Retirement Annuities: Optimization, Analysis and Machine Learning

by

Branislav Nikolić

A Dissertation submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of:
Doctor of Philosophy

Graduate program in Applied Mathematics
York University
Toronto, Ontario

May 2023

Abstract

Over the last few decades, we have seen a steady shift away from Defined Benefit (DB) pension plans to Defined Contribution (DC) pension plans in the United States. Even though a deferred income annuity (DIA) purchased while saving for retirement can pay a guaranteed stream of income for life, practically serving as a pension substitute, several questions arise. Our main contribution is answering the question of purchasing DIAs under the interest rate uncertainty. We pose the question as an optimal control problem, solve its centerpiece Hamilton-Jacobi-Bellman equation numerically, and provide a verification theorem. The result is an optimal DIA purchasing map.

With Cash Refund Income Annuities (CRIA) gaining traction quickly over the past few years, the literature is growing in the area of price sensitivity and its viability when viewed through the lens of key pricing parameters, particularly insurance loading. To that end, we explored the effect of reserving requirements on pricing and have analytically proven that, if accounted for properly at the beginning, reserving requirements would be satisfied at any time during the lifetime of the annuity.

Lower interest rates in the last decade prompted the explosion of fixed indexed annuities (FIAs) in the United States. These popular insurance policies offered a growth component with the addition of a lifetime income provisions. In FIAs, accumulation is achieved through exposure to a variety of indices while offering principal protection guarantees. The vast array of new products and features have created the need for a means of consistent comparisons between FIA products available to consumers. We illustrate that statistical issues in the temporal and cross-sectional return correlations of indices used in FIAs necessitates more sophisticated modelling than is currently employed. We outline few novel approaches to handle these two issues. We model the risk control mechanisms of a class of FIA indices using machine learning. This is done using a small set of core macroeconomic variables as modelling features. This makes for more robust cross-sectional comparisons. Then we outline the properties of a sufficient model for said features, namely 'rough' stochastic volatility.

Dedication

This dissertation is dedicated to my beloved parents, Predrag and Nataša, whose love and encouragement live on in every step I take.

Acknowledgements

"If I have seen further, it is by standing on the shoulders of giants" – Sir Isaac Newton, 1676

The same is true for my PhD journey; I would have not been the same without my own giants, their wisdom, encouragement, and unwavering belief in my abilities along the way.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Thomas Salisbury, whose dedication to my academic, professional and personal growth has been a constant source of inspiration. Tom, your firm guidance, insightful advice, and commitment to excellence have significantly enriched my research. Thank you also for confidently encouraging me to take on adventures you well knew would stand between me and my PhD, but also reeling me back from time to time. Thank you for being that constant force of stability and comfort. Thank you for teaching me how to lean back, scratch my head and think deeply about a math problem.

Thank you to my committee members for their guidance and tireless support during this endeavour, but even more importantly during the entirety of my academic journey and professional career. To Professor Huaxiong Huang, thank you for teaching me to have fun with math and to think beyond numbers, differential equations and numerical schemes, but also thank you for always being there to help when I struggled with them. To Professor Alexey Kuznetsov, thank you for all the hard questions and insightful answers over last 15 years. Having you as a professor, a supervisor and a friend is the most valuable gift York has given me and I am grateful beyond words for that. To my committee chair, Professor Moshe Arye Milevsky, thank you for always pushing me to work harder, go further and think deeper. Thank you for giving me the opportunity and guiding me to become the retirement quant that I am today.

Thank you to my external examiner, Professor Adam Metzler, whose detailed review and insightful comments contributed significantly getting my thesis across the finish line. To Professor Mark Kamstra, thank you for carefully reading my thesis and giving me great ideas on how to apply the results in the world of finance.

A big thank you goes to my colleagues at Cannex Financial Exchanges and The Index Standard for their support and understanding during this balancing act of mine over the last few years. To Lowell Aronoff, Gary Baker and Faisal Habib, thank you for supporting the idea and believing that one can work and go to school at the same time, even when I myself had doubts. To Damian Baboolal, thank you for being the actuary to my quant. Thank you also to Laurence Black and Jay Watson: our discussions over the years encouraged me to explore further and your partnership helped me cross the finish line in stride. Lastly, a very special thank you goes to my dear colleague and friend, my "work sibling", Tamiko Toland, for being my sounding board, writing tutor and true rock over the years.

They say, "it takes a village to raise a child" and I learned firsthand that it's equally true for a PhD. Thank you to all my friends and their families for the support and encouragement during this time. A special thank you to all my "math" friends, whose ideas and suggestions made significant contributions to my graduate work. To my collaborator and friend, Andrew Fleck, thank you for the endless fun while working on our projects over the years and many good laughs along the way. To Mohammad Samani, my dear friend, who taught me so much about what being a true scientist really is, but more importantly who thought me how to program and how not to cook. To Nathan Gold, thank you duuuude from the bottom of my heart. If I were to do this again, I would choose to do it with you every single time. Meeting you and Nataly and sharing this experience with you changed my life forever and made comprehensive exams, late nights and endless projects taste better. You are the only person that I would make a mathematical model for popcorn with.

Thank you to my wonderful family. Without their support and boundless love, none of this would be imaginable, let alone possible. To Maša and Vanja, my nieces and loudest cheerleaders, thank you for your love, encouragement and beautiful smiles that inspired me on this long road. To Miloš, thank you for your sage advice, endless support and encouragement. I have never imagined that having a sister would give me a big brother to look up to and call my own. To Ana, my sister and my lifelong inspiration, thank you for everything that I am today. Your love and care for me means more than any words can ever hold. I love you immeasurably. To Luka and Mina, my everything, thank you being

my guiding stars and greatest motivation. To Maša, my love and my best friend, we did it. Thank you for always believing in me and never sparing the words to share it. I could not have done this without you by my side and I am forever grateful for your patience, support and encouragement. Thank you with all my love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> An annuity is a very serious business; it comes over and over every year, and there is no getting rid of it.
>
> Jane Austen, Sense and Sensibility, Volume I, Chapter 2

Guaranteed lifetime income is a key component of retirement security. Over the last few decades, we have seen a shift away from Defined Benefit (DB) pension plans to Defined Contribution (DC) pension plans in the United States, especially among younger workers. Many retirees rely on government-run pension plans alone, which may not be sufficient to maintain a desirable standard of living during retirement. Retirees may purchase payout annuities to augment other sources of retirement income. However, because consumers and advisors have not universally embraced this approach, a significant income gap remains for many retirees. The shift to DC plans is usually attributed to many firms not being willing or able to fully fund or manage the future liabilities of a DB plan, including market uncertainties and improvements in human longevity, in addition to a higher motility of the workforce compared to the past.

The Setting Every Community Up for Retirement Enhancement (SECURE) Act, passed in late 2019 in the United States created a safe harbor for the selection of annuities in employer retirement plans, removing one of the barriers to adoption of these options. As conversations about worker access to guaranteed lifetime income grow, it is instructive to have frameworks that can guide the use of these solutions.

One common structure considered in practice is a target date fund with an annuity allocation replacing some or all of the fixed income investment. Even though a deferred income annuity (DIA) purchased while saving for retirement can pay a guaranteed stream of income that begins at retirement, this approach alone poses more questions than it answers. What is the optimal time to start purchasing DIAs? What is the right amount compared to other assets inside and outside the plan? How does a non-liquid asset such as DIA compare to fully liquid bonds? These are only a few of the questions that arise when we delve into the issue of increasing retirement security through DIA purchases and which we try to answer in Chapter 2, specifically in the context of stochastic interest rates. We approach the question of DIA purchases as an optimal control problem, solve its centerpiece Hamilton-Jacobi-Bellman equation numerically and provide a verification theorem. The result is an optimal purchasing map.

With Cash Refund Income Annuities (CRIA) gaining traction quickly over the past few years, the literature is growing in the area of price sensitivity and its viability when viewed through the lens of key pricing parameters, particularly insurance loading. To that end, in Chapter 3, we look into the effect of reserving requirements on pricing and we analytically prove that, if accounted for at the beginning, reserving requirements would be satisfied at any time during the lifetime of the annuity.

Lastly, we acknowledge that annuitization rates are low and that either of the two income annuity solutions considered in the first two chapters may not emerge as the dominant solution for lifetime income. With that understanding, in Chapter 4 we focus on Fixed Indexed Annuities (FIA) a popular type of deferred annuity in the United States that may be offered with an income guarantee that is an alternative to annuitization. In order to select an FIA or make an allocation among the indices and strategies available within it,

it is important to understand both the accumulation structure of the FIA, as well as the underlying index features. Moreover, in order to properly compare two or more annuities struck on different underlying indices, the combination of the index and crediting method must be modelled and forecasted in a consistent manner. In this chapter, we examine ways to replicate index performance, pass model obtained index returns through FIA's crediting formulae, and ultimately allow comparison and informed selection. While we tested classical index modeling approaches, we settled on a few machine leaning techniques as a superior solution for inferring index dynamics from the broader market data set.

Chapter 2

Optimal Allocation to Deferred Income Annuities when Interest Rate is Stochastic

2.1 Introduction

Guaranteed lifetime income is a key component of retirement security. Over the last few decades, we have seen a shift away from Defined Benefit (DB) pension plans to Defined Contribution (DC) pension plans in The United States (Employee Benefit Research Institute (2014)), especially among younger workers. Many retirees rely on government-run pension plans alone, which may not be sufficient to maintain a desirable standard of living during retirement. Retirees may purchase payout annuities to augment other sources of retirement income. However, because consumers and advisors have not universally embraced this approach, a significant income gap remains for many retirees. The shift to DC plans is usually attributed to many firms not being willing or able to fully fund or manage the future liabilities of a DB plan, including market uncertainties and improvements in human longevity, in addition to a higher motility of the workforce compared to the past. In a DB plan, future income is guaranteed by the employer, whereas no such guarantee is made in

a DC plan. The employer simply contributes to the plan on behalf of the employee (and oftentimes matches the employee contributions).

The Setting Every Community Up for Retirement Enhancement (SECURE) Act, passed in late 2019 in the United States created a safe harbor for the selection of annuities in employer retirement plans, removing one of the barriers to adoption of these options. As conversations about worker access to guaranteed lifetime income grow, it is instructive to have frameworks that can guide the use of these solutions. One common structure considered in practice is a target date fund with an annuity allocation replacing some or all of the fixed income investment. Even though a deferred income annuity (DIA) purchased while saving for retirement can pay a guaranteed stream of income that begins at retirement, this approach alone poses more questions than it answers. What is the optimal time to start purchasing DIAs? What is the right amount compared to other assets inside and outside the plan? How does a non-liquid asset such as a DIA compare to fully liquid bonds? These are only a few of the questions that arise when we delve into the issue of increasing retirement security through DIA purchases.

Our main objective is to devise a DIA purchasing strategy for someone who seeks to maximize utility of income obtained from both accumulated wealth and a DIA that begins payments at the point of retirement. Our agenda is to answer questions around variable interest rates and how the annuity purchase decision can vary under different circumstances. We believe that, among other factors, interest rates and future uncertainties around them are pivotal factors that should be given special consideration in the decision to purchase an annuity. Additionally, with interest rates at an historic low at the time of this writing (early 2020), we would like to examine under what conditions a DIA purchase still makes sense. We rely on the CIR model that is time-tested in the literature as an appropriate short rate model. To be able to deal with the newly introduced complexity of stochastic interest rates, we choose a simplified personal finance model to exclude both labour income and any existing pensions on the side. This isolates the problem to annuity purchases only. The work therefore distinguishes itself from other portfolio choice literature and instead positions itself firmly among those that examine the timing of annuity purchases.

Our main question is: how much of one's current liquid wealth should be exchanged for a DIA if interest rates are stochastic? Should one make a purchase immediately or wait based on the interest rate environment in conjunction with the level of wealth and retirement income already secured?

Assuming a stochastic interest rate, the argument can be further extended by looking into the pricing mismatch between the real interest rate world and approximations made for annuity pricing. We examine two distinct scenarios. First, we consider a DIA priced fully under the CIR model and then look into sensitivity to model parameters and their inconsistency with ones that went into the annuity price. We explore what factors, such as refundability, speed of reversion, or simply the long-term mean of interest rates, have the biggest impact on the decision to annuitize. Secondly, we apply a more practical approach to pricing. Namely, we price a long-dated bond under the CIR model and translate that into a constant rate that is then, with or without haircut, applied to DIA pricing. We look into any difference in the decision to buy a DIA or not at predetermined points in time prior to retirement. Anecdotally, insurance carriers do stay with simple constant rate models; however, the rates that are used are a product of different loads, mortality assumptions and business-oriented optimizations. A constant-rate approach may seem naïve but serves to illustrate whether annuitization decisions change and by how much with the divergence of the price and value between the insurer and consumer.

2.1.1 Literature review

In his pioneering work, Yaari (1965) determined that everyone looking to maximize their utility of consumption and with no bequest motive should hold life contingent assets over liquid ones. This came from the premise that, out of the pool of participants, the dead will be subsidizing the retirement income streams of the living. As such, the individual can benefit beyond the means of their own assets. With this theoretical and concrete result, more and more researchers turned to the practical question about why still only a small fraction of retirees annuitize portions of their wealth. This is also known as the annuity

puzzle and became a cornerstone of the portfolio choice literature, such as Campbell & Viceira (2002), Neuberger (2003), Cocco et al. (2005), Horneff et al. (2009b), Koijen et al. (2011a), and Hainuat & Deelstra (2014), to name some. In Milevsky & Young (2007), the authors explore the benefits of delaying annuitization by a utility-maximizing agent and leave us with the conclusion that this incentive is different depending on the annuitization structure, gradual vs. "all or nothing," and that even the latter would have its benefits in the absence of bequest motives. Additionally, Blanchett (2015) presents a model to support the purchase of annuities after retirement. What the author calls the "funding ratio" determines how many more units of annuity would be suitable for the individual situation. For the interested reader, the literature review of Dillschneider et al. (2020) provides quite an extensive overview of the field in greater depth.

In addition to the aforementioned classics in the life cycle modeling and portfolio choice literature, we focus our attention on four papers that are closest in nature to ours. Huang, Milevsky & Young (2017) offer insight into the stochastic nature of annuity prices through the single factor of the annuity payout. As we know, this is a function of both interest rate and mortality, as well as any additional frictions that exist in the annuity pricing that separate it from the fair actuarial price. This paper presents a high-level view of the payout region where the annuity purchase makes sense, under the assumptions that said payout yield is indeed mean reverting and expected to evolve as modeled. This approach allows the authors to focus on their main question by removing computational complexity. However, it imposes barriers to practical implementation. Habib et al. (2020) explore a life cycle model that considers many variables in pre- and post-retirement that have impact on the decision to purchase annuity at retirement, whether early or not at all. It considers different mortality schemes portrayed through the refund structure of the annuity, asset allocation decision (if allowed), exogenous income, and pensions. In contrast to Huang, Milevsky & Young (2017), this paper provides a much more robust mechanism when modeling individual financial situations, sources of risk, and decision making, but falls short in the aspect of modelling interest rates. More precisely, it prices annuities under a fixed interest rate. In Charupat et al. (2016), the authors explore responsiveness of annuity prices to changes in interest rates. They define duration for income annuities

and measure it on an historical data set. Their conclusion is that changes in annuity prices induced by interest rate movements are not immediate and fully observable in the short term. Finally, Dillschneider et al. (2020) discuss diversifying away interest rate risk through purchasing annuities gradually in pre-retirement. Welfare gain increases even further with rising interest rates. Their paper is one of the few that incorporates stochastic interest rates while assessing optimal life cycle demand for annuities along with mortality, labour income, and risk arising from financial markets. They opt to use Vasicek over CIR to allow for negative real interest rates and for the sake of modeling and computational simplicity.

Charupat et al. (2016) provides a nice overview of the literature in four main pillars of annuity research: portfolio choice; the annuity puzzle; "money's worth;" and valuation of pension liabilities. We see this work in part belonging to the portfolio choice literature. By equal measure, one can argue that it is fully agnostic to the broad position of annuities within one's portfolio and simply provides guidance as to when it is optimal to turn to life contingent claims over liquid assets in the interest-bearing account. In practice, we know that annuity prices are slow and irregular in responding to interest rate changes. Nevertheless, like the majority of the aforementioned literature, we choose a continuous-time framework and assume that annuity prices will adjust as the interest rates change. We align our work with Huang, Milevsky & Young (2017) in the sense that we are only looking at the choice of the annuity over similar non-contingent assets but not its appropriateness and place within a broader portfolio of stocks and bonds. We dig a bit deeper into the structure of annuity prices and choose to model interest rates stochastically as the main driver of price change. This is also in contrast with our previous work (see Habib et al. (2020)), where we study the annuity purchasing decision in a broader financial planning sense but under a less realistic constant interest rate scenario.

2.1.2 Agenda of the chapter

In section 2.2, we provide a brief overview of DIA pricing, both under constant and stochastic interest rates. In section 2.3, we present wealth dynamics under the CIR interest rate model and set up the problem from a financial planning perspective. Section 2.4 is dedi-

cated to our optimization problem formulation, and development of the HJB equation and its boundary conditions. We touch briefly on solution methodologies and parameter choices for our base case and set direction for analysis. In section 2.5, we present and discuss our findings under various scenarios, while in section 2.6 we look into DIA mispricing and its impact on the optimal decision to annuitize.

2.2 Income annuities

2.2.1 Constant interest rate pricing

An annuity is an insurance product that pays an annuitant (purchaser) an income for the rest of their life. The stochastic present value under a constant interest rate scenario of a pension annuity that pays \$1 is

$$a_x = \int_0^{T_x} e^{-rs} ds = \int_0^{\infty} e^{-rs} \mathbb{1}_{\{T_x \geq s\}} ds \tag{2.1}$$

where T_x is the remaining lifetime random variable. The expected value of the above random variable is known as the Immediate Pension Annuity Factor (IPAF)

$$\bar{a}_x = \mathbf{E}\left[\int_0^{\infty} e^{-rs} \mathbb{1}_{\{T_x \geq s\}} ds\right] = \int_0^{\infty} e^{-rs}{}_s p_x ds \tag{2.2}$$

where ${}_t p_x$ is the t-year probability of survival of an x-year old individual.

A deferred income annuity (DIA) is similar to an immediate income annuity. However it starts paying income after a deferral period τ. With a constant interest rate the expected value of a DIA that pays \$1 after the deferral period is

$${}_\tau \bar{a}_x = \int_\tau^{\infty} e^{-rs}{}_s p_x ds = e^{-\tau r}{}_\tau p_x \int_0^{\infty} e^{-rs}{}_s p_{(x+\tau)} ds \tag{2.3}$$

which can be simplified as

$$ {}_{\tau}\bar{a}_x = e^{-\tau r}\, {}_{\tau}p_x \bar{a}_{(x+\tau)}. \tag{2.4} $$

The above equation is for a DIA which does not offer any refund if the investor dies before retirement. In reality, most DIAs currently available in the market offer a refund, so we want to model the refund aspect of the DIA as well. Consider the following version of a DIA: each unit of DIA that the investor buys is refunded to them should they die, at the value that it is worth on the day of death. For example, if the investor buys two units of DIA (that is to say, they are now entitled to \$2 per year in retirement) at \$100, and then dies in one year when the DIA is worth \$110, then their estate would receive \$220.

With the above refund scheme, there is essentially no financial impact on legacy due to pre-retirement death, since the entire sum of money deposited into the account is returned, with interest. This theoretical DIA will be priced at

$$ {}_{\tau}\tilde{a}_x = \bar{a}_{(x+\tau)} e^{-r\tau}. \tag{2.5} $$

Equation (2.5) is almost identical to equation (2.4); we simply remove the factor ${}_{\tau}p_x$ since that is the probability the refund will not take place.

We will analyze DIAs that interpolate between these extremes. We use a weighted average of the no-refund DIA, and the full-refund DIA, and introduce a parameter Q that takes values between 0 and 1. That is,

$$ {}_{\tau}\tilde{a}^{Q}_x = \bar{a}_{x+\tau} e^{-r\tau} [{}_{\tau}p_x (1-Q) + Q]. \tag{2.6} $$

As we shift Q from 0 to 1, we change the nature of the DIA. A Q value of 0 implies a DIA with no refund (i.e. it includes mortality credits), while a Q value of 1 implies a DIA with a full refund (i.e. it includes no mortality credits).

This representation for refunds was introduced in Habib et al. (2020). It is worth explicitly noting, that our model should be able to incorporate DIA prices that include

any type of built in guarantee or refund, regardless of how they are priced in or how they work.

We use the Gompertz-Makeham Law of Mortality to model human mortality in which, $\lambda(x)$, the Instantaneous Force of Mortality (IFM), is defined as

$$\lambda(x) = \lambda_0 + \frac{1}{b} e^{(x-m)/b} \tag{2.7}$$

where x is the age, λ_0 captures the death rate attributable to accidents, m is the modal value of life[1], and b is the dispersion coefficient. The conditional survival probability is then

$${}_tp_x = e^{-\lambda_0 t + \left(1 - e^{\frac{t}{b}}\right) e^{\frac{x-m}{b}}} \tag{2.8}$$

where once again x is the age of the individual and t is the number of survival years. This analytical representation of survival probabilities can be calibrated to the mortality tables that are readily available from the Society of Actuaries.

2.2.2 Stochastic interest rates

We are interested in modelling the optimal allocation to DIA when interest rates are stochastic, and moreover follow a CIR process. CIR interest rate dynamics are as follows:

$$dr_t = \alpha(\mu - r_t)dt + \sigma\sqrt{r_t}dBt \tag{2.9}$$

Since DIA can be viewed as a portfolio of zero coupon bonds that defaults at the time of death of the annuitant, we derive pricing formula for immediate and deferred annuities under the CIR model by borrowing closed form equations for the pricing of zero coupon bonds. We compute actuarial present values of each \$1 of future income from the annuity.

Since CIR is an affine short rate model, we can write the price of a zero coupon bond as:

[1] strictly speaking, it is the mode only when $\lambda_0 = 0$

$$P(t,T) = \exp[A(t,T) - B(t,T)r_t] \tag{2.10}$$

with $\gamma = \sqrt{\alpha^2 + 2\sigma^2}$, $A(t,T) = \frac{2\alpha\mu}{\sigma^2}\log(\frac{2\gamma\exp[(\gamma+\alpha)(T-t)/2]}{(\gamma+\alpha)(\exp(\gamma(T-t))-1)+2\gamma}$ and $B(t,T) = \frac{2(\exp(\gamma(T-t))-1)}{(\gamma+\alpha)(\exp(\gamma(T-t))-1)+2\gamma}$

We use annuity pricing formulas similar to the constant interest rate case. The Immediate Pension Annuity Factor (IPAF) is

$$\bar{a}_x = \int_0^\infty P(0,s)_s p_x ds \tag{2.11}$$

Similarly, the pricing formula for a DIA, without refund of premium priced in, is

$${}_\tau\bar{a}_x = \int_\tau^\infty P(0,s)_s p_x ds = {}_\tau p_x \int_0^\infty P(0,\tau+s)_s p_{x+\tau} ds \tag{2.12}$$

The price of a DIA that has a refund in the form of commuted value, as described in section 2.2.1 is

$${}_\tau\tilde{a}_x = \int_0^\infty P(0,\tau+s)_s p_{x+\tau} ds \tag{2.13}$$

To account for possible return of premium in case of the annuitant's death prior to retirement and commencement of DIA payments, we borrow the sliding factor model from Habib et al. (2020) as described in 2.2.1 and use the following annuity pricing formula:

$${}_\tau\tilde{a}_x^Q = (1-Q)_\tau\bar{a}_x + Q_\tau\tilde{a}_x \tag{2.14}$$

At this point it is worth noting that annuity pricing formula in (2.14) is the most general version when interest rate is stochastic. For notation simplicity superscript Q will be dropped going forward and every time ${}_\tau\tilde{a}_x$ is used in the equation or expression it will mean exactly ${}_\tau\tilde{a}_x^Q$ from (2.14).

2.3 Model

When it comes to retirement savings accounts, a defined contribution plan (such as a 401(k) plan) allows the plan participant to make a contribution (typically matched by the employer up to a certain limit) that is invested into a portfolio of funds (mutual funds, ETFs, or a portfolio of stocks and bonds) as per the instructions of the plan participant. We restrict these choices to better understand annuitization decisions and the impact of interest rate uncertainty. We allow the plan participant to hold money in the interest-bearing account to purchase DIAs over time. We assume that the interest rate credited to the account value follows a Cox-Ingersoll-Ross (CIR) model, to make it consistent with annuity pricing presented in section 2.2.2. In addition, we consider two methods for the insurance company issuing DIAs to price them, one where annuities are priced under CIR and another where bond prices are obtained via CIR pricing and then used as a proxy for a constant rate in the pricing formula. This is an attempt to capture any potential mispricing inherent in the actual practice of insurers. In contrast with Habib et al. (2020), we only model the pre-retirement phase, where one is able to buy DIAs or postpone the decision until the time of retirement.

Before a fixed retirement date τ, i.e., $t < \tau$, the wealth dynamics for the retirement account W_t are given by

$$dW_t = r_t W_t dt + \nu_t dt - dG_t \tag{2.15}$$

where ν is a steady exogenous deposit stream to the retirement funds, and $G_t \geq 0$ is the amount paid by time t for DIAs with income starting at time τ. So $dG_t \geq 0$. Of course, we require that $dG_t \leq \nu_t dt$ when $W_t = 0$, in order to ensure the existence of a non-negative solution W_t. Additionally, we assume that the investor wants to convert their entire wealth account into immediate or deferred annuity income. This implies that, at the final moment before retirement, they buy an immediate annuity with all the money left over in the account.

The income from the accumulation of DIA, denoted by I_t, is given by

$$dI_t = \frac{dG_t}{{}_{(\tau-t)}\tilde{a}_{(x+t)}} \tag{2.16}$$

We set the problem up with ν as above, but only solve for case $\nu = 0$.

2.4 Optimal allocation to DIAs

The basic question in pre-retirement ($t < \tau$) is to determine the optimal strategy G_t for purchasing annuities, if at all. Since the dynamics for wealth, annuity income and interest rates are Markovian, we can write the value function as follows:

$$J(t, w, i, r) = \max_{G} \mathbb{E}\left[U\left(I_\tau + \frac{W_\tau}{{}_{(\tau)}\tilde{a}_{(x+\tau)}}\right) \middle| I_t = i, W_t = w, r_t = r\right] \tag{2.17}$$

Here the maximum is over all increasing adapted purchasing strategies G on the time interval $[t, \tau]$ that leave the investor with non-negative wealth ($w \geq 0$). $U(x)$ is a utility function. We adopt the Constant Relative Risk Aversion (CRRA) utility model, with $U(x) = \frac{x^{1-\gamma}}{1-\gamma}$, where γ is the subjective risk aversion of a particular investor. We choose the CRRA utility for its separability feature that reduces further computational complexity. The argument of U reflects full annuitization at time τ. Alternatively, we could have imposed a constraint $W_\tau = 0$.

2.4.1 Hamilton-Jacobi-Bellman equation

For dynamics of J we get:

$$dJ = \frac{\partial J}{\partial t}dt + \frac{\partial J}{\partial i}dI_t + \frac{\partial J}{\partial w}dW_t + \frac{\partial J}{\partial r}dr_t + \frac{1}{2}\frac{\partial^2 J}{\partial r^2} < dr >_t \tag{2.18}$$

and

$$dJ = J_t dt + J_i \frac{dG_t}{(\tau - t)\tilde{a}_{(x+t)}} + J_w[(r_t w_t + \nu_t)dt - dG_t] + J_r[\alpha(\mu - r_t)dt + \sigma\sqrt{r_t}dB_t] + \frac{1}{2}J_{rr}\sigma^2 r_t dt, \tag{2.19}$$

at least when G_t is continuous.

This should be a super martingale for the arbitrary choice of G, so both the dG_t and dt terms are ≤ 0. In other words:

$$\frac{J_i}{(\tau - t)\tilde{a}_{(x+t)}} - J_w \leq 0 \tag{2.20}$$

$$J_t + J_w(rw + \nu) + J_r\alpha(\mu - r) + \frac{1}{2}J_{rr}\sigma^2 r \leq 0. \tag{2.21}$$

Note that if $\Delta G > 0$ at t, then J jumps by

$$\Delta J = J(t, W_{t-} - \Delta G, I_{t-} + \frac{\Delta G}{\tilde{a}}, r_t) - J(t, W_{t-}, I_{t-}, r_t) \tag{2.22}$$

so (2.20) also implies that $\Delta J \leq 0$, and further that $\Delta J = 0$ if (2.20) holds with equality over the range of values involved.

Moreover the optimal choice of G^* should make both (2.20) and (2.21) equal to zero along the path of (W_t, I_t, r_t). We now simplify the problem by setting $\nu_t = 0$, i.e. assuming no exogenous income flowing in, namely $\nu_t = 0$. We do so to preserve the scaling and reduce computational complexity of the PDE. We look for scaling with respect to wealth. The economic argument is that the level of wealth on its own should not impact our decision to annuitize but should be taken into consideration only in conjunction with DIA income already secured. This would suggest income scaled by wealth. As much as this would be the preferred argument from the finance and economics stand point, it doesn't come with computational stability for our numerical scheme. We instead scale our PDE with respect to income (already accumulated through DIA purchases) and proceed to find the optimal purchase strategy.

From (2.20) and (2.21) and the terminal condition

$$J(\tau, w, i, r) = \frac{\left(i + \frac{w}{(\tau-t)\tilde{a}_{(x+t)}}\right)^{(1-\gamma)}}{(1-\gamma)} \tag{2.23}$$

we use the following scaling:

$$J(t, w, i, r) = U(i)\phi\left(t, \frac{w}{i}, r\right) \tag{2.24}$$

Where $z = \frac{w}{i}$ and derivatives are as follows:

$$J_w = \phi_z \frac{U(i)}{i} \tag{2.25}$$

$$J_i = \phi_z \left(-z\frac{U(i)}{i}\right) + \phi U^{'}(i) \tag{2.26}$$

$$J_r = \phi_r U(i) \tag{2.27}$$

$$J_{rr} = \phi_{rr} U(i) \tag{2.28}$$

$$J_t = \phi_t U(i) \tag{2.29}$$

The equations become:

$$-\left(1 + \frac{z}{(\tau-t)\tilde{a}_{(x+t)}}\right)\phi_z + \frac{1-\gamma}{(\tau-t)\tilde{a}_{(x+t)}}\phi \leq 0, \gamma \leq 1 \tag{2.30}$$

$$-\left(1 + \frac{z}{(\tau-t)\tilde{a}_{(x+t)}}\right)\phi_z + \frac{1-\gamma}{(\tau-t)\tilde{a}_{(x+t)}}\phi \geq 0, \gamma \geq 1 \tag{2.31}$$

$$\phi_t + \alpha(\mu - r)\phi_r + \frac{1}{2}\sigma^2 r\phi_{rr} + rz\phi_z \leq 0, \gamma \leq 1 \tag{2.32}$$

$$\phi_t + \alpha(\mu - r)\phi_r + \frac{1}{2}\sigma^2 r\phi_{rr} + rz\phi_z \geq 0, \gamma \geq 1 \tag{2.33}$$

If it is not optimal to annuitize, based on the equations 2.32 and 2.33, we have the

following continuation equation:

$$\phi_t + \alpha(\mu - r)\phi_r + \frac{1}{2}\sigma^2 r\phi_{rr} + (rz)\phi_z = 0 \tag{2.34}$$

Conversely, if is optimal to annuitize, the annuitization condition is therefore given by:

$$-\left(1 + \frac{z}{(\tau-t)\tilde{a}_{(x+t)}}\right)\phi_z + \frac{1-\gamma}{(\tau-t)\tilde{a}_{(x+t)}}\phi = 0 \tag{2.35}$$

As a numerical solution, when $\gamma \geq 1$ we take the minimum of (2.33) and (2.35), and, when $\gamma \leq 1$, the maximum of (2.32) and (2.35). When $z \to \infty$ annuitization should be preferred. Intuitively, this means either "infinite" wealth, in which case we force annuitization, or no annuity at all in which case we also mandate buying some.

The solution of the above equation provides the boundary between the annuitization and non-annuitization regions and can be solved for all $t < \tau$. This boundary can be used by the plan participant to develop an optimal strategy. Initially, the plan participant will make sure that they stay at the annuitization boundary. Namely, for a given wealth and income ratio and prevailing level of interest rates, if they are inside the annuitization region, they will compute the amount dG_t that is needed to bring them to the boundary. No action is required if they are outside the annuitization region since DIAs cannot be sold by the plan participant. The plan participant reevaluates their wealth-to-income ratio after purchase or any additional money deposits, DIA level already acquired, and repeats the process periodically.

2.4.2 Verification theorem

We will state a verification theorem, and will sketch the proof, but we will only show existence of pde solutions numerically, not analytically.

Upper Bound: Assume an increasing G, adapted to $\mathcal{F}_t$, a smooth function J satisfying (2.20), (2.21), and the terminal condition (2.23). Assume also that $J_r\sigma\sqrt{r_s}$ is square

integrable. Then

$$E[U(W_\tau, I_\tau, r_\tau) \mid \mathcal{F}_t] \leq J(t, W_t, T_t, r_t) \tag{2.36}$$

In particular, this holds for J as in (2.24) if the corresponding conditions hold for ϕ smooth. Specifically (2.30) and (2.32), or (2.31) and (2.33). The terminal condition for ϕ is found at (2.41)

Proof

Since G is increasing we have $dG_t \geq 0$. Anytime $\Delta G > 0$, the argument at the beginning of Section 2.4.1 shows that $\Delta J \leq 0$, which will work in our favour. So we may assume that G_t is continuous.

We integrate both sides of (2.19) from t to T and get the following:

$$J(\tau, w_\tau, I_\tau, r_\tau) - J(t, w_t, I_t, r_t) =$$
$$\int_t^\tau (J_s + J_w(r_s w_s + \nu) + J_r\alpha(\mu - r_s) + \frac{1}{2}J_{rr}\sigma^2 r_s)ds + \int_t^\tau [\frac{J_i}{(\tau - s)\tilde{a}_{(x+s)}} - J_w]dG_s + \int_t^\tau J_s\sigma\sqrt{r_s}dB_s$$

The first two integrals are less than or equal to zero by hypotheses, and we get the following:

$$J(\tau, w_\tau, I_\tau, r_\tau) \leq J(t, w_t, I_t, r_t) + \int_t^\tau J_r\sigma\sqrt{r_s}dB_s$$

Since $J_r\sigma\sqrt{r_s}$ is assumed square integrable we can take conditional expectation of both sides with respect to filtration $\mathcal{F}_t$.

$$E[J(\tau, w_\tau, I_\tau, r_\tau)|\mathcal{F}_t] \leq E[J(t, w_t, I_t, r_t)|\mathcal{F}_t] + E\left[\int_t^\tau J_r\sigma\sqrt{r_s}dB_s|\mathcal{F}_t\right]$$

On the RHS the first term is $\mathcal{F}_t$ measurable and the second is equal to zero. Using the terminal condition (2.23), we conclude:

$$E[U(w_\tau, I_\tau, r_\tau)|\mathcal{F}_t] \leq J(t, w_t, I_t, r_t)$$

Lower Bound: Now assume there is a smooth function $z = \Phi(t, r)$ and a smooth function

$\phi(t,z,r)$ satisfying the terminal condition (2.41). Assume that (2.34) and (2.30) or (2.31) hold for $z \leq \Phi(t,r)$, and that (2.35) and (2.32) or (2.33) hold for $z \geq \Phi(t,r)$. Assume also that $J_r\sigma\sqrt{r_s}$ is square integrable. Then this choice of J (and ϕ) is optimal, i.e. (2.17) holds.

Proof The argument just given shows $\geq$ in (2.17). So we must just show $\leq$.

First observe that if $Z_t = \frac{W_t}{I_t}$ then

$$dZ_t = Z_t[r_t\,dt - (1 + \frac{1}{(\tau-t)\tilde{a}_{(x+t)}})\frac{dG_t}{I_t}] \tag{2.37}$$

whenever G_t is continuous.

Assume to start with that $Z_0 \leq \Phi(0,r_0)$. Let $\bar{G}_t$ be a solution to the following Skorokhod problem (see Lions & Sznitman (1984) for background and solutions to Skorokhod problems)

- $d\bar{G}_t \geq 0$
- $\bar{G}_t$ is continuous
- Defining $dZ_t = Z_t\left[r_t\,dt - (1 + \frac{1}{(\tau-t)\tilde{a}_{(x=t)}})\,d\bar{G}_t\right]$ makes $Z_t \leq \Phi(t,r_t)\ \forall t \geq 0$
- $d\bar{G}_t = 0$ unless $Z_t = \Phi(t,r_t)$

Now set $dG_t = I_t\,d\bar{G}_t$. As before,

$$\begin{aligned} &J(\tau,W_\tau,I_\tau,r_\tau) - J(t,W_t,I_t,r_t) \\ &= \int_t^\tau (J_s + J_w r_s W_s + J_r\alpha(\mu - r_s) + \frac{1}{2}J_{rr}\sigma^2 r_s)\,ds + \int_t^\tau \left[\frac{J_i}{(T-s)\tilde{a}_{(x+s)}} - J_w\right] dG_s + \int_t^\tau J_r\sigma\sqrt{r_s}\,dB_s. \end{aligned} \tag{2.38}$$

Because $Z_t \leq \Phi(t,r_t)$, the first integral now $= 0$. Since $dG_t = 0$ unless $Z_t = \Phi(t,r_t)$, the second integral now $= 0$. And as before, the conditional expectation of the stochastic

integral = 0. Therefore by our terminal condition,

$$J(t, W_t, I_t, r_t) = E[U(\tau, W_\tau, I_\tau, r_\tau) \mid \mathcal{F}_t] \tag{2.39}$$

as required.

Finally, consider the case that $Z_0 > \Phi(0, r_0)$. In that case, we choose a ΔG_0 to make $Z_{0+} = \Phi(0, r_0)$ and then proceed as above. The only modification required is to observe that $\Delta J = 0$ for $t = 0$, by the discussion at the beginning of Section 2.4.1. QED

2.4.3 Solution methodology

We solve the problem by numerically solving (2.32) or (2.33), using an explicit finite difference scheme. We also assume a CRRA choice of U. Boundary conditions are required for large z and at $z = 0$. Likewise, a terminal condition is required at the terminal (retirement) time τ. In our case we compute annuity prices based on the outlined methodology and use them in the model. This allows us to experiment with pricing parameters and see the effect on the decision to annuitize. In general, annuity prices can be an exogenous input to our model, either taken from the annuity market quotes or via pricing under different regimes.

Boundary conditions:

In terms of terminal and boundary behaviour, we have the following. We start with the terminal condition, at $t = \tau$, that represents utility of income already accumulated plus new income acquired via immediate annuitization of terminal wealth.

$$J(\tau, w, i, r) = \frac{\left(i + \frac{W_\tau}{\bar{a}_{(x+\tau)}}\right)^{(1-\gamma)}}{(1-\gamma)} \tag{2.40}$$

After we apply scaling, we obtain the following:

$$\phi(\tau, z, r) = \left(1 + \frac{Z_\tau}{\bar{a}_{(x+\tau)}}\right)^{(1-\gamma)} \tag{2.41}$$

For $t \leq \tau$, we need only a boundary condition at the maximum value on the spatial grid, $z = Z$. In other words when the ratio of wealth and accumulated income is very large. This can occur if we have a large amount of wealth (in the numerator), or very small amount of accumulated income from DIA purchases (in the denominator). In any event, our assumption is that at that time it would be optimal to buy DIA, in adherence to (2.35). Since we need to compute both the continuation and annuitization equations and compare the two, we do so in the following two-step process.

First, from (2.35) we can compute the first derivative in z and substitute into our continuation PDE (2.34):

$$\phi_t + \alpha(\mu - r)\phi_r + \frac{1}{2}\sigma^2 r\phi_{rr} + (rz)\frac{(1-\gamma)}{(z +_{(\tau-t)} \tilde{a}_{(x+t)})} = 0 \tag{2.42}$$

Secondly, we compute value of annuitization from (2.35)

$$\phi = \phi_z \left(\frac{z +_{(\tau-t)} \tilde{a}_{(x+t)}}{(1-\gamma)}\right) \tag{2.43}$$

Lastly, similar to the interior points, we take a minimum or maximum of the above two equations, depending on the value of γ.

2.4.4 Parameter values

To generate our results we need CIR model parameter assumptions (α, μ, σ), mortality parameters (λ_0, m, b), and a measure of risk aversion (γ). Our base-case CIR parameters are loosely based on preliminary work and presentation from the authors of Dillschneider et al. (2020), with minor alterations to interest rate volatility to achieve initial computational

stability. Specifically, we use $\alpha = 10\%, \mu = 3.5\%$ and $\sigma = 2\%$. In the sensitivity analysis we adjust these parameters and compare results. It is also important to note that we should technically have two sets of CIR parameters. The CIR parameters presented here are the ones used in the optimization and are the real world (P-measure) parameters. Conversely, we should adjust these for the annuity pricing (under the Q-measure) by accounting for the market price of risk. For simplicity of exposition and without loss of generality we ignore this and set market price of risk parameter, say $\lambda = 0$. Our Gompertz-Makeham mortality model parameters are $\lambda_0 = 0.0$, $m = 89.335$, $b = 9.5$, as in Habib et al. (2020).

A variety of studies have estimated the value of γ. One of the earliest papers is the work by Friend & Blume (1975), which has withstood the test of time and provides an empirical justification for constant relative risk aversion. It estimates the value of γ to be between 1 and 2. Feldstein & Ranguelova (2001); Mitchel et al. (1999) in the economics literature have employed values of less than 3. Mankiw & Zeldas (1991); Blake & Burrows (2001); Campbell & Viceira (2002) suggest that risk aversion levels might be higher. On the other hand, to avoid the problem of picking a γ value, Browne et al. (2003) invert the Merton optimum to solve for γ. However, any formulaic approach requires that we have the client's complete financial balance sheet inclusive of financial and real assets. In this work, based on our observations, we choose a γ value of 3 as the baseline value. This allows us to compare and contrast with Habib et al. (2020).

2.5 Results and discussion

We look at DIA allocation decisions at the time of retirement (the final time to make a decision), five years prior to retirement, and ten years ahead of retirement. Other times between these points could be examined, but the movement of the annuitization boundary is fairly consistent and uniform. The graphs represent the wealth-to-income ratio on the vertical axis and the interest rate on the horizontal. An investor would find their ratio and the level of interest rates that is currently prevailing. If they find themselves in the yellow (annuitization) region, it means they are to purchase an annuity at that time.

Within the graph, this would mean moving down vertically until reaching the blue (non-annuitization) region. As they are annuitizing their savings, the wealth-to-income ratio is declining, therefore moving them towards the blue region. In contrast, if they are already in the blue (procrastination) region, no action is needed and it is optimal to wait until the situation changes closer to retirement.

Figure 2.1 illustrates an optimal solution for an investor at age 55 when the DIA is non refundable ($Q = 0$). There are two regions on a wealth-to-DIA ($\frac{W}{I}$) and interest rate (r) plane separated by an annuitization boundary, where annuitization is optimal inside the yellow zone, while it is optimal not to annuitize in the blue region.

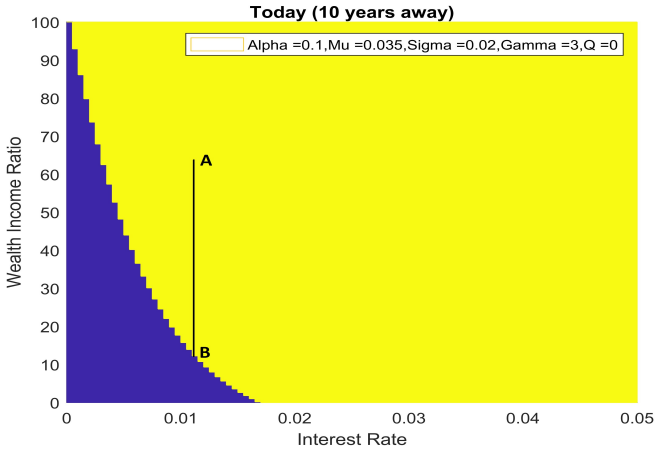


Figure 2.1: Baseline scenario - Annuity decision map at age 55, 10 years prior to retirement, with the annuitization line included. Interest rate levels (r) are displayed on the x-axis while wealth-to-income ratio levels (Z) are on the y-axis. For a given wealth-to-income ratio and interest rate level, if the investor is in the annuitization region (yellow) then they will annuitize a portion of their wealth to arrive at the boundary between annuitization and non-annuitization. On the other hand, if the investor is in the non-annuitization region (blue) then no action will be taken as annuities cannot be sold to bring the client back to the boundary. Baseline parameter values and $Q = 0$.

Figure 2.2 shows a case with the baseline parameter values that all other cases will be compared to. In this scenario, among other parameters, $Q = 0$, and it represents an

annuity where no refund is available in case of the annuitant's death prior to retirement. This value of Q, forms our baseline case that all others would be compared to, in addition to other baseline parameter values. This is observable in the model through the price of the annuity itself by making the annuity less expensive or, put differently, by offering more guaranteed income for the same dollars spent when compared to the annuity with refund capability. In general we observe that as we get closer to retirement, the annuitization region expands because less wealth compared to income (already acquired) is needed to entice annuitization.

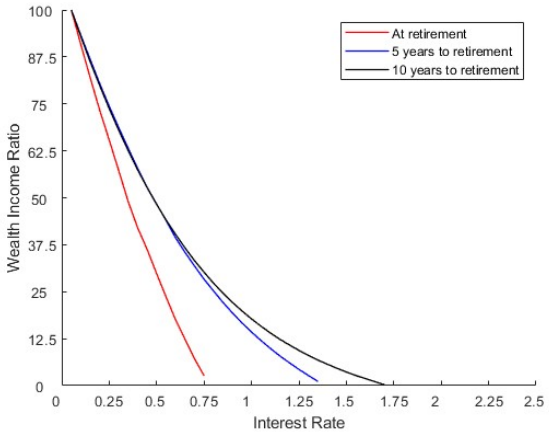


Figure 2.2: Baseline annuity decision map (at age 55, 60, and 65)

As the investor nears retirement, the annuitization boundary moves continuously. Using this model as a strategy to determine DIA purchases, the investor regularly reassesses the appeal of additional annuity purchases to remain at this boundary, making periodical additional purchases if warranted. If the circumstances are such that the wealth-income ratio changes and the investor now lies inside the boundary, they will stop making annuity purchases until they are on the other side of the boundary.

2.5.1 Impact of refundability

We look at two cases along the spectrum of refundability. We use Q to vary the degree of refundability from none ($Q = 0$) to full refund ($Q = 1$). We can use this to understand and calibrate where the true retail analog lies within our theoretical model, but we leave this as a future exercise. Along our scale, we observe that the more refundable a DIA is, the more likely the decision to postpone the annuity purchase when compared to the situation of a non-refundable (life-only) annuity. This is explained by the increase in mortality credits achieved through risk pooling i.e. some participants dying before they reap the full benefit of the annuity and thus transferring their benefit to the living annuitants. This is in line with findings from Habib et al. (2020), which concluded that a non-refundable design leads to more advance purchases. We observe that a higher value of the Q parameter – moving from zero in Figure 2.2, to a half in Figure 2.3, and then to one in Figure 2.4 – makes the blue region larger and promotes waiting to annuitize, possibly all the way to retirement. We see that, at a given interest rate, the ratio of wealth to income needs to be higher to make annuitization enticing. Similarly, we observe postponing purchase at slightly higher rates, at levels where annuitization is always favourable under a non-refundable design. Ceteris paribus, this meets our expectation as one has the opportunity to invest in the equivalent asset with refundable design whereas the non-refundable version has more to offer through mortality credits. However, the magnitude of the changes are small.

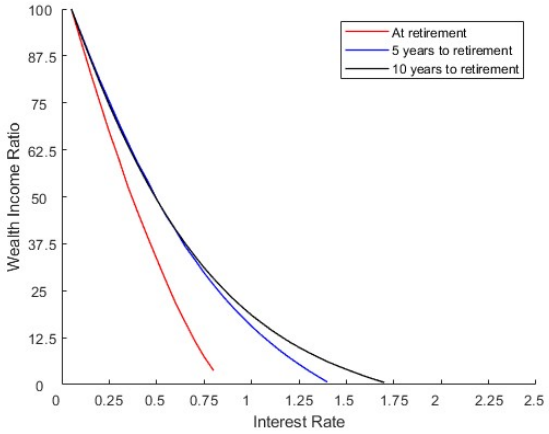


Figure 2.3: Annuity decision map (at age 55, 60, and 65) for $Q = 0.5$ and other baseline parameter values.

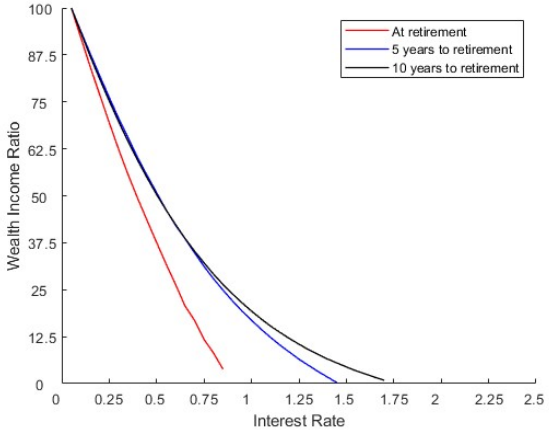


Figure 2.4: Annuity decision map (at age 55, 60, and 65) for $Q = 1$ and other baseline parameter values.

2.5.2 Impact of risk aversion

When it comes to the sensitivity of our results to the variation in level of risk aversion, as measured through the γ parameter, our results are consistent with those from Habib

et al. (2020), as well as both the existing literature and our intuition. The main conclusion is that the annuitization region expands as we increase the coefficient of risk aversion, as shown in Figure 2.6. We see the non-action region shrink at similar levels of interest rates and there are more levels at which it disappears completely. From the intuition and economics standpoint, we understand that a more risk-averse individual would likely lock in annuity pricing at both lower interest rate levels and lower wealth levels given that they are worried more about outliving their assets. On the other hand, a risk taker is willing to wait and see if the alternative would provide a better outcome. Figure 2.5 supports this conclusion when compared to either Figure 2.6 or the baseline case in Figure 2.2.

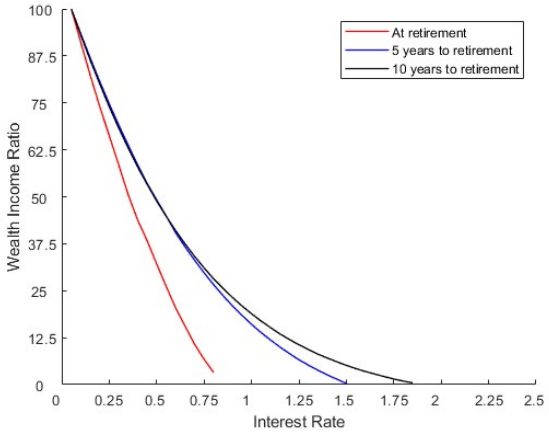


Figure 2.5: Annuity decision map (at age 55, 60, and 65) for $\gamma = 2$ and other baseline parameter values.

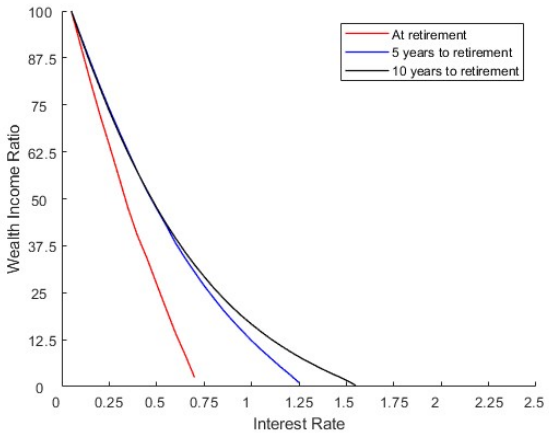


Figure 2.6: Annuity decision map (at age 55, 60, and 65) for $\gamma = 4$ and other baseline parameter values.

2.5.3 Impact of CIR model parameters

Speed of reversion α

Considering the speed at which interest rates converge towards their long-run mean, we observe that, at lower interest rates, one is more likely to defer purchases if the reversion speed is high. Alternatively, looking at Figure 2.7, we see that a high reversion speed requires a higher wealth-to-income ratio to entice annuity purchase. This is natural, since a higher reversion speed returns interest rates to a higher interest rate level sooner, therefore yielding more income for the same premium. The opposite is true if the reversion speed is low, as shown in Figure 2.8. There we see that, regardless of how low the current level of interest rate is, early anuitization is preferred. This is likely due to increased mortality credits that come with earlier annuitization, since the value of the annuity is derived more from mortality credits than interest rates when the current rate is low.

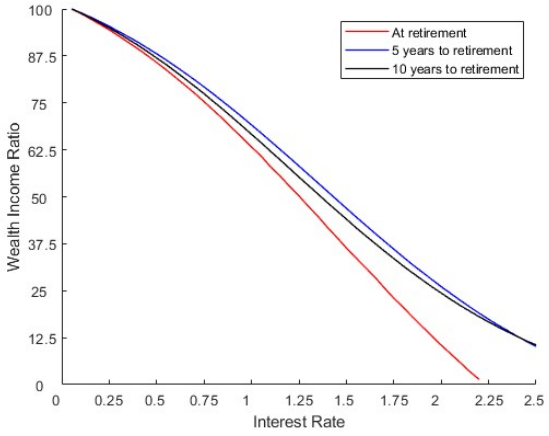


Figure 2.7: Annuity decision map (at age 55, 60, and 65) for $\alpha = 50\%$ and other baseline parameter values.

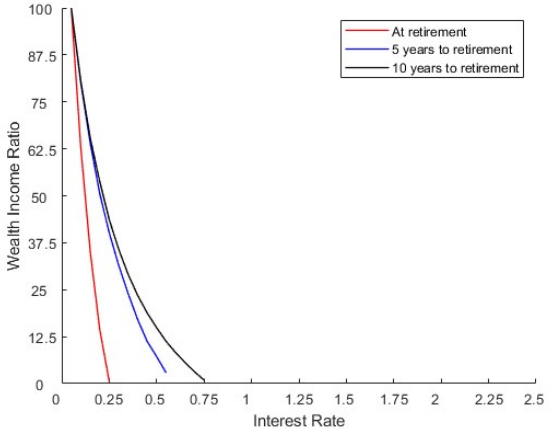


Figure 2.8: Annuity decision map (at age 55, 60, and 65) $\alpha = 5\%$ and other baseline parameter values.

Long-term average interest rate μ

Considering the long-term average interest rate, we see that the annuitization region moves along with it. At the higher levels, depicted in Figure 2.9 the no-action region applies to a

wider set of interest rates and a wider spectrum of wealth levels. We delay annuitization in the expectation that low interest rates will rise towards the long term mean and provide a better income to premium ratio then. Contrary to that, as seen in Figure 2.10, with lower long term mean we see that annuitization is preferred in most cases, mainly due to mortality credits available compared to savings account. It is evident that the changes in the annuitization boundary do not move in equal proportion based on the same amount of increase and decrease.

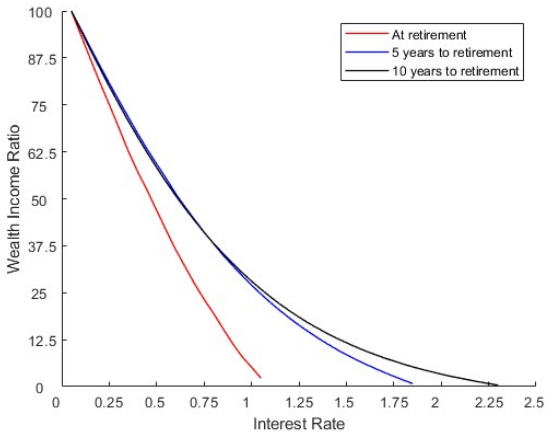


Figure 2.9: Annuity decision map (at age 55, 60, and 65) for $\mu = 4.5\%$ and other baseline parameter values.

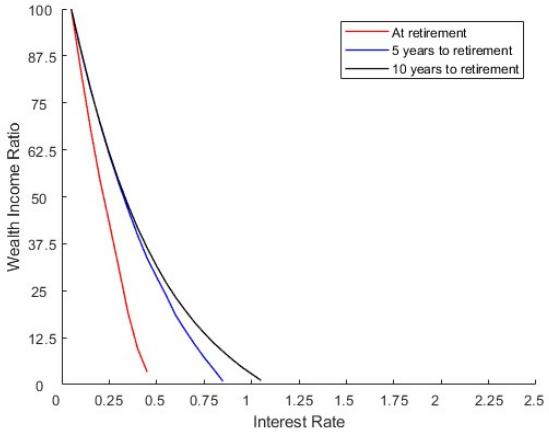


Figure 2.10: Annuity decision map (at age 55, 60, and 65) for $\mu = 2.5\%$ and other baseline parameter values.

Interest rate volatility σ

In regard to the volatility of interest rates, we conclude that the higher the volatility the less appealing annuitization is at lower interest rates. This is shown in Figure 2.11 and explained by the better potential for higher income once interest rates swing up, which in turn provides more income dollars for the same premium. In the case of lower volatility, Figure 2.12, the chance of movement to higher rates, ceteris paribus, is low and, indeed, the preference to annuitize increases to include a wider range of interest rates, than in our base case.

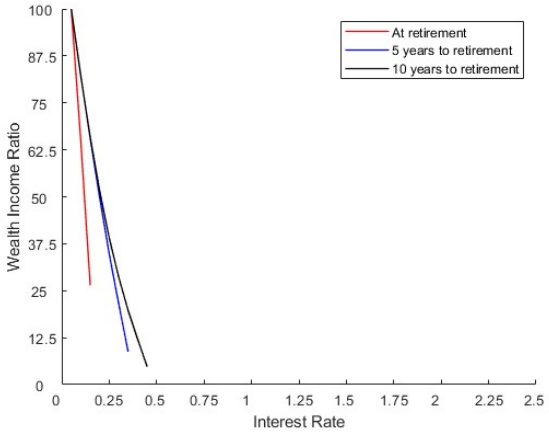


Figure 2.11: Annuity decision map (at age 55, 60, and 65) for $\sigma = 5\%$ and other baseline parameter values.

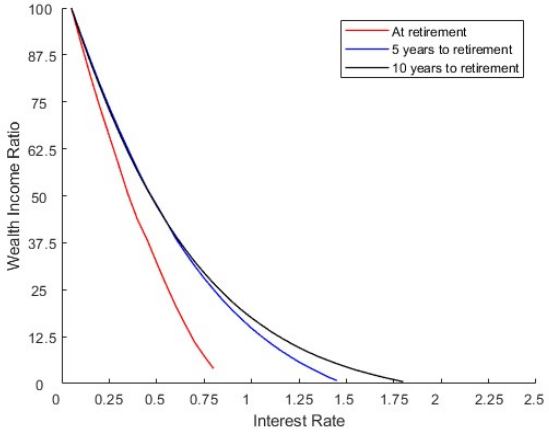


Figure 2.12: Annuity decision map (at age 55, 60, and 65) for $\sigma = 1\%$ and other baseline parameter values.

2.6 Alternative annuity pricing scheme

In this section we assume that insurers may adopt some form of simplified term structure modelling, rather than carrying out a full analysis of rate dynamics. We seek to under-

stand how much approximations affect optimal behaviour in the presence of interest rate uncertainty. For our alternative approach, we price the annuity in the following fashion. First, we price a 30-year zero coupon bond under the CIR model. We choose a 30-year duration to match the average duration of the annuity payout liability from the insurer's point of view. Then, using the inverted continuous compounding bond pricing formula, we extract a constant rate that would yield the same bond price today. Subsequently, we apply this rate within the constant rate annuity pricing framework using the same mortality assumptions. What we see is that one chooses overall to annuitize more under this alternative, fixed rate, pricing scheme. At higher-than-average interest rates, the decision is to annuitize. This is the case regardless of the annuity pricing scheme that we choose, whether or not we can incorporate the knowledge of returning to a long-term average interest rate. The annuitization decision is indistinguishable between the two models. At rates lower than the average, the decision under the alternative pricing scheme is driven only by mortality credits and how far one is from retirement. However, under the CIR model, we know that rates would be moving towards the higher long-term average rate and the decision would be to postpone. In other words, our decision to annuitize is made with knowledge in the CIR world, where we know that a very low rate today means overpricing of the annuity under the alternative scheme. This reflects the issuer ignoring mean reversion in interest rates, their future growth and subsequent drop in annuity prices; therefore the decision is to postpone the purchase of the annuity at this time. At higher rates, the insurer is assuming rates will stay there and therefore under price the annuity, making it very appealing for the investor to annuitize immediately.

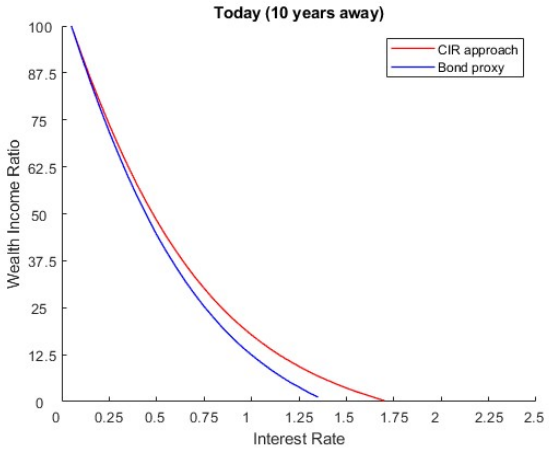


Figure 2.13: Annuity decision map at age 55, under both annuity pricing schemes, for baseline parameter values. CIR (left) vs constant rate (right).

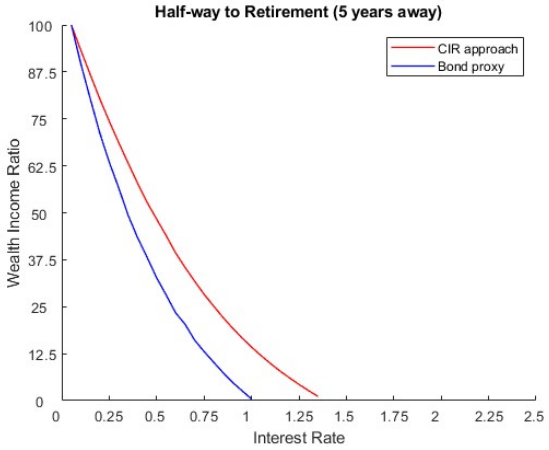


Figure 2.14: Annuity decision map at age 60, under both annuity pricing schemes, for baseline parameter values. CIR (left) vs constant rate (right).

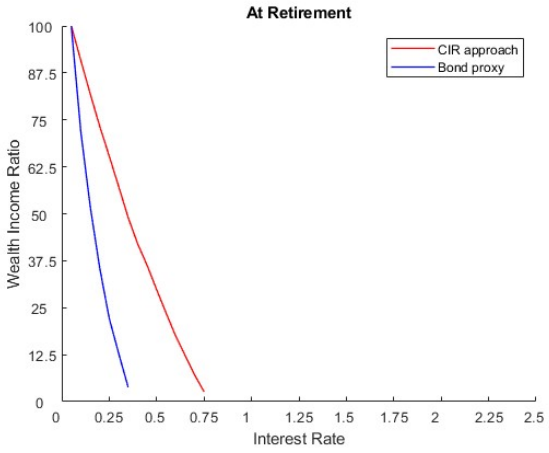


Figure 2.15: Annuity decision map at age 65, under both annuity pricing schemes, for baseline parameter values. CIR (left) vs constant rate (right).

This exercise shows an important feature of our model in which we see the impact of the interest rate model choice and a mismatch between buyer's and issuer's pricing of annuities. We explore how this can further contribute to the purchasing decision in pre-retirement. This is in addition to our previous conjectures from Habib et al. (2020), where we explain how the mismatch in mortality assumptions in annuity pricing and utility of income, combined with constant discount rate mismatch, can contribute to a change of the purchasing decision. Our findings here, focus only on the interest rate modelling and calibration assumptions, something that more practitioners can opine on and include in their planning process.

2.7 Conclusion

In this chapter we explore how one should optimally purchase DIAs within a retirement plan. We develop an optimal allocation strategy using the principles of stochastic control theory and dynamic programming. This lets us decide if it would be optimal to annuitize a portion of investments into a DIA in pre-retirement. Our goal is to understand the impact of interest rate uncertainty on the annuitization decision.

We consider a DIA purchase under both CIR pricing and the more ad hoc approach of CIR bond pricing and constant rate annuity pricing. While the DIA allocation region is smaller under the full CIR strategy, we observe that any potential mismatch between pricing can lead to changes in the optimal decision. Under constant rate pricing, we see the impact of the "knowledge" mismatch about future expected interest rates in addition to implementation complexity between a utility-maximizing agent (i.e. investor) and insurance company.

Our model implicitly assumes that mortality rates are predictable and that the hazard rate is deterministic. There is a growing body of research into stochastic mortality; since mortality and prevailing interest rates combine to drive DIA prices it would be interesting to incorporate these into our model. Another interesting question that arises is whether the optimal annuitization boundary actually equals zero when interest rates are very high, as our numerical experiments suggest, or whether is only asymptotic to zero. We leave these as questions for future research.

Finally, our model shows that periodic purchases of DIAs well before retirement can be attractive. Nevertheless, three factors examined here have a significant impact on the optimal decision: interest rates; wealth levels relative to annuity income already acquired; and the design of the annuity product itself.

Chapter 3

Reserving Argument for Cash Refund Income Annuities

3.1 Introduction

Lifetime income is one of the most important aspects of retirement planning. Annuities naturally play a pivotal role, given their ability to transfer longevity risk from an individual to the insurance company. Every time the risk transfers from one entity to another, a critical question arises whether the risk-taking party will be able to keep its side of the bargain. In the particular case of annuities, this comes down to whether the insurance company will be able to provide lifetime income to its annuitants. Insurance regulators make certain that insurers hold appropriate level of reserves against these liabilities and take measures when those are not adequate. Regulators apply a variety of methods and metrics to the calculation of regulatory capital. This work examines a possible analytical approach to evaluate an insurer's liability relative to cash refund income annuities.

In this chapter, following in the footsteps of the analysis by Milevsky & Salisbury (2022) we take a deeper look into regulatory reserving problem in the case of *Cash Refund Income Annuities*, CRIA for short. In particular, we look into one possible way that a regulator may step in when pricing is involved and whether under that scheme an insurer

that is well covered at annuity issue can rest assured that it will be able to prove that at any time thereafter it will have enough cash on hand to cover outstanding liabilities. Our main result here is that we prove the *reserving theorem* in great generality with respect to mortality assumptions for pricing. We further show that this regulatory approach to pricing always yields a unique price and then explore whether that price is feasibly covered by the insurance loading that the insurance company has to introduce in order to adhere to the regulatory requirement.

3.1.1 Literature review

Cash Refund Income Annuities (CRIA) differ from traditional lifetime income annuities in that they offer the annuitant a refund of the original premium net of payments already received. Starting with Yaari (1965), later extended by Davidoff et al. (2005) and others, research shows that annuitization is optimal in the financial economic sense. The decision is also an emotional one and it is not a surprise that one take at solving the annuity puzzle (see for example Benartzi et al. (2011), Ramsay & Oguledo (2018) and references therein) may be to avoid loss aversion in the event of the annuitant's premature death. This fact alone can potentially explain that over 50% of annuities sold in the U.S. in 2021 were CRIA, estimated through annuity quoting data provided by CANNEX Financial Exchanges Ltd[1].

3.1.2 Agenda for the chapter

The rest of the chapter is structured as follows. In section 3.2 we set up the necessary nomenclature for our problem and provide definitions and relationships between different aspects of pricing and reserving mechanisms. In section 3.3, we formally state the reserving theorem and provide detailed analytical proof followed by the price existence and uniqueness argument. Section 3.4 covers numerical examples for the particular choice of mortality model and as a support for the intuition in section 3.3. Lastly, in section 3.5,

[1]CANNEX is data and analytics firm that provides prices and research about retirement products in North America and until December 2022 my employer.

we summarize our findings and document a few avenues that we uncovered but did not explore in this project.

3.2 Actuarial analysis

3.2.1 Notation and terminology

In this section we try to closely follow the notation presented in Milevsky & Salisbury (2022) which indeed is closely tied with standard actuarial notations for annuity pricing.

- $a(x, r)$ - price of a single premium Life Only Income Annuity (LOIA) at age x under interest rate r paying \$1 in continuous time and without insurance loading. In the absence of loading, this represents an actuarial present value, one discounted by both interest rates and mortality, of the lifetime cash-flow described above.
- $a^*(x, r)$ - price of a single premium Life with Cash Refund Income Annuity (CRIA) at age x under interest rate r, paying \$1 of income in continuous time with the addition of a guaranteed cash-refund at the time of death of the annuitant. Since there is no loading, this also represents the actuarial present value of such arrangement. As expected, $a^*(x, r) > a(x, r)$ due to reduction in mortality credits.
- $\pi \geq 0$ stands for the insurance loading or mark-up from *value* to *price* that the insurance company charges to cover its profits, expenses and in our particular case meet the regulatory reserving requirement at issue. It is worth a mention that in case of LOIA loading applies just as a multiple of the actuarial present value of the annuity, while in the CRIA case it comes in a more complicated fashion into a recursive pricing formula.
- Lastly, T_x represents remaining random lifetime whose density and distribution functions are defined as follows:
  - $f_x(t)$

– $F_x(t) = Pr[T_x \leq t] = 1 - ({}_tp_x)$

For our setup we differentiate between:

– Insurance company pricing mortality - $f^0_x(t)$, with ${}_tp^0_x$ and λ^0_{x+t}

– Observed real mortality- $f_x(t)$, with ${}_tp_x$ and λ_{x+t}

– Regulator mandated insurance mortality- $f'_x(t)$, with ${}_tp'_x$ and λ'_{x+t}

– Composite - $\hat{f}_x(s)$, with ${}_s\hat{p}_x$ and $\hat{\lambda}_{x+s}$ and $\check{f}_x(s)$, with ${}_s\check{p}_x$ and $\check{\lambda}_{x+s}$, which all depend also on variable t, which we suppress from the notation.

Here:

$$ {}_tp^0_x \geq {}_tp_x \geq {}_tp'_x \tag{3.1}$$

$${}_s\hat{p}_x = \begin{cases} {}_sp_x & s \leq t \\ {}_tp_{x\,s-t}p'_{x+t} & s > t \end{cases} \tag{3.2}$$

$$\hat{\lambda}_{x+s} = \begin{cases} \lambda_{x+s} & s \leq t \\ \lambda'_{x+s} & s > t \end{cases} \tag{3.3}$$

$${}_s\check{p}_x = \begin{cases} {}_sp_x & s \leq t \\ {}_tp_{x\,s-t}p^0_{x+t} & s > t \end{cases} \tag{3.4}$$

$$\check{\lambda}_{x+s} = \begin{cases} \lambda_{x+s} & s \leq t \\ \lambda^0_{x+s} & s > t \end{cases} \tag{3.5}$$

For our numerical examples and regulatory annuity pricing formulae we use the Gompertz-Makeham law of mortality in which, $\lambda(x)$, the Instantaneous Force of Mortality (IFM), is defined as

$$\lambda(x) = \lambda_0 + \frac{1}{b}e^{(x-m)/b} \tag{3.6}$$

where x is the age, λ_0 captures the death rate attributable to accidents, $\underline{\text{m}}$ is the modal value of life[2], and b is the dispersion coefficient. The conditional survival

[2]strictly speaking, it is the mode only when $\lambda_0 = 0$

probability is then

$$ {}_tp_x = e^{-\lambda_0 t + \left(1 - e^{\frac{t}{b}}\right) e^{\frac{x-m}{b}}} \tag{3.7} $$

where once again x is the age of the individual and t is the number of survival years. This analytical representation of survival probabilities can be calibrated to the mortality tables that are readily available from the Society of Actuaries.

- Reg_t - actuarial present value of regulatory reserving requirement. It consists of actuarial present value of all outstanding payments for living annuitants for time t onward, plus all the insurance payments of CRIA, paid to those who die afterwards and before their premium was exhausted.

- Res_t - Actuarial present value of insurance company reserves for the remaining liability. It represents the difference of premiums received for annuities that are issued and payout experience under observed cohort mortality until time t for both income and insurance payments.

3.2.2 Life Only Income Annuity

We start by describing the pricing formula for LOIA. It represents an actuarial present value, a value of a continuous cash-flow of \$1 discounted for both interest rate and mortality.

$$ a(x,r) = \int_0^\infty e^{-rt} Pr[T_x \geq t] dt = \int_0^\infty e^{-rt} ({}_tp_x) dt \tag{3.8} $$

Table 3.1 summarizes prices at different ages under different interest rates. They represent the cost of \$1 of lifetime income for three different age groups (55, 65 and 75) and two different interest rate scenarios (2% and 4%).They will serve as a baseline comparison for the price of of a CRIA. Under the Gompertz law of mortality, these can also be obtained via a closed form solution using incomplete Gamma function described in Milevsky (2006).

$Age(x)$	$r = 2\%$	$r = 4\%$
55	22.1261	16.8200
65	17.0438	13.7336
75	11.9162	10.1723

Table 3.1: LOIA prices

3.2.3 Life with Cash Refund Income Annuity

The main topic of this research revolves around reserving for CRIAs and to proceed further we need to obtain a pricing equation and see how it compares both analytically and numerically with the already established LOIA pricing mechanism. For a CRIA without insurance company loading the pricing formula is defined by equation 3.9. It is worth noting that the price of a CRIA is obtained recursively, as its premium is a determinant of its cash-flows.

$$a^*(x,r) = \int_0^\infty e^{-rt}({}_tp_x)dt + \int_0^{a^*(x,r)} (a^*(x,r) - t)e^{-rt}({}_tp_x)\lambda_{(x+t)}dt \tag{3.9}$$

The right hand side of pricing formula in (3.9) is the actuarial present value of the sum of all the payments that annuitant receives while they are alive and the death benefit payable in case of an annuitants death prior to receiving full premium paid for the CRIA.

Since the first integral in (3.9) is equal to $a(x,r)$, the price of our ordinary LOIA defined in (3.8), we observe that second integral serves as the insurance payment tied to the CRIA. That insurance component, that declines with lifetime income payments paid to the annuitant, can be expressed as:

$$a^*(x,r) - a(x,r) = \int_0^{a^*(x,r)} (a^*(x,r) - t)e^{-rt}({}_tp_x)\lambda_{(x+t)}dt \tag{3.10}$$

In order to set up a CRIA pricing algorithm for numerical calculations and further application though out our analytical proofs, we borrow the following definition of CRIA from Milevsky & Salisbury (2022). Those authors provide intuition and a proof that this formulation is equivalent to the one in (3.9). They also show that for this unloaded

$Age(x)$	$r = 2\%$	$r = 4\%$
55	23.7957	17.4711
65	19.5447	14.9022
75	15.1947	11.9716

Table 3.2: CRIA prices

variation of CRIA, there exists unique price and that in the presence of loading there is a lower bound on interest rates for given mortality that makes CRIA pricing feasible.

$$\int_{a^*(x,r)}^{\infty} e^{-rt}({}_tp_x)dt = \int_0^{a^*(x,r)} e^{-rt} r(a^*(x,r) - t)({}_tp_x)dt \tag{3.11}$$

Using the bisection algorithm on (3.11) we obtain the unloaded CRIA prices for a range of ages and interest rates similar to those in section 3.2.2. We observe that CRIA prices in Table 3.2 are relatively higher than those for LOIA in Table 3.1. This is the cost of per dollar income, or put differently a higher premium for the same income is due to the mortality credits forgone and expressed through money back guarantee of CRIA. The spread between two prices is higher at low interest rates, as the present value of the insurance component grows.

3.3 Reserving theorem

3.3.1 Introduction

In this section we take a look at the reserving problem. Reserving is complicated, but a toy model is to assume that the regulator requires the issuer to have enough cash on hand to cover liabilities assuming a slightly worse interest rate and a slightly worse hazard rate than the true one, or at least the one that insurance company uses in their pricing. This would be hard to do for a finite number of contracts, so we opt to do this for an infinite number.

We state and prove reserving theorem in full generality, but here we provide two options that can lead to numerical experiments as well. If the true parameters for interest rate and mortality in the Gompetrz case are r and m, then we could require that the cash on hand at every t is enough to cover the post-t liabilities, for every r' with $|r' - r| < \delta_1$ and m' with $|m' - m| < \delta_2$. (Keeping b at the correct value). This is only one possibility for making hazard rate worst than the original. Another one, that we will explore as well is using models where the difference is in the existence of the external force of mortality, namely Gompertz vs Gompterz-Makeham.

Presumably the worst case is $r' = r - \delta_1$ and $m' = m - \delta_2$, or the introduction of a Makeham constant in the model. To meet the reserving at time $t = 0$, we would just use the price for r' and m'. But we should check that if the condition is met at $t = 0$, then it also holds at later times. A slightly more complicated version would be that the regulator could also require that we use worst-case choices of r' and m' separately for the annuity and insurance components of the CRIA. These would be different values, so we would need to show that a price exists, with this ad hoc choice of mortality parameters. And that being compliant at time $t = 0$ implies compliance at later times. And what conditions on r, m, δ_1, and δ_2 ensure that such a price exists.

Note that this would be an unloaded price. In the presence of loading, the price would have to make the loading viable (with the real r and m), but also to at least cover the unloaded costs with the varied parameters. This might mean that the designated loading amount might not all be collected at time 0, but later, i.e. when the regulator allows it.

3.3.2 Setup

To set the problem up we define two quantities that we are interested in comparing.

First, the present value of reserves held by the insurance company at any time t. This amounts to premiums collected for annuities that have been issued minus all annuity payments paid out to living annuitants in addition to all the refunds issues for those who died up until time t. This is formalized in the equation 3.12 below.

$$Res_t = a^*(x,r) - \left[\int_0^t e^{-rs}{}_sp_x ds + \int_0^{t\wedge a^*(x,r)} e^{-rs}(a^*(x,r)-s)f_x(s)ds\right] \tag{3.12}$$

Second, we define the present value of the regulatory reserve requirement. This consists of all the future payments outstanding to both living annuitants and those who die before receiving their full premium back. Formally, we get equation 3.13.

$$Reg_t = \int_t^{\infty} e^{-rs}{}_s\check{p}_x ds + \int_{t\wedge a^*(x,r)}^{a^*(x,r)} e^{-rs}(a^*(x,r)-s)\hat{f}_x(s)ds \tag{3.13}$$

3.3.3 Statement of the Reserving Theorem

Assume the actuarial nomenclature described in section 3.2.1. We will prove that if $Res_t \geq Reg_t$ at $t = 0$, it will be true that $Res_t \geq Reg_t \forall t$. This means that if the annuity issuer uses the larger of it's and regulator's price for the annuity to collect a premium at issue, the cash on hand will be sufficient to cover the liability going forward as long as the mortality assumptions are in line with equation 3.1.

Regulator price - Existence and Uniqueness

Before we move on to prove our main result, we obtain a price using the regulatory required mortality assumption. Two prices that we define are the insurance company price $(a^*(x,r)_{Ins})$ using insurance company mortality assumption for both the annuity and insurance piece and the regulator price $(a^*(x,r)_{Reg})$ using the same mortality assumption as the insurance company for the annuity piece and its own for the insurance piece of CRIA.

$$a^*(x,r)_{Ins} = \int_0^{\infty} e^{-rs}{}_sp_x^0 ds + \int_0^{a^*(x,r)} (a^*(x,r)-s)e^{-rs}{}_sp_x^0\lambda_{(s+t)}^0 ds \tag{3.14}$$

$$a^*(x,r)_{Reg} = \int_0^{\infty} e^{-rs}{}_sp_x^0 ds + \int_0^{a^*(x,r)} (a^*(x,r)-s)e^{-rs}{}_sp_x'\lambda_{(x+s)}' ds \tag{3.15}$$

In Milevsky & Salisbury (2022), the authors show that when the same mortality parameters are used for both the annuity and insurance piece, a price always exists and further more is unique. They also go to show that under insurance loading, there is a minimum interest rate that makes CRIA pricing feasible.

Intuitively, since the regulator is requiring a more conservative mortality assumption for the insurance piece, it is expected for the price to be higher than the one the insurance company calculates. At first glance, this may look like loading the price, but the key here is that even though the price is higher it is fully reflected in the payout for the annuity, unlike loading where the money is "taken off the table" by the insurance company to fund expenses and profits. Ultimately, we also convert this increase in price to an actual "additional load" that the insurance company can charge for the annuity as a hedge for the regulatory requirement and convert it to profit after the book gets closed.

Regulator price - Existence and Uniqueness - Proof

Without loss of generality, we prove the existence and uniqueness under the Gompertz-Makeham model for analytical simplicity, but as long as the relationship between mortality functions is preserved as defined in section 3.2.1 we should be covered. More precisely, we assume that the insurance company uses a Gompertz model, one where where the human death rate increases exponentially with age, and that regulator insists on the Gompertz-Makeham model that in addition accounts for an age-independent external force of mortality. We defined these hazard rates and survival probabilities in section 3.2.1, but we write them out explicitly here for clarity and their relationship.

$$\lambda^0 = \frac{1}{b} e^{\frac{x-m}{b}} \tag{3.16}$$

$$\lambda' = l + \frac{1}{b} e^{\frac{x-m}{b}} \tag{3.17}$$

$${}_tp'_x = e^{-lt} {}_tp^0_x \tag{3.18}$$

Where l, m, b are external force of mortality, modal value and dispersion coefficient, respectively.

As mentioned in section 3.3.3 the price defined by equation 3.14 always exists. We want to show that same is true for the one defined by equation 3.15.

We start by expressing equation 3.15 in terms of the insurer's mortality assumptions.

$$a^*(x,r)_{Reg} = \int_0^\infty e^{-rs}{}_sp_x^0 ds + \int_0^{a^*(x,r)} (a^*(x,r) - s)e^{-rs}(\lambda^0_{(s+t)} + l)e^{-ls}{}_sp_x^0 ds \tag{3.19}$$

Integrating the second integral by parts and simplifying we arrive to the following expression:

$$\int_0^\infty e^{-rs}{}_sp_x^0 ds - \int_0^{a^*(x,r)} e^{-(r+l)s}{}_sp_x^0 ds = \int_0^{a^*(x,r)} re^{-(r+l)s}(a^*(x,r) - s)_sp_x^0 ds \tag{3.20}$$

This equation closely resembles our equation that we used for numerical calculations, except for the Makeham constant l that appears on both sides. If we set it to 0, we recover equation 3.11.

In order to prove the existence and uniqueness, we define both left and right hand side of 3.11 as functions of a, namely $f(a)$ and $g(a)$, respectively.

We first take a look at $f(a)$:

$$f(0) = \int_0^\infty e^{-rs}{}_sp_x^0 ds \tag{3.21}$$

This is just the price of a LOIA under the insurer's mortality ${}_tp_x^0$, and therefore a constant (C_1).

$$\lim_{a->\infty} f(a) = \lim_{a->\infty} \int_0^\infty e^{-rs}{}_sp_x^0 ds - \int_0^a e^{-(r+l)s}{}_sp_x^0 ds \tag{3.22}$$

Since l is a positive constant, in the limit difference of these two integrals at ∞ amounts to a constant (C_2). In addition $C_2 < C_1$. At this point we note that $f(a)$ is decreasing function starting at C_1 when $a = 0$, and going to C_2 when a goes to ∞.

Second, we examine $g(a)$ and note that $g(0) = 0$. We then compute the derivative in a.

$$\frac{\partial g}{\partial a} = r \int_0^a e^{-(r+l)s} {}_sp_x^0 ds \tag{3.23}$$

Since, the derivative is increasing in a, we note that $g(a)$ is convex and increasing.Therefore, f and g cross at a unique point a, which defines the price a^*.

At this point and going forward, we write $a^*(x, r)$ without the arguments as a^*, for simplicity in notation.

3.3.4 Proof of 3.3.3

Let a^* be the price charged. In practice, this will be teh greater of the two prices defined in (3.14) and (3.15), but we will not assume that explicitly. We want to show that equation 3.12 is always greater than or equal to equation 3.13. To do this rewrite inequality as follows:

$$a^* \geq \int_0^t e^{-rs} {}_sp_x ds + \int_t^\infty e^{-rs} {}_s\check{p}_x ds + \\ + \int_0^{t \wedge a^*} e^{-rs}(a^* - s) f_x(s) ds + \int_{t \wedge a^*}^{a^*} e^{-rs}(a^* - s) \hat{f}_x(s) ds \tag{3.24}$$

Intuitively, this means that the regulator agrees with the insurer on the mortality assumptions used in the annuity portion pricing , but has it's own view of how the embedded insurance piece should be valued.

Given the survival probability relationship described by equation 3.1, we will show the

following $\forall t$:

$$\int_0^\infty e^{-rs}{}_sp_x^0 ds + \int_0^{a^*}(a^*-s)e^{-rs}{}_sp_x'\lambda'_{(x+s)}ds \geq \int_0^t e^{-rs}{}_sp_x ds + \int_t^\infty e^{-rs}{}_s\check{p}_x ds + \int_0^{t\wedge a^*} e^{-rs}(a^*-s)f_x(s)ds + \int_{t\wedge a^*}^{a^*} e^{-rs}(a^*-s)\hat{f}_x(s)ds \tag{3.25}$$

We make the argument in two steps. First we compare the annuity portions in 3.25, followed by the comparison of the insurance part on both sides.

For the annuity portion comparison we want to show the following:

$$\int_0^\infty e^{-rs}{}_sp_x^0 ds \geq \int_0^t e^{-rs}{}_sp_x ds + \int_t^\infty e^{-rs}{}_s\check{p}_x ds \tag{3.26}$$

Which can be rewritten as:

$$\int_0^t e^{-rs}{}_sp_x^0 ds + \int_t^\infty e^{-rs}{}_sp_x^0 ds \geq \int_0^t e^{-rs}{}_sp_x ds + \int_t^\infty e^{-rs}{}_tp_{xs-t}p_{x+t}^0 ds \tag{3.27}$$

Comparing alike integrals we have the following:

$${}_sp_x^0 \geq {}_sp_x \Rightarrow \int_0^t e^{-rs}{}_sp_x^0 ds \geq \int_0^t e^{-rs}{}_sp_x ds \tag{3.28}$$

$${}_sp_x^0 \geq {}_sp_x \Rightarrow \int_t^\infty e^{-rs}{}_sp_x^0 ds \geq \int_t^\infty e^{-rs}{}_sp_x ds \geq \int_t^\infty e^{-rs}{}_tp_{xs-t}p'_{x+t} ds \tag{3.29}$$

Therefore, we have proved the statement in the equation 3.26.

Next, we look into the insurance portion of the pricing equation 3.25. We want to show the following:

$$\int_0^{a^*}(a^*-s)e^{-rs}{}_sp_x'\lambda'_{(x+s)}ds \geq \int_0^{t\wedge a^*} e^{-rs}(a^*-s)f_x(s)ds + \int_{t\wedge a^*}^{a^*} e^{-rs}(a^*-s)\hat{f}_x(s)ds \tag{3.30}$$

By the definition of $\hat{f}_x(t)$ we can rewrite the RHS of equation 3.30 as follows:

$$\int_0^{t\wedge a^*} e^{-rs}(a^*-s)f_x(s)ds + \int_t^{a^*} e^{-rs}(a^*-s)\hat{f}_x(s)ds = \int_0^{a^*} e^{-rs}(a^*-s)\hat{f}_x(s)ds \tag{3.31}$$

Integrating by part on both sides, we get the following:

$$a^* + \int_0^{a^*} e^{-rs}(rs - a^*r - 1)_s p'_x ds \geq a^* + \int_0^{a^*} e^{-rs}(rs - a^*r - 1)_s \hat{p}_x ds \tag{3.32}$$

Eliminating a^* on both sides and multiplying by negative one we get the following:

$$\int_0^{a^*} e^{-rs}(ra^* - rs + 1)_s p'_x ds \leq \int_0^{a^*} e^{-rs}(ra^* - rs + 1)_s \hat{p}_x ds \tag{3.33}$$

One additional thing before we can draw our conclusion is to show that ${}_s\hat{p}_x \geq {}_sp'_x \forall s$.For $s < t$

$${}_s\hat{p}_x = {}_sp_x \geq {}_sp'_x$$

Similarly, for $s > t$

$${}_s\hat{p}_x = {}_tp_x {}_{s-t}p'_x \geq {}_sp'_x {}_{s-t}p'_x = {}_sp'_x$$

Therefore,

$${}_s\hat{p}_x \geq {}_sp'_x \forall s \tag{3.34}$$

Since the integrads in (3.33) are positive and with the survival probability relationship established in (3.34), we have the following:

$${}_s\hat{p}_x \geq {}_sp'_x \Rightarrow \int_0^{a^*} e^{-rs}(ra^* - rs + 1)_s \hat{p}_x ds \geq \int_0^{a^*} e^{-rs}(ra^* - rs + 1)_s p'_x ds \tag{3.35}$$

Therefore, we have proved the statement in (3.30). This completes the proof of statement in (3.25). To finish the proof of our theorem, assume that $Res_0 \geq Reg_0$. By definition, this means that a^* is greater than or equal to the left hand side of (3.25). Therefore it is greater than or equal to the right hand side of (3.25), which is exactly (3.24), i.e. what we we needed to prove.

3.4 Numerical results and discussion

In this section we round off with a numerical example that discusses feasibility of the proposed reserving strategy under the Gompertz-Makeham mortality framework.

As we have shown in section 3.3.3 the price exists and is indeed unique. Now the only question that remains is if this price can be charged by the insurance company and if the regulatory overhead can be charged as the insurance load on the price obtained by equation 3.14.

To verify this, we outline the process as such: first we find CRIA prices for both regulator and insurance company, noting that the regulatory requirement of more conservative mortality for the insurance portion of CRIA makes the prices higher; second, we search for the load on the insurer's price that would make it equal to the regulatory price under its mortality assumptions; lastly, we compare the load required to the loads discussed in Milevsky & Salisbury (2022) to verify that the prices are feasible. We provide a basic sensitivity analysis for the Makeham constant to show that there are limits on the price feasibility.

It is worth noting that we don't delve into what additional load the insurance company needs to account for in pricing for its cost and profits. Whether this "regulatory load" should be part of or in addition to the load, makes no difference for our exercise and we are leaving it as an open question for future consideration.

3.4.1 Pricing parameters

For pricing parameters we choose those outlined in Milevsky & Salisbury (2022), to enable comparison to the results therein and comment on price feasibility under the suggested regulatory load. For the Gompertz mortality model we chose $m = 90$ and $b = 10$. For the Makeham constant we provide results for two values, $\lambda = 0.001$ and $\lambda = 0.01$, the former representing a low instantaneous death rate and the latter one an order of magnitude higher. This can be representative of different demographics and factors such as

$\lambda = 0$	Interest rate		
$Age(x)$	1%	2%	4%
55	29.1138	23.7957	17.4711
65	23.5014	19.5447	14.9022
75	18.0181	15.1947	11.9716
85	12.9016	10.9999	8.9071

Table 3.3: Insurer's price of \$1 of lifetime income

$\lambda = 0$	Interest rate		
$Age(x)$	1%	2%	4%
55	3434.7982	4202.4423	5723.7285
65	4255.0724	5116.4712	6710.3965
75	5549.9804	6581.2377	8353.1281
85	7751.0010	9090.9830	11227.0431

Table 3.4: Insurer's payout per \$100k

employment hazards. We calculate annuity payouts for ages $55, 65, 75$ and 85, under three different interest rates $1\%, 2\%, 4\%$.

3.4.2 Insurer unloaded price

Prices and payout under insurer's mortality are summarized in Tables 3.3 and 3.4, respectively.

3.4.3 Regulator $\lambda_0 = 0.001$

Prices and payouts for our base case for a regulator defined Gompertz-Makeham model, with Makeham constant $\lambda_0 = 0.001$ and the respective implied loads are presented in tables 3.5,3.6 and 3.7. In this case, one with lower external hazard rate in our comparison, we observe that the additional load based on regulatory requirement is not imposing significant price loading for the insurer. Comparing to the results in Milevsky & Salisbury (2022)

$\lambda_0 = 0.001$	Interest rate		
$Age(x)$	1%	2%	4%
55	29.601395	24.06509966	17.59826392
65	23.857161	19.74331588	15.0006935
75	18.259965	15.33025503	12.04130799
85	13.050981	11.08357683	8.95126164

Table 3.5: Regulator's price of $1 of lifetime income

$\lambda_0 = 0.001$	Interest rate		
$Age(x)$	1%	2%	4%
55	$ 3,378.22	$ 4,155.40	$ 5,682.38
65	$ 4,191.61	$ 5,065.01	$ 6,666.36
75	$ 5,476.46	$ 6,523.05	$ 8,304.75
85	$ 7,662.26	$ 9,022.36	$ 11,171.61

Table 3.6: Regulator's annual payout per $100k

where loads range from $5-25\%$ this could be seen as small add to the loads that still make prices feasible under interest rates assumed.

3.4.4 Regulator $\lambda_0 = 0.01$

In the event that the regulator requires more conservative mortality parameters, reflected in our case through higher value of the Makeham constant, to account for riskier living conditions or employment hazards we observe that implied loads rise sharply. At younger

	Interest rate		
$Age(x)$	1%	2%	4%
55	1.10%	0.91%	0.66%
65	0.79%	0.70%	0.54%
75	0.51%	0.47%	0.40%
85	0.28%	0.28%	0.25%

Table 3.7: Implied insurance load $\lambda_0 = 0.001$

$\lambda_0 = 0.01$	Interest rate		
Age	1%	2%	4%
55	35.4565	26.8650	18.8359
65	28.0502	21.7903	15.9549
75	21.0383	16.7120	12.7134
85	14.7140	11.9246	9.3741

Table 3.8: Regulator's price of $1 of lifetime income

$\lambda_0 = 0.01$	Interest rate		
Age	1%	2%	4%
55	2820.3611	3722.3166	5309.0235
65	3565.0338	4589.1933	6267.6778
75	4753.2456	5983.7181	7865.6896
85	6796.2610	8386.0345	10667.7281

Table 3.9: Regulator's annual payout per $100k

ages and low interest rates we observe loads that are in the range where feasibility may come to question.

While regulation cannot make the unload product infeasible, this implies that there is not much room for any additional loading to be imposed and taken immediately out of the reserves, to cover upfront costs.

	Interest rate		
$Age(x)$	1%	2%	4%
55	11.62%	9.88%	7.04%
65	8.09%	7.33%	5.70%
75	5.05%	4.85%	4.12%
85	2.75%	2.77%	2.56%

Table 3.10: Implied insurance load $\lambda_0 = 0.01$

3.5 Conclusion

In this chapter we set out to show that pricing parameters imposed by regulators can indeed introduce some challenges for CRIA issuers, especially if the regulator resorts to requirement modification based on the environmental and employment hazards reflected in different mortality models. Although, we confirm that the CRIA price under "our" regulatory regime always exists, we don't go into details about how this may be implemented by the insurance company that is issuing an annuity. We prove price existence and uniqueness in the Gompertz case, but also provide much a stronger result where the hedging liability at issue time indeed covers any ongoing liability in the general mortality case. Our numerical results show how different views on external hazards affect necessary CRIA loading and their reflection on price feasibility. Potential directions for future research could be proving and setting out conditions of price existence under more general mortality models, as well as ways to meet regulatory requirement over time and balancing in reserving with profits and cost embedded in the insurance load.

Chapter 4

Large Scale Index Modeling for Indexed Annuity Comparison

4.1 Introduction

In the United States, fixed annuities hold an important place in the retirement portfolio. In addition to guaranteed fixed returns that are usually above and beyond other fixed income instruments, they also come with a full array of useful benefits that would not be easy, nor practical, to replicate otherwise e.g. tax deferral, income and death benefit provisions (see NAFA (2022)). In the low interest rate and prolonged market volatility environment, one particular sub-category of fixed annuities, the Fixed Indexed Annuity (FIA), provides two additional benefits in equity market participation and principal protection. The former, albeit sometimes very limited is enabled by the insurance carriers through interest crediting option strategies that are tied to a particular index, while the latter is the guarantee offered by the insurance company via their investments of the premiums in their general account. Selection and allocation to different interest crediting options available within FIAs have become important questions that many near-retirees need an answer to. The National Association of Insurance Commissioners (NAIC) in the United States prescribes (at least in case of FIAs) that illustrations and therefore comparisons be performed via back testing

NAIC (2022). Looking into the most recent, worst and best index performance scenarios, one can definitely understand the "movements" inside the annuity. By no means though can this determine which annuity construct (crediting term + strategy + index) will provide the desired results. Back tests can extend to cover more historical scenarios, but all of this comes down to what is usually referred to as "driving forward looking only in the rear view mirror".

An alternative to back testing would be generating annuity return forecasts based on individual indices and then comparing their point estimates (means, medians, standard deviations etc.). One issue that we identify is, while one can believe that the compared indices may have similar distributions, the annuity crediting strategies transform the returns so much that distributions are no longer comparable. This is easily overcome with side-by-side simulations of different indices and therefore annuity structures. That way we can still compare point estimates but also go further to draw entire return distributions and answer which annuity did better and how often. At this point we arrive at what we believe are the main contributions of this work: doing this comparison, consistently, for many annuities struck on many different indices and with realistic results. We propose the ways to more precisely model esoteric indices (in our case volatility controlled ones) without re-engineering index mechanisms in our simulations, but simply relying on known machine learning (ML) techniques to recreate them. To enable large-scale forecasting, we utilize long-term capital market assumptions that are established in the industry and provided by many institutions. We do this by fitting each of our index models with features from a common "core" economy consisting of major asset classes and their benchmark indices. We also show that, in many cases, this broad exposure works better when compared to the correct set of index features. Finally, we explore what models are best suited for the proper simulation of this common economy.

In order to make a selection or decide on the allocation within an FIA, it is important to understand both the interest crediting structure of the annuity as well as the underlying index features. In order to properly compare two or more annuities whose returns are linked to different underlying indices, the payout and index have to be modelled and forecasted in a consistent manner. Given that the interest crediting formula is essentially a deterministic

function of index returns, we look into ways to replicate index performance, pass obtained returns through annuities and ultimately allow for comparison and informed selection. We discovered there were two main challenges to this i) cross sectional index returns are difficult to model and ii) inter-temporal return dependence in financial markets interacts with index calculations in novel ways. The former problem is a known issue in financial modelling generally. The latter issue of inter-temporal dependence is the motivation for stochastic volatility models. That said, these considerations have not been studied in this context before. As we will demonstrate however they are of critical importance.

It is also important to emphasize what we are not going to address and where and how we chose to make certain assumptions. We will not explore nor elaborate on the merits and appropriateness of annuity crediting strategies. Despite ongoing criticism, with the famous example of Babbel et al. (2010), we will not model annuity parameters (caps, participation rates, spreads, etc.) alongside index simulations and will keep them constant for the duration of calculations.

The organization of this chapter is as follows. The chapter consists of two parts. In 4.2 we explore the cross sectional correlation issue. In section 4.2.1, we study some toy models of annuities and show how cross sectional return correlations among indices are important. In section 4.2.2, we discuss the kinds of indices linked to FIAs. While there are many kinds of indices involving various calculations, we discuss their relative challenges and choose to focus on an extremely popular and easily understood category in Risk Controlled (RC) or volatility managed indices. In section 4.2.3 we discuss our common economy method of side stepping the issue of direct cross correlations by using machine learning on a smaller subset of benchmark indices. Finally in section 4.2.4 we simulate single and multiple annuity rates using the aforementioned methods.

In section 4.3 we will draw on our analysis in the previous section to understand what ways our methods may fall short. As we are replicating the index methodology instead of modelling the index levels directly, we ask what features are important to account for in the underlying index components. In section 4.3.2 we discuss the nature of index return statistics generally and explore the inter-temporal dependence of returns. We then explore stochastic volatility models as a means of accounting for these inter-temporal features

and the relative advantages and disadvantages of well known models. In section 4.3.3 we explore the specific features of risk control models and hypothesise on the behavior we observed in section 4.2. Finally we bolster this hypothesis with results in section 4.3.3. We find that more sophisticated fractional or 'rough' stochastic volatility models are needed to properly account for the effects of the risk control mechanism. We conclude by discussing the implications of this and possible future avenues of research.

At this point, it is also important to bring to readers attention the fact that this chapter is a product of joint work with another PhD student, Andrew Fleck. It is extremely difficult to draw a precise line of how the contribution should be divided, as this was a project that spanned a long period of time and the results presented herein are just a subset of work necessary to get us to this point. That said, the objective of tackling a large-scale annuity comparison problem stemmed from my commercial research interests and nearly decade-long employment at Cannex Financial Exchanges Ltd, where Andrew was a MITACS [1] intern on two occasions, and where the collaboration that led to this project was established.

4.2 Annuity Forecasting at Scale

4.2.1 Motivation and Examples

We will discuss FIAs in more detail in section 4.2.2, but here we present the main FIA building blocks that we consider in our modeling. A growing number of FIAs are linked to volatility controlled or risk controlled (RC) indices which we will denote by I_t, in our example. Basic RC indices simply shift their exposure between an equity index, say the S&P 500 (SPX), and a less volatile fixed income index or even cash. We will refer to these "primary" indices as P_t and B_t, respectively. The return of the RC index is then passed through an annuity crediting formula and this return, which we'll call X_i, and is what the investor experiences in their FIA accumulation account.

[1] https://www.mitacs.ca/en/programs/accelerate

Consider for example, a 10-year FIA using a 55% participation strategy and linked to the S&P 500 Average Daily Risk Control 10% USD Price Return Index (SPXAV10P). In this case, P_t is the SPX and B_t would simply be a cash account. The index provider, in this case S & P, simply tracks index levels of P_t and B_t to construct index levels for I_t. Then finally the annuity carrier applies 55% of the annual returns of I_t for the duration of the FIA.

In summary:

$$\begin{aligned} I_t &= f(\{P_s\}_{s\leq t}, \{B_s\}_{s\leq t}) \\ X_{t_i} &= \max\left(0.55\frac{I_{t_i} - I_{t_i-\delta t}}{I_{t_i-\delta t}}, 0\right) \\ i &= 1, ..., 10 \\ \delta t &= 1 \text{ year} \end{aligned} \tag{4.1}$$

The exact nature of the mapping from the basic indices to the RC index $f(\cdot, \cdot)$ will be discussed in section 4.2.2. For now, it suffice to say that this constant reallocation from a risky asset class to a less risky one in theory keeps the volatility below a prescribed level.

The current state of the art[2] in forecasting FIA returns (i.e. the X_i's) is to model the underlying index returns(i.e. I_t) by a stand-alone Geometric Brownian Motion (GBM) that are then passed through the annuity crediting formula. Let us model the mean effective 10-year return of the annuity in the previous example and consider some problems arising from the results. This is achieved by computing terminal accumulation value at the ten year mark and calculating the annual rate of return that would yield it i.e. Y such that

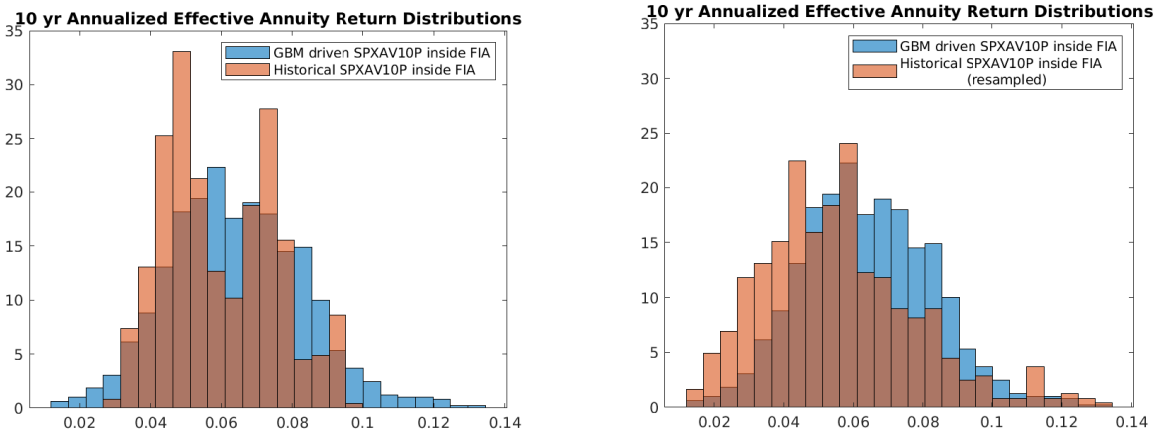
$$Y = \left(\prod_{i=1}^{10}(1 + X_i)\right)^{1/10} - 1$$

.

We present two sets of results (Figure 4.1a), one in which index I_t is modeled by a GBM

[2]Evidenced by Branislav Nikolic's experience while working at CANNEX. For more details see white papers produced there during his tenure.

and another where we rely on the historical data for the index, however both representing returns after these are passed through the annuity crediting formula (i.e. X_{t_i} in 4.1). More precisely, on the historical data we took ten year windows of monthly index return, starting in 1986 (when the SPXAV10P data set begins), and moved it forward in monthly increments to create our data set. The GBM was calibrated to the historical data set of monthly price levels for SPXAV10P obtained from CANNEX.



(a) (b)

Figure 4.1: Histogram of 10 year annualized annuity returns without and with resampling of returns (panels (a) and (b) resp).

	GBM Index	Real Index Returns (Actual)	Real Index Returns (Resampled)
Mean	0.064	0.060	0.056
Std	0.019	0.016	0.023
Skewness	0.286	0.322	0.758

Table 4.1: Annuity return comparison of a direct GBM model for the index with the actual and resampled returns.

On the other hand, compare our previous results to the case where we have taken the sample of historical index returns and re-sampled them with replacement before passing them through the annuity formula (Figure 4.1b). That is, we have removed any dependence

structure that exists through time on the index returns while still using the historical distribution of index returns. Keep in mind the caveat that this historical distribution may not be representative of a stationary distribution, if it exists. Even so, we can at least *visually* see the re-sampled results much closer align to the GBM results.

We may be tempted to think we can salvage the GBM model by simply altering the GBM parameters to give results more consistent with the real distribution. However, this ignores two issues:

Issue #1: Inter-temporal Dependence of Returns While shifting the parameters of the GBM may give us some reasonable values, it ignores any underlying mechanisms as to why such a shift is needed. While the underlying index or P_t (SPX in this case) itself appears to behave *roughly* like a GBM at certain time scales Cont (2001), it is not immediately clear if the behavior of the underlying index or the risk control mechanism is the source of any intertemporal dependence and deviations from the GBM model. Without understanding these issues it would not be prudent to manually alter the GBM parameters to fit historical results as there may be hidden risks in doing so. We will of course study this in much greater detail in section 4.3.

Issue #2: Comparing at Scale In order to answer how often one specific annuity product will outperform another point estimates are often not enough. Assume there are no issues with the GBM methodology for now and consider a new experiment involving annuities being struck on two benchmark indices (there is no RC mechanism under consideration here). Here we have simulated two indices, the S&P 500 (Annual average return (μ) of 0.08 and volatility (σ) of 0.16) and the Russell 2000 ($\mu = 0.11$ and $\sigma = 0.19$), and considered a few annuity strategies in various commonly occurring comparisons. Of course the the higher mean return of Russell translates into greater annuity returns most of the time. How often this is true is reliant on the dependence structure of the index returns (a historical estimate of the correlation between returns gives $\rho = 0.63$) i.e. will not be captured by point estimates alone. This is true in general and any consistent comparisons

will heavily rely on the dependence structure between index returns (see the results of table 4.2). .

	SPX with 7.75% cap		Russell with 25% par		
	Mean	Std	Mean	Std	P(SPX >Russ)
$\rho = -1$	4.6%	1.1%	3.5%	1.1%	69.4%
$\rho = 0$					74.9%
$\rho = 0.63$					82.6%
$\rho = 1$					92.5%

	SPX with 7.75% cap		Russell with 7.25% cap		
	Mean	Std	Mean	Std	P(SPX >Russ)
$\rho = -1$	4.6%	1.1%	4.7%	1.0%	48.3%
$\rho = 0$					48.2%
$\rho = 0.63$					47.9%
$\rho = 1$					44.1%

	SPX with 30% par		Russell with 25% par		
	Mean	Std	Mean	Std	P(SPX >Russ)
$\rho = -1$	3.3%	1.1%	3.5%	1.1%	45.3%
$\rho = 0$					44.6%
$\rho = 0.63$					41.7%
$\rho = 1$					0.0%

Table 4.2: Relative performance of 10 year annuities written on benchmark indices for different strategies and dependence structures.

Today there are approximately 100-150 indices linked to FIA products of relevance, many of which are only a few years old. What this means in practice is that many of them will have only few years of "live" data, compared to back-tested performance. It is also well known that estimating correlations among stock returns is problematic *especially* in the case of many assets (see Bun et al. (2017) for instance) and we believe that this translates to equity indices as well. Indeed this appears to help explain the many theoretical irregularities in real portfolios (Liu & Zeng (2017) or Huang, Zhang & Zhu (2017)).

Compare the following two scenarios i) A sample of 100 indices with 5 years of daily returns and ii) 15 indices with a 10-year daily return universe. Scenario i) is of course the

data we need to work with when studying FIA returns. Scenario ii) motivates our choice of a common economy in section 4.2.3. In both cases let us model the indices with jointly correlated GBMs. In Figures 4.2 and 4.3 we look at the statistics of *estimated* correlation matrices of returns in the two scenarios[3]. Specifically we focus on the distribution of the highest eigenvalue (λ_{max}) in Figure 4.2 and the mean absolute error of the entries in Figure 4.3. In scenario i) in blue clearly we see higher errors within the correlation matrix itself and indeed the sampling distribution for the largest eigenvalue is not representative of the actual value. In scenario ii) in red we see an improvement on both counts. Indeed it is not obvious that we can reliably estimate the dependence of the 100 indices to such a precision that a joint GBM model will meaningfully deliver index returns and therefore annuity comparisons. Our main contribution here is the claim that we can sidestep the issue of fitting 100 indices by moving to scenario ii). Many FIA indices are constructed from the same basic building blocks. For example the whole class of SPX RC indices is constructed from the SPX and fixed income indices alone (see Figure 4.5).

There is of course the issue of accurately replicating the (proprietary) risk control calculations. Our first attempts at doing so using the calculations provided by some carriers were comparative failures. Without access to the exact computations we found we could not perfectly replicate the index even in sample. That is, given the correct components and the formulas provided by the index provider our reconstructed index deviated from the true value. Starting in section 4.2.3 we instead use statistical learning (a.k.a machine learning) techniques to much greater success.

4.2.2 Fixed Index Annuities

When introduced in mid-nineties, in the United States, FIAs (then called Equity Indexed Annuities), were meant to provide owners with interest credit that would outpace banks' certificates of deposit (CDs). Fast forward a couple of decades and what we have is an array of index-linked products that are still there to provide equity index-based interest credits

[3]We chose a random positive definite matrix for the 'true' correlation matrix in either scenario, but the results are universal. See Bun et al. (2017) for instance.

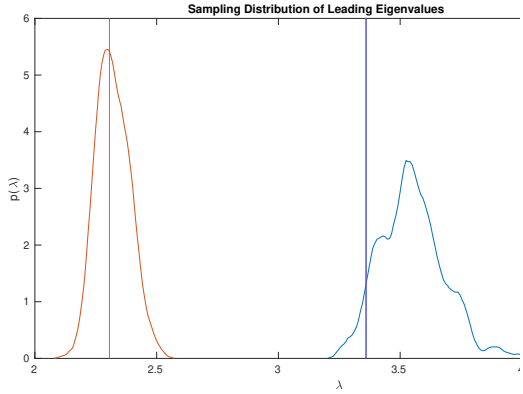


Figure 4.2: Sampling distribution of λ_{max}.

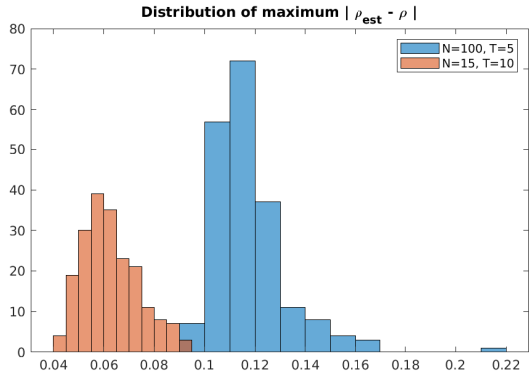


Figure 4.3: Sampling distribution of the MAE for $\rho_{i,j}$.

but also provide principal protection for near or new retirees. Through over a dozen distinct crediting methods in conjunction with the few hundred indices available, these products offer a unique value proposition that needs insights in several areas for the consumer to make an informed choice. Selecting one that is just right for a particular individual is an almost impossible task and there is now a sizable literature exploring product types, crediting strategies, index characteristics, etc (See Alexandrova et al. (2017) and references therein). If annuities are linked to the same index, the choice of crediting strategy boils down to the highest available rate and which crediting method is likely to produce favorable returns given conviction about how well (or not) the index will do in the future. However, when it comes down to understanding how different annuities struck on different indices compare, we have a few ways to go about this.

In terms of index crediting strategies, a popular choice is point-to-point participation with and without a cap. The participation strategy without the cap is the best example for why the indices have to be modeled consistently, as it provides minimal return transformation, while the capped version is the most popular from the sales perspective. The former strategy is what we focus on in this paper.

The point-to-point participation rate strategy measures the performance of the index from one point in time to another, usually starting the day you purchase the contract

until a year later and then repeating the measurement every year. The strategy then uses a "participation rate" that is a percentage of index gains that determines the amount of credited interest, so it increases in proportion to the index returns. The participation rate may be below 100% or above 100%, depending on the nature of the index itself. On the other hand, the point-to-point cap strategy similarly measures the performance of the index from one point in time to another; however the strategy then uses a "cap" that limits the amount of interest, no matter how well the index performs. Different levels of participation are technically available when the cap is used, but are usually set at 100%. As with all FIAs, the interest credited in any year cannot go below 0% no matter how much the index loses. Participation strategy returns mimic the underlying returns better but are usually dampened/enhanced, while a "cap" provides full exposure within the limit Nikolic et al. (2018).

While the ideal scenario would be to address a broad universe of indices available in the market, after initial experimentation we decided to focus on risk controlled indices, as the machinery we built would need different tweaks depending on the index category. Another important reason for this particular choice was the availability of the "live" index history. Once an index goes live, its returns are a true representation of how it works compared to what was intended by the index manufacturer. Absent that, what is available is the "back-tested" or "synthetic history" where the index provider illustrates the behaviour that would have been observed in the past based on the underlying index mechanics and movement of the constituents. This is where we would like to draw attention to a potential pitfall. The index behaviour observed in the "back-test" may be substantially different to the one after index goes live. In that case our model would learn from a not very representative sample. In other words if the index methodology was over fitted to past market conditions, if history doesn't repeat, it will diverge from the expected behaviour and therefore our forecast may be misleading.

One example that could explain some of the idiosyncrasies of more complex index structures is the Barclays Trailblazer Sectors 5 Index (BXIITBZ5). As a "smart beta" index its behaviour is motivated by Modern Portfolio Theory (MPT) and Arbitrage Pricing Theory (APT) and it tries to use momentum in the market to inform index constituent selection

and re-balancing among them (See Barclays Trailblazer Sectors 5 Index (Factsheet) (2022) and Kim & Francis (2013)) . Its ability to recognize and act differently in highly stressed period (e.g. during the opening months of COVID), versus more typical market cycles start causing problems for our machine learning approach. The complexity of the algorithm really means that each instance of return computation is sufficiently novel that an effective training set would need to be prohibitively large, while making certain events out of sample not generalizable. Conversely risk control calculations are consistent and simple enough that even a relatively small training set can be effective, even when we have events out of sample that are in hindsight seen as significant or even seismic.

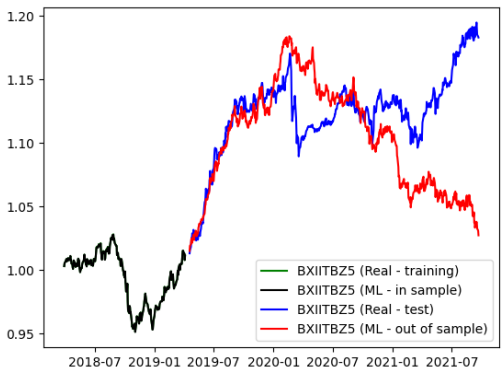
A detailed description of machine learning techniques used is outlined in section 4.2.3. At this point, for illustrative purposes only, we provide the summary of this exercise in Table 4.3. We see that low R^2 values are mainly unaffected with different training/test set ratios in the case of Neural Network and that our estimates are not up to par. This conclusion however, is even possible by visual inspection of figure 4.4.

		Train/Test Split		
		30% / 70%	50% / 50%	70% / 30%
NN Model R^2 (Out of Sample)	SPX10AV10P	0.455	0.664	0.730
	BXIITBZ5	-0.130	-0.109	-0.012

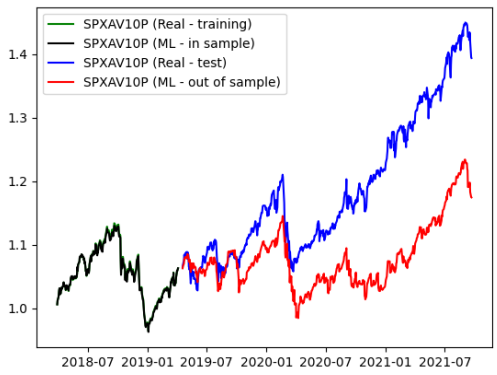
Table 4.3: Comparison of R^2 for out of sample predictors for different training/test data splits

Volatility Controlled Indices

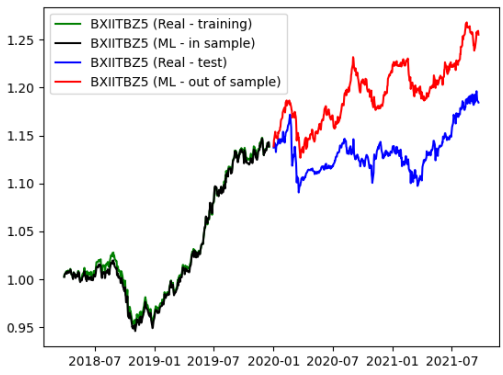
Managing volatility emerged as a noteworthy exercise and rose in prominence particularly within the insurance industry during and right after Financial Crisis in 2008. For securities investors, it was meant to reduce risk and severity of losses, but it enabled variable annuity issuers to offer secondary guarantees on their annuities because of lowered hedging costs, even in the face of the volatile, low interest rate and bond yields period that followed. Managed volatility expanded far beyond this use with its application to index-linked products.



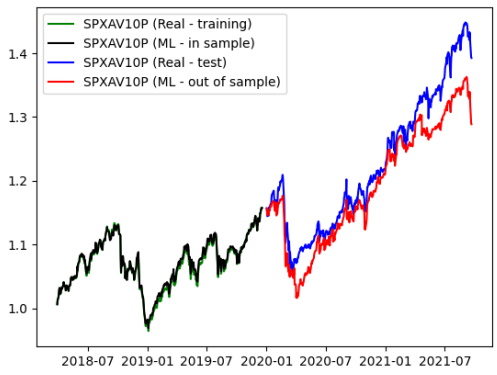
BXIITBZ5 30%/70%



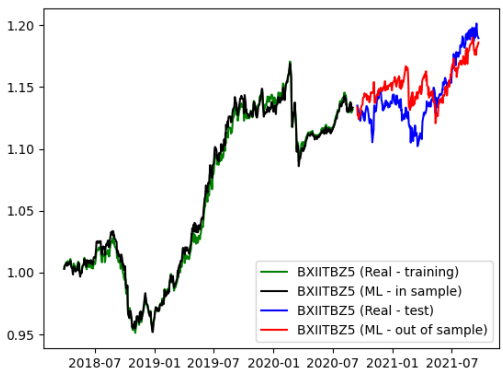
SPXAV10P 30%/70%



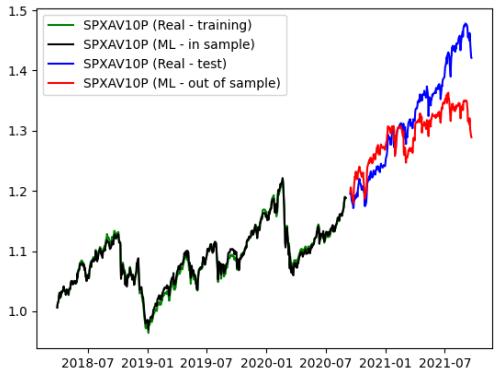
BXIITBZ5 50%/50%



SPXAV10P 50%/50%



BXIITBZ5 70%/30%



SPXAV10P 70%/30%

Figure 4.4: Comparison of NN Models in and out of sample for various train/test splits. Clearly SPXAV10P is more easily captured than BXIITBZ5, when there is less data to train on.

Prior to that, there was very little or no need for them for index-linked structured products, such as structured notes and FIAs. Structured product writers, namely banks and insurance companies, were able to use benchmark indices in their products and still offer competitive crediting rates compared to other fixed income vehicles. As most of these products use options struck on the underlying index to fashion the payouts, both the prevailing interest rate and underlying volatility play significant roles, not only in the option pricing but the variability of the price at any given time. As the low interest rates were there to stay, the option budgets required to provide credits in the FIAs, tied closely to bond yields, were simply not enough to offer attractive payouts using relatively volatile broad market indices. Issuers were able to address this by selecting or creating a less volatile underlying index that would result in more predictable and attractive rates. Lowering and stabilizing the volatility of the underlying index would ensure lower and stable option prices for annuity providers, and therefore better offerings for annuity purchasers. In general, volatility controlled indices operate on the rudimentary principle of moving exposure between high- and low-volatility assets. In its simplest form, this transition takes place between equities and bonds (see figure 4.5). If one is interested in the explanation on a slightly deeper level it instantly gets more complicated as different index providers use different methods to measure and achieve efficient and worthwhile "shifting", while preserving enough of the upside as well. These come in the shapes of volatility targets/levels, pre-determined rule sets and leverage conditions, or a combination thereof to provide an index that allows for attractive option pricing and therefore favourable annuity rates (higher participation, caps and lower spreads). There are a few ways that index providers tackled this problem over the years and with methodologies evolving, or being kept secret. Without extensive work put into reverse engineering it has been very difficult to forecast the performance of these indices, as well as any structured products that use them as the underlying.

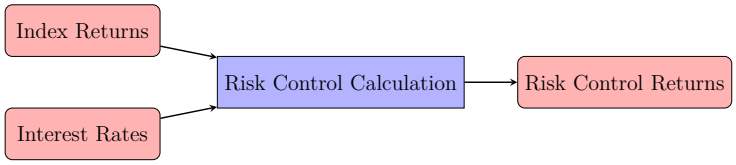


Figure 4.5: Schematic of a Risk Control Index

Consider one generic implementation of a RC index in the following[4]. We say the price level of some equity or commodity index at time t is modelled by a stochastic process P_t. Model the dynamics of P_t with the following SDE:

$$\frac{dP_t}{P_t} = \mu dt + \sigma_t dW_t$$

Where W_t is a standard Wiener process. Note the time dependence on volatility i.e. σ_t. We can model this using a local or stochastic volatility model. RC or volatility control indices take this reality of the stochastic nature of volatility into account and seek to provide the same equity premium exposure of another index while maintaining a volatility around a target value. They do this by shifting exposure between fixed income and equity components. Consider a generic bond (or interest rate index) modelled by another process B_t. For now let's assume $dB_t = rB_t dt$. We can construct a RC index I_t in the following generic way. Using time dependent weights $w_t^{(i)}$ the RC index is a sum of positions in equity and bonds:

$$I_t = w_t^{(1)} B_t + w_t^{(2)} P_t \tag{4.2}$$

$$dI_t = w_t^{(1)} r B_t dt + w_t^{(2)} P_t (\mu dt + \sigma_t dW_t) \tag{4.3}$$

$$\frac{dI_t}{I_t} = \underbrace{\left(\frac{w_t^{(1)} B_t}{w_t^{(1)} B_t + w_t^{(2)} P_t} \right)}_{1-\pi_t} r dt + \underbrace{\left(\frac{w_t^{(2)} P_t}{w_t^{(1)} B_t + w_t^{(2)} P_t} \right)}_{\pi_t} (\mu dt + \sigma_t dW_t) \tag{4.4}$$

$$\frac{dI_t}{I_t} = (r + \pi_t(\mu - r))dt + (\pi_t \sigma_t) dW_t \tag{4.5}$$

We can write this as $dI_t = \mu_t^* I_t dt + \sigma_t^* I_t dW$ where $\mu_t^* = r + \pi_t(\mu - r)$ and $\sigma_t^* = \pi_t \sigma_t$. Assuming the fixed income exposure is fairly close to the theoretical risk free rate the Sharpe ratio of the new index I_t is the same as the original[5]

[4]Inspired by the discussion in S&P Risk Control 2.0 Indices Methodology (2022), reformulated by us using Stochastic Calculus

[5]Considering that $\frac{dI_t}{I_t}$ are the returns (see section 4.3) then $\frac{\mu_t^* - r}{\sigma^*} = \frac{\mu - r}{\sigma_t}$

If we set $\overline{\sigma}$ as the "target volatility", then the natural choice is to set $\pi_t = \frac{\overline{\sigma}}{\hat{\sigma}_t}$ where $\hat{\sigma}_t$ is an estimate of the current volatility. This ensures that the RC index (in theory at least) maintains a constant target volatility of $\sigma_t^* = \frac{\overline{\sigma}}{\hat{\sigma}_t}\sigma_t \approx \overline{\sigma}$. There are of course caveats to this. There are different methodologies in use for calculating $\hat{\sigma}_t$ which won't perfectly capture σ_t. Standard choices include a rolling window of the standard deviation of index returns or an Exponentially Weighted Moving Average (EWMA) model. Many RC indices take a maximum or average of different estimates to smooth the estimates out as well. (Again see S&P Risk Control 2.0 Indices Methodology (2022) for details). Most important for our purposes however is the reality that in order for carriers to provide FIA exposure to an RC index they have to purchase a custom derivative over the counter. The writer of this derivative contract in turn needs to hedge their exposure with options on the original equity index. As a result the RC index cannot be too dissimilar from the original. The result is that when constructing the RC index there are leverage restrictions (denoted K_u) and maximum bond portions (denoted K_d) such that:

$$\pi_t = \max(\min(\hat{\sigma}_t, K_u), K_d) \tag{4.6}$$

Equation 4.6 above will be very relevant in our discussions in section 4.3. Namely we will see that the RC index inherits enough essential features of the original index that any modelling at scale will require some sophisticated stochastic volatility modelling.

Before moving onto the next section we would like to summarize some experiments and results we do not include here for brevity. For the rest of the paper we detail how we used ML to reconstruct RC indices from other components in the market. Given the simplicity of (4.5) and (4.6) it would be natural to ask why we couldn't directly reconstruct the index, no ML required. When we attempted to directly model RC indices using (4.5) and (4.6) and the methods provided to the public in the index fact sheets we found the results to be poor in comparison. In the ML case, as we will see, we provide a set of trailing returns so that the machine can presumably make the necessary volatility calculations. If we provided as an input $\hat{\sigma}_t$ in lieu of trailing days of returns, the ML results underperformed in comparison. This leads us to believe there are idiosyncrasies in the way $\hat{\sigma}_t$ is

estimated (e.g. data cleaning, weekend/seasonal effects etc...) that we do not have access to, therefore ML models are more favourable. Furthermore *in practice* it is simply a more efficient methodology being generic. Our technique does not require novel reproduction of the index methodology associated with each specific RC index.

4.2.3 Index Methodology

The CORE Hypothesis

The key motivating question that we wanted to answer was "what percentage of time one annuity struck on a particular index outperforms another?" To answer this we need to be able to produce the relevant index return distributions (back tested or forecasted) for different annuities and see the overlap or proportion of times one annuity return was higher than the other. In any event we need two pieces of machinery: an index simulator and an annuity return calculator. Given the straightforward and deterministic nature of the latter we can forgo a detailed explanation. In this section we focus on index simulations.

In section 4.2.1 we saw how we could do direct index modeling via stochastic processes with limited success. We also encounter scalability problems when we try to expand this to capture a larger set of annuities. As we said in section 4.2.1 it boils down to the lack of ability to produce a good correlation matrix for the large number of underlying indices. In this section we detail our solution via ML techniques.

A first approach is to "back out" the index methodology function using known index components. This quickly presents a problem, even if the set of index components may be smaller than the set of FIA indices we are interested in (e.g. The SPX is a known component of all SPX RC indices) this is still relying on the ability to simulate *all* known annuity index components with high dimensional correlation matrices. We use this first approach as a benchmark for our "core economy" hypothesis. We spent some time looking into the overarching set of indices that would encompass as much of the economy as possible (from now on referred to simply as CORE), while staying within a reasonable number due to the correlation issues stated in section 4.2.1. We hypothesize that our machine

learning algorithm should be able to "pick up" the index function equally well from the "core economy" set of indices as compared to calibrating it on the set of known index components. With this separation we open the door for future research on how should one optimally model and simulate the economy while using machine learning obtained functions for indices.

One question that remains is the proper choice of the CORE economy. We define the CORE as our set of indices representing asset classes that capture the broader market well (See Table 4.4 for individual indices). From now on, for simplicity, we will refer to the set of equity indices listed in Table 4.4 as CORE Equities, and the set of Fixed Income indices as CORE Bonds. The CORE data was provided by CANNEX and it consists of daily price level index time series from 2010/01/29 to 2021/09/21.

Traditionally, long term capital market assumptions available from investment banks, asset managers and insurance companies (JPM, PGIM, Callan etc.), are provided on an asset class basis, in terms of their returns, volatility and correlations. These are forward looking assumptions and can be used for forecasting in the models. Ideally, these would coincide with same asset classes available in the training data set. For that reason, we chose to represent each asset class with the index benchmark. The chosen index benchmark is preferably a good representation of the asset class but not necessarily an explicit component of the risk control index that we are trying to train on, to avoid over-fitting in the model. This would allow for calibration to the history via the index benchmark and the use of long term capital market assumptions for more consistent forecasting. Ultimately, this is just one possibility and the technology we developed for index calculations should in principle be able to use any other variations of the CORE economy representation (alternative set of indices), subject to size constraints that would make forecasting/projections feasible in regards to correlation issues that we outlined earlier.

Asset Class	Description	Name	Ticker
Equities	US Small Cap Equities	Russell 2000	RTY
	US Small-Mid Cap Equities	Russell 2500 Index	R2500
	US Mid Cap Balance Equities	MSCI US Mid Cap 450 Index	MZUSM
	US Large Cap Equities	Russell 1000	RIY
	US All Cap Equities	Russell 3000	RAY
	World Equities	MSCI World Index	MXWO
	International Equities	MSCI ACWI ex US Index	MXWDU
	Emerging Markets Equities	MSCI Emerging Markets IMI	MXEFIM
Fixed Income	High Quality US Fixed Income	Bloomberg Barclays US Aggregate Bond Index	LBUSTRUU
	Low Quality US Fixed Income	Bloomberg Barclays High Yield Very Liquid Bond Index	I33743US
	World Fixed Income	Bloomberg Barclays Global Aggregate Bond Index	LEGATRUU
	International Fixed Income	Morningstar Global ex-US Core Bond	MGXUSN
	Money Market	Bloomberg Barclays 1-3 Month US Treasury Bill Index	I00078US

Table 4.4: The CORE set of indices

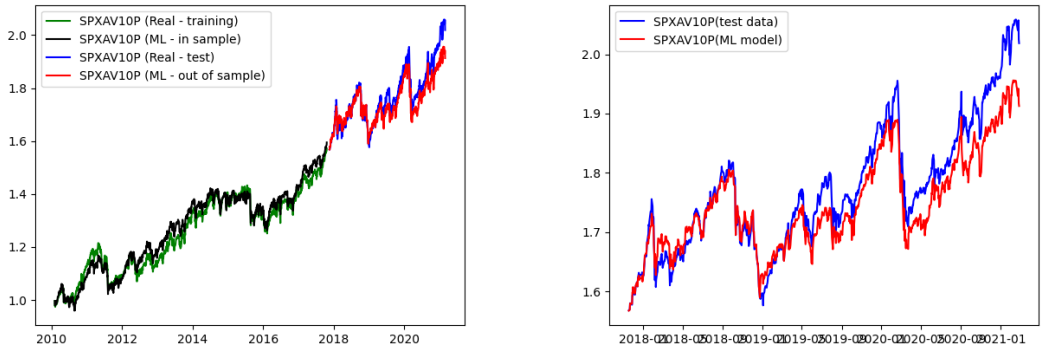
Equity Features	Fixed income Features	Training Set			GPR R^2	NN R^2
		# of Daily Returns	# of Indices	Total # of Returns		
S&P	DGS10	8875	2	17750	0.884	0.878
	I00078US	2807	2	5614	0.807	0.82
	LBUSTRUU	2807	2	5614	0.804	0.816
	Core Bonds	744	6	4464	0.756	0.617
Core Equity	DGS10	3055	9	27495	0.92	0.902
	I00078US	3055	9	27495	0.919	0.898
	LBUSTRUU	3055	9	27495	0.92	0.896
	Core Bonds	918	13	11934	0.907	0.809

Table 4.5: Summary of results, Neural Network with 10 layers, with 10 trailing days as inputs for SPXAV10P. R^2 is calculated *out of sample* (30% of the total data set allocated for training).

Machine Learning Techniques

In examples presented in this section we use the CORE dataset from section 4.2.3 in addition to the S&P itself and the yield on U.S. Treasury Securities at 10-Year Constant Maturity (DSG10). These later additions will serve as the true ingredients for some RC indices as a point of comparison. For our features or predictors we used exclusively daily returns (as this is the true input during index construction) and included 10 trailing days of index returns to allow for any algorithm to compute current volatility levels. Our response variables were the returns for the index of interest.

We started our experimentation with a simple ordinary regression model over index

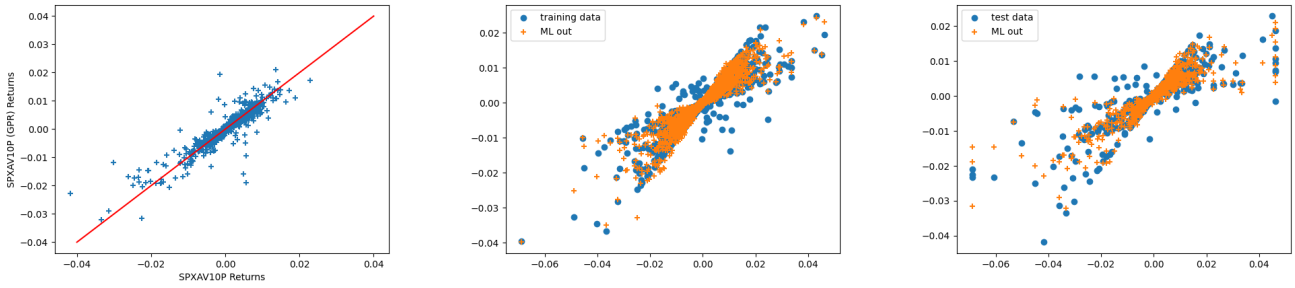


(a) (b)

Figure 4.6: SPXAV10P results with 20 trailing days where (a) is all data and (b) we have zoomed in on the out of sample performance.

returns and unsurprisingly found this wasn't a worthwhile avenue. Looking at equation (4.6) in section 4.2.2 we can see that index exposure will be dynamic and non-linear. We did however discover something that we carried over into more complex ML techniques. Including a set of lagged or trailing returns for the underlying indices produced better results than incorporating traditional volatility calculations in the case of RC indices. Clearly there were idiosyncrasies in the real volatility calculation that were hard to capture with the products stated volatility methodology (e.g EWMA or rolling window). It is unlikely that the provided index documentation is incorrect (albeit possibly vague) rather we suspect our inability to precisely reproduce the index is based on our interpretation of the rules and our dataset.

The natural starting point after a regression model when dealing with a "black box" function like an index calculation was a Neural Network (NN) of the kind repopularized by Hinton (1990) that have become increasingly popular in recent years. For a practical or introductory reference on these and other statistical learning topics see Wüthrich & Merz (2022) and references therin. We used standard stochastic gradient descent (SGD) training with a tanh activation from scikit-learn Developers (n.d.b). Knowing exact index components as features in a NN gave promising results (i.e. the first row of table 4.5). As



(a) (b) (c)

Figure 4.7: Same NN Model as in Figure 4.6 but now comparing returns of (a) the model and real data (b) the in sample index vs SPX returns and (c) the out of sample index vs SPX returns

we discussed in the previous section we were ultimately interested in the performance of the generic CORE economy which delivered comparable if not better performance out of sample.

As we set up our NN with 10 layers, one issue we quickly discovered was that including trailing days of returns meant very wide NNs. For instance, when trying to model the SPXAV10P in Table 4.5 the number of features is 10 trailing days times the number of indices e.g. 130 in the CORE bonds and equity row. Relative to the 10 hidden layer results presented in table 4.5 our NNs are very shallow. In order to remain confident we were not simply 'fitting the data' and accessing the full potential of so-called deep learning[6] to correctly handle novel inputs we wanted to explore larger, deeper networks.

Our first attempt of improving the fit by increasing the dept of the NN from 10 hidden layers to 20 is shown in Figure 4.6. In Figure 4.6 we observe that we can get reasonably good fit in and out of the sample, but cumulative returns, viewed through index price levels start to diverge out of sample. The divergence is mostly caused by a few outliers whose magnitude is not properly captured, even in sample. This nuance is best observed in Figure 4.7b, depicting relatively good agreement of actual returns with the ones produced

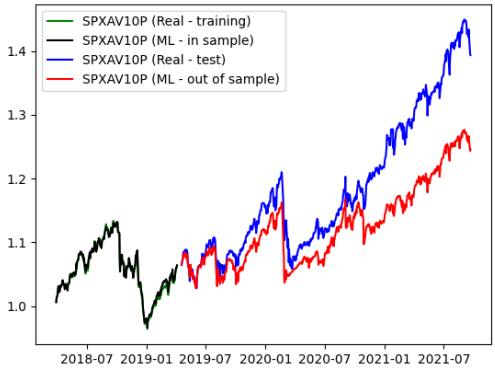
[6]see Nielsen (2015) for information about deep learning and Safran & Shamir (2017) for more info about the depth vs. width trade-off.

via ML (visually, most blue dots are covered with orange ones) in sample, while still having few observations that disagree.

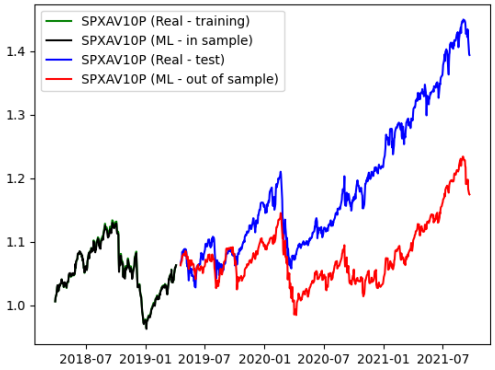
Unfortunately we observed that the NNs in this case started to suffer from vanishing gradients and a low convergence rate (see Hochreiter (1998) and references therein). We claim a more appropriate technique is Gaussian Process Regression (GPR).

We will not reiterate the theory of GPR here, nor explain the implementation we used from scikit-learn Developers (n.d.a) other than to mention we made use of a standard RBF kernel. Suffice it to say that GPR is a technique that can be thought of as a Bayesian stochastic kernel machine or a generalization to Bayesian linear regression accounting for more complex, nonlinear, functional forms. What motivated our use of it here was twofold: computational cost and the theoretical qualities of GPR. Heuristically, the time complexity of SGD is about $\mathcal{O}(dn^2)$[7] where d is the number of parameters and n is the training sample size. By comparison the time complexity of fitting a Gaussian process regressor is $\mathcal{O}(n^3)$. The large NNs in this case make GPR comparable or more practical. Theoretically, a Gaussian process regressor with an RBF kernel can be thought of as the limit of an infinitely deep NN (albeit a RBF network not an MLP, see Domingos (2020)). So we can sidestep the vanishing gradient issue, while hopefully achieving better out of sample performance at the same cost. Indeed Table 4.5 seems to bear this intuition out. Additionally many common financial models are in turn themselves Gaussian processes with a particular kernel. So we can model the CORE indices with some stochastic process and GPR will produce the resulting index models as a jointly distributed stochastic process. For the mathematically inclined GPR provides a common language between potential CORE models and our machine learning models: stochastic processes. We refer the interested reader to Williams & Rasmussen (2006), Lilley & Frean (2005), Domingos (2020) and references therein.

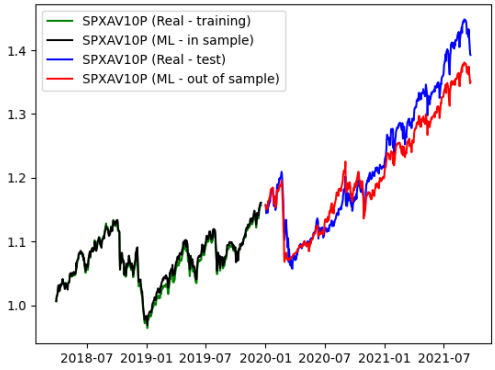
[7] i.e. $\mathcal{O}(d) \times \mathcal{O}(n) \times \mathcal{O}(1/\eta)$ the cost of gradient computation per epoch multiplied by $\mathcal{O}(1/\eta) \approx \mathcal{O}(n)$ where η is the learning rate.



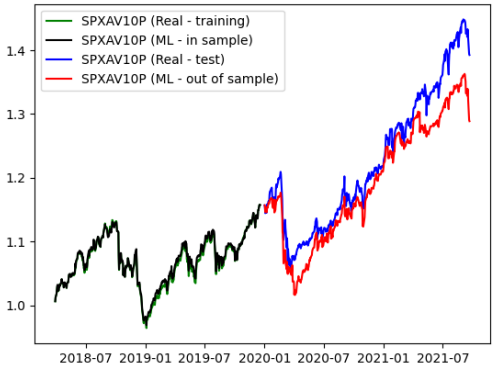
SPXAV10P (GPR) 30%/70%



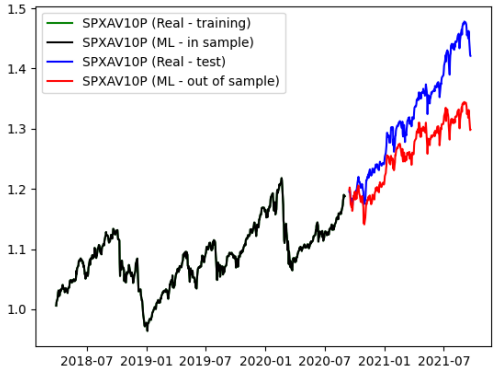
SPXAV10P (NN) 30%/70%



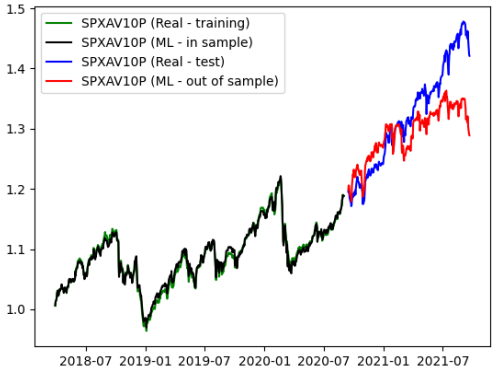
SPXAV10P (GPR) 50%/50%



SPXAV10P (NN) 50%/50%



SPXAV10P (GPR) 70%/30%



SPXAV10P (NN) 70%/30%

Figure 4.8: Comparison of NN and GPR Models for SPXAV10P in and out of sample for various train/test splits.

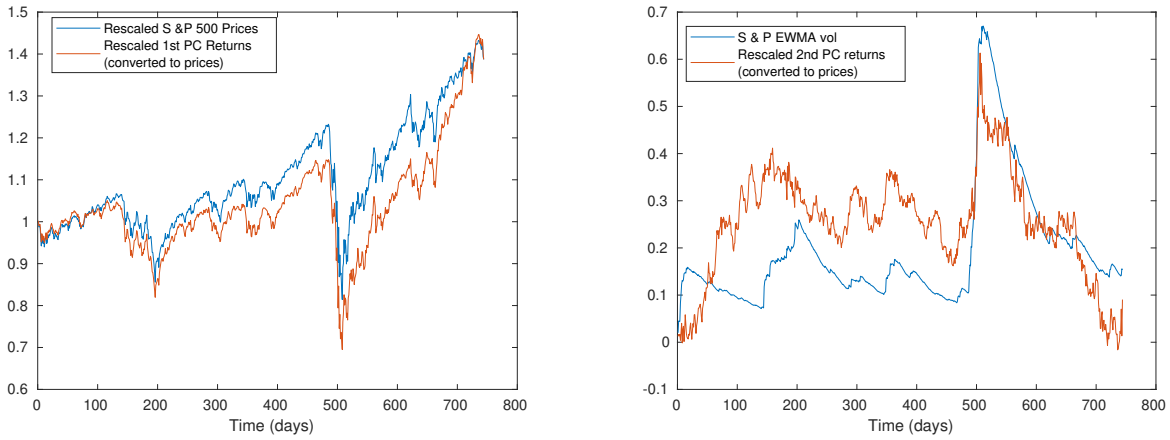
Discussion of Results

From table 4.5 we see that the CORE data set performed fairly well as predictors. Even though more cross validation is required we see the improvement in out of sample performance when compared to the case of known index elements. Additionally, we achieved these results with less trailing days than the actual volatility estimation techniques for these indices.

It is important to note a difference in the results summarized in Table 4.5. As a starting point we took S&P 500 returns and "added" interest rate inputs. In addition, we tested adding our entire set of CORE bond bonds to it and observed that there is no material difference in the quality of the estimate, as measured through R^2 out of sample. This was true regardless of the proxy for interest rates used. Under this data set-up GPR does a lot better than NN with 10 layers and the NN started to suffer from vanishing gradients if we tried to increase the network depth. Alternatively, if we substitute S&P 500 as a feature with the broader set of CORE equity indices, we see notable improvements across the board. Even though, CORE bond selection yields worse results compared to more precise interest rate proxy choices, we are happy with the R^2 of over 0.907 in case of GPR and slightly less in case of NN with 10 layers. In addition, inclusion of more indices to represent the broader equity market seems to solve the issue with vanishing gradients in the impractically deep NN.

We see the stark difference in performance between NN and GPR, and conclude that we obtain the best R^2 out of sample with CORE data (generalized economy) compared to known custom index components used as features. This result is counter intuitive at first and we search for the explanation via principal component analysis. Not surprisingly, we see that the principal components of the CORE data set daily returns corresponding to the two largest eigenvalues are related to "market movements" and "market volatility" (Figure 4.9). This gave us confidence that the effects our ML models were picking up are real and explainable in our context. Despite the relatively short time window of the CORE data (approximately 3 years) the core data set contains a great deal of cross sectional data across many indices as evidenced by the total number of returns column in table 4.5. As

a result we can extract the equivalent information on exposure to a wider set of market factors versus the deeper historical data set of just a few indices.



(a) (b)

Figure 4.9: The principal components corresponding to the largest two eigenvalues and rescaled S&P returns (a) and volatility (b)

4.2.4 Results

A Single Annuity

After establishing the practicality of using the CORE set of indices in conjunction with ML to reproduce risk controlled indices inside FIAs we of course want to use this to accomplish our goal of annuity interest credit forecasting. With simplicity in mind we looked only at participation strategies of the kind we discussed in 4.1. Not only are these easy to reproduce but the resulting return distribution is more closely related to the underlying index distribution. Any underlying issues with the index calculations are easier to spot in these annuity crediting rates.

		GBM	GPR	NN
Actual total returns (i.e. not assuming independence between years)	mean	6.43%	6.05%	5.95%
	std	1.90%	0.68%	0.65%
	skewness	28.56%	44.21%	39.66%
Total returns assuming independence between years	mean	6.43%	6.25%	6.12%
	std	1.90%	1.87%	1.81%
	skewness	28.56%	42.75%	41.98%

Table 4.6: 10 year annualized FIA return statistics

As a base case we simulated the CORE economy (equity and bond indices from table 4.4) using a multivariate GBM [8]. That is we modeled P_t and B_t by a GBM with regards to FIA summary in equation 4.1. These multivariate normal returns were then used as the input into our ML obtained RC index function (f in equation 4.1). The distribution of FIA returns based on this scenario is shown in Figure 4.10 in red and orange, for GPR and NN, respectively. Alternatively, we could simulate the RC index directly as a GBM; that is simulate I_t from equation 4.1 with a GBM directly. In Figure 4.10 the distribution of FIA returns obtained by this direct index simulation via GBM is shown in blue.

Surprisingly, at first glance, this does not translate into comparable cumulative annuity returns over 10-year period. What's surprising is not that the index itself does not seem to evolve according to a GBM (Figure 4.1) but rather that using a GBM based input to an RC index function produced such narrow distributions. Consider one example of the FIA with participation strategy, struck on SPXAV10P and a 55% participation rate over a 10 year accumulation period. Leveraging our ML models described in section 4.2.3 with the CORE Equity and a I00078US as an interest rate input[9] we simulated a sample of annuity returns (Figure 4.10, left panel).

We observe that distributions of FIA returns obtained via ML are tremendously different than those obtained through direct index modeling via the GBM. Despite reproducing the mean we can see the shape of the distributions of FIA returns is very different.

[8] directly estimating the parameter by fitting a multivariate normal to the daily returns in the dataset

[9] i.e. U.S. Treasury Bills 1-3 Months index. As we discussed in 4.2.3 the choice of fixed income exposure seems to make little difference to the final index result

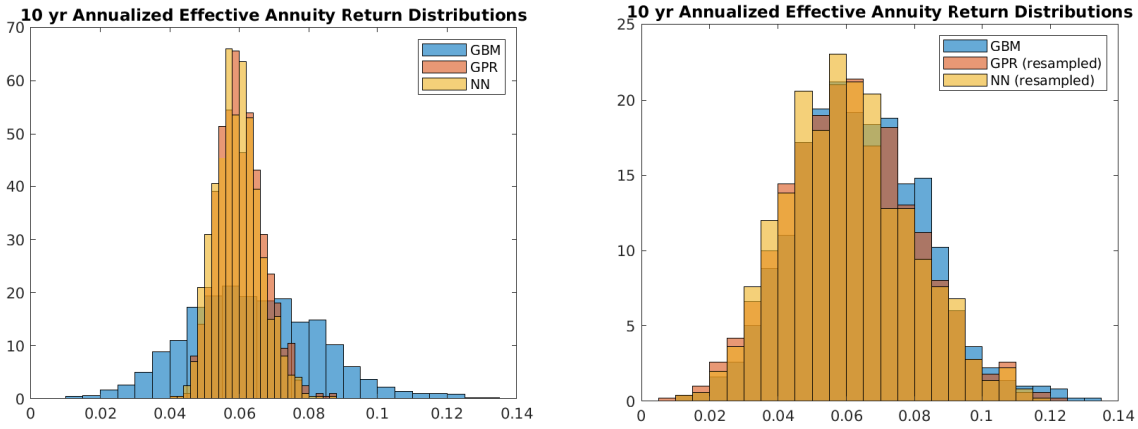


Figure 4.10: FIA return distributions for 1000 simulations for SPXAV10P with 55% participation, alongside the same simulation results with resampled monthly returns before crediting to the FIA. Notice on the left, the relative performance between the two products is driven more by the mean as the ML models have narrower yield distributions.

At first we suspected that our ML models were poorly standardized. Not only this was not the case but we were happy to observe that the annual credited returns are *essentially identical*. In table 4.6 we look at the yield Y computed from annual credited returns X_i such that $(1+Y)^{10} = \prod_{i=i}^{10}(1+X_i)$. We found that the X_i's are correctly distributed but the distributions diverges when looking at total returns i.e $\prod_{i=i}^{10}(1+X_i)$. The only way to reconcile these facts is to realize the X_i's are not independent. To support this we resampled the monthly returns of our ML derived index returns and redid the experiment with the annuity returns (Figure 4.10, right panel).

The intuition is as follows: even though the features we used (underlying economy returns) were simulated via GBMs the ML model is picking up on the volatility calculations in SPXAV10P. Even though, in the case of a pure GBM there is no autocorrelation structure, in our ML case large swings are followed by a higher "estimate" for $\hat{\sigma}$ and π_t (see equation 4.6) dampening returns in the next period. This is the key benefit of the RC indices as marketed. Even though the volatility is constant in our GBM inputs, *sample* volatility is not and so π_t will be non-constant. This will introduce a native auto correlation to the RC index output of our ML model, that is effectively independent of the model for

underlying components.

Compared to the results with the same products in section 4.2.1 we would expect some change in the distribution (perhaps some reduced variance etc). The results of Figure 4.10 on the other hand seem too extreme. Even ignoring the data, consider the following heuristic or 'back of the envelope' argument. The original RC index is constructed to try and achieve a $\sigma = 0.10$, so the participation of 55% should give $0.55\sigma = 0.055$ and the floor at zero percent will again *roughly* half this to give $\approx 0.25\sigma$ at 0.025 or 2.5%, on the same order of magnitude of our observations. But a less than one percent volatility, especially given the high returns of $\approx 6\%$ seems counter-factually low and defies financial sense. So, regardless of the choice of benchmark (direct GBM or historical data) for our results we can see that the shape of our distribution is too narrow and the standard deviation of rates are an order of magnitude too small. These findings motivate the results studied in section 4.3.

Multiple Annuity Results

Before moving onto section 4.3, where we will study the effects of dependent returns, we want to take a moment to explore its importance. Recall the original goal was a comparison between two annuities linked to different indices. In order to do that we need to perform simultaneous simulation of two indices, comprised of different components and with different participation rates. Here we add Standard and Poor's MARC 5% Excess Return Index (SPMARC5P) - similar to the other two indices explored before but with a gold index component as well. For this index, prevailing participation rate (in late 2021) is set at 85% for the strategy. We simulated the same CORE economy with the addition of a gold index since MARC5 has it as a component in hopes that this would improve performance of our estimates and repeated the same procedure as before. As we can see in Figure 4.11 we see similar behaviour as before. It is worth noting that in Figure 4.12 we see a larger difference between the NN and GPR results in the MARC5 case than the SPXAV10P case. This is not too surprising as the MARC5 index is a more complex product and the two methods likely struggle to sufficiently capture the index dynamics to

	SPXAV10P with 55% par		SPMARC5P with 85% par		
	Mean	Std	Mean	Std	P(SPX >Marc5)
GBM	0.059	0.018	0.063	0.016	42.8%
GPR	0.050	0.006	0.065	0.004	2.2%
NN	0.049	0.006	0.057	0.004	11.8%

Table 4.7: Summary statistics for of comparison of annualized returns for two annuities stuck on two different indices

the point that they converge.

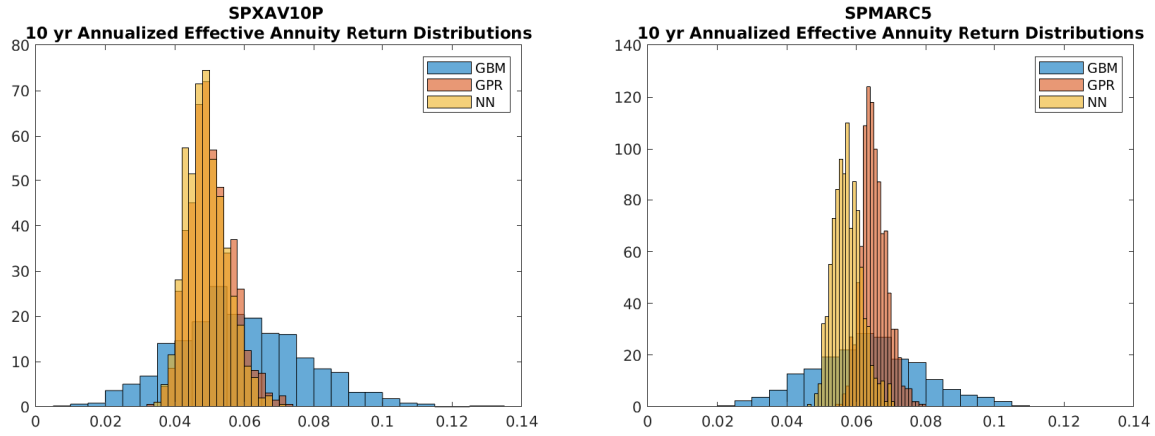


Figure 4.11: FIA return distribution for MARC5 with 85% participation.The results have similar properties to the FIA linked to SPXAV10P .

Not only do we see the same dramatic change in the summary statistics, we also see a dramatic change in relative performance. This is to be expected as the lower variance means narrower yield distribution and therefore less 'overlap' and more dramatic relative performance especially compared to the GBM base case (see Figure 4.12 and Table 4.7).

While this was not what we were hoping to uncover, these results are also very encouraging. An appropriate and well trained ML model was able to capture the very features these indices are marketed on, namely volatility control. However, due to the insufficiently sophisticated "economy" simulation model we see the overreaction of the ML trained index functions. This suggests (as we will see in part 4.3) that it is likely that a more sophisti-

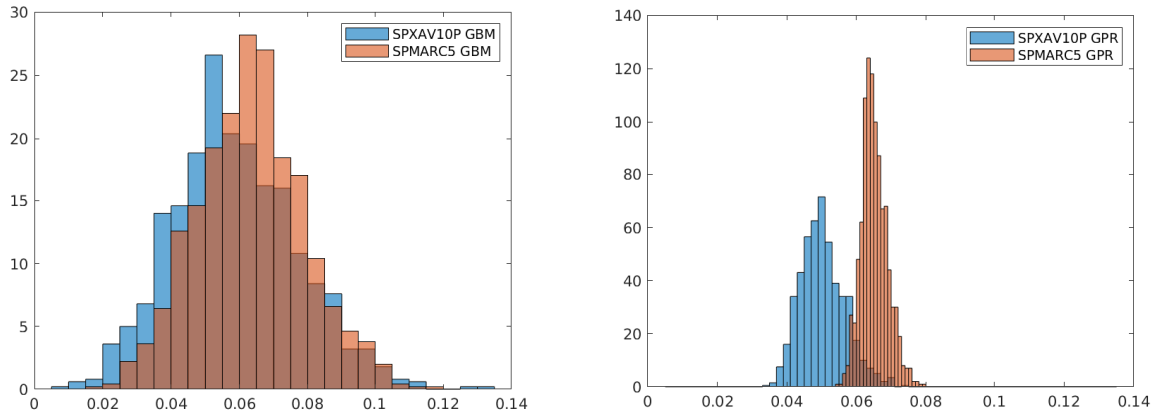


Figure 4.12: The relative performance between the two products is driven more by the mean as the ML models have narrower yield distributions.

cated CORE economy model is needed in place of correlated GBMs. The autocorrelation that affects variance through aggregate returns in the index function has to interact with some autocorrelation native to the CORE set of indices to get more realistic results. These issues and what features the CORE model are needed and how to get proper annuity returns are the topics we explore in section 4.3.

4.3 Risk Control Indices and Volatility Modeling

4.3.1 FIA Index construction and Implications

In this part of the chapter we explain the results of section 4.2.4. To start, we provide some relevant context related to FIA construction. As mentioned, in section 4.2.2 the low interest rate environment that resulted from the Great Financial Crisis was the main motivator for the design of risk-controlled indices. With historically low yields for Treasuries and corporate bonds, insurance providers needed a new product to keep annuity returns competitive. Providing index exposure with a principal guarantee is one way to do this. However, reliance on a broad market index with high volatility (e.g. the SPX) is not tenable in an economic environment where the option budget available under low interest rates renders the crediting rate unattractively low to annuity buyers. The solution widely adopted by the industry were RC indices of the kind discussed in section 4.2.2. Risk control mechanisms modulate exposure to the respective primary indices to make hedge positions cheaper and pricing of the associated options more stable. Ultimately, an entity in the supply chain has to have the underlying index or components on their books somehow. An FIA provider needs to take some of the premiums received for the FIA and purchase an option or portfolio of options giving them the requisite exposure to the index to credit the FIA returns. The option provider in turn needs to hedge this exposure to the index in the open market. A nice schematic of this process is in Watson (2022).

This process greatly constrains the kind of index behaviour we would expect to see. Any successfully marketed index with behavior sufficiently idiosyncratic would very quickly run into capacity issues with hedge providers trying to create exposure and hedging in the open market. Indices therefore must be constructed from components that are very liquid or similarly exposed to broad risk factors to make substitution possible (e.g. hedging the SPX with another 'market proxy'). For instance, in section 4.2.3 we explored (via PCA) the possibility that our CORE economy returns were mostly driven by a market and volatility risk factor. Risk control indices must necessarily be driven by exposure to these same factors; otherwise, the hedge provider would not be able to sell the necessary

options to make a market for FIAs possible. Likewise, the risk control mechanisms cannot be too complex and sensitive or the number of transactions needed to hedge would be cost-prohibitive. We claim the evidence of this is that RC indices still have interesting enough inter-temporal return dependence that, when passing a GBM through the RC mechanism and aggregating returns over an FIA crediting period, we get counterfactual results[10] of the kind in 4.2.4. Put another way, while RC indices may have been motivated in part by FIA products they have a life separate from FIAs and are subject to all the same construction and modelling challenges present with any index. This lies at the heart of our results in this part.

In section 4.2, we discussed how to recreate cross sectional index returns by reproducing the index calculations themselves and passing a smaller CORE economy through them. We saw the simulated economy needed certain essential features that interact with the index calculations. These are features that simple GBM clearly does not possess. In this section, we explore what those essential features may be, and which are sufficient to reproduce reasonable FIA returns. We do this by first studying the stylized facts of indices *in general* and how this affects the aggregation of returns in section 4.3.2. In sections 4.3.2, 4.3.2 and 4.3.2 we apply this to the popular class of stochastic volatility models i.e. the very class of model that inspired the RC mechanisms we are interested in. We then return to RC indices specifically in section 4.3.3 before finally returning to the question of FIA returns in section 4.3.3. A brief conclusion tying the results of sections 4.2 and 4.3 together follows.

4.3.2 Stochastic Volatility: Statistics of Indices

To start, consider that prices are often less useful that the relative growth or loss of an asset. Hence why we employ the notion of a return. Given a time series of prices $\{P_i\}_{i=1}^N$ sampled at regular intervals of time Δt (e.g. a day). We consider two different notions of asset returns, the simple or gross return (R_i) and the logarithmic return (r_i):

[10]It is worth noting briefly that dependence in the aggregation of returns is a known, perhaps under investigated, phenomenon in finance generally. See for instance Schwartz & Whitcomb (1977),Perry (1982),Baltussen et al. (2019) or Xiao et al. (2021)

$$R_i = \frac{P_i - P_{i-1}}{P_{i-1}} \quad r_i = \log\left(\frac{P_i}{P_{i-1}}\right)$$

When the differences between the two are discussed what is emphasized is that logarithmic returns are additive across periods of time and that simple returns are additive across portfolio components. What is often lost is the idea that these are *statistics* of the price data, with all the associated statistical questions and issues. For instance a gross return is clearly most useful when describing a quantity that grows additively or linearly in time whereas a logarithmic return captures the a rate of an exponential growth. Depending on the data, the context and application these two notions of return are not a priori interchangeable. The difference can be captured when considering Geometric Brownian Motion (GBM)- a common null model of financial returns going back to the work of Merton (1969), Samuelson (1973) and the thousands of works derived since:

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t \longrightarrow P_t = P_{t-1} \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma W_{\Delta t}\right\} \tag{4.7}$$

There is an inherently geometric or multiplicative growth to GBMs so that $r_t \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2\Delta t)$ whereas R_t will be a shifted Log-Normal distribution with mean $\exp(\mu\Delta t) - 1$ and variance $[\exp(\sigma^2\Delta t) - 1]\exp(2\mu\Delta t)$. Clearly as a statistic logarithmic returns are favourable as $\frac{r_t}{\Delta t}$ will be scale free and the mean will report a typical portfolio trajectory (for recent discussions of the relevance of this see Peters (2019) and Carr & Cherubini (2022)). Despite this, simple returns are often used in many financial calculation e.g. annuity crediting formulas which is of interest to us.

Fortunately it is the case that at small enough time scales these two notions of return are functionally interchangeable. To see this we need only consider the Taylor approximation to $\exp(x)$ (or equivalently $\log(1 + x)$):

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \exp(r_t) - 1 \approx r_t$$

Consider the justification for the GBM as a suitable null model of financial asset dy-

namics. Regardless of the kind of return considered we ultimately can express longer time returns using a sum of log price ratios:

$$1 + R_T = 1 + \frac{P_T - P_0}{P_0} = \prod_{t=1}^{T} \frac{P_t}{P_{t-1}} = \exp\left(\sum_{t=1}^{T} \log\left(\frac{P_t}{P_{t-1}}\right)\right) \longrightarrow \text{LogNormal by C.L.T}$$

$$r_T = \log\left(\frac{P_T}{P_0}\right) = \log\left(\prod_{t=1}^{T} \frac{P_t}{P_{t-1}}\right) = \sum_{t=1}^{T} \log\left(\frac{P_t}{P_{t-1}}\right) \longrightarrow \text{Normal by CLT}$$

As long as the $\log\left(\frac{P_t}{P_{t-1}}\right)$ terms are sufficiently i.i.d and well behaved we can justify the GBM as a model by invoking the C.L.T. Indeed, at long enough time scales this model does appear sufficient (Cont (2001)).

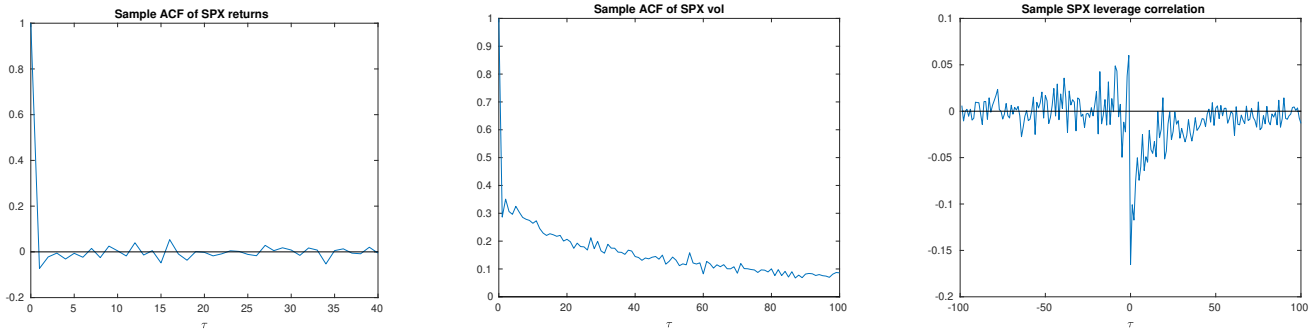
That said, the exact nature of the temporal dependence of returns is the subject of innumerable debates and can in some ways be thought of as *the* open problem of mathematical finance. For insight investigators often turn to statistical autocorrelation functions (ACFs) to provide insight. Consider the autocorrelation of S&P returns (i.e. $COR[r_t, r_{t+\tau}]$) in Figure 4.13a, it suggests that any dependence between returns is nonlinear in nature as returns are not linearly correlated through time. Indeed many nonlinear functions of returns express nontrivial correlations (again see Cont (2001) and references therin). Nonlinear dependence can be expressed as the sum of contributions of correlations between powers of returns via Taylor series. So it is useful to consider products of powers of returns alone. For instance in finance we often consider the magnitude of squared returns e.g the volatility squared (typically denoted by σ_t^2).

Let us define the instantaneous squared volatility at time i by $\hat{\sigma}_i^2 = \frac{r_i^2}{\Delta t}$ (or $\frac{R_i^2}{\Delta t}$ where $R_i \approx r_i$). There are two signifigant nonlinear correlations that appear to be persistent features of many traded assets accross different kinds of markets (see previous references). We demonstrate them here with daily S&P 500 (SPX) returns:

1. Volatility Autocorrelation: In Figure 4.13b we plot the autocorrelation of instanta-

neous volatility in terms of some lag τ i.e. $c(\tau) = COR[\hat{\sigma}_t, \hat{\sigma}_{t+\tau}]$. The results strongly suggest that there is a persistent sizable dependence in volatility across time.

2. Leverage Correlation: In Figure 4.13c we plot $L(\tau) = COR\left[\left(\frac{r_t - \mathbb{E}[r_t]}{\Delta t}\right), \hat{\sigma}^2_{t+\tau}\right]$. Large negative price movements seem to be correlated to increases in volatility that decay after the initial shock. In general the correlation is strongest following negative returns- that is volatility clustering cannot meaningfully 'anticipate' a sudden price drop.



(a) (b) (c)

Figure 4.13: Sample autocorrelation functions of returns (a) volatility (b) and the Leverage Correlation (c)

Higher order correlations may have some impact on prices but for reasons of both statistical quality and modeling convenience we typically restrict ourselves to volatility and leverage. This motivates the common strategy of replacing the volatility σ in (4.7) by a stochastic process σ_t with a positive decaying autocorrelation. In order to induce the leverage correlation typically a dependence between σ_t and W_t is imposed. Now reconsider the issues raised in Figures 4.10 and table 4.6 as well as the analogous results for multiple annuities in section 4.2.4. We want to understand how the variance of returns can depend on the aforementioned correlations and look at the variance of the sum of log returns i.e. $r_T = VAR[\sum_{t=0}^{t} r_t]$.

If we say now that $dP_t = \mu P_t dt + \sigma_t P_t dW_t$ then we simply have:

$$r_T = \log\left(\frac{P_T}{P_0}\right) = \int_0^T d\log(P_s) = \int_0^T \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \int_0^T \sigma_t dW_t$$

i.e. a GBM like formula with a stochastic volatility.

To explain the aggregated variance of index returns we can hopefully decompose $VAR[r_T]$ into a sum of the contributions from autocorrelation and leverage. To do so we need to make some assumptions about the form of σ_t and possible dependence with r_t. For the rest of this section we will study models of the following form:

$$\begin{aligned} r_T &= \int_0^T \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \int_0^T |\sigma_t| dW_t^{(1)} \\ \sigma_t &= m + \lambda \int_{t_0}^t k(t,t') dW_t^{(2)} \\ \rho &= \mathbb{E}[W_t^{(1)} W_t^{(2)}]/t \end{aligned} \tag{4.8}$$

This general framework can be augmented further for realism. For example we could make the change $|\sigma_t| \to f(\sigma_t)$. The choice of $f(\cdot)$ would control the ultimate stationary distribution (if it exists) of volatility e.g., $f(x) = e^x$ would produce log-normal distributions for volatility (Tegnér & Poulsen (2018)). For the remainder of this section we will consider the model presented above and focus on the choice of volatility kernel $k(t,t')$. This kernel will determine the stochastic volatility model under consideration. This model for r_t and σ_t is flexible enough to capture many cases but parsimonious enough to compute the variance we are interested in:

Theorem 4.3.1. *If returns are generated by a model of the form of (4.8), then:*

$$VAR[r_T] = \underbrace{\mathbb{E}\left[\int_0^T \sigma_s^2 ds\right]}_{=\sigma^2 T \text{ in GBM}} + \int_0^T \int_0^T \left(\underbrace{\frac{1}{4} C(t,u)}_{\text{vol. autocorr}} - \underbrace{2\rho\lambda\Lambda k(t,u) H(t-u) \mathbb{E}\left[\sigma_t \sigma_u\right]}_{\text{leverage corr.}} \right) dt du \tag{4.9}$$

Where $C(t,u) = \mathbb{E}[(\sigma_t^2 - \mathbb{E}[\sigma_t^2])(\sigma_u^2 - \mathbb{E}[\sigma_u^2])]$, $\Lambda = P(\sigma_u > 0) - P(\sigma_u \leq 0)$ *and* $H(\cdot)$ *is the Heaviside step function.*

We can simplify equation (4.9) further by making the substitution $C(t,s) = 2(\mathbb{E}\left[\sigma_t\sigma_u\right]^2 - m^4)$. This conveniently means we need to compute only one model correlation. The second, leverage term, in equation (4.9) is so called for the following reason. If we say $r_t \approx \frac{dP_t}{P_t}$ and $\Delta t \approx dt$ then we can show:

$$\mathbb{E}\left[\left(\frac{r_t - \mathbb{E}[r_t]}{\Delta t}\right), \hat{\sigma}_u^2\right] \approx 2\rho k(u,t)\theta(t-u)\mathbb{E}[\sigma_t\sigma_u] \tag{4.10}$$

That is, we can see that the leverage term in (4.9) is proportional to the the leverage correlation $\mathcal{L}(\tau)$. So we have the expression we were looking for: the variance of r_T in terms of the two significant correlations we have discussed.

Many indices conform to the S&P pattern laid out in figures (4.13b) and (4.13c), that is a positive autocorrelation and a negative leverage correlation implying $\rho < 0$. A negative ρ in turn will mean the sum of the correlation terms in (4.9) is positive, so the variance of the T period return distribution will be larger given inter-temporal dependence than otherwise. RC indices are constructed specifically with such a dependence induced by stochastic volatility of the kind in (4.8) in mind. Indeed as the RC mechanism attempts to dampen the shocks of stochastic volatility (shrinking our leverage and autocorrelation terms) we should see smaller T period return variance in the RC index versus the underlying index.

We now apply (4.9) to some well known stochastic volatility models in the form of (4.8) to gain some insight. Keeping in mind that in order to fit the stylized features of index data we should see Leverage correlation and volatility autocorrelation as well as a variance that will be asymptotically linear to conform to a GBM-like behaviour seen at large time scales.

The Ornstein–Uhlenbeck (OU) Model

We will start with perhaps the simplest well known stochastic volatility model: the OU process. This process can be put in the form of our model above via an exponential kernel:

$$d\sigma_t = -\alpha(\sigma - m)dt + \lambda dW_t^{(2)} \Longrightarrow \sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^{(2)} \tag{4.11}$$

So we use the following in equation (4.9):

$$k(t,u) = e^{-\alpha(t-u)}$$
$$\mathbb{E}\left[\sigma_t \sigma_u\right] = m^2 + \left(\frac{\lambda}{2\alpha}\right) e^{-\alpha|t-u|}$$

Let us denote $\left(\frac{\lambda}{2\alpha}\right)$ by θ. For the OU model described in (4.11) we can show that:

$$VAR[r_T] = (m^2+\theta)T + \theta(\theta - 2\lambda\Delta\rho)\left[\frac{T}{2\alpha} - \frac{1-e^{-2\alpha T}}{(2\alpha)^2}\right] + m^2(2\theta - 2\lambda\Delta\rho)\left[\frac{T}{\alpha} - \frac{1-e^{-\alpha T}}{\alpha^2}\right] \tag{4.12}$$

To get a sense of what this means, let us consider how this behaves asymptotically. For a small T approximation take the Taylor series of (4.12):

$$VAR[r_T] \approx (m^2+\theta)T + \alpha\left[2\theta(\theta - 2\lambda\Delta\rho) + m^2(2\theta - 2\lambda\Delta\rho)\right]T^2 + \mathcal{O}(T^3) \tag{4.13}$$

For a negative ρ we will see a superlinear growth in the variance whereas for large T we can see that:

$$VAR[r_T] = \mathcal{O}\left(\left(m^2 + \theta + \frac{\theta(\theta - 2\lambda\Delta\rho)}{2\alpha} + \frac{m^2(2\theta - 2\lambda\Delta\rho)}{\alpha}\right)T\right) \tag{4.14}$$

Eventually for large enough T this will become linear.

It is also worth making note of another asymptotic regime, that is for $\alpha \to \infty$ we can see that:

$$VAR[r_T] = \mathcal{O}((m^2 + \theta)T) \tag{4.15}$$

Where we make use of big-O notation to avoid taking explicit limits (as the $\theta \to 0$). This makes sense, in the large α limit the characteristic scale of the volatility correlation $(1/\alpha)$ goes to zero and the process starts to resemble a GBM. It may also explain why this particular model is as popular as it is in many applications: it allows for the creation of a non-constant volatility surface in options modelling without deviating too much from a GBM. To see this consider the following special case of the OU model, the Heston stochastic volatility model (Heston (1993)).

It is well known that if $m = 0$ a sum of squared OU processes can be used to create a Cox–Ingersoll–Ross (CIR) model- the same process used in the Heston model. Proceeding from our single OU process we have the following correspondence:

$$\begin{aligned} d(\sigma_t^2) &= 2\sigma d\sigma_t + \lambda^2 dt \\ &= \underbrace{(2\alpha)}_{\kappa}\Big(\underbrace{\frac{l^2}{2\alpha}}_{\theta} - \sigma^2\Big)dt + \underbrace{(2\lambda)}_{\epsilon} dW_t^{(2)} \end{aligned}$$

If $m = 0$ then $\sigma_t \sim \mathcal{N}(0, \theta)$ and $\Delta = 0$. Making use of (4.9) or (4.12) we can show the variance of aggregate returns in the Heston model is:

$$\begin{aligned} VAR[r_T] &= \theta T + \left(\theta^2 - \Delta\rho\theta\epsilon\right)\left(\frac{T}{\kappa} - \frac{1 - \exp(-\kappa T)}{\kappa^2}\right) \\ &\approx \theta T + \kappa\theta^2 T^2 + \mathcal{O}(T^2) \end{aligned}$$

Where the second equation is the 'small T' approximation. While we still see the same behaviour as in the OU model the deviations from a GBM are less severe. For a large enough κ the differences are practically inconsequential and the sample variance of returns will scale mostly linearly.

Rough Volatility: The Gatheral or Stein & Stein Model

Many authors have pointed out the insufficiency of the Heston/OU model in a variety of ways (Daniel et al. (2005) and references therein). In particular the fact that volatility autocorrelation decays too rapidly as an exponential, whereas the data suggests something more akin to a power law. Some authors suggest that a sum or nested set of (potentially infinite) OU processes can account for this and reflects the different time scales of market participants (Bochud & Challet (2007), Perelló et al. (2008), Barndorff-Nielsen & Shephard (2001) and Vitali et al. (2019)). Others have attempted to create an entirely new class of stochastic volatility models using tools outside of the standard Ito process. One of these tools is the Fractional Brownian Motion (FBM) denoted by W_t^H:

$$W_t^H = \frac{1}{C_H}\int_{-\infty}^{t}\left((t-s)_+\right)^{H-1/2} - \left((-s)_+\right)^{H-1/2}\Big)\, dW_s$$

Where C_H is a normalization constant. The parameter $H \in (0,1)$, called the Hurst index, controls the degree of mean reversion or momentum the process has. We have chosen the value of $C_H = \int_{-\infty}^{0}\left(1-s\right)^{H-1/2} - \left(s\right)^{H-1/2}\Big)\, dt' + \frac{1}{2H}$ in order to give a convenient expression for variance:

$$VAR[W_t^H - W_u^H] = |t-u|^{2H}$$

And assuming that $t, u > 0$ we can use this to derive the autocorrelation:

$$E[W_t^H W_u^H] = \frac{1}{2}(t^{2H} + u^{2H} - |t-u|^{2H})$$

For $H < 1/2$ the process has negatively correlated values (i.e. mean reverting behaviour), for $H = 1/2$ it is a BM and for $H > 1/2$ it has positively correlated values. For more details see the original work introducing the FBM Mandelbrot & Van Ness (1968) or a more motern text like Holden et al. (2009).

For our purposes we will be studying a model inspired by Comte & Renault (1998) and Gatheral et al. (2018). Although it should be noted that in those references they made use of a log-normal model (i.e. $f(\sigma) = e^{\sigma}$ in 4.8). The model we study here will be a more tractable or 'smoothed' model called the fractional Stein and Stein model after Stein & Stein (1991) by Gatheral[11]. This model is constructed via the OU process, but we replace the BM driving the process with a FBM:

$$d\sigma_t = -\alpha(\sigma - m)dt + \lambda dW_t^H \Longrightarrow \sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H \tag{4.16}$$

Asymptotically at least models of the form of 4.16 will produce the sort of power law behaviour of volatility we see in the data (Cheridito et al. (2003) Theorem 2.3). Unfortunately it is difficult to express this model in the form of (4.8) (see for instance Alòs & Lorite (2021) pg. 121 for a complete discussion). For the often used *truncated fBM* $\tilde{W}_t^H = \frac{1}{C_H} \int_0^t (t-s)^{H-1/2} dW_s$ we can show

$$k(t,u) = k(t-u) = \frac{1}{C_H} \left(\frac{d}{dx}\Big|_{x=(t-u)} \int_0^x (x-z)^{\alpha} e^{\alpha z} dz \right) \tag{4.17}$$

But this comes at the cost of a more difficult to work with covariance function involving hypergeometric integrals.

Let $\tilde{m}(t) = m + e^{-\alpha t}(\sigma_0 - m)$ then:

$$\sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H = \tilde{m}(t) + \lambda \int_{0}^{t} e^{-\alpha(t-s)} dW_s^H$$

In general when looking at financial data α in (4.16) tends to be very small. In fact in

[11] Please see Gatheral et al. (2018) Appendix C for a discussion of the relevant differences.

this regime (months to a few years) we will say α is small enough to make the following approximation:

$$\sigma_t \approx \tilde{m} + \lambda(W_t^H - W_0^H) \tag{4.18}$$

$$\begin{aligned} k(t,s) &= \int_{-\infty}^{t} \left((t-s)_+\right)^{H-1/2} - \left((-s)_+\right)^{H-1/2}\Big) \\ \mathbb{E}\left[\sigma_t \sigma_u\right] &= m^2 + \lambda^2 E[W_t^H W_u^H] \end{aligned}$$

We will also make simplifying assumption that $\Lambda = 1$ i.e we are assuming we've chosen parameters guaranteeing σ_t is positive[12]. With these simplifications we get:

$$\begin{aligned} VAR[r_T] = m^2 \left[T + C_1^{(m)} \lambda^2 T^{2+2H} - C_2^{(m)} \rho \lambda T^{3/2+H}\right] + \\ + \left[\left(\frac{\lambda^2}{1+2H}\right) T^{1+2H} + C_1 \lambda^4 T^{2+4H} - C_2 \rho \lambda^{3/2} T^{3/2+3H}\right] \end{aligned} \tag{4.19}$$

Where:

$$C_1^{(m)} = \frac{1+4H+4H^2}{2(1+H)(1+2H)^2} \qquad C_1 = \frac{2+7H+4H^2}{4(1+4H)(1+2H)^2} - \frac{\sqrt{\pi}\Gamma(2+2H)}{8(2^{4H})\Gamma(3/2+2H)} - \frac{1}{8(1+2H)^2}$$

$$C_2^{(m)} = \frac{8}{3+8H+4H^2} \qquad C_2 = \frac{16H}{(1+2H)^2(1+6H)} + \frac{3\Gamma(1/2+H)\Gamma(1+2H)}{\Gamma(5/2+3H)}$$

This is the small time formula for the T-period variance analogous to (4.13) in the pure OU model. We would now like to know to the large T behavior as in (4.14). Unfortunately

[12]Keep in mind for large enough time we can take into account the OU-style mean rversion as well

as mentioned we cannot give the full model the same treatment analytically, but we can draw some conclusions.

It has been shown by Zeng et al. (2012) that there exists a stationary distribution given by:

$$p(t, \sigma_t|\sigma_0) = \frac{1}{\sqrt{2\pi\gamma^2(t)}} \exp\left\{-\frac{(\sigma_t - \tilde{m}(t))^2}{2\gamma^2(t)}\right\}$$

Where the long term variance is given by $\gamma^2(t) = 2H\lambda^2 e^{-2\alpha t}\int_0^t s^{2H-1}e^{2\alpha s}ds$. If $t \to \infty$ then $\tilde{m} \to m$ and if $H \leq 1/2$ then $\gamma^2(t) = \mathcal{O}(\frac{H\lambda^2}{\alpha}t^{-(1-2H)})$. So we can see that the long term distribution for r_T will become GBM-like with a constant volatility of m. The speed of this convergence is governed by two mean reversion components now: the classical OU reversion governed by α and the FBM reversion governed by H.

Finally consider the regime where we send $H \to 0$ (like what we did for eq (4.15)). Equation (4.19) will be non-informative as $H \neq 0$ in any FBM but we can make the approximation that σ is nearly constant at $\mathbb{E}[\sigma_t^2] \approx m^2 + \lambda^2$ i.e. the lowest order term in (4.19). Consider Figure 4.14 where we have chosen some reasonable model values and compared $VAR[r_T]$ to the "$H = 0$" case of $\sqrt{(m^2 + \lambda^2)T}$.

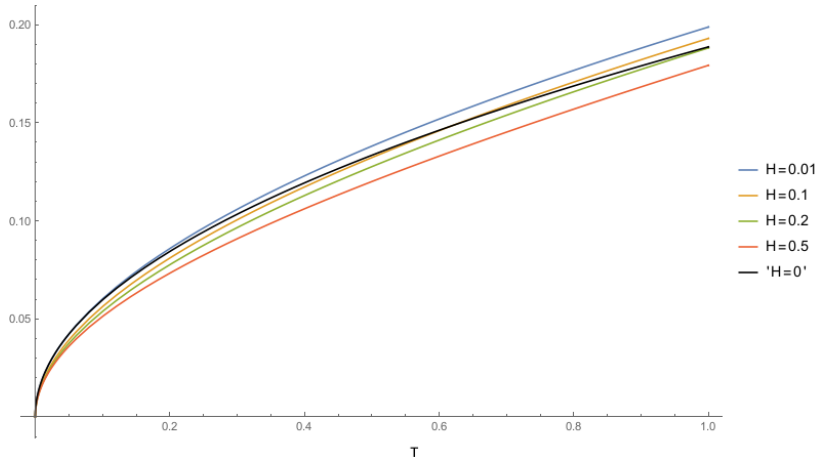


Figure 4.14: Variance for 'small time' r_T in rough model for $\lambda = 0.1, \rho = -0.5, m = 0.16$

Beyond Stochastic volatility?

In a GBM the standard deviation for 'small time' both simple and logarithmic returns scale with $\sqrt{\Delta t}$ i.e. $VAR[R_t] \approx VAR[r_t] = \sigma^2 \Delta t$ until Δt grows too large and this only applies to logarithmic returns. We also saw this in (4.14) and the expression for $\gamma(t)$ in the previous sections for the OU and Gathreal models respectively. Both models with respect to variance scaling eventually become GBM-like. For long enough (> 1 year) time scales as mentioned the GBM serves as a good model for equity asset returns. Indeed we will discuss how RC indices can dampen stochastic volatility resulting in smaller variances when considering the long time emergent GBM behaviour of the index. For now though we would like to consider the possibility of some alternative index models and mechanisms for completeness.

Let us call $S_{n\Delta t}$ and $s_{n\Delta t}$ the standard deviation of simple returns and log returns respectively over $n\Delta t$ time. Consider in Figure 4.15 the ratios $S_{n\Delta t}/S_{\Delta t}$ and $s_{n\Delta t}/s_{\Delta t}$ for the S&P 500 (SPX). We can see that the expected square root scaling does not hold for less than a year and the real relationship appears to be closer to $n^{0.45}$.

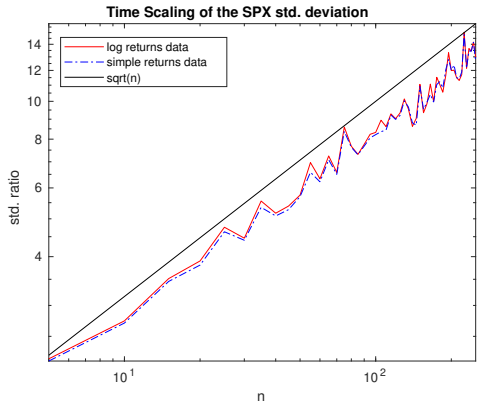


Figure 4.15: Log-Log plot of S and s ratios for $\Delta t = 1$ day.

To infer that it is dependence between small time log returns is the culprit for the anomalous scaling in Figure 4.15 consider a counterfactual presented in Figure 4.16. In this case we have re-sampled across time the log $\left(\frac{P_t}{P_{t-1}}\right)$ terms in order to remove any time dependence. As we can see the re-sampled return data conforms to the square root scaling of a GBM as the CLT convergence is not weakened by dependence in the summands. The difference in the standard deviation at the order of a year can be as much as 20% which will clearly be significant when considering portfolio returns.

The two popular models of the form (4.8) we studied in the previous section frustratingly cannot explain the kind anomalous scaling of the kind in Figures (4.15) and (4.16). We could however in the case of FBM driven model achieve at least a decrease in absolute variance albeit one that grows too quickly (Figure 4.14). For now we will have to be satisfied with this, and as we will see in section 4.3.3 models with this behaviour should prove satisfactory in capturing some essential features of the dependence structure of returns. We take the time here to note *why* this anomalous scaling was so hard to achieve and what if anything can be be done.

In both sections 4.3.2 and 4.3.2 our models possessed a stationary kernel i.e. $k(t,u) = k(t-u)$ producing a stationary process. Indeed in such a case we can show:

$$\mathbb{E}[\sigma^2] = \int_{-\infty}^{t} k^2(t-u)dt' = \int_0^{\infty} k^2(u)du$$

Therefore the first term of (4.9) grows as T. And if $C(t,u) = C(t-u) = (k \star k)(t-u)$ (where $k \star k$ is the convolution of identical copies of the kernel) we have that:

$$\int_0^T \int_0^T C(t,u)dtdu = T\int_{-T}^{T} C(s)\left(1 - \frac{|s|}{T}\right)ds$$

Which necessarily grows at least as fast as T. So with the stationary models of the kind we have used, it's impossible to achieve the scaling of variance we have in Figure 4.15. There are two possible responses to this. First we could abandon the form of (4.8) entirely. Rather than model the volatility separately, just a model the returns directly. So for example the following fractional GBM (studied by Li et al. (2016) and others) would exhibit power law behaviour for $\mathbb{E}[(r_t)^2(r_u)^2]$:

$$P_T = P_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W_T^H\right\}$$

And we would see the anomalous scaling we desire:

$$VAR[R_T] = VAR\left[\sigma W_T^H\right] = \sigma^2 T^{2H}$$

Unfortunately, in this case, return and volatility autocorrelation would be coupled. This implies a kind of non-guassian rough model or a multifractal model (see Mandelbrot et al. (1997) or Bouchaud et al. (2000)). Alternatively we could abandon models based on stochastic processes entirely. For instance agent based models of the kind of (Feng et al. (2012) or Chen et al. (2013)) or even quasi-deterministic models (Orlando et al. (2022) or Guyon & Lekeufack (2022)) have been able to capture many stylized facts of markets.

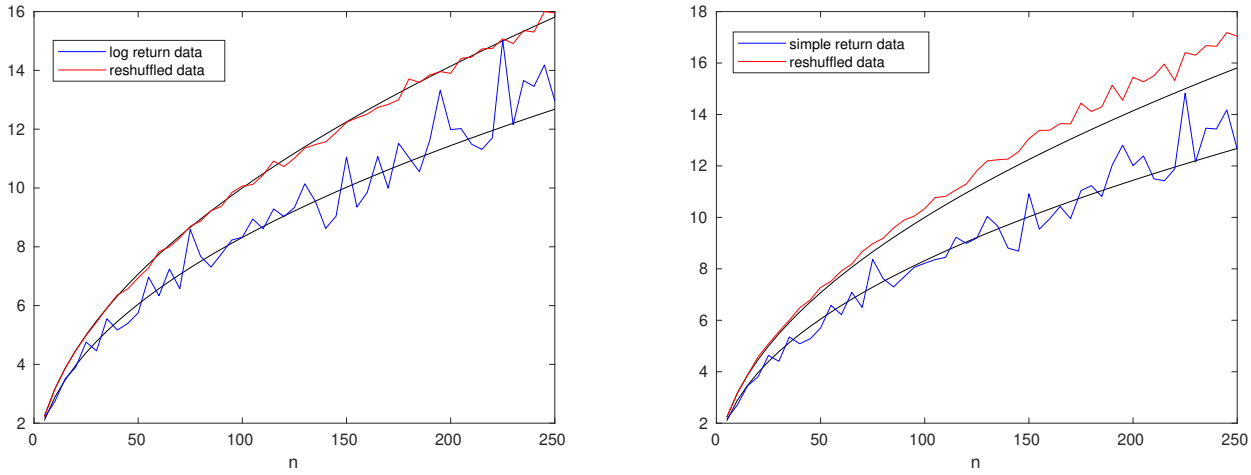
Of particular note is the recent work of Baltussen et al. (2019)), who suggested that in fact aggregated returns *do* appear to display serial correlations. The complex market operations used to hedge index exposure may influence the constituent securities that make up the index; the resulting feedback leading to serial correlations. A more realistic upgrade to (4.8) may introduce a serially correlated substitute for $W^{(1)}$.

For now as we will see, the important thing is to show that stochastic volatility and intertemporal dependence therein can influence and explain the aggregated return variance of an index (and by extension annuity) returns. When modelling Risk Control indices for the purposes of FIAs we can show volatility and leverage correlations of the original index are carried over to the risk control version. This will lead to issues using the naive GBM models no matter how well one can reproduce the Risk Control index calculations. See section 4.2.2 once again for a discussion of the relationship between risk control mechanisms and stochastic volatility in FIA indices.

4.3.3 RC index Simulation

The Role of Dependence

Thus far we have discussed the statistics of the SPX only. The object of study in this chapter however is the SPX RC indices. As we have seen in section 4.2.1, the choice of intertemporal dependence structure of returns for the RC indices is extremely relevant to



(a) (b)

Figure 4.16: Permuted return data for (a) log and (b) simple returns more closely conforms to $n^{0.5}$ as expected vs $n^{0.45}$ (both drawn in black).

determining annuity returns. We have also discussed how the intertemporal dependence structure induced by stochastic volatility will have an impact on the aggregated return statistics of an index. So it is not surprising this is a consideration in the case of RC indices as well.

Consider the volatility autocorrelation and leverage correlations of the SPXAV10P and the SPX together in Figure 4.17 bellow. Recall as well our outline of a generic RC mechanism in section 4.2.2. It is to be expected then that the RC mechanism 'dampens' the volatility autocorrelation. The autocorrelation shows up in the clustering of large volatility events: periods of high volatility follow one another. The RC mechanism when encountering a high volatility event will shift index exposure to bonds lowering the volatility of subsequent returns.

It is the RC index's leverage correlation that is more interesting. If the autocorrelation is damped, we would expect to see something similar in the leverage correlation through some mechanism similar to one in the equation (4.10). Instead, we see essentially the same pattern. It is tempting to think this is simply because of the normalization we have

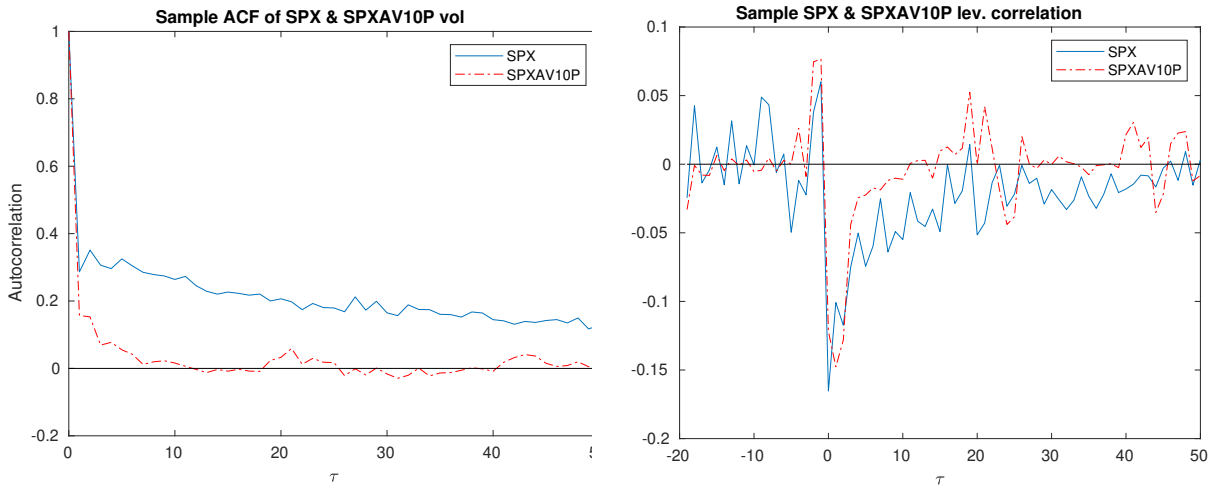


Figure 4.17: As we can see volatility autocorrelation is damped by the RC mechanism (left) while the leverage correlation persists (right)

chosen (indeed the leverage *covariance* is lower in magnitude). Let us perform the following experiment: let us fit a ML model for SPXAV10P as we have done in section 4.2.3 using the SPX and 100078US indices and look at the results when simulating the SPX via a GBM (Figure 4.18). In the language of (4.1) I is the SPXAV10P, P_t is the SPX, B_t 100078US, f is our ML model. Considering the results in Figure 4.18, we can clearly see that the autocorrelation is even lower than the RC index and the leverage correlation is greatly reduced. There does appear to be a small amount of induced leverage - but this could easily be an artifact of the ML model imperfectly fitting the RC mechanism. The GBM of course has no native autocorrelation to dampen and of course has no leverage correlation as volatility is not stochastic. On the other hand, as we have discussed, the observed pattern of SPX volatility autocorrelation appears to have a power law decay and is persistent over long periods of time (longer than the leverage correlation it seems, something also remarked on by Perelló et al. (2004)).

These suggest that, while the RC mechanism does reduce the intertemporal covariance of volatility (e.g. $\mathbb{E}[\sigma_t \sigma_u]$ in equation 4.9), the RC index still has a rich intertemporal dependence of volatility and a stochastic volatility of it's own inherited from the non-risk

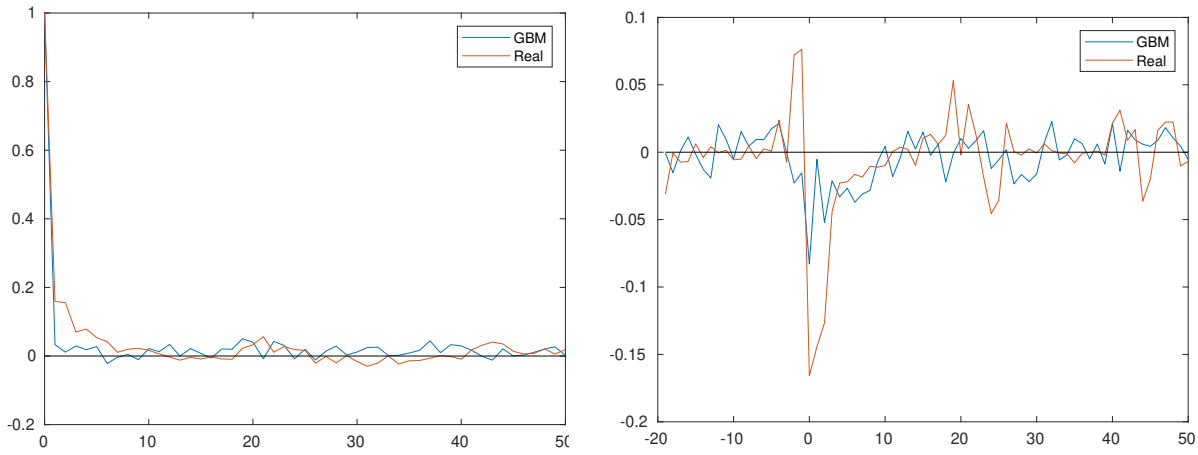


Figure 4.18: As we can see unlike the volatility autocorrelation (left) the RC mechanism will induce a similar leverage correlation when fed GBM generated returns

controlled underlying index. Given the persistence of volatility autocorrelation, it stands to reason that even the most ambitious RC mechanisms would be unable to completely dampen it. It is also a practical matter, as any index carrier cannot to purchase efficiently priced options from an index provider as the provider mitigates volatility risk[13] by adding a risk margin. Effectively, the hedge provider is pricing for volatility higher than the index is designed to maintain, as the trades required to perfectly conform to the RC mechanisms of (4.5) are not practical.

We now return our attention to the results of section 4.2.4 for annuity returns using a GBM economy (CORE or otherwise). For index models with stochastic volatility we would like to see them become GBM-like on longer time scales as we see in the data. We saw two examples of this in sections 4.3.2 and 4.3.2. Through (4.9) we saw that the stochastic volatility dynamics at smaller time scales can aggregate to effect the resultant larger time scale GBM. For instance in equation (4.14) we could see at large T the popular OU/Heston volatility dynamics show the variance is linear in T like a GBM but is larger than it would be otherwise if there was no dependence structure to volatility. Indeed if we were to use a

[13]or Vega risk in the parlance of options traders

model with no intertemporal dependence we would get:

$$VAR[r_T] = \underbrace{\mathbb{E}\left[\int_0^T \sigma_s^2 ds\right]}_{\text{GBM volatility}} + \int_0^T \int_0^T \left(\frac{1}{4}\cancelto{0}{C(t,u)} - \cancelto{0}{\mathcal{L}(t-u)}\right) dtdu$$

While the RC mechanism purports to dampen volatility as we have seen it does not completely eliminate intertemporal covariance as above. In section 4.2.4 we used a GBM has no such dependence. We believe this explains the extreme and suspicious narrowing of the distributions in section 4.2.4. In the current state of the art as demonstrated in section 4.2.1 we can fit the statistics of the resultant RC index directly, but if we want to produce an RC index by passing it's constituent components through a model we need an accurate model for the stochastic volatility that can interact with the RC mechanism.

In the following section we consider two modifications of our models in 4.3.2 and 4.3.2 and attempt to determine what are the necessary features of an accurate stochastic volatility model.

Candidate models for Stochastic Volatility

If true our hypothesis would suggest that any economy/underlying simulation must properly capture the proper decay of volatility autocorrelation such that the RC mechanism doesn't overpower it and deliver unrealistic results like in section 4.2.4. We will test this using the following two models.

Model 1: A FBM model We will employ the model of Gatheral et al. (2018) a version of which we studied in section 4.3.2. We make the change that $f(\sigma) = e^{\sigma}$ as this is more realistic and empirically sound. In discrete daily returns the model takes the form:

$$\begin{aligned} r_{t_i} &= \mu\Delta t + e^{v_{t_i}}\sqrt{\Delta t}\Delta W_{t_i} \\ v_{t_i} &= v_{t_{i-1}} + \alpha(m - v_{t_{i-1}})\Delta t + \lambda\Delta W_{t_i}^H \\ \rho\Delta t(\Delta t)^{2H} &= \mathbb{E}[\Delta W_{t_i}\Delta W_{t_i}^H] \end{aligned} \tag{4.20}$$

Unfortunately estimation within this model is made very complicated by the fractional term. We simply used the parameters found for the SPX in Gatheral et al. (2018). Simulation of an FBM is also difficult to do through some kind of Euler style integration. We prefer to simulate the FBM through spectral methods for accuracy (see Banna et al. (2019)), this means that including a ρ and therefore leverage correlation is very difficult and we will have to eschew this for now. Though as we shall see this will not effect our conclusions.

Model 2: A Moving average model Let us say Δt is one day such that for an N day simulation horizon we have $t - t_0 = N\Delta t$ and $t_0 = t_{i-N} < ... < t_{i-j} < ... < t_i = t$. For a volatility kernel in (4.8) let $k(t_i, t_{i-j}) = k_{i-j}$ where $k_{i-j} \approx 0$ for some $j > q$. We can then make the following approximation for a discrete daily volatility model:

$$\int_{t_0}^{t} k(t, t')dW_t^{(2)} \approx \sum_{j=1}^{q} k(t_i, t_{i-j})\Delta W_{t_j}^{(2)}$$

Notice this is functionally equivalent to a moving average model $MA(q)$ (see Kleiner (1977)). And so fitting the following model with built in MA tools is straightforward:

$$\begin{aligned} r_{t_i} &= \mu\Delta t + e^{v_{t_i}}\sqrt{\Delta t}\Delta W_{t_i}^{(1)} \\ v_{t_i} &= m + \sum_{j=0}^{q} k_{i-j}\Delta W_{t_{i-j}}^{(2)} \\ \rho(\Delta t)^2 &= \mathbb{E}[\Delta W_{t_i}^{(1)}\Delta W_{t_i}^{(2)}] \end{aligned} \tag{4.21}$$

We in fact do this for the SPX and compare the subsequent return distributions in figure 4.19. We can see that the $MA(50)$ volatility model can better reproduce real SPX returns. Notably in the tails of the return distribution where the deviations from normality are driven by high stochastic volatility events.

The question we will try to answer in the next section is how much does this fidelity to the return distribution help when we suspect the real driver of aggregated RC returns is autocorrelatrion. The MA model will by definition only have a exponentially decaying

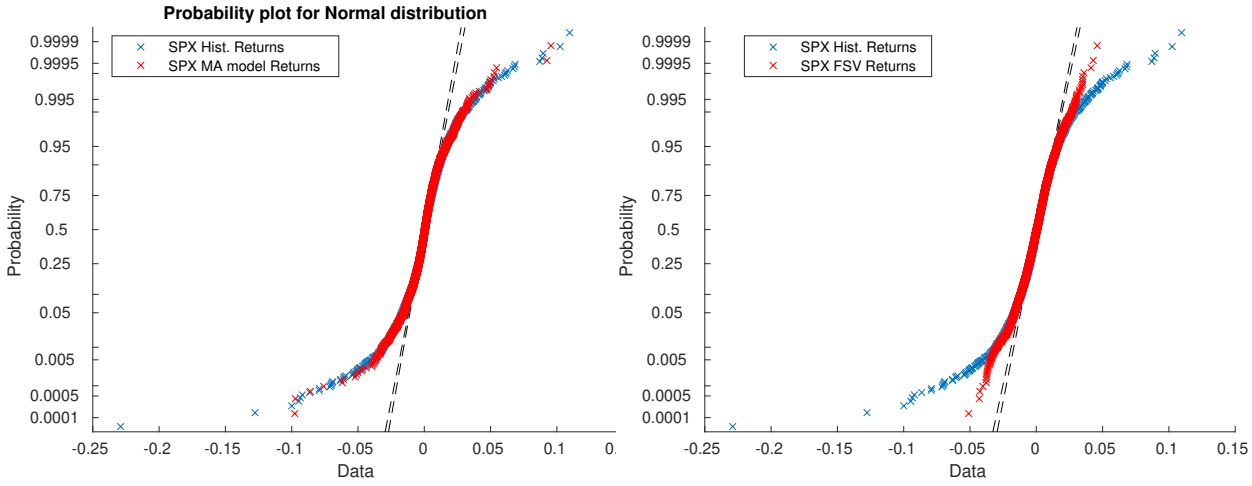


Figure 4.19: As we can see the MA model appears to better reproduce produce the tails of the empirical distribution of returns.

volatility for 50 days. Whereas the FBM model may not fit the returns as well but has a persistent power law autocorrelation that may not be overpowered by the RC mechanism.

Simulation Results

We proceeded by simulating an economy with an SPX component given by equations (4.20) and (4.21) and bond component given by I00078US similar to section 4.2.4. We have compiled the results in table 4.8. Contrary to the discussion[14] in 4.3.2 on the scale of a decade we found the α parameter is actually crucial in calibrating returns. This makes sense as $\alpha = 0.14$ gives an approximate characteristic scale of 5 years. Which is obviously relevant for a 10 year FIA. Recall we discussed in section 4.2.1 the prudence (or not) of using historical RC index results as a benchmark. That said, we can see that a properly calibrated FBM model can capture the standard deviation of historical FIA returns. This despite the fact that they may have a higher skewness; possibly an artifact of no leverage correlation.

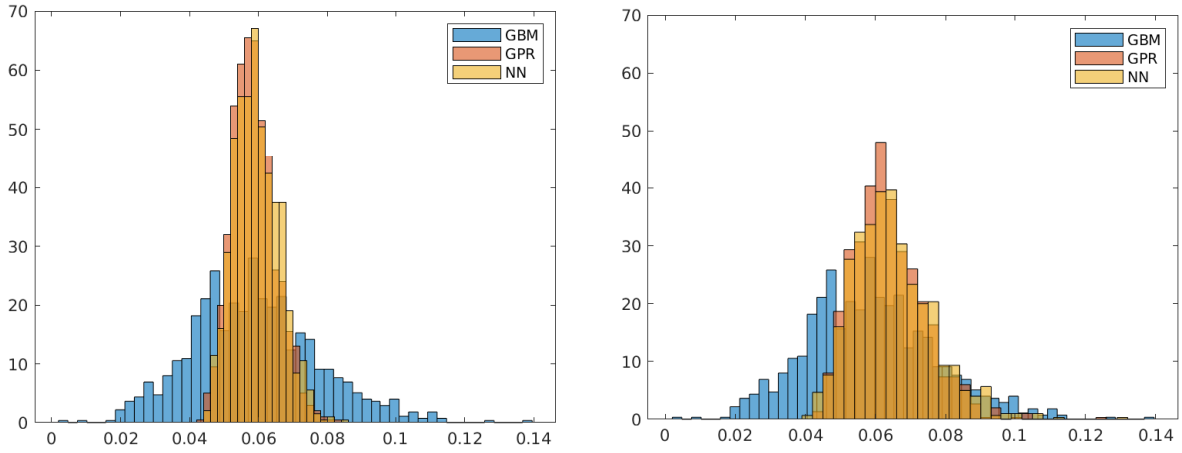
[14]Inspired by the results of Gatheral et al. (2018)

		GBM	Real Returns
Direct RC index model	Mean	0.059	0.060
	Std.	0.019	0.016
	Skewness	0.421	0.322

SPX model		GPR	NN
GBM	Mean	0.054	0.054
	Std.	0.006	0.006
	Skewness	0.394	0.444
MA model ($\rho = -0.4$)	Mean	0.059	0.059
	Std.	0.006	0.006
	Skewness	0.454	0.391
FSV ($H = 0.14$, $\alpha = 0$, $\rho = 0$)	Mean	0.071	0.072
	Std.	0.025	0.029
	Skewness	2.969	2.927
FSV ($H = 0.14$, $\alpha = 0.18$, $\rho = 0$)	Mean	0.064	0.065
	Std.	0.010	0.011
	Skewness	0.850	0.824

Table 4.8: Annuity Return Statistics using and SPX/I00078US GPR model with various stochastic volatility models for SPX

On the other hand the MA(50) model delivers results that are essentially identical to the GBM model studied in section 4.2. Examining Figure 4.20 we can see the extreme narrowing of the annuity return distribution for the MA(50) model annuity returns versus the more heavy tailed FIA return distribution. This despite the fact the tails of the underlying returns better match the underlying in the MA case! This would appear to validate our hypothesis: the RC mechanism interacts with the more realistic power law decay of a FBM model in the proper way whereas the MA model will decay too quickly and the resulting model will be unrealistically narrow in the aggregate like in the GBM case.



(a) (b)

Figure 4.20: Stochastic volatility is very relevant when considering aggregated returns. The MA model (a) behaves very much like our original constant vol model whereas the rough model (b) appears more realistic.

4.4 Conclusion

The main focus of this work was to provide a blueprint for solving the annuity decision problem practically and at scale through simulations. The number of currently employed annuity indices makes direct simulation problematic given the complex dependence structure between them. Furthermore, even within the comparatively simple class of RC indices, direct modelling of the index methodology remains elusive. We were, however, successful in showing that even basic ML applications can outperform more elaborate direct modeling of indices. For instance, ML models seemed to reproduce the index volatility calculations (e.g EWMA) better that the publicly listed methodology. In fact, we showed that we don't even need the constituent underlying assets as features when using the CORE methodology. If the index components are unknown, our CORE methodology proves to be a superior approach in many cases, not simply a healthy alternative. However, outside the class of RC indices, the MARC methodology tended to fail. For example, for MARC5, we had to

add gold and the Barclays index we studied appeared too complex or had novel exposure. Extremely complex indices aside, the fact that there exist large classes of indices where we could pick up and reproduce more reliable returns via machine learning made the exercise worthwhile as a whole. At the very least, this provides a viable avenue for index forecasting where transparency of the index methodology is low.

In fact, the ML methods we used were in some sense "too accurate" as they reproduced the RC mechanisms with enough fidelity that a GBM-based underlying economy model was too simple. Our ML-trained index functions we so good that they "over-reacted" to the random nature of GBM paths. They were anticipating time dependence of volatility and when receiving none "over-dampened" random jumps. This necessitated a more thorough investigation into the relationship between stochastic volatility and the induced return dependence therein and the aggregate distribution of index returns. We identified the major features of stochastic volatility (autocorrelation and leverage effects) and found a way of expressing the variance of aggregate returns with respect to them. While a full accounting of the dependence between returns would require more, we felt that this was sufficient, given the nature of RC calculations. We then explored what kinds of stochastic volatility models would work well together with our ML index models and found that the 'rough' autocorrelation of volatility seems to play a crucial role in in explaining the behavior of RC indices.

This naturally leads us to the first avenue for future research: constructing a rough stochastic volatility economy that will work with our ML models to accurately model the annuity decision problem. Another natural next step is to try and estimate some of the more complex indices. Regardless of volatility models, a more informed choice of the underlying economy (e.g. adding commodities, different duration and credit quality when it comes to fixed income, and some alternatives) would provide much richer data set to train our ML models. Absent that, more complex ML techniques may help. The novelty and complexity of some index strategies may warrant techniques like Natural Language Processing or adversarial learning etc.

Aside from the technical details of modelling, there are products outside FIAs this work can be applied to. Registered index-linked annuities, or RILAs, are very similar to FIAs

but contain a loss profile (see Moenig (2022a)). In terms of product analysis, extending our methodology to help with evaluation and comparison of RILAs is straightforward but possibly even more informative. Because they offer different downside or loss profiles, the option pricing stability arising from RC indices offer genuine benefits but generally suffer from low transparency.; understanding how to forecast these properly could be critical for the industry.

Because RC indices have cheaper options, they also offer participation rates that may exceed 100%-200% that are attractive to clients and are resistant to variation based on prevailing market volatility. Furthermore, the recent rise in interest rates allows for richer option budgets that all index-linked annuities depend upon, including both FIAs and RILAs. Therefore, it is important to understand and be able to compare indexed annuities linked to different, sometimes less familiar indices. Due to their stable and relatively low volatility, RC indices allow for more predictable option pricing and, in a rising interest rate environment, offer attractive participation rates. As it is impossible to research, deconstruct and model all of them, we believe that our methodology opens the door to another way of simulating these products while enabling inclusion of many different structures struck on many different indices.

References

Agram, N. & Øksendal, B. (2019), 'Introduction to white noise, hida-malliavin calculus and applications', arXiv preprint arXiv:1903.02936 .

Albrecht, P. & Maurer, R. (2001), 'Self Annuitization, Ruin Risk in Retirement and Asset Allocation: The Annuity Benchmark'. Working Paper available at http://www.actuaries.org/AFIR/Colloquia/Toronto/Albrecht_Maurer.pdf.

Alexandrova, M., Bohnert, A., Gatzert, N. & Russ, J. (2017), 'Equity-linked life insurance based on traditional products: the case of select products', European Actuarial Journal **7**(2), 379–404.

Alexandrova, M. & Gatzert, N. (2019), 'What do we know about annuitization decisions?', Risk Management and Insurance Review **22**(1), 57–100.

Alòs, E. & Lorite, D. G. (2021), Malliavin calculus in finance: Theory and practice, CRC Press.

B., G. (1825), 'On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies', Philosophical Transactions of the Royal Society **vol. 115**, 513–585.

Babbel, D. F., VanderPal, G. & Marrion, J. (2010), 'Real world index annuity returns', Available at SSRN 1482023 .

Babbel, D. & Merrill, C. (2007), 'Rational Decumulation'. Working Paper, Wharton Financial Institutions Center. Available at http://fic.wharton.upenn.edu/fic/papers/06/0614.pdf.

Baltussen, G., van Bekkum, S. & Da, Z. (2019), 'Indexing and stock market serial dependence around the world', Journal of Financial Economics **132**(1), 26–48.

Banna, O., Mishura, Y., Ralchenko, K. & Shklyar, S. (2019), Fractional Brownian motion: approximations and projections, John Wiley & Sons.

Barclays Trailblazer Sectors 5 Index (Factsheet) (2022), https://indices.cib.barclays/dms/Public%20marketing/BXIITBZ5_2_pager_weekly.pdf, note = Accessed: 24/11/2022.

Barndorff-Nielsen, O. E. & Shephard, N. (2001), 'Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(2), 167–241.

Bauer, D., Kling, A. & Russ, J. (2008), 'A universal pricing framework for guaranteed minimum benefits in variable annuities,', ASTIN Bulletin: The Journal of the IAA **38**(2), 621–651.

Bayraktar, E. & Young, V. (2007), 'Correspondence between Lifetime Minimum Wealth and Utility of Consumption', Finance and Stochastics **vol. 11(2)**(2), 213–236.

Benartzi, S., Previtero, A. & Thaler, R. H. (2011), 'Annuitization puzzles', Journal of Economic Perspectives **25**(4), 143–164.

Bengen, W. (1994), 'Determining Withdrawal Rates Using Historical Data', Journal of Financial Planning **vol. 7(4)**(no. 4), 171–181.

Bernard, C. & Boyle, P. P. (2011), 'A natural hedge for equity indexed annuities', Annals of Actuarial Science **5**(2), 211–230.

Blake, D. & Burrows, W. (2001), 'Survivor Bonds: Helping to Hedge Mortality Risk', The Journal of Risk and Insurance **vol. 68(2)**(no. 2), 339–348.

Blanchett, D. (2015), 'Dynamic Choice and Optimal Annuitization', The Journal of Retirement **3(1)**(1), 38–49. https://doi.org/10.3905/jor.2015.3.1.038.

Blanchett, D., Finke, M. & Nikolic, B. (2021), 'How competitive are income annuity providers over time?', Risk Management and Insurance Review **24**(1), 207–214.

Blanchett, D., Kowara, M. & Chen, P. (2012), 'Optimal Withdrawal Strategy for Retirement Income Portfolios'. http://corporate.morningstar.com/ib/documents/methodologydocuments/optimalwithdrawlstrategyretirementincomeportfolios.pdf.

Bochud, T. & Challet, D. (2007), 'Optimal approximations of power laws with exponentials: application to volatility models with long memory', Quantitative Finance **7**(6), 585–589.

Bodie, Z. (1990), 'Pensions as retirement income insurance', Journal of Economic Literature **28**, 28–49.

Bodie, Z., Detemple, J., Ortuba, S. & Walter, S. (2004), 'Optimal Consumption-Portfolio Choice and Retirement Planning', Journal of Economic Dynamics & Control **vol. 28(6)**(no. 6), 1115–1148.

Bommier, A. & Le Grand, F. (2014), 'Too risk averse to purchase insurance?', Journal of Risk and Uncertainty **48**(2), 135–166.

Bouchaud, J.-P., Potters, M. & Meyer, M. (2000), 'Apparent multifractality in financial time series', The European Physical Journal B-Condensed Matter and Complex Systems **13**(3), 595–599.

Bouchaud, J.-P., Potters, M. et al. (2003), Theory of financial risk and derivative pricing: from statistical physics to risk management, 2nd edn, Cambridge university press.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. & Nesbitt, C. (1997), Actuarial Mathematics, Society of Actuaries, Schaumburg, IL.

Boyer, M. M., Box-Couillard, S. & Michaud, P. C. (2020), 'Demand for annuities: Price sensitivity, risk perceptions, and knowledge', Journal of Economic Behavior and Organization **180**, 883–902.

Boyle, P. & Tian, W. (2008), 'The design of equity-indexed annuities', Insurance: Mathematics and Economics **43**(3), 303–315.

Brown, J., Kling, J., Mullainathan, S. & Wrobel, M. (2008), 'Why don't people insure late-life consumption? a framing explanation of the under-annuitization puzzle', American Economic Review **98**(2), 304–309.

Brown, J. R., Mitchell, O., Poterba, J. & Warshawsky, M. (2001), The Role of Annuity Markets in Financing Retirement, MIT Press, Cambridge.

Browne, S., Milevsky, M. & Salisbury, T. (2003), 'Asset Allocation and the Liquidity Premium for Illiquid Annuities', The Journal of Risk and Insurance **vol. 70(3)**(no. 3), 509–526.

Bun, J., Bouchaud, J.-P. & Potters, M. (2017), 'Cleaning large correlation matrices: tools from random matrix theory', Physics Reports **666**, 1–109.

Campbell, J. & Viceira, L. (2002), Strategic Asset Allocation – Portfolio Choice for the Long-Term Inves Oxford University Press, New York.

Cannon, E. & Tonks, I. (2008), Annuity Markets, Oxford University Press, New York.

Carr, P. & Cherubini, U. (2022), 'Generalized compounding and growth optimal portfolios reconciling kelly and samuelson', The Journal of Derivatives **30**(2), 74–93.

Carrier, G. & Pearson, C. (1976), Partial Differential Equations: Theory and Technique, Academic Press, New York.

Chai, J., Horneff, W., Maurer, R. & Mitchell, O. S. (2011), 'Optimal portfolio choice over the life cycle with flexible work, endogenous retirement, and lifetime payouts', Review of Finance **15**(4), 875–907.

Charupat, N., Kamstra, M. J. & Milevsky, M. A. (2016), 'The sluggish and asymmetric reaction of life annuity prices to changes in interest rates', Journal of Risk and Insurance **83**(3), 519–555.

Chen, J.-J., Zheng, B. & Tan, L. (2013), 'Agent-based model with asymmetric trading and herding for complex financial systems', PloS one **8**(11), e79531.

Chen, P., Ibbotson, R., Milevsky, M. & Zhu, K. (2006), 'Human Capital, Asset Allocation, and Life Insurance', Financial Analysts Journal **vol. 62(1)**(no. 1), 97–109.

Cheridito, P., Kawaguchi, H. & Maejima, M. (2003), 'Fractional ornstein-uhlenbeck processes', Electronic Journal of probability **8**, 1–14.

Clark, S. P. & Dickson, M. (2021), 'Fixed indexed annuities: Dissecting performance expectations', Retirement Management Journal **10**(1), 36–45.

Cocco, J., Gomes, F. & Maenhout, P. (2005), 'Portfolio Choice over the Life-Cycle', Review of Financial Studies **18(2)**(2), 491–533.

Comte, F. & Renault, E. (1998), 'Long memory in continuous-time stochastic volatility models', Mathematical finance **8**(4), 291–323.

Cont, R. (2001), 'Empirical properties of asset returns: stylized facts and statistical issues', Quantitative finance **1**(2), 223.

Cooley, P., Hubbard, C. & Walz, D. (1998), 'Retirement Savings: Choosing a Withdrawal Rate That Is Sustainable', AAII Journal **vol. 20(2)**(no. 2), 16–21.

Daniel, G., Joseph*, N. L. & Brée, D. S. (2005), 'Stochastic volatility and the goodness-of-fit of the heston model', Quantitative Finance **5**(2), 199–211.

Davidoff, T., Brown, J. & Diamond, P. (2005), 'Annuities and individual welfare', American Economic Review **95**(5), 1573–1590.

Davies, J. (1981), 'Uncertain lifetime, consumption and dissaving in retirement', Journal of Political Economy **89**, 561–577.

Dempster, M., Germano, M., Medova, E., Rietbergen, M., Sandrini, F. & Scrowston, M. (2007), 'Designing Minimum Guaranteed Return Funds', Quantitative Finance **vol. 7(2)**(no. 2), 245–256. http://dx.doi.org/10.1080/14697680701264804.

Dickson, D., Hardy, M. & Waters, H. R. (2009), Actuarial mathematics for life contingent risks, Cambridge University Press, U.K.

Dillschneider, Y., Mauer, R. & Schober, P. (2020), 'Dynamic Portfolio Choice with Annuities When the Interest Rate Is Stochastic', TBD **vol. 90**(TBD), TBD.

Domingos, P. (2020), 'Every model learned by gradient descent is approximately a kernel machine', arXiv preprint arXiv:2012.00152 .

Dybvig, P. & Liu, H. (2005), 'Lifetime Consumption and Investment: Retirement and Constrained Borrowing'. Working Paper, Washington University in St. Louis.

Employee Benefit Research Institute (2014), 'What are the trends in U.S. retirement plans?'. Retrieved from https://www.ebri.org/publications/benfaq/index.cfm?fa=retfaq14.

Feldstein, M. & Ranguelova, E. (2001), 'Individual Risk in an Investment-Based Social Security System', American Economic Review **vol. 91(4)**(no. 4), 1116–1125.

Feng, L., Li, B., Podobnik, B., Preis, T. & Stanley, H. E. (2012), 'Linking agent-based models and stochastic models of financial markets', Proceedings of the National Academy of Sciences **109**(22), 8388–8393.

Finkelstein, A. & Poterba, J. (2004), 'Adverse selection in insurance markets: Policyholder evidence from the u.k. annuity market', Journal of Political Economy **112**(1), 183–208.

Finkelstein, A. & Poterba, J. (2014), 'Testing for asymmetric information using unused observables in insurance markets: Evidence from the uk annuity market', Journal of Risk and Insurance **81**(4), 709–734.

Fisher, I. (1930), The Theory of Interest: As Determined by Impatience to Spend Income and Opportun Macmillan, New York.

Friedman, B. M. & Warshawsky, M. J. (1990), 'The cost of annuities: Implications for saving behavior and bequests', The Quarterly Journal of Economics **105**(1), 135–154.

Friend, I. & Blume, M. (1975), 'The Demand for Risky Assets', The American Economic Review **vol. 65(5)**(no. 5), 900–922.

G., S. (2006), 'Optimal Timing of the Annuity Purchase: Combined Stochastic Control and Optimal Stopping Problem', International Journal of Theoretical and Applied Finance **vol. 9(2)**(no. 2), 151–170.

Gaillardetz, P. & Lakhmiri, J. Y. (2011), 'A new premium principle for equity-indexed annuities', Journal of Risk and Insurance **78**(1), 245–265.

Gaillardetz, P., Li, H. Y. & MacKay, A. (2012), 'Equity-linked products: evaluation of the dynamic hedging errors under stochastic mortality', European Actuarial Journal **2**(2), 243–258.

Gatheral, J., Jaisson, T. & Rosenbaum, M. (2018), 'Volatility is rough', Quantitative finance **18**(6), 933–949.

Geoffrey VanderPal, D. (2011), 'Real-world index annuity returns', Journal of Financial Planning **24**(3), 50.

Gerrard, R., Haberman, S. & Vigna, E. (2006), 'The Management of Decumulation Risks in a Defined Contribution Pension Plan', North American Actuarial Journal **10(1)**(1), 84–110.

Gompertz, B. (1825), 'On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies', Philosophical Transactions of the Royal Society of London **115**, 513–583.

Gong, G. & Webb, A. (2010), 'Evaluating the advanced life deferred annuity: An annuity people might actually buy', Insurance: Mathematics and Economics **46**(1), 210–221.

Guyon, J. & Lekeufack, J. (2022), 'Volatility is (mostly) path-dependent', Available at SSRN: https://ssrn.com/abstract=4174589 or http://dx.doi.org/10.2139/ssrn.4174589 .

Haberman, S. & Sibbett, T., eds (1995), History of Actuarial Science (10 volumes), William Pickering, London.

Habib, F., Huang, H., Mauskopf, A., Nikolic, B. & Salisbury, T. (2020), 'Optimal allocation to deferred income annuities', Insurance: Mathematics and Economics **90**, 94–104.

Habib, F., Huang, H. & Milevsky, M. (2017), 'Approximate Solutions to Retirement Spending Problems and the Optimality of Ruin'. Available at SSRN: https://ssrn.com/abstract=2944125 or http://dx.doi.org/10.2139/ssrn.2944125.

Hainuat, D. & Deelstra, G. (2014), 'Optimal Timing for Annuitization, based on Jump-Diffusion Fund and Stochastic Mortality', Journal of Economic Dynamics and Control **44**, 124–146.

Halley, E. (1693), 'An estimate of the degrees of the mortality of mankind, drawn from the curious tables of the births and funerals at the city of breslaw', Philosophical Transactions of the Royal Society of London **17**, 596–610.

Hänggi, P. (1985), 'Stochastic processes applied to physics', Philadelphia: Heiden .

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', The review of financial studies **6**(2), 327–343.

Hida, T. (1976), Analysis of brownian functionals, in 'Stochastic Systems: Modeling, Identification and Optimization, I', Springer, pp. 53–59.

Hinton, G. E. (1990), Connectionist learning procedures, in 'Machine learning', Elsevier, pp. 555–610.

Hochreiter, S. (1998), 'The vanishing gradient problem during learning recurrent neural nets and problem solutions', International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **6**(02), 107–116.

Holden, H., Øksendal, B., Ubøe, J. & Zhang, T. (2009), Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach, Springer Science & Business Media.

Horneff, V., Maurer, R. & Mitchell, O. S. (2020), 'Putting the pension back in 401 (k) retirement plans: Optimal versus default deferred longevity income annuities', Journal of Banking and Finance (in press) **414**.

Horneff, W. J., Maurer, R. H., Mitchell, O. S. & Stamos, M. Z. (2009a), 'Asset allocation and location over the life cycle with investment-linked survival-contingent payouts', Journal of Banking and Finance **33**(9), 1688–1699.

Horneff, W., Maurer, R., Mitchell, O. & Stamos, M. (2009b), 'Asset Allocation and Location over the Life Cycle with Investment-Linked Survival-Contingent Payouts', Journal of Banking and Finance **vol. 33(9)**(9), 1688–1699.

Huang, H. H., Zhang, S. & Zhu, W. (2017), 'Limited participation under ambiguity of correlation', Journal of Financial Markets **32**, 97–143.

Huang, H., Milevsky, M. & Young, V. (2017), 'Optimal purchasing of deferred income annuities when payout yields are mean-reverting', Review of Finance **21**(1), 327–361.

Huang, Y. T., Zeng, P. & Kwok, Y. K. (2017), 'Optimal initiation of guaranteed lifelong withdrawal benefit with dynamic withdrawals', SIAM Journal on Financial Mathematics **8**(1), 804–840.

Inkmann, J., Lopes, P. & Michaelides, A. (2011), 'How deep is the annuity market participation puzzle?', The Review of Financial Studies **24**(1), 279–319.

Israelov, R. & Klein, M. (2016), 'Risk and return of equity index collar strategies', The Journal of Alternative Investments **19**(1), 41–54.

Jimenez-Martin, S. & Sanchez-Martin, A. (2007), 'An Evaluation of the Life Cycle Effects of Minimum Pensions on Retirement Behavior', Journal of Applied Econometrics **vol. 22(5)**(no. 5), 923–950.

J.P.Morgan Asset Management (2015), '2016 Long-Term Capital Market Assumptions'. Retrieved from https://am.jpmorgan.com/gi/getdoc/1383271688187.

J.P.Morgan Asset Management (2017), '2018 Long-Term Capital Market Assumptions'. Retrieved from https://am.jpmorgan.com/gi/getdoc/1383498280832.

Kim, D. & Francis, J. C. (2013), Modern portfolio theory: Foundations, analysis, and new developments, John Wiley & Sons.

Kingston, G. & Thorp, S. J. (2005), 'Annuitization and asset allocation with hara utility', Journal of Pension Economics and Finance **4**(3), 225–248.

Kleiner, B. (1977), 'Time series analysis: Forecasting and control'.

Knoller, C. (2016), 'Multiple reference points and the demand for principal-protected life annuities: An experimental analysis', Journal of Risk and Insurance **83**(1), 163–179.

Koijen, R., Nijman, T. & Werker, B. (2011a), 'Optimal Annuity Risk Management', Review of Finance **15(4)**(4), 799–833.

Koijen, R. S., Nijman, T. E. & Werker, B. J. (2011b), 'Optimal annuity risk management', Review of Finance **15**(4), 799–833.

Konicz, A., Pisinger, D. & Weissensteiner, A. (2016), 'Optimal Retirement Planning with a Focus on Single and Joint Life Annuities', Quantitative Finance **vol. 16(2)**(no. 2), 275–295. http://dx.doi.org/10.1080/14697688.2015.1114361.

LeBaron, B. (2001), 'Stochastic volatility as a simple generator of apparent financial power laws and long memory', Quantitative finance **1**(6), 621.

Lee, D. R. & Carter, L. (1992), 'Modeling and Forecasting U. S. Mortality', Journal of the American Statistical Association **vol. 87(419)**(no. 419), 659–671.

Leung, S. (1994), 'Uncertain Lifetime, the Theory of the Consumer, and the Life Cycle Hypothesis', Econometrica **vol. 62(5)**(no. 5), 1233–1239.

Li, M., Gençay, R. & Xue, Y. (2016), 'Is it brownian or fractional brownian motion?', Economics Letters **145**, 52–55.

Lilley, M. & Frean, M. (2005), Neural networks: a replacement for gaussian processes?, in 'International Conference on Intelligent Data Engineering and Automated Learning', Springer, pp. 195–202.

Lim, B. & Shin, Y. (2010), 'Optimal Investment, Consumption and Retirement Decision with Disutility and Borrowing Constraints', Quantitative Finance **vol. 11(10)**(no. 10), 1581–1592. http://dx.doi.org/10.1080/14697680903369526.

Lin, X. S., Tan, K. S. & Yang, H. (2009), 'Pricing annuity guarantees under a regime-switching model', North American Actuarial Journal **13**(3), 316–332.

Lions, P. L. & Sznitman, A. S. (1984), 'Stochastic differential equations with reflecting boundary conditions', Communications on Pure and Applied Mathematics **37**(4), 511–537.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160370408*

Liu, J. & Zeng, X. (2017), 'Correlation ambiguity and under-diversification', Available at SSRN 2692692 .

MacDonold, B., Jones, B., Morrison, R., Brown, R. & Hardy, M. (2013), 'Research and Reality - A Literature Review on Drawing Down Retirement Financial Savings', North American Actuarial Journal **vol. 17(3)**(no. 3), 181–215.

MacKay, A., Augustyniak, M., Bernard, C. & Hardy, M. R. (2017), 'Risk management of policyholder behavior in equity-linked life insurance', Journal of Risk and Insurance **84**(2), 661–690.

Makeham, W. (1860), 'On the Law of Mortality and the Construction of Annuity Tables', Journal of the Institute of Actuaries and Assurance **vol. 8**, 301–310.

Mandelbrot, B. B., Fisher, A. J. & Calvet, L. E. (1997), 'A multifractal model of asset returns'.

Mandelbrot, B. B. & Van Ness, J. W. (1968), 'Fractional brownian motions, fractional noises and applications', SIAM review **10**(4), 422–437.

Mankiw, N. & Zeldas, S. (1991), 'The Consumption of Stockholders & Nonstockholders', The Journal of Financial Economics **vol. 29**, 97–112.

Masoliver, J. & Perelló, J. (2006), 'Multiple time scales and the exponential ornstein–uhlenbeck stochastic volatility model', Quantitative Finance **6**(5), 423–433.

Medova, E., Murphy, J., Owen, A. & Rehman, K. (2008), 'Individual Asset Liability Management', Quantitative Finance **vol. 8(6)**(no. 6), 547–560. http://dx.doi.org/10.1080/14697680802402691.

Merton, R. (1971), 'Optimum Consumption and Portfolio Rules in a Continuous-Time Model', Journal of Economic Theory **vol. 3(4)**(no. 4), 373–413.

Merton, R. C. (1969), 'Lifetime portfolio selection under uncertainty: The continuous-time case', The review of Economics and Statistics pp. 247–257.

Merton, R. C. (2014), 'The crisis in retirement planning', Harvard Business Review **92**(7), 43–50.

Milevsky, M. A. (2006), The Calculus of Retirement Income: Financial Models for Pension Annuities and Life Insurance, Cambridge University Press, New York.

Milevsky, M. A. & Salisbury, T. S. (2006), 'Financial valuation of guaranteed minimum withdrawal benefits', Insurance: Mathematics and Economics **38**, 21–38.

Milevsky, M. A. & Salisbury, T. S. (2022), 'Refundable income annuities: Feasibility of money-back guarantees', Insurance, mathematics & economics **105**, 175–193.

Milevsky, M. & Huang, H. (2011), 'Spending Retirement on Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates', Financial Analysts Journal **vol. 67(2)**(no. 2), 45–58.

Milevsky, M., Moore, K. & Young, V. (2006), 'Asset Allocation and Annuity-Purchase Strategies to Minimize the Probability of Financial Ruin', Mathematical Finance **16(4)**(4), 647–671.

Milevsky, M. & Robinson, C. (2005), 'A Sustainable Spending Rate without Simulation', Financial Analysts Journal **vol. 61(6)**(no. 6), 89–100.

Milevsky, M. & Young, V. (2007), 'Annuitization and Asset Allocation', Journal of Economic Dynamics and Control **vol. 31(9)**(no. 9), 3138–3177.

Mitchel, O., Poterba, J., Warshawsky, M. & Brown, J. (1999), 'New Evidence on the Money's Worth of Individual Annuities', American Economic Review **vol. 89(5)**(no. 5), 1299–1318.

Mitchell, O. S., Poterba, J. M., Warshawsky, M. J. & Brown, J. R. (1999), 'New evidence on the money's worth of individual annuities', American Economic Review **89**(5), 1299–1318.

Modigliani, F. (1988), 'The role of intergenerational transfers and life cycle saving in the accumulation of wealth', Journal of Economic Perspectives **2**(2), 15–40.

Moenig, T. (2022a), 'It's rila time: An introduction to registered index-linked annuities', Journal of Risk and Insurance **89**(2), 339–369.

Moenig, T. (2022b), 'It's rila time: An introduction to registered index-linked annuities', Journal of Risk and Insurance **89**(2), 339–369.

Moenig, T. & Bauer, D. (2016), 'Revisiting the risk-neutral approach to optimal policyholder behavior: A study of withdrawal guarantees in variable annuities', Review of Finance **20**(2), 759–794.

Moenig, T. & Xu, C. (2022), 'Valuing lifetime withdrawal guarantees in rilas', Available at SSRN 4021905 .

Moore, K. S. (2009), 'Optimal surrender strategies for equity-indexed annuity investors', Insurance: Mathematics and Economics **44**(1), 1–18.

Moore, K. S. & Young, V. R. (2003), 'Pricing equity-linked pure endowments via the principle of equivalent utility', Insurance: Mathematics and Economics **33**(3), 497–516.

Morton, K. & Mayer, D. (2005), Numerical Solution of Partial Differential Equations: An Introduction, Cambridge University Press, Cambridge, UK.

NAFA (2022), 'Helping consumers understand fias', https://nafa.com/wp-content/uploads/NAFA_EduSeries_WhyFIA.pdf. Accessed: 12/13/2022.

NAIC (2022), 'Naic model laws', https://content.naic.org/cipr-topics/naic-model-laws. Accessed: 24/11/2022.

Neuberger, A. (2003), 'Annuities and the Optimal Investment Decision'.

Ngai, A. & Sherris, M. (2011), 'Longevity risk management for life and variable annuities: The effectiveness of static hedging using longevity bonds and derivatives', Insurance: Mathematics and Economics **49**(1), 100–114.

Nielsen, M. A. (2015), Neural networks and deep learning, Vol. 25, Determination press San Francisco, CA, USA.

Nikolic, B., Toland, T. & Baboolal, D. (2018), 'Accumulation value of fixed annuities (myga & fia): Understanding yields by product design [white paper]', Cannex Financial Exchanges Limited .

Novikov, E. A. (1965), 'Functionals and the random-force method in turbulence theory', Sov. Phys. JETP **20**(5), 1290–1294.

Orlando, G., Bufalo, M. & Stoop, R. (2022), 'Financial markets' deterministic aspects modeled by a low-dimensional equation', Scientific reports **12**(1), 1–13.

Pashchenko, S. (2013), 'Accounting for non-annuitization', Journal of Public Economics **98**, 53–67.

Paulson, M. & Goin, G. (2021), 'Controllable factors and uncontrollable risks in a retirement spending plan: Using a registered indexed linked annuity to mitigate risk exposure', Retirement Management Journal **10**(1), 59–66.

Perelló, J., Masoliver, J. & Bouchaud, J.-P. (2004), 'Multiple time scales in volatility and leverage correlations: a stochastic volatility model', Applied Mathematical Finance **11**(1), 27–50.

Perelló, J., Sircar, R. & Masoliver, J. (2008), 'Option pricing under stochastic volatility: the exponential ornstein–uhlenbeck model', Journal of Statistical Mechanics: Theory and Experiment **2008**(06), P06010.

Perry, P. R. (1982), 'The time-variance relationship of security returns: Implications for the return-generating stochastic process', The Journal of Finance **37**(3), 857–870.

Peters, O. (2019), 'The ergodicity problem in economics', Nature Physics **15**(12), 1216–1221.

Pfau, W. & Kitces, M. (2014), 'Reducing Retirement Risk with a Rising Equity Glide Path', Journal of Financial Planning **vol. 27(1)**(1), 38–45.

Pitacco, E. (2016), 'Guarantee structures in life annuities: A comparative analysis', The Geneva Papers on Risk and Insurance-Issues and Practice **41**(1), 78–97.

Poterba, J. (2014), 'Retirement security in an aging population', American Economic Review **104**(5), 1–30.

Poterba, J. M. (1997), 'The history of annuities in the united states', National Bureau of Economic Research (NBER), Working Paper **6001**.

Poterba, J. M. & Solomon, A. (2021), 'Discount rates, mortality projections, and money's worth calculations for us individual annuities', National Bureau of Economic Research (NBER), Working Paper **28557**.

Promislow, S. D. (2006), Fundamentals of actuarial mathematics, John Wiley & Sons, Toronto.

Pye, G. (2012), The Retrenchment Rule: When It's Too Late to Save More for Retirement, GBP Press, New York.

Que, Q. & Belkin, M. (2016), Back to the future: Radial basis function networks revisited, in 'Artificial intelligence and statistics', PMLR, pp. 1375–1383.

Ramsay, C. M. & Oguledo, V. I. (2018), 'The annuity puzzle and an outline of its solution', North American actuarial journal **22**(4), 623–645.

Reichling, F. & Smetters, K. (2015), 'Optimal annuitization with stochastic mortality and correlated medical costs', American Economic Review **105**(11), 3273–3320.

Rogers, L. (2013), Optimal Investment, Springer Press, Cambridge.

Safran, I. & Shamir, O. (2017), Depth-width tradeoffs in approximating natural functions with neural networks, in 'International conference on machine learning', PMLR, pp. 2979–2987.

Samuelson, P. (1969), 'Lifetime Portfolio Selection by Dynamic Stochastic Programming', Review of Economics and Statistics **vol. 51**(no. 3), 239–246.

Samuelson, P. A. (1973), 'Mathematics of speculative price', Siam Review **15**(1), 1–42.

Schmidhuber, J. (2015), 'Deep learning in neural networks: An overview', Neural networks **61**, 85–117.

Schwartz, R. A. & Whitcomb, D. K. (1977), 'The time-variance relationship: Evidence on autocorrelation in common stock returns', The Journal of Finance **32**(1), 41–55.

scikit-learn Developers (n.d.a), 'Gaussian process regression (gpr)', https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html. Accessed: 24/11/2022.

scikit-learn Developers (n.d.b), 'Multi-layer perceptron regressor', https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html. Accessed: 24/11/2022.

Scott, J., Sharpe, W. & Watson, J. (2008), 'The 4% Rule – At What Price?'. Available at http://www.standford.edu/~sfsharpe/retecon/4percent.pdf.

Scott, M. (2013), 'Applied stochastic processes', Lecture Notes .

Sheldon Lin, X. & Tan, K. S. (2003), 'Valuation of equity-indexed annuities under stochastic interest rates', North American Actuarial Journal **7**(4), 72–91.

Sheshinski, E. (2008), The Economic Theory of Annuities, Princeton University Press, Princeton.

Sheshinski, E. (2010), 'Refundable annuities (annuity options)', Journal of Public Economic Theory **12**(1), 7–21.

Society of Actuaries (2014), 'RP-2014 Mortality Tables'. Retrieved from https://www.soa.org/Research/Experience-Study/pension/research-2014-rp.aspx.

S&P Risk Control 2.0 Indices Methodology (2022), https://www.spglobal.com/spdji/en/documentmethodologies/methodology-sp-risk-control-2-indices.pdf, note = Accessed: 24/11/2022.

Srokowski, T. & Kamińska, A. (2004), 'Stochastic equation for a jumping process with long-time correlations', Physical Review E **70**(5), 051102.

Stein, E. M. & Stein, J. C. (1991), 'Stock price distributions with stochastic volatility: an analytic approach', The review of financial studies **4**(4), 727–752.

Steinorth, P. & Mitchell, O. S. (2015), 'Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits', Insurance: Mathematics and Economics **64**, 246–258.

Stout, R. & Mitchell, J. (2006), 'Dynamic Retirement Withdrawal Planning', Financial Services Review **15(2)**(2), 117–131.

Tegnér, M. & Poulsen, R. (2018), 'Volatility is log-normal—but not for the reason you think', Risks **6**(2), 46.

VanderPal, G. A. (2008), 'Equity index annuities', Journal of Personal Finance **7**(2).

Viswanathan, G., Fulco, U., Lyra, M. & Serva, M. (2003), 'The origin of fat-tailed distributions in financial time series', Physica A: Statistical Mechanics and its Applications **329**(1-2), 273–280.

Vitali, S., Budimir, I., Runfola, C. & Castellani, G. (2019), 'The role of the central limit theorem in the heterogeneous ensemble of brownian particles approach', Mathematics **7**(12), 1145.

Wang, M. C. & Uhlenbeck, G. E. (1945), 'On the theory of the brownian motion ii', Reviews of modern physics **17**(2-3), 323.

Watson, J. (2022), 'When indices are cut: What withdrawals teach about risk-control index design', https://blogs.cfainstitute.org/investor/2021/11/15/when-indices-are-taken-away-what-withdrawals-teach-about-risk-control-index-design/. Accessed: 10/11/2022.

Wettstein, G., Munnell, A., Hou, W. & Gok, N. (2021), 'The value of annuities', SSRN: https://ssrn.com/abstract=3797822 .

Williams, C. K. & Rasmussen, C. E. (2006), Gaussian processes for machine learning, MIT press Cambridge, MA.

Wu, Y., Wang, H., Zhang, B. & Du, K.-L. (2012), 'Using radial basis function networks for function approximation and classification', International Scholarly Research Notices **2012**.

Wüthrich, M. V. & Merz, M. (2022), Statistical foundations of actuarial learning and its applications, Springer Nature.

Xiao, Y., Tang, Y. & Lee, C. F. (2021), Impacts of time aggregation on beta value and r 2 estimations under additive and multiplicative assumptions: Theoretical results and empirical evidence, in 'Handbook Of Financial Econometrics, Mathematics, Statistics, And Machine Learning', World Scientific, pp. 3947–3984.

Xu, W., Chen, Y., Coleman, C. & Coleman, T. F. (2018), 'Moment matching machine learning methods for risk management of large variable annuity portfolios', Journal of Economic Dynamics and Control **87**(1-20).

Yaari, M. E. (1965), 'Uncertain lifetime, life insurance, and the theory of the consumer', The Review of Economic Studies **32**(2), 137–150.

Zeng, C., Chen, Y. & Yang, Q. (2012), 'The fbm-driven ornstein-uhlenbeck process: Probability density function and anomalous diffusion', Fractional Calculus and Applied Analysis **15**(3), 479–492.

Zhang, S.-Q., Wang, F. & Fan, F.-L. (2022), 'Neural network gaussian processes by increasing depth', IEEE Transactions on Neural Networks and Learning Systems .