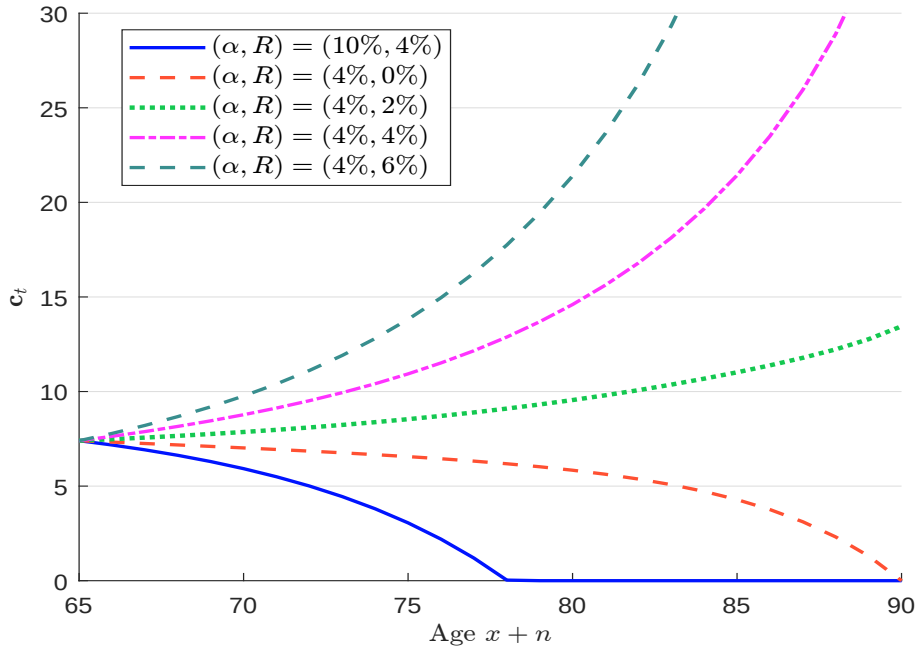
- The wealth accumulation effect
  The first term reflects the impact of delaying annuitization on the wealth available at retirement. If $R > \alpha$, wealth grows faster than it is withdrawn, implying that this term is positive. Postponing annuitization in this case increases the amount of wealth available for conversion into an annuity, which tends to raise post-annuitization consumption. Conversely, when $R < \alpha$, withdrawals dominate investment growth, reducing the wealth available at annuitization and weakening the case for delay.

- The annuity price effect
  The second term captures changes in the annuity factor. Since annuity prices generally decrease with age because fewer future payments are expected, we typically have $\partial_n \ln a(x+n, t_0+n) < 0$. As a result, delaying annuitization allows the individual to purchase a larger annuity income per unit of wealth, which increases post-annuitization consumption[54].

[54] Mortality improvements over time may partially offset this effect. If longevity is expected to increase, annuities purchased at later dates may become more expensive because insurers anticipate longer payment periods.

The overall impact of delaying annuitization depends on the balance between these two opposing forces. When investment returns are high enough relative to the withdrawal rate, both effects support postponement. Higher returns increase the wealth available at annuitization, and the decline in annuity prices with age allows for the purchase of a higher income per unit of wealth. In this case, delaying annuitization leads to higher post-annuitization consumption. However, when withdrawals substantially exceed investment returns, the erosion of wealth can outweigh the benefit of lower annuity prices. In this scenario, postponing annuitization can reduce post-annuitization consumption, making earlier annuitization more appealing. This trade-off is clearly illustrated in Table 17. When investment returns are low ($R = 0\%$ or $R = 1\%$), delaying annuitization from age 75 to 85 decreases post-annuitization consumption. For example, when $R = 0\%$, post-annuitization consumption decreases from 6.560 at $t_a = 75$ to 4.286 at $t_a = 85$. In contrast, when investment returns are high ($R = 5\%$), delaying annuitization substantially increases post-annuitization consumption. $\boldsymbol{c}_t$ rises from 12.308 at $t_a = 75$ to 28.516 at $t_a = 85$. Figure 32 illustrates the evolution of post-annuitization consumption with respect to age $x + n$ for different set of parameters (α, R), while Figure 33 shows the components of the post-annuitization consumption[55].

Figure 32: Post-annuitization consumption with respect to age $x + n$ (France, male, 2023, LC WLS, $W_0 = 100$, $\mathcal{L} = 0\%$)



Remark 16 *In practice, investment returns R_t are stochastic rather than constant. For instance, we generally assume that $R_t \sim \mathcal{N}(\mu, \sigma^2)$. In this case, the wealth W_t becomes stochastic, implying that the post-annuitization consumption $\boldsymbol{c}_t$ for $t > t_a$ is stochastic as well. For example, if $R_t = \alpha = 4\%$, we previously found that $\boldsymbol{c}_t = 10.933$ and $\boldsymbol{c}_t = 9.369$ for male and female individuals, respectively. Let us now assume that $\mu = 4\%$ and $\sigma = 10\%$. We obtain the probability density function shown in Figure 34.*

[55] To isolate the pure annuity pricing effect, we report the additional consumption generated by delayed annuitization relative to the fix-now strategy. This adjustment removes therefore the portion of consumption that would be obtained under immediate annuitization.

Figure 33: Components of post-annuitization consumption (France, male, 2023, LC WLS, $W_0 = 100$, $\alpha = R = 4\%$, $\mathcal{L} = 0\%$)

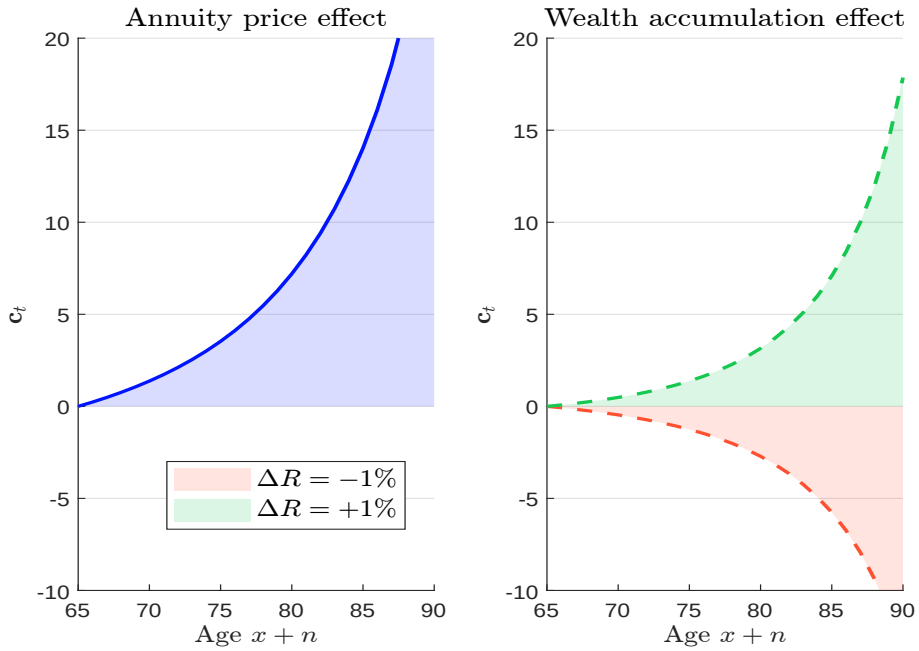
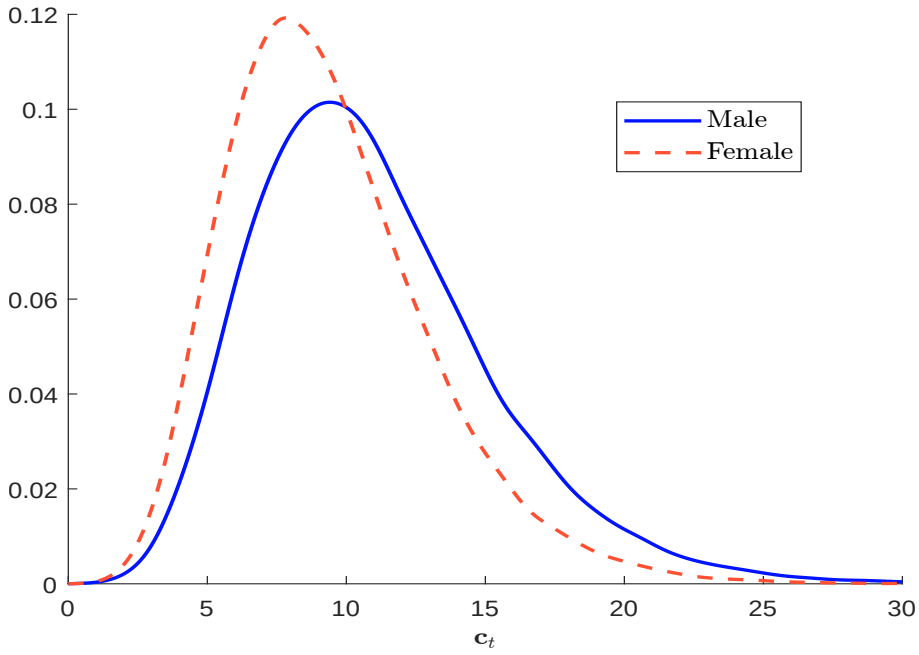


Figure 34: Probability density function of post-annuitization consumption (France, male, 2023, LC WLS, age 75, $W_0 = 100$, $\alpha = 4\%$, $\mu = 4\%$, $\sigma = 10\%$, $\mathcal{L} = 0\%$)



The bequest motivation An advantage of the flex-to-fix strategy is the possibility of generating a bequest. Recall that if the individual dies before the annuitization date, their heirs receive the accumulated wealth. In contrast, if death occurs after annuitization, no wealth is transferred to the heirs. As discussed in the introduction, many individuals choose not to annuitize because they wish to leave part of their wealth to their heirs. We have seen that the bequestable wealth is equal to W_{t^-} if death occurs during the period $[t-1, t[$. Moreover, the probability that an individual dies between t and $t+1$, given that they were alive at the retirement date t_0, is equal to:

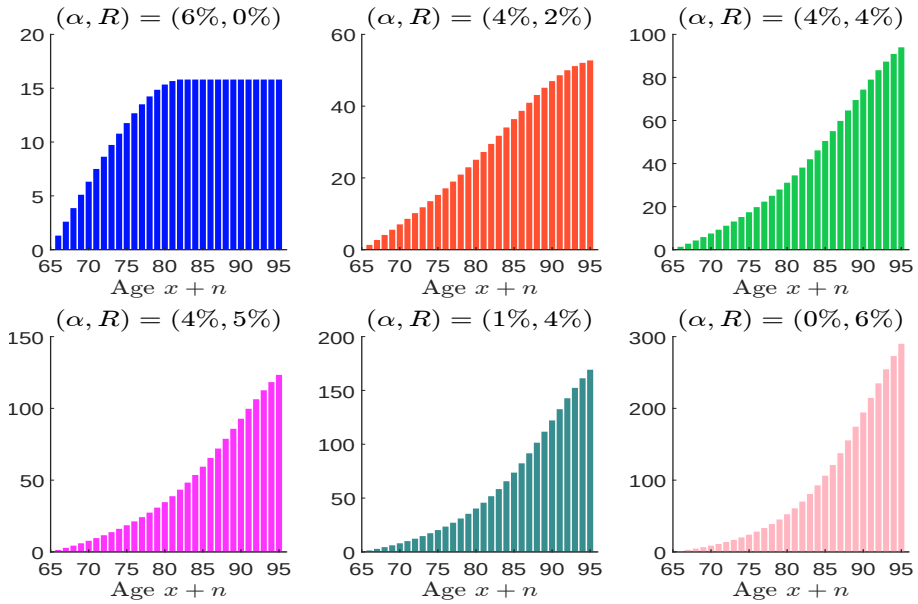
$$\begin{aligned}
\mathrm{PD}_t &= \Pr\{\text{individual is alive at time } t\} \times \Pr\{\text{individual dies between } t \text{ and } t+1\} \\
&= p(x, t_0, t-t_0)\, q(x+t-t_0, t) \\
&= p(x, t_0, h)\, q(x+h, t_0+h)
\end{aligned}$$

where $h = t - t_0$. If we assume a static mortality table — $q(x,t) = q(x,t_0)$, we obtain[56] $\mathrm{PD}_t = p(x, t_0, h)\, q(x+h, t_0)$. Finally, we deduce that the expected bequest ratio of an individual aged x at retirement date t_0 who chooses to annuitize at date t_a is equal to:

$$\mathbb{E}\left[\frac{\mathfrak{B}_{t_a}}{W_0}\right] = \frac{1}{W_0} \sum_{t=t_0+1}^{t_a} \mathrm{PD}_{t-1}\, \mathbb{E}\left[W_{t^-}\right]$$

Figure 35 shows examples of the expected bequest ratio. It is an increasing function of the annuitization age. Moreover, we see that the annuitization age is by far the most important determinant. Although portfolio returns and withdrawal rates have a noticeable impact, their influence remains secondary compared with the timing of annuitization. Finally, a pronounced gender gap emerges (see Figure 94 on page 175). This difference is driven by the mortality probabilities $q(x,t)$.

Figure 35: Expected bequest ratio (in %) by annuitization age (France, male, 2023, LC WLS)



[56] In Appendix C.9 on page 147, we show that PD_t is a proper probability distribution.

In two extreme cases, the optimal annuitization decision is straightforward. If an individual has no bequest motive and wants to eliminate longevity risk as soon as possible, the optimal annuitization age is the retirement age. Conversely, if the individual absolutely prioritizes leaving their wealth to their heirs, then the optimal annuitization age is effectively infinite, meaning that the individual never annuitizes[57]. More realistic situations lie between these two extremes. Many individuals value both the possibility of leaving a bequest and the protection against longevity risk provided by an annuity. In this case, the individual's preference can be interpreted as follows. They would like their heirs to inherit part of their wealth if they die relatively early (for example, before age 75), but they would also like to secure a lifelong income if they live to older ages, when health and long-term care expenses may become significant. To formalize this trade-off, suppose that the individual specifies a target expected bequest $\mathfrak{B}^\star$. The objective is then to determine the annuitization date t_a that generates an expected bequest equal to this target, or as close as possible:

$$\mathbb{E}\left[\mathfrak{B}_{t_a}\right] \approx \mathfrak{B}^\star$$

Using the expected bequest function derived above, the optimal annuitization date is defined as:

$$t_a^\star = t_0 + n^\star = \begin{cases} t_0 & \text{if } \mathfrak{B}^\star = 0 \\ \inf\left\{t_a : \mathbb{E}\left[\dfrac{\mathfrak{B}_{t_a}}{W_0}\right] \geq \dfrac{\mathfrak{B}^\star}{W_0}\right\} & \text{if } 0 < \mathfrak{B}^\star \leq \mathbb{E}\left[\mathfrak{B}_{t_0+H}\right] \\ \infty & \text{if } \mathfrak{B}^\star > \mathbb{E}\left[\mathfrak{B}_{t_0+H}\right] \end{cases}$$

The first case corresponds to an individual with no bequest objective, who annuitizes immediately at retirement. The second case identifies the earliest annuitization date that achieves the desired expected bequest. The last case arises when the target bequest exceeds the maximum attainable expected bequest. In that situation, the target cannot be achieved through any finite annuitization date, and the individual optimally chooses never to annuitize.

Figure 95 on page 176 reports the continuous solution[58] for $t_a^\star$ across several values of $(\alpha, R, \mathfrak{B}^\star)$. When the target $\mathfrak{B}^\star$ is low, the optimal annuitization age is close to the retirement age (Panels (a) and (b)). When the target $\mathfrak{B}^\star$ is high, the optimal strategy is to never annuitize, except for large values of R combined with small values of α (Panels (e) and (f)). More generally, we have the following properties:

$$\frac{\partial\, t_a^\star}{\partial\, \alpha} > 0,\ \frac{\partial\, t_a^\star}{\partial\, R} < 0 \text{ and } \frac{\partial\, t_a^\star}{\partial\, \mathfrak{B}^\star} > 0$$

The annuitization time is an increasing function of α: a larger withdrawal rate delays annuitization in order to accumulate more wealth (Figure 37). It is a decreasing function of R, since a higher investment return allows the target to be reached more quickly (Figure 97 on page 178). Finally, $t_a^\star$ increases with the target $\mathfrak{B}^\star$: a more ambitious objective requires delaying annuitization (Figure 98 on page 178).

[57] In some cases, the initial wealth W_0 is intended solely for inheritance purposes, and the individual relies on other retirement income sources to finance consumption during retirement. The retirement vehicle is then used primarily to take advantage of the favorable tax treatment associated with long-term savings and wealth transmission. In this case, the decumulation phase becomes an accumulation phase since no withdrawals are made from the portfolio ($\alpha = 0$). The objective is no longer to finance retirement spending but rather to maximize the value of the wealth transferred to the heirs. As a result, the individual may rationally choose a higher allocation of risky assets to increase the expected terminal wealth and, consequently, the expected bequest. This situation corresponds to the sixth panel of Figure 35, where $\alpha = 0$ and the expected investment return R is equal to 6%.

[58] For each pair (α, R), we compute the cubic spline interpolation $\mathbb{E}\left[\mathfrak{B}_t\right] = f_{\text{cs}}\left(t; \alpha, R\right)$. Given a target $\mathfrak{B}^\star$, we then obtain the optimal value $t_a^\star = f_{\text{cs}}^{-1}\left(\mathfrak{B}^\star; \alpha, R\right) := g\left(\mathfrak{B}^\star, \alpha, R\right)$. Unlike the discrete solution, which exhibits jumps in $t_a^\star$, this approach yields a continuous function of $t_a^\star$.

Figure 36: Optimal annuitization date when $\mathfrak{B}^\star = 50\%$ (France, male, 2023, LC WLS)

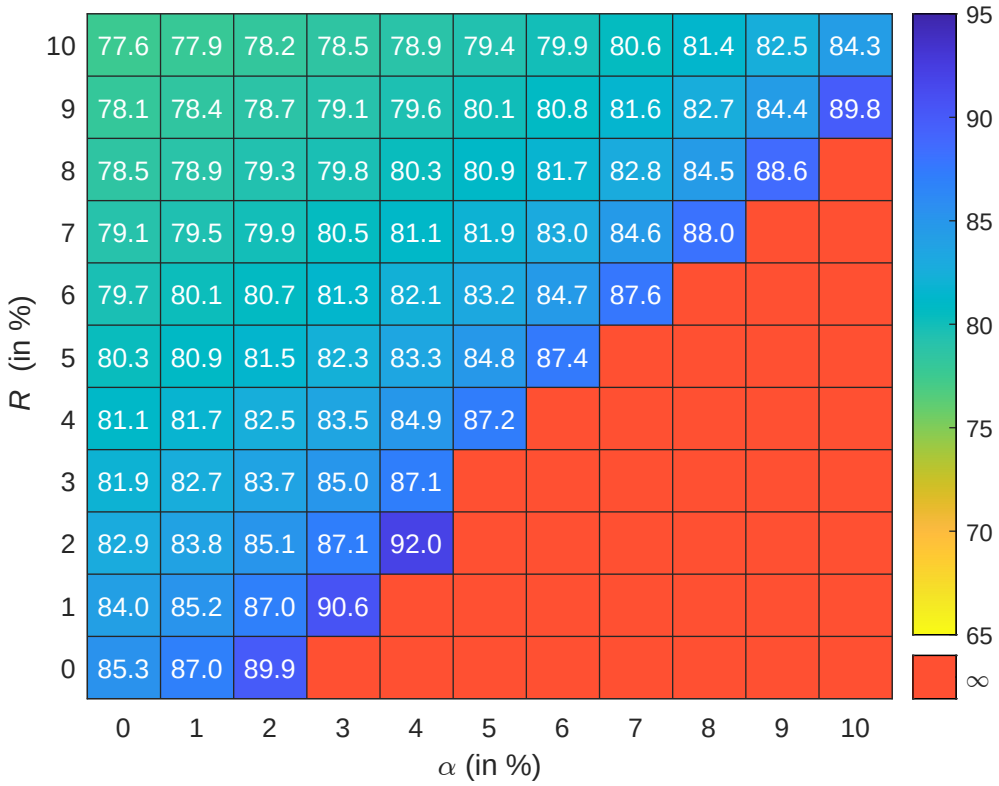
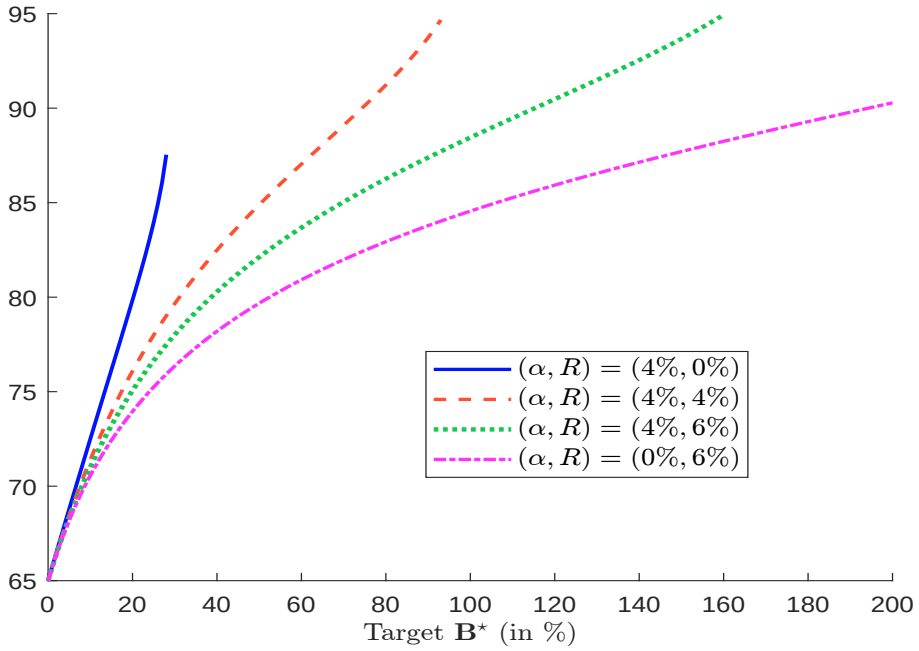


Figure 37: Optimal annuitization date with respect to $\mathfrak{B}^\star$ (France, male, 2023, LC WLS)



The objective function We consider a classical consumption-bequest optimization problem (Samuelson, 1969; Merton, 1969, 1971). In this framework, an individual has two competing objectives: maximizing lifetime consumption and maximizing the bequest left to their heirs. Let $\mathcal{U}(n) = \mathcal{C}(n) + \psi\mathcal{B}(n)$ denote the objective (risk-neutral utility) function, where $\mathcal{C}(n)$ is the stochastic lifetime consumption component, $\mathcal{B}(n)$ is the stochastic bequest component, and $\psi \geq 0$ is a preference parameter governing the trade-off between consumption and bequest. The optimization problem of the individual is:

$$\begin{aligned} n^{\star} &= \underset{n\in\{0,\ldots,H-1\}\cup\{\infty\}}{\arg\max}\ \mathbb{E}\left[\mathcal{U}(n)\right] \\ &= \underset{n\in\mathcal{N}}{\arg\max}\ \mathbb{E}\left[\mathcal{C}(n)\right] + \psi\mathbb{E}\left[\mathcal{B}(n)\right] \end{aligned}$$

where $\mathcal{N} = \{0,\ldots,H-1\}\cup\{\infty\}$. The optimal annuitization date is then obtained by evaluating the objective function at each admissible annuitization age and selecting the value of n that maximizes the expected utility. In general, the quantities $\mathbb{E}\left[\mathcal{C}(n)\right]$ and $\mathbb{E}\left[\mathcal{B}(n)\right]$ cannot be computed analytically. Therefore, we replace them with their Monte Carlo estimates:

$$\begin{cases} \mathbb{E}\left[\mathcal{C}(n)\right] = \dfrac{1}{n_S}\displaystyle\sum_{s=1}^{n_S}\mathcal{C}^{(s)}(n) \\ \mathbb{E}\left[\mathcal{B}(n)\right] = \dfrac{1}{n_S}\displaystyle\sum_{s=1}^{n_S}\mathcal{B}^{(s)}(n) \end{cases}$$

where n_S is the number of Monte Carlo simulations, and $\mathcal{C}^{(s)}(n)$ and $\mathcal{B}^{(s)}(n)$ are the simulated components generated by the s^{th} simulation scenario.

Now, let us define $\mathcal{C}^{(s)}(n)$ and $\mathcal{B}^{(s)}(n)$. The fixed-date problem is evaluated using n_S simulation paths from the initial valuation date t_0. Each path combines a simulated systematic mortality-factor path and a simulated liquid-return path, while individual death times are simulated separately for outcome diagnostics. The simulated mortality rates are defined using the Lee-Carter model:

$$q^{(s)}(x, t_0+h) = 1 - \exp\left(-m^{(s)}(x, t_0+h)\right)$$

for $h = 1,\ldots,H$, where $\ln m^{(s)}(x, t_0+h) = \alpha_x + \beta_x\kappa^{(s)}_{t_0+h}$ and $\kappa^{(s)}_{t_0+h}$ is simulated using the ARIMA(1,1,0) model fitted on the training period ending at time t_0. For the investment returns, we have: $R^{(s)}_{t_0+h} = \mu + \sigma\varepsilon^{(s)}_{t_0+h}$ where $\varepsilon^{(s)}_{t_0+h} \sim \mathcal{N}(0,1)$. The n_S jointly simulated paths $\left\{R^{(s)}_{t_0+h}, \kappa^{(s)}_{t_0+h}, m^{(s)}(x, t_0+h), q^{(s)}(x, t_0+h)\right\}_{h=1}^{H}$ are used to evaluate pre-annuitization wealth dynamics and mortality evolution. For each simulated path s, outcomes up to t_0+n determine the mortality state and the available wealth at the conversion date. The annuity price at t_0+n is then evaluated using only the available information at that date. Given the state $\kappa^{(s)}_{t_0+n}$, we generate $n_{S'}$ future mortality factor continuations. For $h = 1,\ldots,H-n$, these continuations follow the same fitted ARIMA(1,1,0) transition law. At $n = 0$, the conditioning state is the state fitted at t_0. The annuity price is estimated by:

$$a^{(s)}(x+n, t_0+n) = \frac{1}{n_{S'}}\sum_{s'=1}^{n_{S'}} a^{(s,s')}(x+n, t_0+n) = \frac{1}{n_{S'}}\sum_{s'=1}^{n_{S'}}\sum_{h=1}^{H-n}\frac{p^{(s,s')}(x+n, t_0+n, h)}{(1+\varrho_{t_0})^h}$$

where $p^{(s,s')}(x+n, t_0+n, h)$ is the h-year survival probability obtained from conditional mortality path s', given the mortality state reached under path s at t_0+n. We recall that

the annual annuity income purchased at the conversion date is:

$$Y_{t_0+n}^{(s)} = \frac{W_{t_0+n}^{(s)}}{(1+\mathcal{L})a^{(s)}\left(x+n, t_0+n\right)}$$

Let $\tilde{\varrho}_t$ denote the discount rate of the objective function[59]. In Appendix C.10 on page 148, we show that the discounted lifetime consumption component is:

$$\mathcal{C}^{(s)}(n) = \sum_{j=1}^{n} \frac{p^{(s)}\left(x, t_0, j\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \boldsymbol{c}_{t_j}^{(s)} + \frac{p^{(s)}\left(x, t_0, n\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^n} Y_{t_n}^{(s)} \tilde{a}^{(s)}\left(x+n, t_0+n\right) \tag{17}$$

where $t_j = t_0 + j$, $p^{(s)}\left(x, t_0, j\right)$ denotes the survival probability from age x at t_0 to age $x+j$ along mortality path s, and $\tilde{a}^{(s)}\left(x+n, t_0+n\right)$ is the annuity priced with the discount rate $\tilde{\varrho}_{t_0}$:

$$\tilde{a}^{(s)}\left(x+n, t_0+n\right) = \frac{1}{n_{S'}} \sum_{s'=1}^{n_{S'}} \sum_{h=1}^{H-n} \frac{p^{(s,s')}\left(x+n, t_0+n, h\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^h}$$

The first term of Equation (17) values withdrawals from the liquid account before annuitization, weighted by the survival probability to each withdrawal date. The second term values the annuity income purchased at t_0+n, conditional on survival to the conversion date, and discounted back to t_0. The discounted bequest component is:

$$\mathcal{B}^{(s)}(n) = \sum_{j=1}^{n} \frac{p^{(s)}\left(x, t_0, j-1\right) - p^{(s)}\left(x, t_0, j\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^j} W_{t_j^-}^{(s)} \tag{18}$$

where $W_{t_j^-}^{(s)}$ is the bequest at time $t_j = t_0 + j$, since death occurs before the scheduled withdrawal at date t_j. When $n = \infty$, no annuity is purchased, and the lifetime consumption and bequest components become:

$$\mathcal{C}^{(s)}(n) = \sum_{j=1}^{H} \frac{p^{(s)}\left(x, t_0, j\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \boldsymbol{c}_{t_j}^{(s)} \tag{19}$$

and:

$$\mathcal{B}^{(s)}(n) = \sum_{j=1}^{H} \frac{p^{(s)}\left(x, t_0, j-1\right) - p^{(s)}\left(x, t_0, j\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^j} W_{t_j^-}^{(s)} + \frac{p^{(s)}\left(x, t_0, H\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^H} W_{t_H}^{(s)} \tag{20}$$

Alternatively, the expected objective value $\mathbb{E}\left[\mathcal{U}(n)\right]$ can be estimated by simulating individual lifetimes directly, rather than weighting outcomes using the path-specific survival probabilities $p^{(s)}\left(x, t_0, j\right)$. This alternative method is described in Appendix C.11.

Numerical results The empirical analysis considers a French retiree aged 65 at the end of 2023, with initial wealth normalized to $W_0 = 100$ and a limiting age of $x^+ = 95$. Mortality is modeled using the Lee-Carter Poisson specification, with the period factor projected through an ARIMA(1,1,0) model. The mortality model is fitted using observations from 1950 onwards over ages 50–95, and the period-factor forecast uses a 15-year lookback window. The annuity loading is set to $\mathcal{L} = 5\%$. The pricing and objective discount rates are set to $\varrho = 2\%$ and $\tilde{\varrho} = 2\%$, respectively. The baseline withdrawal rate is $\alpha = 4\%$, and the bequest weight varies over $\psi \in \{0, 0.25, 0.5, 0.75, 1\}$. The grid analysis spans $\mu \in \{2\%, 3\%, \ldots, 8\%\}$

[59] The annuity discount rate ϱ and the discount rate entering the retiree's objective $\tilde{\varrho}$ need not coincide.

and $\sigma \in \{0\%, 2\%, \ldots, 10\%, 12\%, 15\%\}$. The numerical analysis relies on $n_S = 5\,000$ Monte Carlo simulations, each combining an ARIMA mortality-factor path with an independent Gaussian return path. For each candidate conversion date, we generate $n_{S'} = 1\,000$ ARIMA mortality-factor continuations conditional on the state reached at that date. We also draw $n_R = 100$ individual lifetimes from the path-specific death-time distribution for each outer path. The n_S outer mortality and investment paths are held constant across candidate annuitization dates. The n_R lifetime draws are also reused across candidates, whereas the $n_{S'}$ mortality continuations are generated conditionally for each conversion-date state. Table 19 summarizes these calibration and simulation choices and defines the common numerical environment used throughout the results below.

Table 19: Numerical design for the fixed-date actuarial analysis

Quantity	Baseline specification
Country, valuation date, issue age	France, 2023, age 65
Limiting age	$x^+ = 95$
Mortality model	LC Poisson
Mortality forecast	ARIMA(1,1,0) with linear trend
Calibration window	1950–2023 with ages 50–95
Period-factor lookback	15 years
Initial wealth	$W_0 = 100$
Pricing discount rate	$\varrho = 2\%$
Objective discount rate	$\tilde{\varrho} = 2\%$
Withdrawal rate	$\alpha = 4\%$
Loading factor	$\mathcal{L} = 5\%$
Bequest weight	$\psi \in \{0, 0.25, 0.5, 0.75, 1\}$
Investment return	$\mu \in \{2\%, 3\%, \ldots, 8\%\}$, $\sigma \in \{0\%, 2\%, \ldots, 12\%, 15\%\}$
Number of simulations	$n_S = 5\,000$, $n_{S'} = 1\,000$, $n_R = 100$

In Table 20, we report the optimal annuitization delay $n^\star$ for five representative parameter sets (μ, σ): $(2\%, 0\%)$, $(3\%, 2\%)$, $(4\%, 6\%)$, $(5\%, 10\%)$, $(6\%, 15\%)$, corresponding to Sharpe ratios of ∞, 0.50, 0.33, 0.30 and 0.27, respectively. This choice reflects the decreasing Sharpe ratio pattern commonly observed in financial markets — higher expected returns are associated with higher volatility and lower Sharpe ratio. The table reports also the following statistics:

- $\Delta\mathcal{U}(n) = \dfrac{\sum_{s=1}^{n_S} \mathcal{U}^{(s)}(n)}{\sum_{s=1}^{n_S} \mathcal{U}^{(s)}(0)} - 1$ is the expected gain in the objective function relative to the fix-now strategy.
- $\bar{P}_{\mathcal{U}}(n) = \frac{1}{n_S} \sum_{s=1}^{n_S} \mathbb{1}\left\{\mathcal{U}^{(s)}(n) > \mathcal{U}^{(s)}(0)\right\}$ is the probability of outperformance relative to the fix-now strategy. $P_{\mathcal{U}}(n) = \frac{1}{n_S n_R} \sum_{s=1}^{n_S} \sum_{r=1}^{n_R} \mathbb{1}\left\{\mathcal{U}^{(s,r)}(n) > \mathcal{U}^{(s,r)}(0)\right\}$ includes the uncertainty on the survival time.
- $Q_\alpha(n)$ and $\mathrm{CTE}_\alpha(n)$ are the quantile and the conditional tail expectation of the series $\Delta\mathcal{U}^{(s,r)}(n) = \mathcal{U}^{(s,r)}(n) - \mathcal{U}^{(s,r)}(0)$.
- $P_{\boldsymbol{\tau}}(n) = \frac{1}{n_S} \sum_{s=1}^{n_S} p^{(s)}\left(x, t_0, \boldsymbol{\tau}^{(s)} - t_0\right) \mathbb{1}\left\{\boldsymbol{\tau}^{(s)} \leq t_0 + n\right\}$ is the probability of exhausting the liquid account before annuitization[60].

[60] We have $\boldsymbol{\tau}^{(s)} = \inf\left\{t_0 + j \leq t_0 + n : W^{(s)}_{t_0+n} \leq \epsilon\right\}$ where $\epsilon = 10^{-12}$.

Table 20: Results of the fixed-date problem (France, 2023, LC Poisson, $\alpha = 4\%$)

ψ	Sex	μ (%)	σ (%)	$n^\star$	$\Delta\mathcal{U}(n^\star)$ (%)	$\bar{P}_{\mathcal{U}}(n^\star)$ (%)	$P_{\mathcal{U}}(n^\star)$ (%)	$Q_\alpha(n^\star)$	$\mathrm{CTE}_\alpha(n^\star)$	$P_{\boldsymbol{\tau}}(n^\star)$ (%)
0.00	Female	2	0	0						
		3	2	10	3.7	74.0	62.0	−9.8	−10.9	0.0
		4	6	15	14.7	74.0	57.9	−17.1	−20.2	0.0
		5	10	17	29.6	73.0	54.8	−22.3	−31.8	0.0
		6	15	20	49.2	66.9	46.1	−36.4	−49.2	3.3
0.00	Male	2	0	0						
		3	2	0						
		4	6	10	5.1	57.9	45.8	−18.9	−24.6	0.0
		5	10	13	15.0	60.8	45.1	−28.3	−38.2	0.0
		6	15	17	29.1	57.8	38.0	−43.1	−56.9	1.3
0.25	Female	2	0	0						
		3	2	13	5.7	83.4	68.9	−11.5	−13.6	0.0
		4	6	17	18.3	78.4	69.6	−19.1	−21.5	0.0
		5	10	20	35.3	75.8	68.1	−23.1	−32.7	0.3
		6	15	22	58.0	69.9	63.7	−36.9	−49.7	4.5
0.25	Male	2	0	0						
		3	2	6	0.8	56.7	46.6	−10.1	−11.0	0.0
		4	6	13	9.1	65.3	59.3	−22.6	−26.8	0.0
		5	10	17	21.8	66.3	60.0	−30.6	−40.6	0.0
		6	15	19	38.7	62.9	58.2	−43.7	−57.9	2.3
0.50	Female	2	0	0						
		3	2	16	8.5	91.6	68.2	−14.6	−16.8	0.0
		4	6	20	23.2	83.8	73.2	−21.8	−23.6	0.0
		5	10	23	43.0	79.9	73.7	−24.5	−33.8	1.0
		6	15	24	69.7	73.6	69.7	−37.6	−50.0	5.9
0.50	Male	2	0	0						
		3	2	12	3.7	73.2	49.5	−18.5	−20.1	0.0
		4	6	18	15.2	74.3	64.9	−29.8	−32.0	0.0
		5	10	21	30.9	73.2	67.7	−34.0	−42.4	0.5
		6	15	22	51.5	68.5	65.6	−44.6	−58.8	3.9
0.75	Female	2	0	10	0.4	100.0	13.5	−10.0	−11.1	0.0
		3	2	22	12.9	98.0	69.9	−19.9	−23.1	0.0
		4	6	26	30.8	89.8	78.4	−26.9	−28.1	0.2
		5	10	29	55.3	85.1	79.6	−29.2	−35.6	3.3
		6	15	29	88.9	78.1	74.3	−38.4	−50.2	8.3
0.75	Male	2	0	0						
		3	2	20	9.3	93.8	58.9	−29.1	−31.9	0.0
		4	6	23	24.3	85.4	70.8	−36.4	−37.8	0.0
		5	10	26	44.0	81.6	74.3	−40.1	−46.0	1.8
		6	15	29	70.2	75.1	72.1	−46.7	−59.9	6.8
1.00	Female	2	0	∞	5.0	100.0	42.9	−19.6	−19.6	0.0
		3	2	∞	21.7	100.0	79.3	−9.6	−13.4	0.0
		4	6	∞	44.4	94.6	87.1	−13.8	−22.7	0.8
		5	10	∞	74.7	89.6	85.6	−22.8	−34.5	3.5
		6	15	∞	116.0	81.9	80.0	−38.4	−50.2	8.6
1.00	Male	2	0	∞	5.0	100.0	44.9	−34.3	−34.4	0.0
		3	2	∞	20.0	99.8	65.7	−22.0	−26.2	0.0
		4	6	∞	39.6	94.4	81.4	−23.5	−33.4	0.6
		5	10	∞	65.1	88.6	82.2	−31.3	−44.2	2.5
		6	15	∞	98.5	80.8	78.0	−46.7	−59.9	6.9

Baseline fixed-date results Table 20 reports the optimal fixed-date annuitization rule for the five representative investment environments when the withdrawal rate is set to $\alpha = 4\%$. Three broad patterns emerge from the results. First, the optimal annuitization date is highly sensitive to investment opportunities. Immediate annuitization is optimal mainly in the low-return deterministic environment $(\mu, \sigma) = (2\%, 0\%)$, with only a few exceptions in the grid[61], whereas delayed annuitization becomes increasingly attractive as expected returns rise. For both sexes, the optimal delay tends to increase with μ, reflecting the growing value of retaining investment flexibility before converting wealth into a life annuity. The associated gains in the objective function are economically significant. For example, when $\psi = 0.75$, the expected improvement relative to immediate annuitization increases from 12.9% to 88.9% for females and from 9.3% to 70.2% for males when moving from $(3\%, 2\%)$ to $(6\%, 15\%)$. Second, stronger bequest motives lead to later annuitization. Holding investment conditions constant, the optimal conversion date increases monotonically with ψ. For instance, under $(\mu, \sigma) = (3\%, 2\%)$, the optimal delay for females rises from 10 years when $\psi = 0$ to 13, 16 and 22 years when $\psi = 0.25$, 0.50 and 0.75, respectively. Similar patterns are observed in the other investment environments. When $\psi = 1$, annuitization is never optimal for either sex. This suggests that a retiree who values consumption and bequest equally prefers to maintain liquid wealth throughout retirement rather than exchange it for an irreversible life annuity. Third, females generally postpone annuitization longer than males. The difference is modest in low-return environments, but it becomes more pronounced as investment opportunities improve. For example, when $\psi = 0.25$, the optimal delays for females are 13, 17, 20 and 22 years across the four risky portfolios, compared with 6, 13, 17 and 19 years for males. These longer delays are consistent with females' lower mortality rates and correspondingly longer expected retirement horizons. This raises the value of deferring conversion and purchasing the annuity at a later, more favorable point.

Table 20 also illustrates the differences between the scenario-level probability of outperformance $\bar{P}_{\mathcal{U}}(n^\star)$ and the realized-lifetime probability $P_{\mathcal{U}}(n^\star)$. In all configurations, $\bar{P}_{\mathcal{U}}(n^\star)$ is greater than $P_{\mathcal{U}}(n^\star)$. The former averages utilities using path-specific survival probabilities, whereas the latter incorporates simulated individual lifetimes. Therefore, introducing realized longevity adds idiosyncratic mortality risk and reduces the frequency with which delayed annuitization outperforms immediate annuitization. Interestingly, the realized-lifetime probabilities are not monotonic in portfolio risk. For females with $\psi = 0.75$, $P_{\mathcal{U}}(n^\star)$ increases from 69.9% under $(3\%, 2\%)$ to 79.6% under $(5\%, 10\%)$, before declining to 74.3% under $(6\%, 15\%)$. A similar pattern is observed for males. Finally, downside risk remains substantial even when delayed annuitization is optimal. The lower-tail measures deteriorate as portfolio volatility increases. For females with $\psi = 0.75$, the 5% conditional tail expectation $\text{CTE}_{5\%}(n^\star)$ falls from -23.1 under $(3\%, 2\%)$ to -50.2 under $(6\%, 15\%)$. Thus, larger expected gains from delaying annuitization are accompanied by increasingly severe losses in adverse realized outcomes. Nevertheless, the probability $P_{\boldsymbol{\tau}}(n^\star)$ of exhausting the liquid account before annuitization remains relatively low, generally below 10% even under the most aggressive delay strategies.

Figure 38 shows how the lifetime consumption and bequest components vary with the conversion date. Under $(\mu, \sigma) = (2\%, 0\%)$, the lifetime consumption component steadily declines as annuitization is postponed, falling from 95.2 at $n = 0$ to 74.6 at $n = 29$. The profile changes when the expected portfolio return is higher. The consumption component first rises with deferral, reaching a maximum that occurs progressively later as μ increases. In the four risky representative cases, the maximum occurs around $n = 10$, 15, 17, and 20, with corresponding peak values of 98.8, 109.3, 123.4, and 142.1, respectively. The bequest

[61] For instance, at $\psi = 0.75$, the optimal delay for females under $(2\%, 0\%)$ is $n^\star = 10$ rather than 0. At $\psi = 0$, the optimal delay for males under $(3\%, 2\%)$ is also $n^\star = 0$.

Figure 38: Fixed-date objective components — $\mathbb{E}\left[\mathcal{C}\left(n\right)\right]$, $\mathbb{E}\left[\mathfrak{B}\left(n\right)\right]$ and $\mathbb{E}\left[\mathcal{U}\left(n\right)\right]$ (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

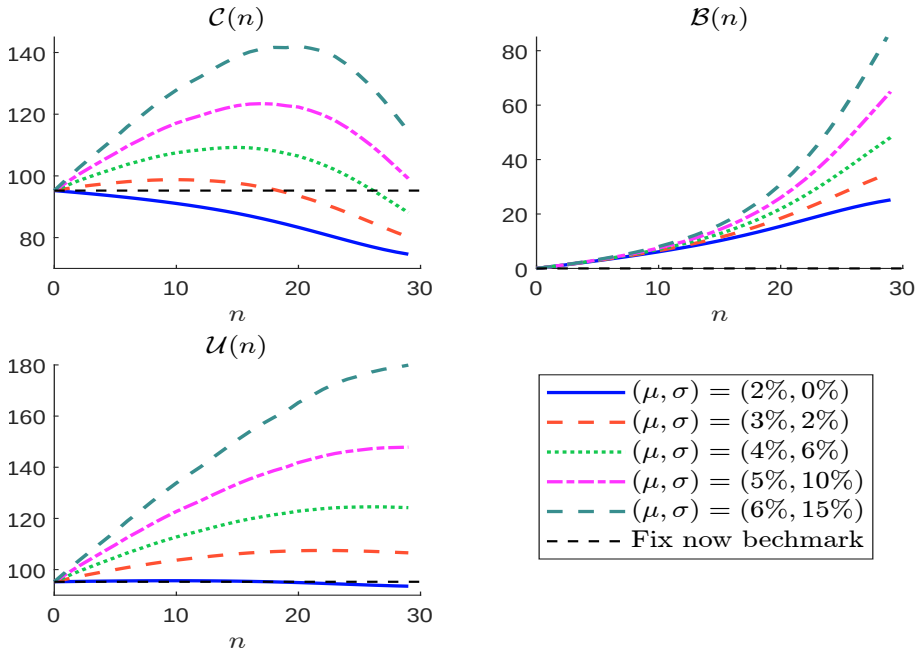
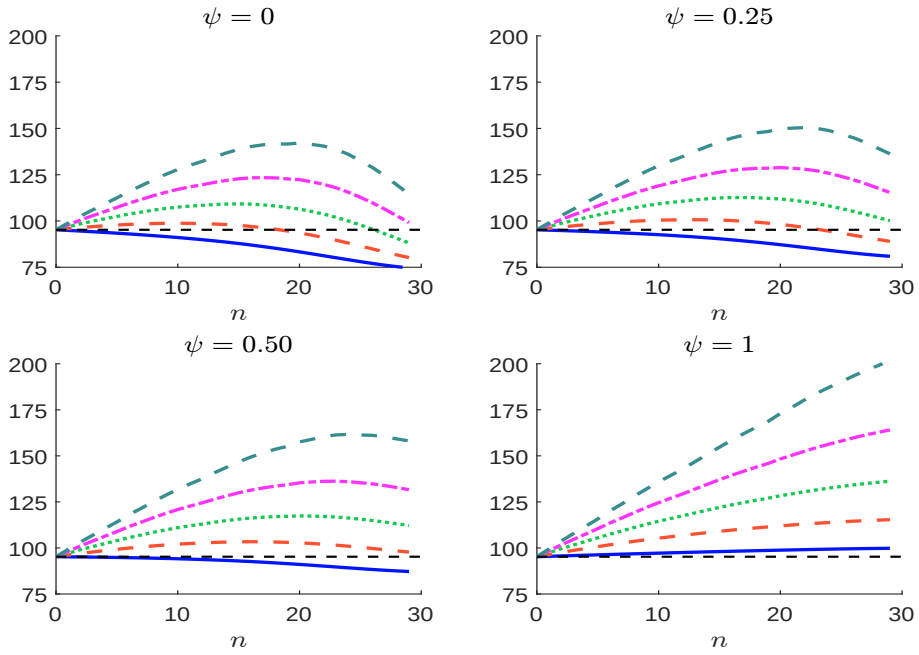
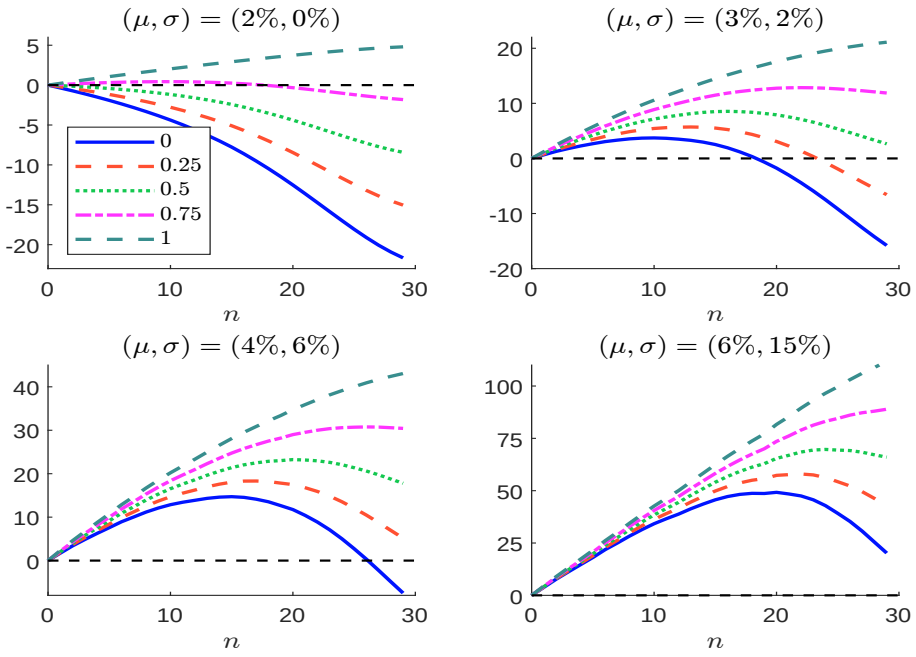


Figure 39: Objective function $\mathbb{E}\left[\mathcal{U}\left(n\right)\right]$ (France, female, 2023, LC Poisson, $\alpha = 4\%$)



component has a simpler shape. It increases with the conversion date in every case, since wealth remains liquid and therefore bequestable for longer. At $n = 29$, the estimated component ranges from 25.2 under $(2\%, 0\%)$ to 87.3 under $(6\%, 15\%)$. Higher expected returns therefore allow liquid wealth to support consumption for a longer period of time and also leave more wealth exposed to the bequest term. Combining the two components yields the objective function $\mathbb{E}\left[\mathcal{U}(n)\right]$, whose shape depends heavily on the bequest weight ψ. A low value of ψ gives more weight to the hump-shaped consumption profile, so the optimum tends to be near the consumption peak. In contrast, a high value of ψ gives more weight to the monotonically increasing bequest profile, pushing the optimum later[62]. At $\psi = 1$, annuitization is eliminated altogether. This trade-off is precisely what generates the delay patterns reported in Table 20. The same qualitative patterns hold for males (see Figure 99 on page 179).

Figure 40: Expected gain $\Delta\mathcal{U}(n)$ in % (France, female, 2023, LC Poisson, $\alpha = 4\%$)



The resulting objective gains are reported in Figure 39. With no bequest motive ($\psi = 0$), the deterministic case favors the fix-now strategy, whereas finite deferral is already optimal in each of the four risky representative cases. For females, the best finite delays are 10, 15, 17, and 20 years, with objective gains of 3.7%, 14.7%, 29.6%, and 49.2%, respectively. At $\psi = 0.75$, the corresponding delays increase to 22, 26, 29, and 29 years, and the gains rise to 12.9%, 30.8%, 55.3%, and 88.9%, respectively. The investment environment (μ, σ) largely determines the size of the gain from remaining liquid, while a stronger bequest motive tends to extend the preferred deferral horizon and increase the resulting gain.

Figure 41 plots the realized lifetime outperformance probability $P_{\mathcal{U}}(n)$ against n for females and the five values of ψ, across the representative investment environments. The dashed horizontal line at 50% marks the point beyond which delaying annuitization outperforms the fix-now strategy in a majority of simulated outcomes. In the determinis-

[62]The cases $\psi = 0$, $\psi = 0.25$, $\psi = 0.50$, and $\psi = 1$ are shown in Figure 39. We only report the objective function $\mathbb{E}\left[\mathcal{U}(n)\right]$ since the lifetime consumption and bequest components do not change.

Figure 41: Probability of outperformance $P_{\mathcal{U}}(n)$ in % (France, female, 2023, LC Poisson, $\alpha = 4\%$)

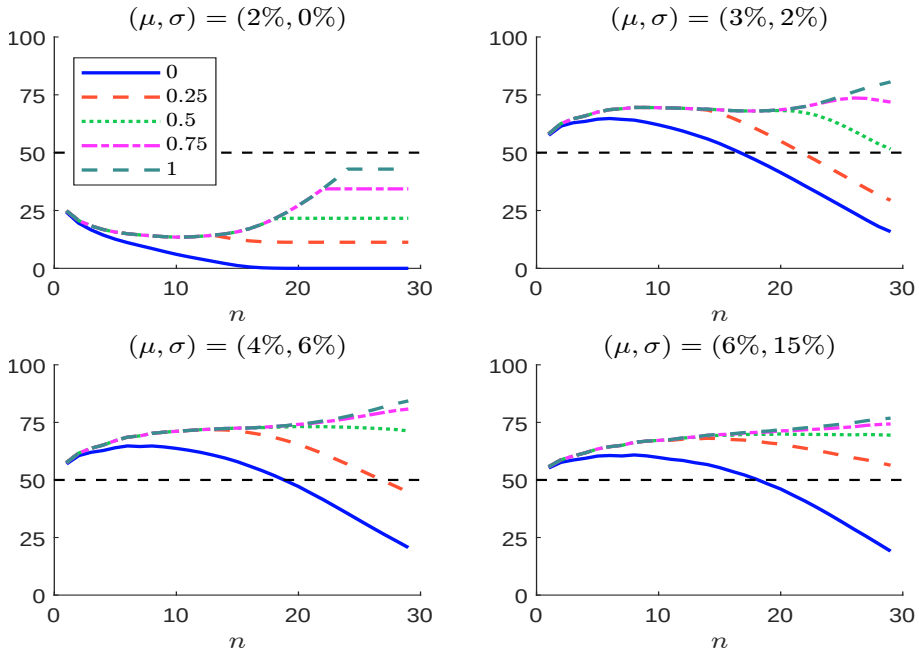
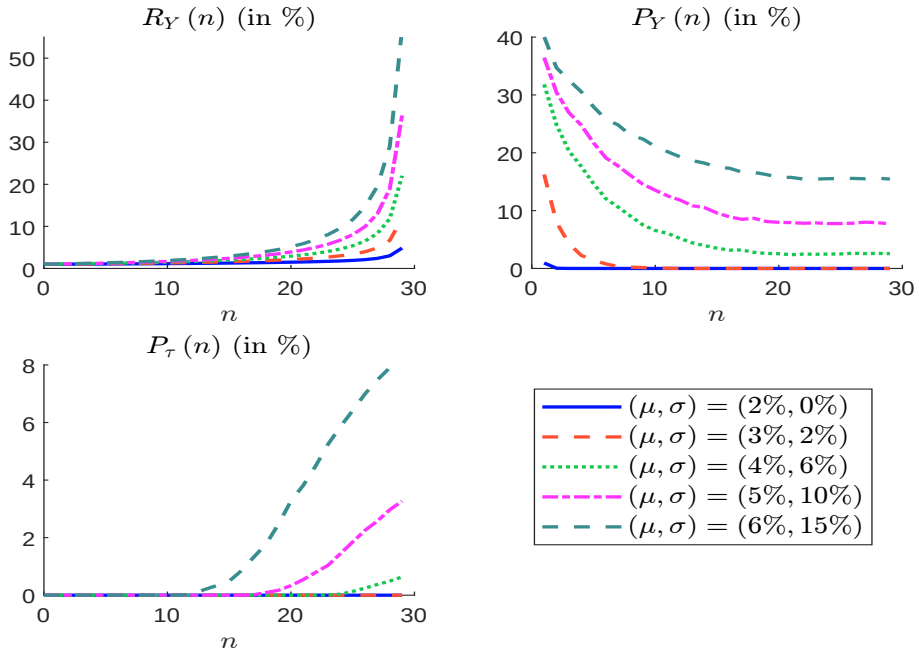


Figure 42: Fixed-date outcome diagnostics (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)



tic case $(\mu, \sigma) = (2\%, 0\%)$, outperformance probabilities remain modest throughout. For $\psi \in \{0, 0.25\}$, $P_{\mathcal{U}}(n)$ declines from about 20–25% at short horizons to nearly zero as n increases. Thus, delaying annuitization is unfavorable on nearly every simulated path. For $\psi \in \{0.5, 0.75, 1\}$, the probability first declines and then recovers as n increases, reflecting the growing importance of the bequest term. However, it never crosses the 50% threshold. Even with the strongest bequest motive, deferring annuitization in this riskless environment remains unfavorable on a majority of simulated paths, which is consistent with the modest $P_{\mathcal{U}}(n^\star)$ value of 42.9% reported in Table 20 for $\psi = 1$. The picture changes markedly once investment risk is introduced. In the three risky environments, $P_{\mathcal{U}}(n)$ rises sharply over the first few years, peaking at 65–70% around $n = 5$ to 10, regardless of ψ. Beyond this point, the curves diverge according to the strength of the bequest motive. For low values of ψ, the probability steadily declines as n increases and eventually falls below the 50% threshold. Thus, beyond a certain horizon, immediate annuitization outperforms deferral on most simulated paths, even though, the corresponding optimal delay $n^\star$ is comparatively short. For high values of ψ, however, the probability remains close to its peak or rises again as n increases. Thus, deferral continues to outperform immediate annuitization on most paths well beyond $n = 20$. This pattern is consistent with the substantially longer optimal delays and higher $P_{\mathcal{U}}(n^\star)$ values reported for $\psi = 0.75$ and $\psi = 1$ in Table 20. The divergence between the low- and high-ψ curves becomes more pronounced as μ and σ increase, mirroring the widening gap in optimal delays documented previously.

Future annuity income provides another way of understanding the consequences of deferral. Figure 42 reports the annuity-income ratio conditional on surviving to the conversion date[63]:

$$R_Y(n) = \mathbb{E}\left[\frac{Y_{t_0+n}}{Y_0}\middle| \text{alive at } t_0+n\right]$$

At the optimal conversion age, $R_Y(n^\star)$ equals 1.2%, 2.5%, 6.5%, 36%, and 57.0% across the five representative cases. The large ratios in the final two configurations reflect the wealth accumulated along the surviving paths and the higher annuity payout available at older conversion ages. However, as these figures are conditional survivor averages, they should not be interpreted as guaranteed increases in retirement income. Figure 42 also reports the downside probability of income[64]:

$$P_Y(n) = \Pr\left\{Y_{t_0+n} < Y_0 \mid \text{alive at } t_0+n\right\}$$

At the selected conversion dates, $P_Y(n^\star)$ is zero in the first two representative cases. It remains low at $(4\%, 6\%)$, where the estimate is approximately 2.6%, but rises to 7.7% at $(5\%, 10\%)$ and to 15.5% at $(6\%, 15\%)$. The high conditional mean incomes obtained in the more aggressive portfolios are therefore accompanied by an increasing risk that the income ultimately purchased falls short of the initial reference level Y_0. Finally, the third panel shows the probability of exhaustion $P_{\boldsymbol{\tau}}(n)$. At the baseline withdrawal rate, exhaustion is limited for most of the relevant deferral range. It is absent throughout the finite-delay range

[63] We use the following estimator:

$$R_Y(n) = \frac{\sum_{s=1}^{n_S} p^{(s)}(x, t_0, n)\, Y_{t_0+n}^{(s)}}{Y_0 \sum_{s=1}^{n_S} p^{(s)}(x, t_0, n)}$$

[64] We use the following estimator:

$$P_Y(n) = \frac{\sum_{s=1}^{n_S} p^{(s)}(x, t_0, n)\, \mathbb{1}\left\{Y_{t_0+n} < Y_0\right\}}{\sum_{s=1}^{n_S} p^{(s)}(x, t_0, n)}$$

in the first two representative cases. For $n = 29$ the female exhaustion probabilities are approximately 0.6%, 3.3%, and 8.3% in the three most volatile cases. For moderate delays, the main source of uncertainty therefore comes from investment and lifetime outcomes rather than outright depletion of the liquid account. Exhaustion only becomes a material risk when a volatile portfolio is kept liquid until very old ages.

Table 21: Optimal policy counts over the (μ, σ) grid (France, 2023, LC Poisson, $\alpha = 4\%$)

ψ	Female			Male		
	Fix-now	Flex-to-fix	Never	Fix-now	Flex-to-fix	Never
0.00	8	48	0	16	40	0
0.25	6	50	0	8	48	0
0.50	2	54	0	8	48	0
0.75	0	56	0	8	48	0
1.00	0	0	56	0	0	56

Table 22: Optimal policy counts over the (ψ, σ) grid (France, 2023, LC Poisson, $\alpha = 4\%$)

μ	Female			Male		
	Fix-now	Flex-to-fix	Never	Fix-now	Flex-to-fix	Never
2%	16	16	8	32	0	8
3%	0	32	8	8	24	8
4%	0	32	8	0	32	8
5%	0	32	8	0	32	8
6%	0	32	8	0	32	8
7%	0	32	8	0	32	8
8%	0	32	8	0	32	8

Table 23: Optimal policy counts over the (μ, ψ) grid (France, 2023, LC Poisson, $\alpha = 4\%$)

σ	Female			Male		
	Fix-now	Flex-to-fix	Never	Fix-now	Flex-to-fix	Never
0%	3	25	7	5	23	7
2%	3	25	7	5	23	7
4%	2	26	7	5	23	7
6%	2	26	7	5	23	7
8%	2	26	7	5	23	7
10%	2	26	7	5	23	7
12%	1	27	7	5	23	7
15%	1	27	7	5	23	7

The full (μ, σ, ψ) grid provides a broader view of the policy regions. Tables 21, 22, and 23 summarize the distribution of optimal policies across the parameter space according to bequest weight ψ (56 cases), expected return μ (42 cases), and volatility σ (35 cases), respectively. The bequest motive emerges as the primary determinant of the qualitative form of the optimal policy. When $\psi = 0$, immediate annuitization (fix-now) is optimal in only eight of the 56 female cases and 16 of the 56 male cases, while the remaining configurations select a finite delayed-annuitization strategy (flex-to-fix). As the bequest weight increases, the region associated with immediate annuitization steadily shrinks. At the baseline value

Figure 43: Optimal annuitization date (France, female, 2023, LC Poisson, $\alpha = 4\%$)

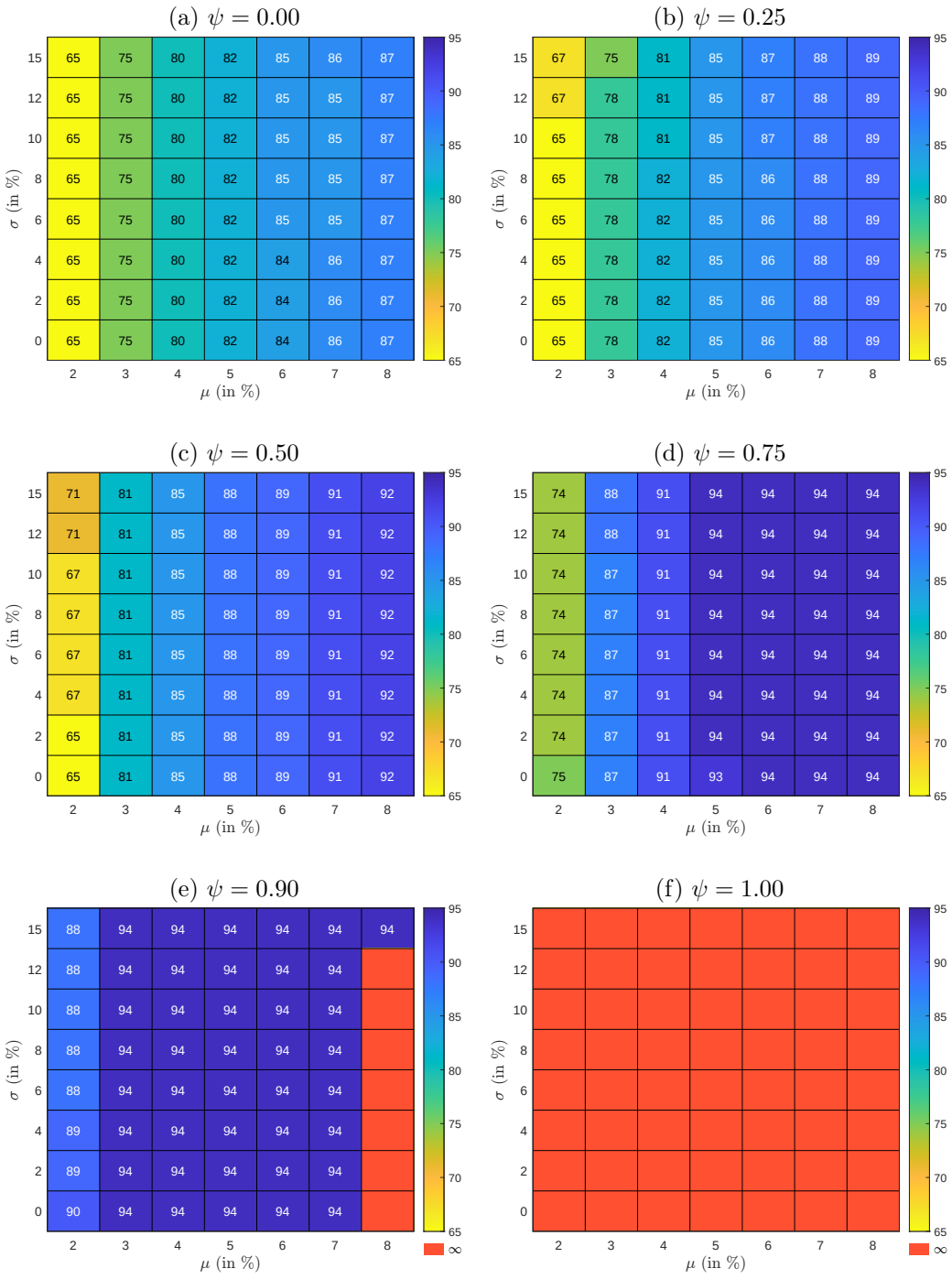
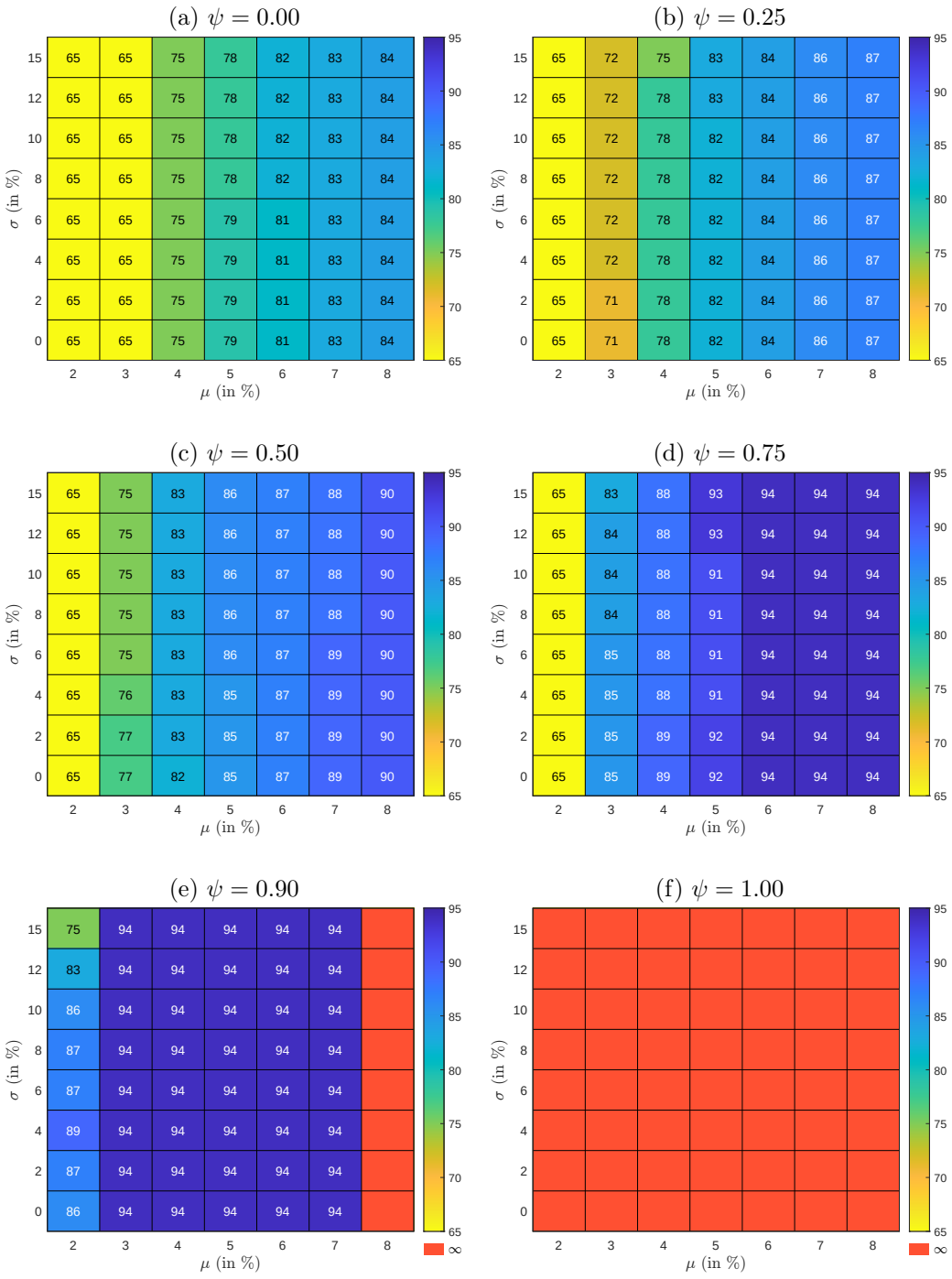


Figure 44: Optimal annuitization date (France, male, 2023, LC Poisson, $\alpha = 4\%$)



$\psi = 0.75$, the flex-to-fix strategy is optimal in all 56 female cases. For males, only the eight configurations corresponding to $\mu = 2\%$ remain in the fix-now region, while the other 48 cases select flex-to-fix. At the extreme value $\psi = 1$, the fix-never strategy becomes optimal throughout the entire grid for both sexes, indicating that a retiree who places equal value on consumption and bequest always prefers to retain liquid wealth rather than convert it into an irreversible life annuity. Investment returns also have substantial influence, although less pronounced than the bequest motive. Immediate annuitization is concentrated almost exclusively in the low-return region of the parameter space, whereas higher expected returns generally favor delayed annuitization. By contrast, volatility has only a limited effect on the qualitative policy classification. Changes in σ modify the preferred annuitization date within the flex-to-fix region, but they rarely alter the nature of the optimal strategy itself. Overall, the results suggest a clear hierarchy of effects on the optimal policy:

$$\psi \succ \mu \succ \sigma$$

The bequest motive has the strongest impact, expected returns have a secondary but economically significant effect, and volatility plays a comparatively minor role. Figures 43 and 44 complement this analysis by displaying the optimal annuitization delay in each grid cell. These heatmaps provide a visual representation of the policy regions and illustrate how the optimal conversion date evolves across the parameter space.

Figure 45: Investment return and volatility sensitivity of lower-tail objective loss and realized lifetime outperformance for the best finite flex-to-fix candidates (France, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

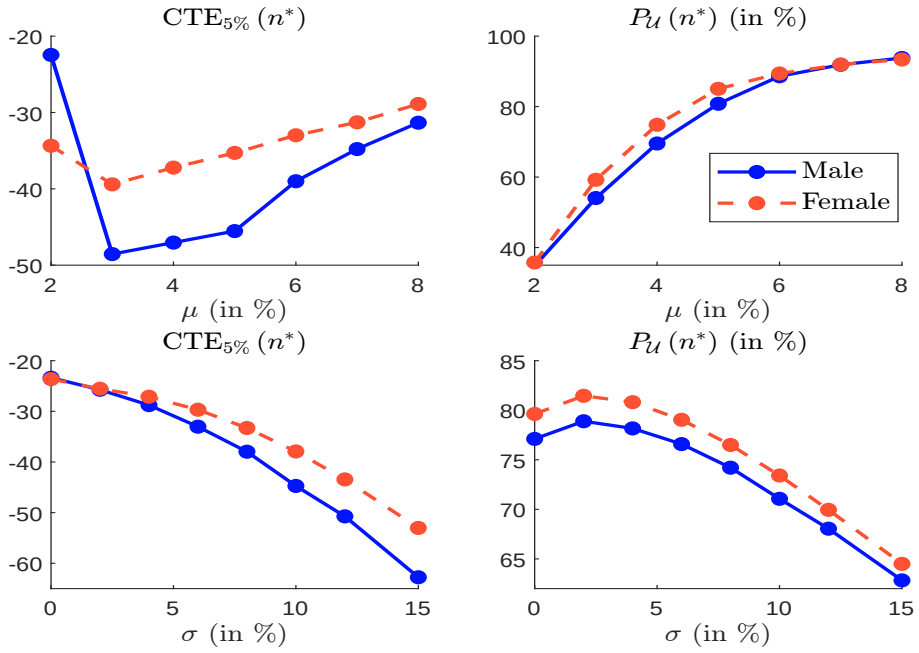


Figure 45 focuses instead on the realized risk dimension. For each (μ, σ) cell, we first select the best finite flex-to-fix candidate n^*, and then average the resulting risk measures across σ for each value of μ (top panels), and across μ for each value of σ (bottom panels). The effect of μ is not uniform across the two risk measures. The average $\text{CTE}_{5\%}\,(n^*)$ is non-monotonic. It worsens sharply between $\mu = 2\%$ and $\mu = 3\%$ (from -22.5 to -48.5 for males,

and from -34.4 to -39.4 for females), before improving steadily as μ increases, reaching -31.3 for males and -28.9 for females at $\mu = 8\%$. By contrast, average $P_{\mathcal{U}}(n^*)$ increases monotonically with μ for both sexes, from roughly 35% at $\mu = 2\%$ to 93% at $\mu = 8\%$, with the male and female curves converging at high μ. Higher expected returns therefore increase the likelihood that the best finite rule outperforms immediate annuitization, even though they do not uniformly reduce lower-tail losses. The effect of σ is more consistent across the two measures. Average $\text{CTE}_{5\%}(n^*)$ declines steadily and without reversal as σ increases. It falls from about -23.6 to -53.0 for females and from about -23.4 to -62.8 for males between $\sigma = 0\%$ and $\sigma = 15\%$, with the gap between the sexes widening at higher volatility. Average $P_{\mathcal{U}}(n^*)$ instead rises slightly before declining. For females, it rises from 79.6% at $\sigma = 0\%$ to a peak of 81.5% at $\sigma = 2\%$, then falls to 64.5% at $\sigma = 15\%$. Males follow the same pattern at a slightly lower level throughout, falling to 62.8% at $\sigma = 15\%$. Thus, volatility has a much clearer effect on the distribution of realized outcomes than on the expected objective of the selected finite policy.

Sensitivity analysis with respect to the withdrawal rate α We now conduct a sensitivity analysis with respect to the withdrawal rate α. We recall that the benchmark value $\alpha = 4\%$ is motivated by the empirical analysis of Bengen (1994, 1996, 1997), who developed the historical backtesting methodology for determining safe withdrawal rates. This rule was subsequently extended and popularized by Cooley *et al.* (1998). However, the safe withdrawal rate depends on the characteristics of asset returns (μ, σ). In what follows, we consider α to be in the range of $\{1\%, 1.5\%, 2\%, \ldots, 8\%\}$. Throughout this sensitivity analysis, the bequest weight is held fixed at $\psi = 0.75$, while the five previously considered representative return configurations are retained.

Table 24: Normalized withdrawal rates (in %)

α		1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000
$\tilde{\alpha}$	Female	0.187	0.375	0.562	0.749	0.937	1.124	1.311	1.499
	Male	0.167	0.334	0.500	0.667	0.834	1.001	1.168	1.334

Remark 17 *To make spending levels comparable across sexes or clusters, the withdrawal rate can be normalized using the loaded fix-now annuity-income rate:*

$$\tilde{\alpha} = \frac{\alpha}{Y_0/W_0}$$

The immediate annuity-income rates are $Y_0/W_0 = 5.3378\%$ for females and 5.9950% for males. Table 24 reports the resulting normalized withdrawal rates. By construction, $\tilde{\alpha} = 1$ corresponds to $\alpha = Y_0/W_0 := \alpha^\star$, i.e., to a withdrawal rate exactly equal to the fix-now annuity income. Therefore, we anticipate a change in behavior once $\tilde{\alpha} > 1$, since the retiree then withdraws more than the immediate annuity could deliver. This places additional strain on the liquid account relative to the fix-now benchmark.

Table 25 reports the optimal fixed-date rule for the five representative investment environments across the withdrawal rate grid. The deterministic case $(\mu, \sigma) = (2\%, 0\%)$ behaves very differently from the four risky cases, and the transition is sharp rather than gradual. For $\alpha \leq 5\%$, *i.e.*, as long as $\tilde{\alpha} < 1$, the deterministic case generates essentially no exhaustion risk. The policy is either fix-now or a short finite delay, and $P_{\boldsymbol{\tau}}(n^\star)$ remains at 0.0% throughout for both sexes. However, once α reaches 6% — where $\tilde{\alpha}$ crosses the unit threshold (1.12 for females and 1.00 for males) — the deterministic case abruptly switches to a

Table 25: Sensitivity to the withdrawal rate (France, 2023, LC Poisson, $\psi = 0.75$)

α (%)	$\tilde{\alpha}$	Sex	μ (%)	σ (%)	Policy	$n^\star$	$\Delta\mathcal{U}(n^\star)$ (%)	$P_{\mathcal{U}}(n^\star)$ (%)	$\text{CTE}_\alpha(n^\star)$	$P_{\boldsymbol{\tau}}(n^\star)$ (%)
1	0.19	Female	2	0	Fix-now	0				
			3	2	F2F	20	13.0	70.4	−67.5	0.0
			4	6	F2F	26	36.9	77.0	−87.3	0.0
			5	10	F2F	29	70.4	79.2	−94.7	0.0
			6	15	F2F	29	115.7	74.7	−94.7	0.0
1	0.17	Male	2	0	Fix-now	0				
			3	2	F2F	18	8.7	65.5	−71.8	0.0
			4	6	F2F	23	28.1	71.6	−89.8	0.0
			5	10	F2F	26	54.2	74.0	−99.3	0.0
			6	15	F2F	29	89.0	72.3	−103.5	0.0
4	0.75	Female	2	0	F2F	10	0.4	13.5	−11.1	0.0
			3	2	F2F	22	12.9	69.9	−23.1	0.0
			4	6	F2F	26	30.8	78.4	−28.1	0.2
			5	10	F2F	29	55.3	79.6	−35.6	3.3
			6	15	F2F	29	88.9	74.3	−50.2	8.3
4	0.67	Male	2	0	Fix-now	0				
			3	2	F2F	20	9.3	58.9	−31.9	0.0
			4	6	F2F	23	24.3	70.8	−37.8	0.0
			5	10	F2F	26	44.0	74.3	−46.0	1.8
			6	15	F2F	29	70.2	72.1	−59.9	6.8
5	0.94	Female	2	0	F2F	16	1.1	16.6	−9.7	0.0
			3	2	F2F	24	12.9	64.5	−13.3	0.0
			4	6	F2F	26	28.7	78.2	−24.7	6.2
			5	10	F2F	29	50.2	78.8	−35.3	12.7
			6	15	F2F	29	80.2	73.9	−49.2	18.4
5	0.83	Male	2	0	Fix-now	0				
			3	2	F2F	21	9.6	51.9	−21.6	0.0
			4	6	F2F	24	23.0	70.3	−34.8	3.0
			5	10	F2F	28	40.6	74.8	−45.8	9.7
			6	15	F2F	29	64.1	71.8	−59.5	15.4
6	1.12	Female	2	0	F2F	21	2.5	42.9	−19.6	69.4
			3	2	F2F	29	12.9	64.8	−17.2	56.2
			4	6	F2F	27	26.5	79.4	−25.8	30.6
			5	10	F2F	29	45.2	79.8	−35.6	28.6
			6	15	F2F	29	71.9	75.6	−48.3	31.0
6	1.00	Male	2	0	F2F	21	0.6	44.9	−34.4	55.2
			3	2	F2F	28	10.0	57.9	−30.5	42.0
			4	6	F2F	25	21.7	73.2	−37.5	20.6
			5	10	F2F	28	37.2	76.3	−46.5	22.8
			6	15	F2F	29	58.2	73.4	−58.9	26.3
8	1.50	Female	2	0	F2F	15	3.5	42.9	−19.6	85.3
			3	2	Never	∞	11.6	57.4	−18.9	
			4	6	F2F	29	22.0	73.1	−26.8	74.6
			5	10	F2F	29	36.1	75.1	−35.5	62.0
			6	15	F2F	29	57.1	72.3	−46.7	55.5
8	1.33	Male	2	0	F2F	15	2.1	44.9	−34.4	74.5
			3	2	F2F	21	9.5	53.7	−32.4	70.6
			4	6	F2F	28	18.7	66.1	−38.9	62.8
			5	10	F2F	28	30.7	71.0	−46.9	52.4
			6	15	F2F	29	47.5	69.8	−57.8	48.0

substantial exhaustion probability: 69.4% for females and 55.2% for males at $n^\star = 21$. Since returns are riskless in this case, exhaustion is driven purely by the deterministic shortfall between the withdrawal rate α and the fixed return μ. The sharp jump in $P_{\boldsymbol{\tau}}(n^\star)$ around $\tilde{\alpha} = 1$ therefore illustrates the behavioral threshold introduced in the above remark. Females and males differ systematically in this dimension. At every common value of α, the probability of female exhaustion exceeds that of males, both in the deterministic case (*e.g.*, 85.3% versus 74.5% at $\alpha = 8\%$) and across risky configurations (*e.g.*, 55.5% versus 48.0% at $\alpha = 8\%, \sigma = 15\%$). Two effects compound here. First, females face a lower loaded fix-now annuity-income rate (5.3378% versus 5.9950% for males), so a given α translates to a higher normalized withdrawal rate $\tilde{\alpha}$ for females than for males. Second, females also have a higher probability of surviving to any given horizon, which mechanically increases the survival-weighted exhaustion measure $P_{\boldsymbol{\tau}}(n^\star)$.

The withdrawal rate significantly impacts the depletion risk. For instance, moving from $\alpha = 5\%$ to $\alpha = 8\%$ at $\sigma = 10\%$ causes the female exhaustion probability to rise from 12.7% to 62.0%, and the male probability to rise from 9.7% to 52.4%. This effect dominates all other effects documented in the paper. Neither the bequest weight nor the investment environment generates swings of this magnitude alone, confirming that the withdrawal rate is an important factor in the risk of running out of liquid wealth before conversion. Our previous ordering in terms of impact can now be completed and becomes $\alpha \succ \psi \succ \mu \succ \sigma$. Finally, higher expected returns offset some of the additional strain created by a higher withdrawal rate, but do not eliminate it. At $\alpha = 6\%$, the female exhaustion probability falls from 69.4% at $\sigma = 0\%$ to 28.6% at $\sigma = 10\%$, then rises to 31.0% at $\sigma = 15\%$. The pattern for males is similar, falling from 55.2% to 20.6% at $\sigma = 6\%$ then rising modestly at higher volatility. The same offsetting effect is visible at $\alpha = 8\%$. Even in the most favorable risky configuration, roughly half of the simulated retirees are expected to exhaust their liquid account before conversion. Since the optimal delay is nearly always close to the maximum admissible horizon in these high-withdrawal cases ($n^\star \in \{28, 29\}$ in most rows), it illustrates that combining a long deferral with an unsustainably high withdrawal rate remains a substantial source of risk, even when the underlying portfolio offers an attractive expected return.

Figures[65] 46 and 103 confirm the threshold behavior documented in Table 25 for the deterministic case. The optimal delay remains at the fix-now floor for low withdrawal rates. Then, it rises steeply around $\alpha^\star$, before eventually decreasing again as spending becomes unsustainable, even at long horizons. The finer grid reveals two features that are not visible in the five-point table. First, the transition around $\alpha^\star$ is not merely a large jump between two adjacent rows in the table, but a genuinely steep. Second, risky configurations are far less sensitive to α over most of the grid. The optimal delay for the $(5\%, 10\%)$ and $(6\%, 15\%)$ configurations sits close to the upper boundary $n = 29$, across almost the entire spending range. Only the lower-return risky case $(3\%, 2\%)$ shows a visible upward drift with α.

Figures 47 and 104 show that the ranking of $P_{\mathcal{U}}(n^\star)$ across the five investment environments is not stable over the spending grid. At low withdrawal rates, the risky configurations cluster tightly together with $P_{\mathcal{U}}(n^\star)$ close to its plateau. Meanwhile, the deterministic case sits well below, confirming that delaying annuitization at low spending outperforms fix-now in a minority of realized lifetimes in the riskless case even though the expected gain remains positive. As α rises past $\alpha^\star$, the deterministic case catches up sharply, and eventually approaches the risky curves from below, while the lower-return risky case $(3\%, 2\%)$ drifts downward over the same range.

[65] The male figures are located in Appendix on pages 181–182.

Figure 46: Optimal annuitization delay $n^\star$ across withdrawal rates α (France, female, 2023, LC Poisson, $\psi = 0.75$)

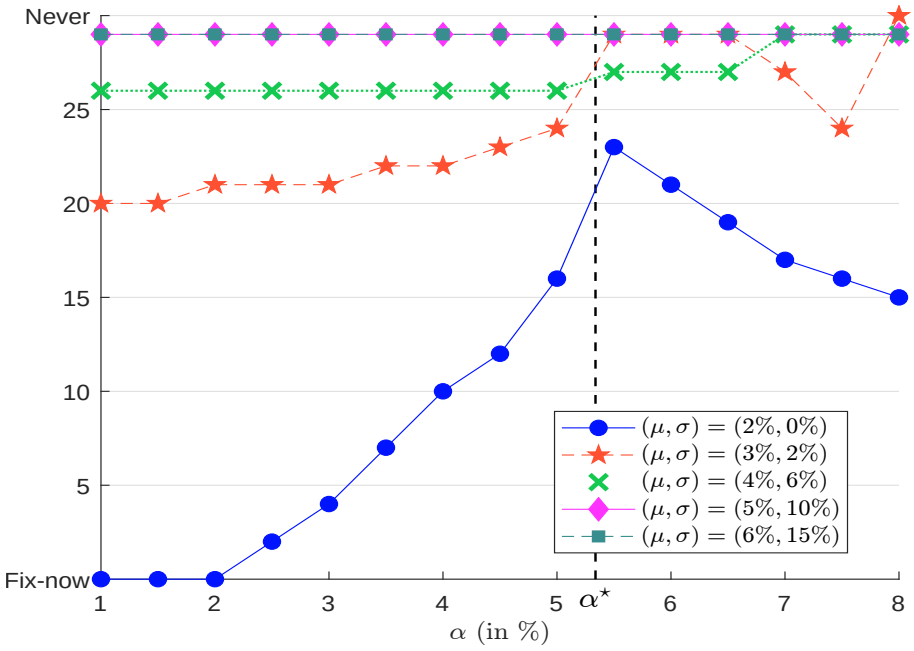


Figure 47: Realized lifetime outperformance probability $P_{\mathcal{U}}(n^\star)$ across withdrawal rates α (France, female, 2023, LC Poisson, $\psi = 0.75$)

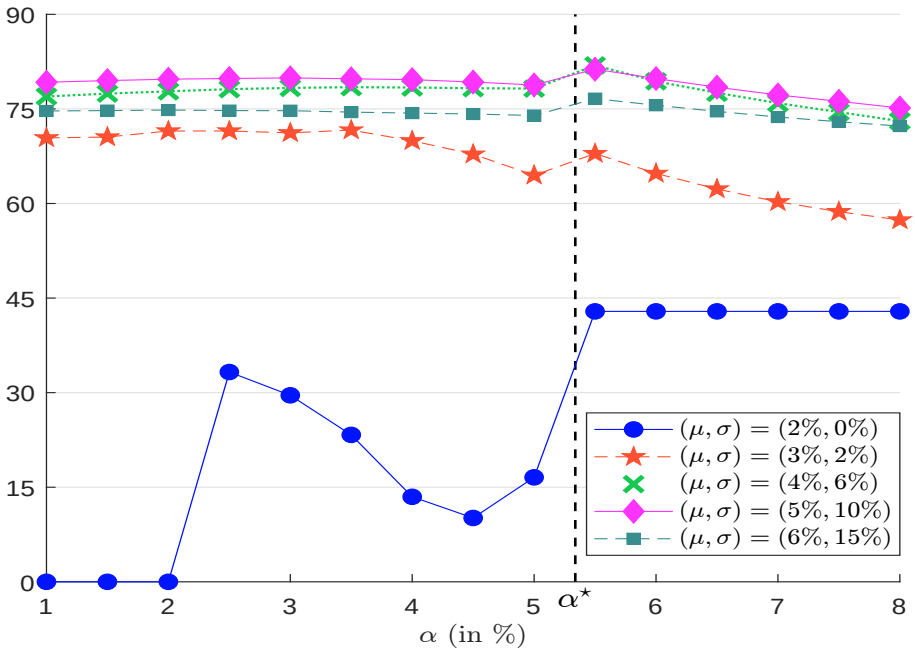
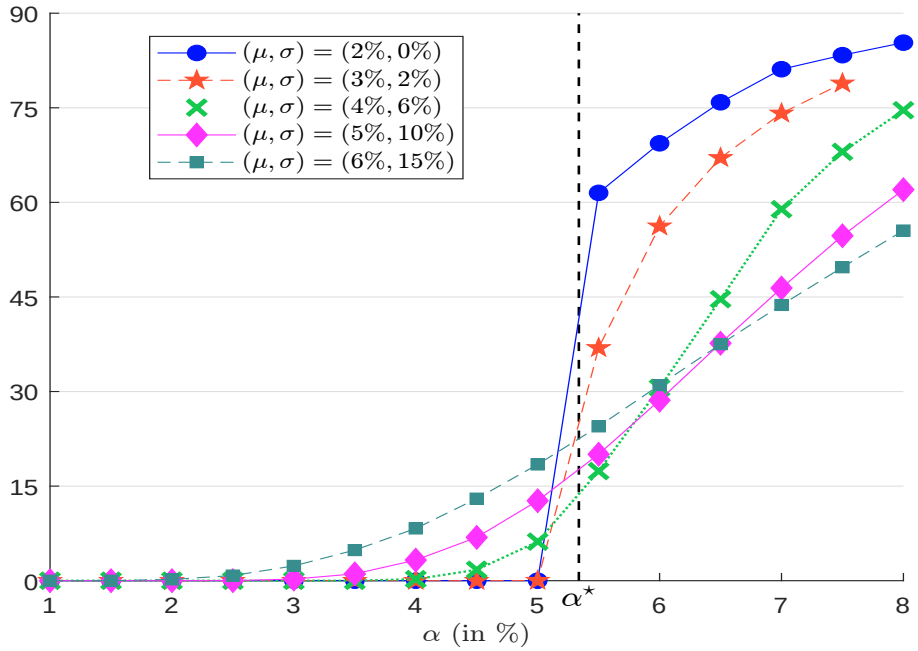


Figure 48: Probability $P_{\boldsymbol{\tau}}(n^\star)$ of liquid account depletion before annuitization across withdrawal rates α (France, female, 2023, LC Poisson, $\psi = 0.75$)



Figures 48 and 105 confirm the sharp, threshold-like rise in exhaustion risk for the deterministic case documented in the table, and show that this transition is essentially a step rather than a gradual increase. This transition occurs within roughly half a percentage point of $\alpha^\star$. By contrast, for the risky configurations, the increase in exhaustion probability with α is smoother and more gradual across the entire grid, with no comparable discontinuity around $\alpha^\star$.

Taken together, the fixed-date results show that the four parameters have distinct and unequal roles, which can be summarized by the ordering $\alpha \succ \psi \succ \mu \succ \sigma$. The withdrawal rate α has the largest overall impact. Once spending is high enough relative to the fix-now annuity income, exhaustion risk rises sharply, and the optimal conversion date may change significantly. The bequest weight ψ systematically pushes conversion later, not by improving portfolio performance but by raising the value of keeping wealth liquid and bequestable. Its effect on the optimal delay is comparable in magnitude across most investment environments. The expected return μ plays another significant role. It increases the value of waiting and extends the selected deferral period. However, this effect is generally smaller than that of ψ and is sensitive to the withdrawal rate. Higher α can amplify or offset the benefit of a higher μ, depending on the region of the spending grid. Volatility σ has the weakest effect on broad policy regions — it rarely changes the selected range of delays — but it is not irrelevant. It strongly influences the distribution of outcomes around the optimal policy through realized-lifetime outperformance, lower-tail losses, future annuity-income risk, and pre-conversion exhaustion. However, this ordering is not fixed across the entire parameter space. The withdrawal rate's dominance is concentrated in the region where $\tilde{\alpha}$ approaches or exceeds one. Below that threshold, its effect on the optimal delay and exhaustion risk is comparatively muted. In that low-spending region, the bequest weight and the expected return become the main drivers of the optimal conversion date, while volatility continues to govern the risk around it.

4.3 Dynamic analysis

Building on our previous work on accumulation and decumulation strategies (Bruder *et al.*, 2023, 2025), we now examine the flex-to-fix strategy within a dynamic optimization framework. Several formulations of the annuitization problem have been proposed in the literature. In particular, Milevsky and Young (2007) distinguish between two institutional arrangements. Under the *all-or-nothing* arrangement, annuitization occurs at a single date and involves the retiree's entire remaining wealth. The sole decision is therefore the timing of the conversion. Under the *anything-anytime* arrangement, annuity purchases are unrestricted. Any fraction of wealth may be annuitized at any date, and each purchase is irreversible. In this section, we consider two related settings. In the first, the retiree invests in a diversified portfolio and may annuitize only once, at the beginning of a year. The objective is therefore to determine the optimal year to convert all remaining wealth into a life annuity. This specification can be viewed as a discrete-time version of the all-or-nothing arrangement. In the second setting, we study a dynamic asset-allocation problem with a target annuitization date T (*e.g.*, age 75). The retiree allocates wealth among three assets: a risky asset S_t, a risk-free asset B_t, and an annuity asset A_t, which generates a continuous income stream. Annuity purchases are irreversible, but they may occur progressively over time. At the target date T, all remaining financial wealth is converted into an annuity. In this framework, a single large annuitization decision is replaced by a sequence of smaller and gradual purchases. This reduces the dependence of retirement outcomes on the level of wealth at a particular conversion date. In both settings, the model incorporates a consumption stream $\boldsymbol{c}_t$. As argued in Bruder *et al.* (2023), many retirees require a minimum level of expenditure to meet essential living needs. For this reason, treating $\boldsymbol{c}_t$ as a fully flexible control variable is not appropriate. Instead, we assume that consumption is deterministic and specified at time t_0 for all future dates t. This assumption reflects the idea that retirees enter retirement with a planned spending path that serves as a baseline for their decumulation strategy.

4.3.1 Optimal annuitization date

Mathematical formulation The fixed-date analysis selects the conversion date at t_0 and cannot respond to subsequent mortality or investment outcomes. Now, we allow the annuitization decision to be reconsidered annually. At each date $t_j = t_0 + j$ with $j = 0, \ldots, H-1$, an individual who is alive and has not previously annuitized must choose between converting all remaining liquid wealth into an immediate life annuity or remaining liquid for one more year. Annuitization is irreversible, which is not the case of continuation (withdrawal). Therefore, the conversion date is a stopping time determined by the state observed along the retirement path.

Annuitization problem The timing within the year is the same as in the fixed-date analysis. At t_0, the first decision is made with the initial wealth W_0, before any liquid withdrawal. At each subsequent decision date t_j, the return over the preceding year has been realized, and the scheduled withdrawal has been paid. Therefore, the decision is based on the post-withdrawal wealth W_{t_j}. If the individual annuitizes at t_j, the balance W_{t_j} is converted immediately, and the first annuity payment is received one year later, conditional on survival. If the individual remains liquid, W_{t_j} is invested for one more year. If death occurs during that year before the next scheduled withdrawal, the bequest is the pre-withdrawal balance $W^-_{t_{j+1}} = W_{t^-_{j+1}}$. The state variables at decision date t_j are[66] $X_{t_j} = \left(W_{t_j}, \kappa_{t_j}, \Delta\kappa_{t_j}\right)$. The

[66]The date t_j determines the individual's age, calendar year, and remaining horizon, so it need not be included as an additional stochastic state variable.

wealth component determines the resources available for future withdrawals, the bequest, and the conversion. Under the fitted ARIMA(1,1,0) specification, the mortality state evolves according to the following system of equations:

$$\begin{cases} \Delta\kappa_{t_j+1} = \hat{c}_\kappa + \hat{\phi}_\kappa \Delta\kappa_{t_j} + \epsilon_{t_j+1} \\ \kappa_{t_j+1} = \kappa_{t_j} + \Delta\kappa_{t_j+1} \\ \epsilon_{t_j+1} \sim N\left(0, \hat{\sigma}_\kappa^2\right) \end{cases} \tag{21}$$

where $\hat{c}_\kappa$, $\hat{\phi}_\kappa$ and $\hat{\sigma}_\kappa$ are the estimated parameters of the ARIMA(1,1,0) process. Therefore, the conditional law of future mortality depends on the observed mortality history through $\left(\kappa_{t_j}, \Delta\kappa_{t_j}\right)$. Assuming liquidity is maintained, wealth evolves based on the investment return earned over the following year and the subsequent scheduled withdrawal. The investment return between t_j and $t_j + 1$ is $R_{t_j+1} \sim \mathcal{N}\left(\mu, \sigma^2\right)$ where μ is the expected annual return and σ is the annual volatility. Starting from the post-withdrawal wealth W_{t_j}, the realized return first produces the pre-withdrawal wealth $W^-_{t_j+1}$. Then, the individual attempts to withdraw the fixed annual amount $w_0 = \alpha W_0$, where α is the withdrawal rate expressed as a proportion of the initial wealth W_0. However, actual consumption $\boldsymbol{c}_{t_j+1}$ cannot exceed the available account balance. The financial transition is therefore:

$$\begin{cases} R_{t_j+1} = \mu + \sigma\varepsilon_{t_j+1} \\ \varepsilon_{t_j+1} \sim \mathcal{N}(0,1) \\ W^-_{t_j+1} = \left(1 + R_{t_j+1}\right) W_{t_j} \\ \boldsymbol{c}_{t_j+1} = \min\left(w_0, W^-_{t_j+1}\right) \\ W_{t_j+1} = W^-_{t_j+1} - \boldsymbol{c}_{t_j+1} \end{cases} \tag{22}$$

where $W^-_{t_j+1}$ is the wealth after the investment return and before the scheduled withdrawal, $\boldsymbol{c}_{t_j+1}$ is the amount actually consumed, and W_{t_j+1} is the remaining liquid wealth immediately after the withdrawal and available at the next annuitization decision. If $W^-_{t_j+1} < w_0$, the individual consumes the entire remaining account. In this case, $\boldsymbol{c}_{t_j+1} = W^-_{t_j+1}$ and $W_{t_j+1} = 0$. Conditional on remaining in drawdown, zero wealth is then absorbed. The return innovation ε_{t_j+1} is assumed to be independent of the mortality innovation ϵ_{t_j+1}.

Annuitization and continuation values Let $V\left(X_{t_j}\right)$ denote the objective value at time t_j, given that the individual is alive and liquid in state $X_{t_j} = \left(W_{t_j}, \kappa_{t_j}, \Delta\kappa_{t_j}\right)$. For a mortality state $\left(\kappa_{t_j}, \Delta\kappa_{t_j}\right)$, the unloaded annuity pricing factors are:

$$a\left(x+j, t_0+j\right) = \mathbb{E}\left[\sum_{h=1}^{H-j} \frac{p\left(x+j, t_0+j, h\right)}{\left(1+\varrho_{t_0}\right)^h} \middle| \kappa_{t_j}, \Delta\kappa_{t_j}\right]$$

and:

$$\tilde{a}\left(x+j, t_0+j\right) = \mathbb{E}\left[\sum_{h=1}^{H-j} \frac{p\left(x+j, t_0+j, h\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^h} \middle| \kappa_{t_j}, \Delta\kappa_{t_j}\right]$$

The expectations in these equations are over future systematic mortality innovations conditional on the current mortality state. If the current liquid wealth is W_{t_j}, immediate annuitization generates the following annual income:

$$Y_{t_j} = \frac{W_{t_j}}{(1+\mathcal{L})a\left(x+j, t_0+j\right)}$$

The value of annuitization is therefore:

$$V^{\text{annuitization}}\left(X_{t_j}\right) = Y_{t_j}\tilde{a}\left(x+j, t_0+j\right) = \frac{W_{t_j}}{(1+\mathcal{L})} \cdot \frac{\tilde{a}\left(x+j, t_0+j\right)}{a\left(x+j, t_0+j\right)} \tag{23}$$

Once annuitization is selected, no liquid account remains, and no subsequent annuitization decision is available. If the individual chooses to withdraw, the return and mortality state evolve over the following year. Let $q_{t_j+1} = q\left(x+j, t_0+j+1; \kappa_{t_j+1}\right)$ be the one-year death probability implied by the mortality state reached at t_j+1. If the individual survives, they receive $\boldsymbol{c}_{t_j+1}$ and reach X_{t_j+1}. If the individual dies before withdrawal, the bequest is $W^{-}_{t_j+1}$. The continuation value is therefore:

$$V^{\text{continuation}}\left(X_{t_j}\right) = \mathbb{E}\left[\frac{1-q_{t_j+1}}{1+\tilde{\varrho}_{t_0}}\left(\boldsymbol{c}_{t_j+1} + V\left(X_{t_j+1}\right)\right) + \psi\frac{q_{t_j+1}}{1+\tilde{\varrho}_{t_0}}W^{-}_{t_j+1} \mid X_{t_j}\right] \tag{24}$$

Optimal solution The Bellman equation is:

$$V\left(X_{t_j}\right) = \max\left\{V^{\text{annuitization}}\left(X_{t_j}\right), V^{\text{continuation}}\left(X_{t_j}\right)\right\} \qquad \text{for } j = 0, \ldots, H-1 \tag{25}$$

Accordingly, the optimal annuitization time is:

$$n^{\star} = \inf\left\{j \in \{0, \ldots, H-1\} : V^{\text{annuitization}}\left(X_{t_j}\right) \geq V^{\text{continuation}}\left(X_{t_j}\right)\right\} \tag{26}$$

where $n^{\star} = \infty$ if the annuitization region is never reached. At the limiting age x^{+}, there are no remaining annuity payments or subsequent decisions. The remaining liquid wealth is valued as a terminal bequest:

$$V\left(X_{t_H}\right) = \psi W_{t_H} \tag{27}$$

Every predetermined fixed-date rule is feasible in Equation (25), since it can be reproduced by continuing before the specified conversion date and annuitizing at that date for every state. Consequently, the dynamic value cannot be less than the value of any such fixed rule when both are evaluated under the same model and numerical approximation.

Remark 18 *The stopping problem is solved using backward induction on date-specific state meshes for the state variable X_{t_j}. The mortality coordinates are derived from the Gaussian distribution implied by Equation (21), while the wealth domain is obtained from the distribution generated by remaining liquid. Conditional annuity factors are computed backward on the mortality mesh. Gaussian conditional expectations are evaluated using Gauss-Hermite quadratures. The annuity factor recursions use a nine-point rule, while the continuation value in Equation (24) uses a tensor product of seven return nodes and seven mortality nodes. Annuity factor surfaces and value functions are interpolated bilinearly and trilinearly, respectively. Any transition that falls outside the finite next-period state domain is projected coordinatewise to the nearest boundary before interpolation. Appendix C.12 provides information on the numerical solution.*

Numerical results The dynamic analysis uses the same actuarial and financial calibration as the fixed-date analysis, but the results reported below are restricted to females. We recall that the retiree is aged $x = 65$ in France at the end of 2023, with limiting age $x^{+} = 95$, withdrawal rate $\alpha = 4\%$, annuity loading $\mathcal{L} = 5\%$, pricing discount rate $\varrho_{t_0} = 2\%$, and objective discount rate $\tilde{\varrho}_{t_0} = 2\%$. Mortality is modeled using the Lee-Carter Poisson specification with a period factor projected through the previously described ARIMA(1,1,0). The same investment grids $\mu \in \{2\%, 3\%, 4\%, 5\%, 6\%, 7\%, 8\%\}$ and

$\sigma \in \{0\%, 2\%, 4\%, 6\%, 8\%, 10\%, 12\%, 15\%\}$ are used, as well as the same bequest weights $\psi \in \{0, 0.25, 0.50, 0.75, 0.90, 1\}$ .The full experiment therefore contains $7 \times 8 \times 6 = 336$ dynamic stopping problems. Unless otherwise indicated, the baseline dynamic configuration is $(\mu, \sigma, \psi) = (4\%, 6\%, 0.75)$. For each grid cell, the date-specific wealth domains are constructed using 800 finite-volume cells. The Bellman value function is represented on 61 wealth nodes, together with 13 nodes for κ_{t_j} and 11 nodes for $\Delta\kappa_{t_j}$. Conditional annuity factors use a nine-point Gauss-Hermite rule, while the continuation expectation in Equation (24) uses seven return nodes and seven mortality nodes. The full-grid distributional results use $n_S = 5\,000$ evaluation paths per cell. The state-dependence diagnostics reported below use a separate sample of 50 000 paths and 20 equal-count state bins at each decision age. Table 26 summarizes the numerical design.

Table 26: Numerical design for the dynamic annuitization problem

Quantity	Baseline specification
Country, valuation date, issue age	France, 2023, age 65
Sex	Female
Limiting age	$x^+ = 95$
Mortality model	LC Poisson
Mortality forecast	ARIMA(1,1,0) with linear trend
Calibration window	1950–2023 with ages 50–95
Period-factor lookback	15 years
Initial wealth	$W_0 = 100$
Pricing discount rate	$\varrho = 2\%$
Objective discount rate	$\tilde{\varrho} = 2\%$
Withdrawal rate	$\alpha = 4\%$
Loading factor	$\mathcal{L} = 5\%$
Bequest weight	$\psi \in \{0, 0.25, 0.5, 0.75, 1\}$
Investment return	$\mu \in \{2\%, 3\%, \ldots, 8\%\}$, $\sigma \in \{0\%, 2\%, \ldots, 12\%, 15\%\}$
Baseline specification	$\mu = 4\%$, $\sigma = 6\%$, $\psi = 0.75$
Wealth-domain construction	800 finite-volume cells
Bellman state mesh	$61 \times 13 \times 11$
Gauss-Hermite orders	9 for annuity factors, 7×7 for continuation
Full-grid evaluation paths	$n_S = 5\,000$ per cell
State-dependence evaluation paths	50 000
State-dependence bins	20 equal-count bins per age

In this section, we analyze the dynamic stopping problem by comparing three initial objective values: V_0^{now}, obtained under immediate annuitization (the fix-now strategy), V_0^{fix}, obtained under the optimal fixed-date (flex-to-fix) strategy from the previous section[67], and V_0^{dyn} obtained under the fully dynamic optimal-stopping strategy. These three values are weakly ordered by construction: $V_0^{\text{dyn}} \geq V_0^{\text{fix}} \geq V_0^{\text{now}}$, since the dynamic policy encompasses every fixed-date rule and the optimal fixed-date encompasses the special case $n^\star = 0$. We compute two relative-value measures: $\Delta V_{\text{fix/now}} = \left(V_0^{\text{fix}}/V_0^{\text{now}} - 1\right)$ and $\Delta V_{\text{dyn/fix}} = \left(V_0^{\text{dyn}}/V_0^{\text{fix}} - 1\right)$. The first quantity measures the value of choosing the optimal fixed conversion date instead of annuitizing immediately. The second quantity measures the additional value of conditioning the exact conversion date on the realized state, rather than committing to it in advance.

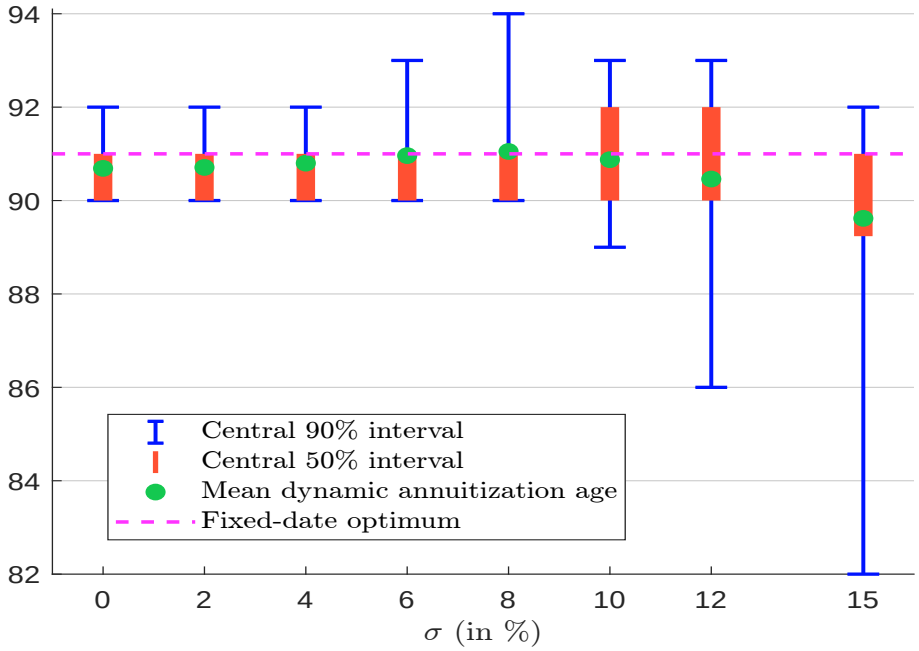
[67]The selected fixed-date strategy is re-evaluated using the same state transition, quadrature, interpolation, and terminal convention as the dynamic strategy, so that the three values are directly comparable.

Table 27: Fixed-date and dynamic annuitization results (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

μ (%)	σ (%)	$x + n^\star_{\text{fix}}$	$\mathbb{E}\left[x + n^\star \mid n^\star < \infty\right]$	$\Pr\{n^\star < \infty\}$ (%)	$\Delta V_{\text{fix/now}}$ (%)	$\Delta V_{\text{dyn/fix}}$ (%)
2	0	75	74.71	100.0	0.43	0.0086
3	2	87	86.99	100.0	12.68	0.0159
4	6	91	90.96	100.0	30.06	0.0244
5	10	94	93.28	100.0	53.38	0.0391
6	15	94	93.10	100.0	84.56	0.0178

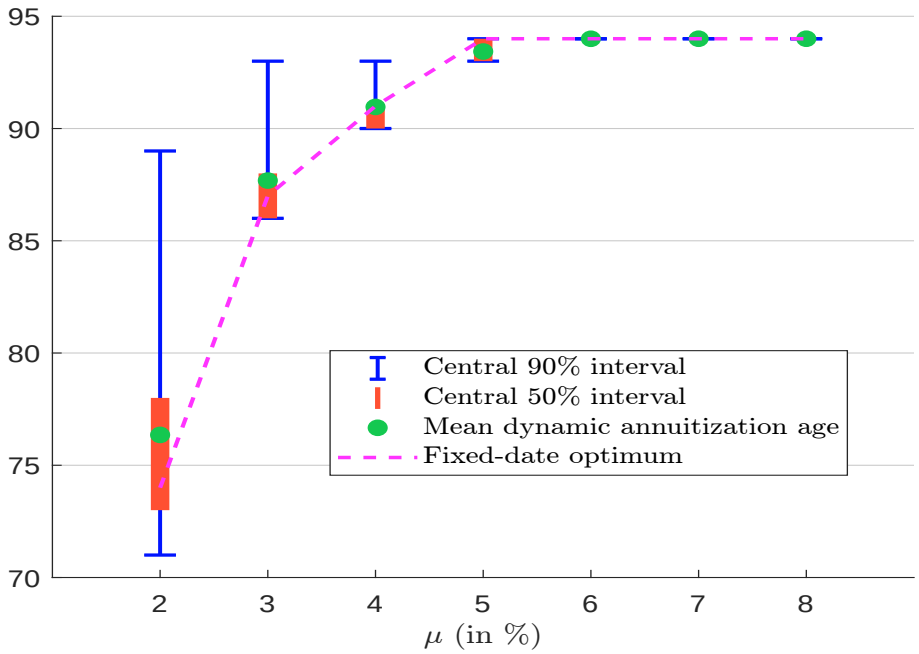
Table 27 reports results for the five representative investment environments used in the fixed-date analysis, including the probability of finite annuitization and the mean stopping age conditional on annuitization together with the two value-gain measures $\Delta V_{\text{fix/now}}$ and $\Delta V_{\text{dyn/fix}}$. The economically important contrast in Table 27 is between these two value gains. Moving from fix-now to an appropriate fixed annuitization horizon generates substantial gains that grow with expected returns across the representative investment environments. By contrast, allowing the exact stopping date to depend on the realized state adds only a few basis points in these cases. At the baseline $(4\%, 6\%)$ calibration, for example, the selected fixed-date rule improves the objective by 30.06% relative to fix-now[68], whereas dynamic timing adds a further 0.0244% relative to the selected fixed rule. The main value improvement therefore comes from choosing the broad annuitization horizon, rather than from adjusting the exact conversion year after observing subsequent states.

Figure 49: Location and dispersion of the dynamic annuitization age relative to the fixed-date optimum (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$, $\mu = 4\%$)



[68]The fixed-date gains in Table 27 differ slightly from the corresponding Monte Carlo estimates in the fixed-date section, because the selected fixed policy is re-evaluated here using the same backward operator as the dynamic policy. This common evaluation makes $\Delta V_{\text{dyn/fix}}$ a comparison between strategies under the same numerical operator.

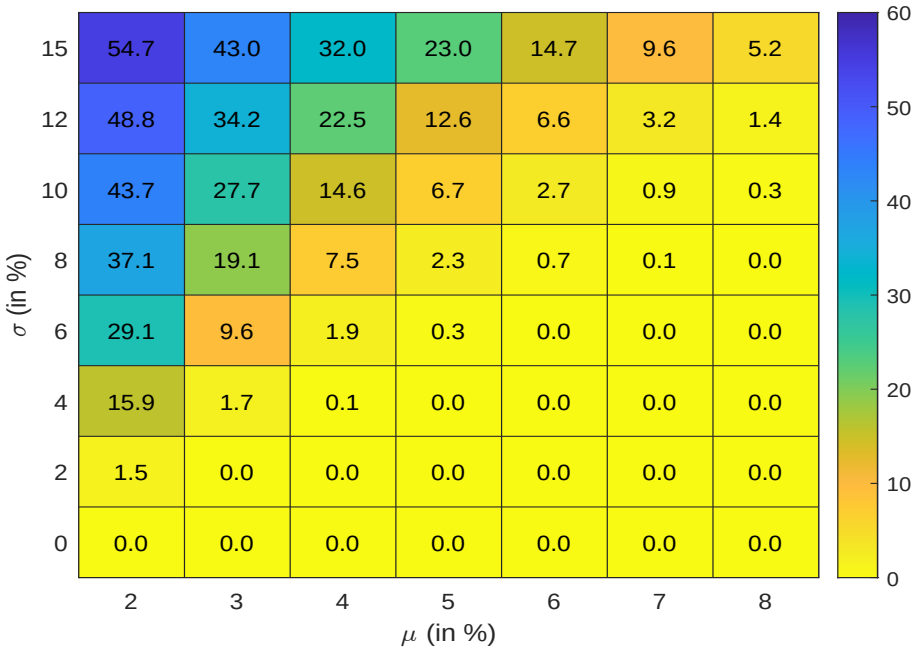
Figure 50: Location and dispersion of the dynamic annuitization age relative to the fixed-date optimum (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$, $\sigma = 6\%$)



The small incremental value does not mean that the dynamic policy selects a fixed conversion age for every path. Figures 49 and 50 report the distribution of dynamic annuitization ages across two slices of the baseline return specification. In all configurations displayed in the figures, $\Pr\{n^\star < \infty\} = 1$. The green circle indicates the mean dynamic stopping age. The thick red vertical segment contains the central 50% of stopping ages. The thin blue segment contains the central 90%. The dashed violet line shows $x + n^\star_{\text{fix}}$. Figure 49 shows that volatility has little effect on the location of the fixed-date optimum when $\mu = 4\%$. The selected fixed conversion age remains at 91 throughout the volatility grid, and the mean dynamic age remains close to 91 at moderate volatility. The main change occurs in the dispersion of the stopping date. The central 90% interval is $[90, 92]$ when $\sigma = 0\%$, compared to $[82, 92]$ when $\sigma = 15\%$. At the latter volatility, however, the central 50% interval remains $[90, 91]$. Thus, the increased dispersion at high volatility comes mainly from a lower tail of paths on which substantially earlier annuitization becomes optimal rather than from a broad widening of the whole distribution. Figure 50 shows the stronger effect of the expected return on the location of the stopping decision. With σ fixed at 6%, the fixed conversion age rises from 74 at $\mu = 2\%$ to 87, 91, and 94 at $\mu = 3\%$, 4%, and 5%, respectively. The mean dynamic age follows the same progression. At $\mu = 2\%$, the central 50% interval is $[73, 78]$ and the central 90% interval is $[71, 89]$. At the baseline $\mu = 4\%$, they contract to $[90, 91]$ and $[90, 93]$, respectively. When $\mu \geq 6\%$, the dynamic stopping age essentially converges to age 94 across the remaining three grid points. This convergence reflects the finite modeling horizon since age 94 is the final admissible conversion age before $x^+ = 95$. Once the expected return is high enough that delay is almost always preferred, the terminal boundary itself becomes the binding constraint on the stopping decision. Over these slices, μ primarily shifts the location of the stopping region, while σ has a stronger effect on the dispersion of state-contingent stopping dates around that region.

Remark 19 *The preceding comparison concerns configurations for which $n^{\star}_{\text{fix}}$ is finite. Figure 51 considers the special case in which the optimal predetermined strategy is fix-never. At $\psi = 1$, the fixed-date analysis selects fix-never in all 56 return configurations. However, under the dynamic strategy, finite annuitization still occurs in some realized states. This shows that fix-never is not the optimal action along every path that can subsequently be reached. The highest frequency of finite annuitization occurs at $(\mu, \sigma) = (2\%, 15\%)$, where 54.7% of the evaluation paths eventually annuitize. Nevertheless, the corresponding dynamic gain relative to the fixed fix-never strategy is only 0.0891%. Therefore, fix-never is a quasi-optimal ex ante commitment rule in these cells*[69].

Figure 51: Probability of finite dynamic annuitization when the optimal fixed-date strategy is fix-never (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 1.00$)



We examine the state dependence of the policy at the baseline calibration: $(\mu, \sigma, \psi) = (4\%, 6\%, 0.75)$. For a fixed mortality state, the wealth dimension of the stopping boundary can be identified by the sign of $V^{\text{annuitization}}(X_{t_j}) - V^{\text{continuation}}(X_{t_j})$, which is positive in the annuitization region and negative in the continuation region. Table 28 reports the wealth level at which this difference changes sign for each age when $(\kappa_{t_j}, \Delta\kappa_{t_j})$ is held at the nearest mesh point to its date-specific conditional mean. The switching boundary falls sharply over the main annuitization ages. At age 89, annuitization at the central mortality state requires wealth above approximately 3.6 times the initial wealth. One year later, the threshold is close to 1.6 times the initial wealth. By ages 91 and 92, the threshold falls well below initial wealth. The concentration of dynamic stopping around ages 90–92 is therefore associated with a rapid expansion of the annuitization region across the wealth dimension[70]. This wealth dependence is consistent with the presence of the fixed withdrawal $w_0 = \alpha W_0$ in the continuation branch. Since the scheduled withdrawal is fixed relative to initial wealth

[69] However, this does not imply that annuitization is a dominated action in every state that can subsequently be reached.

[70] One factor is that the maximum age has been fixed at $x^{+} = 95$ years.

rather than current wealth, its weight in $V^{\text{continuation}}\left(X_{t_j}\right)$ decreases as current wealth X_{t_j} increases. As a result, $V^{\text{continuation}}$ does not scale proportionally with current wealth. This type of non-homogeneity produces a wealth-dependent switching boundary.

Table 28: Approximate wealth boundary of the annuitization region at the central mortality state under the baseline dynamic calibration (France, female, 2023, LC Poisson, $\mu = 4\%$, $\sigma = 6\%$, $\alpha = 4\%$, $\psi = 0.75$)

Decision age	Switching interval for W_{t_j}/W_0	
	Lower bound	Upper bound
89	3.54	3.61
90	1.59	1.65
91	0.345	0.414
92	0.216	0.288

Figures 52 and 53 evaluate the stopping rule along 50 000 simulated baseline paths. At each decision age, the states of individuals who are both alive and liquid are separated into 20 bins of equal count. The solid lines show the fraction of states in each bin that Equation (25) prescribes annuitization, and the shaded regions are 95% Wilson confidence intervals. Figure 52 confirms the wealth dependence observed in Table 28. At age 89, annuitization remains rare across the realized wealth distribution, with an estimated probability of only 8.3%, even in the highest wealth bin. At age 90, the probability increases from essentially zero in the lowest bins to 55.3% in the highest bin. At age 91, the range is approximately 0.1% to 95.7%. By age 92 the highest wealth bins annuitize with essentially a probability of 100%. Therefore, the simulated states display the same rapid expansion of the annuitization region across wealth as the Bellman boundary. Figure 53 shows that the decision also varies with the Lee-Carter mortality state. At age 89, annuitization is close to zero over most of the observed κ_{t_j} distribution, reaching approximately 41.1% in the highest bin. At age 90, it rises to 85.8% in the highest bin. At age 91, the probability is approximately 86.2% for some of the intermediate-to-high κ_{t_j} range. However, the relation is not globally monotone. For example, at age 92, the estimated probability declines from approximately 81.3% in the lowest observed κ_{t_j} region to 36.5% in the highest bin. This pattern should be interpreted jointly with the other state variables rather than in isolation. Part of this joint dependence has a simple explanation. Since $\varrho_{t_0} = \tilde{\varrho}_{t_0}$, the two conditional annuity factors in Equation (23) coincide, and we have $V^{\text{annuitization}}\left(X_{t_j}\right) = W_{t_j}/\left(1+\mathcal{L}\right)$. Conditional on current wealth, $V^{\text{annuitization}}$ does not depend on the mortality state at all. Therefore, any residual dependence of the stopping decision on κ_{t_j}, conditional on wealth, must operate entirely through $V^{\text{continuation}}\left(X_{t_j}\right)$. In addition, the curves in Figure 53 are marginal bin averages over the population that remains alive and liquid at each age. Therefore, they integrate over different realized values of W_{t_j} and $\Delta\kappa_{t_j}$. The figure shows that the stopping boundary depends on the mortality state, but the binned curves should not be interpreted as a one-dimensional monotonic effect of κ_{t_j} alone.

Figures 54 and 55 examine the incremental value of dynamic timing over the return grid at the baseline bequest weight $\psi = 0.75$. Figure 54 shows $\Delta V_{\text{dyn/fix}}$ as asset volatility increases. The gain is non-decreasing in σ for each expected-return level in the full grid, and the effect is largest at low expected returns. For example, when $\mu = 2\%$, the gain increases from 0.0086% at $\sigma = 0\%$ to 0.3309% at $\sigma = 15\%$. At higher expected returns, however, the selected fixed conversion date is close to the upper boundary of the admissible horizon, which limits the scope for further timing adjustments. Even at its largest, however, the magnitude remains modest relative to $\Delta V_{\text{fix/now}}$, and it stays small over most of the parameter space.

Figure 52: Conditional probability of dynamic annuitization by wealth states at selected ages (France, female, 2023, LC Poisson, $\mu = 4\%$, $\sigma = 6\%$, $\alpha = 4\%$, $\psi = 0.75$)

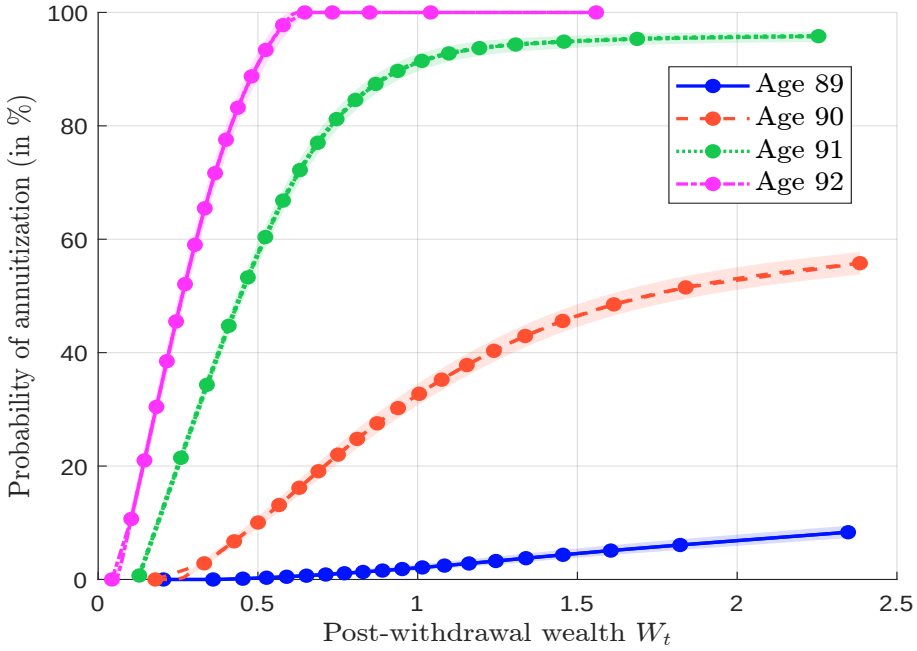


Figure 53: Conditional probability of dynamic annuitization by mortality states at selected ages (France, female, 2023, LC Poisson, $\mu = 4\%$, $\sigma = 6\%$, $\alpha = 4\%$, $\psi = 0.75$)

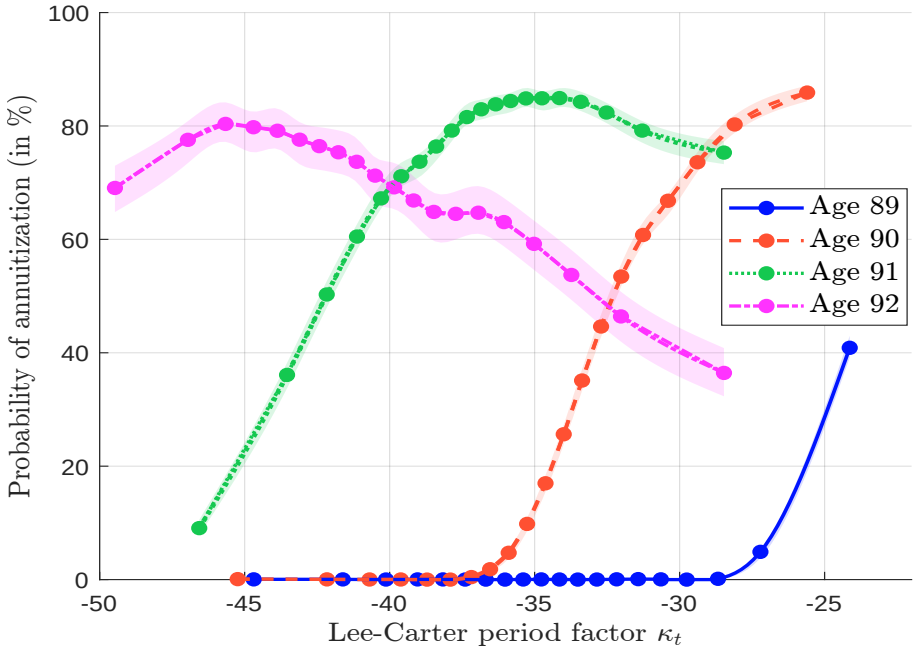


Figure 54: Volatility and value of timing flexibility (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

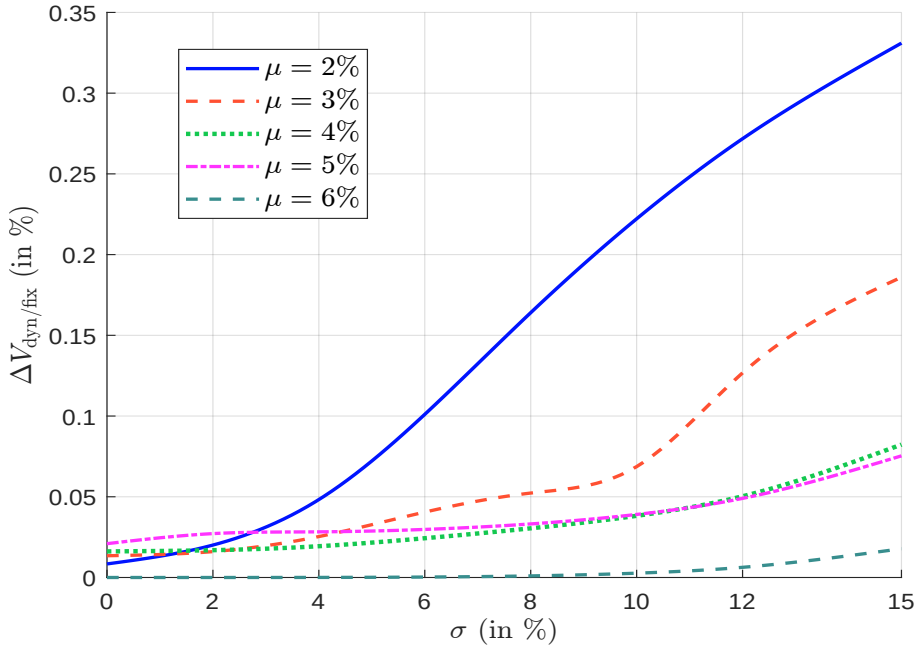
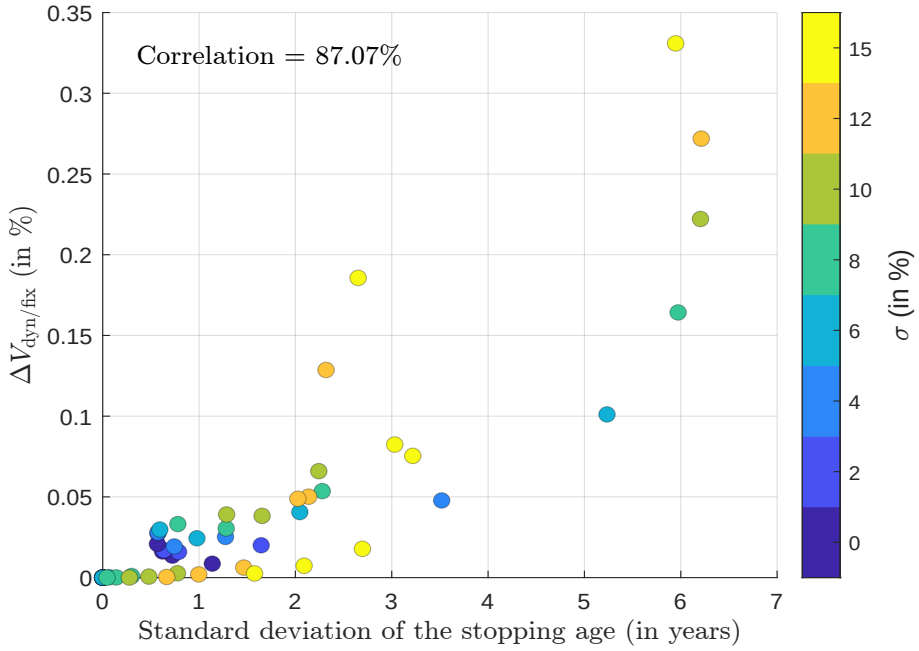


Figure 55: Dispersion of the dynamic stopping age and value of timing flexibility (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)



Across the 56 return configurations at $\psi = 0.75$, the median value of $\Delta V_{\text{dyn/fix}}$ is 0.0174%, the 75$^{\text{th}}$ percentile is 0.0424%, and 90% of the cells have a gain below 0.1149%. Figure 55 relates the dynamic gain to the standard deviation of the finite stopping age. The cross-sectional correlation is approximately 87%. Configurations with more dispersed stopping dates tend to have larger gains from state-dependent timing. However, this association is descriptive rather than causal since both stopping-age dispersion and the value gain are endogenous outcomes of the same underlying parameter configuration and stopping problem.

Table 29 examines the bequest-weight dimension while holding the return specification at baseline values $(\mu, \sigma) = (4\%, 6\%)$. As ψ increases, the fixed-date optimum moves progressively outward from age 80 when $\psi = 0$ to age 91 under the baseline value $\psi = 0.75$ and to age 94 when $\psi = 0.90$. The incremental value of dynamic timing decreases across these configurations, ranging from approximately 0.056% at low bequest weights to nearly zero at the upper end of the grid. Even though the fixed benchmark is already at the last admissible conversion age, 98.12% of dynamic paths still annuitize at $\psi = 0.90$. By contrast, at $\psi = 1$, the fixed-date strategy switches discretely to fix-never, and only 1.88% of dynamic paths annuitize under the baseline return specification. As shown in Figure 51, this last probability depends highly on the investment environment and can be substantially larger when expected returns are low and volatility is high.

Table 29: Bequest weight sensitivity of the fixed benchmark and dynamic timing value at $(\mu, \sigma) = (4\%, 6\%)$ (France, female, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

ψ	$x + n^{\star}_{\text{fix}}$	$\Pr\{n^{\star} < \infty\}$ (%)	$\mathbb{E}\left[x + n^{\star} \mid n^{\star} < \infty\right]$	$\Delta V_{\text{dyn/fix}}$ (%)
0.00	80	100.00	79.91	0.0558
0.25	82	100.00	82.11	0.0552
0.50	85	100.00	85.24	0.0429
0.75	91	100.00	90.96	0.0244
0.90	94	98.12	93.96	0.0002
1.00	Never	1.88	91.93	0.0000

Taken together, the dynamic analysis adds one key qualification to the fixed-date results. While a well-chosen predetermined annuitization horizon captures most of the expected objective value from delayed conversion, it does not reproduce the state structure of the optimal stopping policy. Realized wealth and mortality states can move individual paths to enter the annuitization region at significantly different ages, and annuitization can occur *ex post* even when the optimal predetermined rule is fix-never. Nevertheless, the additional expected value of this state-contingent timing is small over most of the parameter grid.

4.3.2 Optimal asset allocation with a target date

The mathematical problem In this section, we adopt the second approach within a singular stochastic control framework. We assume that the investor allocates wealth[71] X_t between a risky asset S_t, a risk-free bond B_t, and a continuous annuity income Y_t. The risky asset follows a geometric Brownian motion with expected return μ and volatility σ:

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t$$

where W_t is a standard Brownian motion. The risk-free asset earns a constant interest rate r: $\mathrm{d}B_t = rB_t \,\mathrm{d}t$. We assume that a_t is the capital cost required to purchase one unit of

[71] We use the notation X_t for the wealth because W_t is reserved to designate a Brownian motion in continuous time.

continuous annuity income at time t and that $Z = \{Z_t\}_{t\in[t_0,T)}$ is the cumulative amount spent on purchasing annuities up to time t:

$$\mathrm{d}Y_t = \frac{1}{\breve{a}_t}\mathrm{d}Z_t$$

where $\breve{a}_t = (1+\mathcal{L})\, a_t$. Let X_{t^-} and Y_{t^-} denote the state variables immediately before any potential lump-sum transaction at time t. We suppose that the investor enters the strategy with an exogenous stream of mandatory pension income $\bar{y} > 0$ and since Z is non-decreasing, we have $Y_t \geq \bar{y}$. Terminal annuitization at time T, whereby all remaining wealth is used to purchase an annuity, is not part of the diffusion dynamics on $[t_0, T)$, but instead enters as a terminal condition. It induces a jump in the annuity income at the target date:

$$Y_T = Y_{T^-} + \frac{X_{T^-}}{\breve{a}_T} \qquad \text{and} \qquad X_T = 0$$

The control strategy consists of the pair (α_t, Z_t), where α_t is the proportion of liquid wealth invested in the risky asset. Let $\mathcal{A}(t,x,y)$ denote the set of admissible control pairs (α, Z) starting from time t with initial liquid wealth $X_{t^-} = x$ and initial annuity income $Y_{t^-} = y$. Specifically, any cumulative annuity expenditure process Z is modeled as an adapted, nondecreasing, and right-continuous process. This structure reflects the fact that annuity contracts can be purchased with instantaneous lump sums ($\Delta Z_t > 0$) but they cannot be sold or converted back into cash. Under this singular stochastic control framework, the investor's wealth process X_t evolves according to the following stochastic differential equation:

$$\mathrm{d}X_t = \Big(\big(r + \alpha_t(\mu - r)\big)\, X_{t^-} - \boldsymbol{c}_t + Y_{t^-}\Big)\,\mathrm{d}t + \alpha_t X_{t^-}\sigma \mathrm{d}W_t - \mathrm{d}Z_t \tag{28}$$

where $X_t \geq 0$, $Y_t \geq \bar{y}$, $\alpha \in [0,1]$, $Z_t \geq 0$ and $\boldsymbol{c}_t \geq 0$ is the predetermined deterministic consumption flow.

In what follows, we focus on the expected utility maximization framework, and we denote by $\mathcal{U}(x)$ the utility function. To evaluate the optimality of the controls (α_t, Z_t), we must define an objective function that captures the investor's lifetime utility. Our model incorporates three distinct scenarios based on mortality and investment risks:

- Death at s before the target date T: the investor leaves behind a lump sum of liquid wealth X_{s^-}.
- Ruin at $\boldsymbol{\tau}$ before the target date T: the investor depletes all liquid wealth, and the existing annuity income can no longer cover the predetermined consumption target. Following this failure, the investor permanently abandons the original consumption schedule and consumes the entire existing annuity income for the rest of their life.
- Survival to the target date T: All remaining liquid wealth is irreversibly converted into an annuity, ensuring a steady stream of income for the rest of the investor's life.

Before stating the objective function, we define the following key parameters: λ_t is the instantaneous hazard rate[72] at time t, ϱ is the investor's subjective discount rate[73] (time preference), w is a weighting parameter that scales the utility of bequest relative to consumption, and b is an exogenous subsistence bequest floor. Since the consumption target

[72] The probability of an individual surviving from time t to s is given by $\exp\left(-\int_t^s \lambda_u \,\mathrm{d}u\right)$.

[73] We assume that $\tilde{\varrho} = \varrho$.

is not guaranteed to be met along every investment path, we define the consumption-plan failure time as follows:

$$\boldsymbol{\tau} = \inf\left\{s \in [t,T) : X_s = 0, Y_s < \boldsymbol{c}_s\right\}$$

For notation simplicity, we define the following factors: $D(t,s) = \exp\left(-\int_t^s (\varrho + \lambda_u)\,\mathrm{d}u\right)$ is the effective discount factor, $K_t = \int_t^\infty D(t,s)\,\mathrm{d}s$ is the utility annuity factor, and $\mathfrak{B}(t) = \int_t^T D(t,s)\,\lambda_s \mathcal{U}(wb)\,\mathrm{d}s$ captures the bequest utility at the subsistence floor wb. If $\boldsymbol{\tau} < T$, the investor irrevocably enters a prescribed fallback arrangement. The original consumption schedule is permanently abandoned, no further investment or annuity purchase decisions are made, liquid wealth remains at zero, and the entire existing annuity income $Y_{\boldsymbol{\tau}}$ is consumed for the remainder of life. The fallback value, conditional on being alive at time t and receiving annuity income y, is therefore $G(t,y) = K_t \mathcal{U}(y) + \mathfrak{B}(t)$, where the first term captures the value of consuming the existing annuity income over the remaining lifetime, and the second term captures the bequest utility associated with death prior to the target date T.

Let V denote the full value function. The investor maximizes the expected discounted utility of three payoffs over the admissible controls $(\alpha, Z) \in \mathcal{A}(t,x,y)$:

$$\begin{aligned} V(t,x,y) \quad = \quad & \sup_{(\alpha,Z)\in\mathcal{A}(t,x,y)} \mathbb{E}\Bigg[\underbrace{\int_t^{\boldsymbol{\tau}\wedge T} D(t,s)\left(\mathcal{U}(\boldsymbol{c}_s) + \lambda_s \mathcal{U}\left(\xi\left(b + X_{s^-}\right)\right)\right)\mathrm{d}s}_{\text{Consumption and mortality-weighted bequest before } T} + \\ & \underbrace{\mathbb{1}\{\boldsymbol{\tau} < T\}\, D(t,\boldsymbol{\tau})\, G(\boldsymbol{\tau}, Y_{\boldsymbol{\tau}})}_{\text{Ruin before } T} + \\ & \underbrace{\mathbb{1}\{\boldsymbol{\tau} \geq T\}\, D(t,T)\, K_T \mathcal{U}\left(Y_{T^-} + \frac{X_{T^-}}{\breve{a}_T}\right)}_{\text{Survive until time } T \text{ and annuitize the entire amount}} \Bigg] \end{aligned}$$

where $\mathcal{U}(x) = \dfrac{x^{1-\eta}}{1-\eta}$ $(\eta \geq 0, \eta \neq 1)$ is the constant relative risk aversion (CRRA) utility function. The introduction of the exogenous bequest floor $b > 0$ carries economic and empirical significance for retirement wealth modeling:

- Bequest as a luxury good
  In the classical CRRA framework, leaving zero bequest yields infinite negative utility ($\mathcal{U}(0) = -\infty$). Economically, this pathological property forces bequests to act as an extreme necessity, implying that an impoverished retiree would rather starve than leave no financial wealth to their heirs. By adopting the utility function $\mathcal{U}\left(\xi\left(b + x\right)\right)$, the utility of leaving no liquid wealth remains finite at $\mathcal{U}(\xi b)$. As shown by De Nardi (2004) and Lockwood (2018), this formulation mathematically recasts bequests as luxury goods. Impoverished retirees prioritize personal subsistence and annuitization, whereas wealthier individuals exhibit a strong propensity to leave bequests, consistent with empirical patterns of wealth decumulation.

- Restoring rational portfolio and annuitization behavior
  Without the floor b, the unbounded marginal utility of bequests as $x \to 0^+$ creates an artificial barrier against liquidity depletion ($\mathcal{U}'(\xi x) \to \infty$). This numerical artifact induces pathological risk aversion, preventing retirees from making optimal lump-sum annuity purchases out of fear of momentarily reaching zero liquid wealth. The subsistence floor stabilizes marginal incentives, permitting rational, uninhibited participation in the annuity market.

The Hamilton-Jacobi-Bellman variational inequality Solving the optimization problem is relatively tricky. Appendix C.13 on page 152 describes the different steps to obtain the solution. In particular, we show that the state space is partitioned by two complementary regimes:

1. a waiting or continuation region, with no annuitization;
2. an annuitization or action region, with a lump-sum annuity purchase.

In this context, the solution is given by the following HJB variational inequality[74]:

$$\max\left\{V_t+\sup_{\alpha\in[0,1]}\left\{\mathcal{L}^{\alpha_t}V\right\}-(\varrho+\lambda_t)V(t,x,y)+\lambda_t\mathcal{U}\left(\xi(b+x)\right)+\mathcal{U}(\boldsymbol{c}_t),\frac{1}{\breve{a}_t}V_y-V_x\right\}=0$$

subject to the terminal annuitization condition at the target date T:

$$V(T,x,y)=\mathcal{U}\left(y+\frac{x}{\breve{a}_T}\right)\int_T^\infty e^{-\int_T^s(\varrho+\lambda_u)\,\mathrm{d}u}\,\mathrm{d}s$$

and the Dirichlet condition:

$$\begin{cases}V(t,0,y)=G(t,y)=K_t\mathcal{U}(y)+\mathfrak{B}(t)\\ y<\boldsymbol{c}_t\end{cases}$$

Numerical results In this section, we conduct numerical tests to illustrate the properties of the HJBVI dynamic model. Our focus is on how the wait/action partition of the state space evolves with age and the strength of the bequest motive. Prior to this, we estimate the instantaneous hazard rate λ_t at age t and calculate the annuity price. To avoid introducing stochastic factors into the HJBVI equation, we use the Gompertz annuity pricing model (GAPM) described in Appendix C.5 on page 143. In the numerical implementation, we replace the pure Gompertz law with the Gompertz-Makeham specification: $\lambda_t=e^{-\varphi t}\left(A+B\exp\left(c_\lambda(t_0+t)\right)\right)$, where $t_0=65$ denotes the initial age and φ is the cohort improvement factor. The model is calibrated to HMD France 2023 male data, and the annuity price is computed by numerical integration. Maximum likelihood estimation on single-year death counts and exposures for ages 50–105 yields the following values: $A=3.22\times10^{-3}$, $B=4.33\times10^{-6}$, $c_\lambda=0.1167$, and $\varphi=0.0191$ per year. The financial and preference parameters are set as follows: the risk-free rate is $r=2\%$, the risky-asset expected return is $\mu=4\%$, the volatility is $\sigma=15\%$, the relative risk aversion is $\eta=2.5$, the bequest floor is $b=0$, the consumption target is $\boldsymbol{c}_t=4\%$, the minimum annuity income is $y_0=\bar{y}=2\%$. An actuarial load $\mathcal{L}$ of 20% is applied at both the initial purchase and the terminal deadline. The retirement age is $t=65$ years, while the flex-to-fix horizon is $T=80$ years (ages 65–80). With these parameters, the unconstrained Merton portfolio allocation is $\bar{\alpha}=(\mu-r)/\left(\eta\sigma^2\right)\approx35.6\%$.

Figures 56 and 57 display the wait/action partition of the (x,y) state space at six ages, under a weak ($\xi=1$) and a strong ($\xi=0.1$) bequest motives[75]. The wait region (blue)

[74] $\mathcal{L}^{\alpha_t}$ is the second-order partial differential operator:

$$\mathcal{L}^{\alpha_t}V=\left((r+\alpha_t(\mu-r))\,x-\boldsymbol{c}_t+y\right)V_x+\frac{1}{2}\alpha_t^2\sigma^2x^2V_{xx}$$

[75] Since $\mathcal{U}\left(\xi(b+x)\right)=\xi^{1-\eta}\mathcal{U}(b+x)$ and $\eta>1$, the multiplicative weight $\xi^{1-\eta}$ is decreasing in ξ. A lower value of ξ therefore corresponds to a stronger bequest motive.

Figure 56: Regime maps in the (x, y) state space under weak bequest motive ($\xi = 1$)

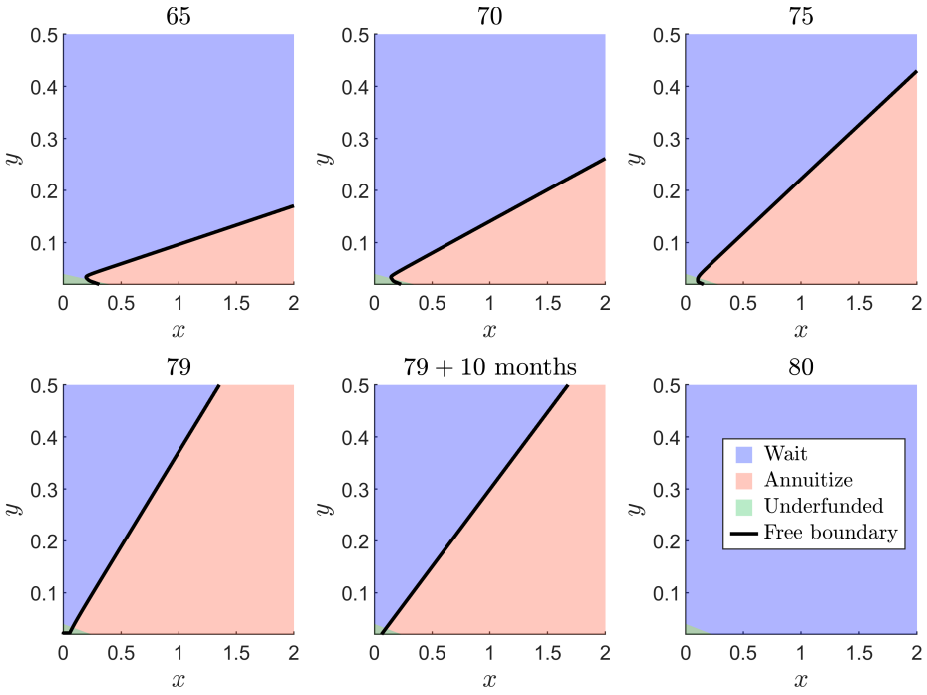
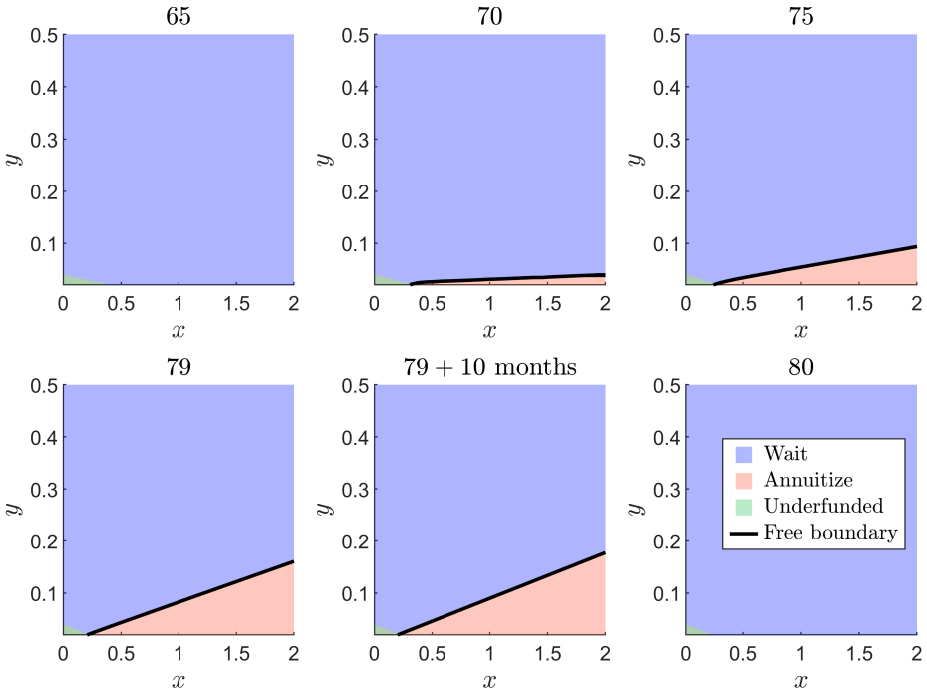


Figure 57: Regime maps in the (x, y) state space under strong bequest motive ($\xi = 0.1$)



occupies the upper left of each panel, where the gradient obstacle is slack and the retiree retains liquid wealth. The action region (red) lies in the lower right, where converting part of the liquid wealth into a life annuity income is optimal. The white region marks states with $x < \breve{a}_t \left(\boldsymbol{c} - y\right)^+$, meaning converting all liquid wealth would still leave annuity income below the consumption target. It is shaded separately for readability, but no additional constraint is imposed there, and the regime indicator is computed in the white region exactly as elsewhere. Between ages 65 and 79, the action region expands. The free boundary $x^\star\left(t, y\right)$ shifts leftward as the target date approaches, lowering the wealth threshold above which immediate annuitization becomes optimal. Over this range, the boundary is close to being linear in y, which is consistent with the value function being approximately homogeneous of degree $1-\eta$ once the bequest floor b and the present value of the consumption plan are small relative to total resources. In the final time steps, the pattern reverses and the action region contracts, as shown in the bottom panels of Figures 56 and 57. This is the region in which the obstacle is closest to degenerating. At $t = T$, the terminal condition $V\left(T, x, y\right) = K_T \mathcal{U}\left(y + x/\breve{a}_T\right)$ satisfies $\breve{a}_T^{-1} V_y - V_x \equiv 0$ identically. Thus, voluntary annuitization an instant before the deadline and mandatory annuitization at the deadline are equivalent. As $t \to T$, the distinction between the two therefore vanishes, and the retiree has nothing to gain by annuitizing early rather than waiting for the deadline mechanism to enforce conversion. In this terminal zone, the dominant strategy is to retain liquid wealth until the last possible moment. The two bequest specifications differ sharply. Under the weak motive ($\xi = 1$), a substantial action region is already present by age 65. Under the strong motive ($\xi = 0.1$), the action region is essentially absent before age 75. The marginal value of liquid wealth as a bequest asset exceeds the mortality credit over almost the entire reported window, and annuitization is deferred until the shrinking time horizon eliminates that advantage. Within the green region, the wait/action distinction survives, but its economic content changes. There, annuitization narrows the consumption gap without closing it, and the continuation strategy amounts to holding a risky portfolio in the hope of reaching the funding threshold before liquid wealth is exhausted.

Figures 58 and 59 plot the free boundary $x^\star\left(t, y\right)$ against age for four fixed annuity income levels $y \in \{0.03, 0.04, 0.05, 0.06\}$, under weak and strong bequest motives. Under the weak bequest motive ($\xi = 1$), the boundary lies within the range $[0.05, 0.50]$ throughout and is defined from age 65 onward. Even moderate liquid wealth justifies voluntary annuitization at the start of the flex window. Under the strong bequest motive ($\xi = 0.1$), the boundary is much higher, ranging from about 0.35 to over 2.0, with no boundary appearing before ages 69–72 depending on y. This means that no admissible wealth level triggers voluntary annuitization during the first few years. In both cases, $x^\star\left(t, y\right)$ declines with age as the target date approaches, which is consistent with the expansion of the action region documented in Figures 56 and 57. The decline is steeper under the strong motive. In the weak motive case, the curves gradually fall and remain close to parallel between ages 65 and 79. In the strong motive case, the decline is rapid. In both figures, $x^\star\left(t, y\right)$ turns back up in the last few time steps, which we interpret as a preference for waiting until near the time horizon T.

Figures 60, 61, and 62 plot the optimal risky asset allocation $\alpha^\star\left(t, x, y\right)$ as a function of liquid wealth x, at ages 65, 70, and 75, respectively. We consider four annuity income levels $y \in \{0.03, 0.04, 0.05, 0.06\}$ and a weak bequest motive ($\xi = 1$). The curve before the circle symbol corresponds to the wait region (continuation), while the curve after the circle symbol defines the action region (annuitization). Dashed vertical lines mark the free boundary $x^\star\left(t, y\right)$. For all income levels, $\alpha^\star\left(t, x, y\right)$ converges to the Merton solution $\bar{\alpha} = \left(\mu - r\right) / \left(\eta \sigma^2\right) \approx 35.6\%$ as $x \to \infty$. For a given x, the optimal allocation $\alpha^\star\left(t, x, y\right)$ is increasing in y. The intuition draws on the classic portfolio separation theorem. Annuity income y constitutes a riskless, lifelong cash flow that functions as an implicit bond position

Figure 58: Free boundary $x^\star(t, y)$ under weak bequest motive ($\xi = 1$)

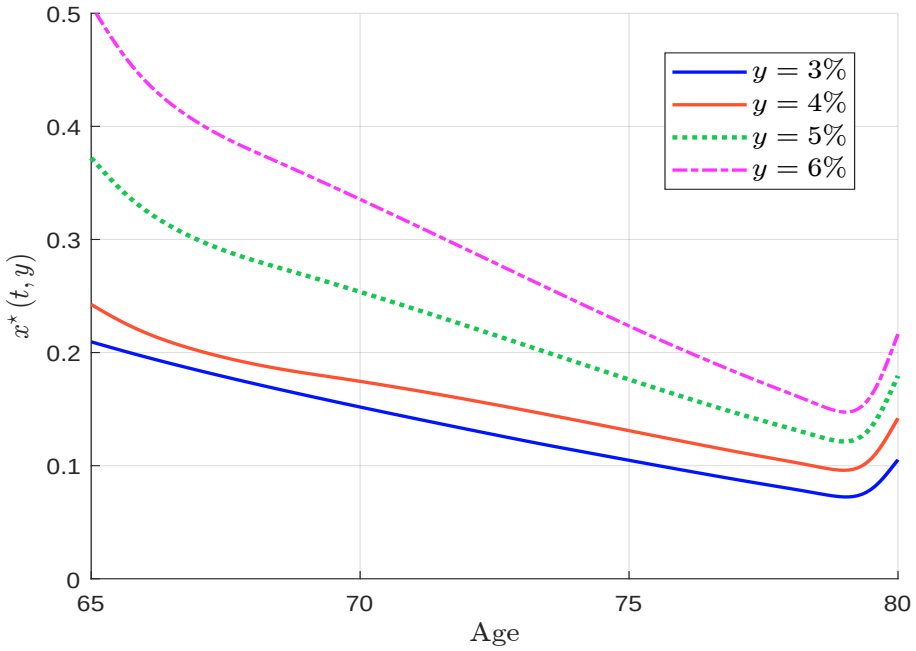


Figure 59: Free boundary $x^\star(t, y)$ under strong bequest motive ($\xi = 0.1$)

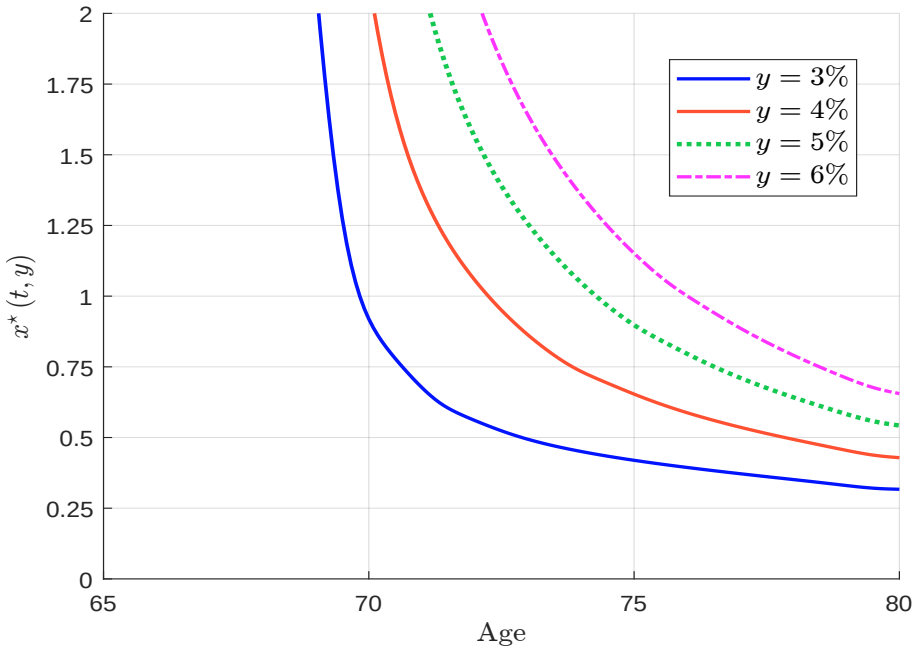


Figure 60: Optimal risky asset allocation $\alpha^\star(t,x,y)$ at age 65 ($\xi = 1$)

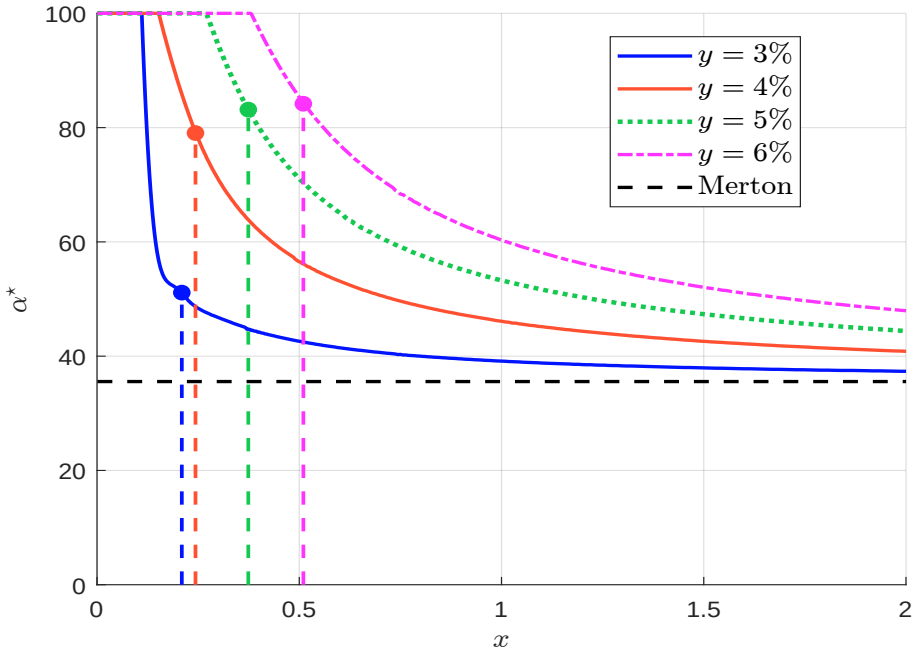


Figure 61: Optimal risky asset allocation $\alpha^\star(t,x,y)$ at age 70 ($\xi = 1$)

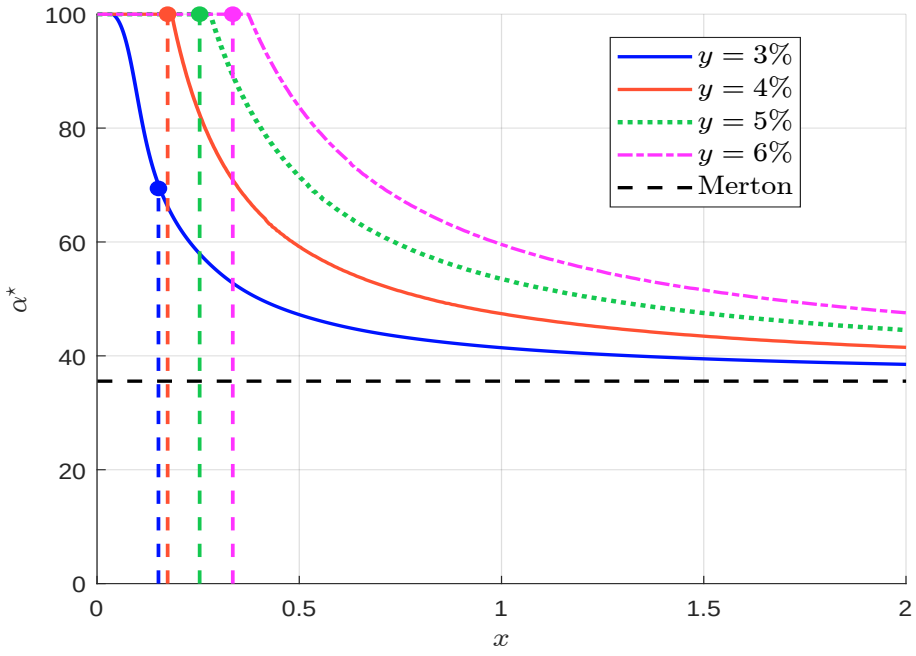


Figure 62: Optimal risky asset allocation $\alpha^\star(t,x,y)$ at age 75 ($\xi = 1$)

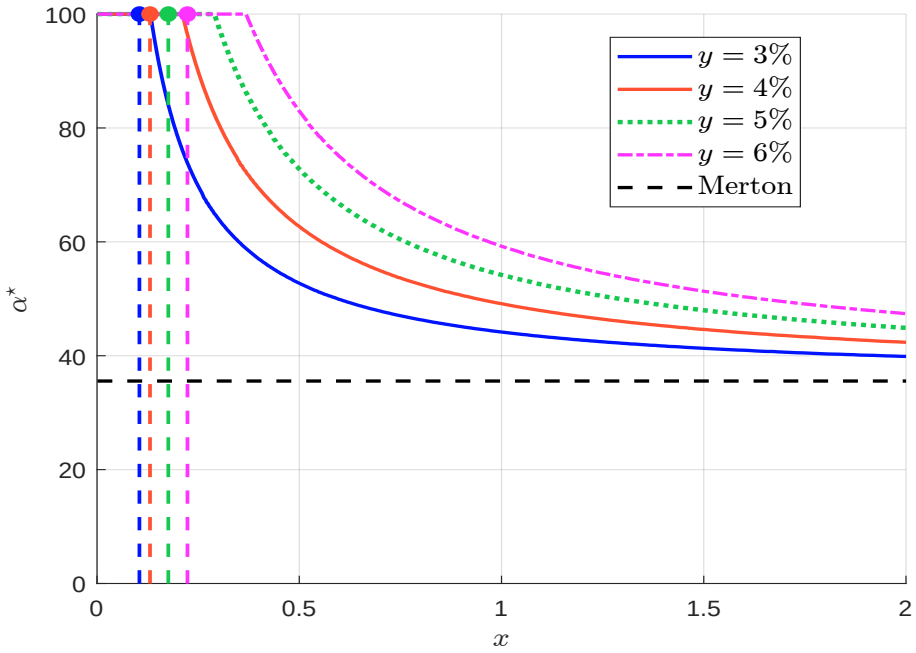
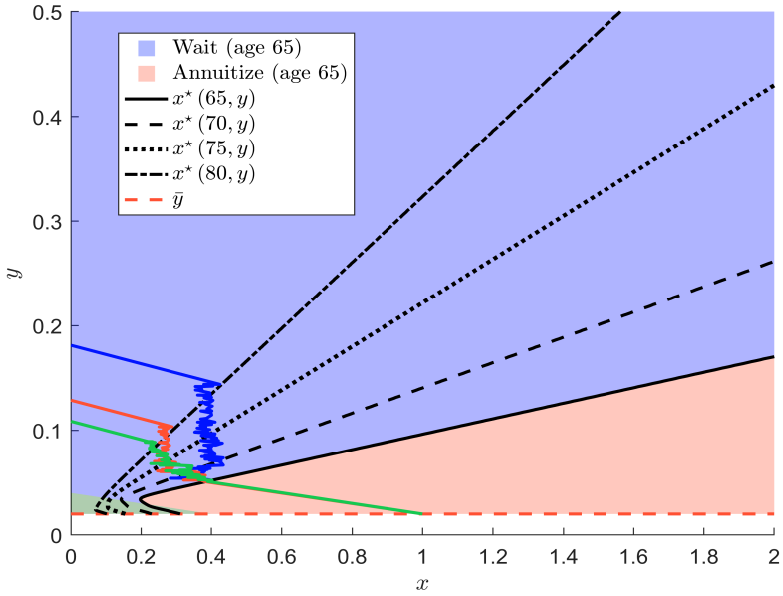
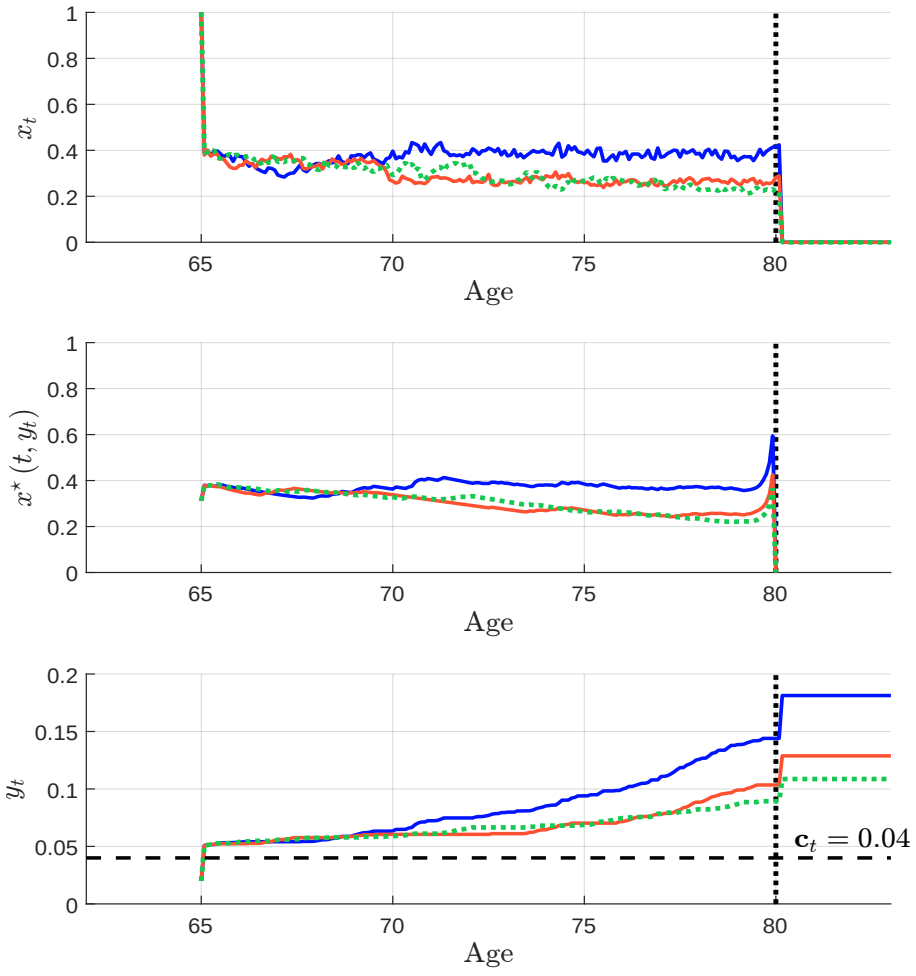


Figure 63: Phase portrait of a representative sample path under weak bequest motive ($\xi = 1$)



within the retiree's overall wealth. A retiree who already holds a large implicit bond position needs to invest less of their liquid portfolio safely to reach a given overall risk exposure. Therefore, he tilts the liquid portfolio further toward risky assets. In the case $y_t \gg \boldsymbol{c}_t$, consumption is fully covered by income, regardless of portfolio outcomes, and liquid wealth is managed almost purely for growth and bequest. This pushes $\alpha^\star(t,x,y)$ well above the Merton solution for moderate wealth x. Conversely, when $y_t < \boldsymbol{c}_t$, the consumption gap creates an effective liability that the portfolio must help service. This reduces the retiree's appetite for financial risk relative to the case where income is sufficient.

Figure 64: Time series of liquid wealth x_t, free boundary $x^\star(t,y_t)$ and annuity income y_t under weak bequest motive ($\xi = 1$) — same trajectories as Figure 63



Figures 63 and 64 illustrate the flex-to-fix strategy along three representative simulated paths starting at $x_0 = 1$ and $y_0 = \bar{y} = 2\%$ at age 65. They correspond to the blue, red and green lines, respectively. Figure 63 traces the trajectory in the (x,y) phase plane overlaid on the wait/action partition at age 65 and the free boundaries at ages 65, 70, 75, and 78. The starting point lies well inside the action region since the initial liquid wealth exceeds the threshold $x^\star(65,\bar{y})$ by a wide margin. The optimal policy therefore prescribes an immediate lump-sum purchase, moving the state along the annuitization ray $z_t \mapsto \left(x_t - z_t, y_t + z_t/\breve{a}_t\right)$

until it reaches the free boundary. Thereafter, the path evolves in the wait region, with x_t driven by portfolio returns and the net income flow $y_t - \boldsymbol{c}_t$. As the target date approaches, the free boundary shifts leftward, causing the path to repeatedly meet it. Each contact triggers a further small purchase. Thus, the state is reflected along the boundary rather than crossing it.

Figure 64 shows the time series of the three simulated paths. In the top panel, the liquid wealth x_t tracks the boundary value $x^\star(t, y_t)$ (reported in the middle panel) closely throughout the flex window. The mechanism is one-sided, since the singular control only acts when x_t reaches the boundary from above. Thus, wealth is prevented from rising far beyond the threshold, but is free to fall below it. In this case, the state simply remains in the wait region until subsequent portfolio returns restore it. At the target date $T = 80$ years, the remaining wealth is fully converted and x_t drops to zero. In the bottom panel, y_t rises in steps during the flex window, each step corresponding to one contact with the boundary. Then, at $T = 80$ years, it jumps to its terminal level y_{80}, which is well above the consumption target $\boldsymbol{c}_t = 4\%$. After $T = 80$ years, income y_t is constant, and consumption $\boldsymbol{c}_t$ is entirely financed by the annuity income stream.

Simulation results The model developed in Section 4.3.2 yields an analytically characterized free boundary, but it does not reveal how the optimal flex-to-fix policy compares with simpler, rule-based annuitization strategies in different market environments. Therefore, we complement the theoretical analysis with a Monte Carlo simulation study. Both the model and the simulation involve numerous parameters, including the risk-aversion coefficient η, the risky asset expected return μ, the volatility σ, the loading schedule $\mathcal{L}_t$, and the initial wealth x_0, among others. Each of these parameters can materially influence the quantitative outcomes. Our parameter choices reflect plausible values for a representative retiree, but we do not claim that these numerical magnitudes are universally applicable. The simulation should therefore be regarded as a stylized illustration intended to demonstrate the qualitative properties of the strategies rather than as a calibrated forecast. We consider three parameter scenarios, labeled $\mathcal{P}_1$ (high volatility), $\mathcal{P}_2$ (baseline) and $\mathcal{P}_3$ (low volatility).We

Table 30: Financial parameter scenarios

Scenario	σ	μ	$\mu - r$	η	Merton solution $\bar{\alpha}$
$\mathcal{P}_1$ (high volatility)	20%	7%	5%	4	31.3%
$\mathcal{P}_2$ (baseline)	12%	5%	3%	4	52.1%
$\mathcal{P}_3$ (low volatility)	8%	4%	2%	4	78.1%

assume a constant Sharpe ratio $(\mu - r)/\sigma = 0.25$ with a risk-free rate of $r = 2\%$ (see Table 30). The three values of σ span the range typically observed for diversified investment portfolios: 8% (defensive allocation), 12% (balanced allocation), and 20% (dynamic allocation). Initial liquid wealth is normalized to $x_0 = 1$ and the consumption target is set to $\boldsymbol{c}_t = 0.04$. The initial annuity income is[76] $y_0 \approx 0$, representing an individual who begins retirement with no pre-existing annuity. Moreover, we consider a loading schedule that rises linearly from 20% at retirement (age 65) to 30% at the deadline (age 80):

$$\mathcal{L}_t = 0.20 + 0.10\frac{t - t_0}{T - t_0}$$

where $T = 80$ years. This increasing loading reflects the tendency of insurers to price longevity risk more conservatively at older ages. The instantaneous hazard rate is cali-

[76] In fact, we set $y_0 = 0.001$ to avoid a degenerate grid endpoint.

brated to the 2023 Human Mortality Database (HMD) French male data (HMD, 2026), and simulation paths extend to a maximum age of 110 years. Beyond this age, all surviving paths are censored.

Table 31: Strategies compared in the simulation study

ID	Name	Description
$\mathcal{S}_1$	HJBVI ($\xi = 1$)	Optimal flex-to-fix policy with a weak bequest motive; invests at $\alpha^\star(t, x, y)$ and purchases annuities whenever (t, x, y) enters the action region
$\mathcal{S}_2$	HJBVI ($\xi = 0.1$)	Identical to $\mathcal{S}_1$, but with a stronger bequest motive; the stronger bequest motive raises the marginal value of liquid wealth as a bequest asset, shifting the free boundary outward and delaying annuitization
$\mathcal{S}_3$	Never	Never purchases annuities; invests at the constant Merton allocation $\bar{\alpha}$ throughout retirement
$\mathcal{S}_4$	Immediate	Converts all liquid wealth to annuity income at age 65; no further portfolio investment takes place
$\mathcal{S}_5$	Annuitize-73	Invests at $\bar{\alpha}$ from age 65 to 73, then converts all remaining liquid wealth to annuity income at age 73
$\mathcal{S}_6$	Annuitize-80	Invests at $\bar{\alpha}$ from age 65 to 80, then converts all remaining liquid wealth at the deadline (age 80)
$\mathcal{S}_7$	Annuitize-10%	Invests at $\bar{\alpha}$ from age 65 to 80 and allocates 10% of current liquid wealth each year from age 65 to 79 to purchase annuity (15 annual purchases), and converting any residual wealth at age 80

Table 31 lists seven decumulation strategies. The first two implement the HJBVI optimal policy, and the remaining five serve as rule-based benchmarks. All seven strategies use the same sample of mortality paths and Brownian increments within a given parameter scenario ($\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$). After a full conversion ($\mathcal{S}_4$, $\mathcal{S}_5$, $\mathcal{S}_6$, and $\mathcal{S}_7$), the allocation of risky assets is set to zero. However, surplus income $y_t > \boldsymbol{c}_t$ continues to accumulate as liquid wealth at a rate of $y_t - \boldsymbol{c}_t$. Consequently, the bequest $x_{\boldsymbol{\tau}}$ need not be zero, even for the immediate strategy $\mathcal{S}_4$. For each strategy and parameter scenario, we calculate the fraction P_{85} of paths ruined before age 85 (short-horizon solvency risk), the fraction P_{95} of paths ruined before age 95 (long-horizon solvency risk), the average bequest at death $\mathbb{E}[x_{\boldsymbol{\tau}}]$, the average bequest conditional on dying before age 80 $\mathbb{E}\left[x_{\boldsymbol{\tau}} \middle| \boldsymbol{\tau} < 80\right]$, and the mean annuity income $\mathbb{E}[y_{80}]$ at age 80, evaluated immediately after any conversion on that date. Tables 32–34 report these five performance metrics across the seven strategies. The point estimates are computed from $n_S = 50\,000$ simulated paths and are accompanied by 95% confidence intervals.

These results reveal several key observations, mainly around the following four themes: (1) the dominance of HJBVI strategies in preventing ruin, (2) the invariance of the Sharpe ratio and sensitivity to the HJBVI regime, (3) annuity income at age 80, and (4) the performance of rule-based strategies, with $\mathcal{S}_7$ being the best benchmark. We discuss these themes below.

Tables 32–34 show that $\mathcal{S}_1$ records zero ruin for every scenario and threshold. By dynamically tracking the HJB free boundary, the optimal strategy triggers annuity purchases whenever the state variables enter the action region. This provides self-insurance against longevity and market risk. The bequest-oriented variant $\mathcal{S}_2$ relaxes this protection marginally, with a ruin probability of 0.2% in $\mathcal{P}_1$ and $\mathcal{P}_2$, and 0.1% in $\mathcal{P}_3$, in exchange for a substantially

Table 32: Simulation results — $\mathcal{P}_1$ ($\sigma = 20\%$, $\mu = 7\%$, $\bar{\alpha} = 31.3\%$)

Strategy	P_{85} (%)	P_{95} (%)	$\mathbb{E}[x_{\boldsymbol{\tau}}]$	$\mathbb{E}\left[x_{\boldsymbol{\tau}} \middle\vert \boldsymbol{\tau} < 80\right]$	$\mathbb{E}[y_{80}]$ (%)
$\mathcal{S}_1$ – $\xi = 1$	0.0 (±0.00)	0.0 (±0.00)	0.434 (±0.003)	0.057 (±0.001)	8.1 (±0.02)
$\mathcal{S}_2$ – $\xi = 0.1$	0.2 (±0.04)	0.2 (±0.04)	0.490 (±0.003)	0.149 (±0.002)	7.7 (±0.02)
$\mathcal{S}_3$ – Never	0.0 (±0.00)	1.4 (±0.10)	0.886 (±0.004)	0.269 (±0.004)	0.1 (±0.00)
$\mathcal{S}_4$ – Immediate	0.0 (±0.00)	0.0 (±0.00)	0.316 (±0.001)	0.031 (±0.001)	5.1 (±0.00)
$\mathcal{S}_5$ – Annuitize-73	2.1 (±0.13)	2.1 (±0.13)	0.478 (±0.003)	0.129 (±0.003)	6.2 (±0.01)
$\mathcal{S}_6$ – Annuitize-80	3.6 (±0.16)	3.6 (±0.16)	0.607 (±0.004)	0.269 (±0.004)	7.7 (±0.03)
$\mathcal{S}_7$ – Annuitize-10%	0.2 (±0.04)	0.2 (±0.04)	0.440 (±0.003)	0.127 (±0.002)	7.4 (±0.02)

Table 33: Simulation results — $\mathcal{P}_2$ ($\sigma = 12\%$, $\mu = 5\%$, $\bar{\alpha} = 52.1\%$)

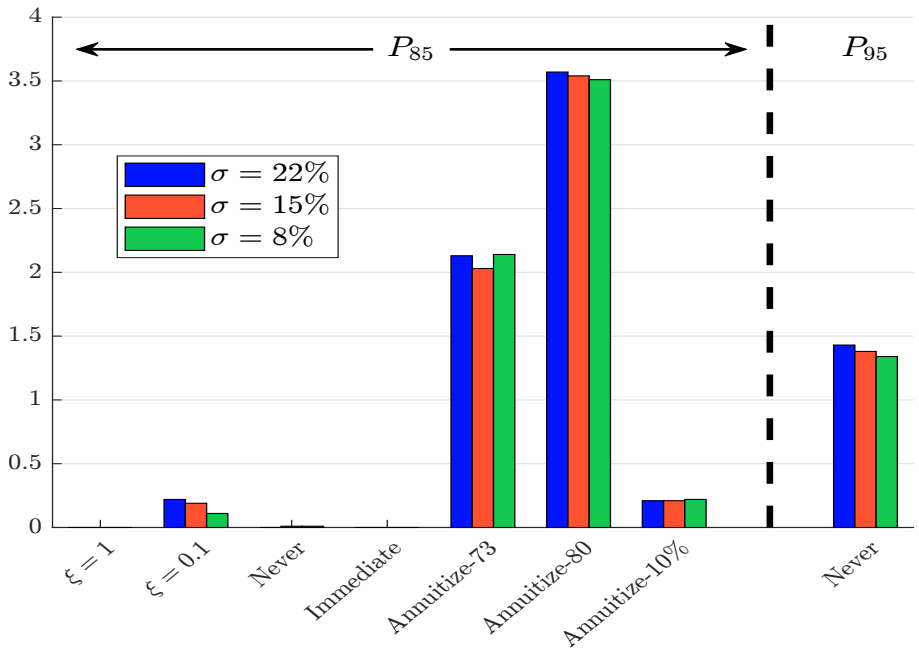
Strategy	P_{85} (%)	P_{95} (%)	$\mathbb{E}[x_{\boldsymbol{\tau}}]$	$\mathbb{E}\left[x_{\boldsymbol{\tau}} \middle\vert \boldsymbol{\tau} < 80\right]$	$\mathbb{E}[y_{80}]$ (%)
$\mathcal{S}_1$ – $\xi = 1$	0.0 (±0.00)	0.0 (±0.00)	0.420 (±0.003)	0.073 (±0.001)	7.7 (±0.01)
$\mathcal{S}_2$ – $\xi = 0.1$	0.2 (±0.04)	0.2 (±0.04)	0.495 (±0.003)	0.148 (±0.002)	7.7 (±0.02)
$\mathcal{S}_3$ – Never	0.0 (±0.01)	1.4 (±0.10)	0.889 (±0.004)	0.267 (±0.004)	0.1 (±0.00)
$\mathcal{S}_4$ – Immediate	0.0 (±0.00)	0.0 (±0.00)	0.317 (±0.002)	0.030 (±0.001)	5.1 (±0.00)
$\mathcal{S}_5$ – Annuitize-73	2.0 (±0.12)	2.0 (±0.12)	0.483 (±0.004)	0.129 (±0.003)	6.2 (±0.01)
$\mathcal{S}_6$ – Annuitize-80	3.5 (±0.16)	3.5 (±0.16)	0.612 (±0.004)	0.267 (±0.004)	7.7 (±0.03)
$\mathcal{S}_7$ – Annuitize-10%	0.2 (±0.04)	0.2 (±0.04)	0.445 (±0.003)	0.127 (±0.002)	7.4 (±0.02)

Table 34: Simulation results — $\mathcal{P}_3$ ($\sigma = 8\%$, $\mu = 4\%$, $\bar{\alpha} = 78.1\%$)

Strategy	P_{85} (%)	P_{95} (%)	$\mathbb{E}[x_{\boldsymbol{\tau}}]$	$\mathbb{E}\left[x_{\boldsymbol{\tau}} \middle\vert \boldsymbol{\tau} < 80\right]$	$\mathbb{E}[y_{80}]$ (%)
$\mathcal{S}_1$ – $\xi = 1$	0.0 (±0.00)	0.0 (±0.00)	0.390 (±0.002)	0.079 (±0.001)	7.4 (±0.01)
$\mathcal{S}_2$ – $\xi = 0.1$	0.1 (±0.03)	0.1 (±0.03)	0.485 (±0.003)	0.152 (±0.002)	7.6 (±0.02)
$\mathcal{S}_3$ – Never	0.0 (±0.01)	1.3 (±0.10)	0.887 (±0.004)	0.268 (±0.004)	0.1 (±0.00)
$\mathcal{S}_4$ – Immediate	0.0 (±0.00)	0.0 (±0.00)	0.317 (±0.002)	0.031 (±0.001)	5.1 (±0.00)
$\mathcal{S}_5$ – Annuitize-73	2.1 (±0.13)	2.1 (±0.13)	0.483 (±0.003)	0.131 (±0.003)	6.3 (±0.01)
$\mathcal{S}_6$ – Annuitize-80	3.5 (±0.16)	3.5 (±0.16)	0.610 (±0.004)	0.268 (±0.004)	7.7 (±0.03)
$\mathcal{S}_7$ – Annuitize-10%	0.2 (±0.04)	0.2 (±0.04)	0.444 (±0.003)	0.127 (±0.002)	7.4 (±0.02)

larger expected bequest, particularly among paths that terminate before age 80. $\mathcal{S}_2$ delivers $\mathbb{E}\left[x_{\boldsymbol{\tau}} | \boldsymbol{\tau} < 80\right] \approx 0.148$–$0.152$, compared to 0.057–0.079 for $\mathcal{S}_1$. This contrast illustrates the economic meaning of ξ. A lower weight shifts the free boundary outward, delaying annuity purchases and leaving more liquid wealth available to heirs. The immediate strategy $\mathcal{S}_4$ also achieves zero ruin, but at the cost of nearly zero pre-deadline bequest (0.031) and complete sacrifice of capital market growth. By contrast, $\mathcal{S}_3$ exhibits markedly different ruin probabilities before ages 85 and 95 (0.0% vs. 1.4% for $\mathcal{P}_1$ and $\mathcal{P}_2$, 0.0% vs. 1.3% for $\mathcal{P}_3$), confirming that the never annuitize strategy provides no protection against extreme longevity risk despite offering the highest bequest ($\mathbb{E}\left[x_{\boldsymbol{\tau}}\right] \approx 0.886$–$0.889$) and liquidity. Figure 65 shows the ruin probabilities P_{85} and P_{95} for each strategy across the three parameter scenarios. This figure summarizes the first two columns of Tables 32–34.

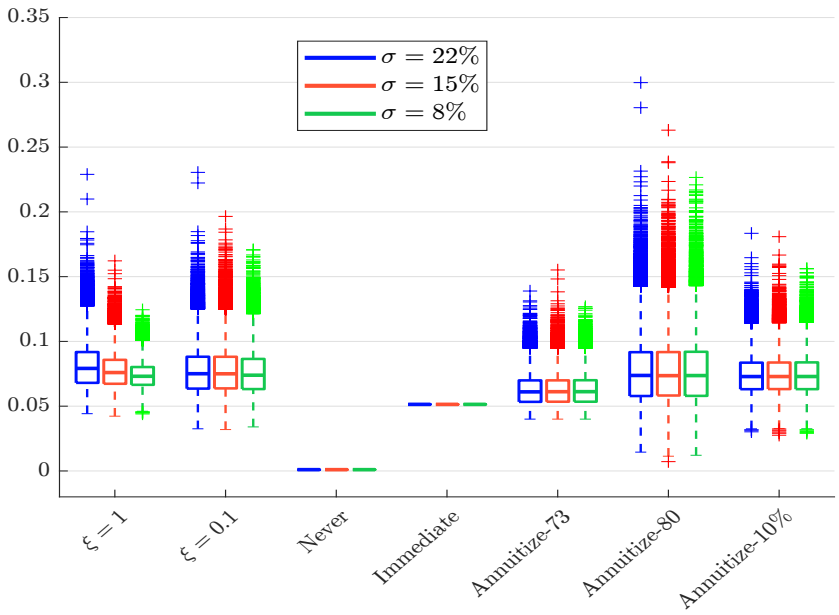
Figure 65: Ruin probabilities P_{85} and P_{95}



All three parameter scenarios have the same Sharpe ratio $(\mu - r)/\sigma = 0.25$. Therefore, any strategy that invests at the constant Merton fraction $\bar{\alpha}$ faces identical normalized wealth dynamics regardless of whether σ equals 20%, 12%, or 8%. As a result, the rule-based strategies $\mathcal{S}_3$, $\mathcal{S}_5$, $\mathcal{S}_6$, and $\mathcal{S}_7$ produce nearly identical ruin probabilities and bequests across $\mathcal{P}_1$–$\mathcal{P}_3$ and the small residual differences ($\mathcal{S}_5$: 2.1%, 2.0% and 2.1%, $\mathcal{S}_6$: 3.6%, 3.5% and 3.5%; $\mathcal{S}_7$: 0.2% throughout) lie within Monte Carlo sampling uncertainty and carry no structural interpretation. The HJBVI strategies, whose allocation $\alpha^\star(t, x, y)$ is not restricted to $\bar{\alpha}$, do respond to the volatility regime. $\mathcal{S}_1$ achieves $\mathbb{E}[y_{80}] = 8.1\%$ in $\mathcal{P}_1$, but only 7.4% in $\mathcal{P}_3$, while continuing to deliver zero ruin in all cases. This variation reflects the fact that the free boundary $x^\star(t, y)$ depends on σ and μ individually, not only through their ratio. The optimal policy therefore adapts its annuitization pace to the prevailing market environment in a way that a constant-fraction rule cannot replicate.

The expected annuity income at age 80 reflects the cumulative impact of investment returns and annuitization choices over a 15-year period. $\mathcal{S}_6$ achieves an expected income $\mathbb{E}[y_{80}]$ of 7.7% across all three scenarios. It retains the full risky portfolio until conversion and

Figure 66: Probability distribution of annuity income at age 80



captures the equity risk premium throughout the time horizon. Strategy $\mathcal{S}_1$ closely tracks strategy $\mathcal{S}_6$ while incurring zero ruin, demonstrating that the optimal policy recaptures most of the return advantage through well-timed partial purchases. Strategy $\mathcal{S}_4$ is anchored at 5.1% (deterministic and independent of σ). Converting at age 65 buys a long-dated annuity cheaply but forgoes 15 years of investment returns. The difference between $\mathcal{S}_4$ and $\mathcal{S}_1/\mathcal{S}_2$ represents the opportunity cost that $\mathcal{S}_4$ incurs by annuitizing immediately. Figure 66 shows the probability distribution of annuity income y_{80} at age 80 for each strategy across the three parameter scenarios $\mathcal{P}_1$–$\mathcal{P}_3$. Strategies $\mathcal{S}_5$ and $\mathcal{S}_6$, which hold a full Merton portfolio until the conversion date, exhibit substantial dispersion in y_{80}. Since their final annuity income depends almost entirely on the market state at a single point in time, these strategies create concentrated market-timing risk. By contrast, the HJBVI strategies $\mathcal{S}_1$ and $\mathcal{S}_2$ progressively lock in realized gains through partial annuity purchases. At age 80, their distribution is tighter and centered at a higher level than that of the deterministic benchmark $\mathcal{S}_4$ (which converts immediately at age 65). This illustrates the primary advantage of the optimal flex-to-fix policy. It captures much of the equity risk premium over the accumulation horizon while insulating the ultimate income level from adverse market shocks near the deadline.

Among the five rule-based strategies, $\mathcal{S}_7$ (annuitize-10%) has the most favorable risk-bequest profile. It limits ruin to 0.2% across all scenarios, maintains a pre-deadline bequest (0.127), comparable to Strategy $\mathcal{S}_5$ (0.129–0.131), and provides an annuity income of 7.4% at age 80, close to the optimal strategy $\mathcal{S}_1$ (7.4%–8.1%). Annual partial purchases mitigate exposure to adverse market conditions at conversion dates, thus avoiding the concentrated pre-conversion drawdown risk affecting Strategies $\mathcal{S}_5$ and $\mathcal{S}_6$. Compared to $\mathcal{S}_3$, $\mathcal{S}_7$ substantially reduces extreme longevity risk. Compared to $\mathcal{S}_4$, it preserves meaningful capital-market upside. Moreover, the staggered purchase schedule mitigates interest rate and adverse selection risks at each transaction, making Strategy $\mathcal{S}_7$ practical and implementable as a decumulation rule.

5 Conclusion

In this paper, we examine the issue of decumulation, *i.e.*, the conversion of accumulated wealth into retirement income, in the context of longevity and market risk. This process occurs within a framework of three competing motives: retaining liquidity, securing guaranteed income, and preserving a bequest. Empirical evidence shows that this problem cannot be reduced to preventing the premature exhaustion of savings. Retirees may preserve substantial capital and spend relatively little in order to act as a buffer against unexpected expenditure and market losses, but also because liquid wealth itself constitutes a bequest asset. Unlike annuity income, capital can be transferred to heirs, and every annuity purchase permanently forecloses that possibility. Nevertheless, the demand for guaranteed lifetime income remains significant, even though observed retirement choices are much more concentrated in flexible-access and drawdown products than in annuities. This contrast is consistent with the annuity puzzle. Low annuity uptake does not necessarily imply that retirees do not value guaranteed income, because annuitization is irreversible and simultaneously reduces the wealth available for unexpected expenditures and bequests. Private annuitization decisions must also be considered alongside the lifetime income already provided by public and occupational pension systems. Consequently, the economic comparison between drawdown and annuitization depends on the relative importance of liquidity, bequests, and guaranteed income rather than expected returns alone.

Mortality is a key factor in this problem because survival probabilities influence the length of retirement spending and the value of income based on longevity. This study uses the Human Mortality Database (HMD) to document long-term mortality trends and compare alternative stochastic mortality models. The data reveal a significant decrease in mortality over the past century, as well as notable variations among countries, genders, and age groups. For instance, the long-term decline in the central death rate has considerably slowed for individuals aged 75 and older since 2010, with some countries experiencing a slight increase. Additionally, mortality increases rapidly at older ages, making it difficult to model these dynamics at very advanced ages. Thus, reliable retirement planning requires mortality models that accurately capture age-specific mortality rates and how they change over time. Our empirical comparison confirms the importance of mortality model structure for this task. Forecast accuracy depends on how age and time effects are represented. For French data and retirement-relevant ages, the Lee-Carter (LC) model outperforms the Cairns-Blake-Dowd (CBD) model, both in-sample and out-of-sample. Within the LC family, switching from weighted least squares to Poisson or negative binomial specifications results in smaller changes in forecast accuracy than switching from LC to CBD. Therefore, the structural representation of mortality dynamics is more important for forecasting than the precise statistical specification used for estimation. Additionally, forecast errors increase with the projection horizon, and forecasting becomes more difficult when the underlying mortality trend changes.

In Section 3, we convert mortality projections into discounted survival-contingent cash flows to translate projected survival paths into annuity values. Because there are insufficient observable market prices for standardized longevity-linked instruments, it is not possible to calibrate the market price of systematic longevity risk. Therefore, we use actuarial expected present values under the physical measure, meaning that the resulting annuity values are based on real-world mortality forecasts. Annuity values increase with improved survival over time and are higher for females. They also decrease with older issue ages when discounting is held constant. The mortality specification also visibly affects valuation. LC produces higher annuity values than CBD for both females and males in the 2023 age-65 comparison. Introducing stochastic mortality has a minimal effect on the mean actuarial

value[77]. However, it reveals dispersion and tail risk in the annuity value distribution that are relevant for risk measurement and capital requirements. Annuity values are determined by demographic and financial conditions. Mortality improvement changes survival probabilities gradually and persistently. In contrast, interest rates can fluctuate substantially over short periods. Therefore, a constant discount rate cannot accurately reflect observed variations in market annuity payout yields over time. Using prevailing long-term interest rates for discounting produces dynamics that are much closer to those observed in practice. Thus, year-to-year variation in annuity values is primarily associated with the financial component, while changes in mortality occur over a longer time horizon. The same pooling mechanism that generates the mortality credit also determines the mortality risk structure of an annuity portfolio. We distinguish between idiosyncratic mortality risk, which decreases as the number of contracts grows, due to the diversification of individual survival outcomes across the pool, and systematic longevity risk, which remains because all annuitants are exposed to the same underlying population mortality path. The results confirm this distinction. Initially, individual mortality dominates the risk of a very small pool. However, as the number of annuitants grows, the systematic mortality component becomes increasingly important, and it dominates the liability variance of large annuity portfolios. Thus, although annuity pooling generates mortality credits and diversifies individual mortality outcomes, it leaves insurers uncertain about the future evolution of population mortality.

Section 4 brings these mortality and annuity results into the decumulation problem itself. We focus on the transition between two complementary retirement income strategies: (1) drawdown, which preserves liquidity, market exposure, and bequest potential, and (2) annuitization, which transfers longevity risk and provides guaranteed lifetime income. This trade-off motivates the *flex-to-fix* approach, in which retirees remain invested and liquid initially before converting part or all of their remaining wealth into an annuity at a later age. The flex-to-fix strategy has two limiting cases: *fix-now*, where all wealth is annuitized immediately at retirement, and *fix-never*, where the retiree remains in pure drawdown throughout retirement and never annuitizes. The full range of delayed conversion strategies that the flex-to-fix framework is designed to study lies between these two extremes. We examine this framework from two complementary perspectives. The static approach evaluates the trade-off under a predetermined conversion rule that is fixed in advance and independent of how wealth, mortality, or markets evolve. This approach asks which annuitization age, chosen once and for all, best serves a given retiree. The dynamic approach, on the other hand, allows the conversion decision to respond to the retiree's actual situation at each point in retirement. This approach asks when annuitization becomes optimal given the actual wealth and income states attained. Together, these two perspectives establish a benchmark for the value of delaying annuitization under a simple rule, and characterize how much a fully state-contingent policy can improve upon that benchmark.

The static analysis examines the transition between drawdown and annuitization from two perspectives. The mortality credit approach compares the return on a survival-contingent pool embedded in an annuity with the return available from a drawdown portfolio. The actuarial approach compares lifetime consumption and bequest outcomes generated by alternative predetermined conversion dates. The mortality credit analysis provides a simple economic rationale for delaying annuitization. Indeed, mortality credits are relatively small at younger retirement ages, but they increase sharply with age. This makes the pooling advantage of annuities progressively more valuable over time. For French data, the crossover point occurs at an optimal annuitization age between 75 and 80 when the drawdown portfolio earns a 4% return. However, immediate annuitization is optimal when the assumed

[77]The mean stochastic value is nearly identical to the deterministic value, with a Jensen gap below five basis points.

return falls below 2%. Improvements in longevity have historically shifted this crossover point toward later ages[78]. However, the implied conversion age remains highly sensitive to the assumed return on the drawdown portfolio. The fixed-date actuarial analysis confirms and enhances these findings by evaluating full lifetime consumption and bequest outcomes. The analysis shows that there is no single optimal annuitization age across all retiree preferences and investment environments. The optimal policy may range from immediate annuitization (fix-now) to delayed annuitization (flex-to-fix) to maintaining liquidity throughout retirement (fix-never). For instance, targeting an expected bequest of 50% of initial wealth pushes the optimal conversion age beyond 75. Generally, higher expected investment returns increase the value of waiting, and a stronger bequest motive delays conversion by increasing the value of retaining liquid and bequestable wealth. In the fixed-date analysis, the bequest motive generally influences the optimal policy more than the expected return on the drawdown portfolio. The withdrawal rate becomes the dominant driver in the higher-spending region where planned spending approaches or exceeds the income available from immediate annuitization, causing exhaustion risk to rise sharply. However, at lower withdrawal rates, the bequest motive and expected return become the principal determinants of the optimal delay. Finally, asset volatility has the lowest effect on the choice between fix-now, flex-to-fix, and fix-never, but materially affects the risk around the selected policy. Higher volatility produces more severe lower-tail losses, and at sufficiently high levels, it reduces the probability that delayed annuitization outperforms immediate annuitization over the retiree's realized lifetime. In the fixed-date analysis, females also tend to annuitize later than males, which is consistent with their lower mortality rates and longer expected retirement horizons.

Taken together, these results can be summarized by the following ordering of the four primary parameters in terms of their relative importance:

$$\alpha \succ \psi \succ \mu \succ \sigma$$

where α is the withdrawal rate of the drawdown strategy, ψ is the bequest weight in the utility function, and μ and σ are the expected return and volatility of the investment portfolio, respectively. The withdrawal rate α has the largest overall impact. Once spending is sufficiently high relative to immediate annuity income, exhaustion risk rises sharply and significantly alters the optimal conversion date. The bequest weight ψ systematically delays conversion across nearly all investment environments by raising the value of liquid wealth. The expected return μ further extends the optimal deferral period, though its effect interacts closely with the withdrawal rate. Volatility σ has the weakest effect on the primary policy selection[79]. Yet, it remains critical in governing the dispersion of outcomes around the optimal strategy. However, it should be noted that this parameter hierarchy is not static across the entire parameter space. For instance, the withdrawal rate α dominates in regions where the normalized withdrawal rate $\bar{\alpha} = \alpha / \left(Y_0/W_0\right)$ approaches or exceeds one, while below this threshold its effect on exhaustion risk and delay timing is minimal, leaving bequest weight ψ and expected return μ as the primary drivers of optimal timing. Meanwhile volatility σ continues to dictate downside risk. These findings demonstrate that, even under a predetermined conversion rule, the timing of annuitization depends heavily on the interplay between spending requirements, bequest preferences, and the investment opportunity set.

Our research into dynamic approaches establishes a key finding. The optimal annuitization decision is fundamentally a state-dependent stopping problem rather than a predetermined calendar rule. The irreversibility of annuity purchases plays a central role. Since

[78] The crossover point was between ages 60 and 65 at the beginning of the 1990s.
[79] It rarely changes the selected range of delays.

annuity contracts cannot be reversed, each purchase permanently reduces a retiree's future flexibility. Therefore, there is an option value in waiting because maintaining liquid wealth preserves access to financial markets and the ability to leave a bequest. This option is worth preserving as long as mortality credits and market excess returns do not sufficiently compensate for the loss of liquidity. Thus, the optimal policy is threshold-based rather than age-triggered, which is often recommended by practitioners. Under this structure, annuitization only occurs if the marginal value of annuity income exceeds the marginal value of waiting. In this paper, we investigate two dynamic approaches based on this conceptual framework. The first approach extends the static actuarial fixed-date annuitization problem by modeling the pure timing dimension as an optimal stopping time formulation. Assuming annuitization will eventually occur as a single lump-sum conversion, when is the optimal time to annuitize? This setup allows us to use a dynamic programming framework and Bellman's principle of optimality to study the relationship between mortality uncertainty and wealth dynamics. Two main results emerge from this analysis. First, a well-chosen predetermined annuitization horizon captures most of the expected objective value available from delayed conversion. At baseline calibration, the optimal fixed-date rule outperforms the fix-now strategy by around 30%. However, allowing the exact conversion date to depend on the realized state adds only a few basis points to that gain. Therefore, the additional expected value of state-contingent timing is minimal across most of the parameter grid. This value is largest in high-spending or high-volatility configurations, where the timing decision matters most. Second, the state-dependent nature of the optimal policy remains economically significant even when it adds little expected value. Realized wealth and mortality paths can move individual retirees into the annuitization region at materially different ages. Finite annuitization can occur ex post, even when the optimal predetermined strategy is fix-never. Even when considering the strongest bequest motive, where every fixed-date configuration selects the fix-never rule, more than half of the simulated paths annuitize under the dynamic policy in the riskiest investment environment. Therefore, the fix-never strategy remains the optimal ex ante commitment when the bequest motive is strong. However, ex post, it can be outperformed by the flex-to-fix strategy.

The previous approach uses a binary formulation in which wealth is either fully annuitized or retained in the drawdown investment portfolio on a single conversion date. Our second approach is more dynamic. It relaxes the all-or-nothing constraint by modeling a gradual transition from drawdown to annuity income. This transition is subject to a mandatory target date and is modeled using a Hamilton-Jacobi-Bellman variational inequality (HJBVI). This approach more closely aligns with the flex-to-fix concept. Throughout the flex window, the retiree maintains a mixed portfolio of liquid assets and annuities and continuously rebalances between the two. The target date naturally serves as a commitment device, transforming the open-ended optimal stopping problem into a finite-horizon singular control problem with a deterministic terminal condition. By incorporating irreversibility and a forced deadline into the same framework, the HJBVI model captures the flexibility that younger retirees value and the longevity protection that is required as the horizon shortens. The most distinctive structural feature of this analysis is the *free boundary* separating the wait and action regions in the wealth-income state space, which is composed of liquid wealth and current annuity income. Unlike a simple age trigger, this boundary considers the interaction between liquid wealth, current annuity income, and the time remaining until the target date. A wealthy retiree should annuitize not because of their age, but rather because the marginal utility of converting additional liquid wealth into guaranteed income is greater than the marginal utility of retaining that wealth for inheritance or investment growth. Conversely, a retiree with ample guaranteed income relative to their spending needs — such as a substantial pre-existing pension or existing annuityincome — can maintain their

liquid portfolio. The guaranteed income provides insurance, while the liquid wealth primarily serves as a reserve for bequests or emergencies. A second key insight is the *implicit bond component*. According to standard portfolio theory, a retiree's optimal equity share is determined by their risk aversion and the market Sharpe ratio. In the flex-to-fix framework, existing annuity income acts as an implicit risk-free bond position within total wealth. As the retiree accumulates more guaranteed income through progressive purchases, the implicit bond component of total wealth increases. Consequently, the liquid portfolio must tilt toward risky assets to restore the desired overall risk exposure. This effect is not a behavioral anomaly but rather the direct consequence of applying optimal portfolio separation to total wealth rather than financial wealth alone. The practical implication is counterintuitive. A retiree should generally increase their equity allocation as they annuitize more since the annuity insures against downside risk. This has direct consequences for product design. Decumulation products that mechanically reduce risk in the liquid portfolio as annuity income increases will likely result in suboptimal total risk exposure.

The simulation study confirms these mechanisms. Fixed-date strategies with constant-mix allocations are subject to concentrated market-timing risk before a conversion date, whether it is at age 73 or 80[80]. If equity markets perform poorly in the quarters immediately before conversion, the retiree locks in a permanently low income level with no possibility of recovery. Across our simulated scenarios, the two annuitize-at-a-single-date benchmarks expose retirees to significant ruin risk before age 85 — approximately 2% for annuitizing at age 73 and 3.5% for annuitizing at age 80 — whereas the optimal HJBVI policy does not experience ruin. This is achieved by spreading out purchases over the entire flex window and progressively locking in realized gains through partial conversions each time the free boundary is reached. By the target date, the retiree has secured a substantial stream of guaranteed income through well-timed partial purchases. The residual mandatory conversion affects only a small fraction of the total wealth. The resulting income distribution on the target date is significantly tighter and centered at a higher level than under the immediate strategy or fixed-date rules. It closely tracks the income delivered by holding a full risky portfolio until the target date but without an exposure to ruin. Spreading purchases evenly across the flex window is the predetermined rule that comes closest to replicating this behavior. This reduces the probability of ruin to approximately 0.2%, capturing most of the equity risk premium. Nevertheless, this rule remains dominated by the fully state-contingent HJBVI policy. The bequest motive emerges as the primary determinant of the location of the free boundary, which affects the overall pace of annuitization. A retiree who strongly desires to leave liquid wealth to their heirs will have a substantially higher threshold for voluntary annuitization. This retiree will postpone purchases until a shrinking time horizon forces them to act. They will reach the target date with a larger pre-conversion bequest than an otherwise identical retiree without a bequest motive. In our baseline simulation, moving from a weak to a strong bequest motive triples the average bequest left to heirs who die before the target date. The dynamic model also reveals how the bequest motive influences the equity proportion of the liquid portfolio and the timing of each partial purchase within the flex window.

However, the HJBVI framework relies on two assumptions that simplify the model and identify areas for future research. First, the model assumes a single risky asset, but real-world retirement portfolios are multi-asset. Multi-asset allocation is not addressed in this study, even though it remains a key element of decumulation strategy design in practice, since different asset classes have distinct risk-return profiles and interact differently with the annuitization decision. Second, annuity prices are computed under a constant discount

[80]These ages serve as two benchmarks for annuitizing on a single date. The target date is also set to age 80.

rate ignoring interest rate risk, even though annuity prices are highly sensitive to the term structure of interest rates. Rising long-term interest rates lower annuity prices and make conversion more attractive, while falling rates raise prices and may justify delaying conversion. The most immediate extension that addresses these limitations is a multi-asset HJBVI model with a stochastic interest rate. This model would address asset allocation within the liquid portfolio and demonstrate the effect of annuity purchases and bond duration choices on each other. Since an annuity behaves like a long-duration bond, each purchase alters the retiree's overall interest rate exposure. Thus, the duration of bonds held in the liquid portfolio must be adjusted accordingly. Because the annuity price is stochastic, the model generates new hedging demands. This could justify early partial annuitization even for retirees with strong bequest motives, as it would serve as a hedge against falling rates.

Finally, this study raises the broader question of how accumulation and decumulation strategies interact. Currently, these phases and their respective strategies are typically designed separately. Glide paths in target date funds exemplify this issue. These glide paths mechanically de-risk the portfolio as retirement approaches, following an established schedule regardless of the decumulation strategy the retiree will use. However, this choice is not neutral. For instance, a retiree who annuitizes immediately upon retiring (fix-now) faces a different risk profile than a retiree who never annuitizes to preserve an inheritance for their heirs (fix-never). Failing to consider this distinction can result in genuine investment paradoxes. An individual may accumulate under a glide path that reduces equity exposure to 10% by retirement, and then in retirement, may resume accumulating risk exposure[81]. In other words, decumulation can be a phase of renewed accumulation. A de-risking schedule calibrated solely to the retirement date may be miscalibrated for the period that follows. This observation suggests a natural extension of the present analysis. The challenge lies in combining accumulation and decumulation into a single investment strategy, rather than treating the retirement date as a fixed boundary between two independently optimized problems. Such a unified *lifecycle investing* framework would explicitly consider retirees' ultimate objectives, such as income security, liquidity, and bequests, instead of treating the accumulation and decumulation phases separately. In doing so, this framework would transform the retirement investing problem into a *goal-based investing* approach, in which the same objectives that motivate the flex-to-fix strategy developed in this paper would guide portfolio composition both before and after retirement.

[81] For instance, a substantial first-pillar public pension could provide essential income, enabling the remaining liquid wealth to be invested more aggressively. Alternatively, this wealth could be earmarked for inheritance rather than near-term consumption.

References

Agnew, J. R., Anderson, L. R., Gerlach, J. R., and Szykman, L. R. (2008). Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults. *American Economic Review*, 98(2), pp. 418-422.

Alho, J. M. (2000). Discussion on "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications", *North American Actuarial Journal*, 4(1), pp. 91-93.

Amundi Investment Solutions (2026). *Decoding Investors*. Annual Report 2026, 100 pages.

Antolin, P. (2008). Ageing and the Payout Phase of Pensions, Annuities and Financial Markets. *OECD Working Papers on Insurance and Private Pensions*, 29, 17 pages.

Banks, J., and Crawford, R. (2022). Managing Retirement Incomes. *Annual Review of Economics*, 14(1), pp. 181-204.

Barrieu, P., Bensusan, H., El Karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y. (2012). Understanding, Modelling and Managing Longevity Risk: Key Issues and Main Challenges. *Scandinavian Actuarial Journal*, 2012(3), pp. 203-231.

Bavafa, H., Mukherjee, A., and Welch, T. Q. (2026). Asset Decumulation in Retirement: Patterns, Predictors, and the Role of Financial Literacy. *Geneva Papers on Risk and Insurance — Issues and Practice*, 51(3), pp. 682-711.

Benartzi, S., Previtero, A., and Thaler, R. H. (2011). Annuitization Puzzles. *Journal of Economic Perspectives*, 25(4), pp. 143-164.

Bengen, W. P. (1994). Determining Withdrawal Rates Using Historical Data. *Journal of Financial Planning*, 7(4), pp. 171-180.

Bengen, W. P. (1996). Asset Allocation for a Lifetime. *Journal of Financial Planning*, 9(4), pp. 58-67.

Bengen, W. P. (1997). Conserving Client Portfolios During Retirement, Part III. *Journal of Financial Planning*, 10(6), pp. 84-97.

Bernheim, B. D. (1991). How Strong Are Bequest Motives? Evidence Based on Estimates of the Demand for Life Insurance and Annuities. *Journal of Political Economy*, 99(5), pp. 899-927.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Zeldes, S. P. (2014). What Makes Annuitization More Appealing?. *Journal of Public Economics*, 116, pp. 2-16.

Biffis, E. (2005). Affine Processes for Dynamic Mortality and Actuarial Valuations. *Insurance: Mathematics and Economics*, 37(3), pp. 443-468.

Blake, D., and Burrows, W. (2001). Survivor Bonds: Helping to Hedge Mortality Risk. *Journal of Risk and Insurance*, 68(2), pp. 339-348.

Blake, D., and Cairns, A. J. (2021). Longevity Risk and Capital Markets: The 2019–20 Update. *Insurance: Mathematics and Economics*, 99, pp. 395-439.

Blake, D., Cairns, A. J. G., and Dowd, K. (2006). Living with Mortality: Longevity Bonds and other Mortality-linked Securities. *British Actuarial Journal*, 12, pp. 153-197.

Blake, D., Cairns, A. J., Dowd, K., and Kessler, A. R. (2019). Still Living with Mortality: The Longevity Risk Transfer Market after One Decade. *British Actuarial Journal*, 24, e1, pp. 1-80.

Booth, H., Hyndman, R. J., Tickle, L., and De Jong, P. (2006). Lee-Carter Mortality Forecasting: A Multi-country Comparison of Variants and Extensions. *Demographic Research*, 15, pp. 289-310.

Brown, J. R., and Poterba, J. M. (2000). Joint Life Annuities and Annuity Demand by Married Couples. *Journal of Risk and Insurance*, 67(4), pp. 527-553.

Boyle, P. and Webb, S. (2022). The "Flex First, Fix Later" Pension — Is This the Future of Retirement?. *LCP on point*.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and Economics*, 31(3), pp. 373-393.

Brown, J. R. (2001). Private Pensions, Mortality Risk, and The Decision to Annuitize. *Journal of Public Economics*, 82(1), pp. 29-62.

Brown, J. R., Kling, J. R., Mullainathan, S., and Wrobel, M. V. (2008). Why Don't People Insure Late-Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle. *American Economic Review*, 98(2), pp. 304-309.

Bruder, B., Caudamine, L., Do, H., Gisimundo, V., Granjon, C., and Xu, J. (2023). Optimal Decumulation Strategies for Retirement Solutions. *Amundi Working Paper*, 149, available at *SSRN*, 4652428, 57 pages.

Bruder, B., Roncalli, T., Schittly, C., and Xu, J. (2025). Retirement Accumulation Strategies with Real Assets and Inflation Risk. *Amundi Working Paper*, 180, available at *SSRN*, 5669210, 125 pages.

Butt, S., Chapman, B., Cottis Black, O., Dyer, A., and Palmer, S. (2025). *Planning and Preparing for Later Life 2024*. Department for Work and Pensions, DWP Research Report No. 1103.

Cairns, A. J., Blake, D., and Dowd, K. (2006a). A Two-factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4), pp. 687-718.

Cairns, A. J., Blake, D., and Dowd, K. (2006b). Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin*, 36(1), pp. 79-120.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A Quantitative Comparison of Stochastic Mortality Models using Data from England and Wales and the United States. *North American Actuarial Journal*, 13(1), pp. 1-35.

Callaghan, M., Ralston, D., and Kay, C. (2020). *Retirement Income Review: Final Report*. The Australian Government the Treasury, July, 648 pages.

CANNEX (2026). CANNEX PAY Index. www.cannex.com/index.php/services/united-state/benchmarks-indices/pay-index (report available at www.cannex.com/public/CANNEXPAYIndexUSA.pdf).

Cannon, E., and Tonks, I. (2004). UK Annuity Rates, Money's Worth and Pension Replacement Ratios 1957–2002. *Geneva Papers on Risk and Insurance — Issues and Practice*, 29(3), pp. 371-393.

Cannon, E.. and Tonks, I. (2008). *Annuity Markets*. Oxford University Press.

Carriere, J. F. (1999). No-arbitrage Pricing for Life Insurance and Annuities. *Economics Letters*, 64(3), pp. 339-342.

Carter, L. R. (1996). Forecasting U.S. Mortality: A Comparison of Box-Jenkins ARIMA and Structural Time Series Models. *The Sociological Quarterly*, 37(1), pp. 127-144.

Clare, A., Seaton, J., Smith, P. N., and Thomas, S. (2024). The Science of Flexible Retirement Choices: Switching Retirement Savings Into an Annuity. *Journal of Retirement*, 11(4), pp. 59-75.

Cooley, P. L., Hubbard, C. M., and Walz, D. T. (1998). Retirement Savings: Choosing a Withdrawal Rate That Is Sustainable. *American Association of Individual Investors Journal*, 20(2), pp. 16-21.

Corebridge Financial (2026). *The Decumulation Planning Gap: Findings from a Survey of Americans Ages 45 to 79*. Corebridge Financial (Survey conducted by Greenwald Research).

Dahl, M. (2004). Stochastic Mortality in Life Insurance: Market Reserves and Mortality-linked Insurance Contracts. *Insurance: Mathematics and Economics*, 35(1), pp. 113-136.

Davidoff, T., Brown, J. R., and Diamond, P. A. (2005). Annuities and Individual Welfare. *American Economic Review*, 95(5), pp. 1573-1590.

De Nardi, M. (2004). Wealth Inequality and Intergenerational Links. *Review of Economic Studies*, 71(3), pp. 743-768.

Deaton, A. (2005). Franco Modigliani and the Life-cycle Theory of Consumption. *Banca Nazionale del Lavoro Quarterly Review*, 58(233-234), pp. 91-107.

Delwarde, A., Denuit, M., and Partrat, C. (2007). Negative Binomial Version of the Lee-Carter Model for Mortality Forecasting. *Applied Stochastic Models in Business and Industry*, 23(5), pp. 385-401.

Denuit, M., Hieber, P., and Robert, C. Y. (2022). Mortality Credits within Large Survivor Funds. *ASTIN Bulletin*, 52(3), pp 813-834.

Dickson, D. C., Hardy, M. R., and Waters, H. R. (2020). *Actuarial Mathematics for Life Contingent Risks*. Third edition, Cambridge University Press.

Dushi, I., and Webb, A. (2004). Household Annuitization Decisions: Simulations and Empirical Analyses. *Journal of Pension Economics & Finance*, 3(2), pp. 109-143.

European Commission and Social Protection Committee (2024). *Pension Adequacy Report — Current and Future Income Adequacy in Old Age in the EU*. Volume I, June, 198 pages.

Eurostat (2024). Pensions and Labour Market Participation — Main Characteristics. *Statistics Explained*, 27 pages.

Financial Conduct Authority (2018). *Retirement Outcomes Review: Final Report*. FCA Report, June, 78 pages.

Financial Conduct Authority (2025). *Financial Lives 2024 Survey — Pensions: Selected Findings*. FCA Report, May.

Financial Conduct Authority (2026). *Retirement Income Market Data 2025/26*. FCA report, September.

Finkelstein, A., and Poterba, J. (2004). Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market. *Journal of Political Economy*, 112(1), pp. 183-208.

Forsyth, P., and Labahn, G. (2007). Numerical Methods for Controlled Hamilton-Jacobi-Bellman PDEs in Finance. *Journal of Computational Finance*, 11, pp. 1-44.

Geddes, T., Hu, M. L., Johnson, H., Lothrop, G. (2026). *2024 Retirement Risk Survey: Report of Findings*. Society of Actuaries Research Institute, April, 109 pages.

Gerrard, R., Haberman, S., and Vigna, E. (2004). Optimal Investment Choices Post-retirement in a Defined Contribution Pension Scheme. *Insurance: Mathematics and Economics*, 35(2), pp. 321-342.

Goldman Sachs Asset Management (2025). *New Economics of Retirement*. Retirement Survey & Insights Report 2025, 38 pages.

Gong, G., and Webb, A. (2010). Evaluating the Advanced Life Deferred Annuity — An Annuity People Might Actually Buy. *Insurance: Mathematics and Economics*, 46(1), pp. 210-221.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hershfield, H. E., Shu, S., Brown, J. R., Hurwitz, A., Milevsky, M. A., Mitchell, O. S., and Toland, T. (2026). The Annuity Puzzle Revisited: Barriers, Behavior, and Policy Paths to Lifetime Income. *NBER*, 35145.

Horneff, W. J., Maurer, R. H., and Stamos, M. Z. (2008). Optimal Gradual Annuitization: Quantifying the Costs of Switching to Annuities. *Journal of Risk and Insurance*, 75(4), pp. 1019-1038.

Holzmann, R., and Hinz, R. (2005). *Old Age Income Support in the 21st Century — An International Perspective on Pension Systems and Reform*. World Bank, 246 pages.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press.

Hu, W. Y., and Scott, J. S. (2007). Behavioral Obstacles in the Annuity Market. *Financial Analysts Journal*, 63(6), pp. 71-82.

HMD (2026). *Human Mortality Database*. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org.

Inkmann, J., Lopes, P., and Michaelides, A. (2011). How Deep Is the Annuity Market Participation Puzzle?. *Review of Financial Studies*, 24(1), pp. 279-319.

Kopczuk, W., and Lupton, J. P. (2007). To Leave or Not to Leave: The Distribution of Bequest Motives. *Review of Economic Studies*, 74(1), pp. 207-235.

Lee, R. D. (2000). The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North American Actuarial Journal*, 4(1), pp. 80-91.

Lee, R. D., and Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), pp. 659-671.

Lockwood, L. M. (2012). Bequest Motives and the Annuity Puzzle. *Review of Economic Dynamics*, 15, pp. 226-243.

Lockwood, L. M. (2018). Incidental Bequest and the Choice to Self-insure Late-Life Risks. *American Economic Review*, 108(9), pp. 2513-2550.

Ludkovski, M., and Young, V. R. (2008). Indifference Pricing of Pure Endowments and Life Annuities under Stochastic Hazard and Interest Rates. *Insurance: Mathematics and Economics*, 42(1), pp. 14-30.

Lusardi, A., and Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, 52(1), pp. 5-44.

Martellini, L., and Milhau, V. (2020). *Advances in Retirement Investing*. Cambridge University Press.

Merton, R. C. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *Review of Economics and Statistics*, 51(3), pp. 247-257.

Merton, R. C. (1971). Optimum Consumption and Portfolio Rules in a Continuous-Time Model. *Journal of Economic Theory*, 3(4), pp. 373-413.

Merton, R. C. (2014). The Crisis in Retirement Planning. *Harvard Business Review*, 92(7/8), pp. 43-50.

Milevsky, M. A. (1998). Optimal Asset Allocation towards the End of the Life Cycle: To Annuitize or Not to Annuitize?. *Journal of Risk and Insurance*, 65(3), pp. 401-426.

Milevsky, M. A. (2001). Optimal Annuitization Policies: Analysis of the Options. *North American Actuarial Journal*, 5(1), pp. 57-69.

Milevsky, M. A. (2006). *The Calculus of Retirement Income: Financial Models for Pension Annuities and Life Insurance*. Cambridge University Press.

Milevsky, M. A. (2013). *Life Annuities: An Optimal Product for Retirement Income*. Research Foundation of CFA Institute, 150 pages.

Milevsky, M. A., and Huang, H. (2011). Spending Retirement on Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates. *Financial Analysts Journal*, 67(2), pp. 45-58.

Milevsky, M. A., Moore, K. S., and Young, V. R. (2006). Asset Allocation and Annuity-Purchase Strategies to Minimize the Probability of Financial Ruin. *Mathematical Finance*, 16(4), pp. 647-671.

Milevsky, M. A., and Promislow, S. D. (2001). Mortality Derivatives and the Option to Annuitise. *Insurance: Mathematics and Economics*, 29(3), pp. 299-318.

Milevsky, M. A., Promislow, S. D., and Young, V. R. (2006). Killing the Law of Large Numbers: Mortality Risk Premiums and the Sharpe Ratio. *Journal of Risk and Insurance*, 73(4), pp. 673-686.

Milevsky, M. A., and Salisbury, T. S. (2015). Optimal Retirement Income Tontines. *Insurance: Mathematics and Economics*, 64, pp. 91-105.

Milevsky, M. A., and Young, V. R. (2007). Annuitization and Asset Allocation. *Journal of Economic Dynamics and Control*, 31(9), pp. 3138-3177.

Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., and Brown, J. R. (1999). New Evidence on the Money's Worth of Individual Annuities. *American Economic Review*, 89(5), pp. 1299-1318.

Modigliani, F. (1986). Life Cycle, Individual Thrift, and the Wealth of Nations. *American Economic Review*, 76(3), pp. 297-313.

OECD (2022a). *Financial Planning and Financial Education for Old Age in Times of Change*. OECD Publishing, Paris, July, 36 pages.

OECD (2022b). *OECD Pensions Outlook 2022*. OECD Publishing, Paris, December, 201 pages.

OECD (2024). *OECD Pensions Outlook 2024: Improving Asset-backed Pensions for Better Retirement Outcomes and More Resilient Pension Systems*. OECD Publishing, Paris, 200 pages.

OECD (2025). *Pensions at a Glance 2025: OECD and G20 Indicators*. OECD Publishing, Paris.

Olivieri, A. (2001). Uncertainty in Mortality Projections: An Actuarial Perspective. *Insurance: Mathematics and Economics*, 29(2), pp. 231-245.

Pashchenko, S. (2013). Accounting for Non-annuitization. *Journal of Public Economics*, 98, pp. 53-67.

Peijnenburg, K., Nijman, T., and Werker, B. J. (2016). The Annuity Puzzle Remains a Puzzle. *Journal of Economic Dynamics and Control*, 70, pp. 18-35.

Peijnenburg, K., Nijman, T., and Werker, B. J. M. (2017). Health Cost Risk: A Potential Solution to the Annuity Puzzle?. *Economic Journal*, 127(603), pp. 1598-1625.

Pensions Europe (2025). *Decumulation in Focus: Understanding the Payout Phase*. Report, April, 48 pages.

Pensions Policy Institute (2025). *Assessing the UK Retirement Income Market: Defaults, Active Choices, Innovation, and the existing gaps and challenges for Delivering Value for Money (VfM)*. Report, April, 38 pages.

Pitacco, E., Denuit, M., Heberman, S., and Olivieri, A. (2009). *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press.

Poterba, J. M., and Solomon, A. (2025). New Evidence on the Money's Worth of Immediate and Deferred Individual Annuities. *NBER Working Paper*, 28557.

Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Rabitti, G., and Borgonovo, E. (2020). Is Mortality or Interest Rate the Most Important Risk in Annuity models? A Comparison of Sensitivity Analysis Methods. *Insurance: Mathematics and Economics*, 95, pp. 48-58.

Reichling, F., and Smetters, K. (2015). Optimal Annuitization with Stochastic Mortality Probabilities and Correlated Medical Costs. *American Economic Review*, 105(11), pp. 3273-3320.

Samuelson, P. A. (1969). Lifetime Portfolio Selection by Dynamic Stochastic Programming. *Review of Economics and Statistics*, 51(3), pp. 239-246.

Scott, J. S., Sharpe, W. F., and Watson, J. G. (2009). The 4% Rule — At What Price?. *Journal of Investment Management*, 7(3), pp. 31-48.

Sharpe, W. (2017). In Ritholtz, B., Tackling the Nastiest, Hardest Problem in Finance, *Bloomberg Opinion*, June 5, 2017.

Stabile, G. (2006). Optimal Timing of the Annuity Purchases: A Combined Stochastic Control and Optimal Stopping Problem. *International Journal of Theoretical and Applied Finance*, 9(2), pp. 151-170.

Stamos, M. Z. (2008). Optimal Consumption and Portfolio Choice for Pooled Annuity Funds. *Insurance: Mathematics and Economics*, 43(1), pp. 56-68.

Thaler, R. (2019). In Mogelof, E., Martel, R., and Sharon, A., Financial Advisors and Retirement: The Decumulation Dilemma, *PIMCO Insights*, October 28, 2019.

The Investment Association (2026). *UK Pensions Policy Update*. February, 20 pages.

Thinking Ahead Institute (2026). *Global Pensions Assets Study 2026*. Annual Report.

United Nations (2024). *World Population Prospects: The 2024 Revision*. Department of Economic and Social Affairs, Population Division, July.

Vaupel, J. W., Villavicencio, F., and Bergeron-Boucher, M. P. (2021). Demographic Perspectives on the Rise of Longevity. *Proceedings of the National Academy of Sciences*, 118(9), e2019536118, 10 pages.

Waring, M. B., and Siegel, L. B. (2015). The Only Spending Rule Article You Will Ever Need. *Financial Analysts Journal*, 71(1), pp. 91-107.

Wilmoth, J. R. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. *Technical Report*, Department of Demography, University of California, Berkeley.

Wilmoth, J. R., and Lundström, H. (1996). Extreme Longevity in Five Countries: Presentation of Trends with Special Attention to Issues of Data Quality. *European Journal of Population*, 12(1), pp. 63-93.

Yaari, M. E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *Review of Economic Studies*, 32(2), pp. 137-150.

Young, V. R., and Zariphopoulou, T. (2002). Pricing Dynamic Insurance Risks using the Principle of Equivalent Utility. *Scandinavian Actuarial Journal*, 2002(4), pp. 246-279.

A Notations

- $a(x)$ is the value of a discrete annuity-immediate paying one monetary unit per year since age x.
- $\bar{a}(x)$ is the value of a continuous annuity paying one monetary unit per year since age x.
- $\ddot{a}(x)$ is the value of a discrete annuity-due paying one monetary unit per year since age x.
- $a(x,t)$ is the value of the annuity paid by an individual aged x at time t.
- $\breve{a}(x,t) = (1+\mathcal{L})\, a(x,t)$ is the value of the annuity including the loading factor paid by an individual aged x at time t.
- α is the withdrawal rate.
- α_x is the age-specific average log-mortality level in the Lee-Carter model.
- β_x is the age-specific loading (sensitivity to the period effect) in the Lee-Carter model.
- $\boldsymbol{c}_t$ is the consumption at date t.
- $D(x,t)$ is the number of deaths at age x during year t.
- $E(x,t)$ is the exposure-to-risk at age x during year t.
- $H = x^+ - x$ is the maximum projection horizon.
- $i_{x,t}$ is the internal rate of return associated to an annuity premium $a(x,t)$.
- κ_t is the time-varying index in the Lee-Carter model.
- $\hat{\kappa}_t$ is the expected value of κ_t.
- $\kappa_t^{(s)}$ is the simulated value of κ_t.
- $\kappa_t^{(1)}$ is the level parameter in the CBD model.
- $\kappa_t^{(2)}$ is the slope parameter in the CBD model.
- $\mathcal{L}$ is the annuity loading factor.
- $\text{logit}(x) = \ln\left(\dfrac{x}{1-x}\right)$ is the logit transformation.
- $m(x,t)$ is the central death rate at age x during time t.
- $p(x,t,h)$ is the probability of individual, alive at age x in year t, survives at age $x+h$.
- $\pi(x,t) = i_{x,t} - R_t^h$ is the risk premium.
- $q(x,t)$ is the one-year death probability, *i.e.*, the probability that an individual aged x at time t dies within one year.
- R_t^h is the interest rate for the period $[t, t+h]$.
- R_t is the investment return at date t.

- ϱ_t is the discount rate for computing the annuity price.
- $\tilde{\varrho}_t$ is the discount rate for computing the utility function.
- $\boldsymbol{\tau}$ is the survival time.
- t is the calendar time index.
- t_0 is the retirement date.
- $t_j = t_0 + j$.
- t_a (or $t^\star$) is the annuitization date.
- $\mathcal{U}(x)$ is the utility function.
- $w_0 = \alpha W_0$ is the withdrawal amount.
- W_t is the wealth at date t.
- W_{t^-} (or W_t^-) is the pre-withdrawal wealth.
- x is the age, measured in years.
- Y_t is the annuity income.
- $y_{x,t} = B/a(x,t)$ is the annuity payout yield.

B Data

B.1 Human mortality database

The HMD contains period and cohort mortality data for 41 countries: Australia, Austria, Belarus, Belgium, Bulgaria, Canada, Chile, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Korea, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan, the United Kingdom, the United States, and Ukraine. The available data variables are:

- Period data: birth counts, death counts, population estimates, exposure-to-risk, age-specific death rates, age-standardized death rates, life tables (male, female, and both sexes), and life expectancy at birth;
- Cohort data: exposure-to-risk, death rates, life tables (male, female, and both sexes), and life expectancy at birth.

The length of the data period varies considerably from one country to another. For instance, France has one of the longest series (1816–2023), while Germany has one of the shortest (1990–2020).

B.2 Data cleaning

B.2.1 Surface smoothing

For each category (female, male, both), a smoothed reference surface is built using valid non-open-ended cells. Rather than smoothing $\ln\left(D\left(x,t\right)/E\left(x,t\right)\right)$ directly, the procedure applies the offset transformation:

$$y\left(x,t\right)=\ln\left(\frac{D\left(x,t\right)+0.5}{E\left(x,t\right)}\right)$$

This avoids numerical problems when death counts are zero and stabilizes the log transformation at low counts. The smoothed reference surface is constructed in four successive steps:

1. Rolling median in age: We apply a centered 5-point moving median in age.
2. Rolling median in year: We apply a centered 5-point moving median in year.
3. Moving average in age: We apply a centered 3-point moving average in age.
4. Moving average in year: We apply a centered 3-point moving average in year.

We denote by $\hat{y}\left(t,x\right)$ the final smoothed log-mortality surface. Figure 67 shows the raw log-mortality rates $y\left(x,t\right)$, while Figure 68 displays the smoothed log-mortality rates $\hat{y}\left(x,t\right)$. We observe that $\hat{y}\left(x,t\right)$ is considerably more regular than $y\left(x,t\right)$. To better illustrate the smoothing procedure, we plot some specific slices of the surface in Figures 69 and 70. For example, we see that the smoothed log-mortality rate increases with age in 1970, which is not the case with the raw data. We also observe that the smoothed rates are lower than the raw rates in 1944. This is expected since mortality increased during the World War II. However, this does not mean that an upward trend was forming during these years. Rather, it reflects a temporary war-related spike that the smoothing procedure correctly attenuates. Similarly, we observe that log-mortality is relatively regular at ages 55 and 75, while it becomes more irregular at age 90 and highly noisy at age 100. This illustrates the value of the smoothing procedure from the perspective of trend extraction and forecasting.

Figure 67: Log-mortality surface (France, female)

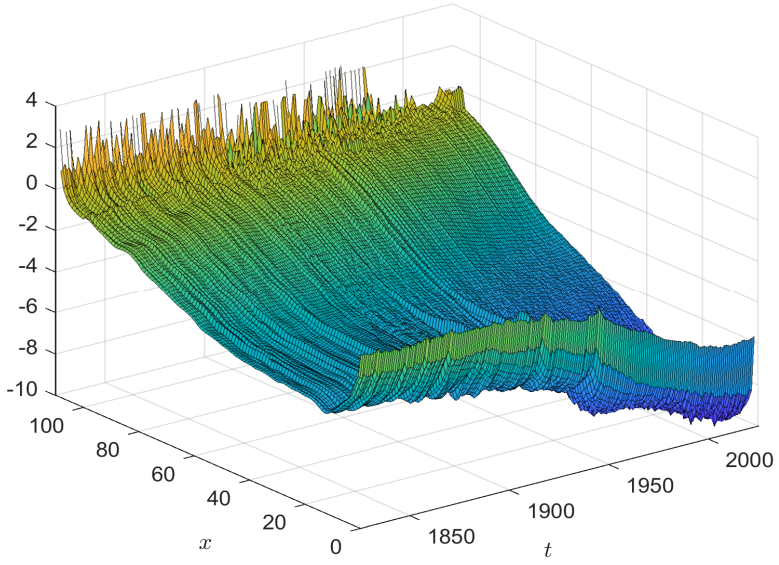


Figure 68: Smoothed log-mortality surface (France, female)

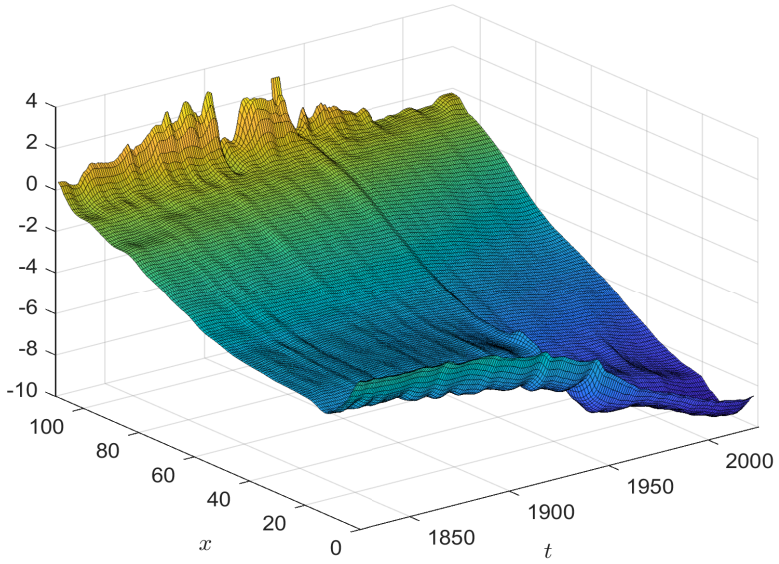


Figure 69: Log-mortality rates for the years 1850, 1944, 1970 and 2000 (France, female)

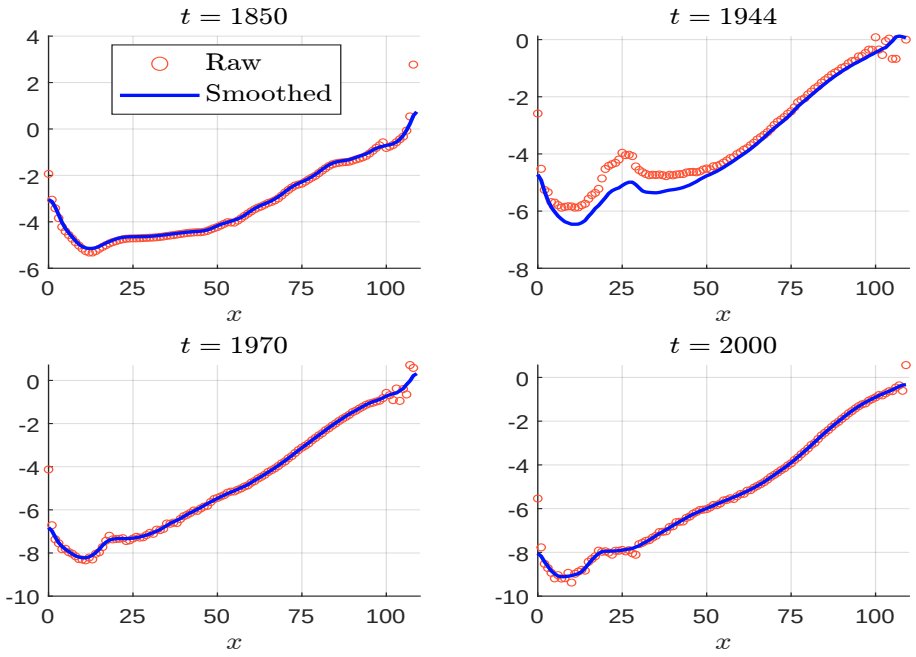
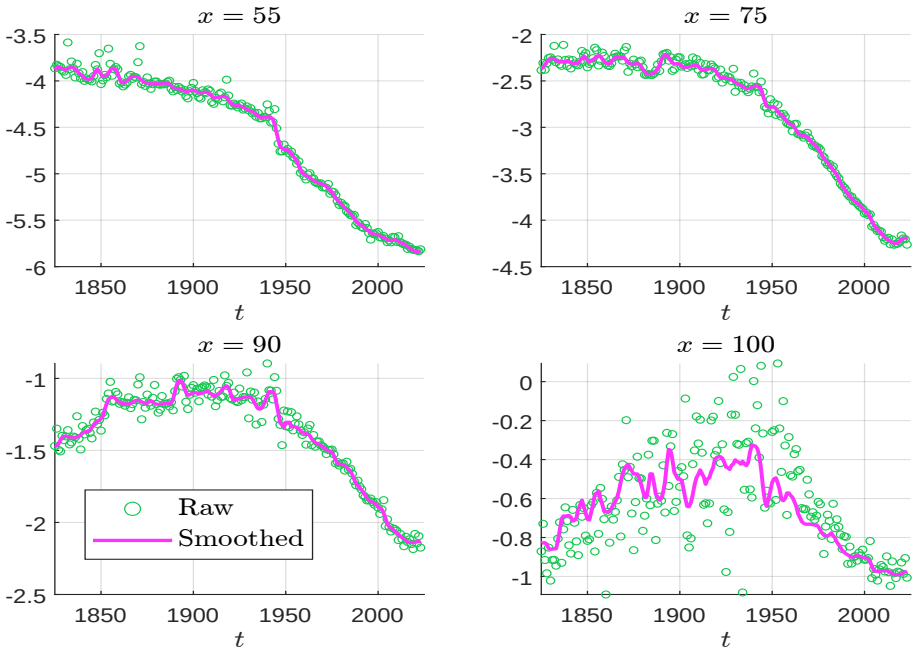


Figure 70: Log-mortality rates for the ages 55, 75, 90 and 100 (France, female)



B.2.2 Outlier detection and replacement

The associated fitted death count is $\hat{D}(x,t) = E(x,t)\, e^{\hat{y}(t,x)}$. Deviation from this reference surface is measured by the Poisson-style standardized residual:

$$r(t,x) = \frac{D(x,t) - \hat{D}(x,t)}{\sqrt{\hat{D}(x,t) + \varepsilon}}$$

where $\varepsilon = 10^{-8}$. Large values of $\left|r(t,x)\right|$ indicate that the observed death count differs substantially from the local reference value. However, residual size alone is not sufficient for flagging because sparse high-age data can generate large residuals through ordinary stochastic variation. Therefore, the cleaning procedure also measures local curvature in the observed log-mortality surface. The age and time curvatures are defined as $c_x(x,t) = y(x,t) - \frac{1}{2}\left(y(x-1,t) + y(x+1,t)\right)$ and $c_t(x,t) = y(x,t) - \frac{1}{2}\left(y(x,t-1) + y(x,t+1)\right)$. These second-difference measures identify cells whose log mortality is unusually sharp relative to neighbouring ages or years. Each curvature measure is standardized using the robust z-score:

$$z(x,t) = 0.6745 \frac{c(x,t) - \text{median}\, c(x,t)}{\text{MAD}\, c(x,t)}$$

The cleaning procedure combines the residual $r(t,x)$, the age-curvature z-score $z_x(x,t)$ and the time-curvature z-score $z_t(x,t)$ as follows:

- For each age x, the exposure floor is the 5th percentile of exposure among valid cells at that age. A cell is eligible for automatic flagging only if its exposure is greater than this floor.
- Among eligible cells, the baseline flagging rule[82] is $\left(\left|r(t,x)\right| > 4 \text{ and } \left|z_x(x,t)\right| > 3\right)$ or $\left(\left|r(t,x)\right| > 4 \text{ and } \left|z_t(x,t)\right| > 3\right)$.
- For flagged cells, the observed death count $D(x,t)$ is replaced by the fitted local value from the smoothed reference surface: $\hat{D}(x,t) = E(x,t)\, e^{\hat{y}(t,x)}$.

[82] For ages above 100, the thershold $\left|r(t,x)\right| > 4$ is replaced by $\left|r(t,x)\right| > 5$.

C Mathematical appendix

C.1 Calibration of the Lee-Carter model

As explained by Lee and Carter (1992), the model is underdetermined. Indeed, if $(\alpha_x, \beta_x, \kappa_t)$ is a solution, then $(\alpha_x - \beta_x c, \beta_x, \kappa_t + c)$ and $\left(\alpha_x, \beta_x c, \kappa_t / c\right)$ are also solutions for any constant c. This is why the parameters are generally standardized such that $\sum_{x=0}^{n-1} \beta_x = 1$ and $\sum_{t=1}^{T} \kappa_t = 0$. Lee and Carter (1992) propose using the low-rank approximation given by the Eckart-Young theorem to obtain the least squares solution. In this case, the solution is given by the following equations:

- The coefficient α_x is the mean over time of the log-mortality rates:

$$\alpha_x = \frac{1}{T} \sum_{t=1}^{T} \ln m\,(x,t)$$

- We perform the singular value decomposition of the $n \times T$ matrix A,whose entries are $A_{x,t} = \ln m\,(x,t) - \alpha_x$:

$$A = USV^\top$$

- The loading vector corresponds to the first left singular vector, and the time index is proportional to the first right singular vector: $\beta_x = U_{x+1,1}$ and $\kappa_t = S_{1,1} V_{t,1}$.
- The solution $(\alpha_x, \beta_x, \kappa_t)$ is then rescaled to $\left(\alpha_x, \beta_x / c, \kappa_t c\right)$ with $c = \sum_{x=0}^{n-1} \beta_x$.

The second approach suggested by Wilmoth (1993) is to consider directly the weighted least squares problem:

$$\begin{aligned} \hat{\theta} &= \arg\min \sum_{x=0}^{n-1} \sum_{t=1}^{T} w_{x,t} \left(\ln m\,(x,t) - \alpha_x - \beta_x \kappa_t\right)^2 \\ \text{s.t.} &\begin{cases} \sum_{x=0}^{n-1} \beta_x = 1 \\ \sum_{t=1}^{T} \kappa_t = 0 \end{cases} \end{aligned}$$

In this case, we can apply the iterative WLS algorithm:

1. We initialize the parameters as follows:

$$\begin{cases} \alpha_x = \dfrac{\sum_{t=1}^{T} w_{x,t} \ln m\,(x,t)}{\sum_{t=1}^{T} w_{x,t}} \\ \beta_x = 1/n \\ \kappa_t = 0 \end{cases}$$

2. We update the age-specific level, given β_x and κ_t fixed:

$$\alpha_x = \frac{\sum_{t=1}^{T} w_{x,t} \left(\ln m\,(x,t) - \beta_x \kappa_t\right)}{\sum_{t=1}^{T} w_{x,t}}$$

3. We update the time-varying index by regressing $\ln m\,(x,t) - \alpha_x$ on β_x, given α_x and β_x fixed:

$$\kappa_t = \frac{\sum_{x=0}^{n-1} w_{x,t} \beta_x \left(\ln m\,(x,t) - \alpha_x\right)}{\sum_{x=0}^{n-1} w_{x,t} \beta_x^2}$$

4. We update the age-specific loading by regressing $\ln m(x,t) - \alpha_x$ on κ_t, given α_x and κ_t fixed:
$$\beta_x = \frac{\sum_{t=1}^{T} w_{x,t}\kappa_t \left(\ln m(x,t) - \alpha_x\right)}{\sum_{t=1}^{T} w_{x,t}\kappa_t^2}$$

5. We normalize the parameters such that $\sum_{x=0}^{n-1}\beta_x = 1$ and $\sum_{t=1}^{T}\kappa_t = 0$. Then, we return to Step 2 and iterate until the convergence of the residual sum of squares $\text{RSS}(\theta) = \sum_{x=0}^{n-1}\sum_{t=1}^{T} w_{x,t}\left(\ln m(x,t) - \alpha_x - \beta_x\kappa_t\right)^2$.

C.2 The constant force approximation

Let x and t denote exact age and continuous calendar time. The force of mortality (or hazard rate) $\lambda(x,t)$ is defined as the instantaneous conditional rate of death:
$$\lambda(x,t) = \lim_{\Delta t \to 0} \frac{\Pr\left\{\text{death in } [t, t+\Delta t) \mid \text{alive at } t, \text{ age } x\right\}}{\Delta t}$$

To relate the central death rate to $\lambda(x,t)$, we define $N(x,t)\,\mathrm{d}x$ as the number of individuals alive at calendar time t whose exact age lies in $[x, x+\mathrm{d}x)$. The exposure to risk over the age-time cell $[x, x+1)\times[t, t+1)$ is then:
$$E(x,t) = \int_t^{t+1}\int_x^{x+1} N(y,s)\,\mathrm{d}y\,\mathrm{d}s$$

while the corresponding expected number of deaths over the same cell is:
$$\mathbb{E}\left[D(x,t)\right] = \int_t^{t+1}\int_x^{x+1} \lambda(y,s)\,N(y,s)\,\mathrm{d}y\,\mathrm{d}s$$

The population counterpart of the observed central death rate $m(x,t) = D(x,t)/E(x,t)$ is therefore:
$$\frac{\mathbb{E}\left[D(x,t)\right]}{E(x,t)} = \frac{\int_t^{t+1}\int_x^{x+1} \lambda(y,s)\,N(y,s)\,\mathrm{d}y\,\mathrm{d}s}{\int_t^{t+1}\int_x^{x+1} N(y,s)\,\mathrm{d}y\,\mathrm{d}s}$$

This shows that, at the population level, the central death rate is an exposure-weighted average of the force of mortality over the relevant age and calendar-year interval. Following Cairns *et al.* (2009), the constant force approximation assumes that the force of mortality is constant over each unit interval of integer age and calendar year: $\lambda(x+u, t+v) = \lambda(x,t)$ for $0 \le u, v < 1$. Substituting this approximation into the above expression gives:
$$\frac{\mathbb{E}\left[D(x,t)\right]}{E(x,t)} = \lambda(x,t)\,\frac{\int_t^{t+1}\int_x^{x+1} N(y,s)\,\mathrm{d}y\,\mathrm{d}s}{\int_t^{t+1}\int_x^{x+1} N(y,s)\,\mathrm{d}y\,\mathrm{d}s} = \lambda(x,t)$$

Thus, under the constant force approximation, the central death rate and the force of mortality are equal: $m(x,t) = \lambda(x,t)$. Under the same approximation, we can derive the one-year survival probability:
$$\begin{aligned} p(x,t) &= 1 - q(x,t) \\ &= \exp\left(-\int_0^1 \lambda(x+u, t+u)\,\mathrm{d}u\right) \\ &= \exp\left(-m(x,t)\right) \end{aligned}$$

The corresponding one-year death probability is therefore $q(x,t) = 1 - \exp\left(-m(x,t)\right)$.

C.3 Optimal weighting schemes for estimating the CBD model

Since $Y(x,t) = g\big(m(x,t)\big)$ where $g(m) = \ln(e^m - 1)$, we have:

$$g'(m) = \frac{e^m}{e^m - 1} = \frac{1}{1 - e^{-m}}$$

and[83]:

$$\text{var}\big(Y(x,t)\big) \approx \left(\frac{1}{1 - e^{-m(x,t)}}\right)^2 \text{var}\big(m(x,t)\big) = \frac{\text{var}\big(m(x,t)\big)}{q^2(x,t)}$$

The optimal weights are then equal to:

$$w(x,t) = \frac{1}{\text{var}\big(Y(x,t)\big)} = \frac{q^2(x,t)}{\text{var}\big(m(x,t)\big)}$$

Recall that the Poisson LC model assumes $D(x,t) \sim \mathcal{P}\big(\mu(x,t)\big)$. It follows that:

$$\text{var}\big(m(x,t)\big) = \text{var}\left(\frac{D(x,t)}{E(x,t)}\right) = \frac{\mu(x,t)}{E^2(x,t)}$$

We deduce that:

$$w_{\mathcal{P}}(x,t) = \frac{q^2(x,t)\,E^2(x,t)}{\mu(x,t)} \tag{29}$$

For the negative binomial LC model, $D(x,t) \sim \mathcal{NB}\left(\frac{1}{\delta}, \frac{\delta\mu(x,t)}{1+\delta\mu(x,t)}\right)$. It follows that:

$$\text{var}\big(m(x,t)\big) = \frac{\mu(x,t) + \delta\mu^2(x,t)}{E^2(x,t)}$$

and:

$$w_{\mathcal{NB}}(x,t) = \frac{q^2(x,t)\,E^2(x,t)}{\mu(x,t) + \delta\mu^2(x,t)} \tag{30}$$

Equations (29) and (30) require estimating $\mu(x,t) = E(x,t)\,\mathbb{E}\big[m(x,t)\big]$. By replacing $\mu(x,t)$ by the empirical estimator $\hat{\mu}(x,t) = E(x,t)\,m(x,t)$, we obtain:

$$w_{\mathcal{P}}(x,t) = \frac{q^2(x,t)\,E(x,t)}{m(x,t)} \tag{31}$$

and:

$$w_{\mathcal{NB}}(x,t) = \frac{q^2(x,t)\,E(x,t)}{m(x,t)\big(1 + \delta E(x,t)\,m(x,t)\big)} \tag{32}$$

[83] Let X and Y be two random variables such that $Y = g(X)$. We have:

$$\text{var}(Y) \approx \left(\frac{\partial\, g\big(\mathbb{E}[X]\big)}{\partial\, x}\right)^2 \text{var}(X)$$

C.4 CBD outlier detection

An additional outlier detection procedure is applied to the CBD model because the mortality profile is estimated separately for each year as a straight line in age. Therefore, an isolated observation can shift the estimated level $\kappa_t^{(1)}$ or distort the estimated slope $\kappa_t^{(2)}$, particularly at extreme ages where regression leverage is highest. The procedure combines two complementary criteria. The first criterion identifies observations inconsistent with the fitted age profile in year t, based on studentized residuals:

$$r(x,t) = \sqrt{\frac{w(x,t)}{\hat{\sigma}_t^2\left(1-h(x,t)\right)}}\left(Y(x,t) - \hat{\kappa}_t^{(1)} - \hat{\kappa}_t^{(2)}(x-\bar{x})\right)$$

where $\hat{\sigma}_t^2$ is the estimated residual variance of the annual CBD regression, and $h(x,t)$ is the diagonal element, corresponding to age x, of the hat matrix[84] $\mathbf{H}_t$, which measures the leverage of an observation in the annual regression. A cell is flagged when $\left|r(x,t)\right| > 3.5$. The second criterion uses the time-curvature robust z-score $z(x,t)$, which is defined in the initial cleaning procedure. A cell is provisionally flagged when $\left|z(x,t)\right| > 4.5$. To preserve genuine systematic mortality shocks, these flags are removed when more than 10% of observations in the same year satisfy this criterion. Among cells entering the preliminary CBD calibration, the flagging rule is $\left|r(x,t)\right| > 3.5$ or $\left|z(x,t)\right| > 4.5$, subject to the systematic-year safeguard described above. Flagged cells receive zero weight when the CBD model is re-estimated. For the equal-weight CBD calibration of the French data for ages 60–109 and years 1816–2023, the studentized residual and time-curvature criteria flag 186 and 48 female cells, respectively, with 17 cells common to both rules, giving 217 distinct outliers. For males, the corresponding figures are 197, 43 and 20, giving 220 distinct outliers.

C.5 Gompertz annuity pricing model (GAPM)

Milevsky (2013) provides a closed-form annuity expression under Gompertz mortality law. We note that the valuation is actuarial and not risk-neutral. Under the Gompertz law, the survival probability from age x to age $x+t$ is given by:

$$S(x,t) = \exp\left(e^{(x-m)/b}\left(1-e^{t/b}\right)\right)$$

where m is the modal age at death and $b>0$ is the dispersion parameter. Let $A_x = e^{(x-m)/b}$. The value of a continuous annuity is:

$$\bar{a}(x) = \int_0^\infty e^{-\varrho t}\exp\left(A_x\left(1-e^{t/b}\right)\right)\mathrm{d}t$$

Using the change of variable $y = A_x e^{t/b}$, we deduce that $t = b\ln\dfrac{y}{A_x}$, $\mathrm{d}t = \dfrac{b}{y}\mathrm{d}y$ and:

$$\begin{aligned}
\bar{a}(x) &= \int_{A_x}^\infty \left(\frac{y}{A_x}\right)^{-\varrho b}\exp(A_x - y)\frac{b}{y}\mathrm{d}y \\
&= be^{A_x}A_x^{\varrho b}\int_{A_x}^\infty y^{-\varrho b-1}e^{-y}\mathrm{d}y \\
&= be^{A_x}A_x^{\varrho b}\Gamma(-\varrho b, A_x)
\end{aligned}$$

[84] Let $\mathbf{X}_t$ denote the design matrix whose row corresponding to age x is $(1, x-\bar{x})$, and let W_t denote the diagonal matrix of weights $w(x,t)$. The hat matrix is $\mathbf{H}_t = W_t^{1/2}\mathbf{X}_t\left(\mathbf{X}_t^\top W_t\mathbf{X}_t\right)^{-1}\mathbf{X}_t^\top W_t^{1/2}$.

where $\Gamma(\alpha, x)$ is the upper incomplete gamma function. For a continuous annuity with a guaranteed payment period g, the value is the sum of a certain annuity over $(0, g)$ and a deferred survival-contingent annuity thereafter:

$$\begin{aligned}
\bar{a}(x, g) &= \int_0^g e^{-\varrho t}\,\mathrm{d}t + \int_g^\infty e^{-\varrho t} S(x,t)\,\mathrm{d}t \\
&= \frac{1-e^{-\varrho g}}{\varrho} + e^{-\varrho g}\int_0^\infty e^{-\varrho u} S(x, u+g)\,\mathrm{d}u \\
&= \frac{1-e^{-\varrho g}}{\varrho} + e^{-\varrho g} b e^{A_x} A_{x+g}^{\varrho b}\Gamma\left(-\varrho b, A_{x+g}\right) \\
&= \frac{1-e^{-\varrho g}}{\varrho} + b e^{A_x} A_x^{\varrho b}\Gamma\left(-\varrho b, A_{x+g}\right) \\
&= \frac{1-e^{-\varrho g}}{\varrho} + \frac{b\Gamma\left(-\varrho b, \exp\left((x+g-m)/b\right)\right)}{\exp\left(-\varrho(x-m) - \exp\left((x-m)/b\right)\right)}
\end{aligned}$$

The corresponding discrete annuity is:

$$\bar{a}(x, g) = \sum_{t=1}^{g} \frac{1}{(1+\varrho)^t} + \sum_{t=g+1}^{\omega-x} \frac{p(x,t)}{(1+\varrho)^t} \qquad \text{for } 0 \le g \le \omega - x$$

where ω is the limiting age.

C.6 Joint scenario simulation of the Lee-Carter model parameters and mortality innovations

The Lee-Carter model is defined by the following regression model:

$$\ln m(x,t) = \alpha_x + \beta_x \kappa_t + \varepsilon(x,t) \tag{33}$$

The matrix form of Equation (33) is:

$$y = \alpha \mathbf{1}_T^\top + \beta \kappa^\top + \varepsilon$$

where $y = \left(y_{x,t}\right)$ is a $n \times T$ matrix with elements $y_{x,t} = \ln m(x,t)$, $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$, $\beta = (\beta_0, \ldots, \beta_{n-1})$, $\kappa = (\kappa_1, \ldots, \kappa_T)$ and $\varepsilon = \left(\varepsilon_{x,t}\right)$ is a $n \times T$ matrix with elements $\varepsilon_{x,t} = \varepsilon(x,t)$. We obtain[85]:

$$\begin{aligned}
\mathrm{vec}(y) &= \mathrm{vec}\left(\alpha \mathbf{1}_T^\top + \beta \kappa^\top + \varepsilon\right) \\
&= (\mathbf{1}_T \otimes I_n)\,\alpha + (\kappa \otimes I_n)\,\beta + \mathrm{vec}(\varepsilon) \\
&= \begin{pmatrix} (\mathbf{1}_T \otimes I_n) & (\kappa \otimes I_n) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \mathrm{vec}(\varepsilon)
\end{aligned}$$

or:

$$Y = X\theta + U$$

[85] We use the property $\mathrm{vec}\left(ABC^\top\right) = (C \otimes A)\,\mathrm{vec}(B)$.

where $Y = \text{vec}(y)$, $X = \begin{pmatrix} (\mathbf{1}_T \otimes I_n) & (\kappa \otimes I_n) \end{pmatrix}$, $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and $U = \text{vec}(\varepsilon)$. Since $\varepsilon(x,t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, it follows that $\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma_\varepsilon^2 \left(X^\top X\right)^{-1}\right)$ where:

$$\left(X^\top X\right)^{-1} = \begin{pmatrix} \frac{1}{T} I_n & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \frac{1}{\sum_{t=1}^T \kappa_t^2} I_n \end{pmatrix}$$

because of the identification constraint[86] $\sum_{t=1}^T \kappa_t = 0$. Let $\left(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\right)$ be the estimated parameters. We then have $\hat{U} = Y - \hat{X}\hat{\theta}$ where $\hat{X} = \begin{pmatrix} (\mathbf{1}_T \otimes I_n) & (\hat{\kappa} \otimes I_n) \end{pmatrix}$ and $\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$. We deduce that[87]:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{U}^\top \hat{U}}{nT - 2n - 2}$$

We obtain the following scenario generator:

1. Consider the estimated values $\left(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\right)$ and compute $\hat{X}$, $\hat{\theta}$, $\hat{U}$ and $\hat{\sigma}_\varepsilon^2$.

2. Generate scenarios of the age-specific parameters $\left(\alpha_x^{(s)}, \beta_x^{(s)}\right)$ as follows:

$$\begin{pmatrix} \alpha^{(s)} \\ \beta^{(s)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \hat{\sigma}_\varepsilon^2 \left(\hat{X}^\top \hat{X}\right)^{-1}\right)$$

   and standardize the values of $\beta_x^{(s)}$.

3. Simulate scenarios of the mortality innovations $\varepsilon^{(s)}(x, t+h)$ such that:

$$\varepsilon^{(s)}(x, t+h) \sim \mathcal{N}\left(0, \hat{\sigma}_\varepsilon^2\right)$$

4. Generate scenarios of the time index $\kappa_{t+h}^{(s)}$ using the ARIMA(1,1,0) process.

Remark 20 *This Monte Carlo generator implicitly assumes that $(\alpha_x, \beta_x, \kappa_t)$ and $\varepsilon(x,t)$ are independent, which holds by construction. It also assumes that (α_x, β_x) and κ_t are independent, an assumption that is not theoretically guaranteed.*

[86] We have:

$$X^\top X = \begin{pmatrix} \left(\mathbf{1}_T^\top \otimes I_n\right)(\mathbf{1}_T \otimes I_n) & \left(\mathbf{1}_T^\top \otimes I_n\right)(\kappa \otimes I_n) \\ \left(\kappa^\top \otimes I_n\right)(\mathbf{1}_T \otimes I_n) & \left(\kappa^\top \otimes I_n\right)(\kappa \otimes I_n) \end{pmatrix} = \begin{pmatrix} \left(\mathbf{1}_T^\top \mathbf{1}_T\right) \otimes I_n & \left(\mathbf{1}_T^\top \kappa\right) \otimes I_n \\ \left(\kappa^\top \mathbf{1}_T\right) \otimes I_n & \left(\kappa^\top \kappa\right) \otimes I_n \end{pmatrix}$$

Since $\sum_{t=1}^T \kappa_t = 0$, $X^\top X$ is block diagonal:

$$X^\top X = \begin{pmatrix} T \cdot I_n & \left(\sum_{t=1}^T \kappa_t\right) I_n \\ \left(\sum_{t=1}^T \kappa_t\right) I_n & \left(\sum_{t=1}^T \kappa_t^2\right) I_n \end{pmatrix} = \begin{pmatrix} T \cdot I_n & 0 \\ 0 & \left(\sum_{t=1}^T \kappa_t^2\right) I_n \end{pmatrix}$$

[87] The values of $\hat{\kappa}_t$ are assumed to be known with certainty.

C.7 Expression of the best estimate of the portfolio liability

Let $D_{i,g,h} = \mathbb{1}\{\boldsymbol{\tau}_{i,g} > t + h\}$ be the indicator variable equal to 1 if the i^{th} individual in group g survives to payment horizon h, and 0 otherwise. The total number of survivors in group g at horizon h is:

$$N_{g,h} = \sum_{i=1}^{N_g} D_{i,g,h}$$

Taking the conditional expectation of $N_{g,h}$ given $\mathcal{F}_t$ yields:

$$\mathbb{E}\left[N_{g,h} \mid \mathcal{F}_t\right] = \sum_{i=1}^{N_g} \mathbb{E}\left[D_{i,g,h} \mid \mathcal{F}_t\right] = \sum_{i=1}^{N_g} p\left(x_g, t, h\right) = N_g p\left(x_g, t, h\right)$$

where $p\left(x_g, t, h\right)$ is the probability that an individual aged x_g at the valuation date t survives to the payment horizon h. It follows that:

$$\begin{aligned}
\bar{L}\left(t\right) &= \mathbb{E}\left[L\left(t\right) \mid \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\sum_{g=1}^{n_G}\sum_{h=1}^{H_g} \frac{B}{\left(1+\varrho_t\right)^h} N_{g,h} \,\middle|\, \mathcal{F}_t\right] \\
&= \sum_{g=1}^{n_G}\sum_{h=1}^{H_g} \frac{B}{\left(1+\varrho_t\right)^h} \mathbb{E}\left[N_{g,h} \mid \mathcal{F}_t\right] \\
&= \sum_{g=1}^{n_G} N_g \left(B \sum_{h=1}^{H_g} \frac{p\left(x_g, t, h\right)}{\left(1+\varrho_t\right)^h}\right) \\
&= \sum_{g=1}^{n_G} N_g a\left(x_g, t\right)
\end{aligned}$$

C.8 Computation of the conditional variance $\text{var}^{(s)}\left(x_g, t\right)$

Let $\text{var}^{(s)}\left(x_g, t\right)$ be the conditional variance of a single contract in group g under scenario s. We denote by $\boldsymbol{\tau}_g$ the survival time of a representative annuitant in the g^{th} group. Using the contract present value in Equation (8) on page 40, we have:

$$\text{var}^{(s)}\left(x_g, t\right) = \text{var}\left(A\left(x_g, t\right) \mid \mathcal{F}_t, \mathcal{P}^{(s)}\right)$$

where:

$$A\left(x_g, t\right) = \sum_{h=1}^{H_g} \frac{B}{\left(1+\varrho_t\right)^h} \mathbb{1}\left\{\boldsymbol{\tau}_g > t + h\right\}$$

To evaluate this variance, we first consider the conditional second moment. We have:

$$\begin{aligned}
A^2\left(x_g, t\right) &= \sum_{h=1}^{H_g}\sum_{k=1}^{H_g} \frac{B^2}{\left(1+\varrho_t\right)^{h+k}} \mathbb{1}\left\{\boldsymbol{\tau}_g > t + h\right\} \mathbb{1}\left\{\boldsymbol{\tau}_g > t + k\right\} \\
&= \sum_{h=1}^{H_g}\sum_{k=1}^{H_g} \frac{B^2}{\left(1+\varrho_t\right)^{h+k}} \mathbb{1}\left\{\boldsymbol{\tau}_g > t + \max\left(h, k\right)\right\}
\end{aligned}$$

It follows that:

$$\begin{aligned}\mathbb{E}\left[A^2\left(x_g,t\right)\mid\mathcal{P}^{(s)}\right] &= \sum_{h=1}^{H_g}\sum_{k=1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{h+k}}\mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\tau}_g > t+\max\left(h,k\right)\right\}\mid\mathcal{P}^{(s)}\right]\\ &= \sum_{h=1}^{H_g}\sum_{k=1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{h+k}}p^{(s)}\left(x_g,t,\max\left(h,k\right)\right)\\ &= \sum_{h=1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{2h}}p^{(s)}\left(x_g,t,h\right)+2\sum_{h=1}^{H_g-1}\sum_{k=h+1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{h+k}}p^{(s)}\left(x_g,t,k\right)\end{aligned}$$

Since $\mathbb{E}\left[A\left(x_g,t\right)\mid\mathcal{P}^{(s)}\right]=a^{(s)}\left(x_g,t\right)$, the expression of the conditional variance is:

$$\begin{aligned}\mathrm{var}^{(s)}\left(x_g,t\right) &= \mathbb{E}\left[A^2\left(x_g,t\right)\mid\mathcal{F}_t,\mathcal{P}^{(s)}\right]-\mathbb{E}^2\left[A\left(x_g,t\right)\mid\mathcal{F}_t,\mathcal{P}^{(s)}\right]\\ &= \sum_{h=1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{2h}}p^{(s)}\left(x_g,t,h\right)+2\sum_{h=1}^{H_g-1}\sum_{k=h+1}^{H_g}\frac{B^2}{\left(1+\varrho_t\right)^{h+k}}p^{(s)}\left(x_g,t,k\right)-\\ &\quad\left(a^{(s)}\left(x_g,t\right)\right)^2\end{aligned}$$

C.9 Probability mass of the death-time distribution under a static mortality table

Recall that $q\left(x,t\right)$ denotes the one-year death probability. Under a static mortality table, we have $q\left(x,t\right)=q\left(x,t_0\right)$ for all relevant ages x and calendar dates t. For an individual aged x at the initial date t_0, the probability of dying during the interval $(t,t+1]$, conditional on being alive at t_0, is:

$$\mathrm{PD}_t = p\left(x,t_0,t-t_0\right)q\left(x+t-t_0,t_0\right)$$

Let $h=t-t_0$. Since $p\left(x,t_0,h+1\right)=p\left(x,t_0,h\right)\left(1-q\left(x+h,t_0\right)\right)$, it follows that:

$$\mathrm{PD}_t = p\left(x,t_0,h\right)-p\left(x,t_0,h+1\right)$$

For any $H\geq 1$, the sum of probability of dying is $\sum_{t=t_0}^{t_0+H}\mathrm{PD}_t = 1-p\left(x,t_0,H+1\right)$. Taking $H\to\infty$, we obtain:

$$\sum_{t=t_0}^{\infty}\mathrm{PD}_t = 1-\lim_{H\to\infty}p\left(x,t_0,H+1\right)$$

and:

$$\sum_{t=t_0}^{\infty}\mathrm{PD}_t = 1 \Leftrightarrow \lim_{H\to\infty}p\left(x,t_0,H\right)=0$$

Thus, the death probabilities form a proper probability distribution when the probability of surviving indefinitely is zero. In practice, however, mortality tables are only available up to a maximum age x^+. For an individual aged x at time t_0, the last one-year mortality interval represented in the table begins at age x^+. Setting $H=x^+-x$, the finite-horizon identity becomes:

$$\sum_{t=t_0}^{t_0+x^+-x}\mathrm{PD}_t + p\left(x,t_0,x^+-x+1\right)=1$$

Moreover, we have:

$$p\left(x, t_0, x^+ - x + 1\right) = \prod_{y=x}^{x^+} \left(1 - q(y, t_0)\right)$$

The final term is the probability that the individual survives all of the one-year mortality intervals represented in the table. In other words, it is the probability mass associated with death dates that occur after the maximum age covered by the available mortality table.

C.10 Computation of the lifetime consumption component $\mathcal{C}^{(s)}(n)$

For a finite annuitization date $t_0 + n$, the discounted lifetime consumption component is:

$$\mathcal{C}^{(s)}(n) = \sum_{j=1}^{H} \frac{p^{(s)}(x, t_0, j)}{(1 + \tilde{\varrho}_{t_0})^j} \boldsymbol{c}_{t_j}^{(s)} \tag{34}$$

where $t_j = t_0 + j$, $p^{(s)}(x, t_0, j)$ is the survival probability from age x at t_0 to age $x + j$ along mortality path s, and $\boldsymbol{c}_{t_0+j}^{(s)}$ is actual consumption at date $t_0 + j$. Before and at the annuitization date, consumption is financed by withdrawals from the liquid account, whereas after annuitization it is equal to annual annuity income:

$$\boldsymbol{c}_{t_j}^{(s)} = \begin{cases} \min\left(w_{t_j}, W_{t_j^-}^{(s)}\right) & \text{for } j \leq n \\ Y_{t_n}^{(s)} & \text{for } j > n \end{cases}$$

where $t_n = t_0 + n$. It follows that:

$$\begin{aligned} \mathcal{C}^{(s)}(n) &= \sum_{j=1}^{n} \frac{p^{(s)}(x, t_0, j)}{(1 + \tilde{\varrho}_{t_0})^j} \boldsymbol{c}_{t_j}^{(s)} + \sum_{j=n+1}^{H} \frac{p^{(s)}(x, t_0, j)}{(1 + \tilde{\varrho}_{t_0})^j} Y_{t_n}^{(s)} \\ &= \sum_{j=1}^{n} \frac{p^{(s)}(x, t_0, j)}{(1 + \tilde{\varrho}_{t_0})^j} \boldsymbol{c}_{t_j}^{(s)} + \frac{Y_{t_n}^{(s)}}{(1 + \tilde{\varrho}_{t_0})^n} \sum_{h=1}^{H-n} \frac{p^{(s)}(x, t_0, n + h)}{(1 + \tilde{\varrho}_{t_0})^h} \end{aligned}$$

Survival from age x at t_0 to age $x+n+h$ can be decomposed into survival to the annuitization date and conditional survival thereafter:

$$p^{(s)}(x, t_0, n + h) = p^{(s)}(x, t_0, n)\, p^{(s)}(x + n, t_0 + n, h)$$

We deduce that:

$$\begin{aligned} \mathcal{C}^{(s)}(n) &= \underbrace{\sum_{j=1}^{n} \frac{p^{(s)}(x, t_0, j)}{(1 + \tilde{\varrho}_{t_0})^j} \boldsymbol{c}_{t_j}^{(s)}}_{\text{Withdrawals}} \\ &\quad + \underbrace{\frac{p^{(s)}(x, t_0, n)}{(1 + \tilde{\varrho}_{t_0})^n} Y_{t_0+n}^{(s)} \left(\sum_{h=1}^{H-n} \frac{p^{(s)}(x + n, t_0 + n, h)}{(1 + \tilde{\varrho}_{t_0})^h} \right)}_{\text{Post-annuitization consumption}} \end{aligned} \tag{35}$$

The term in brackets is the actuarial value, at the annuitization date, of one unit of annual post-annuitization consumption when future payments are discounted using the discount rate entering the retiree's objective.

When the objective discount rate coincides with the annuity pricing rate ($\varrho_{t_0} = \tilde{\varrho}_{t_0}$), the term in brackets is exactly the annuity pricing factor:

$$\sum_{h=1}^{H-n} \frac{p^{(s)}\left(x+n, t_0+n, h\right)}{\left(1+\tilde{\varrho}_{t_0}\right)^h} = \sum_{h=1}^{H-n} \frac{p^{(s)}\left(x+n, t_0+n, h\right)}{\left(1+\varrho_{t_0}\right)^h} = a^{(s)}\left(x+n, t_0+n\right)$$

Since we have:

$$Y_{t_0+n}^{(s)} a^{(s)}\left(x+n, t_0+n\right) = \frac{W_{t_0+n}^{(s)}}{1+\mathcal{L}}$$

the lifetime consumption component simplifies to:

$$\mathcal{C}^{(s)}(n) = \underbrace{\sum_{j=1}^{n} \frac{p^{(s)}\left(x, t_0, j\right)}{\left(1+\varrho_{t_0}\right)^j} \boldsymbol{c}_{t_j}^{(s)}}_{\text{Withdrawals}} + \underbrace{\frac{p^{(s)}\left(x, t_0, n\right)}{\left(1+\varrho_{t_0}\right)^n} \frac{W_{t_n}^{(s)}}{1+\mathcal{L}}}_{\text{Net wealth}} \tag{36}$$

The first term values pre-annuitization withdrawals, weighted by the survival probability to each withdrawal date. The second term values post-annuitization consumption, conditional on survival to the conversion date and discounted back to t_0 (Equation (35)). When the annuity-pricing and objective discount rates coincide, this second term reduces to discounted net wealth after loading (Equation (36)).

C.11 Survival time approach of the fixed-date annuitization problem

Let $\boldsymbol{\tau}^{(s)}$ be the survival time conditional on the s^{th} simulated scenario. The associated survival function $\mathbf{S}^{(s)}\left(t_j\right)$ is:

$$\mathbf{S}^{(s)}\left(t_j\right) = p^{(s)}\left(x, t_0, j\right)$$

In particular, we have $\mathbf{S}^{(s)}\left(t_0\right) = p^{(s)}\left(x, t_0, 0\right) = 1$, $\Pr\left\{t_0 + j - 1 < \boldsymbol{\tau}^{(s)} \leq t_0 + j \mid \mathcal{F}^{(s)}\right\} = \mathbf{S}^{(s)}\left(t_{j-1}\right) - \mathbf{S}^{(s)}\left(t_j\right) = p^{(s)}\left(x, t_0, j-1\right) - p^{(s)}\left(x, t_0, j\right)$ and $\Pr\left\{\boldsymbol{\tau}^{(s)} > t_0 + H \mid \mathcal{F}^{(s)}\right\} = \mathbf{S}^{(s)}\left(t_H\right) = p^{(s)}\left(x, t_0, H\right)$. For each mortality path s, we generate n_R independent realizations $\boldsymbol{\tau}^{(s,r)}$ of the future lifetime using the probability integral transform. We then compute the (s,r)-value of the utility function. For $1 \leq n \leq H-1$, we have:

$$\begin{aligned}
\mathcal{U}^{(s,r)}(n) \quad = \quad & \sum_{j=1}^{n} \frac{\boldsymbol{c}_{t_j}^{(s)}}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t_j\right\} + \\
& \sum_{j=n+1}^{H} \frac{Y_{t_0+n}^{(s)}}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t_j\right\} + \\
& \psi \sum_{j=1}^{n} \frac{W_{t_j^-}^{(s)}}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \mathbb{1}\left\{t_{j-1} < \boldsymbol{\tau}^{(s,r)} \leq t_j\right\}
\end{aligned}$$

$n = 0$ and $n = \infty$ are two special cases:

$$\mathcal{U}^{(s,r)}(0) = \sum_{j=1}^{H} \frac{Y_{t_0}^{(s)}}{\left(1+\tilde{\varrho}_{t_0}\right)^j} \mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t_j\right\}$$

and:

$$\begin{aligned}\mathcal{U}^{(s,r)}(\infty) &= \sum_{j=1}^{H} \frac{\boldsymbol{c}_{t_j}^{(s)}}{(1+\tilde{\varrho}_{t_0})^j} \mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t_j\right\} + \\ &\quad \psi \sum_{j=1}^{H} \frac{W_{t_j^-}^{(s)}}{(1+\tilde{\varrho}_{t_0})^j} \mathbb{1}\left\{t_{j-1} < \boldsymbol{\tau}^{(s,r)} \leq t_j\right\} + \\ &\quad \psi \frac{W_{t_H}^{(s)}}{(1+\tilde{\varrho}_{t_0})^H} \mathbb{1}\left\{\boldsymbol{\tau}^{(s,r)} > t_H\right\}\end{aligned}$$

Under this setting, the optimization problem becomes:

$$n^\star = \underset{n\in\{0,\ldots,H-1\}\cup\{\infty\}}{\arg\max} \frac{1}{n_S \times n_R} \sum_{s=1}^{n_S}\sum_{r=1}^{n_R} \mathcal{U}^{(s,r)}(n)$$

This formulation replaces survival-probability weighting by explicit simulation of individual death times. As n_R increases, the estimator converges to the expected utility conditional on each mortality scenario, thus providing a Monte Carlo approximation of the original optimization problem.

C.12 Numerical solution of the dynamic annuitization problem

The dynamic problem is defined by the state $X_{t_j} = \left(W_{t_j}, \kappa_{t_j}, \Delta\kappa_{t_j}\right)$. Since these variables are continuous, the Bellman equation is evaluated using date-specific numerical grids. The mortality grid is obtained from the conditional distribution of the ARIMA state, and the wealth grid is constructed from the distribution generated by remaining in drawdown. First, we describe these grids, and then we describe the numerical evaluation of the conditional expectations that enter the Bellman equation.

C.12.1 State-space discretization

Mortality state Under the fitted ARIMA(1,1,0) model, the mortality transition can be written as[88]:

$$\Theta_{j+1} = A_\kappa \Theta_j + c_\kappa + u_\kappa \varepsilon_{j+1}$$

where $\Theta_j := \Theta_{t_j} = \left(\kappa_{t_j}, \Delta\kappa_{t_j}\right)$, $\varepsilon_{j+1} \sim \mathcal{N}\left(0, \hat{\sigma}_\kappa^2\right)$, and A_κ, c_κ, and u_κ are determined by the estimated ARIMA parameters. Since the mortality state is known at t_0 (the retirement date) and the innovations are Gaussian, the state Θ_j remains bivariate Gaussian at every subsequent date. Writing $m_j = \mathbb{E}\left[\Theta_j\right]$ and $\Sigma_j = \text{var}\left(\Theta_j\right)$, the first two moments satisfy:

$$\begin{cases} m_{j+1} = A_\kappa m_j + c_\kappa \\ \Sigma_{j+1} = A_\kappa \Sigma_j A_\kappa^\top + u_\kappa \hat{\sigma}_\kappa^2 u_\kappa^\top \end{cases}$$

with $m_0 = (\kappa_{t_0}, \Delta\kappa_{t_0})$ and $\Sigma_0 = \mathbf{0}_{2\times 2}$. These moments determine the mortality domain at each date. We use a rectangular region containing 99.5% of the joint Gaussian distribution. This region is represented by 13 equally spaced values of κ_{t_j} and 11 values of $\Delta\kappa_{t_j}$. At $j = 0$ (or $t_j = t_0$), the mortality state is observed and the grid reduces to the observed point.

[88] Using the index j is equivalent to using the time index $t_j = t_0 + j$ since we are in discrete time. Moreover, it directly indicates the row number of vectors or matrices built to solve the Bellman equation.

Wealth state The wealth domain is based on the distribution obtained when the individual remains in income drawdown throughout the horizon. Annuitization removes a path from the liquid state space, therefore this always-liquid distribution contains the necessary wealth states for the stopping problem. Let $\mathbf{F}_j(w)$ denote the distribution of post-withdrawal wealth W_{t_j}, and define the one-period transition kernel $\mathcal{K}_j(\omega \mid w) = \Pr\{W_{t_{j+1}} \leq \omega \mid W_{t_j} = w\}$. Using the definition of the wealth transition, we have:

$$\mathcal{K}_j(\omega \mid w) = \Phi\left(\frac{(\omega + w_0)/w - 1 - \mu}{\sigma}\right)$$

where $w_0 = \alpha W_0$. The probability distribution at date t_{j+1} follows from the Chapman-Kolmogorov equation:

$$\begin{cases} \mathbf{F}_{j+1}(\omega) = p_{j,0} + \int_0^\infty \mathcal{K}_j(\omega \mid w)\, \mathrm{d}\mathbf{F}_j(w) \\ p_{j,0} = \Pr\{W_{t_j} = 0\} \end{cases}$$

The first term accounts for wealth paths that have already reached zero. This remains an absorbing state while the individual is in drawdown. When $\sigma = 0$, the wealth transition is deterministic. The positive part of the distribution is propagated numerically on a finite-volume grid. If $W_{j,m}$ denotes the representative wealth in cell m, and $p_{j,m}$ its probability mass, then the Chapman-Kolmogorov equation is approximated by:

$$\hat{\mathbf{F}}_{j+1}(\omega) = p_{j,0} + \sum_{m=1}^{n_{cell}} p_{j,m} \mathcal{K}_j(\omega \mid W_{j,m})$$

We use $n_{cell} = 800$ cells. At each date, the upper limit of the wealth domain used in the Bellman equation is the 99.95% quantile of $\hat{\mathbf{F}}_j$, and the value function is stored on 61 equally spaced wealth points between zero and this upper limit.

C.12.2 Numerical evaluation of the Bellman equation

The Bellman recursion involves two numerical expectations. The first expectation is used to calculate the conditional annuity factors, while the second expectation gives the value of remaining liquid for one more year.

Conditional annuity factors The conditional pricing factor is calculated backward on the mortality grid. Let $\{z_m, \varpi_m\}_{m=1}^q$ denote the nodes and weights of a standard normal Gaussian quadrature rule. For a current mortality state $(\kappa, \Delta\kappa)$, node m gives the next mortality state $(\kappa_{j+1,m}, \Delta\kappa_{j+1,m})$ and the corresponding one-year survival probability $\mathbf{S}_{j+1,m}$. The pricing-factor recursion is approximated by:

$$a_j(\kappa, \Delta\kappa) = \frac{1}{1 + \varrho_{t_0}} \sum_{m=1}^{q} \varpi_m \mathbf{S}_{j+1,m} \left(1 + Y_{j+1} a_{j+1}(\kappa_{j+1,m}, \Delta\kappa_{j+1,m})\right)$$

where Y_{j+1} denotes the bilinear interpolation of income on the next-date mortality grid. The objective-payment factor $\tilde{a}_j$ is calculated in the same way by replacing the pricing discount rate ϱ_{t_0} with the objective discount rate $\tilde{\varrho}_{t_0}$. A nine-point quadrature rule is used for these one-dimensional mortality expectations. In the baseline specification, we set $\varrho_{t_0} = \tilde{\varrho}_{t_0} = 2\%$, so the pricing and objective factors coincide up to numerical precision. The annuitization value consequently reduces to $V^{\text{annuitization}}(w, \kappa, \Delta\kappa) = w/(1 + \mathcal{L})$.

Continuation value The continuation value requires an expectation over both the return and mortality innovations, which are assumed to be independent. Let $\{z'_r, \varpi'_r\}_{r=1}^{q'}$ and $\{z''_m, \varpi''_m\}_{m=1}^{q'}$ denote the quadrature rules for the return and mortality innovations, respectively. For each return node r, the financial transition gives $W^-_{j+1,r}$, $\boldsymbol{c}_{j+1,r}$, and $W_{j+1,r}$. For each mortality node m, the ARIMA transition gives $\kappa_{j+1,m}$, $\Delta\kappa_{j+1,m}$, and the one-year death probability $q_{j+1,m}$. For a pair of nodes (r, m), we write $X_{j+1,r,m} = \left(W_{j+1,r}, \kappa_{j+1,m}, \Delta\kappa_{j+1,m}\right)$. The continuation value is approximated by:

$$
\begin{aligned}
V^{\text{continuation}}(w, \kappa, \Delta\kappa) \quad = \quad & \sum_{r=1}^{q'} \sum_{m=1}^{q''} \varpi'_r \varpi''_m \frac{1 - q_{j+1,m}}{1 + \tilde{\varrho}_{t_0}} \left(\boldsymbol{c}_{j+1,r} + V_{j+1}\left(X_{j+1,r,m}\right)\right) + \\
& \psi \sum_{r=1}^{q'} \sum_{m=1}^{q''} \varpi'_r \varpi''_m \frac{q_{j+1,m}}{1 + \tilde{\varrho}_{t_0}} W^-_{j+1,r}
\end{aligned}
$$

We use $q' = q'' = 7$, which gives 49 return-mortality node pairs at each state. Starting from the terminal value $V_H(w, \kappa, \Delta\kappa) = \psi w$, the recursion proceeds backward from $j = H - 1$ to $j = 0$. At each state, annuitization is chosen when $V^{\text{annuitization}}(w, \kappa, \Delta\kappa) \geq V^{\text{continuation}}(w, \kappa, \Delta\kappa)$. Otherwise, the individual remains liquid.

C.12.3 Numerical specification

Table 35 summarizes the numerical specifications used in the baseline solution. Since the numerical domains are finite, a quadrature transition may occasionally fall outside the next-date grid. In this case, the relevant coordinate is projected to the nearest boundary before interpolation, rather than being extrapolated outside the state domain.

Table 35: Baseline numerical specification for the dynamic annuitization problem

Component	Quantity	Baseline value
Mortality domain	Joint probability coverage	99.5%
Mortality grid	κ nodes	13
Mortality grid	$\Delta\kappa$ nodes	11
Wealth distribution	Auxiliary cells	800
Wealth domain	Probability coverage	99.95%
Wealth grid	Wealth nodes	61
Annuity-factor expectation	Mortality quadrature order	9
Continuation expectation	Return quadrature order	7
Continuation expectation	Mortality quadrature order	7
Annuity-factor surfaces	Interpolation	Bilinear
Value function	Interpolation	Trilinear
Out-of-domain states	Treatment	Nearest-boundary projection

C.13 The Hamilton-Jacobi-Bellman variational inequality

Since the admissible singular control process $Z = \{Z_t\}_{t\in[0,T)}$ allows for discontinuous jump interventions, the classical HJB partial differential equation must be generalized into a HJB variational inequality. The state space (t, x, y) is naturally partitioned into two distinct operational regions: the continuation region, where no annuities are purchased ($\mathrm{d}Z_t = 0$), and the action region, where an instantaneous lump-sum annuity purchase is optimal ($\mathrm{d}Z_t > 0$).

C.13.1 The continuation region

Suppose that at state (t, x, y), it is optimal not to purchase an annuity ($\mathrm{d}Z_t = 0$). In this case, annuity income remains strictly constant ($\mathrm{d}Y_t = 0$, *i.e.*, $Y_{t+\mathrm{d}t} = Y_t = y$), and the system evolves purely under the diffusion dynamics:

$$\mathrm{d}X_t = \Big(\big(r + \alpha_t\left(\mu - r\right)\big)\, X_t - \boldsymbol{c}_t + y\Big)\, \mathrm{d}t + \alpha_t \sigma X_t\, \mathrm{d}W_t$$

Applying ItÃt's lemma to $V\left(t + \mathrm{d}t, X_{t+\mathrm{d}t}, Y_{t+\mathrm{d}t}\right)$ along the pure diffusion path yields:

$$\begin{aligned}
(*) \quad &= \quad V\left(t + \mathrm{d}t, X_{t+\mathrm{d}t}, Y_{t+\mathrm{d}t}\right) \\
&= \quad V\left(t, x, y\right) + V_t\, \mathrm{d}t + V_x\, \mathrm{d}X_t + \frac{1}{2} V_{xx}\, \left(\mathrm{d}X_t\right)^2 \\
&= \quad V\left(t, x, y\right) + \left(V_t + \Big(\big(r + \alpha_t\left(\mu - r\right)\big)\, x - \boldsymbol{c}_t + y\Big)\, V_x + \frac{1}{2}\alpha_t^2 \sigma^2 x^2 V_{xx}\right)\, \mathrm{d}t + \\
&\quad \alpha_t \sigma x V_x\, \mathrm{d}W_t
\end{aligned}$$

where $V_t = \partial_t V, V_x = \partial_x V$ and $V_{xx} = \partial_x^2 V$. We define the second-order partial differential operator $\mathcal{L}^{\alpha_t}$ as:

$$\mathcal{L}^{\alpha_t} V = \Big(\big(r + \alpha_t\left(\mu - r\right)\big)\, x - \boldsymbol{c}_t + y\Big)\, V_x + \frac{1}{2}\alpha_t^2 \sigma^2 x^2 V_{xx}$$

It follows that:

$$V\left(t + \mathrm{d}t, X_{t+\mathrm{d}t}, Y_{t+\mathrm{d}t}\right) = V\left(t, x, y\right) + \left(V_t + \mathcal{L}^{\alpha_t} V\right)\, \mathrm{d}t + \alpha_t \sigma x V_x \mathrm{d}W_t$$

and:

$$\mathbb{E}_t\left[V\left(t + \mathrm{d}t, X_{t+\mathrm{d}t}, Y_{t+\mathrm{d}t}\right)\right] = V\left(t, x, y\right) + \left(V_t + \mathcal{L}^{\alpha_t} V\right)\, \mathrm{d}t \tag{37}$$

According to Bellman's principle of optimality, since withholding intervention ($\mathrm{d}Z_t = 0$) is locally optimal within the continuation region, the current value function $V\left(t, x, y\right)$ equals the supremum of the expected discounted continuation value over all portfolio choices $\alpha_t \in [0, 1]$:

$$\begin{aligned}
V\left(t, x, y\right) \quad = \quad \sup_{\alpha \in [0,1]} \Bigg\{ & \mathbb{P}\left[\text{die in } \mathrm{d}t\right] \times \text{bequest payoff} + \\
& \mathbb{P}\left[\text{survive in } \mathrm{d}t\right] \times \text{utility of consumption over } \mathrm{d}t + \\
& \mathbb{P}\left[\text{survive in } \mathrm{d}t\right] \times \text{best continuation from } t + \mathrm{d}t \Bigg\}
\end{aligned}$$

Mathematically, we get:

$$\begin{aligned}
V\left(t, x, y\right) \quad = \quad \sup_{\alpha \in [0,1]} \Bigg\{ & \mathbb{P}\left[\text{die in } \mathrm{d}t\right] \times e^{-\varrho\, \mathrm{d}t} \mathcal{U}\left(\xi\left(b + x\right)\right) + \\
& \mathbb{P}\left[\text{survive in } \mathrm{d}t\right] \times e^{-\varrho\, \mathrm{d}t} \mathcal{U}\left(\boldsymbol{c}_t\right)\, \mathrm{d}t + \\
& \mathbb{P}\left[\text{survive in } \mathrm{d}t\right] \times e^{-\varrho\, \mathrm{d}t} \mathbb{E}_t\left[V\left(t + \mathrm{d}t, X_{t+\mathrm{d}t}, Y_{t+\mathrm{d}t}\right)\right] \Bigg\}
\end{aligned}$$

Given survival up to t, the hazard rate λ_t gives, over a window of length $\mathrm{d}t$:

$$\begin{cases} \mathbb{P}\left[\text{die in } \mathrm{d}t\right] = \lambda_t \,\mathrm{d}t + o\left(\mathrm{d}t\right) \\ \mathbb{P}\left[\text{survive in } \mathrm{d}t\right] = 1 - \lambda_t \,\mathrm{d}t + o\left(\mathrm{d}t\right) \end{cases}$$

where $\lim_{\mathrm{d}t \to 0} \dfrac{o\left(\mathrm{d}t\right)}{\mathrm{d}t} = 0$. Using these probabilities and the conditional expectation from Equation (37), we have:

$$\begin{aligned} V\left(t,x,y\right) &= \sup_{\alpha \in [0,1]} \Bigg\{ e^{-\varrho\,\mathrm{d}t}\left(\lambda_t\,\mathrm{d}t\right)\mathcal{U}\left(\xi\left(b+x\right)\right) + e^{-\varrho\,\mathrm{d}t}\left(1-\lambda_t\,\mathrm{d}t\right)\mathcal{U}\left(\boldsymbol{c}_t\right)\,\mathrm{d}t + \\ &\qquad e^{-\varrho\,\mathrm{d}t}\left(1-\lambda_t\,\mathrm{d}t\right)\left(V\left(t,x,y\right) + \left(V_t + \mathcal{L}^{\alpha_t}V\right)\,\mathrm{d}t\right) \Bigg\} + o\left(\mathrm{d}t\right) \end{aligned}$$

Applying the first-order Taylor expansion $e^{-\varrho\,\mathrm{d}t} = 1 - \varrho\,\mathrm{d}t + o\left(\mathrm{d}t\right)$ and collecting terms up to order $\mathrm{d}t$, we obtain:

$$\begin{aligned} V\left(t,x,y\right) &= V\left(t,x,y\right) + \sup_{\alpha \in [0,1]} \Bigg\{ \lambda_t \mathcal{U}\left(\xi\left(b+x\right)\right) + \mathcal{U}\left(\boldsymbol{c}_t\right) - \left(\varrho+\lambda_t\right)V\left(t,x,y\right) + \\ &\qquad \left(V_t + \mathcal{L}^{\alpha_t}V\right) \Bigg\}\,\mathrm{d}t + o\left(\mathrm{d}t\right) \end{aligned}$$

It follows that:

$$\lambda_t \mathcal{U}\left(\xi\left(b+x\right)\right) + \mathcal{U}\left(\boldsymbol{c}_t\right) - \left(\varrho+\lambda_t\right)V\left(t,x,y\right) + V_t + \sup_{\alpha \in [0,1]} \left\{\mathcal{L}^{\alpha_t}V\right\} = 0 \tag{38}$$

At the target date T, the value function simply equals the utility of fully annuitized wealth scaled by the utility annuity factor:

$$V\left(T^-,x,y\right) = K_T \mathcal{U}\left(y + \frac{x}{\breve{a}_T}\right)$$

where:

$$K_T = \int_T^\infty e^{-\int_T^s \left(\varrho+\lambda_u\right)\mathrm{d}u}\,\mathrm{d}s$$

In the event of ruin, we have the Dirichlet condition $V\left(t,0,y\right) = G\left(t,y\right) = K_t\mathcal{U}\left(y\right) + \mathfrak{B}\left(t\right)$ and $y < \boldsymbol{c}_t$. Within the continuation region, the optimal portfolio allocation $\alpha_t^\star$ can be derived by maximizing the terms inside the supremum:

$$\begin{aligned} \alpha_t^\star &= \arg\max_{\alpha_t \in [0,1]} \Big(\left(r + \alpha_t\left(\mu - r\right)\right)x - \boldsymbol{c}_t + y\Big)V_x + \frac{1}{2}\alpha_t^2\sigma^2x^2V_{xx} \\ &= \arg\max_{\alpha_t \in [0,1]} \alpha_t\left(\mu - r\right)xV_x + \frac{1}{2}\alpha_t^2\sigma^2x^2V_{xx} \\ &= -\frac{\left(\mu - r\right)V_x}{x\sigma^2V_{xx}} \end{aligned} \tag{39}$$

C.13.2 The action region

Suppose that at state (t, x, y), it is optimal for the investor to exert a singular control by transferring an instantaneous lump sum of wealth $\mathrm{d}Z > 0$ to purchase annuities. Since this transaction occurs instantaneously at time t, no calendar time elapses ($\mathrm{d}t = 0$). Thus, neither discounting nor intermediate consumption or mortality occurs during the jump. The state variables exhibit a discontinuous jump:

$$(t, x, y) \longrightarrow \left(t, x - \mathrm{d}Z, y + \frac{\mathrm{d}Z}{\breve{a}_t}\right)$$

According to Bellman's principle of optimality, the pre-jump value function cannot be strictly less than the value function evaluated at the post-jump state under any admissible transfer $\mathrm{d}Z > 0$:

$$V(t, x, y) \geq V\left(t, x - \mathrm{d}Z, y + \frac{\mathrm{d}Z}{\breve{a}_t}\right)$$

Using a first-order Taylor expansion on the right-hand side around (t, x, y), we obtain:

$$V(t, x, y) \geq V(t, x, y) - \mathrm{d}Z \cdot V_x(t, x, y) + \frac{\mathrm{d}Z}{\breve{a}_t} V_y(t, x, y) + o(\mathrm{d}Z)$$

Taking the limit as $\mathrm{d}Z \to 0^+$, we derive the gradient obstacle condition:

$$\frac{1}{\breve{a}_t} V_y(t, x, y) - V_x(t, x, y) \leq 0 \tag{40}$$

Within the action region where an immediate lump-sum annuity purchase is strictly optimal, Equation (40) holds with equality:

$$V_x(t, x, y) = \frac{1}{\breve{a}_t} V_y(t, x, y)$$

From an economic standpoint, this condition states that everywhere in the state space, the marginal utility of retaining one dollar of liquid wealth (V_x) must be greater than or equal to the marginal utility of converting that dollar into a continuous annuity income ($V_y/\breve{a}_t$).

C.13.3 The unified HJB variational inequality

Since an admissible control strategy requires the investor to behave optimally in every state (t, x, y), the state space naturally partitions into complementary waiting and intervening regimes.

- Within the continuation region, withholding intervention is locally optimal. Thus, the second-order partial differential equation holds with exact equality — Equation (38) — while purchasing an immediate lump-sum annuity in this region would be suboptimal. Consequently, the gradient obstacle condition remains non-positive — (Equation 40). Thus, we have:

$$\begin{cases} V_t + \sup_{\alpha \in [0,1]} \{\mathcal{L}^{\alpha_t} V\} - (\varrho + \lambda_t) V(t, x, y) + \lambda_t \mathcal{U}\left(\xi(b + x)\right) + \mathcal{U}(\boldsymbol{c}_t) = 0 \\ \dfrac{1}{\breve{a}_t} V_y(t, x, y) - V_x(t, x, y) \leq 0 \end{cases}$$

- Within the action region, executing an instantaneous lump-sum annuity purchase is strictly optimal. The gradient obstacle condition holds with exact equality, perfectly balancing the marginal utilities of liquid wealth and annuity income. Conversely, delaying intervention would be suboptimal. Thus, we have:

$$\begin{cases} V_t + \sup_{\alpha\in[0,1]}\left\{\mathcal{L}^{\alpha_t}V\right\} - (\varrho+\lambda_t)\,V(t,x,y) + \lambda_t\mathcal{U}\left(\xi\,(b+x)\right) + \mathcal{U}\left(\boldsymbol{c}_t\right) \leq 0 \\ \dfrac{1}{\breve{a}_t}V_y(t,x,y) - V_x(t,x,y) = 0 \end{cases}$$

Combining these complementary regimes across the entire state space, the value function $V(t,x,y)$ is characterized as a viscosity solution of the following HJB variational inequality:

$$\max\left\{ V_t + \sup_{\alpha\in[0,1]}\left\{\mathcal{L}^{\alpha_t}V\right\} - (\varrho+\lambda_t)\,V(t,x,y) + \lambda_t\mathcal{U}\left(\xi\,(b+x)\right) + \mathcal{U}\left(\boldsymbol{c}_t\right), \frac{1}{\breve{a}_t}V_y - V_x \right\} = 0 \tag{41}$$

subject to the terminal annuitization condition at the target date T:

$$V(T,x,y) = \mathcal{U}\left(y + \frac{x}{\breve{a}_T}\right)\int_T^\infty e^{-\int_T^s(\varrho+\lambda_u)\,\mathrm{d}u}\,\mathrm{d}s$$

and the Dirichlet condition:

$$\begin{cases} V(t,0,y) = G(t,y) = K_t\mathcal{U}(y) + \mathfrak{B}(t) \\ y < \boldsymbol{c}_t \end{cases}$$

C.13.4 The free boundary problem

Equation (41) splits the state space into two regions, which are separated by an endogenous two-dimensional free boundary surface $x^\star(t,y)$, where:

$$\begin{cases} V_t + \sup_{\alpha\in[0,1]}\left\{\mathcal{L}^{\alpha_t}V\right\} - (\varrho+\lambda_t)\,V(t,x,y) + \lambda_t\mathcal{U}\left(\xi\,(b+x^\star)\right) + \mathcal{U}\left(\boldsymbol{c}_t\right) = 0 \\ \dfrac{1}{\breve{a}_t}V_y - V_x = 0 \end{cases} \tag{42}$$

System (42) defines the free boundary as the set of points where both the continuation and action conditions hold with equality. In practice, however, we cannot solve this system of equations directly to recover $x^\star(t,y)$ for two fundamental reasons:

- Circular dependency
  The terms V_t, V_x, and V_y are not explicit parameters and depend on the unknown global value function $V(t,x,y)$. To compute V and its derivatives, we must first know the location of the boundary. Conversely, locating the boundary requires the derivatives of V. Therefore, we cannot algebraically isolate $x^\star(t,y)$ from this circularity.

- Degeneracy
  Economically, the gradient condition $\dfrac{1}{\breve{a}_t}V_y - V_x = 0$ implies that the marginal utility of purchasing an annuity equals the marginal utility of retaining liquid wealth. Thus, the value function is strictly invariant along the segment of the transaction ray contained in the action region $z \mapsto \left(x - z, y + \dfrac{z}{\breve{a}_t}\right)$. Because an instantaneous lump-sum purchase involves no passage of calendar time, V remains constant along the entire jump trajectory, so its directional derivative along the annuitization direction $\left(-1, 1/\breve{a}_t\right)$

vanishes at every intermediate point of the jump path. Equivalently, the non-local intervention operator $\mathcal{M}V(t,x,y) = \sup_{z\in[0,x]} V\left(t, x-z, y+\frac{z}{\breve{a}_t}\right)$ is flat along the transaction ray. Consequently, the arg max is not locally unique within the action region.

Since we cannot explicitly find the root $x^\star(t,y)$ using Equation (42), we adopt a global numerical approach. We solve for the value function $V(t,x,y)$ globally over a discretized state space, computing two quantities at each grid point:

1. The value of waiting in the continuation region, given by the operator:

$$V_t + \sup_{\alpha\in[0,1]} \{\mathcal{L}^{\alpha_t} V\} - (\varrho+\lambda_t) V(t,x,y) + \lambda_t \mathcal{U}\left(\xi(b+x)\right) + \mathcal{U}(\boldsymbol{c}_t)$$

2. The value of making an immediate optimal lump-sum purchase, given by the non-local intervention operator:

$$\mathcal{M}V(t,x,y) = \sup_{z\in[0,x]} V\left(t, x-z, y+\frac{z}{\breve{a}_t}\right)$$

Once the value function has converged over the grid, the free boundary is recovered *ex post* as the interface between the set on which $z=0$ is the unique maximizer in $\mathcal{M}$:

$$\begin{cases} V_t + \sup_{\alpha\in[0,1]} \{\mathcal{L}^{\alpha_t} V\} - (\varrho+\lambda_t) V(t,x,y) + \lambda_t \mathcal{U}\left(\xi(b+x)\right) + \mathcal{U}(\boldsymbol{c}_t) = 0 \\ \frac{1}{\breve{a}_t} V_y - V_x < 0 \end{cases}$$

and the set on which a maximizer is attained at some $z>0$, where the discrete gradient condition holds with equality. The boundary is not obtained by solving the equations directly, but rather by identifying it as the active set of the discrete obstacle problem. The optimal transfer $\mathrm{d}Z^\star$ is the largest maximizer in $\mathcal{M}$, with $\mathrm{d}Z^\star = x$ corresponding to the corner solution of full annuitization.

C.14 Numerical solution of the HJB variational inequality

To implement the numerical scheme, we solve this free-boundary problem using the finite difference method combined with Howard's policy-iteration algorithm (Howard, 1960; Puterman, 2014). Let N_t, N_x, and N_y denote the number of mesh intervals for t, x, and y, respectively. We define $\Delta t = T/N_t$, $\Delta x = x_{\max}/N_x$, and $\Delta y = (y_{\max} - \bar{y})/N_y$ as the mesh spacings. Let $V_{i,j}^m \approx V\left(t_m, x_i, y_j\right)$ be the approximate solution at the grid point $\left(t_m, x_i, y_j\right)$, where[89] $t_m = m\cdot\Delta t$, $x_i = i\cdot\Delta x$, and $y_j = \bar{y} + j\cdot\Delta y$. Accordingly, the policy now comprises two ingredients: the continuous allocation α_t and the discrete exercise regime between waiting and annuitizing. The Howard's algorithm alternates between a policy-evaluation step, in which both ingredients are frozen and a linear system is solved, and a policy-improvement step, in which both are updated.

C.14.1 Policy evaluation

In the policy-evaluation step, we discretize the wait operator and gradient obstacle.

[89] We have $m = 0,1,\ldots,N_t$, $i = 0,1,\ldots,N_x$, and $j = 0,1,\ldots,N_y$.

Discretization of the wait operator Because annuity income is frozen throughout the continuation region ($\mathrm{d}Y_t = 0$), Equation (38) contains no derivative with respect to y. Given a fixed candidate control $\alpha_{i,j}^m$, the wait-region equation reads, for a fixed slice j:

$$V_t + \bar{A}(x)\, V_{xx} + \bar{B}(x,y)\, V_x + \bar{C}V + \bar{D}(x) = 0 \tag{43}$$

where:

$$\begin{cases} \bar{A}_i \equiv \bar{A}(x_i) = \dfrac{1}{2}\left(\alpha_{i,j}^m\right)^2 \sigma^2 x_i^2 \\ \bar{B}_{i,j} \equiv \bar{B}(x_i, y_j) = \left(r + \alpha_{i,j}^m (\mu - r)\right) x_i - \boldsymbol{c}_{t_m} + y_j \\ \bar{C} = -(\varrho + \lambda_{t_m}) \\ \bar{D}_i \equiv \bar{D}(x_i) = \mathcal{U}(\boldsymbol{c}_{t_m}) + \lambda_{t_m} \mathcal{U}\left(\xi(b + x_i)\right) \end{cases}$$

Each annuity slice j therefore constitutes an independent one-dimensional problem in x, so no boundary condition is required in the y direction for the wait operator. We define $\mathcal{F}^m = \bar{A}\partial_x^2 + \bar{B}\partial_x + \bar{C}$ with all coefficients evaluated at t_m. To solve this PDE, we use a fully implicit finite difference scheme. Since the terminal condition is known at $t_{N_t} = T$, we solve the PDE backward in time. The backward step from the known future state at t_{m+1} to the current state at t_m approximates the time derivative as $V_t \approx (V_{i,j}^{m+1} - V_{i,j}^m)/\Delta t$. The fully implicit scheme evaluates the spatial operator at t_m:

$$\frac{V_{i,j}^{m+1} - V_{i,j}^m}{\Delta t} + \left(\bar{A}_i \partial_{xx} V_{i,j}^m + \bar{B}_{i,j} \partial_x V_{i,j}^m + \bar{C} V_{i,j}^m\right) + \bar{D}_i = 0$$

The spatial derivatives are approximated using central differences for diffusion and an upwind scheme for the drift to ensure monotonicity:

$$\partial_x^2 V_{i,j}^m \approx \frac{V_{i+1,j}^m - 2V_{i,j}^m + V_{i-1,j}^m}{\Delta x^2}$$

and:

$$\partial_x V_{i,j}^m \approx \begin{cases} \dfrac{V_{i+1,j}^m - V_{i,j}^m}{\Delta x} & \text{if } \bar{B}_{i,j} \geq 0 \\ \dfrac{V_{i,j}^m - V_{i-1,j}^m}{\Delta x} & \text{if } \bar{B}_{i,j} < 0 \end{cases}$$

By setting $s_i = \bar{A}_i \Delta t / \Delta x^2$, the discrete equation for row i on slice j can be written in tridiagonal matrix form:

$$\underbrace{(I - \Delta t\, \mathcal{F}^m)}_{M_j^m} V_j^m = \underbrace{V_j^{m+1} + \Delta t\, \bar{D}}_{\mathrm{RHS}_j^m} \tag{44}$$

where $V_j^m = \left(V_{0,j}^m, V_{1,j}^m, \ldots, V_{N_x,j}^m\right)$. The coefficients of M_j^m are given by:

$$M_j^m = I - \Delta t\, \mathcal{F}^m = \begin{bmatrix} d_0 & u_0 & 0 & \cdots & 0 \\ l_1 & d_1 & u_1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & l_{N_x-1} & d_{N_x-1} & u_{N_x-1} \\ 0 & \cdots & 0 & l_{N_x} & d_{N_x} \end{bmatrix}$$

and:

- When $\bar{B}_{i,j} \geq 0$: $l_i = -s_i$, $d_i = 1 + 2s_i + \bar{B}_{i,j}\dfrac{\Delta t}{\Delta x} - \bar{C}\Delta t$, $u_i = -s_i - \bar{B}_{i,j}\dfrac{\Delta t}{\Delta x}$.
- When $\bar{B}_{i,j} < 0$: $l_i = -s_i + \bar{B}_{i,j}\dfrac{\Delta t}{\Delta x}$, $d_i = 1 + 2s_i - \bar{B}_{i,j}\dfrac{\Delta t}{\Delta x} - \bar{C}\Delta t$, $u_i = -s_i$.

At the left boundary ($i = 0$), the diffusion coefficient vanishes ($\bar{A}_0 = 0$, hence $s_0 = 0$) and the equation degenerates to a first-order transport equation whose drift reduces to $\bar{B}_{0,j} = y_j - \boldsymbol{c}_{t_m}$. Since the consumption stream is planned at date t_0 and must remain financeable out of available resources, we adopt the convention that the effective consumption rate is capped at y_j.

- When $\bar{B}_{0,j} \geq 0$, the characteristic points into the domain and no boundary condition is required. The forward (upwind) difference $\partial_x V^m_{0,j} \approx \left(V^m_{1,j} - V^m_{0,j}\right)/\Delta x$ involves interior nodes only, and the interior scheme with $s_0 = 0$ yields $l_0 = 0$, $d_0 = 1 + \bar{B}_{0,j}\dfrac{\Delta t}{\Delta x} - \bar{C}\Delta t$, $u_0 = -\bar{B}_{0,j}\dfrac{\Delta t}{\Delta x}$.
- When $\bar{B}_{0,j} < 0$, the characteristic points out of the domain while diffusion vanishes, so the state is absorbed. The failure time $\boldsymbol{\tau}$ is triggered and the problem stops at the fallback value $G(t, y)$. This is a Dirichlet condition:

$$V^m_{0,j} = G\left(t_m, y_j\right) = K_{t_m}\mathcal{U}\left(y_j\right) + \mathfrak{B}\left(t_m\right) \tag{45}$$

  imposed by overwriting the row as $l_0 = 0$, $d_0 = 1$, $u_0 = 0$ with right-hand side $G\left(t_m, y_j\right)$.

At the right boundary ($i = N_x$), the liquid-wealth domain is unbounded and must be truncated at $x = x_{\max}$. Following Forsyth and Labahn (2007), we derive the artificial boundary condition from an approximation valid for large wealth, place the truncation point well outside the region of interest, and confirm its adequacy through a domain sensitivity analysis. We take this approximation to be the exact value of an explicit admissible policy: hold only the risk-free asset, make no further annuity purchases on $[t, T)$, fund the planned consumption stream, and convert the entire remaining balance into an annuity at T. Under this policy, the wealth path is deterministic:

$$\mathrm{d}X^{\mathrm{ref}}_s = \left(rX^{\mathrm{ref}}_s + y - \boldsymbol{c}_s\right)\mathrm{d}s$$

It follows that:

$$X^{\mathrm{ref}}_s(t, y) = e^{r(s-t)}x_{\max} + \int_t^s e^{r(s-u)}\left(y - \boldsymbol{c}_u\right)\mathrm{d}u \tag{46}$$

for $s \in [t, T]$. Its value function is given by the one-dimensional quadrature:

$$\begin{aligned} g_{\mathrm{ref}}(t, y) \quad = \quad & \Phi(t) + \int_t^T D(t, s)\,\lambda_s\mathcal{U}\left(\xi\left(b + X^{\mathrm{ref}}_s(t, y)\right)\right)\mathrm{d}s \\ & + D(t, T)\,K_T\mathcal{U}\left(y + \frac{X^{\mathrm{ref}}_T(t, y)}{\breve{a}_T}\right) \end{aligned} \tag{47}$$

where $\Phi(t) = \int_t^T D(t, s)\,\mathcal{U}\left(\boldsymbol{c}_s\right)\mathrm{d}s$ solves:

$$\begin{cases} \Phi'(t) - (\varrho + \lambda_t)\,\Phi(t) + \mathcal{U}\left(\boldsymbol{c}_t\right) = 0 \\ \Phi(T) = 0 \end{cases}$$

Prescribing $V^m_{N_x,j} = g_{\text{ref}}\left(t_m, y_j\right)$ amounts to solving a restricted problem in which the investor retains the full control set until wealth first reaches the boundary, and follows the reference policy thereafter. This restricted problem satisfies the same HJB variational inequality as the original one in the interior, its boundary condition is exact by construction, and its value is a lower bound on V (*i.e.*, $g_{\text{ref}} \leq V$), since the restricted control set is a subset of the original admissible set. The boundary row is therefore set to $l_{N_x} = 0$, $d_{N_x} = 1$, $u_{N_x} = 0$, and $\text{RHS}_{N_x} = g_{\text{ref}}\left(t_m, y_j\right)$.

Discretization of the gradient obstacle The annuitization constraint $V_x - \frac{1}{\breve{a}_t} V_y \geq 0$ strictly couples the x and y dimensions. It is discretized with one-sided differences aligned with the direction along which a lump-sum purchase displaces the state, $(x, y) \mapsto \left(x - \Delta x, y + \Delta x/\breve{a}_t\right)$, that is, backward in x and forward in y:

$$\underbrace{\frac{V^m_{i,j} - V^m_{i-1,j}}{\Delta x} - \frac{1}{\breve{a}_{t_m}} \frac{V^m_{i,j+1} - V^m_{i,j}}{\Delta y}}_{\mathcal{G}^m_{i,j}} = 0$$

By rearranging the terms to isolate the target node $V^m_{i,j}$, we have:

$$V^m_{i,j} - w_x V^m_{i-1,j} = w_y V^m_{i,j+1}$$

where $w_x = \dfrac{\breve{a}_{t_m}\,\Delta y}{\breve{a}_{t_m}\,\Delta y + \Delta x}$ and $w_y = \dfrac{\Delta x}{\breve{a}_{t_m}\,\Delta y + \Delta x}$. To integrate this into the existing linear solver framework, we map this relation into the same tridiagonal matrix structure used for the wait operator:

$$M^m_j V^m_j = \underbrace{w_y V^m_{j+1}}_{\text{RHS}^m_j} \tag{48}$$

where the coefficients are given by $l_i = -w_x$, $d_i = 1$, $u_i = 0$, and:

$$M^m_j = \begin{bmatrix} d_0 & u_0 & 0 & \cdots & 0 \\ l_1 & d_1 & u_1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & l_{N_x-1} & d_{N_x-1} & u_{N_x-1} \\ 0 & \cdots & 0 & l_{N_x} & d_{N_x} \end{bmatrix}$$

Since the calculation of $\mathcal{G}^m_{i,j}$ involves the nodes $(i-1, j)$ and $(i, j+1)$, it reaches outside the computational grid at two places only: the lower wealth boundary $i = 0$ and the upper annuity boundary $j = N_y$. Fortunately, both of these out-of-grid dependencies correspond to economic states where singular jump interventions are strictly inadmissible, allowing us to close the numerical system seamlessly. We discuss these boundary conditions in the next paragraph.

C.14.2 Policy improvement

In the policy-improvement step, we update both the continuous control $\alpha^m_{i,j}$ and the discrete exercise regime $R_{i,j} \in \{0, \text{Wait}; 1, \text{Annuitize}\}$ based on the current guess of the value function. At a given time step m, the controls are updated across the grid as follows.

1. Continuous control update
   The optimal risky allocation $\alpha_{i,j}^m$ is evaluated using the finite difference approximations of the derivatives of V^m:

$$\alpha_{i,j}^m = \min\left(1, \max\left(0, -\frac{(\mu - r)\,\partial_x V_{i,j}^m}{x_i \sigma^2 \partial_x^2 V_{i,j}^m}\right)\right)$$

2. Discrete regime update
   We evaluate the gradient obstacle $\mathcal{G}_{i,j}^m$ using V^m to determine whether the marginal utility of purchasing annuities exceeds that of retaining liquidity ($V_x - V_y/\breve{a}_t < 0$). To respect physical liquidity constraints and eliminate out-of-bounds grid references at the boundaries ($i = 0$ and $j = N_y$), the discrete regime indicator $R_{i,j}$ is assigned across the entire state space as follows:

$$R_{i,j} = \begin{cases} 0 & \text{if } i = 0 \quad \text{(zero liquidity: physical intervention impossible)} \\ 0 & \text{if } j = N_y \quad \text{(annuity ceiling: no further lump-sum transfers into annuities)} \\ 1 & \text{if } i \in \{1, \ldots, N_x\},\ j \in \{0, \ldots, N_y - 1\} \text{ and } \mathcal{G}_{i,j}^m < 0 \quad \text{(annuitize)} \\ 0 & \text{otherwise} \quad \text{(wait)} \end{cases}$$

By explicitly setting $R_{0,j} \equiv 0$, we reflect the physical reality that an investor with zero liquid wealth ($x_0 = 0$) cannot execute an instantaneous lump-sum annuity purchase ($\Delta Z > 0$). Similarly, setting $R_{i,N_y} \equiv 0$ enforces the economic condition that once the guaranteed income stream reaches the computational domain ceiling $y_{\max}$, no further annuitization is permitted. Numerically, this natural economic gating ensures that the gradient obstacle stencil never references the undefined exterior nodes $(-1, j)$ or $(i, N_y + 1)$. Once the policy (α, R) is fixed, we construct the mixed linear system to solve for the updated value function V^m. The equation for each row i in slice j is chosen dynamically based on $R_{i,j}$:

- When $R_{i,j} = 0$, the row uses the tridiagonal coefficients of the wait operator and asymptotic Dirichlet condition derived in the policy-evaluation step.
- When $R_{i,j} = 1$, the value function is determined by enforcing the gradient obstacle equality.

C.14.3 The algorithmic loop

The numerical solution is computed backward in time from the terminal date T to the initial date 0. After initializing the terminal value function $V_{i,j}^{N_t} = \mathcal{U}\left(y_j + \frac{x_i}{\breve{a}_T}\right) \int_T^\infty e^{-\int_T^s (\varrho + \lambda_u)\,\mathrm{d}u}\,\mathrm{d}s$ for all (i, j), the scheme marches backward across each time step t_m ($m = N_t - 1, N_t - 2, \ldots, 0$) using Howard's policy-iteration algorithm:

- Initialization
  At the start of time step t_m, we initialize the candidate value function using the converged solution from the future time step: $V^{m,(0)} = V^{m+1}$.
- Howard policy-iteration loop
  For iteration index $k = 0, 1, 2, \ldots$, we update the policy and evaluate the value function until convergence. In each iteration k, we execute the following steps:
  - for $j = N_y$ down to 0
    - Policy Improvement

· Evaluate the optimal portfolio allocation $\alpha_{i,j}^{(k+1)}$ across all interior wealth nodes using the numerical spatial derivatives of the current guess $V_j^{m,(k)}$.
· At the upper boundary slice $j = N_y$ and the lower wealth boundary $i = 0$, we strictly enforce the continuation regime by setting $R_{i,j}^{(k+1)} = 0$.
· For all strictly positive wealth nodes ($i \in \{1, \ldots, N_x\}$ and $j < N_y$), we evaluate the discrete gradient obstacle condition $\mathcal{G}_{i,j}^m$ using the newly updated adjacent slice $V_{j+1}^{m,(k+1)}$ and the current guess $V_j^{m,(k)}$. We assign the discrete regime indicator:

$$R_{i,j}^{(k+1)} = \begin{cases} 1 \text{ (annuitize)} & \text{if } \mathcal{G}_{i,j}^m < 0 \\ 0 \text{ (wait)} & \text{otherwise} \end{cases}$$

∗ Policy Evaluation
We construct a tridiagonal linear system $M_j V_j^{m,(k+1)} = \text{RHS}$ dynamically based on the updated regime vector $R_j^{(k+1)}$.

· For nodes where $R_{i,j}^{(k+1)} = 0$, row i of matrix M_j is filled with the implicit finite difference coefficients (l_i, d_i, u_i) of the continuation operator (Equation (44)), and the right-hand side is set to $\text{RHS}_i = V_{i,j}^{m+1} + \bar{D}_i \, \Delta t$. At the upper wealth boundary ($i = N_x$), we enforce the asymptotic Dirichlet condition by setting row N_x to $(l_{N_x}, d_{N_x}) = (0, 1)$, and $\text{RHS}_{N_x} = g_{\text{ref}}\left(t_m, y_j\right)$.
· For nodes where $R_{i,j}^{(k+1)} = 1$, row i is replaced by the gradient obstacle coefficients (l_i, d_i, u_i) from Equation (48), and the right-hand side is set to the known boundary term $\text{RHS}_i = w_y V_{i,j+1}^{m,(k+1)}$.
· Because the assembled matrix M_j strictly preserves diagonally dominant M-matrix properties under both regimes, we solve the tridiagonal system directly to obtain the updated slice $V_j^{m,(k+1)}$.

– Convergence check
After all slices $j \in \{0, \ldots, N_y\}$ have been updated, we compute the maximum infinity-norm error over the entire two-dimensional spatial grid:

$$\text{err} = \max_{0 \le i \le N_x,\, 0 \le j \le N_y} \left| V_{i,j}^{m,(k+1)} - V_{i,j}^{m,(k)} \right|$$

If $\text{err} < \epsilon$, the policy-iteration loop terminates, and we accept $V^m = V^{m,(k+1)}$ as the converged solution at time t_m.

C.14.4 Recovery of the optimal lump-sum transfer $\Delta Z^\star$

Once the free boundary surface $x^\star(t, y)$ is mapped, the optimal singular transfer $\Delta Z^\star$ for any initial state (x_i, y_j) inside the action region ($R_{i,j}^m = 1$) is determined geometrically by the reflection principle. The lump-sum purchase $\Delta Z^\star$ is obtained by solving the scalar fixed-point equation:

$$\Delta Z^\star = x_i - x^\star\left(t_m, y_j + \frac{\Delta Z^\star}{\breve{a}_{t_m}}\right) \tag{49}$$

D Additional results

D.1 Tables

Table 36: Impact of the calibration set on the value of the annuity contract (France)

Calibration set	Age	Female				Male			
		1950	1975	2000	2020	1950	1975	2000	2020
1950–2023	60	13.88	16.05	17.56	18.16	11.53	12.92	15.30	16.22
50–95	75	6.73	8.35	10.34	10.91	5.60	6.48	8.35	9.42
$x^+ = 95$	90	2.05	2.44	2.86	3.09	1.84	2.08	2.45	2.68
1950–2023	60	13.91	16.18	17.82	18.63	11.54	12.95	15.40	16.41
30–105	75	6.77	8.48	10.66	11.45	5.61	6.52	8.47	9.65
$x^+ = 105$	90	2.35	2.95	3.75	4.18	2.06	2.38	2.95	3.34
1825–2023	60	13.88	16.05	17.42	17.65	11.53	12.92	15.27	15.85
50–95	75	6.73	8.35	10.34	10.48	5.60	6.48	8.35	9.26
$x^+ = 95$	90	2.05	2.44	2.86	3.06	1.84	2.08	2.45	2.67

Table 37: Mortality credit in % (France, male, LC WLS, $r = 0$)

Age	1900	1930	1950	1970	1990	2000	2010	2015	2020	2023
65	4.74	4.70	3.75	3.35	2.21	1.89	1.45	1.29	1.26	1.32
70	7.55	7.26	5.83	5.15	3.40	2.82	2.14	1.83	1.75	1.86
75	12.28	11.80	9.36	8.16	5.48	4.39	3.42	2.90	2.68	2.82
80	19.11	19.07	15.73	13.46	9.36	7.59	6.04	5.18	4.69	4.87
85	32.00	30.95	25.17	22.05	15.75	13.52	10.97	9.67	9.10	9.33
90	46.43	48.15	38.82	35.17	26.21	23.78	20.59	18.64	17.94	18.53

Table 38: Mortality credit in % (France, female, LC WLS, $r = 0$)

Age	1900	1930	1950	1970	1990	2000	2010	2015	2020	2023
65	4.05	3.25	2.27	1.54	0.85	0.73	0.61	0.57	0.59	0.62
70	6.44	5.38	3.85	2.67	1.41	1.18	0.96	0.85	0.85	0.91
75	10.16	9.07	6.61	4.82	2.63	2.12	1.64	1.43	1.40	1.49
80	16.25	15.01	11.71	8.82	5.30	4.19	3.21	2.81	2.71	2.86
85	26.32	24.23	19.63	15.68	10.42	8.60	6.75	5.92	5.75	6.01
90	37.49	38.14	33.14	26.80	19.46	16.81	13.92	12.51	12.28	12.83

D.2 Figures

Figure 71: Comparison of estimates using SVD and WLS methods (France, female, Lee-Carter model)

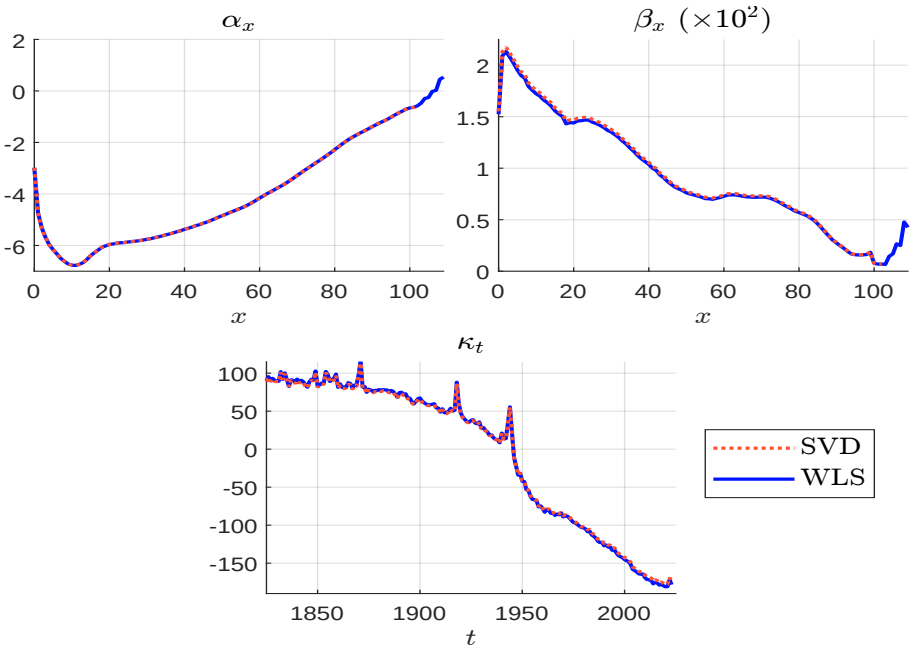


Figure 72: Comparison of estimates using SVD and WLS methods (France, male, Lee-Carter model)

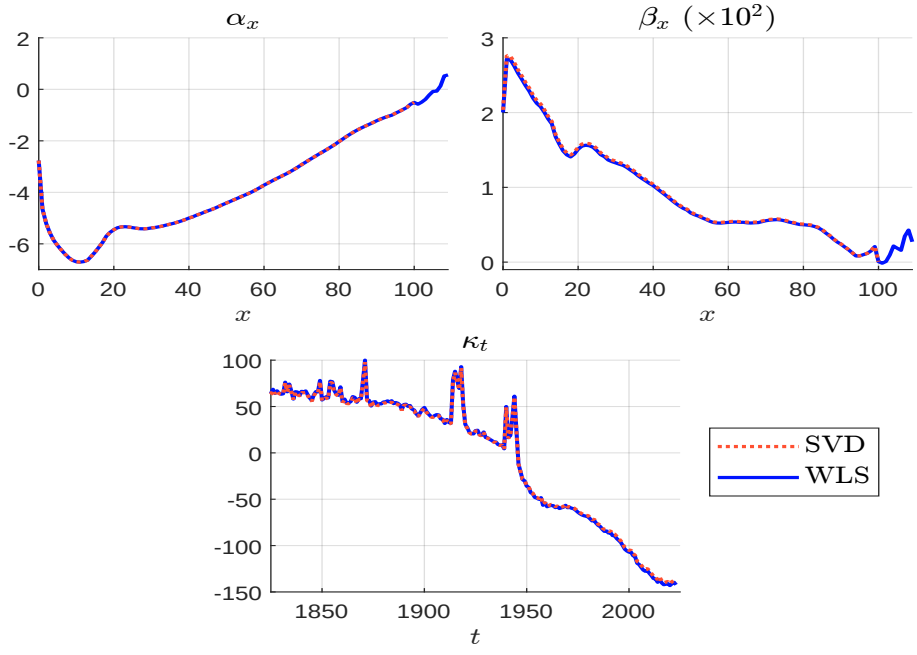


Figure 73: Poisson LC model estimation (France, female)

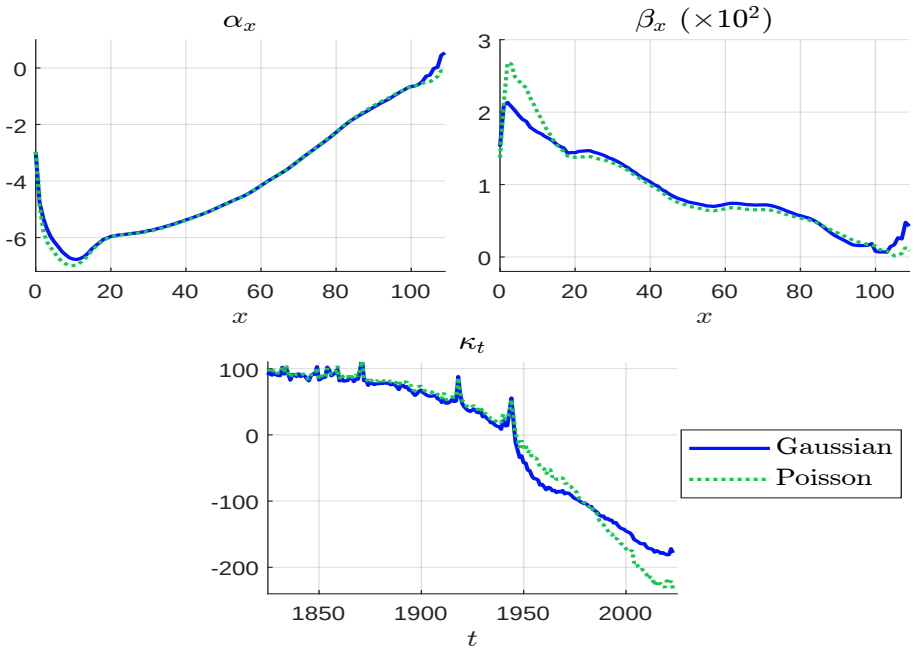


Figure 74: Poisson LC model estimation (France, male)

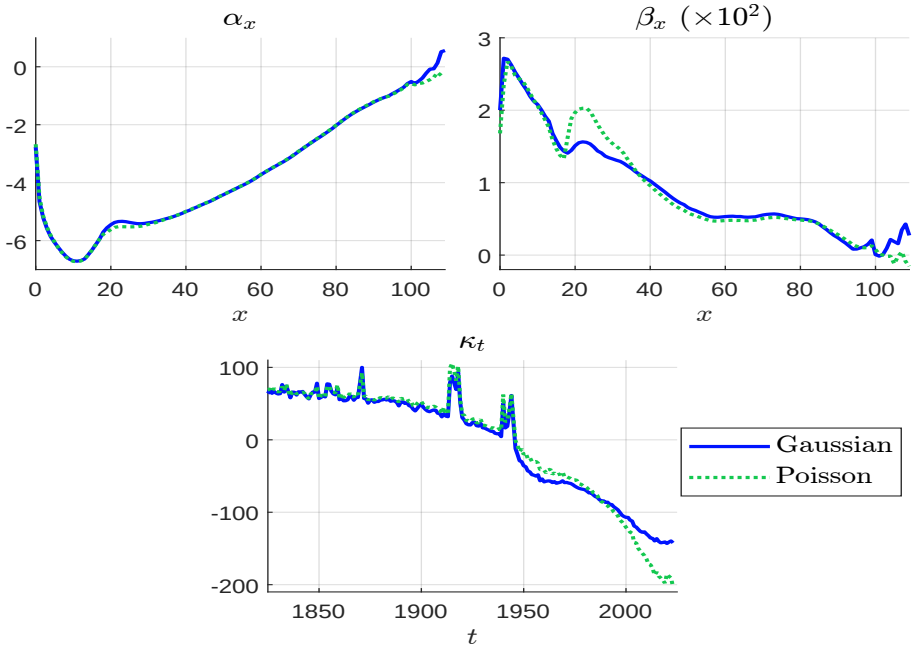


Figure 75: Mean absolute error of the central death rate (Poisson Lee-Carter model, France)

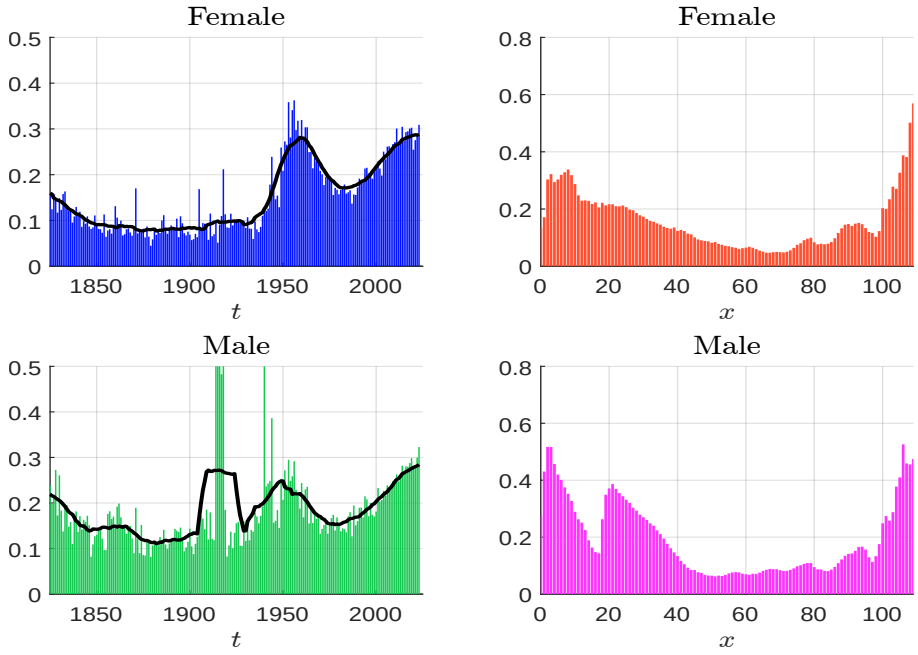


Figure 76: Negative binomial LC model estimation (France, female)

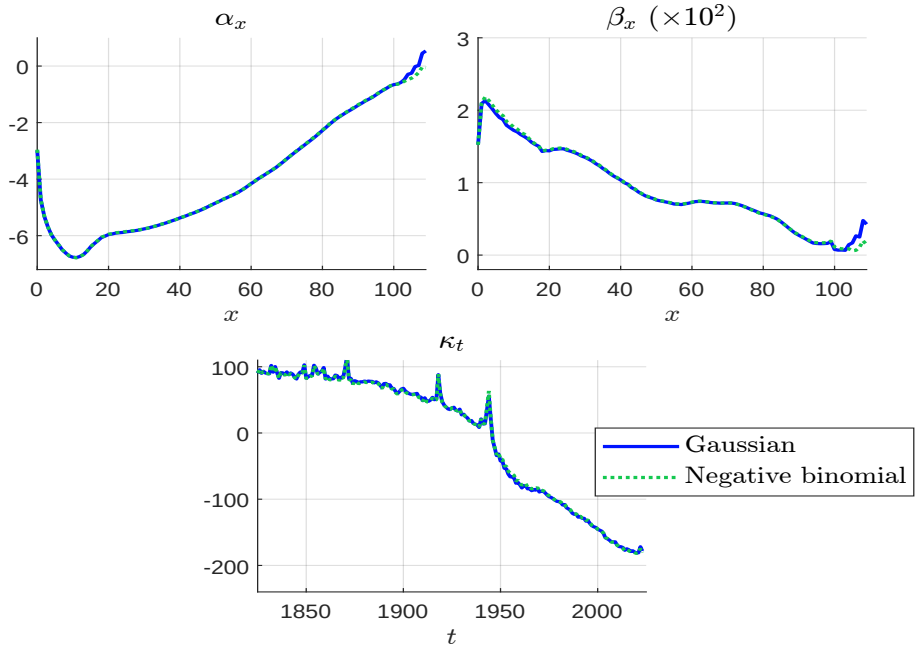


Figure 77: Negative binomial LC model estimation (France, male)

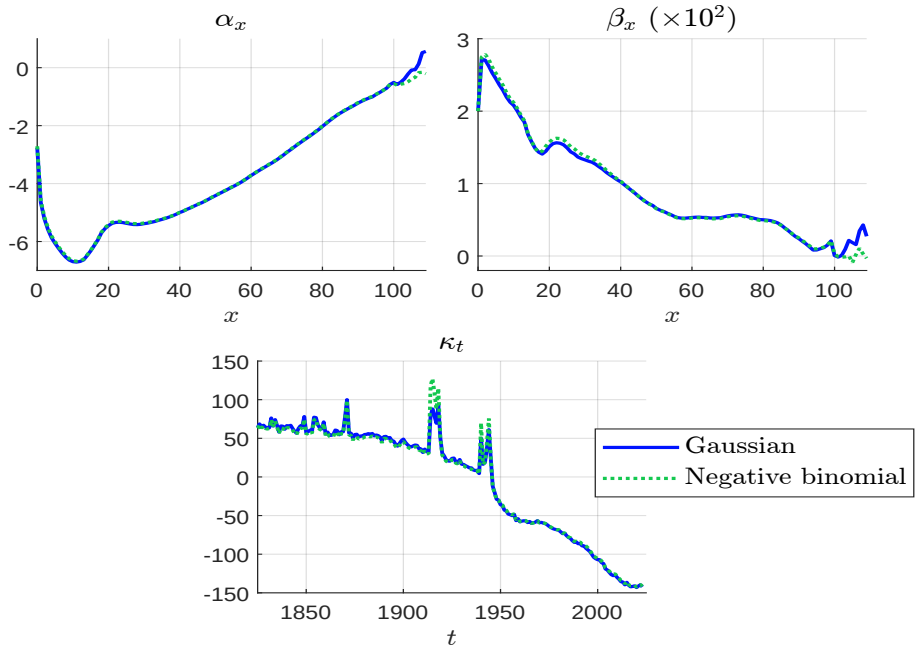


Figure 78: Mean absolute error of the central death rate (Negative binomial Lee-Carter model, France)

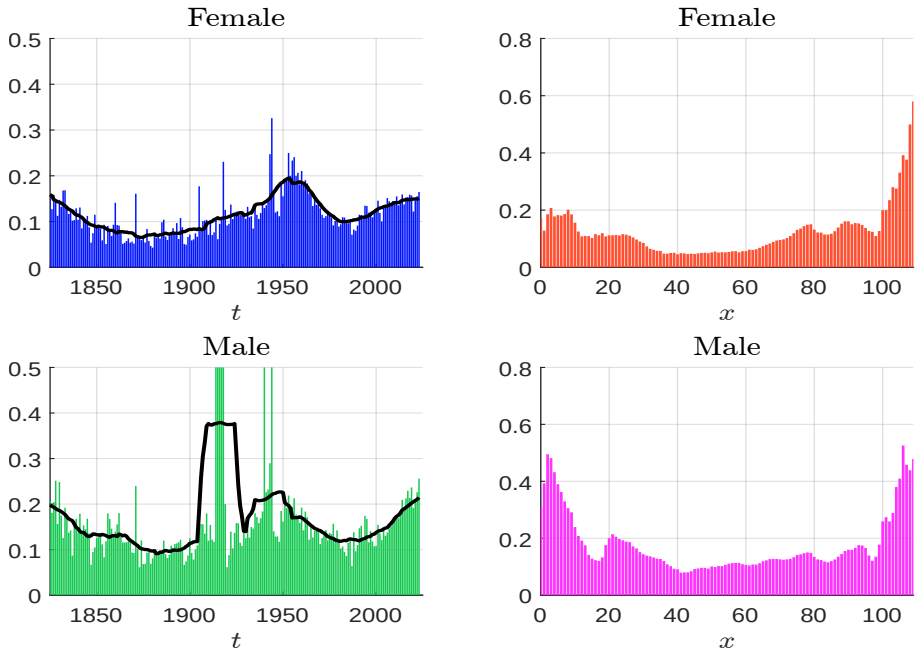


Figure 79: Estimates κ_1 and κ_2 (France, 60–100)

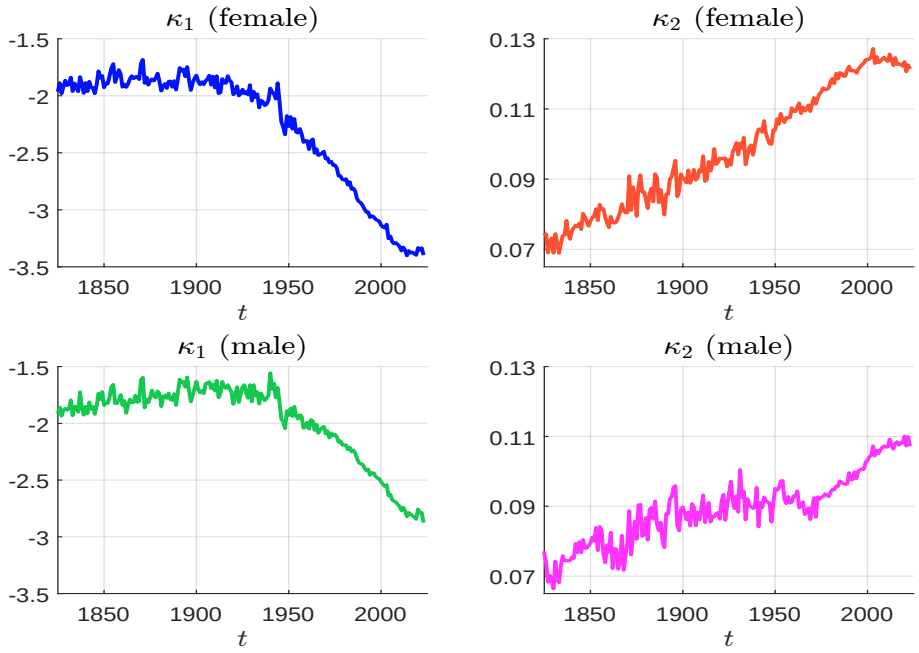


Figure 80: Estimates κ_1 and κ_2 (France, 60–109)

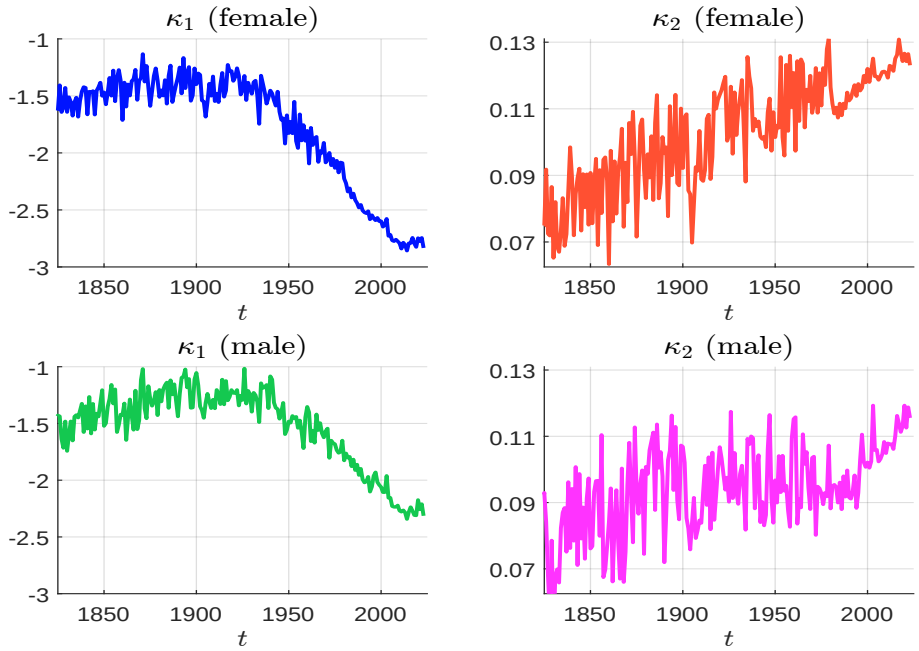


Figure 81: WLS estimated values of κ_1 (France, female, 60–109, outlier correction)

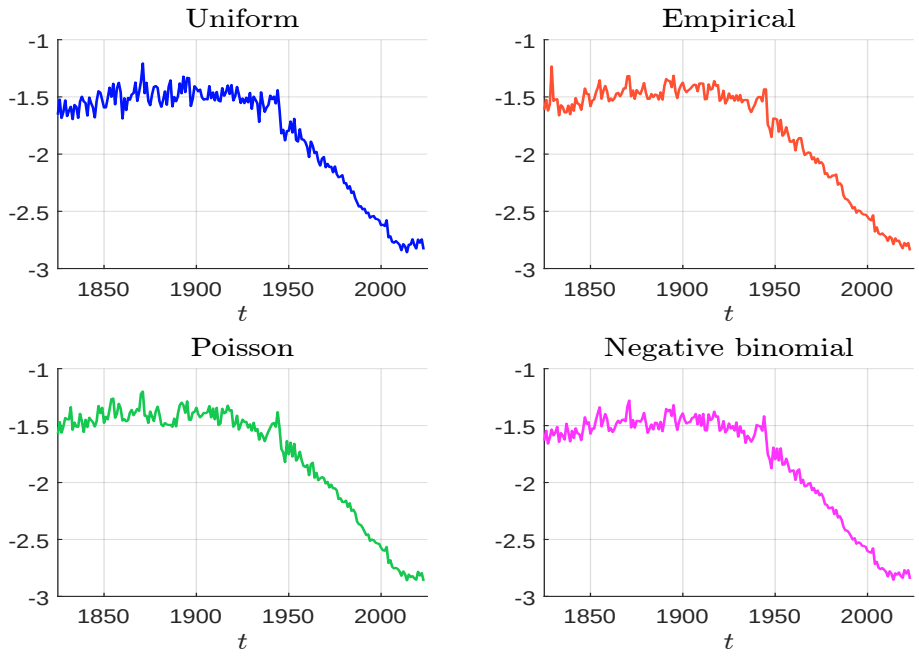


Figure 82: WLS estimated values of κ_2 (France, female, 60–109, outlier correction)

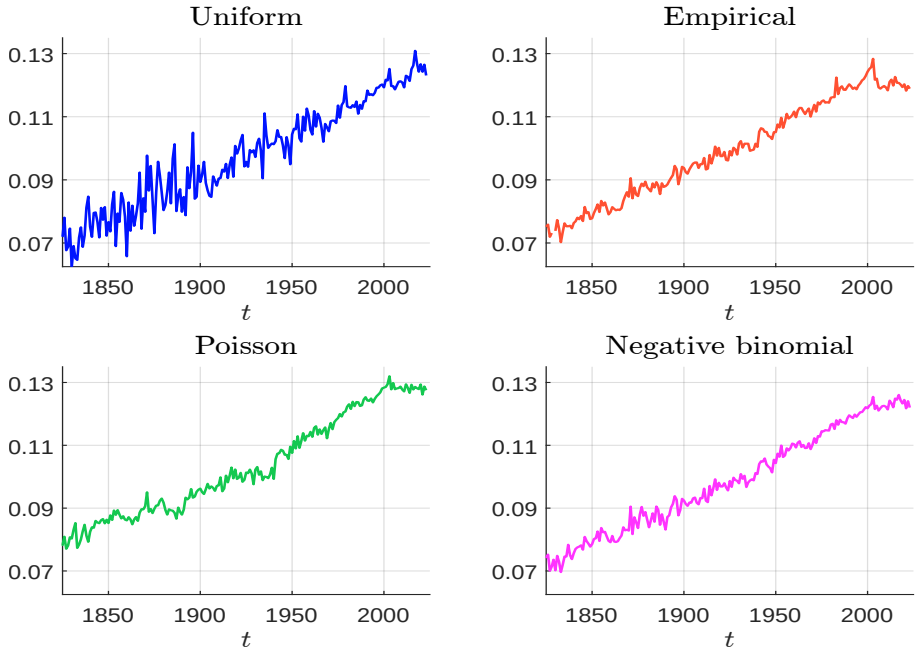


Figure 83: ARIMA forecast of Lee-Carter time index κ_t (France, male, WLS)

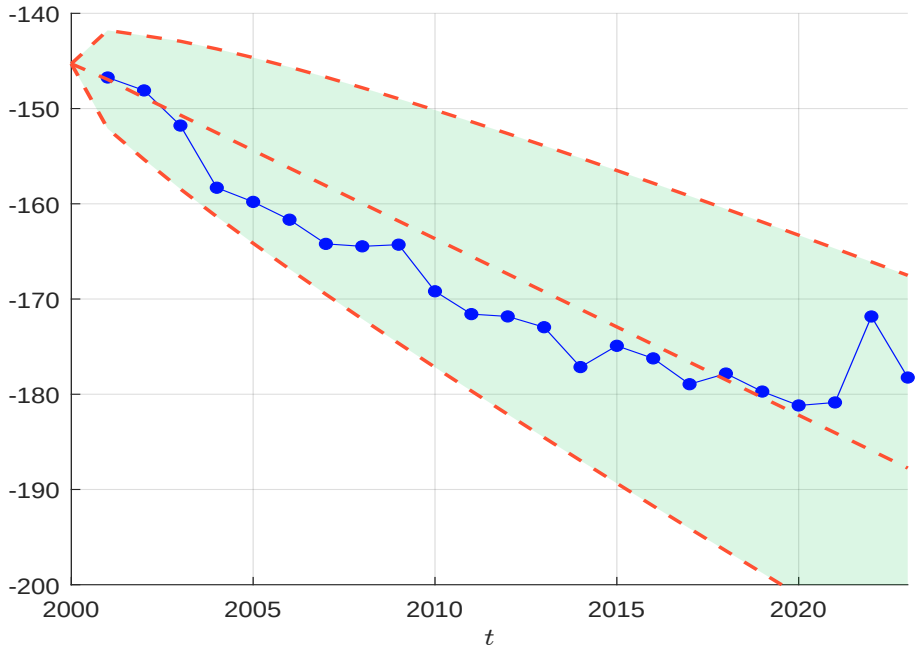


Figure 84: Time evolution of the value of the annuity contract (France, age 60, $\varrho_t = 3\%$)

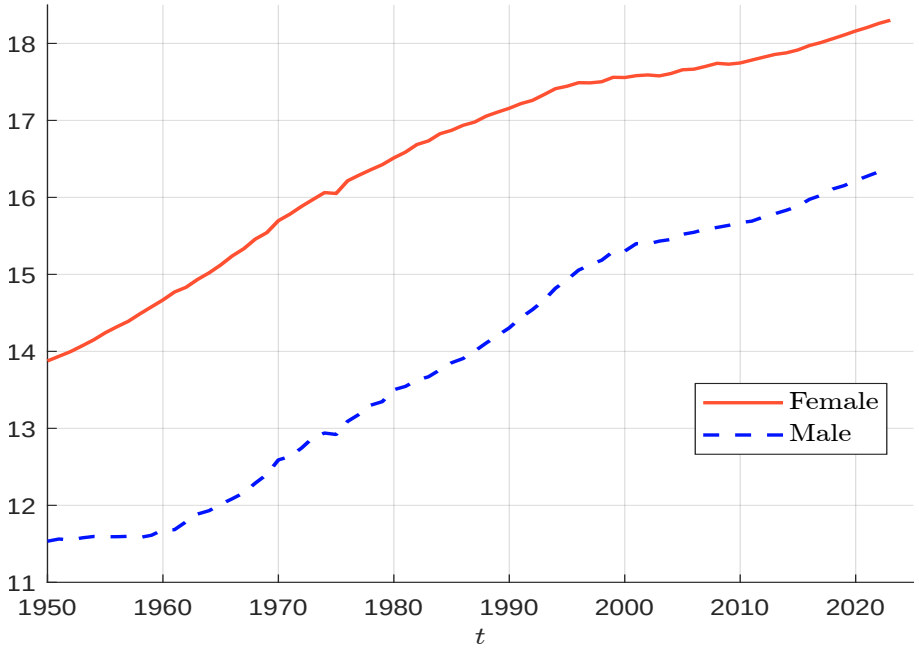


Figure 85: Impact of the discount rate on the value of the annuity contract (France, age 60, year 2000)

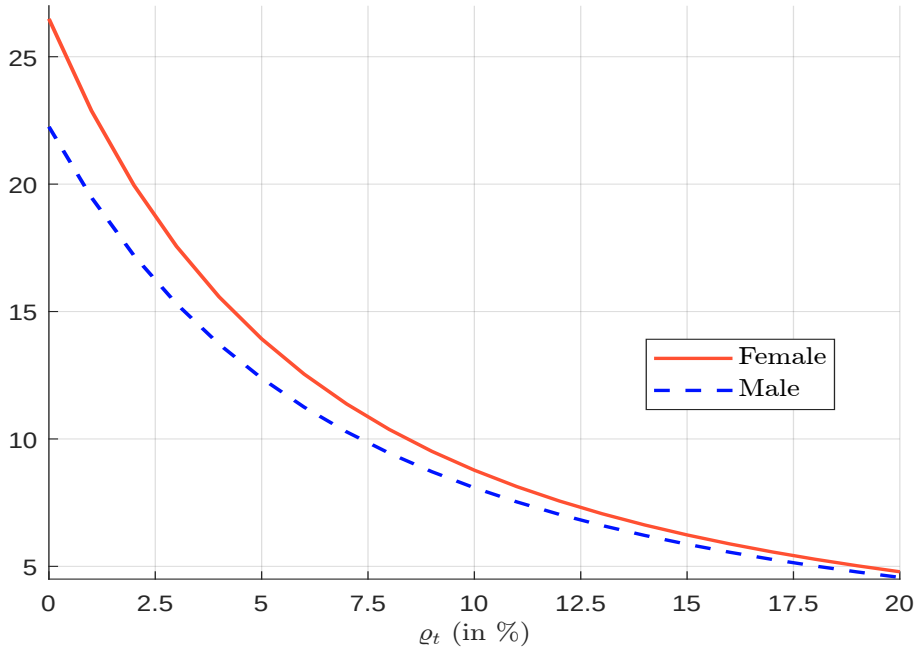


Figure 86: Impact of the age on the value of the annuity contract (France, year 2000, $\varrho_t = 3\%$)

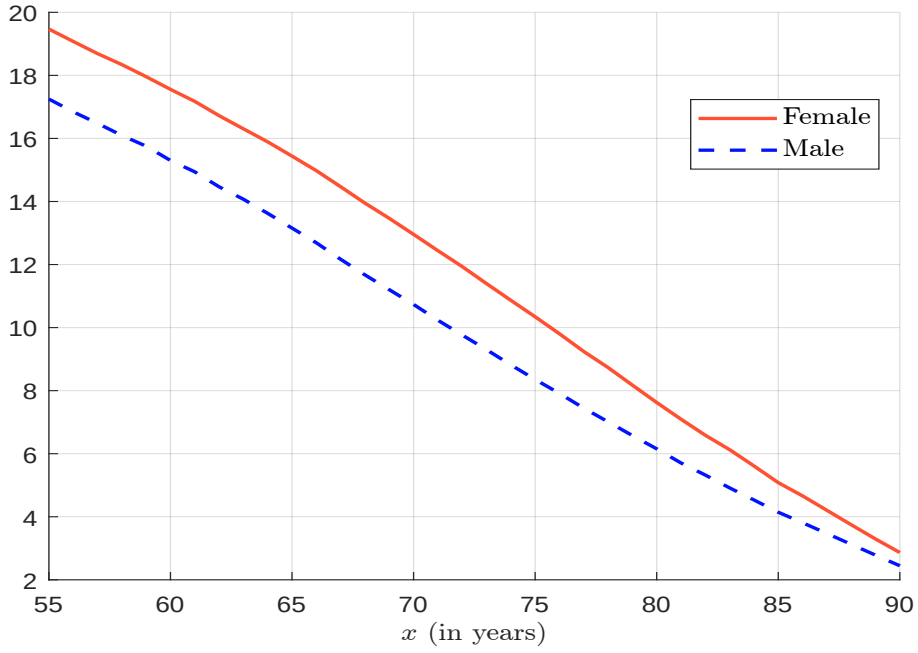


Figure 87: Simulation of the Lee-Carter time index κ_t (France, female, WLS, valuation year 2023)

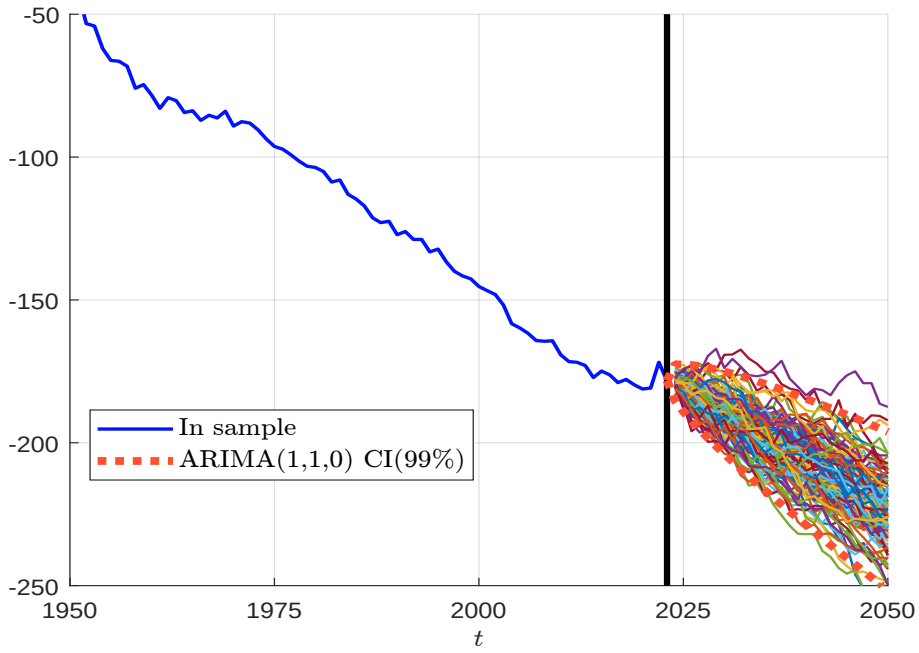


Figure 88: Time evolution of the value of the annuity contract (France, age 60, $\varrho_t = R_t^{10Y}$)

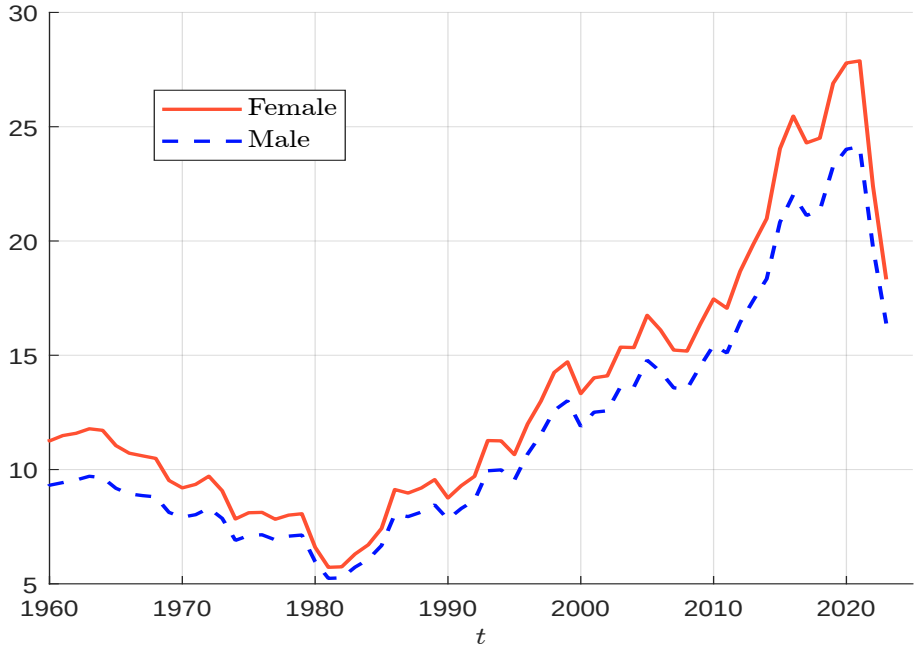


Figure 89: Valuation gap analysis (France, male, age 60, baseline model = LC WLS)

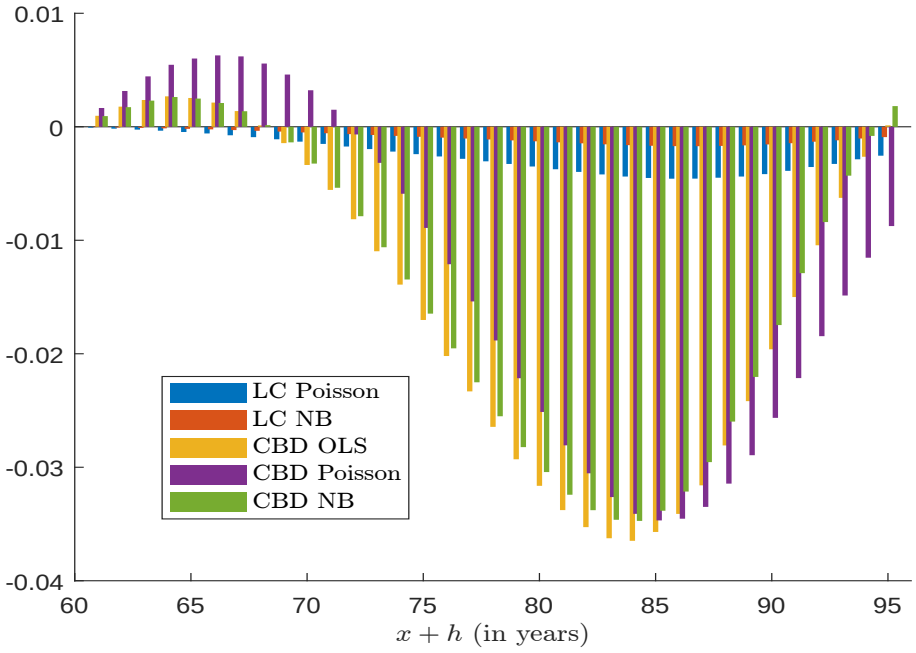


Figure 90: Cumulative gap analysis (France, male, age 60, baseline model = LC WLS)

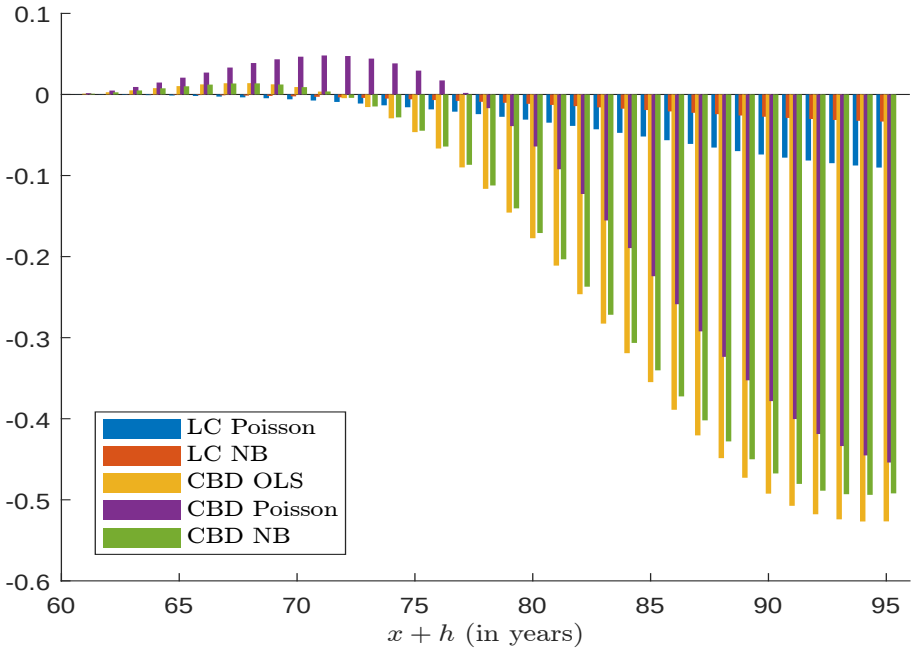
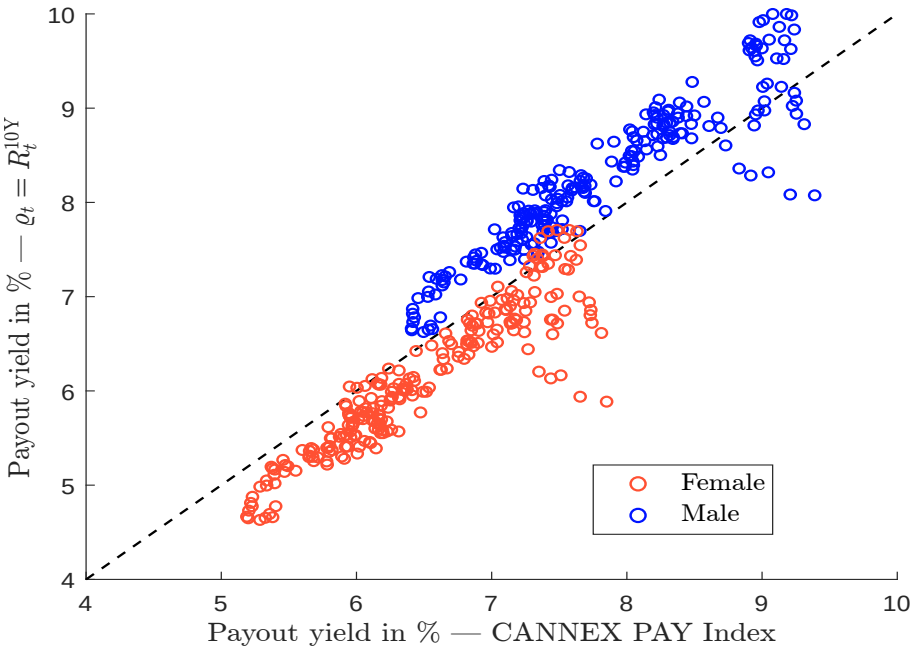


Figure 91: Scatter plot of annualized payout rates (USA, 2005–2025, CANNEX PAY Index vs. $\varrho_{t,h} = R_t^{10Y}$)



Source: CANNEX (2026) & Authors' calculations.

Figure 92: Scatter plot of annualized payout rates (USA, 2005–2025, $\varrho_{t,h} = R_t^{10Y}$ vs. $\varrho_{t,h} = R_t^h$)

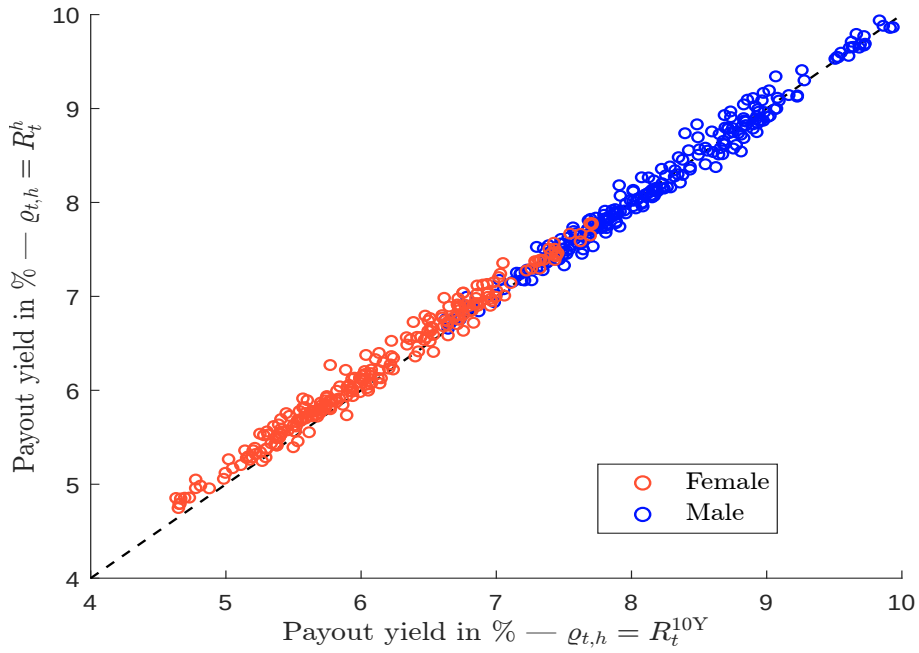


Figure 93: Evolution of the conversion age $x^\star$ in years (France, LC WLS, $r = 0$, $R = 2\%$)

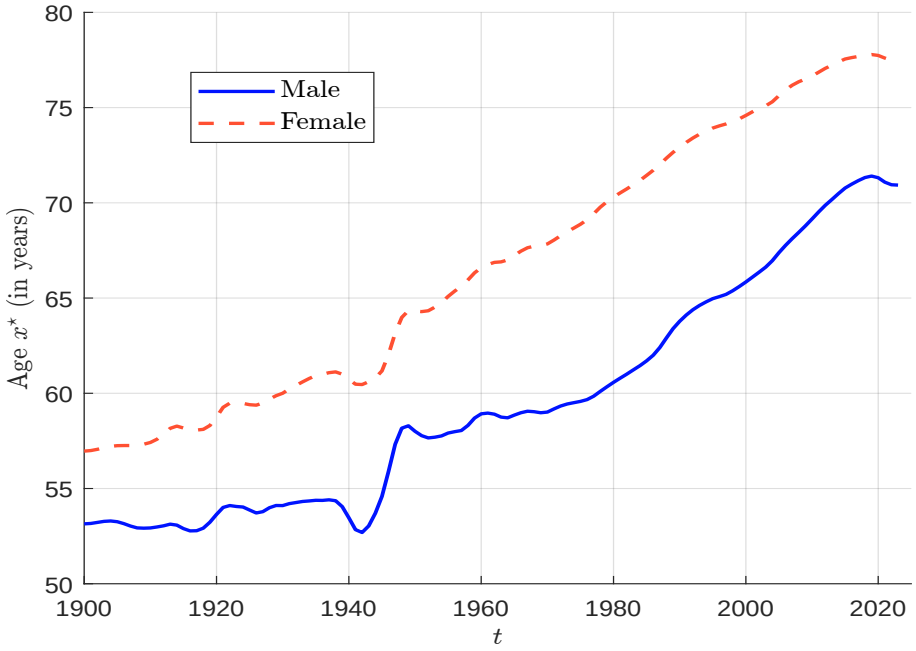


Figure 94: Expected bequest ratio (in %) by annuitization age (France, female, 2023, LC WLS)

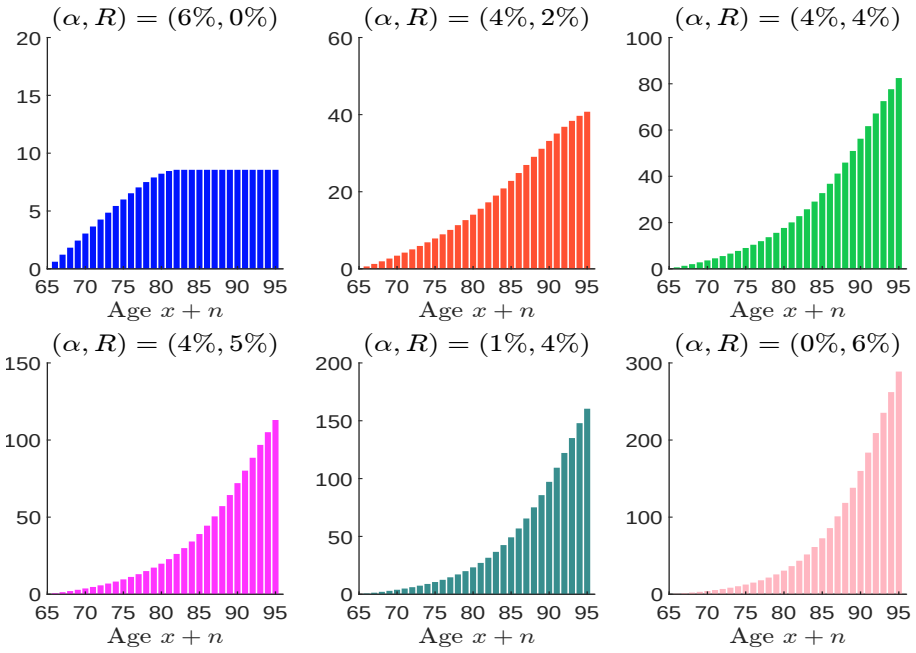


Figure 95: Optimal annuitization date (France, male, 2023, LC WLS)

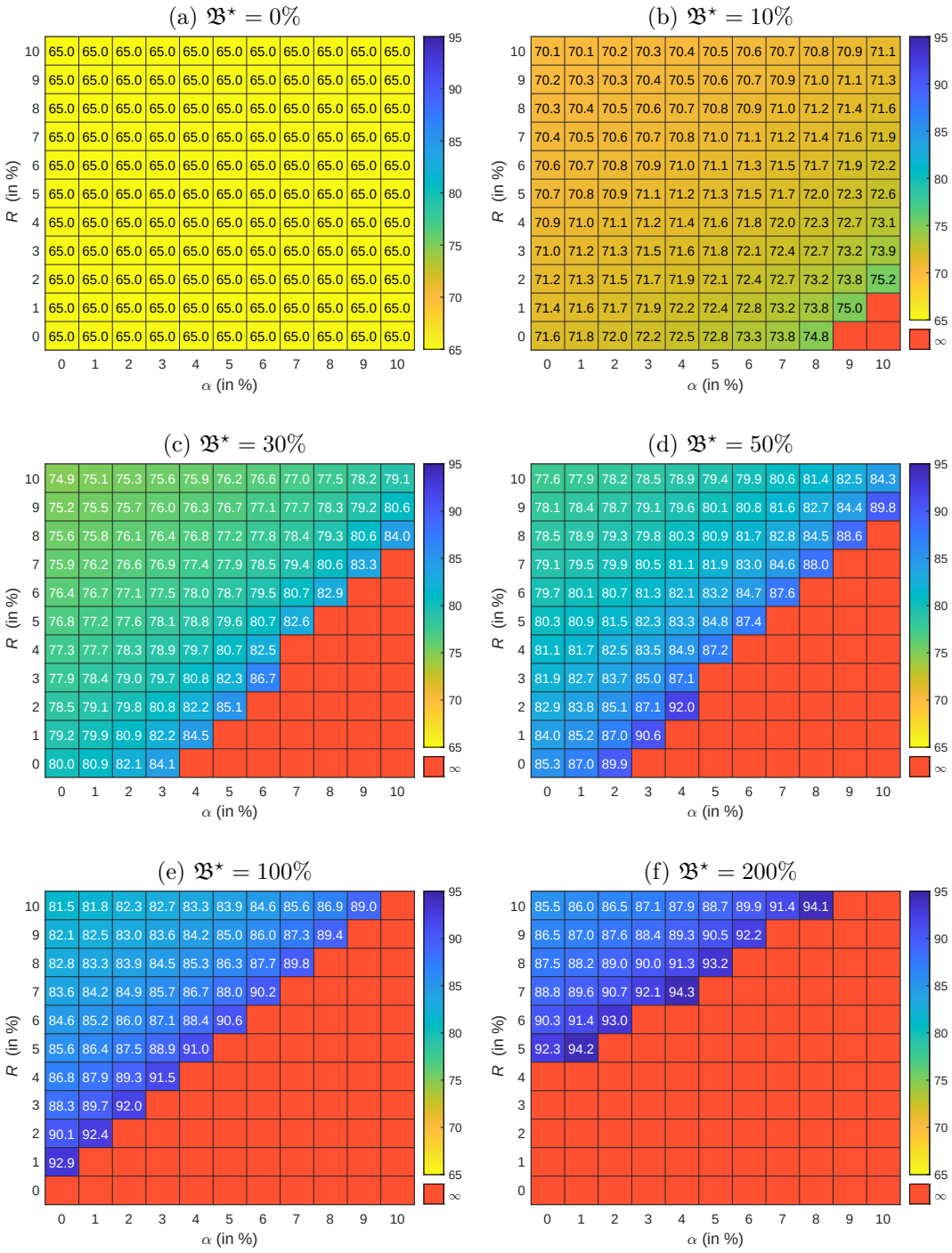


Figure 96: Optimal annuitization date (France, female, 2023, LC WLS)

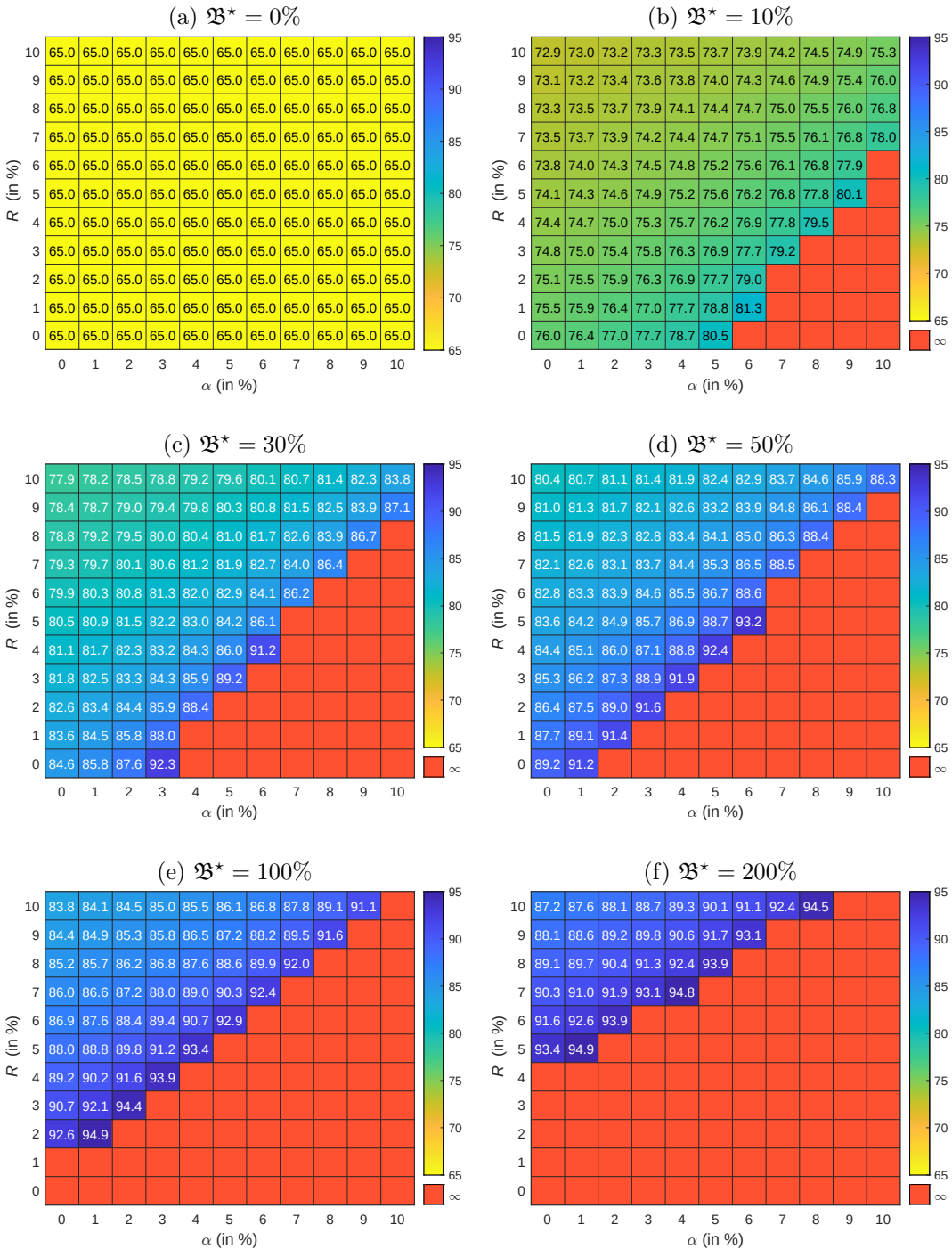


Figure 97: Optimal annuitization date with respect to α (France, male, 2023, LC WLS)

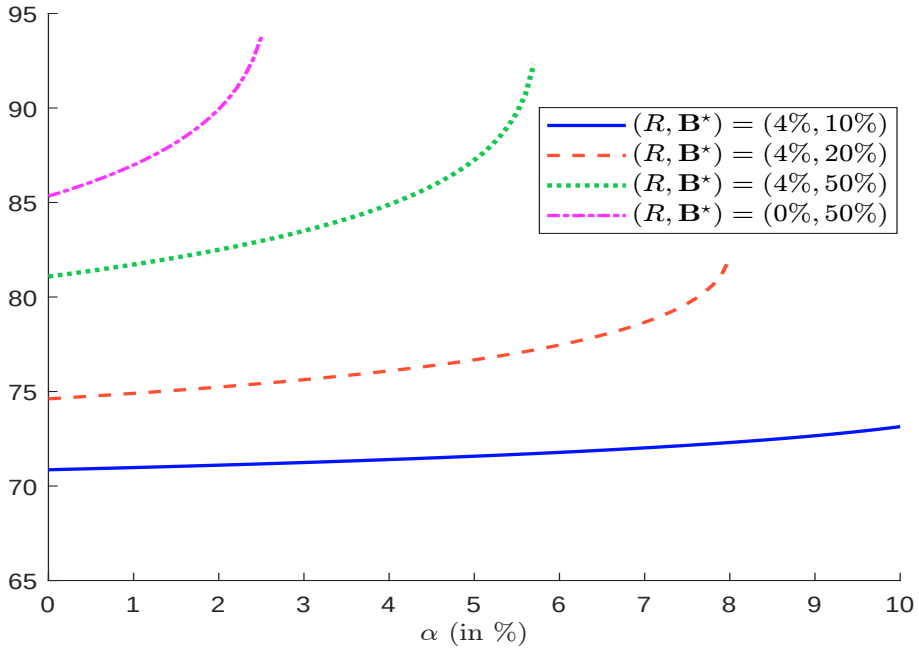


Figure 98: Optimal annuitization date with respect to R (France, male, 2023, LC WLS)

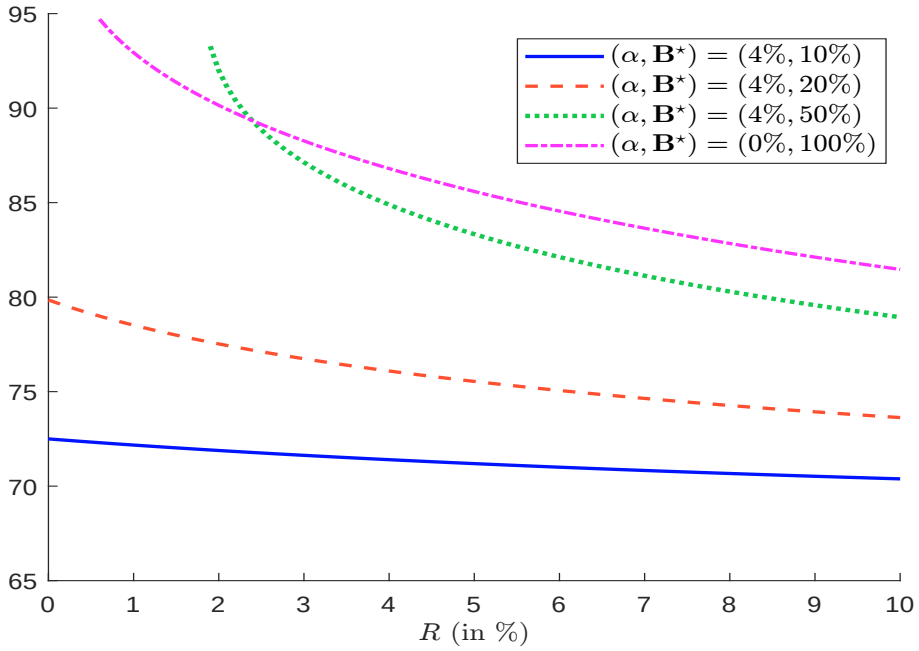


Figure 99: Fixed-date objective components — $\mathbb{E}\left[\mathcal{C}\left(n\right)\right]$, $\mathbb{E}\left[\mathfrak{B}\left(n\right)\right]$ and $\mathbb{E}\left[\mathcal{U}\left(n\right)\right]$ (France, male, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

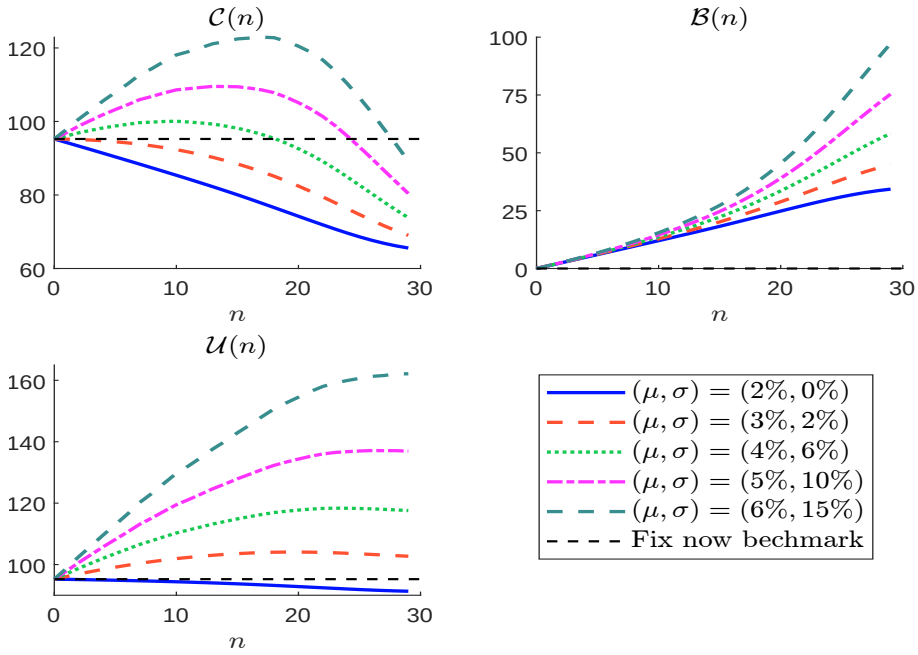


Figure 100: Expected gain $\Delta\mathcal{U}\left(n\right)$ in % (France, male, 2023, LC Poisson, $\alpha = 4\%$)

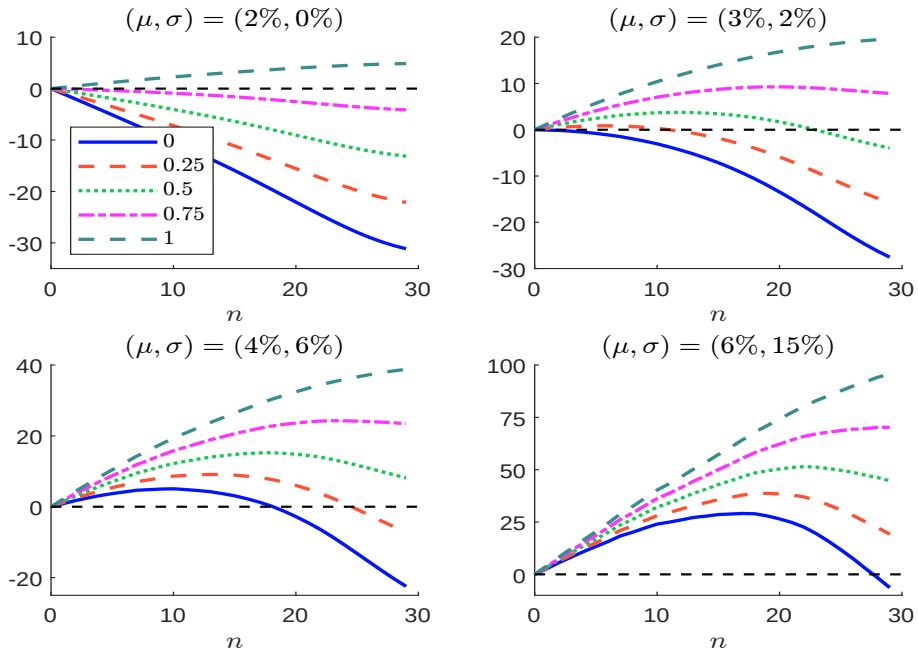


Figure 101: Probability of outperformance $P'_{\mathcal{U}}(n)$ in % (France, male, 2023, LC Poisson, $\alpha = 4\%$)

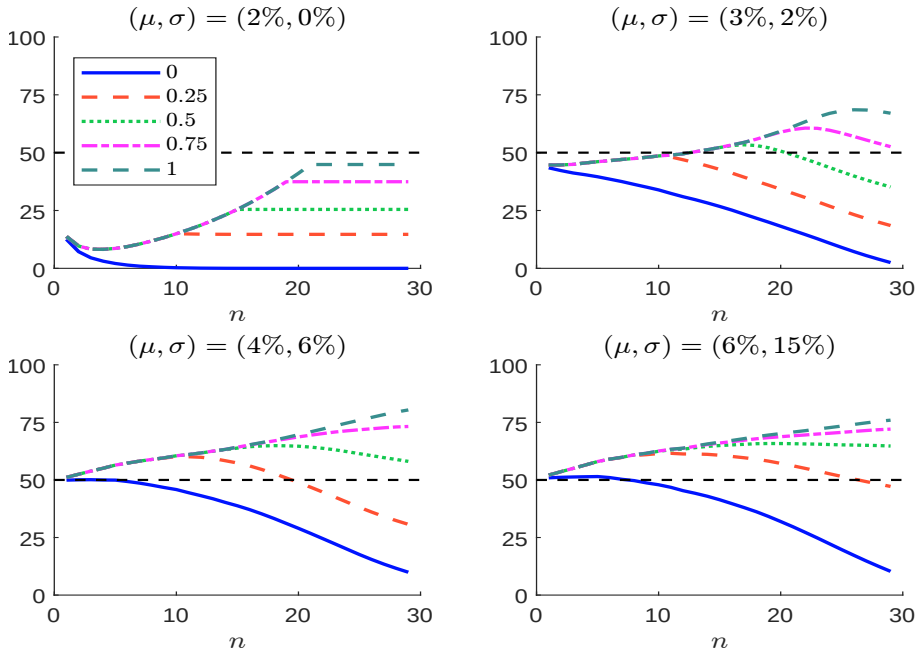


Figure 102: Fixed-date outcome diagnostics (France, male, 2023, LC Poisson, $\alpha = 4\%$, $\psi = 0.75$)

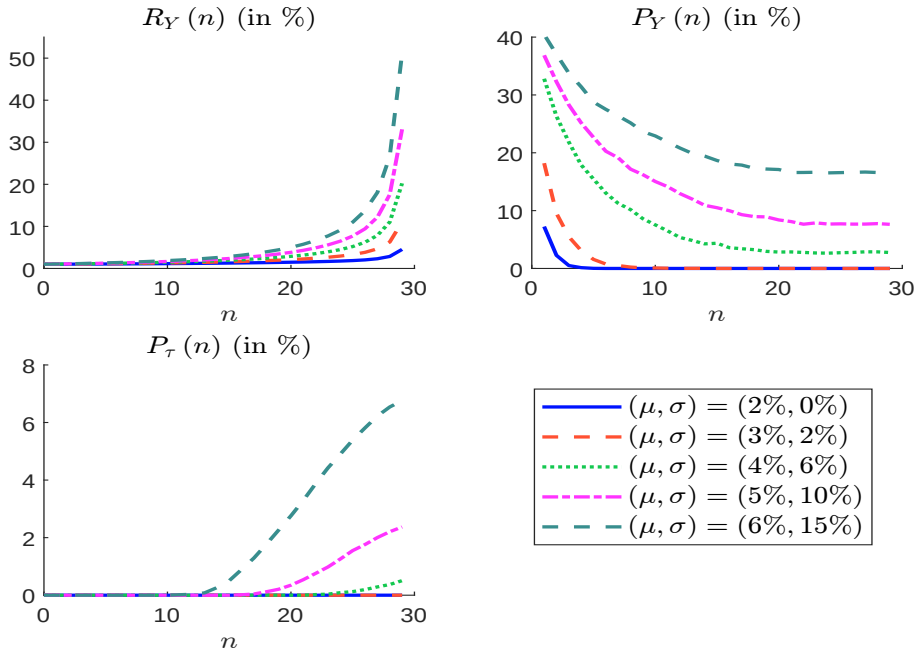


Figure 103: Optimal annuitization delay $n^\star$ across withdrawal rates α (France, male, 2023, LC Poisson, $\psi = 0.75$)

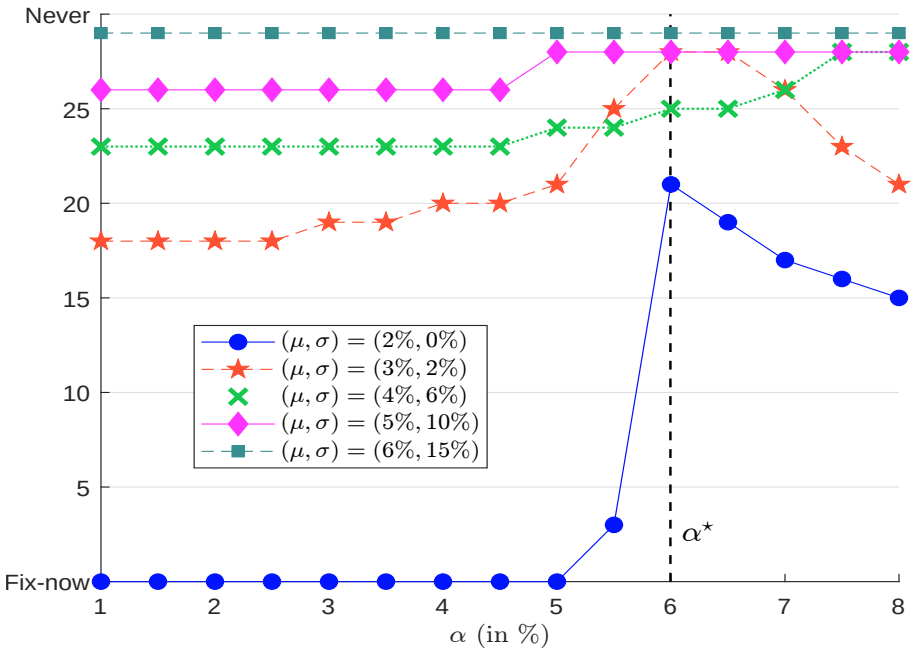


Figure 104: Realized lifetime outperformance probability $P_{\mathcal{U}}(n^\star)$ across withdrawal rates α (France, male, 2023, LC Poisson, $\psi = 0.75$)

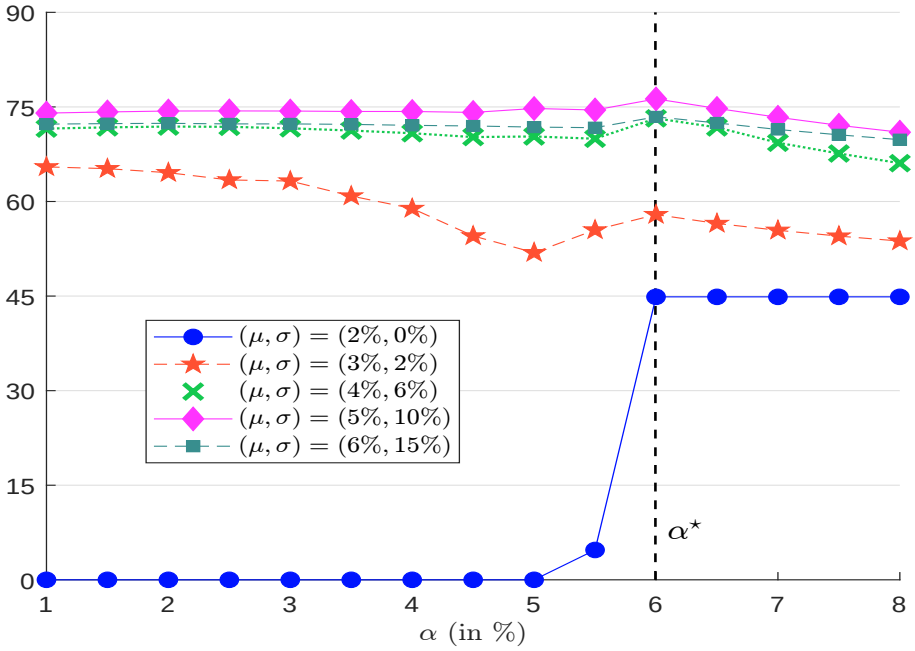


Figure 105: Probability $P_{\boldsymbol{\tau}}(n^\star)$ of liquid account depletion before annuitization across withdrawal rates α (France, male, 2023, LC Poisson, $\psi = 0.75$)

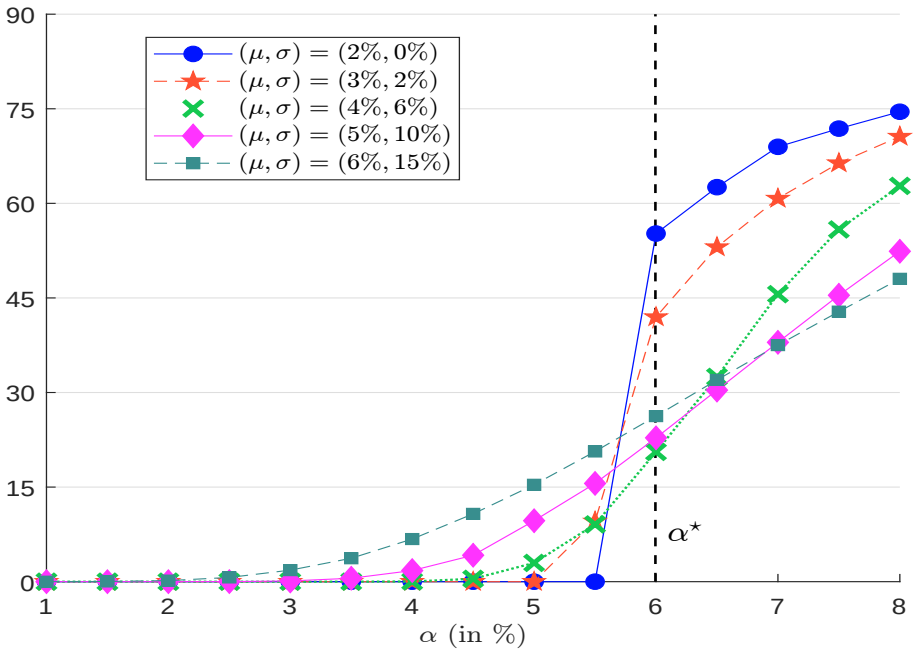


Figure 106: Optimal risky asset allocation under strong bequest motive ($\xi = 0.1$)

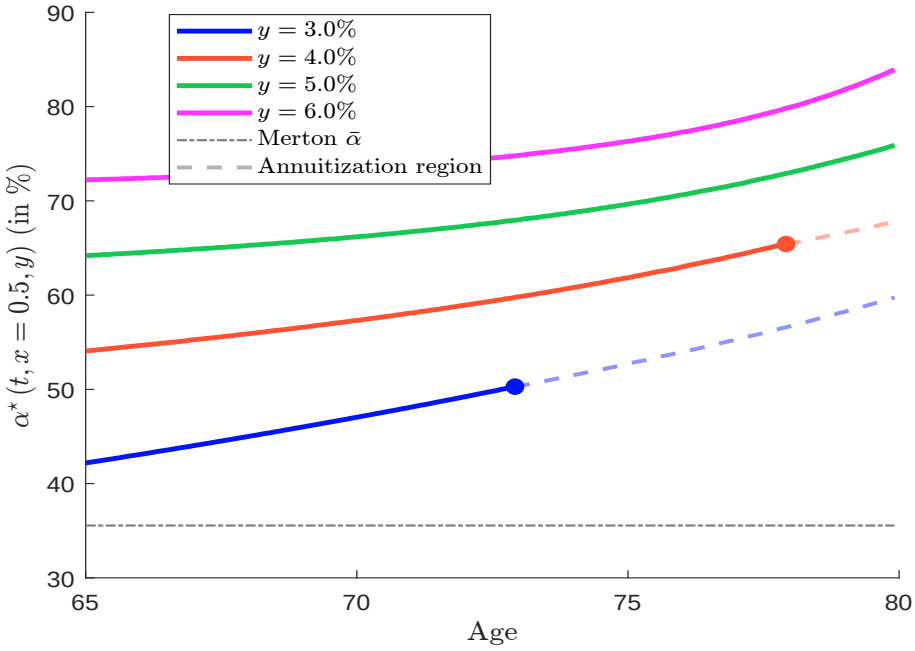


Table of Contents

1 Introduction **2**

2 Mortality modeling **15**
2.1 Mortality data . . . 15
2.1.1 Data description . . . 15
2.1.2 Data cleaning . . . 17
2.2 Lee-Carter model . . . 17
2.2.1 Model specification and calibration . . . 18
2.2.2 Extension of the Lee-Carter model . . . 23
The Poisson LC model . . . 23
The negative binomial LC model . . . 24
2.3 Cairns-Blake-Dowd model . . . 26
2.3.1 Model specification . . . 26
2.3.2 Model calibration . . . 26
2.4 Model backtesting . . . 29
2.4.1 Forecasting . . . 30
2.4.2 In-sample backtesting . . . 32
2.4.3 Out-of-sample backtesting . . . 33

3 Annuity modeling **35**
3.1 Literature review . . . 36
3.1.1 Insurance premium, actuarial valuation and annuity . . . 36
3.1.2 Contingent claim pricing and risk-neutral valuation . . . 37
3.1.3 The issue of incomplete markets . . . 38
3.1.4 Market development and practice . . . 38
3.2 Annuity pricing . . . 39
3.2.1 Contract value . . . 39
3.2.2 Deterministic valuation . . . 40
3.2.3 Stochastic valuation . . . 41
3.2.4 Empirical analysis . . . 41
Average annuity values . . . 41
Probability distribution of annuity values . . . 44
Model comparison . . . 47
3.2.5 Discount rate, risk premium and mortality credit . . . 50
Evidence of time-varying discount rate . . . 50
Computation of the risk premium . . . 51
Link to the mortality credit . . . 53
3.3 Portfolio risk analysis . . . 54
3.3.1 Portfolio liability modeling . . . 54
3.3.2 Risk measurement . . . 55
Estimation of the variance components . . . 55
Estimation of the probability distribution . . . 56
3.3.3 Empirical results . . . 56
Idiosyncratic vs. systematic risk . . . 56
Tail risk measures . . . 58
Model risk . . . 61

4 Optimal decumulation strategy **62**
4.1 Paradigm shift in decumulation and the flex-to-fix strategy . . . 62
4.2 Static analysis . . . 65
4.2.1 The mortality credit approach . . . 65
4.2.2 The actuarial approach . . . 68
Fixed-date annuitization problem . . . 68
Illustration of the flex-to-fix strategy . . . 70
The bequest motivation . . . 75
The objective function . . . 78
Numerical results . . . 79
Baseline fixed-date results . . . 82
Sensitivity analysis with respect to the withdrawal rate α . . . 91
4.3 Dynamic analysis . . . 96
4.3.1 Optimal annuitization date . . . 96
Mathematical formulation . . . 96
Annuitization problem . . . 96
Annuitization and continuation values . . . 97
Optimal solution . . . 98
Numerical results . . . 98
4.3.2 Optimal asset allocation with a target date . . . 106
The mathematical problem . . . 106
The Hamilton-Jacobi-Bellman variational inequality . . . 109
Numerical results . . . 109
Simulation results . . . 116

5 Conclusion **121**

Bibliography **127**

A Notations **134**

B Data **136**
B.1 Human mortality database . . . 136
B.2 Data cleaning . . . 136
B.2.1 Surface smoothing . . . 136
B.2.2 Outlier detection and replacement . . . 139

C Mathematical appendix **140**
C.1 Calibration of the Lee-Carter model . . . 140
C.2 The constant force approximation . . . 141
C.3 Optimal weighting schemes for estimating the CBD model . . . 142
C.4 CBD outlier detection . . . 143
C.5 Gompertz annuity pricing model (GAPM) . . . 143
C.6 Joint scenario simulation of the Lee-Carter model parameters and mortality innovations . . . 144
C.7 Expression of the best estimate of the portfolio liability . . . 146
C.8 Computation of the conditional variance $\text{var}^{(s)}\left(x_g, t\right)$ . . . 146
C.9 Probability mass of the death-time distribution under a static mortality table 147
C.10 Computation of the lifetime consumption component $\mathcal{C}^{(s)}(n)$ . . . 148
C.11 Survival time approach of the fixed-date annuitization problem . . . 149
C.12 Numerical solution of the dynamic annuitization problem . . . 150

C.12.1 State-space discretization . . . 150
Mortality state . . . 150
Wealth state . . . 151
C.12.2 Numerical evaluation of the Bellman equation . . . 151
Conditional annuity factors . . . 151
Continuation value . . . 152
C.12.3 Numerical specification . . . 152
C.13 The Hamilton-Jacobi-Bellman variational inequality . . . 152
C.13.1 The continuation region . . . 153
C.13.2 The action region . . . 155
C.13.3 The unified HJB variational inequality . . . 155
C.13.4 The free boundary problem . . . 156
C.14 Numerical solution of the HJB variational inequality . . . 157
C.14.1 Policy evaluation . . . 157
Discretization of the wait operator . . . 158
Discretization of the gradient obstacle . . . 160
C.14.2 Policy improvement . . . 160
C.14.3 The algorithmic loop . . . 161
C.14.4 Recovery of the optimal lump-sum transfer $\Delta Z^\star$ . . . 162

D Additional results **163**
D.1 Tables . . . 163
D.2 Figures . . . 164