Canadian Institute of Actuaries | Institut canadien des actuaires

Research paper

The Impact of Longevity Annuity Provision on Canadian Retirement Income Planning

Prepared by:
Rui Zhou, ACIA, University of Melbourne
Johnny S.-H. Li, University of Waterloo
Kenneth Zhou, ACIA, University of Waterloo

September 2020

Document 220135

Ce document est disponible en français

The Impact of Longevity Annuity Provision on Canadian Retirement Income Planning

Rui Zhou, Johnny S.-H. Li, Kenneth Zhou

A longevity annuity is a deferred annuity where payments start very late in life; i.e., well after the normal retirement age. By transferring the risk of outliving retirement savings at high ages to annuity providers, a longevity annuity provides annuitants with enhanced later-life financial security. In this report, we examine the impact of longevity annuity provision on retirement income planning based on Canada's tax rules and retirement system. A dynamic life cycle framework is developed to study welfare increases and consumption pattern changes resulting from the provision of longevity annuities to Canadian retirees. To determine the optimal choices in the life cycle model, we propose a modified general endogenous grid method (GEGM) which addresses the non-differentiability problem arising from realistic tax rules and a realistic retirement system. This life cycle framework and modified GEGM are further applied to explore how individuals in different social classes respond to the access to longevity annuities.

1 Introduction

With the rapid mortality improvement we have been experiencing in recent years, there is a high likelihood of retirees underestimating their life expectancy and outliving their retirement assets. Volatile investment return exacerbates the problem by the potential poor growth of retirement savings. A long-recognized solution to secure retirement income is an annuity. Since the seminal work of Yaari (1965), economists have agreed on the substantial social welfare benefits of immediate annuities. Economists have further concurred that retirees should convert most or all of their assets into annuities at retirement. However, rates of voluntary annuitization have remained extremely low, a phenomenon known as the "annuity puzzle."

Milevsky (2005) introduces the longevity annuity, also known as the advanced-life delayed annuity (ALDA), as a new way to provide lifetime retirement income. A longevity annuity is a deferred annuity where payments start very late in life; i.e., well after the normal retirement age. By transferring the risk of outliving retirement savings at high ages to annuity providers, a longevity annuity provides annuitants with enhanced later-life financial security. Since the payments start at very advanced ages, the cost of a longevity annuity is much lower than that of an immediate annuity, hence reducing the loss of liquidity due to annuitization.

Longevity annuities were first made accessible to 401(k) plans in the US in 2014. More than 10 A-rated (according to A.M. Best) or better insurers[1] in the US currently provide the product. Although the sales of longevity annuities have recently taken off, their share in the annuity

[1] www.forbes.com/sites/mattcarey/2018/08/13/qlac-pricing-improves-as-more-insurers-offer-product-bond-yields-rise/#43dd501b36e0

market is still small. The 2019 Canadian federal budget also introduces ALDAs and allows them to be purchased under the registered retirement saving plans (RRSPs), registered retirement income funds (RRIFs), deferred profit-sharing plans, pooled registered pension plans and defined contribution registered pension plans (RPPs).

Existing research on longevity annuities is mainly conducted in the US context. Gong and Webb (2010) study the money worth of a longevity annuity and compare retirement wealth decumulation strategies with and without a longevity annuity. Their results show that a longevity annuity provides a substantial proportion of longevity risk protection offered by an immediate annuity at much lower cost. However, the assumptions used in this paper are restrictive, such as a fixed percentage of wealth converted to a longevity annuity, deterministic mortality projections, constant interest rate, and constant investment return. More recently, Pfau, Tomlinson, and Vernon (2016) analyze whether using a longevity annuity in combination with systematic withdrawal plans produces higher expected income with the same amount of risk, compared to using single-premium immediate annuities or guaranteed lifetime withdrawal benefit annuities. Horneff, Maurer, and Mitchell (2016) develop a realistic life cycle model for a US individual and measure welfare improvements due to including longevity annuities in the 401(k) plan payout. The researchers conclude that the access to a longevity annuity boosts welfare by 5–20% and the optimal percentage of a plan asset converted to a longevity annuity is 8–15%. While these studies provide valuable information for Canadian retirees, Canada's tax rules, retirement system, and health care system differ from those of the US These differences may significantly affect conclusions.

In this report, we examine the impact of longevity annuity provision on retirement income planning based on realistic Canadian tax rules and Canadian-specific demographic assumptions. More specifically, we develop a life cycle model for retirement income planning that allows stochastic models for mortality experience and investment returns. The optimal consumption and saving decisions are determined by maximizing the lifetime utility of a representative Canadian male.

Life cycle model is a standard economic approach to examine lifetime choices including consumption, saving, investment, and labour supply. Horneff et al. (2016) formulated a life cycle model to measure welfare improvements for US retirees due to including longevity annuities in the 401(k) plan payout. Our study, with its focus on Canadian retirees, differs from that of Horneff et al. (2016). Our life cycle model incorporates Canadian tax rules and applies stochastic mortality models calibrated to Canadian historical mortality rates. We adopt the widely used Lee-Carter model (Lee and Carter, 1992) for Canadian national mortality data and the augmented comment factor model (Li and Lee, 2005) for CPP pension group mortality.

To solve the optimization problem, we propose a modified GEGM. The EGM, first proposed by Carroll (2006), requires significantly less computation time compared to the traditional value function iteration (VFI) approach by avoiding iterative numerical integrations. Druedahl and Jørgensen (2017) generalize EGM to address several challenges arising in complex life cycle models including insufficient first-order conditions (FOCs), no prior information on whether constraints are binding, and irregular endogenous grid. A key step of the GEGM is to divide the optimization problem into segments and evaluate the value function by segment. Druedahl and

Jørgensen (2017) assume that the value function is differentiable in each segment. However, the realistic tax rules and retirement income system assumed in our analysis introduce kinks in the wealth process and hence non-differentiable points. To allow for a non-differentiable value function, we propose a new segmentation approach based on the threshold values used for personal tax and retirement income.

Using the proposed life cycle model and modified GEGM, we examine welfare increases and consumption pattern changes resulting from the provision of longevity annuities to Canadian retirees. We further study how individuals in different socio-economic classes respond to the access to longevity annuities. Socio-economic classes affect not only the income but also the mortality experience of an individual. In this analysis, we examine the relation between socio-economic classes and mortality experience exhibited in the Canada Pension Plan (CPP) mortality data and incorporate this relation in the life cycle model. The CPP mortality data categorize pensioners into 11 groups by the CPP pension amount received, and provide historical mortality experience for each pension group. Since the CPP pension amount is determined by the income earned and the number of years worked during the working life, the pension groups can be viewed as an indicator of the pensioners' socio-economic classes.

The remainder of the report is organized as follows: Section 2 describes the Canadian retirement income system; Section 3 looks at the longevity annuity in depth; Section 4 fits the widely used Lee–Carter model to Canadian mortality data; Section 5 develops the life cycle model; Section 6 discussess the GEGM approach; Section 7 demonstrates the simulated consumption and saving patterns with and without longevity annuity access for a Canadian male; Section 8 examines the impact of socio-economic class on mortality experience and incorporates this impact into the assessment of utility gain from longevity annuity access; and Section 9 concludes the report.

2 Canadian retirement income system

2.1 The three pillars

Canadian retirement income typically comes from three resources:

- The CPP or Quebec Pension Plan (QPP).
- Old Age Security (OAS).
- Employer-sponsored pension plans and personal savings and investments.

The three resources are often referred to as the three pillars of Canadian retirement income system. The three pillars differ in funding sources, tax treatments, and how and when payments are taken from them. These differences significantly affect how individuals save for their retirement and how they spend their savings during retirement. Therefore, we describe the three pillars in detail in the following subsections. In particular, we focus on a Canadian citizen who has lived and worked in Ontario, Canada, during his entire lifetime.

2.2 CPP/QPP

Almost every person over the age of 18 who works in Canada outside of Quebec and earns more than a basic exemption amount must contribute to the CPP. The CPP operates throughout Canada, except in Quebec, where the QPP provides similar benefits. Since we perform analysis for a Canadian citizen living in Ontario in this paper, we only provide details about the CPP.

The CPP contribution amount is determined by basic exemption amount, pensionable earnings, and contribution rate:

- Basic exemption amount: The basic exemption amount is $3,500 in 2018. If the annual earnings are below $3,500, no CPP contribution is made.
- Pensionable earnings: Pensionable earnings are equal to the annual earnings capped at a maximum amount which is set each January based on increases in the average wage in Canada over the past year. The maximum pensionable earnings in 2018 are $55,900.
- Contribution rate: The contribution rate on the pensionable earnings is 9.9% in 2018, split equally between employer and employee. If self-employed, the self-employed person pays the full 9.9%.

Denote the average wage in year t by AW_t and the annual earnings of an employee in year t by Y_t. Assume that basic exemption amount and contribution rate remain unchanged in the future years. The annual CPP contribution made by the employee or the employer in year t can be expressed as follows:

$$0.5 \times 9.9\% \times \max\left(\min\left(Y_t, 55900\,\frac{AW_{t-1}}{AW_{2017}}\right) - 3500, 0\right).$$

As a result, the maximum contribution for employer or employee in 2018 is $2,593.80 each.

The standard age from which the CPP retirement pension can be taken is the month after the 65th birthday. However, the CPP pension can be taken from as early as age 60 with a reduced pension amount or after age 65 with an increased pension amount. The CPP pension amount is determined by how many years of CPP contributions are made during employment and how much are the contributions. In general, the longer and the higher the CPP contributions, the greater the CPP pension will be.

The CPP pension amount calculation permits dropping out for some periods in which individuals may have relatively low or no earnings. There are two common drop-outs: the general drop-out and the Child Rearing Provision. The general drop-out allows up to 17% of the base contributory period of the lowest earnings to be dropped from the calculation. This provision applies to all CPP contributors. The Child Rearing Provision permits dropping out for the periods with lower earnings due to caring for a dependent child under the age of seven. An example of a CPP pension amount calculation is shown in Appendix A.

In practice, the CPP pension amount increases with the Consumer Price Index (CPI), but remains unchanged when the CPI decreases. In this paper, we assume no inflation, and hence the CPP pension amount remain unchanged. There are other benefits under the CPP program, such as disability benefit and survivor's pension. We only consider the CPP retirement pension since our numerical analysis will be based on a healthy single individual.

2.3 OAS

The OAS program, which includes the OAS pension, Guaranteed Income Supplement (GIS), Allowance, and Allowance for the Survivor, is funded by the general revenues of the Government of Canada. There is no direct contribution to the program from an individual.

Both the OAS pension and GIS depend on the previous year's income. The maximum monthly OAS pension amount regardless of marital status is $589.59 in 2018. If the annual net world income in the previous year exceeds the threshold amount ($75,910 for 2018), recovery tax will be deducted from the OAS pension amount at the rate of 15 cents for each dollar of income above the threshold until the OAS pension amount reaches 0. Both the maximum OAS pension amount and the threshold amount increase with CPI. In practice, the monthly OAS pension amount and the threshold amount are adjusted quarterly and annually respectively. For convenience, we assume that both are adjusted annually. The OAS pension amount received in year t can be written as:

$$\max\left(12 \times 589.59 \frac{CPI_{t-1}}{CPI_{2017}} - 0.15\max\left(Y_t - 75910 \frac{CPI_{t-1}}{CPI_{2017}}, 0\right), 0\right).$$

It is easy to verify that the maximum annual income to receive the OAS pension is $123,077 for 2018. Note that the OAS pension amount is treated as taxable income.

The GIS is a monthly non-taxable benefit to OAS pension recipients who have a low income and are living in Canada. The amount of the GIS received depends on their marital status and the previous year's income. The maximum monthly GIS payment amount for a single pensioner is $880.61. The actual amount of GIS is determined using a set of complex rate tables[2] and decreases almost linearly with income (excluding the OAS pension and GIS) received. The maximum annual income to receive the GIS pension is $17,880 for a single pensioner. We perform a linear regression of GIS pension amounts shown in the GIS rate table on their corresponding income levels. We find that the GIS pension received in year t can be approximated as:

$$\max\left(12 \times 880.61 \frac{CPI_{t-1}}{CPI_{2017}} - 0.5909Y_t, 0\right).$$

The Allowance may be received by an individual who is 60–64 years of age with a spouse or common-law partner receiving the OAS pension and eligible for the GIS. The Allowance for the Survivor amount may be received by a widowed individual who is 60–64 years of age. We do not provide details about the allowance and Allowance for the Survivor amount since our analysis focuses on a single individual.

2.4 Employer-sponsored pension plans and personal savings and investments

Employer-sponsored pension plans, such as a group RRSP or an RPP, are contributed to regularly by both employee and employer. There are two main types of employer-sponsored pension plans: defined contribution plans and DB plans. Our analysis considers only defined contribution plans. An individual can also establish his individual RRSP account with his bank.

[2] www.canada.ca/content/dam/canada/employment-social-development/migration/documents/assets/portfolio/docs/en/cpp/oas/sv-oas-jul-sept-2019.pdf

The total annual contribution to an RPP and RRSP is limited to 18% of earned income in the previous year up to a maximum amount, which is set at $26,230 in 2018. In this report, we use RPP/RRSP to represent both employer-sponsored pension plans and individual RRSPs for simplicity. Since contributions made to RPPs/RRSPs are tax-deductible, RPPs/RRSPs provide significant tax incentives for high-income earners.

December 31 of the year that an individual turns 71 years old is the last day that contributions can be contributed to an RRSP. In the year an individual turns 71 years old, the RPP/RRSP has to be:

- withdrawn;
- transferred to an RRIF; or
- used to purchase an annuity.

Withdrawing all the RPP/RRSP at once is often not the best strategy because withdrawals are treated as income, and personal income tax applies. There is no withholding tax when money is transferred directly to an RRIF or used to purchase an annuity. However, income tax needs to be paid when the payment of the RRIF or annuity starts. A minimum RRIF withdrawal, the amount of which depends on the age of retiree, must be made every year once converted. Earnings in an RRIF are tax-free.

Besides RRSPs, another registered saving plan with tax incentives is the Tax-Free Savings Account (TFSA) program which began in 2009. A Canadian who is 18 and older can invest money in a TFSA tax-free throughout their lifetime. While contributions to a TFSA are not tax-deductible, income earned in the account is generally tax-free, even when it is withdrawn. TFSA contributions are limited to $5,500 for the year 2018. The TFSA annual room limit is indexed to inflation and rounded to the nearest $500. Unused TFSA contribution room can be accumulated to future years. The TFSA contribution room is determined by the annual TFSA dollar limit plus unused TFSA contribution room from the previous year and any withdrawals made from the TFSA in the previous year.

3 Longevity annuity

The longevity annuity, first proposed by Milevsky (2005), as mentioned above, is a deep-deferred annuity that begins to make annuity payments late in life. It has the benefit of low costs and income security for high ages. Assume that an individual retires when he turns 65 at the beginning of year t and converts a portion of his retirement savings into a longevity annuity, which begins payment from the age of x_{a}. Denote the central death rate of an x-year-old in year t by $m_{x,t}$. Assuming constant force of mortality between integer ages, the death probability of this individual in year t is $q_{x,t} = 1 - e^{-m_{x,t}}$. The probability of this individual aged x at the beginning of year t surviving T years can be written as:

$$ {}_{T}S_{x,t} = \prod_{j=0}^{T-1} (1 - q_{x+j,t+j}). $$

The present value of a longevity annuity issued to an x-year-old at the beginning of year t with a $1 payment in advance starting from age x_a can be written as:

$$ {}_{x_a - x|}\ddot{a}_{x,t} = \sum_{i=x_a-x}^{\omega-x-1} (1+r)^{-i}{}_{i}S_{x,t}. $$

where r is the rate of return earned by the annuity provider and ω is the limiting age. We assume that no one lives beyond 100 years old and thus $\omega = 100$.

A longevity annuity provides a significant tax benefit. An ALDA will make regular payments no later than the end of the year the annuitant reaches age 85 under the 2019 Canadian federal budget plan, while a registered annuity starts payment no later than the end of the year the annuitant turns age 71 under existing rules. The late payments provide tax deferral since annuity payments are taxed at the personal income marginal tax rate. In addition, the calculation of minimum withdrawal from registered accounts excludes the asset value converted to an ALDA. Therefore, an ALDA purchase reduces the minimum withdrawal required from registered accounts, thereby providing further tax deferral. The ALDA purchase will be limited to 25% of the registered account value at the end of the previous year and also subject to a lifetime $150,000 limit which will be indexed to inflation. Similar tax treatment and purchase limits apply to a longevity annuity allowed in the 401(k) plan in the US.

4 Mortality modelling

4.1 Lee–Carter model

Mortality rates determine the life expectancy of an individual and the cost of a longevity annuity. As a result, mortality rates play an important role in the choices that an individual makes for retirement planning. To depict the mortality dynamics, we consider the widely used Lee–Carter model. The Lee–Carter model in its original form (Lee and Carter, 1992) can be expressed mathematically as

$$\ln(m_{x,t}) = \beta_x^{(0)} + \beta_x^{(1)}\kappa_t, \tag{1}$$

where $m_{x,t}$ denotes the central death rate at age x and in year t, $\beta_x^{(0)}$ is the average level of mortality (in log scale) over time, $\beta_x^{(1)}$ is the age-specific sensitivity to the time-varying factor, and κ_t, which governs the dynamics of central death rates at all ages. The parameters $\beta_x^{(0)}$, $\beta_x^{(1)}$, and κ_t can be estimated by maximum likelihood estimation (MLE) assuming that the number of deaths for age x in year t, $D_{x,t}$, follows Poisson distribution with mean equal to $E_{x,t}m_{x,t}$ where $E_{x,t}$ is the number of exposure-to-risk for age x in year t. The loglikelihood function can be written as

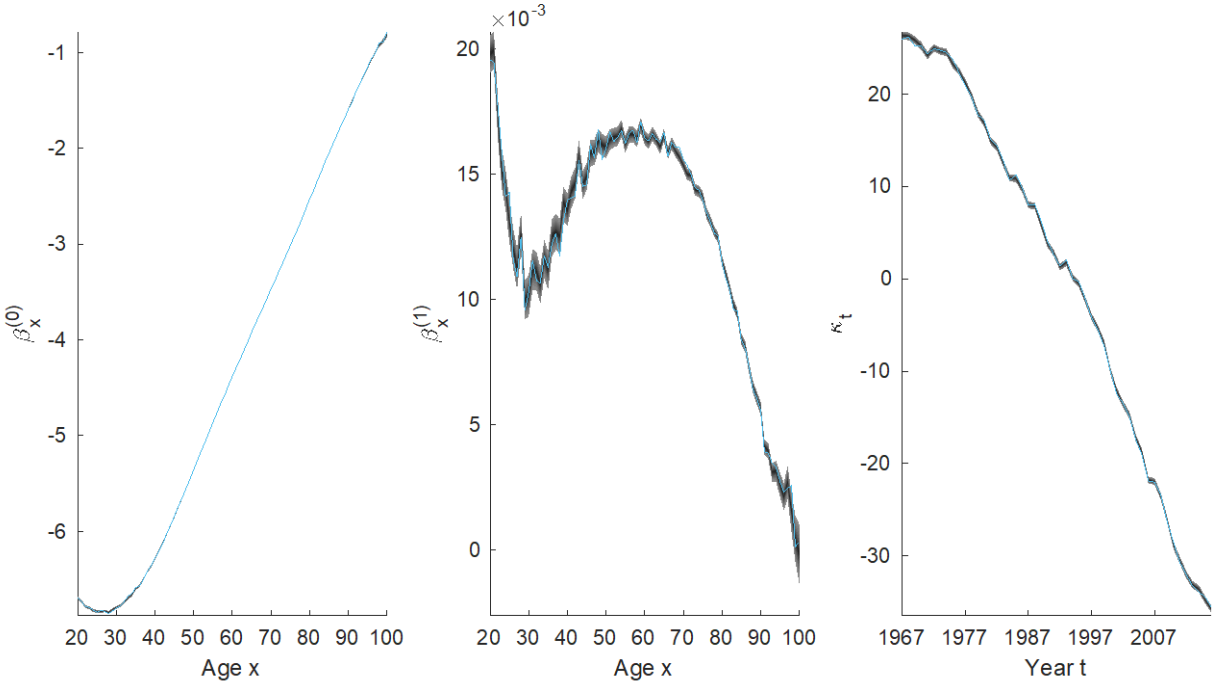
$$\sum_{t=t_0}^{t_1}\sum_{x=x_0}^{x_1} D_{x,t}\ln[E_{x,t}m_{x,t}] - E_{x,t}m_{x,t} - \ln[D_{x,t}!],$$

where $[t_0, t_1]$ is the sample period and $[x_0, x_1]$ is the sample age range and $D_{x,t}!$ is the factorial function of $D_{x,t}$.

4.2 Mortality data

We use two sets of mortality data in our analysis. The first set is the Canadian male mortality data obtained from the Human Mortality Database[3] (HMD) for the sample age range of 20–100 and the sample period of 1967–2016. The second set is the CPP male mortality data by pension groups for the age range of 65–89 and the sample period of 1991–2015. In this section, we use the HMD data to illustrate the estimation of a Lee–Carter model. We will provide more details about the CPP data in Section 8.1.

Figure 1: Lee–Carter parameter estimates for Canadian males using HMD data



The estimated parameters using the HMD data and corresponding far charts that represent the 90% confidence intervals of the parameter estimates are shown in Figure 1. The confidence intervals of the parameter estimates are obtained using the bootstrapping method proposed by Brouhns et al. (2005) with 1000 bootstrapped samples. Each sample of death count at age x in year t is generated as a random number from Poisson distribution with mean equal to the actual number of deaths at age x in year t. The confidence interval of the estimated $\beta_x^{(0)}$ is very narrow and close to unobservable while the confidence interval of the estimated $\beta_x^{(1)}$ is much wider. The estimated κ_t shows a downward trend, indicating that mortality rates decrease over time.

We further model the series of estimated κ_t by a time-series process. Since the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of $\kappa_{t+1} - \kappa_t$ are both insignificant for lags up to 10 as shown in Figure 2, we select random walk with drift to model k_t. This simple model for k_t makes the optimization of the life cycle model more

[3] https://www.mortality.org/

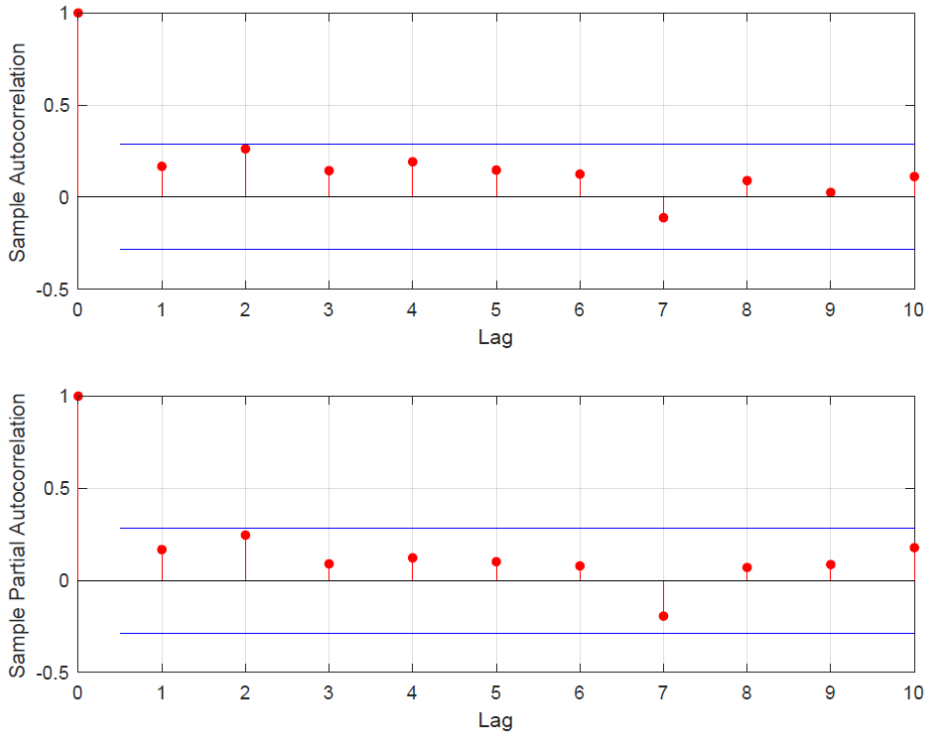
manageable. The random walk with drift model for k_t can be written as:

$$\kappa_{t+1} = \kappa_t + \mu + \sigma z_{t+1},$$

where μ and σ are constants, and $\{z_t\}$ is a sequence of i.i.d. standard normal random variables.

The estimated μ and σ using the original sample are -1.2668 and 0.9027 respectively. The 90% confidence intervals of the estimated μ and σ using the bootstrapping method are $[-1.2790, -1.2544]$ and $[0.8311, 0.9756]$ respectively.

Figure 2: Sample ACF and PACF plots for $\kappa_{t+1} - \kappa_t$



Since the estimated age components $\beta_x^{(0)}$ and $\beta_x^{(1)}$ do not change over time, the uncertainty of mortality forecast solely comes from the uncertainty in the forecast of κ_t. To simulate mortality paths for future time period $t \in [t_1 + 1, t_1 + n]$, we first simulate values of z_t for $t \in [t_1 + 1, t_1 + n]$ and then calculate the values of corresponding κ_t and $m_{x,t}$.

5 A life cycle model

5.1 The setup

We consider a Canadian male who turns 25 at the beginning of 2018 and lives in Ontario throughout his lifetime. To simplify notations, we represent the calendar year 2018 with year 1. This individual earns an annual salary of Y_t in year t which is paid upfront at the beginning of the year. Assuming that we know exactly how much this individual earns each year, there is no uncertainty in Y_t. Both he and his employer contribute 4.55% (half of the 9.9% contribution

rate in 2018) of his pensionable earnings (calculated based on his annual salary) to the CPP. His employer contributes another 5% of his annual salary into the RPP, which is a defined contribution plan. The individual does not contribute to the RPP, but he contributes a certain amount of his annual salary to his RRSP subject to the total RPP/RRSP contribution limit of $0.18Y_t$. This amount will be determined in the following subsections via the optimization of the life cycle model. For simplicity, we do not allow RRSP contribution room to be carried forward[4].

Without loss of generality, we assume the average wage and CPI remain unchanged in future years. Therefore, the dollar values in this report can be viewed as real dollar values or inflation-adjusted values. The CPP contribution made by either the employee or employer can be simplified as follows:

$$cCPP_t = 0.5 \times 9.9\% \times \max(\min(Y_t, 55900) - 3500, 0).$$

As an employee, this individual also pays employment insurance (EI) premiums. EI premiums are determined by EI rates and insurable earnings up to a maximum amount. The EI premium rate paid by the employee is $1.66 per $100 of insurable earnings, and the maximum insurable earnings are $51,700 in 2018. The employer pays 1.4 times the amount of the employee's premiums. The EI premiums paid by the employee can be expressed as:

$$EI_t = \frac{1.66 \times \max(Y_t, 51700)}{100}.$$

This individual pays tax on his earnings. The combined federal and Ontario tax brackets and tax rates including surtaxes for Ontario residents are shown in Table 1. We denote the total tax payable by Tax_t. When calculating personal income tax, we exclude from his annual earnings the CPP contribution, employee RPP and individual RRSP contribution, and EI premiums, which are all tax-deductible.

In the first 30 years of his working life, he is assumed to spend 30% of his annual earnings on housing, including rent/mortgage payments, heating and cooling, and property taxes. Given that the Canada Mortgage and Housing Corporation normally restricts the gross debt service (GDS) ratio to 35% and the total debt service (TDS) ratio to 42%[5], 30% of annual earnings on housing is a reasonable estimate. After the mortgage is paid off in 30 years, we assume that his housing expense is reduced to 10% of annual earnings. We use H_t to denote the housing expense in year t:

$$H_t = \begin{cases} 30\% \times Y_t, & \text{if } 1 \le t \le 30 \\ 10\% \times Y_t, & \text{if } t > 30 \end{cases}.$$

The remaining earnings after deducting payments for EI, the CPP, RPP/RRSPs, tax, and housing expenses can be either consumed or saved in a TFSA.

[4] If we allow RRSP contribution room to accumulate, an additional state variable is needed in the life cycle model and complicates the computation.

[5] The GDS ratio is the percentage of the gross annual income that goes to "mortgage expenses" – the principal, interest, property taxes, and heating costs, plus fees for condominium maintenance. The TDS ratio evaluates the gross annual income needed for all debt payments: housing, credit cards, personal loans, and car loans.

Table 1: Combined federal and Ontario tax brackets and tax rates including surtaxes

2018 taxable income	Other income	Capital gains	Canadian dividends	
			Eligible	Non-eligible
First $42,960	20.05%	10.03%	−6.86%	7.81%
Over $42,960 up to $46,605	24.15%	12.08%	−1.20%	12.57%
Over $46,605 up to $75,657	29.65%	14.83%	6.39%	18.95%
Over $75,657 up to $85,923	31.48%	15.74%	8.92%	21.07%
Over $85,923 up to $89,131	33.89%	16.95%	12.24%	23.87%
Over $89,131 up to $93,208	37.91%	8.95%	17.79%	28.53%
Over $93,208 up to $144,489	43.41%	21.70%	25.38%	34.91%
Over $144,489 up to $150,000	46.41%	23.20%	29.52%	38.39%
Over $150,000 up to $205,842	7.97%	23.98%	31.67%	40.20%
Over $205,842 up to $220,000	51.97%	25.98%	37.19%	44.84%
Over $220,000	53.53%	26.76%	39.34%	46.65%

The individual retires when he turns 65 and converts a portion of his RPP/RRSP to a longevity annuity which makes regular payments from age 80. To avoid the minimum RRIF withdrawal requirement and maximize the withdrawal flexibility as much as possible, the individual only converts the remaining RPP/RRSP to an RRIF when he turns 71 years old. Once the RPP/RRSP is converted to a RRIF, a minimum RRIF withdrawal requirement applies to the value of his RRIF excluding the portion converted to a longevity annuity. Besides withdrawal from the RPP/RRSP/RRIF, the individual also receives a CPP pension, OAS pension, and GIS starting from age 65. Note that both the CPP pension and OAS pension are taxable, but the GIS is not. We denote the CPP pension, OAS pension, and GIS amount in year t by $pCPP_t$, OAS_t, and GIS_t respectively.

5.2 Resources and decisions

We divide the individual's financial resources at the beginning of year t into liquid resources m_t and illiquid resources n_t. Let us denote y_t as the net income of year t. For an individual in his work life, we have

$$y_t = Y_t - cCPP_t - EI_t - Tax_t - H_t, \qquad \text{for } t \leq 40.$$

Note that annual income Y_t for year t is earned at the beginning of the year and the deduction of CPP contribution, EI premium, income tax, and housing expenses occurs at the same time as income payment. The liquid resources m_t at the beginning of year t during the working life include the TFSA balance from the previous year and the net income y_t earned at the beginning of this year. The illiquid resources n_t refer to the RPP/RRSP balance at the beginning of year t, which includes the employer's RPP contributions and the individual's RRSP

contributions from previous years.

Two decisions are made at the beginning of year t on how to spend the net income y_t: consumption c_t and RRSP contribution d_t. The remaining amount will be invested into a TFSA as liquid resources. The RRSP contribution is subject to a contribution limit. Since the employer contributes 5% of pre-tax earnings to an RPP and unused RRSP contribution room is not carried forward, the maximum contribution to an RRSP in year t is $(0.18 - 0.05)Y_t$. We also impose the condition that no RRSP can be withdrawn before retirement. Therefore, the post-decision RPP/RRSP balance at time t, b_t can be written as

$$b_t = n_t + d_t,$$

where $0 \leq d_t \leq 0.13Y_t$. The post-decision liquid resources for year t, a_t, can be expressed as

$$a_t = m_t - c_t - d_t + g(d_t),$$

where $g(d_t)$ is the tax refund arriving immediately after contributing d_t to the RRSP and $g(d_t) < d_t$. The post-decision liquid resources a_{t-1} are assumed to be invested into a TFSA and earn a tax-free investment return. For simplicity, we assume that there is no annual contribution limit on the TFSA.

For both the RPP/RRSP and TFSA, the individual follows a simple investment strategy: $(100 - \text{age})\%$ in stocks and $\text{age}\,\%$ in fixed-income securities or cash, and earns the same investment return. In practice, pension assets are invested long-term and thus generally earn a higher return than liquid assets such as a TFSA. Taking the investment return of a TFSA into account, the pre-decision liquid resources at the beginning of year $t + 1$ can be written as

$$m_{t+1} = e^{R_t} a_t + y_{t+1},$$

where R_t is the rate of investment return for both liquid and illiquid assets in year t. The pre-decision RRSP balance at the beginning of year $t + 1$ is

$$n_{t+1} = e^{R_t} b_t + 0.05Y_{t+1},$$

where $0.05Y_{t+1}$ is the 5% employer contribution on the pre-tax earnings Y_{t+1}.

When the individual turns 65 at time 40, he retires and determines the portion of his RPP/RRSP assets that should be used to purchase longevity annuity with regular payments starting at age 80. We assume that the longevity annuity is fairly priced and thus its price is equal to the expected present value of the future annuity benefits. Let the annual payment of the purchased longevity annuity be denoted by L. The price of the longevity annuity purchase is equal to ${}_{15|}\ddot{a}_{65,41} \cdot L$, where ${}_{15|}\ddot{a}_{65,41}$ is the present value of a 15-year deferred annuity of \$1 paid in advance issued to a 65-year-old male at the beginning of year 41. The annuity payment starts at the beginning of year 56 if the annuitant is still alive at the payment date.

At the beginning of year t during retirement, the individual's income includes a CPP pension, an OAS pension, a GIS, and a longevity annuity payment if applicable. Deducting the OAS recovery tax, personal income tax, and housing expenses from the income, we obtain the individual's net income y_t and

$$y_t = pCPP_t + OAS_t + GIS_t + L \cdot 1_{\{t \geq 56\}} - Tax_t - H_t, \quad \text{for } t \geq 41.$$

The amount of OAS pension and GIS included in the net income is determined based on the total income from the CPP pension and longevity annuity. The impact of RRSP/RRIF withdrawal on the OAS pension and GIS amount will be considered separately. As previously defined, the liquid assets m_t include the TFSA balance and the net income and n_t represents the balance of the RPP/RRSP/RRIF. After making decisions on consumption c_t and RPP/RRSP/RRIF withdrawal d_t, the post-decision assets at the beginning of year t are expressed as

$$a_t = m_t - c_t - d_t + g(d_t),$$
$$b_t = n_t + d_t - {}_{15|}\ddot{a}_{65,41} \cdot L \times 1_{\{t=41\}},$$

where d_t is negative and subject to the minimum RRIF withdrawal requirement after the individual turns 71, and $g(d_t)$ is also negative and represents the additional income tax that needs to be paid for the RPP/RRSP/RRIF withdrawal and the reduction of the OAS pension and GIS amount due to the RPP/RRSP/RRIF withdrawal.

The resources prior to making decisions at the beginning of year $t+1$ are then given by

$$m_{t+1} = e^{R_t} a_t + y_{t+1},$$
$$n_{t+1} = e^{R_t} b_t.$$

5.3 Maximizing expected lifetime utility

The expected lifetime utility of a 25-year-old individual at time 0 is defined as follows:

$$\sum_{t=1,\dots 75} \beta^{t-1} \mathbb{E}[u(c_t)],$$

where β is the time-preference discount factor and $u(.)$ is the utility function. We assume that the utility function has constant relative risk aversion, and hence

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where γ is the relative risk-aversion parameter.

Our objective is to find the decisions (i.e., the values of c_t and d_t) that maximize the expected lifetime utility. To formulate the optimal decision rules, we apply Bellman's Principle of Optimality (Bellman, 1957) which states:

> Principle of Optimality: An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

At the beginning of year t, for $1 \le t \le 40$, the resources m_t and n_t and the most recent mortality experience κ_{t-1} are known and set as state variables, based on which choices are made. Applying the Principle of Optimality, we obtain the Bellman equation expressed as follows:

$$V_t(m_t, n_t, \kappa_{t-1}) = \max_{c_t, d_t} u(c_t) + \beta \mathbb{E}\left[p_{25+t-1,t} V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t) | m_t, n_t, \kappa_{t-1} \right],$$

s.t.

$$a_t = m_t - c_t - d_t + g(d_t),$$

$$
\begin{aligned}
&b_t = n_t + d_t,\\
&m_{t+1} = e^{R_t} a_t + y_{t+1},\\
&n_{t+1} = e^{R_t} b_t + 0.05 Y_{t+1},\\
&y_t = Y_t - EI_t - cCPP_t - Tax_t - H_t,\\
&\kappa_t = \kappa_{t-1} + \mu + \sigma z_t,\\
&c_t > 0,\\
&0 \le d_t \le ud_t,\\
&c_t + d_t - g(d_t) \le m_t,
\end{aligned}
$$

where $V_t(m_t, n_t, \kappa_{t-1})$ is the value function representing the maximum obtainable value given the current state; $p_{25+t-1,t} = 1 - q_{25+t-1,t}$; and $ud_t = 0.13 Y_t$ is the maximum employee pension contribution. The constraint $c_t + d_t - g(d_t) \le m_t$ implies that $a_t \ge 0$ and one cannot borrow to consume. The value function $V_t(m_t, n_t, \kappa_{t-1})$ comprises the utility of the current period consumption $u(c_t)$ and the discounted continuation value $\beta \mathbb{E}\left[p_{25+t-1,t} V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t) | m_t, n_t, \kappa_{t-1}\right]$. The Bellman equation defines a recursive relation for the value function. Solving this equation requires us to start from the terminal year and work backward till the current year. Therefore, we should first formulate the Bellman equation for $41 \le t \le 75$ and then revisit the equation-solving.

At the retirement age 65 (beginning of year 41), we assume that the individual first determines the proportion of the RPP/RRSP to be converted to a longevity annuity and hence the annual longevity annuity payment L. Since future decisions of consumption and pension withdrawal depend on the payment expected from the longevity annuity purchase, L is set as a state variable for year 42 and onward. During retirement years, the retiree decides the consumption c_t and pension withdrawal d_t at the beginning of each period. The withdrawal can be from an RPP/RRSP prior to age 71 and RRIF after age 71. The Bellman equation for year $41 \le t \le 75$ can be written as:

$$
\begin{aligned}
&V_t(m_t, n_t, \kappa_{t-1}, L)\\
&=\\
&\begin{cases}
\max\limits_{c_t, d_t, L} & u(c_t) + \beta \mathbb{E}\left[p_{25+t-1,t}\, V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t, L) | m_t, n_t, \kappa_{t-1}\right], & t = 41,\\
\max\limits_{c_t, d_t} & u(c_t) + \beta \mathbb{E}\left[p_{25+t-1,t}\, V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t, L) | m_t, n_t, \kappa_{t-1}, L\right], & 42 \le t \le 74,
\end{cases}
\end{aligned}
$$

s.t.

$$
\begin{aligned}
&a_t = m_t - c_t - d_t + g(d_t),\\
&b_t = n_t + d_t - {}_{15|}\ddot{a}_{65,41} \cdot L \cdot 1_{\{t=41\}},\\
&m_{t+1} = e^{R_t} a_t + y_{t+1},\\
&n_{t+1} = e^{R_t} b_t,
\end{aligned}
$$

$$y_t = pCPP_t + OAS_t + GIS_t + L \cdot 1_{\{t \geq 56\}} - Tax_t - H_t,$$
$$k_t = k_{t-1} + \mu + \sigma z_t,$$
$$c_t \geq 0,$$
$$ld_t \leq |d_t| \leq n_t,$$
$$0 \leq c_t + d_t - g(d_t) \leq m_t$$
$$0 \leq_{15|} \ddot{a}_{65,41} \cdot L \leq 0.25\, n_{41},$$

where ld_t is the minimum withdrawal required from an RRIF and $0 \leq_{15|} \ddot{a}_{65,41} \cdot L \leq 0.25\, n_{41}$ is the 25% limit on a longevity annuity purchase.

Finally, we consider year 75, at the beginning of which the individual turns 99. Since he will die for certain before the limiting age 100 and any wealth remaining after his death does not yield any utility, he should withdraw all the assets in the RRIF and consume all the wealth he has. Therefore, the optimal choices for $t = 75$ are

$$c_t^* = m_t + n_t - g(n_t),$$
$$d_t^* = -n_t.$$

The post-decision assets are $a_{75} = 0$ and $b_{75} = 0$. Since there is no continuation value, the value function for $t = 75$ can be written as

$$V_t(m_t, n_t, \kappa_{t-1}, L) = u(c_t^*) = u(m_t + n_t - g(n_t)).$$

To solve the Bellman equation, we start from the final year of the individual's life and iterate backward to the current year. The optimal choices for year 75 have simple analytic form. However, such analytic optimal solutions do not exist for other years due to the complexity of the problem. Traditionally, the optimal choices are determined numerically using VFI (see Horneff et al., 2016), which constructs an exogenous grid for pre-decision-state variables and obtains the optimal choices for each point on the exogenous grid for years 74, 73, ..., 1 sequentially. Iterative guesses of the optimal choices are made for each point on the grid. For each guess, we evaluate the value function $V_t(m_t, n_t, \kappa_{t-1}, L)$ which involves calculating expectation with numerical integration and interpolation of next year's value function. The guesses are performed iteratively using brute force, Newton's method, or other methods. The iteration stops when the change of value function is smaller than a pre-specified tolerance value. Due to the iterative evaluations of the value function, VFI is very time-consuming.

6 A modified general endogenous grid method

6.1 Solving the optimization problem with the endogenous grid method

Carroll (2006) proposes the EGM, which avoids the iterative search in VFI and hence requires significantly less computation time. The fundamental idea of the EGM is to specify an exogenous grid over the post-decision state instead of over the pre-decision states. The value function can be rewritten as follows:

$$\tilde{V}_t(a_t, b_t, \kappa_{t-1}) = \max_{c_t, d_t} u(c_t) + \beta w_t(a_t, b_t, \kappa_{t-1}) \quad (2)$$

where $w_t(a_t, b_t, \kappa_{t-1})$ is the continuation value function as defined in Appendix B. Suppose that the continuation value function is differentiable with respect to c_t and d_t and the choices constraints are unbinding. The unique optimal choices can be found by solving the FOCs if the optimization objective function is concave and the choice set is convex.

Using the EGM approach, we first construct an exogenous grid for post-decision-state (a_t, b_t, κ_{t-1}). For each point on this grid, we evaluate the continuation value function and its derivatives and then determine the optimal choices. Since the pre-decision-state variables are endogenously determined, the grid constructed from their values are called an endogenous grid. A more detailed description of EGM can be found in Appendix B. The benefit of the EGM is that the optimal choices are found without numerical optimization, which is iterative and time-consuming. The continuation value, which is the expectation over the next-period variables, is only taken once for each exogenous grid point. In contrast, the numerical optimization used in VFI requires iterative procedures that calculate expectation in each iteration.

Druedahl and Jørgensen (2017) discuss the three challenges in generalizing the EGM to multi-dimensional models with non-convexities and constraints.

- Irregular endogenous grids: The endogenous pre-decision grid points are unevenly spaced due to the non-linearlity of Euler equations.
- Non-sufficient FOCs: When the objective function is not concave, FOCs are only necessary but not sufficient. There can be multiple points that satisfy FOCs.
- No prior knowledge on where the constraints are binding[6]: In a multi-dimensional model with multiple constraints, there is no prior knowledge about the location in the state space where constraints are binding.

Druedahl and Jørgensen (2017) propose a generalized EGM to tackle these challenges. Their method does not require sufficient FOCs or prior knowledge about whether the constraints are binding. It provides optimal choices for points on a regular exogenous grid over pre-decision states. The GEGM divides the optimization problem into segments in which it is known whether choices are constrained and the value function is evaluated. Druedahl and Jørgensen (2017) demonstrate that the GEGM is approximately 20 times faster than a highly optimized implementation of VFI with both methods designed to achieve a given level of precision. A detailed comparison of the speed and accuracy between GEGM and VFI can be found in Druedahl and Jørgensen (2017).

For a working individual at time t, for $t < 40$, we take the following procedures to determine optimal choices for the points on a regular exogenous grid over pre-decision states:

1. Construct a common regular exogenous grid $\mathcal{G}_t^{m,n,\kappa}$ over the pre-decision state (m_t, n_t, κ_{t-1}).
2. For each segment
    a. Construct a common regular grid $\mathcal{G}_t^{m,n,\kappa}$ over the post-decision state

[6] A constraint is considered to be binding if changing it also changes the optimal solution, while a constraint that does not affect the optimal solution is non-binding.

(a_t, b_t, κ_{t-1}).

b. Determine the optimal choices for each point in the common regular grid $\mathcal{G}_t^{a,b,\kappa}$ using FOCs.

c. Determine the pre-decision state corresponding to each point on the post-decision grid $\mathcal{G}_t^{a,b,\kappa}$. Construct the endogenous grid on the pre-decision state (m_t, n_t, κ_{t-1}).

d. Determine optimal choices and evaluate $V_t(m_t, n_t, \kappa_{t-1})$ for each point in the exogenous grid $\mathcal{G}_t^{m,n,\kappa}$ by interpolating the endogenous grid.

3. For each point on the common exogenous grid $\mathcal{G}_t^{m,n,\kappa}$, the optimal choice is selected from the segment with the highest value of choice.

To solve the Bellman equation for a retiree at time t, for $t < 40$, we include the longevity annuity payment L as a state variable. We construct a common regular exogenous grid $\mathcal{G}_t^{m,n,\kappa,L}$ over the pre-decision state $(m_t, n_t, \kappa_{t-1}, L)$ and a common regular exogenous grid $\mathcal{G}_t^{a,b,\kappa,L}$ over the post-decision state $(a_t, b_t, \kappa_{t-1}, L)$. The remaining steps are the same as those described above for the working individual. The Bellman equation is solved backwards from age 99 to 98, 97, …, 25 such that the lifetime utility is maximized.

6.2 Segments

Evaluation by segment is the key step in GEGM to address the challenge of no prior knowledge on where the constraints are binding. The segmentation should ensure that the binding constrains can be identified in each segment.

For a working individual, there are two choices: consumption and RRSP contribution. When c_t is unconstrained, the optimal consumption is always positive since zero consumption leads to infinitely negative utility. Therefore, $c_t > 0$ is unbinding. Assuming that the individual cannot borrow to consume, consumption should be a value such that $a_t \geq 0$. For the RRSP contribution, we assume that it is always non-negative and it has to be smaller than $ud_t = 0.13Y_t$.

The GEGM algorithm proposed by Druedahl and Jørgensen (2017) assumes that the value function is differentiable w.r.t. c_t and d_t. If this assumption holds in our setup, we should consider six segments:

1. $a_t = 0$ and $d_t = 0$ (both c_t and d_t are constrained).
2. $a_t = 0$ and $d_t = ud_t$ (both c_t and d_t are constrained).
3. $a_t = 0$ and $d_t \in (0, ud_t)$ (c_t is constrained, but d_t is not).
4. $a_t > 0$ and $d_t = 0$ (d_t is constrained, but c_t is not).
5. $a_t > 0$ and $d_t = ud_t$ (d_t is constrained, but c_t is not).
6. $a_t > 0$ and $d_t \in (0, ud_t)$ (neither choice is constrained).

For a retiree, $d_t < 0$ for pension withdrawal and d_t is subject to the minimum withdrawal requirement. The interval of d_t is thus divided into $d_t = -b_t$, $d_t = ud_t$, and $d_t \in (-b_t, ud_t)$ where $-ud_t$ is the minimum withdrawal amount.

6.3 Accommodating non-differentiable value function

Personal income tax is a piecewise linear function of income due to income tax thresholds presented in Table 1. For a working individual, the RRSP contribution is tax-deductible. For a retired individual, an RRSP withdrawal is treated as taxable income. In addition, the OAS pension and GIS amount are reduced when other income received by a retired individual exceeds certain threshold values. These threshold values lead to a number of kinks in $g(d_t)$. As a result, $g(d_t)$ is also a piecewise linear function of d_t and the value function is not always differentiable w.r.t. d_t.

To address the differentiability problem, we further divide the domain of d_t into smaller intervals. Suppose that there are $n+1$ income threshold values, $\{h_0, h_1, \dots, h_n\}$, where $h_0 = 0$. The tax rate that applies to the income bracket (h_j, h_{j+1}) is rt_{j+1}. The tax rate of rt_{n+1} applies to the income bracket of (h_n, ∞). It is obvious that $g(d_t)$ is not differentiable if $Y_t - d_t$ falls at one of the threshold values.

Assume that $Y_t \in (h_i, h_{i+1})$. Let $\hat{n} = \min_j$ s.t. $Y_t - h_j \leq ud_t$. We divide the domain of d_t into the following segments:

- Segment d.1: $(0, Y_t - h_i)$
- Segment d.2: $(Y_t - h_i, Y_t - h_{i-1})$
- Segment d.3: $(Y_t - h_{i-1}, Y_t - h_{i-2})$
- ...
- Segment d.$(i - \hat{n} + 1)$: $(Y_t - h_{\hat{n}+1}, Y_t - h_{\hat{n}})$
- Segment d.$(i - \hat{n} + 2)$: $(Y_t - h_{\hat{n}}, ud_t)$
- Segment d.$(i - \hat{n} + 3)$: 0
- Segment d.$(i - \hat{n} + 4)$: $Y_t - h_i$
- Segment d.$(i - \hat{n} + 5)$: $Y_t - h_{i-1}$
- ...
- Segment d.$(2(i - \hat{n} + 2))$: $Y_t - h_{\hat{n}}$
- Segment d.$(2(i - \hat{n} + 2) + 1)$: ud_t

$g(d_t)$ is a linear and differentiable function of d_t on each of the segments d.1 to d.$(i - \hat{n} + 2)$, and $g_d(d_t) = rt_{i+1}, rt_i, \dots, rt_{\hat{n}}$ respectively. However, $g(d_t)$ is not differentiable at $d_t = 0, Y_t - h_i, \dots, Y_t - h_{\hat{n}}, ud_t$.

For each segment of d_t, we consider both constrained c_t ($a_t = 0$) or non-constrained c_t ($a_t > 0$). Therefore, there are $4(i - \hat{n}) + 6$ segments for c_t and d_t in total. We will evaluate the value function and make choices for each segment. The choices leading to the highest value function among all segments are the optimal solution. More details about the value function evaluation by segment are presented in Appendix C. For a retiree, a similar approach is taken to allow non-differentiable value function.

6.4 Simulating scenarios

With the optimal choices and value function determined for all possible points on the common regular grid $\mathcal{G}_t^{m,n,\kappa,L}$ or $\mathcal{G}_t^{m,n,\kappa}$, we can use simulation and interpolation to study the pattern of future consumption and RRSP deposit/withdrawal by following the procedures below:

1. Let $t = 1$. We have $m_t = y_0$ and $n_t = 0$. κ_{t-1} is known.
2. Determine the optimal choices c_t and d_t using linear interpolation over the optimal choices obtained for the common regular grid $\mathcal{G}_t^{m,n,\kappa}$. Simulate investment return and mortality rate for year t and calculate m_{t+1}, n_{t+1}, and κ_t based on the simulated values.
3. Repeat Step 3 for $t = 2,3, \ldots, 40$.
4. For $t = 41$, determine the optimal choices c_t, d_t, and L using linear interpolation over the optimal choices obtained for the common regular grid $\mathcal{G}_t^{m,n,\kappa}$. Simulate investment return and mortality rate for year t and calculate m_{t+1}, n_{t+1}, and κ_t based on the simulated values.
5. For $t = 41,42, \ldots 74$, determine the optimal choices c_t and d_t, and L using linear interpolation over the optimal choices obtained for the common regular grid $\mathcal{G}_t^{m,n,\kappa,L}$. Simulate investment return and mortality rate for year t and calculate m_{t+1}, n_{t+1}, and κ_t based on the simulated values.
6. For $t = 75$, optimal choices are $d_t = -n_t$ and $c_t = m_t - d_t + g(d_t)$.

7 Results based on Canadian male mortality data

7.1 Simulated lifetime utility and behaviours

The life cycle setup assumes that the individual turns age 25 at the beginning of 2018 and the value of κ_t for year 2017 ($t = 0$) is known at the outset. However, the HMD Canadian mortality data only range up to year 2016. To fill the gap, we use the mean forecast of κ_t for year 2017 as a known value.

We assume that the annual effective risk-free interest rate is 1% and the annual return of the risky assets follows log-normal distribution $LN(0.04,0.2)$. The time-preference parameter β is set to 0.97. The risk-aversion parameter γ is set to 5. We use 10-point Gaussian quadrature to approximate the expectation when calculating the continuation value. The number of points for the common grid is 10 for each dimension.

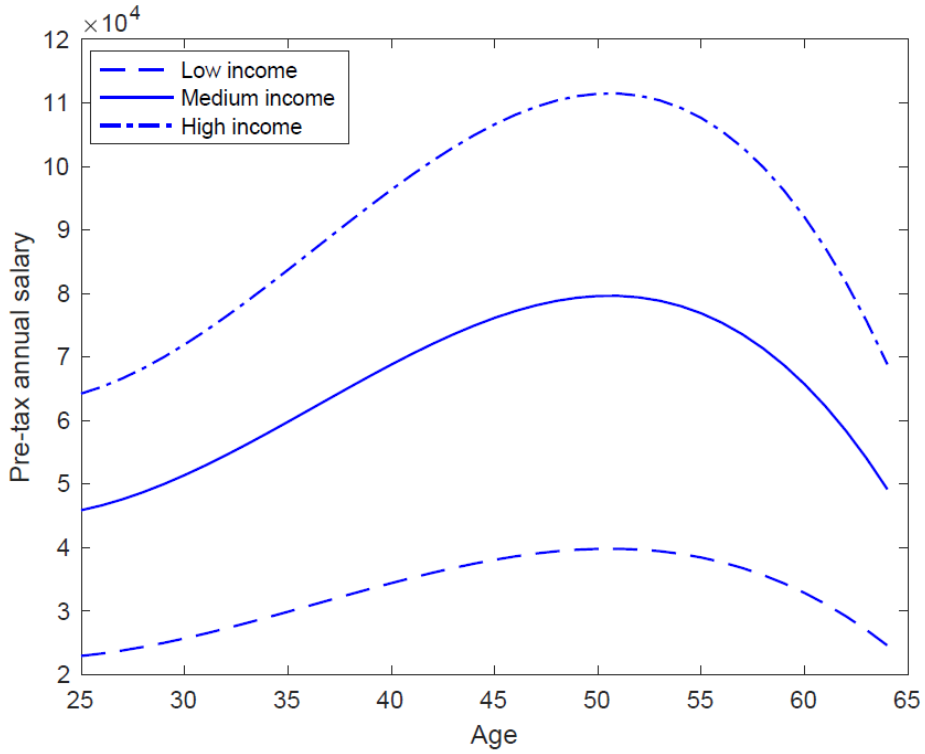
The individual earns the Canadian average salary for his age. His pre-tax salary Y_t in year t at age $25 + t - 1$ follows a cubic function. The cubic-function parameters are estimated by matching the average wage, salaries, and commissions by age group for Canadian males in 2017, which are shown in Table 2. By minimizing the squared errors, we obtain the following estimated salary function:

$$Y_{25+t} = 140{,}600 - 10{,}128(25 + t) + 329(25 + t)^2 - 3.017(25 + t)^3. \quad (3)$$

Table 2: Average wage, salaries, and commissions for Canadian males by age groups in 2017

Age group	Wages, salaries, and commissions
-34	51,400
-44	68,800
-54	79,600
-64	65,700

Figure 3: Estimated average salary by age



We consider individuals with three different salary levels for comparison purposes:

- Medium income: the individual has the salary pattern expressed by Equation (5) and shown as the solid line in Figure 3.
- Low income: the individual earns half of the medium income at all ages with the salary pattern shown as the dashed line in Figure 3.
- High income: the individual earns 40% more than the medium income at all ages with the salary pattern shown as the dash-dotted line in Figure 3.

We also compare the cases with and without access to a longevity annuity. Therefore, a total of six cases are compared. For each case, we construct state-variable grids and determine optimal choices for all the points on the grids. We then simulate mortality and investment scenarios, calculate the optimal choices under the simulated scenarios, and compute the lifetime utility of

the individual.

Table 3 shows the average simulated lifetime utility for the six cases. The simulated lifetime utility of the individual with a medium income and access to a longevity annuity is averaged at $-1.4835e{-}17$, which is significantly higher than $-1.7282e{-}17$, the lifetime utility for the individual with a medium income but no access to a longevity annuity. Similarly, there is significant utility gain for the individual with a high income when access to a longevity annuity is allowed. However, the utility gain for the individual with a low income is marginal. Therefore, the individual with a low income does not benefit from access to a longevity annuity. We also observe that both medium- and high-income individuals convert an average of 25% of RRSP assets to a longevity annuity, which is the upper limit we impose on the purchase of a longevity annuity. The low-income individual converts an average of a mere 2% of RRSP assets to a longevity annuity, thereby indicating that the benefit of a longevity annuity is minimal and demand is low among low-income individuals.

Table 3: Maximum lifetime utility with and without longevity annuity access

Salary level	Low	Medium	High
Optimal conversion rate	1.89%	24.99%	25%
Utility with access ($\times 10^{-18}$)	-195.37	-14.835	-4.829
Utility w/o access ($\times 10^{-18}$)	-195.85	-17.282	-5.765

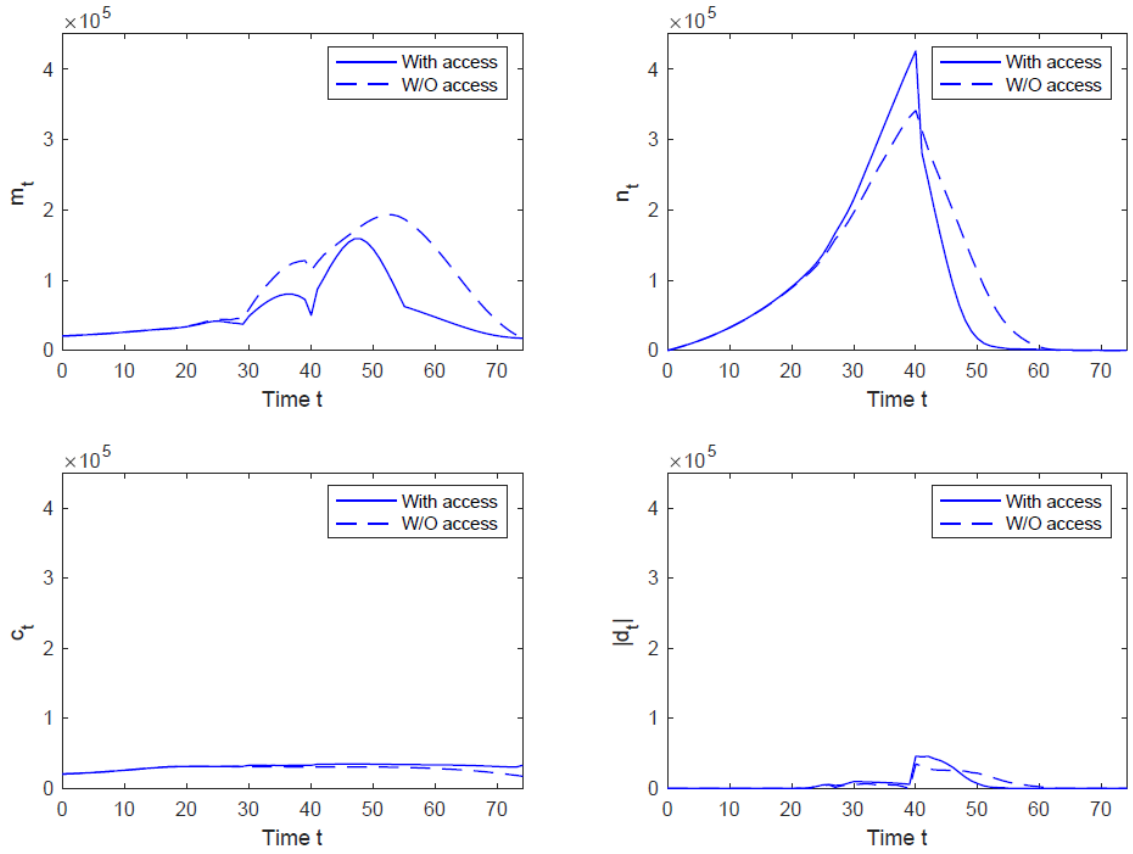
Low income individuals do not purchase or benefit from longevity annuity because longevity annuity payment may significantly reduce GIS amount. The maximum annual income to receive GIS pension is $17,880 for a single pensioner. The total of CPP pension and minimum withdrawal from a registered account for a low income individual is much lower than $17,880. Per dollar of longevity annuity payment reduces GIS amount by approximately 59 cents until the GIS amount reaches zero. The loss of GIS amount outweighs the benefit of longevity annuity.

Figure 4 shows the average values of simulated state variables and optimal choices for the individual with a medium income and with/without access to a longevity annuity. The dashed (solid) lines correspond to the case without (with) access to a longevity annuity. The right-bottom panel of Figure 4 displays the absolute value of d_t. We notice that the dashed and solid lines start to deviate from each other from time 23 or age 48. With longevity annuity access, the individual chooses to contribute more into his RRSP. Therefore, sim_n becomes higher than sim_n^{wo} after age 48. At age 65, sim_n drops below sim_n^{wo} because a portion of the RRSP is converted to a longevity annuity. We also observe changes in both the dashed and solid lines around age 55. At age 55, the mortgage is paid off and thus more liquid resources are available for consumption and RRSP contributions. The individual with no access to a longevity annuity deposits more into his TFSA while the one with access saves more in his RRSP.

Prior to retirement, there is no obvious difference in consumption between the two cases. After retirement, we observe that the individual with longevity annuity access always consumes more than the one without access. The individual with access to a longevity annuity withdraws

more of his pension fund than the one without access between age 65 and 70, but significantly less after age 70. After age 70, he consumes more from his TFSA. At age 80, his RRSP has been depleted and his consumptions at higher ages are supported by his longevity annuity, savings, CPP pension, OAS pension, and GIS.

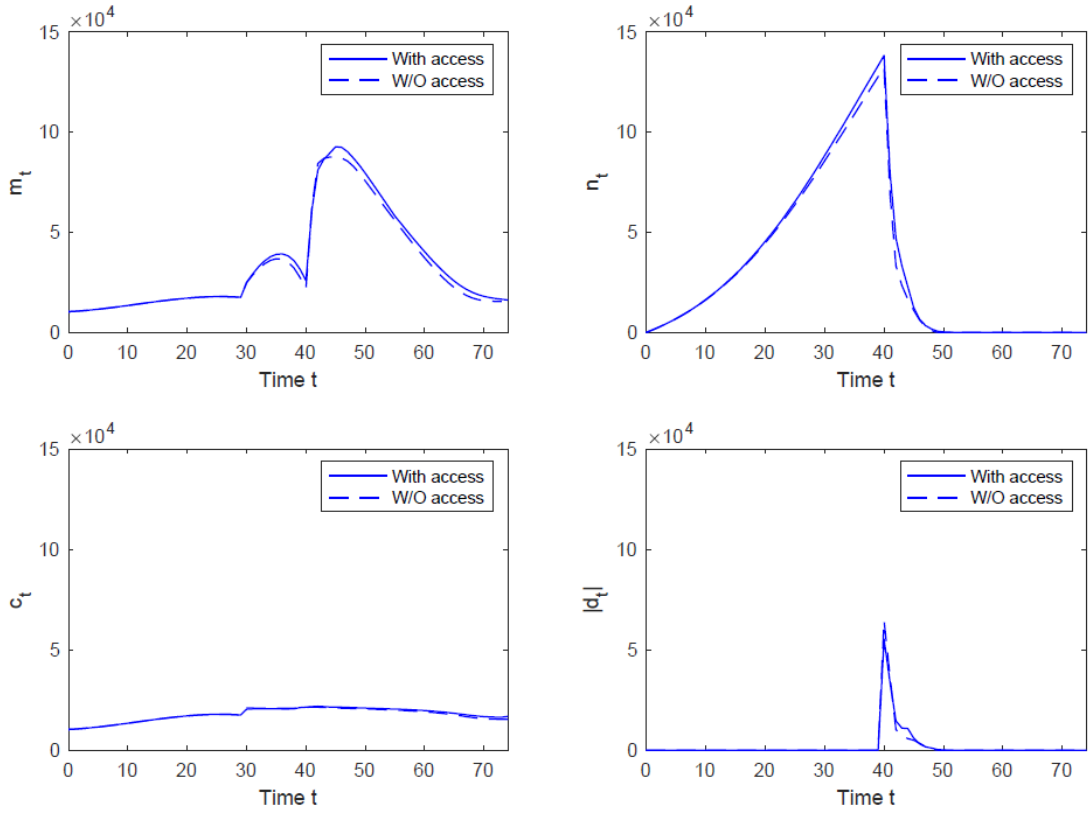
Figure 4: Comparing the simulated scenarios for the medium-income individual with and without access to a longevity annuity



In Figure 5, we plot the simulation results for the low-income individual with and without access to longevity annuity. We observe that the difference due to the longevity annuity provision is negligible, thereby verifying our conclusion that a longevity annuity does not significantly benefit low-income individuals. We also notice that the individual withdraws a large amount of his RRSP/RPP in the first few years of retirement. This is due to the fact that the GIS amount is reduced by approximately 59 cents per extra dollar of annual income until the GIS amount hits zero. The maximum annual income to receive the GIS pension is $17,880 for a single pensioner. The individual is better off withdrawing his RRSP/RPP in large lump sums such that he only loses his GIS in the years that he makes large withdrawals. Low-income individuals have on average approximately $136,600 in their RRSP/RPP at age 65. Let us consider two withdrawal strategies. The first strategy is withdrawing $17,075 every year for eight years. Assuming that the RRSP/RPP withdrawal is the only income besides the OAS pension, he will receive $478 of GIS per year for these eight years. The second strategy is withdrawing $68,300 per year in the first two years and thus depleting his RRSP/RPP account in

two years. He receives no GIS for the first two years but the full amount of $10,567 per year for the remaining six years. Although more personal income tax needs to be paid when a large lump sum of RRSP/RPP is withdrawn using the second strategy, the benefit of extra GIS received outweighs the negative tax impact.

Figure 5: Comparing the simulated scenarios for the low-income individual with and without access to a longevity annuity



The analysis above shows that the provision of a longevity annuity significantly improves the lifetime utility of medium- and high-income individuals. To quantify the improvement in dollar values, we analyse how much cash infusion the individual without access to a longevity annuity needs at age 25 to achieve the same lifetime utility as the individual with access. The individual with a medium income and no longevity annuity access needs an $11,880 cash infusion at age 25 in order to match the utility of the medium-income individual with access. However, the low-income individual without access only needs a $33 cash infusion at age 25 in order to match the utility of the low-income individual with access.

In Table 4, we present the optimal conversion rates at various income levels. We observe that the optimal conversion rate only starts to decrease when an individual earns an income lower than 70% of medium income.

Table 4: Optimal conversion rate for individual at different income levels, assuming zero premium loading

Income (% of medium income)	50	55	60	70	80	90	100
Optimal conversion rate	0%	19.50%	21.07%	24.24%	24.86%	24.97%	25.00%

7.2 Sensitivity tests

7.2.1 Numbers of grid points in GEGM

The numbers of grid points for $m_t,\ n_t,\ \kappa_t,\ L,\ a_t$ and b_t used in the exogenous regular grids are $n_m = 20,\ n_n = 20,\ n_\kappa = 10,\ n_L = 10,\ n_a = 40$, and $n_b = 40$ respectively. We examine how the optimal conversion rate and lifetime utility change with the number of grid points in Table 5. While increasing or decreasing the numbers of grid points has no impact on the simulated average optimal conversion rate, the maximized lifetime utility changes by a small amount.

Table 5: Optimal conversion rates and lifetime utilities for a medium income individual using different grid sizes in GEGM

$(n_m, n_n, n_\kappa, n_L, n_a, n_b)$	Utility ($\times 10^{-18}$)	Optimal conversion rate
(20,20,20,20,40,40)	−14.835	25.00%
(30,30,10,10,90,90)	−14.894	25.00%
(10,10,10,10,10,10)	−15.352	25.00%

7.2.2 Premium loading

We expect that the demand or the optimal conversion rate decreases with the price of longevity annuity. In Table 6, we consider four loading levels for a medium income individual. When 20% loading is applied, the longevity annuity with \$1 regular payment is sold for $\$(1.2_{15|}\ddot{a}_{65,41})$ and the optimal conversion rate remains at 25%. Due to the limit of 25% conversion rate, the individual always purchases the maximum amount allowed under the case of 0%-25% loading. When the loading is further increased to 50% and 100%, we observe that the conversion rate reduces. Given that the actual loadings are usually lower than 20%, we should expect the medium income individual to convert 25% of their RRSP/RPP to longevity annuity.

Table 6: Optimal conversion rate for medium income individuals under different levels of longevity annuity premium loading

Premium loading	0%	20%	50%	100%
Optimal conversion rate	25.00%	24.99%	24.75%	13.95%

7.2.3 Conversion limit

In the above analysis, we set the maximum conversion rate to 25%. For medium and high-income individuals, the optimal conversion rate hits the 25% limit in most of the simulated scenarios. In Table 7, we increase the conversion limit to 35%, 50%, and 75%. We observe that the optimal conversion rate increases with the conversion limit. When the limit is set to 35%, the average simulated optimal conversion rate is 34.95%, thereby indicating that the majority of simulated optimal conversion rates hits the upper limit. When the conversion limit increases from 50% to 75%, the change in the conversion rate is marginal, thereby indicating that most of the simulated optimal conversion rates lie below 50%.

Table 7: Optimal conversion rates under different conversion limits for a medium income individual

Conversion limit	25%	35%	50%	75%
Optimal conversion rate	25.00%	34.95%	39.25%	39.27%

7.2.4 Mortality level

Longevity annuity is priced under the assumption of population mortality curve. However, individuals have more information about their own health and thus use their perceived future mortality when making decisions. We revisit the life cycle model using perceived mortality for ages 65 to 99. The perceived mortality rate at age x in year t, denoted by $\widehat{m}_{x,t}$, is assumed to be a multiple of $m_{x,t}$ and is expressed as follows:

$$\begin{aligned}\widehat{m}_{x,t} &= \psi \cdot m_{x,t} \\ &= \psi \cdot e^{\beta_x^{(0)} + \beta_x^{(1)} \kappa_t} \\ &= \psi \cdot e^{\ln c + \beta_x^{(0)} + \beta_x^{(1)} \kappa_t} \\ &= \psi \cdot e^{\widehat{\beta}_x^{(0)} + \beta_x^{(1)} \kappa_t},\end{aligned}$$

where $\hat{\beta}_x^{(0)} = \ln c + \beta_x^{(0)}$. The life cycle model setup remains unchanged except that we replace $\beta_x^{(0)}$ by $\hat{\beta}_x^{(0)}$.

Table 8 presents the optimal conversion rates using perceived mortalities. Under the three scenarios of perceived mortality and 25% conversion limit, the average simulated conversion rate is always 25% for medium income individuals. When the conversion limit is increased to 50%, we observe that individuals with lower perceived mortality purchase significantly more longevity annuity.

Table 8: Optimal conversion rates using perceived mortalities for ages 65–99

ψ		0.5	1	1.5
Optimal conversion rate	25% limit	25.00%	25.00%	25.00%
	50% limit	44.19%	39.25%	35.05%

7.2.5 Risk aversion parameter

Table 9 presents the optimal conversion rates using different risk aversion parameters. When the conversion limit is set to 25%, relative risk aversion does not change in the average simulated optimal conversion rate for medium income individuals. When the conversion limit is set to 50%, we observe that less risk averse individuals purchase more longevity annuity.

Table 9: Optimal conversion rates under various assumptions of relative risk aversion for medium income individuals

γ		3	5	7
Optimal conversion rate	(25% limit)	25.00%	25.00%	25%
	(50% limit)	43.71%	39.25%	36.95%

8 Results based on CPP mortality data

8.1 CPP mortality data

We obtained the CPP mortality data by pension groups from the Office of the Superintendent of Financial Institutions Canada. The mortality data include exposure and death counts for males and females at the age range of 60–120 in years 1967–2015. The pension groups used for segregating the data are as follows:

1. 0% to 9% of the maximum pension.
2. 10% to 19% of the maximum pension.
3. 20% to 29% of the maximum pension.
4. 30% to 39% of the maximum pension.
5. 40% to 49% of the maximum pension.
6. 50% to 59% of the maximum pension.
7. 60% to 69% of the maximum pension.
8. 70% to 79% of the maximum pension.
9. 80% to 89% of the maximum pension.
10. 90% to 99% of the maximum pension.

11. Greater than or equal to 100% of the maximum pension[7].

8.2 Model and estimation

In Section 6, we consider individuals with different pre-retirement income levels. The CPP pension received by the low pre-retirement-income individual is approximately 59% of the maximum CPP pension. Therefore, the low-income individual belongs to the pension group 6. The CPP pension received by the medium-income individual is approximately 98% of the maximum pension, thereby falling into the pension group 10. The CPP pension received by the high-income individual is the maximum pension, thereby belonging to the pension group 11. In the following, we model the mortality of pensioners in pension groups 6, 10, and 11.

We consider the augmented common factor (ACF) model proposed by Li and Lee (2005). The model is structured as follows:

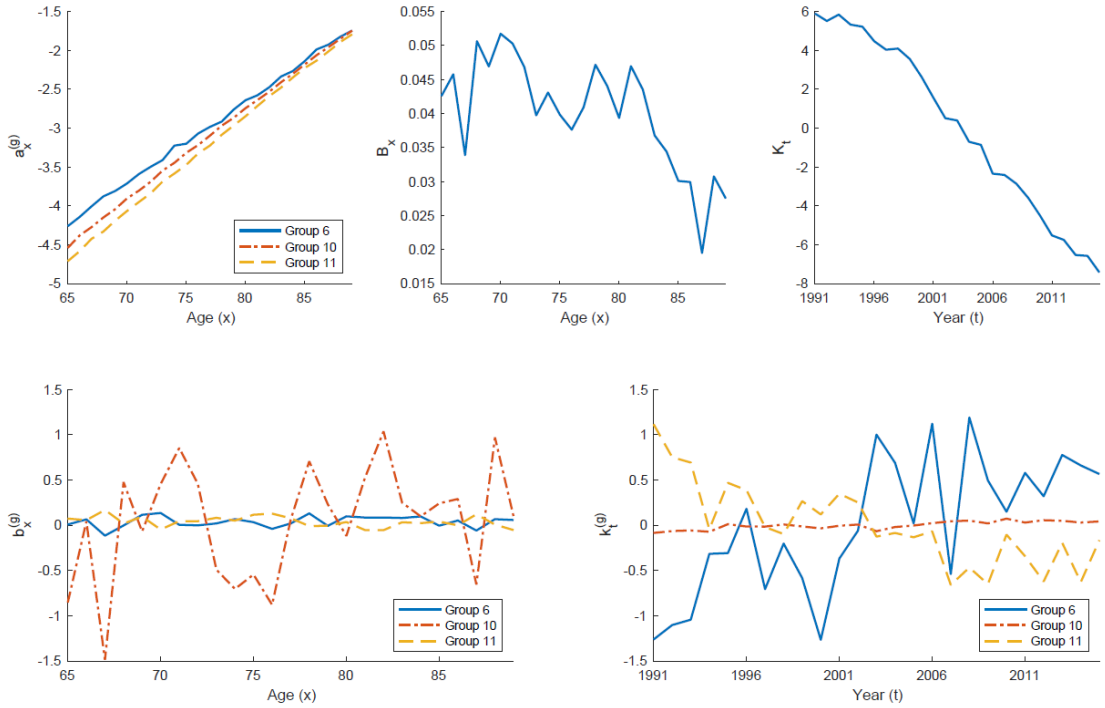
$$\ln(m_{x,t}^{(g)}) = a_x^{(g)} + B_x K_t + b_x^{(g)} k_t^{(g)},$$

where $m_{x,t}^{(g)}$ is pension group g's central death rate at age x and in year t, $a_x^{(g)}$ is the age-specific parameter indicating pension group g's average mortality at age x, K_t is the time-varying index that affects all pension groups to be modelled, $k_t^{(g)}$ is the time-varying index that is specific to pension group g, and B_x and $b_x^{(g)}$ respectively are age-specific parameters reflecting the sensitivity of $\ln(m_{x,t}^{(g)})$ to K_t and $k_t^{(g)}$.

The ACF model is estimated using the MLE method. Figure 6 shows the estimates of $a_x^{(g)}$, B_x, K_t, $b_x^{(g)}$ and $k_t^{(g)}$ fitted to the CPP male data in pension groups 6, 10, and 11 using the sample period of 1991–2015 and age range of 65–89.

[7] There are two ways for a pensioner to have more than 100% of the maximum pension: (1) convert pre-existing disability benefits into the CPP at retirement, and (2) convert a pre-existing survivor pension into the CPP at retirement.

Figure 6: Estimates of $a_x^{(g)}$, B_x, K_t, $b_x^{(g)}$, and $k_t^{(g)}$ fitted to the CPP male data



8.3 Prediction

Following Li and Lee (2005), we assume a random walk with drift for K_t:

$$K_t = \mu + K_{t-1} + \sigma\omega_t,$$

where μ is the drift term and $\{\omega_t\}$ is a sequence of i.i.d. standard normal random variables. For each $k_t^{(g)}$, we assume a first-order autoregressive process:

$$k_t^{(g)} = \phi_0^{(g)} + \phi_1^{(g)} k_{t-1}^{(g)} + \sigma^{(g)}\zeta_t^{(g)},$$

where $\phi_0^{(g)}$, $\phi_1^{(g)}$, and $\sigma^{(g)}$ are constants, and $\{\zeta_t^{(g)}\}$ is a sequence of i.i.d. standard normal random variables. We further assume that $\zeta_t^{(g)}$ and ω_t are independent with each other and $\zeta_t^{(i)}$ and $\zeta_t^{(j)}$ for $i \neq j$ are also independent. Table 6 shows the estimates of $\phi_0^{(g)}$, $\phi_1^{(g)}$, σ, μ, and $\sigma^{(g)}$. Figure 7 shows the prediction intervals of K_t, $k_t^{(6)}$, $k_t^{(10)}$, $k_t^{(11)}$ for $t = 2016, \ldots, 2040$.

Table 6: Estimates of $\boldsymbol{\phi_0^{(g)}}$, $\boldsymbol{\phi_1^{(g)}}$, $\boldsymbol{\sigma^{(g)}}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ fitted to the CPP male data

Pension group	6	10	11
$\phi_0^{(g)}$	0.0073	0.0019	-0.0148
$\phi_1^{(g)}$	0.5408	0.7553	0.7554
$\sigma^{(g)}$	0.4003	0.0009	0.0974
μ	-0.4755		
σ	0.2306		

Let $q_{x,t}^{(g)} = 1 - \exp(-m_{x,t}^{(g)})$ represent the probability that an individual from pension group g who turns age x at the beginning of year t dies in year t. The T-year survival probability for this individual can be written as

$$ {}_TS_{x,t}^{(g)} = \prod_{j=0}^{T-1} (1 - q_{x+j,t+j}^{(g)}). $$

Figure 8 shows the prediction intervals of ${}_TS_{65,2016}^{(6)}$, ${}_TS_{65,2016}^{(10)}$, ${}_TS_{65,2016}^{(11)}$ for $T = 1, \ldots, 25$.

Figure 7: Prediction intervals of K_t, $k_t^{(6)}$, $k_t^{(10)}$, $k_t^{(11)}$

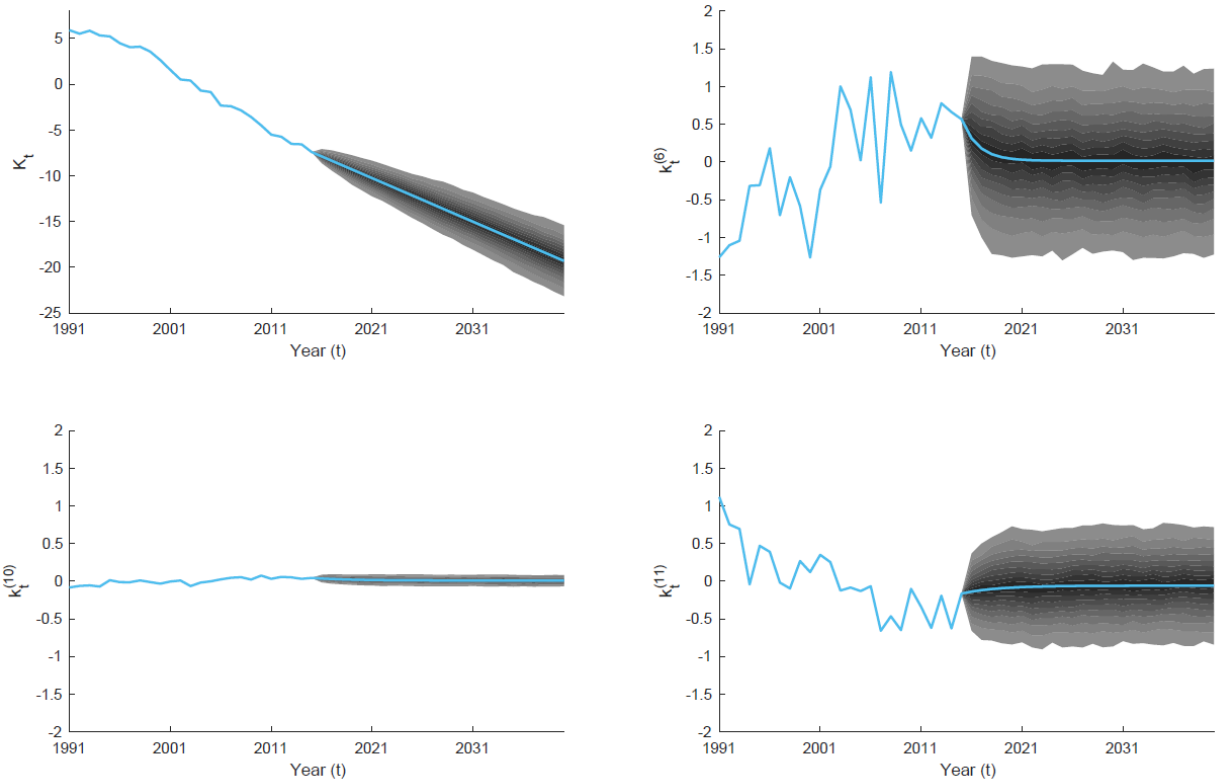
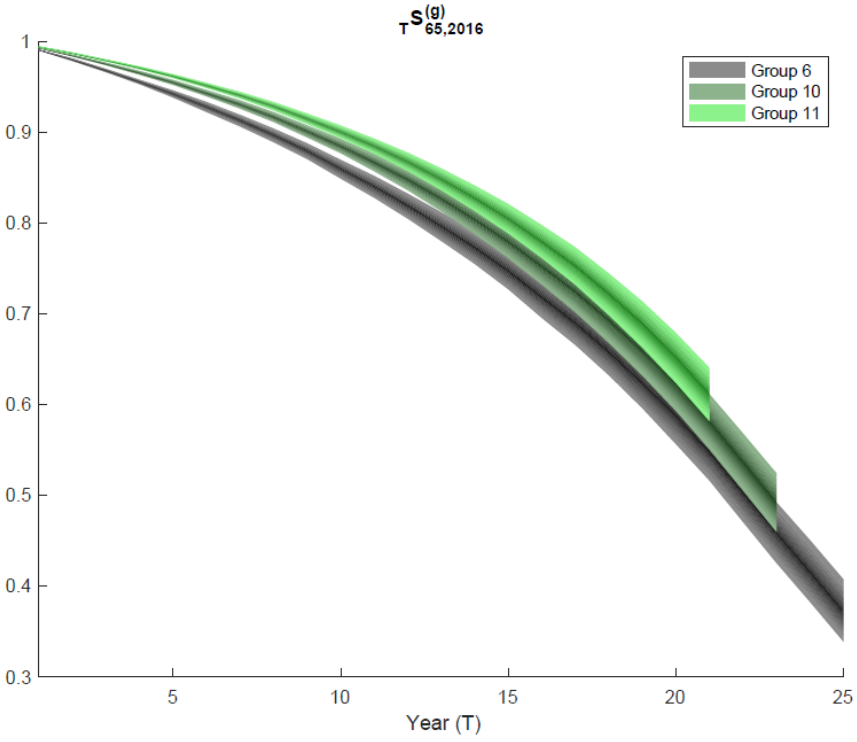


Figure 8: Prediction intervals of ${}_TS_{65,2016}^{(6)}$, ${}_TS_{65,2016}^{(10)}$, ${}_TS_{65,2016}^{(11)}$ for $T = 1, \dots, 25$



8.4 Incorporating the impact of income class on mortality

In Section 7, we analyse the utility gain of longevity annuity access for individuals with different pre-retirement income levels but the same mortality experience. However, the study of CPP mortality data indicates that lower mortality is associated with higher income levels. Therefore, it is necessary to incorporate the impact of income level on mortality in the life cycle model.

Since the multi-population model described in the previous section includes two time-varying indexes, one common index for all pension groups and one individual index for a specific pension group, both indexes should be treated as state variables in the life cycle model. To minimize the change in the model formulation, we redefine κ_t in the life cycle model by $\kappa_t = [K_t, k_t^{(g)}]$. The optimization problem remains unchanged except that the constraints for κ_t are replaced by the follows:

$$\kappa_t = [K_t, k_t^{(g)}],$$
$$K_t = \mu + K_{t-1} + \sigma\omega_t,$$
$$k_t^{(g)} = \phi_0^{(g)} + \phi_1^{(g)} k_{t-1}^{(g)} + \sigma^{(g)}\zeta_t^{(g)}.$$

The CPP mortality data range from ages 60 to 120. Since there are no data available for younger ages, we begin the analysis from the retirement age 65. Assume that an individual turns 65 at the beginning of 2018. Let us denote the individual's liquid assets and RRSP at the beginning of 2018 as m_0 and n_0 respectively. For each income level, we set m_0 and n_0 to the average simulated liquid asset and RRSP of an individual with the same income level and without longevity annuity access at age 65 from Section 7. We evaluate the individual's lifetime utility (from ages 65 to 100) under five scenarios:

1. Low income and experiencing Group 6 mortality.
2. Low income and experiencing Group 10 mortality.
3. Medium income and experiencing Group 10 mortality.
4. High income and experiencing Group 10 mortality.
5. High income and experiencing Group 11 mortality.

Since Group 10 mortality is close to the average mortality of all groups, we can view scenarios 2 to 4 as the cases that do not consider the mortality difference among different income levels. The comparisons between scenarios 1 and 2 and between scenarios 4 and 5 demonstrate how the incorporation of the relation between mortality and income level affects our results.

The results of the five scenarios are shown in Table 7. The low-income individual is better off without a longevity annuity purchase in both scenarios 1 and 2 while he achieves slightly lower utility in scenario 2 than in scenario 1. We observed that Group 10 mortality is lighter than Group 6 mortality in the previous section. As a result, the individual lives longer and the longevity annuity price is higher in scenario 2 than in scenario 1. Since the low-income individual does not purchase any longevity annuity, the decrease of utility in scenario 2 is due

to the fact that he has to live on the same amount of initial wealth for a longer period. In scenario 3, the medium-income individual still purchases the maximum amount of longevity annuity under his group-specific mortality. The access to a longevity annuity improves his utility significantly. In scenarios 4 and 5, the high-income individual also purchases the maximum amount of longevity annuity. When the high-income individual is assumed to experience Group 10 mortality (scenario 4) instead of Group 11 mortality (scenario 5), he pays a lower price for a longevity annuity and his expected life expectancy is shorter. Therefore, a higher lifetime utility can be achieved in scenario 4.

Based on the scenarios considered above, ignoring the mortality difference among income levels leads to utility underestimation for low-income individuals and overestimation for high-income individuals. The incorporation of mortality difference among income levels does not appear to affect the percentage converted to a longevity annuity. This observation may be due to the coarse categorization of income levels used in the comparison.

Table 7: The impact of longevity annuity access for retirees in different pension groups

Scenario	#1	#2	#3	#4	#5
Income	Low		Medium	High	
Pension group	6		10	11	
m_0	24,819		118,040	172,599	
n_0	137,100		345,570	628,516	
Mortality group	6	10	10	10	11
Conversion rate	0	0	25%	25%	25%
Utility with access ($\times 10^{-18}$)	-24.752	-27.116	-3.291	-0.892	-1.000
Utility w/o access ($\times 10^{-18}$)	-24.752	-27.116	-6.043	-2.085	-2.196

9 Conclusion

In this report, we investigate the benefit of longevity annuity access for Canadians with different income levels. We develop a life cycle model to determine the optimal consumption, pension contribution/withdrawal, and conversion to a longevity annuity. Our life cycle model, built on realistic tax rules and the Canadian retirement income system, also incorporates stochastic mortality rates and investment returns. To optimize the lifetime utility, we propose a modified GEGM which can handle the non-differentiable problem caused by realistic tax rules and the Canadian retirement income system.

Assuming the same mortality experience, we find that individuals with medium and high incomes greatly benefit from the access to a longevity annuity while low-income individuals benefit only marginally. A longevity annuity provides significant tax deferral for wealthy individuals with sufficient retirement income from other sources, because annuity payments begin at high ages and longevity annuity purchase reduces the minimum withdrawal from

registered accounts required for age 71 and above. Low-income individuals do not benefit as much since they have insufficient retirement income from other sources and are better off withdrawing more than the minimum requirement from registered accounts.

We further consider the impact of income class on mortality experience using the CPP mortality data. We observe that individuals receiving a higher CPP pension generally experience lower mortality. Incorporating the group-specific mortalities, the optimal longevity annuity purchase does not change for the three CPP pension groups considered. However, we do observe small changes in the maximum utility achievable. The utility for high-income individuals decreases with the use of group-specific mortalities since a higher price is paid for longevity annuity and the individuals live on the same amount of wealth for a longer expected lifetime. The utility for low-income individuals increases with the use of group-specific mortalities since the individuals live on the same amount of wealth for a shorter expected lifetime.

Low income individuals do not purchase longevity annuity because longevity annuity payment may significantly reduce GIS amount. To make longevity annuity beneficial for low income individuals, we may consider changing the calculation rule of GIS amount or modifying the payment pattern of longevity annuity. Some variations of longevity annuity have been proposed, including the variable investment-linked deferred annuity (Maurer et al., 2013) and ruin-contingent life annuity (Huang et al., 2009, 2014). In the future research, experimenting various GIS calculation rules and variations of longevity annuity is warranted with the aim of greater benefit for the low income class.

The retirement age is fixed at 65 in this study. A possible direction of future research is to consider various retirement ages and examine how longevity annuity purchase changes with retirement age. We may also include retirement age as a decision variable and study the impact of longevity annuity access on the choice of retirement age.

Acknowledgement

This research was funded by the Canadian Institute of Actuaries through the Academic Research Grant Program.

Appendix A Calculation of CPP pension payment

Denote year t's maximum pensionable earnings by YMPE $_t$. We show an example of CPP pension amount calculation for an individual who begins to receive a CPP pension when he turns 65 on January 1, 2018. We assume that the maximum general drop-out period is applied for this individual. Denote $\mathbb{A}$ as the list of contribution years that are not dropped out by the general drop-out. The following steps are taken to calculate the CPP pension amount: [(i)]

1. Calculate the number of base contributory months. The base contribution period is 47 years, from age 18 to age 65. Allowing 17% of the base contributory period (eight years) with lowest earnings to be dropped, the number of base contributory months is $NCM = (65 - 18 - 8) \times 12 = 468.$
2. Calculate unadjusted pensionable earnings in year t, UPE $_t$, for $t \in \mathbb{A}$:
$$UPE_t = \min(Y_t, YMPE_t)$$
3. Calculate adjusted pensionable earnings in year t, APE $_t$, for $t \in \mathbb{A}$:
$$APE_t = \frac{UPE_t}{YMPE_t} \times \frac{\sum_{t=2014}^{2018} YMPE_t}{5}$$
4. Calculate total adjusted pensionable earnings (TAPE)
$$TAPE = \sum_{t \in \mathbb{A}} APE_t$$
5. Calculate the monthly CPP retirement pension in 2018
$$0.25 \times \frac{TAPE}{NCM}$$

Since the CPP pension amount is indexed to the CPI, the monthly CPP retirement pension in a future year t for $t > 2018$ is $0.25 \times \frac{TAPE}{NCM} \times \frac{CPI_t}{CPI_{2018}}$.

Appendix B Endogenous grid method

Carroll (2006) proposes the EGM, which avoids the iterative search in VFI and hence requires significantly less computation time. The fundamental idea of the EGM is to specify an exogenous grid over the post-decision state instead of over the pre-decision states. Recall that the value function $V_t(m_t, n_t, \kappa_{t-1})$ for $t \leq 40$ is a function of the pre-decision-state (m_t, n_t, k_{t-1}). To illustrate the idea of EGM, we rewrite the value function into a function of post-decision-state (a_t, b_t, k_{t-1}) as follows:

$$
\begin{aligned}
& V_t(m_t, n_t, \kappa_{t-1}) \\
&= \max_{c_t, d_t} u(c_t) + \beta \mathbb{E}\left[p_{25+t-1,t}\, V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t) | m_t, n_t, \kappa_{t-1}\right] \\
&= \max_{c_t, d_t} u(c_t) + \beta \mathbb{E}[f(\kappa_t) V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t) | m_t, n_t, \kappa_{t-1}], \\
&= \max_{c_t, d_t} u(c_t) \\
&+ \beta \mathbb{E}[f(\kappa_{t-1} + \Delta\kappa_t) V_{t+1}(e^{R_t} a_t + y_{t+1}, e^{R_t} b_t + 0.05 Y_{t+1}, \kappa_{t-1} + \\
&\Delta\kappa_t) | a_t, b_t, \kappa_{t-1}],
\end{aligned}
$$

where $\Delta\kappa_t = \mu + \sigma z_t$ and $f(\kappa_t) = p_{25+t-1,t} = e^{-m_{25+t-1,t}} = e^{-e^{\beta^{(0)}_{25+t-1} + \beta^{(1)}_{25+t-1} \kappa_t}}$. In the last equality, we change the conditioned variables from pre-decision-state variables (m_t, n_t, κ_{t-1}) to post-decision-state variables (a_t, b_t, κ_{t-1}) because (a_t, b_t) are a deterministic functions of (m_t, n_t) given the choices (c_t, d_t).

Since R_t and $\Delta\kappa_t$ are random variables independent from the post-decision-state variables and y_{t+1} and Y_{t+1} are known values, the continuation value is a function of the post-decision-state variables. Denoting the continuation value function by $w_t(a_t, b_t, \kappa_{t-1})$, where

$$
\begin{aligned}
w_t(a_t, b_t, \kappa_{t-1}) = \mathbb{E}[f(\kappa_{t-1} + \Delta\kappa_t) V_{t+1}(R^a_{t+1} a_t + y_{t+1}, b_t R^b_{t+1}, \kappa_{t-1} + \\
\Delta\kappa_t) | a_t, b_t, \kappa_{t-1}],
\end{aligned}
$$

the value function can be rewritten as follows:

$$
\tilde{V}_t(a_t, b_t, \kappa_{t-1}) = \max_{c_t, d_t} u(c_t) + \beta w_t(a_t, b_t, \kappa_{t-1}). \tag{4}
$$

Note that a_t and b_t are functions of c_t and d_t due to the budget constraints. Suppose that the continuation value function is differentiable with respect to c_t and d_t and the choices constraints are unbinding. Using the FOCs, the optimal choices satisfy the following equations:

$$
c^*_t = \left[\beta w_{t,a}(a_t, b_t, k_t)\right]^{-\frac{1}{\gamma}}, \tag{5}
$$

$$
d^*_t = g_d^{-1}\left(\frac{w_{t,a}(a_t, b_t, k_t) - w_{t,b}(a_t, b_t, k_t)}{w_{t,a}(a_t, b_t, k_t)}\right), \tag{6}
$$

where

$$
w_{t,a}(a_t, b_t, k_t) = \frac{\partial w_t(a_t, b_t, k_t)}{\partial a_t},
$$

$$
w_{t,b}(a_t, b_t, k_t) = \frac{\partial w_t(a_t, b_t, k_t)}{\partial b_t},
$$

$$g_d(d_t) = \frac{\partial g(d_t)}{\partial d_t},$$

and $g_d^{-1}(.)$ is the inverse function of $g_d(.)$. A proof of equations (5) and (6) is provided in Appendix D. Assuming that the objective function is concave and the choice set is convex, the FOCs are sufficient conditions for global maximum and c_t^* and d_t^* must be the unique optimal choices. Using the EGM approach, we first construct an exogenous grid for post-decision-state (a_t, b_t, κ_{t-1}). For each point on this grid, we evaluate the continuation value function and its derivatives and then determine the optimal choices c_t^* and d_t^* using equations (5) and (6). Finally, we use the budget constraint to back out the values of the pre-decision-state variables corresponding to the optimal choices and post-decision-state variables as follows:

$$m_t^* = a_t + c_t^* + d_t^* - g(d_t^*),$$
$$n_t^* = b_t - d_t^*.$$

Since the pre-decision-state variables are endogenously determined, the grid constructed from their values are called an endogenous grid. The benefit of the EGM is that the optimal choices are found without numerical optimization, which is iterative and time-consuming. The continuation value, which is the expectation over the next-period variables, is only taken once for each exogenous grid point. In contrast, the numerical optimization used in VFI requires iterative procedures that calculate expectation in each iteration.

Appendix C Evaluating the value function by segments

C.1 Constructing exogenous grids

For year t, we first construct a common regular exogenous grid $\mathcal{G}_t^{m,n,\kappa}$ over the pre-decision states m_t, n_t, and κ_{t-1}. The goal of the algorithm is to evaluate $V_t(m_t, n_t, \kappa_{t-1})$ for each point in the exogenous grid $\mathcal{G}_t^{m,n,\kappa}$. The algorithm runs backward from age 99 to 98, 97, …, 25. Therefore, the value function for year $t+1$, $V_{t+1}(m_{t+1}, n_{t+1}, \kappa_t)$, must be evaluated for points on the grid $\mathcal{G}_{t+1}^{m,n,\kappa}$.

We construct another common regular grid $\mathcal{G}_t^{a,b,\kappa}$ over post-decision states a_t, b_t, and κ_{t-1} and calculate $w_t(a_t, b_t, \kappa_{t-1})$ and its derivatives $w_{t,a}(a_t, b_t, \kappa_{t-1})$ and $w_{t,b}(a_t, b_t, \kappa_{t-1})$ on this grid using numerical integration and interpolation.

C.2 Value function evaluation

For each segment, we determine the values of $V_t(m_t, n_t, \kappa_{t-1})$ for all the points on the grid $\mathcal{G}_t^{m,n,\kappa}$.

Case 1. Segments with both c_t and d_t constrained: $a_t = 0$ and $d_t = \hat{d} \in \{0, Y_t - h_j, Y_t - h_{j-1}, \dots, Y_t - h_{\hat{n}+1}, Y_t - h_{\hat{n}}\}$

Since both choices are constrained, there is no need for EGM and we can directly work with the regular gird $\mathcal{G}_t^{m,n,\kappa}$. Given a point $(\hat{m}_t, \hat{n}_t, \hat{\kappa}_{t-1})$ on the exogenous grid $\mathcal{G}_t^{m,n,\kappa}$, we use the budget constraints to determine the consumption choice:

$$\hat{c}_t = \hat{m}_t - \hat{d} + g(\hat{d}).$$

The corresponding post-decision states are

$$\hat{a}_t = 0,$$
$$\hat{b}_t = \hat{n}_t + \hat{d}.$$

Interpolating the values of $w_t(a_t, b_t, \kappa_{t-1})$ on the common grid $\mathcal{G}_t^{a,b,\kappa}$, we can calculate $w_t(\hat{a}_t, \hat{b}_t, \hat{\kappa}_{t-1})$ and hence the value of $V_t(\hat{m}_t, \hat{n}_t, \hat{\kappa}_{t-1})$ by

$$V_t(\hat{m}_t, \hat{n}_t, \hat{\kappa}_{t-1}) = u(\hat{c}_t) + \beta w_t(\hat{a}_t, \hat{b}_t, \hat{\kappa}_{t-1}).$$

Case 2. Segments with d_t constrained only: $a_t > 0$ and $d_t = \hat{d} \in \{0, Y_t - h_j, Y_t - h_{j-1}, \dots, Y_t - h_{\hat{n}+1}, Y_t - h_{\hat{n}}\}$

Given a point $(\tilde{a}_t, \tilde{b}_t, \tilde{\kappa}_{t-1})$ on the common grid $\mathcal{G}_t^{a,b,\kappa}$, we can calculate corresponding optimal consumption choice $\tilde{c}_t$ using the FOCs as follows:

$$\tilde{c}_t = \left[\beta w_{t,a}(\tilde{a}_t, \tilde{b}_t, \tilde{k}_{t-1})\right]^{-\frac{1}{\gamma}}$$

Using the budget constraints, we can invert the values of the corresponding endogenous pre-decision states:

$$\tilde{m}_t = \tilde{c}_t + \hat{d} - g(\hat{d}),$$

$$\tilde{n}_t = \tilde{b}_t - \hat{d}.$$

The value function $V_t(\tilde{m}_t, \tilde{n}_t, \tilde{\kappa}_{t-1})$ can also be evaluated. Therefore, for all points in the common grid over post-decision states $\mathcal{G}_t^{a,b,\kappa}$, we have node sets containing optimal choices, pre-decision states, and evaluated value functions. Note that the pre-decision states in the node sets are endogenously calculated rather than exogenously given and thus they form an irregular endogenous grid. In order to compute the value function for the regular common grid $\mathcal{G}_t^{m,n,\kappa}$, we follow the steps proposed in Druedahl and Jørgensen (2017):

1. Local triangulation:
   a) Compute a set of simplexes such that no data points on $\mathcal{G}_t^{a,b,\kappa}$ are contained in any circumspheres of the simplexes. The set of simplexes forms the Delaunay triangulation.
   b) Determine the corresponding simplexes mapped into (m_t, n_t, κ_{t-1}) space.
   c) Construct the simplexes' bounding boxes in (m_t, n_t, κ_{t-1}) space.
2. Interpolation to common pre-decision-state grid $\mathcal{G}_t^{m,n,\kappa}$ and first upper envelope. For each bounding box, do the following:
   a) Find the nodes in the common pre-decision-state grid $\mathcal{G}_t^{m,n,\kappa}$ inside the bounding box.
   b) Find candidate choices for each state-variable node $(m_t, n_t, \kappa_{t-1}) \in \mathcal{G}_t^{m,n,\kappa}$ inside the bounding box, using barycentric interpolation. To limit the use of extrapolation, we do not consider points with any barycentric weights less than -0.125.
   c) Calculate the implied value function of these interpolated choices.
   d) Update the optimal choice if no previous set of choices have been found yielding a higher value-of-choice.

The above steps evaluate the value function for the nodes in the common grid $\mathcal{G}_t^{m,n,\kappa}$.

Case 3. Segments with only c_t constrained: $a_t = 0$ and $d_t \in (ld, ud)$

The way we segment d_t ensures that $g(d_t)$ is differentiable for $d_t \in (ld, ud)$ and $g_d(d_t)$ is a constant. Let

$$\tilde{v}_t(a_t, b_t, \kappa_{t-1}) = u(c_t) + \beta w_t(a_t, b_t, \kappa_{t-1}).$$

Since $a_t = 0$, we have

$$c_t = m_t - d_t + g(d_t)$$

We obtain the first-order derivative for d_t as follows:

$$\frac{\partial \tilde{v}_t(a_t,b_t,k_t)}{\partial d_t} = \frac{\partial u(c_t)}{\partial c_t}\frac{\partial c_t}{\partial d_t} + \beta\left[w_{t,a}(a_t,b_t,k_t)\frac{\partial a_t}{d_t} + w_{t,b}(a_t,b_t,k_t)\frac{\partial b_t}{d_t}\right]$$
$$= c_t^{-\gamma}[-1 + g_d(d_t)] + \beta w_{t,b}(0,b_t,k_t)$$

Note that since $a_t = 0$, we have $\frac{\partial a_t}{d_t} = 0$. Fixing b_t and c_t, the above derivative is a constant since $g_d(d_t)$ is a constant for $d_t \in (ld, ud)$. Therefore, there is no FOC for d_t. In addition, $\tilde{v}_t(a_t, b_t, k_t)$ approaches its maximum when d_t approaches either lb or ub. Since $\tilde{v}_t(a_t, b_t, k_t)$ is defined and continuous at ld and ud, the value obtainable in this case is always smaller than those in Case 1. Therefore, this case should be discarded.

Case 4. Segments with no constraints: $a_t > 0$ and $d_t \in (ld, ud)$

Given a node $(\tilde{a}_t, \tilde{b}_t, \tilde{\kappa}_{t-1})$ on the common grid $\mathcal{G}_t^{a,b,\kappa}$, we use the FOCs to determine the candidate optimal choices:

$$\tilde{c}_t = \left[\beta w_{t,a}(\tilde{a}_t, \tilde{b}_t, \tilde{k}_t)\right]^{-\frac{1}{\gamma}}$$

However, there is no FOC for d_t. Let us revisit the first-order derivative for d_t when c_t is unconstrained:

$$\frac{\partial \tilde{v}(a_t, b_t, \kappa_t)}{\partial d_t} = \beta\left[\frac{\partial w_t(a_t, b_t, \kappa_{t-1})}{\partial a_t}\frac{\partial a_t}{\partial d_t} + \frac{\partial w_t(a_t, b_t, \kappa_{t-1})}{\partial b_t}\frac{\partial b_t}{\partial d_t}\right]$$
$$= \beta\left[w_{t,a}(a_t, b_t, \kappa_t)(-1 + g_d(d_t)) + w_{t,b}(a_t, b_t, \kappa_t)\right]$$

Since $g_d(d_t)$ is a constant, the first order derivative is also a constant. The value function is in fact a linear function of d_t in this segment given the exogenous post-decision grid point $(\tilde{a}_t, \tilde{b}_t, \tilde{\kappa}_{t-1})$ and approaches its maximum when d_t approaches either lb or ub. Since the value function is defined and continuous at lb and ub, the maximum value function obtainable in this case is always lower than those obtained in Case 2. Therefore, this case should also be discarded.

C.3 Second upper envelope over segments

Combining the four cases discussed above, the segments we consider should always have constrained d_t. Therefore, the total number of segments is $2(i - \hat{n} + 3)$. For each segment, we evaluated $V_t(m_t, n_t, \kappa_{t-1})$ for all the nodes on the common grid $\mathcal{G}_t^{m,n,\kappa}$. As a final step, we apply another envelope over all the segments. For each node on the common grid $\mathcal{G}_t^{m,n,\kappa}$, the optimal choice is chosen as the choice from the segment with the highest value-of-choice.

Appendix D Deriving first order conditions

Assume that the continuation value function is differentiable with respect to c_t and d_t. The optimal choices should satisfy the FOCs:

$$c_t^{-\gamma} + \beta \frac{\partial w_t(a_t,b_t,\kappa_{t-1})}{\partial a_t} \frac{\partial a_t}{\partial c_t} = 0,$$

$$\beta \left[\frac{\partial w_t(a_t,b_t,\kappa_{t-1})}{\partial a_t} \frac{\partial a_t}{\partial d_t} + \frac{\partial w_t(a_t,b_t,\kappa_{t-1})}{\partial b_t} \frac{\partial b_t}{\partial d_t} \right] = 0.$$

Since

$$a_t = m_t - c_t - d_t + g(d_t),$$

$$b_t = n_t + d_t,$$

we have

$$\frac{\partial a_t}{\partial c_t} = -1$$

$$\frac{\partial a_t}{\partial d_t} = -1 + \frac{\partial g(d_t)}{\partial d_t}$$

$$\frac{\partial b_t}{\partial d_t} = 1.$$

Given an exogenous grid point (a_t, b_t, k_t), the optimal choice can be found by solving the two equations. Assume that the value function is differentiable with respect to c_t and d_t. Using the FOCs, the optimal choices can be written as follows:

$$\hat{c}_t = \left[\beta w_{t,a}(a_t, b_t, k_t) \right]^{-\frac{1}{\gamma}},$$

$$\hat{d}_t = g_d^{-1} \left(\frac{w_{t,a}(a_t,b_t,k_t) - w_{t,b}(a_t,b_t,k_t)}{w_{t,a}(a_t,b_t,k_t)} \right),$$

where

$$w_{t,a}(a_t, b_t, k_t) = \frac{\partial w_t(a_t,b_t,k_t)}{\partial a_t},$$

$$w_{t,b}(a_t, b_t, k_t) = \frac{\partial w_t(a_t,b_t,k_t)}{\partial b_t},$$

$$g_d(d_t) = \frac{\partial g(d_t)}{\partial d_t},$$

and $g_d^{-1}(.)$ is the inverse function of $g_d(.)$. Assume that the first-order derivatives are convex, the FOCs are both necessary and sufficient. Therefore, $\hat{c}_t$ and $\hat{d}_t$ must be the unique optimal choice for the exogenous grid point (a_t, b_t, k_t).

References

Bellman, R.E. (1957). *Dynamic Programming.* Princeton University Press, Princeton.

Brouhns, N., Denuit, M., and Keilegom, I.V. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 2005(3), 212–224.

Carroll, C.D. (2006). The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems. *Economics Letters*, 91(3), 312–320.

Druedahl, J., and Jørgensen, T. (2017). A General Endogenous Grid Method for Multi-Dimensional Models with Non-Convexities and Constraints. *Journal of Economic Dynamics and Control*, 74(C), 87–107.

Gong, G., and Webb, A. (2010). Evaluating the Advanced Life Deferred Annuity – An Annuity People Might Actually Buy. *Insurance: Mathematics and Economics*, 46(1), 210–221.

Horneff, V., Maurer, R., and Mitchell, O.S. (2016). Putting the Pension Back in 401(k) Plans: Optimal versus Default Longevity Income Annuities. *NBER Working Paper.* No. 22717. Downloaded at www.nber.org/papers/w22717

Huang, H., Milevsky, M.A., and Salisbury, T. (2009). A Different Perspective on Retirement Income Sustainability: The Blueprint for a Ruin Contingent Life Annuity (RCLA). *Journal of Wealth Management*, 11(4), 89–96.

Huang, H., Milevsky, M., and Salisbury, T. (2014). Valuation and Hedging of the Ruin-Contingent Life Annuity (RCLA). *Journal of Risk and Insurance*, 81(2), 367–395.

Lee, R.D., and Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), 659–671.

Li, N., and Lee, R. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee–Carter Method. *Demography*, 42(3), 575–594.

Maurer, R. , Mitchell, O.S., Rogalla, R., and Kartashov, V. (2013). Lifecycle Portfolio Choice With Systematic Longevity Risk and Variable Investment-Linked Deferred Annuities. *Journal of Risk and Insurance*, 80(3), 649–676.

Milevsky, M.A. (2005). Real Longevity Insurance with a Deductible: Introduction to Advanced-Life Delayed Annuities (ALDA). *North American Actuarial Journal*, 9(4), 109–122.

Pfau, W., Tomlinson, J., and Vernon, S. (2016). Optimizing Retirement Income Solutions in Defined Contribution Retirement Plans. Downloaded at http://longevity3.stanford.edu/optimal-retirement-income-solutions-in-defined-contribution-retirement-plans/

Yaari, M.E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *Review of Economic Studies*, 32(2), 137–150.