8-7-2012

Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications

Nan Zhu
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/rmi_diss

Recommended Citation
Zhu, Nan, "Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications."
Dissertation, Georgia State University, 2012.
https://scholarworks.gsu.edu/rmi_diss/30

This Dissertation is brought to you for free and open access by the Department of Risk Management and Insurance at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Risk Management and Insurance Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
PERMISSION TO BORROW

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or publish this dissertation may be granted by the author or, in his/her absence, the professor under whose direction it was written or, in his absence, by the Dean of the Robinson College of Business. Such quoting, copying, or publishing must be solely for the scholarly purposes and does not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential gain will not be allowed without written permission of the author.

Nan Zhu
NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Nan Zhu
35 Broad Street NW, 11th Floor, Atlanta, GA 30303

The director of this dissertation is:

Daniel Bauer
Department of Risk Management and Insurance
35 Broad Street NW, 11th Floor, Atlanta, GA 30303
Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications

BY

Nan Zhu

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY

ROBINSON COLLEGE OF BUSINESS

2012
ACCEPTANCE

This dissertation was prepared under the direction of the Nan Zhu Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctoral of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

H. Fenwick Huss, Dean

DISSERTATION COMMITTEE

Daniel Bauer
Conrad S. Ciccotello
Shiferaw Gurmu
Richard D. Phillips
Ajay Subramanian
ABSTRACT

Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications

BY

Nan Zhu

July 11, 2012

Committee Chair: Daniel Bauer

Major Academic Unit: Department of Risk Management and Insurance

My doctoral thesis “Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications” addresses economic and mathematical aspects pertaining to uncertainties in human lifetimes. More precisely, I commence my research related to life insurance markets in a methodological direction by considering the question of how to forecast aggregate human mortality when risks in the resulting projections is important. I then rely on the developed method to study relevant applied actuarial problems. In a second strand of research, I consider the uncertainty in individual lifetimes and its influence on secondary life insurance market transactions.

Longevity risk is becoming increasingly crucial to recognize, model, and monitor for life insurers, pension plans, annuity providers, as well as governments and individuals. One key aspect to managing this risk is correctly forecasting future mortality improvements, and this topic has attracted much attention from academics as well as from practitioners. However, in the existing literature, little attention has been paid to accurately modeling the uncertainties associated with the obtained forecasts, albeit having appropriate estimates for the risk in mortality projections, i.e. identifying the transiency of different random sources affecting the projections, is important for many applications.
My first essay “Coherent Modeling of the Risk in Mortality Projections: A Semi-Parametric Approach” deals with stochastically forecasting mortality. In contrast to previous approaches, I present the first data-driven method that focuses attention on uncertainties in mortality projections rather than uncertainties in realized mortality rates. Specifically, I analyze time series of mortality forecasts generated from arbitrary but fixed forecasting methodologies and historic mortality data sets. Building on the financial literature on term structure modeling, I adopt a semi-parametric representation that encompasses all models with transitions parameterized by a Normal distributed random vector to identify and estimate suitable specifications. I find that one to two random factors appear sufficient to capture most of the variation within all of our data sets. Moreover, I observe similar systematic shapes for their volatility components, despite stemming from different forecasting methods and/or different mortality data sets. I further propose and estimate a model variant that guarantees a non-negative process of the spot force of mortality. Hence, the resulting forward mortality factor models present parsimonious and tractable alternatives to the popular methods in situations where the appraisal of risks within medium or long-term mortality projections plays a dominant role.

Relying on a simple version of the derived forward mortality factor models, I take a closer look at their applications in the actuarial context in the second essay “Applications of Forward Mortality Factor Models in Life Insurance Practice.” In the first application, I derive the Economic Capital for a stylized UK life insurance company offering traditional product lines. My numerical results illustrate that (systematic) mortality risk plays an important role for a life insurer’s solvency. In the second application, I discuss the valuation of different common mortality-contingent embedded options within life insurance contracts. Specifically, I present a closed-form valuation formula for Guaranteed Annuity Options within traditional endowment policies, and I demonstrate how to derive the fair option fee for a Guaranteed Minimum Income Benefit within a Variable Annuity Contract based on Monte Carlo simulations. Overall my results exhibit the advantages of forward mortality factor models in terms of their simplicity and compatibility with classical life contingencies theory.

The second major part of my doctoral thesis concerns the so-called life settlement market, i.e. the secondary market for life insurance policies. Evolving from so-called “viatical settlements” popular in the late 1980s that targeted severely ill life insurance policyholders, life settlements generally involve senior insureds with below average life expectancies. Within such a transaction, both the liability of future contingent premiums and the benefits of a life insurance contract are transferred from the policyholder to a life settlement company, which may further securitize a bundle of these contracts in the capital market.

One interesting and puzzling observation is that although life settlements are advertised as a high-return investment with a low “Beta”, the actual market systematically underperformed relative to expectations. While the common explanation in the literature for this gap between anticipated and realized returns falls on the allegedly meager quality of the underlying life expectancy estimates, my third essay “Coherent Pricing of Life Settlements

\[^1\text{Published in The Geneva Papers on Risk and Insurance – Issues and Practice, Vol. 36 (2011).}\]
"Under Asymmetric Information" proposes a different viewpoint: The discrepancy may be explained by *adverse selection*. Specifically, by assuming information with respect to policyholders’ health states is asymmetric, my model shows that a discrepancy naturally arises in a competitive market when the decision to settle is taken into account for pricing the life settlement transaction, since the life settlement company needs to shift its pricing schedule in order to balance expected profits. I derive practically applicable pricing formulas that account for the policyholder’s decision to settle, and my numerical results reconfirm that—depending on the parameter choices—the impact of asymmetric information on pricing may be considerable. Hence, my results reveal a new angle on the financial analysis of life settlements due to asymmetric information.

Hence, all in all, my thesis includes two distinct research strands that both analyze certain economic risks associated with the uncertainty of individuals’ lifetimes—the first at the aggregate level and the second at the individual level. My work contributes to the literature by providing both new insights about how to incorporate lifetime uncertainty into economic models, and new insights about what repercussions—that are in part rather unexpected—this risk factor may have.
DEDICATION

To Yaxian, My Better Half
I owe my advisor, Professor Daniel Bauer greatly, for his most patient, insightful, and encouraging guidance throughout my Ph.D. study. Without him I could never have finished this work. I would also like to thank my committee members, Professors Conrad Ciccotello, Shiferaw Gurmu, Richard Phillips, and Ajay Subramanian for their instructive suggestions. In particular, my gratitude goes to Professor Richard Phillips and Professor Ajay Subramanian for their continuous encouragements for my research and generous funding for my conference presentations.

Best wishes to my Ph.D. fellows, Jin Gao, Hae Won Jung, Fan Liu, Jimmy Martinez, Thorsten Moenig, and Xue Qi for all the laughters and tears we shared over the years. Without you folks I could not recognize the Department of Risk Management and Insurance at Georgia State University as my second home.

Last, I want to thank my wife, Yaxian Li, and my parents, Xia Zang and Xinmin Zhu, from the deepest bottom of my heart. Thank you for supporting me unconditionally, correcting me when necessary, and believing in me at all times.
# TABLE OF CONTENTS

**DEDICATION** ................................................. ix  
**ACKNOWLEDGEMENTS** ........................................ x  
**LIST OF FIGURES** .......................................... xiv  
**LIST OF TABLES** ............................................. xv  

**Chapter 1** **COHERENT MODELING OF THE RISK IN MORTALITY PROJECTIONS: A SEMI-PARAMETRIC APPROACH** xvii  

1.1 Introduction ........................................... xvii  
1.2 A Factor Model of Mortality Forecasts .................. xxv  
  1.2.1 Mortality Forecasts ................................ xxv  
  1.2.2 Data ................................................... xxix  
  1.2.3 Factor Analysis ....................................... xxxii  
  1.2.4 Simple Factor Models ................................ xl  
1.3 Forward Mortality Factor Models ....................... xlv  
  1.3.1 Theory ................................................ xlv  
  1.3.2 Econometrical Approach .............................. xlvii  
  1.3.3 Maximum Likelihood Estimation .................... liv  
  1.3.4 Should the Self-consistency Condition be Imposed? lx
Chapter 2  APPLICATIONS OF FORWARD MORTALITY FACTOR MODELS IN LIFE INSURANCE PRACTICE

2.1 Introduction ............................................. lxxiv
2.2 Forward Mortality Factor Models .............................. lxxvii
2.3 Application I: Economic Capital in Internal Models .......... lxxxiii
  2.3.1 Model Framework .................................. lxxxiii
  2.3.2 A Stylized Life Insurance Company ..................... lxxxvi
  2.3.3 Implementation .................................... lxxxviii
  2.3.4 Results ........................................... xciii
2.4 Application II: Valuation of Annuitization Options ............ xcvii
  2.4.1 Valuation of Guaranteed Annuity Options .......... xcvii
  2.4.2 Valuation of Guaranteed Minimum Income Benefits ........ cii
  2.4.3 Implementation and Results ........................ ciii
2.5 Conclusion ......................................... cvii
2.6 Appendix: Principle Component Analysis ...................... cviii

Chapter 3  COHERENT PRICING OF LIFE SETTLEMENTS UNDER ASYMMETRIC INFORMATION  .............. cxi
LIST OF FIGURES

Figure 1.1 Examples of $P$-spline Projections, female US data . . . . . . . . . xxxi

Figure 1.2 Eigenvectors for the ENW data, Lee-Carter model . . . . . . . . . xxxv

Figure 1.3 Eigenvectors for the ENW data, CBD-Perks Model . . . . . . . . . xxxviii

Figure 1.4 Eigenvectors for the ENW data, $P$-spline method . . . . . . . . . xlii

Figure 1.5 Fitting $C_1(x)$ with Logistic-Gompertz function, USA data . . . . lii

Figure 1.6 Fitting $C_2(x)$ with Logistic-Gompertz function, male population, Lee-
Carter method . . . . . . . . . . . . . . . . . . . . . . . . . . . . liv

Figure 1.7 Simulated Life Expectancy After One Year . . . . . . . . . . . . . lxix

Figure 2.1 Principal Component Analysis, Estimation, and Projection . . . . lxxxii

Figure 2.2 Empirical CDF of Economic Capital . . . . . . . . . . . . . . . . xciv

Figure 3.1 Population Survival Probabilities Projected from the Lee-Carter Method cxxviii

Figure 3.2 Empirical CDFs of Individual Survival Probabilities . . . . . . . cxxix

Figure 3.3 Equilibriums under Symmetric Information . . . . . . . . . . . . . cxxx

Figure 3.4 Asymmetric Behavior of Policyholders . . . . . . . . . . . . . . . cxxxii

Figure 3.5 Equilibrium Offer Prices . . . . . . . . . . . . . . . . . . . . . . cxxxiv
LIST OF TABLES

Table 1.1 The Six Largest Eigenvalues: Lee-Carter Model . . . . . . . . . . xxxvi
Table 1.2 The Six Largest Eigenvalues: CBD-Perks Model . . . . . . . . . . xxxix
Table 1.3 The Six Largest Eigenvalues: P-spline Method . . . . . . . . . . xli
Table 1.4 Tests for i.i.d. and Normality . . . . . . . . . . . . . . . . . . . . xliv
Table 1.5 Sample Means and Standard Variances . . . . . . . . . . . . . . . xliv
Table 1.6 The MLE results for female population – Lee-Carter Model . . . . lvii
Table 1.7 The MLE results for male population – Lee-Carter Model . . . . lviii
Table 1.8 The MLE results for female population – CBD-Perks Model . . . lviii
Table 1.9 The MLE results for male population – CBD-Perks Model . . . . lix
Table 1.10 The MLE results for female population – P-spline Model . . . . lix
Table 1.11 The MLE results for male population – P-spline Model . . . . lix
Table 1.12 MLE without Self-Consistency Condition . . . . . . . . . . . . . . lxiii
Table 1.13 Comparison of Mean and Standard Variance . . . . . . . . . . . lxiii
Table 1.14 Non-negative estimation: U.S. female & Lee-Carter . . . . . . . lxvii
Table 2.1 Calibrated parameters from the MLE method . . . . . . . . . . . lxxxi
Table 2.2  Estimated capital market parameters  

Table 2.3  Portfolio for the company, $E = 2,000,000$  

Table 2.4  Economic capital for the stylized company based on different risk measures  

Table 2.5  Calculations of VaR$_{99\%}$  

Table 2.6  Valuation of GAOs  

Table 2.7  Fair option fee $\phi^*$
Chapter 1

COHERENT MODELING OF THE RISK IN MORTALITY PROJECTIONS: A SEMI-PARAMETRIC APPROACH

1.1 Introduction

Having appropriate estimates for the risk in mortality projections is important in many respects. For instance, a retiree’s personal financial planning decisions will be affected by his longevity prospects as well as associated uncertainties. For life insurers, pension plans, and other similar institutions, the risk in mortality projections directly translates into risk in their liabilities. And even for governments, the extent and the organization of inter-generational risk sharing depend on the riskiness of aggregate mortality trends. However, while there exist a growing number of scientific papers on forecasting human mortality, these contributions have bestowed little attention upon the uncertainty associated with the resulting projections. More specifically, the ubiquitous approach relies on past mortality data to furnish a projection of the mortality experience for subsequent periods—possibly involving error terms—that matches the past observations in some optimal sense. In this paper, we take a different approach by directly focusing on the risk in projections: Given a forecasting methodology, we assess the inherent risks by analyzing time series of forecasts generated based on a rolling window of annual mortality data, and—in doing so—develop suitable models.
Clearly, the two approaches are theoretically equivalent since a stochastic model for mortality experience in every period implies a stochastic mortality forecast and vice versa. However, the distinction is relevant for the econometrical approach and, therefore, is important for the specification and estimation process. This is widely recognized for interest rate models: There, specifying the short rate is equivalent to specifying models for the entire term structure of interest rates; yet any meaningful empirical approach requires the consideration of the entire yield curve data and not only observations on the short end. In particular, the cross-sectional view is important to identify the persistence and transiency of different random sources.

Hence, in brief the goal of this paper is to carry over conventional approaches and techniques for specifying and estimating term structure models to the mortality context and to apply the resulting models. However, there are some profound differences to interest rate modeling. First, in addition to the “term”-dimension, there is an “age”-dimension to be taken into account, and the two enter the relevant equations dissimilarly. Moreover, insurance prices—from which risk-adjusted mortality projections could potentially be stripped—are obfuscated by insurance expenses, idiosyncrasies of the particular insured population, and/or credit risk. Generational life tables, which serve as the actuarial basis for pricing the policies, on the other hand, are compiled infrequently and come with potential inconsistencies in the data or the compilation process over time.

As a resolution to the latter point, we base our considerations on mortality forecasts that are generated from different windows of past mortality experience for ten different populations using a fixed projection methodology. More precisely, we rely on three popular—
yet very different—forecasting approaches that underlie the compilation of most existing
generational life tables: (1) The Lee-Carter approach (Lee and Carter (1992)), (2) the CBD-
Perks model (Cairns et al. (2006b)), and (3) the $P$-spline method (Currie et al. (2004)). It is
important to note, however, that we do not adopt the assumptions on the forecasting errors
associated with the approaches but view them as methodologies to generate (deterministic)
forecasts. Therefore, our approach may be interpreted as an attempt to devise a “stochastic
wrap” around existing mortality forecasting approaches that captures the risk in the resulting
projections in a coherent manner.

We are thus given a time series of mortality forecasts in the form of expected survival
probabilities as objects on some space of functions in two variables, namely (current) age
and (forecasting) time horizon. Dynamic stochastic models can then be formulated via a
stochastic (differential) equation in this function space. By initially restricting ourselves to
time-homogeneous models with Gaussian innovations and by adequately transforming the
observations, we can conduct a factor analysis to determine the number of drivers of the
mortality projections and their shapes. We find that one factor explains the majority—and
in many cases the vast majority—of the variation in the data. Moreover, the shape
of the associated eigenvector as a function of term and age is highly systematic and very
similar across different populations and forecasting methodologies. As may be expected, it is
increasing in the age, but is also increasing in the term if we hold the age constant. Thus, in
analogy to the common connotation in interest rate modeling, we refer to the first factor as
the *slope factor*. The shape of eigenvector associated with the second principal component
is more diverse though still relatively systematic. Notably, here the sign differs between the
short and the long end of the term structure in most cases, so that we refer to it as the *twist factor*. By regressing the transformed data on the leading principal components, we obtain simply forecasting models for mortality projections that can be estimated by OLS.\(^1\)

However, the resulting factor models do not necessarily account for *cross-sectional restrictions* that originate from their interpretation as forecasts. This is again in analogy to interest rate modeling, where cross-sectional restrictions enter in the form of *no-arbitrage restrictions* (see e.g. Piazzesi (2010)). This feature can be captured by so-called *forward mortality models*, which impose the *self-consistency condition* that expected values of future forecasts should align with the current forecasts. By relying on results from the financial mathematics literature on the question when these types of models can be finite-dimensionally realized with Gaussian transitions, i.e. when transitions from one to the next forecast can be parameterized by a finite-dimensional Normal random vector, we arrive at a semi-parametric representation that encompasses all such models. This allows us to identify suitable models by expressing the principal component(s) from the factor analysis in terms of this semi-parametric representation. In particular, we find that the observed shapes can be represented by few parameters. To account for the cross-sectional relationship resulting from the self-consistency condition, we devise a maximum likelihood approach, where the underlying Gaussian distribution yields a particularly simple formulation.

In this context, it is necessary to point out that within the interest rate literature, there is a debate on whether cross-sectional shall be imposed when forecasting yields (see, among others, Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006), and Christensen et

\(^1\)Devising factor models by relying on the leading principal components is also popular in interest rate modeling. See e.g. Diebold and Li (2006) or Joslin et al. (2011).
al. (2011)). Recent contributions by Joslin et al. (2011) and Duffee (2011) add key insights to this discussion. More specifically, Joslin et al. (2011) show that for Gaussian term structure models without any restrictions on risk premium dynamics, no-arbitrage restrictions are irrelevant to estimating factor dynamics, whereas Duffee (2011) even asserts that no-arbitrage restrictions are unnecessary to estimate the cross-sectional mapping. It is important to note that the former argument does not apply in our context since there is no embedded change of measure, i.e. a local expectation hypothesis holds due to the interpretation of the data as forecasts. Hence, the common intuition that cross-sectional restrictions generally improve the efficiency of estimates—unlike in interest rate modeling as shown by Joslin et al. (2011)—actually pertains in our case (see e.g. Piazzesi (2010)). However, the argument from Duffee (2011) that the restrictions only bite for the cross-sectional estimation if they are inconsistent with the true cross-sectional patterns in the data prevails. Putting together these two lines of reasoning allows us to conclude that in the mortality setting, imposing the cross-sectional restrictions will improve efficiency but should not invalidate the estimates without imposing them if the cross-sectional restrictions hold in the data, i.e. if the forecasts are self-consistent. In particular, we can test the self-consistency of each forecasting approach. We find that for U.S. female data, only the Lee-Carter approach produces self-consistent forecasts whereas for the other two approaches self-consistency is rejected. Thus, our approach endorses the Lee-Carter method for devising (deterministic) mortality projections, and—in conjunction—the two approaches yield coherent, parsimonious, and tractable models for stochastically forecasting mortality projections in this case.
Despite the advantages in tractability, the confinement to Gaussian distribution entails the theoretical shortcoming that realizations of survival probabilities—with a small probability—will exceed unity. To rectify this shortcoming, we present a non-negative variation of the proposed Gaussian model. More precisely, by relying on a square-root affine process, we propose a non-negative spot force model that has similar characteristics to the spot force model associated with the Gaussian model. We estimate the resulting model via an unscented Kalman filter approach.

We apply the resulting models to derive confidence intervals for future life expectancies, and compare them to corresponding quantities based on the underlying mortality forecasting approaches. We find that for U.S. female data and Lee-Carter as the underlying forecasting approach, the confidence intervals do not differ much among the simple factor, the self-consistent forward mortality model, and the non-negative spot force model. However, these confidence intervals are considerably wider than those produced by the basic Lee-Carter approach with error terms, especially for younger ages where mortality estimates in the far future are important. This underscores the primary motivation of our approach, namely that conventional mortality forecasting approaches fail to accurately capture the risk in mortality forecasts.

In the companion paper Zhu and Bauer (2011a), we take a closer look at applications of the resulting models in the life insurance context, where the advantages of our formulation are particularly apparent. More specifically, we analyze the calculation of Economic Capital.

---

2Our two-dimensional process is driven by a one-dimensional Brownian motion so that it does not fall in the affine class as it is defined by some authors. See e.g. Piazzesi (2010).
and the valuation of mortality-contingent embedded options in insurance contracts. Our results illustrate the economic significance of the risk in mortality forecasts.

**Related Literature**

Clearly our paper is related to stochastically forecasting mortality, where the ubiquitous approach is to rely on the error estimates related to single-period projections. For instance, by simulating the mortality evolution over a specified time horizon, it is possible to derive confidence intervals for multi-period survival probabilities, (cohort) life expectancies, and similar quantities. However, it is important to realize that the underlying error estimates—and particularly their shapes—were derived to as accurately as possible forecast the next period’s mortality experience; this does not necessarily imply that these error terms are suitable to appraise the risk within medium- to long-term projections. For specific mortality forecasting models, this issue has been pointed out before in the statistical and demographic literature in the context of trend or parameter uncertainty (see e.g. Currie (2010) or Dowd et al. (2010)). Also, this observation is related to the concept of *recalibration risk* that was recently raised in the context of hedging longevity risk (see Cairns (2012) for details).

The most widely used mortality forecasting methodology in academic research, within the life insurance industry, and also by official entities such as the US Census Bureau or the United Nations is the Lee-Carter model (Lee and Carter (1992)). As indicated above, in the paper we also rely on the Lee-Carter method for generating (deterministic) mortality forecasts. In addition, we make use of a simple version of the CBD-Perks model proposed by Cairns et al. (2006b) without cohort effects and the $P$-spline method from Currie et al.
(2004). For an overview on stochastic mortality models, we refer to Booth (2006) and Cairns et al. (2008). The models considered in the paper generally are so-called forward mortality models (cf. Cairns et al. (2006a), Bauer et al. (2008a, 2010b, 2012)), while the resulting models—or more specifically the associated spot mortality models for each cohort—in turn fall into the general class of affine processes (Duffie et al. (2000, 2003)). In the mortality context, affine models have been previously considered by, among others, Biffis (2005).

Furthermore, as emphasized in several places throughout this introduction, this paper is closely related to the literature on modeling the term structure of interest rates. More precisely, our factor analysis for mortality rates resembles—and was inspired by—the analogous approach for yields from Litterman and Scheinkman (1991); the models for mortality forecasts, i.e. forward mortality models, are structurally similar as the forward interest rate models in Heath et al. (1992); and our specific finite-dimensional realizations rely on examples given in Björk and Gombani (1999). Also, our estimation and specification techniques in many places are inspired by corresponding results from interest rate modeling. Specifically, we rely on ideas from Diebold and Li (2006), Duffee (2002, 2011), Joslin et al. (2011), and Piazzesi (2010).

A number of papers have studied the impact of mortality risk in the economic literature. Lee and Tuljapurkar (1998) construct stochastic forecasts of the population, productivity growth, and interest rates to assess the long-run finances of Social Security (OASDI). Auerbach and Lee (2005) analyze and compare how different public pension structures spread the demographic and economic risks across generations. In a recent contribution, Cocco and Gomes (2012) investigate the implications of mortality risk on individuals’ lifetime utilities
and the benefits of mortality-linked financial assets such as longevity bonds by solving the associated life-cycle model with mortality risk. However, all these papers use conventional approaches such as the Lee-Carter model to generate stochasticity in mortality, and thus may not be able to fully comprehend the economic impact of mortality risk. It is thus to our interest to apply our mortality models in the same circumstances and compare our results with the previous findings. We leave this as an imperious future research.

Outline of the Paper

The remainder of this paper is organized as follows: Section 1.2 defines mortality forecasts, introduces the data, and conducts the factor analysis. Section 1.3 introduces our forward mortality factor models under the Gaussian assumption, and derives appropriate function specifications as well as parameter estimations. Section 1.4 discusses the non-negative extension of the model, while applications are presented in Section 1.5. Finally, Section 1.6 concludes.

1.2 A Factor Model of Mortality Forecasts

1.2.1 Mortality Forecasts

Similar to other papers on demographic modeling, we start our considerations from the data. However, here, instead of relying on realized “annual age-specific death rates” (cf. p. 659 in Lee and Carter (1992)) leading to models for mortality experience, we assume that we are given a time series of age-specific—or rather cohort-specific—mortality-forecasts as expectations to survive for a certain period of time. More specifically, we assume that at
time \( t \), we are given \( \{ \tau p_x(t) \mid (\tau, x) \in \mathcal{C} \} \), where \( \tau p_x(t) \) denotes the probability for an \( x \)-year old to survive for \( \tau \) periods until time \( t + \tau \) based on the information at—or, more precisely, up to—time \( t \), and \( \mathcal{C} \) denotes a (large) collection of term/age combinations.\(^3\) Our goal is to propose dynamic stochastic models for mortality forecasts \( \{ \tau p_x(t) \mid (\tau, x) \in \mathcal{C} \}_{t \geq 0} \).

Understanding the risk in mortality forecasts is important in many regards. For instance, the expected future lifetime for an \( x \)-year old at time \( t \), which is an important metric for financial decisions at the individual and the societal level, is a functional of the mortality forecasts:

\[
\hat{e}_x(t) = \int_0^\infty \tau p_x(t) \, d\tau.
\]

For insurance companies and pension plans, on the other hand, the expected present value of their liabilities is given via mortality forecasts. For example, the expected discounted payoff of a life annuity on an \( x \)-year old at time \( t \) is

\[
\hat{a}_x(t) = \sum_{k=0}^\infty p(t, k) \nu p_x(t),
\]

where \( p(t, \tau) \) is the time \( t \) price of a zero coupon bond maturing at time \( t + \tau \). Similarly, other actuarial present values can be expressed in terms of \( \{ \tau p_x(t) \} \).

A (continuous-time) model for \( \{ \tau p_x(t) \mid (\tau, x) \in \mathcal{C} \}_{t \geq 0} \) can be formulated via a stochastic (differential) equation on a suitable function space. However, since the specific monotonicity and boundedness requirements of the forecasted probabilities lead to complications

\(^3\)Our notation is based on *International Actuarial Notation*, where ‘\( p \)’ generally denotes survival probabilities, but we extend it to introduce time dependence.
in their modeling, it is easier to work with the transformed objects

\[ \mu_t(\tau, x) = -\frac{\partial}{\partial \tau} \log \{e p_x(t)\}, \]

the so-called *forward force of mortality*, which we interpret as an element of some Hilbert space \( \mathcal{H} \) of continuous functions (we refer to Bauer et al. (2012) for details on these models and examples of suitable function spaces). We start by considering *time-homogeneous*, Gaussian models of the type

\[ d\mu_t = (A\mu_t + \Lambda)\, dt + \Sigma dW_t, \quad (1.1) \]

where \((W_t)\) is a \(d\)-dimensional Brownian motion, \((\Lambda) \in \mathcal{H}, \Sigma \in L(\mathbb{R}^d, \mathcal{H})\), and \(A\) is the infinitesimal generator of a strongly continuous semigroup \((S_t)\) that coincides with the translation semigroup of left shifts in the first and right shifts in the second variable for \(x \geq t \geq 0\), i.e.

\[ (S_t f)(\tau, x) = f(\tau + t, x - t), \quad 0 \leq t \leq x. \]

In particular, we obtain \(A = \frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\) on the domain of \(A, \text{dom}(A)\) (cf. Lemma 3.1 in Bauer et al. (2012)). While the focus on Gaussian models limits generally and is associated with the possibility of survival probabilities exceeding unity, these shortcomings are counterbalanced by econometrical tractability. Generalizations that rectify the theoretical deficiencies are presented in Section 1.4.
Assume generational mortality data $\tau p_x(t_j)$ is given for different evaluation dates $t_j$, $j = 1, \ldots, N$. In addition, let $l$ denote a lag time, and choose a sub-collection $\tilde{C} \subset C$, $|\tilde{C}| = K$, such that for $(\tau, x) \in \tilde{C}$, $(\tau + l, x), (\tau + t_j + 1 - t_j, x - t_j + 1 + t_j), (\tau + l + t_j + 1 - t_j, x - t_j + 1 + t_j) \in C$, $\forall j \in \{1, 2, \ldots, N - 1\}$. For each $(\tau, x) \in \tilde{C}$, $j \in \{1, 2, \ldots, N - 1\}$, define

$$F_l(t_j, t_{j+1}, (\tau, x)) = -\log \left\{ \frac{\tau + l + t_j + 1 - t_j}{\tau + l + t_j + 1 - t_j} \cdot \frac{\tau p_x(t_{j+1})}{\tau p_x(t_{j+1})} \cdot \frac{\tau p_x(t_{j+1})}{\tau p_x(t_{j+1})} \right\}.$$ 

Conceptually, $F_l(t_j, t_{j+1}, (\tau, x))$ measures the log change of the one-year marginal survival probability for the period $[t_{j+1} + \tau, t_{j+1} + \tau + l]$ from projection at time $t_{j+1}$ relative to time $t_j$, for an—at time $t_{j+1}$—$x$-year old individual. The motivation for this definition is provided by the following proposition:

**Proposition 1**

The vectors

$$\bar{F}_l(t_j, t_{j+1}) = \left( \omega(\tau_1, x_1) \times \frac{F_l(t_j, t_{j+1}, (\tau_1, x_1))}{\sqrt{t_{j+1} - t_j}} \right)_{(\tau, x) \in \tilde{C}} = \left( \omega(\tau_1, x_1) \times \frac{F_l(t_j, t_{j+1}, (\tau_1, x_1))}{\sqrt{t_{j+1} - t_j}}, \ldots, \omega(\tau_K, x_K) \times \frac{F_l(t_j, t_{j+1}, (\tau_K, x_K))}{\sqrt{t_{j+1} - t_j}} \right),$$

$j = 1, 2, \ldots, N - 1$ are independent and Gaussian distributed.\(^4\) If the data is equidistant with $t_{j+1} - t_j = \Delta$, they are identically distributed.

\(^4\)By scaling the datapoints by $\frac{1}{\sqrt{t_{j+1} - t_j}}$, we ascertain that the vectors $\bar{F}_l(t_j, t_{j+1})$, $j = 1, 2, \ldots, N - 1$, are also approximately i.i.d. when relying on non-equidistant data.
A proof is provided in Section 1.7. Here, $\omega(\cdot, \cdot)$ is an evaluation weighting function for each $(\tau, x) \in \tilde{C}$, which is introduced for the general consideration that different weights may be assigned to different term/age combinations. For example, we may choose $\omega(\tau, x)$ such that $\partial \omega(\tau, x)/\partial \tau < 0$ reflecting a preference of the near future over the far future. In one extreme case, if we assume $\omega(\tau, x) = 0$, $\forall \tau \geq 2$, then our approach resembles a model for mortality experience.

The i.i.d. structure of $\tilde{F}_l(t_j, t_j + \Delta)$, $j = 1, 2, \ldots, N - 1$, can now be exploited, for instance via a factor analysis.

1.2.2 Data

We utilize historical mortality data sets from five representative developed countries/regions: England & Wales (ENW), France (FRA), Japan (JPN), United States (USA), and West Germany (FRG) as available from the Human Mortality Database for both male and female populations.\textsuperscript{5} For each set of the data, we apply the three aforementioned forecasting methods to generate generational mortality tables: the Lee-Carter approach, the CBD-Perks model, and the $P$-spline method. More specifically, within each dataset, we will have generational mortality tables compiled for twenty-two consecutive years (1986-2007) each using historical mortality data of the past thirty years (that is, for the table of 1986, data from 1956-1985 is employed; for the table of 1987, data from 1957-1986 is employed; etc.) from year 1956 to 2006.\textsuperscript{6} The lag time $l$ is chosen at 1\textsuperscript{7} and for each set of the data, $\tilde{C}$ is chosen as large as

\textsuperscript{5}Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.
\textsuperscript{6}This is the largest intersecting period of available data for all five selected countries/regions.
\textsuperscript{7}We found no systematic deviation for $l$ chosen at higher values, such as 5.
possible. While we introduce the weighting function \( \omega(\tau, x) \) for a possible distinction of different term/age combinations, in the following analysis we assume \( \omega(\tau, x) = 1, \forall \tau, x \), since no prior information is available on which weighting function is more appropriate.

Due to the distinct underlying assumptions, for different forecasting methodologies, different age ranges are used in the generation of generational mortality tables. For the Lee-Carter approach, we use ages ranging from 0 to 95 in the analysis,\(^8\) which yields \( K = 4560 \). In the estimation, instead of the original approach we use the modified weighted-least-squares algorithm from Wilmoth (1993) and further adjust \( \kappa_t \) by fitting a Poisson regression model to the annual number of deaths at each age (cf. Booth et al. (2002)).

The CBD-Perks model was originally proposed to evaluate the evolution of the mortality curve past a certain age (60, for example) by observing an approximately linear relationship between the logit of mortality rates and ages. However, this pattern is generally invalid for very young ages. Therefore, for the CBD-Perks generations, we use a reduced age range from 25-95, which leads a smaller \( K = 2485 \).

While the \( P \)-spline method is proposed as an alternative forecasting method that does not make strong assumptions on the functional form of the mortality surface, its direct application sometimes leads to bizarre results. For example, Figure 1.1(a) and 1.1(b) display the projected log mortality rates at different ages from the \( P \)-spline method by minimizing the Bayesian information criterion (BIC), based on female USA mortality data from year 1960-1989 and 1964-1993, respectively. From the figure, we observe that although the projections in the former case are all well behaved, in the latter case the log mortality rates

\(^8\)We discard data for extremely high ages due to the limited number of exposures and poor data quality.
Figure 1.1. Examples of $P$-spline Projections, female US data
for all sample ages (unrealistically) increase, and even exceed 0 for ages 71 and 95, which is obviously undesirable. One way to avoid this problem is to target a fixed value of the degree of freedom ($df$) instead of minimizing the BIC. More specifically, by fixing $df$ at 20, we observe reasonable shapes from all generations. Similarly as in the CBD-Perks model, we use a reduced age range from 25 to 95 with $K = 2485$.

1.2.3 Factor Analysis

With $\Delta = t_{j+1} - t_j = 1$, Proposition 1 implies that $\bar{F}_l(t_j, t_{j+1})$ are i.i.d. Gaussian so that we can write

$$\bar{F}_l(t_j, t_{j+1}) = a + bZ_j + \epsilon_j,$$  \hspace{1cm} (1.2)

with coefficients $a \in \mathbb{R}^K$, $b \in \mathbb{R}^{K \times d}$, factors $Z_j \in \mathbb{R}^d$ with $\mathbb{E}(Z_j) = 0$ and $\text{Cov}(Z_j) = I_{d \times d}$, and an error term $\epsilon_j \in \mathbb{R}^K$ with $\mathbb{E}(\epsilon_j) = 0$ and $\text{Cov}(\epsilon_j) = \text{diag}(\psi_1, \ldots, \psi_K)$.

Estimates of $a$, $b$, and the number of factors, $d$, can be obtained from a principal component analysis on the time series of $\bar{F}_l(t_j, t_{j+1})$, $j = 1, 2, \ldots, N - 1$. This is akin to the fixed income literature (see e.g. Litterman and Scheinkman (1991) or Rebonato (1998)), and several authors have taken a similar approach to the analysis of period mortality data (see e.g. Lee and Carter (1992) or Njenga and Sherris (2009)). However, thus far, there has been no attempt to analyze generational mortality data in order to identify the drivers of the entire age/term structure of mortality. The procedure is standard: We first compute the

\footnote{We thank Professor I.D. Currie for suggesting this in an email communication.}
empirical covariance matrix of $F_i(t_j, t_{j+1})$, $\hat{\Sigma}$, then decompose it as

$$\hat{\Sigma} = U \times \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix} \times U'^\prime = \sum_{\nu=1}^{K} \lambda_{\nu} u_{\nu} u_{\nu}',$$

where $U = (u_1, u_2, \cdots, u_K)$ is an (orthogonal) matrix consisting of the eigenvectors of $\hat{\Sigma}$, and $\lambda_{\nu}$, $\nu = 1, 2, \ldots, K$, are the corresponding eigenvalues in decreasing order. We then pick the $d$ greatest eigenvalues that explain the majority of the variation in the data, e.g. we choose $d$ such that

$$\frac{\sum_{\nu=1}^{d} \lambda_{\nu}}{\sum_{\nu=1}^{K} \lambda_{\nu}} \geq \xi,$$

where $\xi$ is a given threshold.
Notice that the resulting approximative covariance matrix is

\[
(u_1, \cdots, u_d, 0, \cdots, 0) \times \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_d
\end{pmatrix} \begin{pmatrix}
u_1' \\
u_2' \\
\vdots \\
u_d'
\end{pmatrix}
\]

\[
= \sum_{\nu=1}^{d} \lambda_{\nu} u_{\nu} \times u_{\nu}' = \text{Cov} \left( \sum_{\nu=1}^{d} u_{\nu} \sqrt{\lambda_{\nu}} Z_{\nu,j} \right),
\]

where \(Z_{\nu,j}\) are i.i.d. (scalar) standard Normal random variables, \(\nu, j \in \{1, \ldots, d\} \times \{1, \ldots, N-1\}\). Hence, isolating the first \(d\) eigenvalues suggests the representation

\[
\bar{F}_l(t_j, t_{j+1}) = \mathbb{E} \left[ \bar{F}_l(t_j, t_{j+1}) \right] + \sum_{\nu=1}^{d} u_{\nu} \sqrt{\lambda_{\nu}} Z_{\nu,j} + \epsilon_j,
\]

i.e. \(a = \mathbb{E} \left[ \bar{F}_l(t_j, t_{j+1}) \right]\) and \(b = (u_1 \sqrt{\lambda_1}, \ldots, u_d \sqrt{\lambda_d})\) in Equation (1.2). In what follows, we conduct the factor analysis on each data set with generations from all three forecasting methodologies, respectively.

**The Lee-Carter Approach** Table 1.1 shows the six greatest eigenvalues \((\lambda_{\nu}, \nu = 1, \ldots, 6)\) for different populations under the Lee-Carter approach. From the table, we observe that in all data sets, the first (largest) eigenvalue plays the dominant role in explaining the
total variation. Moreover, by comparing data sets between genders, we find that higher (absolute) variances emerge from the male population, whereas the first eigenvalue has greater explanatory power (higher weight) for the female population in all sample sets.

The eigenvectors associated the two largest eigenvalues as functions of $\tau$ and $x$ for England & Wales are displayed in Figure 1.2 (for other countries, very similar shapes are
<table>
<thead>
<tr>
<th>Country</th>
<th>Factor</th>
<th>Female Population</th>
<th>Male Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Value</td>
<td>Percentage</td>
</tr>
<tr>
<td><strong>England &amp; Wales</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$6.98 \times 10^{-2}$</td>
<td>96.95%</td>
<td>6.46 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$1.00 \times 10^{-3}$</td>
<td>1.42%</td>
<td>3.50 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$4.87 \times 10^{-4}$</td>
<td>0.68%</td>
<td>1.60 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$3.13 \times 10^{-4}$</td>
<td>-</td>
<td>1.31 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$2.13 \times 10^{-4}$</td>
<td>-</td>
<td>7.89 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$1.65 \times 10^{-4}$</td>
<td>-</td>
<td>6.59 $\times 10^{-4}$</td>
</tr>
<tr>
<td><strong>France</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$3.87 \times 10^{-2}$</td>
<td>95.51%</td>
<td>4.75 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$7.63 \times 10^{-4}$</td>
<td>1.88%</td>
<td>4.70 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$6.07 \times 10^{-4}$</td>
<td>1.50%</td>
<td>2.71 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$2.84 \times 10^{-4}$</td>
<td>-</td>
<td>1.32 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$1.04 \times 10^{-4}$</td>
<td>-</td>
<td>7.84 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$6.19 \times 10^{-5}$</td>
<td>-</td>
<td>5.04 $\times 10^{-4}$</td>
</tr>
<tr>
<td><strong>Japan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$1.70 \times 10^{-2}$</td>
<td>92.80%</td>
<td>3.84 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$5.86 \times 10^{-4}$</td>
<td>3.19%</td>
<td>1.20 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$3.62 \times 10^{-4}$</td>
<td>1.98%</td>
<td>9.46 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$1.84 \times 10^{-4}$</td>
<td>-</td>
<td>5.17 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$1.38 \times 10^{-4}$</td>
<td>-</td>
<td>3.02 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$4.99 \times 10^{-5}$</td>
<td>-</td>
<td>2.08 $\times 10^{-4}$</td>
</tr>
<tr>
<td><strong>United States</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$4.50 \times 10^{-2}$</td>
<td>93.13%</td>
<td>7.51 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$1.80 \times 10^{-3}$</td>
<td>3.70%</td>
<td>8.30 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$6.15 \times 10^{-4}$</td>
<td>1.27%</td>
<td>2.60 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$4.79 \times 10^{-4}$</td>
<td>-</td>
<td>1.91 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$2.29 \times 10^{-4}$</td>
<td>-</td>
<td>8.57 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$2.07 \times 10^{-4}$</td>
<td>-</td>
<td>4.22 $\times 10^{-4}$</td>
</tr>
<tr>
<td><strong>West Germany</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$3.79 \times 10^{-2}$</td>
<td>95.17%</td>
<td>4.02 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$7.87 \times 10^{-4}$</td>
<td>1.97%</td>
<td>6.20 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$4.06 \times 10^{-4}$</td>
<td>1.02%</td>
<td>2.82 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$3.05 \times 10^{-4}$</td>
<td>-</td>
<td>1.30 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$2.60 \times 10^{-4}$</td>
<td>-</td>
<td>1.01 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$1.66 \times 10^{-4}$</td>
<td>-</td>
<td>7.80 $\times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 1.1. The Six Largest Eigenvalues: Lee-Carter Model
exhibited and are thus omitted to keep the presentation concise).\textsuperscript{10} We observe that the structure of the first principal component is primarily governed by an increasing age/term effect, which may be the key reason for its dominant role in explaining the variation of the generational mortality data. Moreover, we find that the forward forces of mortality for high ages in the far future appear to be more volatile than those in the near future, a feature that is not captured by most mortality forecasting approaches—particularly mean-reverting ones. Therefore, we refer to the first factor as the \textit{slope factor}.

As for the second principal component, for some data sets (e.g. female ENW) it looks rather unsystematic, whereas for others (e.g. male JPN) we observe a consistently over time decreasing influence that even generates an inverse relationship for higher ages in the near and the far future (generally this factor is more clearly observed from male population data sets). It is therefore referred to as the \textit{twist factor}.

Considering both the weights of eigenvalues and shapes of eigenvectors, in what follows we choose the number of drivers, $d$, to be 1 for female populations, and 2 for male populations in the Lee-Carter case.

\textbf{The CBD-Perks Model} Table 1.2 shows the six greatest eigenvalues for different populations under the CBD-Perks model. From the table, we observe that the first eigenvalue takes an even more dominant role in all data sets compared with the Lee-Carter approach, and that there is no considerable difference between male and female populations in the ex-

\textsuperscript{10}Notice that here we transform $u_i$, $i = 1, 2$, from a $K \times 1$ vector to an upper-triangular matrix with row and column representing $\tau$ and $x$, respectively.
planning power of the first eigenvector. However, we still observe that the male populations still possess higher absolute variations across all selected countries/regions.

The eigenvectors associated with the two largest eigenvalues as functions of $\tau$ and $x$ for England & Wales are displayed in Figure 1.3. We find that the first principal component exhibits essentially the same shape as in the Lee-Carter case across all data sets, which implies that the dominant slope factor appears to be independent of the underlying forecasting
<table>
<thead>
<tr>
<th>Country</th>
<th>Factor</th>
<th>Female Population Value</th>
<th>Female Population Percentage</th>
<th>Male Population Value</th>
<th>Male Population Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>England &amp; Wales</td>
<td>λ₁</td>
<td>$1.36 \times 10^{-1}$</td>
<td>98.62%</td>
<td>$2.53 \times 10^{-1}$</td>
<td>98.80%</td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>$1.70 \times 10^{-3}$</td>
<td>1.25%</td>
<td>$2.80 \times 10^{-3}$</td>
<td>1.09%</td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>$1.71 \times 10^{-4}$</td>
<td>0.12%</td>
<td>$2.53 \times 10^{-4}$</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>λ₄</td>
<td>$9.63 \times 10^{-6}$</td>
<td>-</td>
<td>$2.66 \times 10^{-5}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>$1.95 \times 10^{-6}$</td>
<td>-</td>
<td>$7.98 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₆</td>
<td>$5.98 \times 10^{-8}$</td>
<td>-</td>
<td>$2.51 \times 10^{-7}$</td>
<td>-</td>
</tr>
<tr>
<td>France</td>
<td>λ₁</td>
<td>$2.75 \times 10^{-2}$</td>
<td>97.49%</td>
<td>$1.03 \times 10^{-1}$</td>
<td>97.60%</td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>$6.37 \times 10^{-4}$</td>
<td>2.26%</td>
<td>$2.20 \times 10^{-3}$</td>
<td>2.06%</td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>$6.41 \times 10^{-5}$</td>
<td>0.23%</td>
<td>$3.26 \times 10^{-4}$</td>
<td>0.31%</td>
</tr>
<tr>
<td></td>
<td>λ₄</td>
<td>$5.06 \times 10^{-6}$</td>
<td>-</td>
<td>$2.07 \times 10^{-5}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>$1.62 \times 10^{-6}$</td>
<td>-</td>
<td>$7.58 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₆</td>
<td>$3.29 \times 10^{-8}$</td>
<td>-</td>
<td>$1.13 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td>Japan</td>
<td>λ₁</td>
<td>$2.72 \times 10^{-2}$</td>
<td>96.93%</td>
<td>$1.21 \times 10^{-1}$</td>
<td>98.11%</td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>$7.96 \times 10^{-4}$</td>
<td>2.84%</td>
<td>$2.20 \times 10^{-3}$</td>
<td>1.80%</td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>$5.02 \times 10^{-5}$</td>
<td>0.18%</td>
<td>$9.70 \times 10^{-5}$</td>
<td>0.08%</td>
</tr>
<tr>
<td></td>
<td>λ₄</td>
<td>$1.09 \times 10^{-5}$</td>
<td>-</td>
<td>$1.40 \times 10^{-5}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>$1.05 \times 10^{-6}$</td>
<td>-</td>
<td>$1.61 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₆</td>
<td>$2.12 \times 10^{-7}$</td>
<td>-</td>
<td>$2.61 \times 10^{-7}$</td>
<td>-</td>
</tr>
<tr>
<td>United States</td>
<td>λ₁</td>
<td>$2.09 \times 10^{-2}$</td>
<td>97.73%</td>
<td>$4.88 \times 10^{-2}$</td>
<td>97.49%</td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>$3.49 \times 10^{-4}$</td>
<td>1.63%</td>
<td>$8.16 \times 10^{-4}$</td>
<td>1.63%</td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>$1.31 \times 10^{-4}$</td>
<td>0.61%</td>
<td>$4.23 \times 10^{-4}$</td>
<td>0.84%</td>
</tr>
<tr>
<td></td>
<td>λ₄</td>
<td>$3.60 \times 10^{-6}$</td>
<td>-</td>
<td>$8.54 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>$1.35 \times 10^{-6}$</td>
<td>-</td>
<td>$6.16 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₆</td>
<td>$5.18 \times 10^{-8}$</td>
<td>-</td>
<td>$9.43 \times 10^{-7}$</td>
<td>-</td>
</tr>
<tr>
<td>West Germany</td>
<td>λ₁</td>
<td>$5.81 \times 10^{-2}$</td>
<td>98.54%</td>
<td>$2.33 \times 10^{-1}$</td>
<td>98.91%</td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>$7.39 \times 10^{-4}$</td>
<td>1.25%</td>
<td>$2.00 \times 10^{-3}$</td>
<td>0.85%</td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>$1.13 \times 10^{-4}$</td>
<td>0.19%</td>
<td>$4.95 \times 10^{-4}$</td>
<td>0.21%</td>
</tr>
<tr>
<td></td>
<td>λ₄</td>
<td>$5.75 \times 10^{-6}$</td>
<td>-</td>
<td>$4.46 \times 10^{-5}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>$2.12 \times 10^{-6}$</td>
<td>-</td>
<td>$6.73 \times 10^{-6}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>λ₆</td>
<td>$1.97 \times 10^{-7}$</td>
<td>-</td>
<td>$4.39 \times 10^{-7}$</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1.2. The Six Largest Eigenvalues: CBD-Perks Model
methodology. For the second principal component, similarly, a twisted shape is observed across all data sets. However, the explanation power for the twist factor is relatively small.

Considering both the weights of eigenvalues and shapes of eigenvectors, we assume \( d \) to be 1 for both female and male populations in the CBD-Perks case.

**The \( P \)-spline Method** Table 1.3 shows the six greatest eigenvalues for different populations under the \( P \)-spline method. From the table, we observe that similarly to the CBD-Perks case, the first eigenvalue takes a highly dominant role in all cases, and that there is no considerable difference between male and female population in the explanation power of the first principal component.

Again, the eigenvectors for the two largest eigenvalues as functions of \( \tau \) and \( x \) for England & Wales are displayed in Figure 1.4, and we observe that again both the first and the second principal components exhibit similar shapes (“slope” and “twist”). Since the first eigenvalue takes an extremely dominant role, we choose \( d = 1 \) for both female and male populations in the \( P \)-spline case.

**1.2.4 Simple Factor Models**

From the factor analysis, we can devise simple factor models that can be used as simple, easy-to-estimate mortality forecasting methodologies. Specifically, notice that Equation (1.3) is essentially a regression equation with unknown \( Z_{\nu,j} \). By assuming that the first \( d \) principle components are portrayed by the model without error (see e.g. Diebold and Li (2006) or Joslin et al. (2011) for similar approaches in interest rate modeling) and with the property that \( u_\nu \),
<table>
<thead>
<tr>
<th>Country</th>
<th>Factor</th>
<th>Female Population</th>
<th>Male Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>Percentage</td>
<td>Value</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$1.40 \times 10^{-1}$</td>
<td>99.77%</td>
<td>$1.11 \times 10^{-1}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$2.88 \times 10^{-4}$</td>
<td>0.21%</td>
<td>$1.00 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$2.57 \times 10^{-5}$</td>
<td>-</td>
<td>$1.34 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$1.07 \times 10^{-5}$</td>
<td>-</td>
<td>$2.72 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$1.44 \times 10^{-6}$</td>
<td>-</td>
<td>$3.93 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$8.03 \times 10^{-7}$</td>
<td>-</td>
<td>$1.12 \times 10^{-6}$</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$1.32 \times 10^{-1}$</td>
<td>99.27%</td>
<td>$1.32 \times 10^{-1}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$9.04 \times 10^{-4}$</td>
<td>0.68%</td>
<td>$8.51 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$4.81 \times 10^{-5}$</td>
<td>-</td>
<td>$2.61 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$1.25 \times 10^{-5}$</td>
<td>-</td>
<td>$4.62 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$1.34 \times 10^{-6}$</td>
<td>-</td>
<td>$8.96 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$1.24 \times 10^{-6}$</td>
<td>-</td>
<td>$3.31 \times 10^{-6}$</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$2.05 \times 10^{-2}$</td>
<td>99.31%</td>
<td>$1.44 \times 10^{-1}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$9.57 \times 10^{-5}$</td>
<td>0.46%</td>
<td>$1.20 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$4.14 \times 10^{-5}$</td>
<td>-</td>
<td>$2.64 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$2.74 \times 10^{-6}$</td>
<td>-</td>
<td>$2.57 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$2.40 \times 10^{-6}$</td>
<td>-</td>
<td>$6.73 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$6.07 \times 10^{-7}$</td>
<td>-</td>
<td>$9.59 \times 10^{-7}$</td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$1.74 \times 10^{-1}$</td>
<td>99.19%</td>
<td>$5.60 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$1.11 \times 10^{-3}$</td>
<td>0.62%</td>
<td>$9.61 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$2.91 \times 10^{-4}$</td>
<td>-</td>
<td>$1.36 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$3.14 \times 10^{-5}$</td>
<td>-</td>
<td>$9.68 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$9.63 \times 10^{-6}$</td>
<td>-</td>
<td>$5.66 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$1.04 \times 10^{-6}$</td>
<td>-</td>
<td>$3.17 \times 10^{-6}$</td>
</tr>
<tr>
<td>West Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$1.22 \times 10^{-1}$</td>
<td>99.62%</td>
<td>$4.00 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$3.12 \times 10^{-4}$</td>
<td>0.26%</td>
<td>$1.94 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>$1.46 \times 10^{-4}$</td>
<td>-</td>
<td>$9.59 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>$7.96 \times 10^{-6}$</td>
<td>-</td>
<td>$9.98 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>$1.31 \times 10^{-6}$</td>
<td>-</td>
<td>$2.51 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>$3.23 \times 10^{-7}$</td>
<td>-</td>
<td>$2.61 \times 10^{-7}$</td>
</tr>
</tbody>
</table>

Table 1.3. The Six Largest Eigenvalues: P-spline Method
Figure 1.4. Eigenvectors for the ENW data, $P$-spline method
\( \nu = 1, \ldots, d \), are orthogonal to each other, we have

\[
Y_\nu(j) \triangleq (u_\nu \sqrt{\lambda_\nu})^T \bar{F}_l(t_j, t_{j+1}) = (u_\nu \sqrt{\lambda_\nu})^T \mathbb{E}[\bar{F}_l] + \lambda_\nu Z_{\nu,j}, \ \nu = 1, \ldots, d. \tag{1.4}
\]

Therefore, Equation (1.3) can be modified as

\[
\bar{F}_l(t_j, t_{j+1}) = \mathbb{E}[\bar{F}_l(t_j, t_{j+1})] + \sum_{\nu=1}^{d} \frac{u_\nu}{\sqrt{\lambda_\nu}} [Y_\nu(j) - (u_\nu \sqrt{\lambda_\nu})^T \mathbb{E}[\bar{F}_l(t_j, t_{j+1})]] + \epsilon_j \triangleq \tilde{m} + \sum_{\nu=1}^{d} \tilde{s}_\nu \times Y_\nu(j) + \epsilon_j, \tag{1.5}
\]

which is a regression equation of \( \bar{F}_l(t_j, t_{j+1}) \) on the known \( Y_\nu(j) \), and the constant and linear coefficients \( \tilde{m}, \tilde{s}_\nu, \nu = 1, \ldots, d \) can be easily obtained from OLS regression.

The above factor model can then be used to generate forecasts of mortality projections: From the i.i.d. normality of \( Z_{\nu,j} \) we know that \( Y_\nu(j) \) are also i.i.d. Normal distributed, and we denote the directly calculated sample mean and standard error of \( Y_\nu(j) \) as \( (\mu_{Y,\nu}^s, \sigma_{Y,\nu}^s) \). Therefore, we can simulate \( Y_\nu(N) \sim N(\mu_{Y,\nu}^s, \sigma_{Y,\nu}^s) \). A forecast is then given by

\[
\bar{F}_l(t_N, t_{N+1}) = \tilde{m} + \sum_{\nu=1}^{d} \tilde{s}_\nu \times Y_\nu(N),
\]

from which together with known \( \tau p_x(t_N) \), we can derive \( \tau p_x(t_{N+1}) \), i.e. we can simulate the generation table at time \( t_{N+1} \).
We use the female USA data for an illustration, where we choose \( d = 1 \) as indicated above. First, we test if the sample data is actually i.i.d. (Ljung-Box test) and Normal distributed (Jarque-Bera test), where we use a confidence level of 95%. The test results are displayed in Table 1.4 for all three underlying mortality projection methodologies. From the table, we see that the i.i.d. assumption is rejected for the data set that are generated under the \( P \)-spline method, which suggests possible serial correlation in \( \{Y(j)\} \). We then calculate \( \{Y(j)\} \) under each projection methodology, and regress \( \tilde{m} \) and \( \tilde{s}_\nu \). The associated \( (\mu^*_Y, \sigma^*_Y) \) as well as the 95% confidence intervals are displayed in Table 1.5.
1.3 Forward Mortality Factor Models

1.3.1 Theory

While the simple factor models proposed in the previous section present simply, easy-to-estimate approaches to forecasting mortality projections, they do not account for the inherent structure that arises from the interpretation of the data as forecasts. More precisely, for the forecasts to be self-consistent, the expected value of forecasts should align with the projection engrained in the cross-section of the data. For instance, the expected value of next year’s realized survival rates should coincide with the projection in this year’s generational mortality table.

So-called forward mortality models adhere to this relationship. In what follows, we briefly outline the relevant theory borrowing from Bauer et al. (2012). Subsequently, we demonstrate how these results—in conjunction with the results from the previous section—can be employed to develop simple parametric models for mortality forecasts.

Mathematically, self-consistency of a dynamic model for mortality forecasts takes the form of a martingale property. More specifically, expected realized mortality rates should align with the given forecasts.\(^{11}\)

\[
\mathbb{E}_t \left[ \exp \left\{ - \int_0^T \mu_s(0, x_0 + s) \, ds \right\} \right] = \exp \left\{ - \int_0^t \mu_s(0, x_0 + s) \, ds \right\} T_{-t \, p_{x_0 + t}}(t),
\]

i.e. \( \left( \exp \left\{ - \int_0^t \mu_s(0, x_0 + s) \, ds \right\} T_{-t \, p_{x_0 + t}}(t) \right)_{t \geq 0} \) are martingales. This yields a self-consistency condition akin to the well-known HJM (drift) condition for forward-interest rate models (cf.\(^{11}\) As usual in this context, \( \mathbb{E}_t \) denotes the expectation operator based on the information up to time \( t \).
Cor. 3.1 in Bauer et al. (2012)) which links the drift component and the volatility component of \( \mu_t(\tau, x) \):

\[
\alpha(\tau, x) = \sigma(\tau, x) \times \int_0^\tau \sigma'(s, x) \, ds.
\]  

(1.7)

As before, we are interested in factor models

\[
\mu_t(\tau, x) = G(\tau, x; Z_t),
\]

where \( G \) is a known deterministic function and \( Z_t \) is some convenient finite-dimensional random variable (so that \( (Z_t)_{t \geq 0} \) is some convenient stochastic process). Proposition 4.1 in Bauer et al. (2012) shows that for the time-homogeneous Gaussian models considered here, the volatility structure must necessarily be of the form

\[
\sigma(\tau, x) = C(x + \tau) \times \exp \{ M \tau \} \times N,
\]

(1.8)

where \( N \in \mathbb{R}^{m \times d} \), \( M \in \mathbb{R}^{m \times m} \), and \( C' \in C^1([0, \infty), \mathbb{R}^m) \); the factor model is then given by

\[
\mu_t(\tau, x) = \mu_0(\tau+t, x-t) + \int_0^t \alpha(\tau+t-s, x-t+s) \, ds + C(x+\tau) \exp \{ M \tau \} \int_0^t \exp \{ M (t-s) \} N \, dW_s. 
\]

(1.9)

For a given number of drivers \( d \) from the factor analysis, the above semi-parametric representation can then be conveniently employed to investigate suitable functional forms.
1.3.2 Econometrical Approach

We starting by noting the following proposition which will prove to be convenient in what follows.\footnote{A similar result for forward interest rate models is provided in Angelini and Herzel (2005).}

**Proposition 2**

Let $\sigma(\tau, x) = (\sigma_1(\tau, x), \ldots, \sigma_d(\tau, x))$, where each function $\sigma_i(\tau, x)$ is of the form

$$
\sigma_i(\tau, x) = C_i(x + \tau) \times \exp \{ M_i \tau \} \times N_i, \tag{1.10}
$$

$C_i(\cdot) \in \mathbb{R}^{1 \times m_i}$, $M_i \in \mathbb{R}^{m_i \times m_i}$, $N_i = \mathbb{R}^{m_i \times 1}$, $m_i \in \mathbb{N}$, $i = \{ 1, 2, \ldots, d \}$. Then $\sigma(\tau, x)$ is also of the form (2.3), i.e. the model implied by $\sigma(\tau, x)$ allows for a Gaussian realization, where $C(x) = [C_1(x), \ldots, C_d(x)]$, $M = \text{diag} \{ M_1, \ldots, M_d \}$, and $N = \text{diag} \{ N_1, \ldots, N_d \}$.

A proof is provided in Section 1.7. Proposition 2 essentially allows us to treat each independent factor separately.

With some basic manipulations, we obtain

$$
F_l(t_j, t_{j+1}, (\tau, x)) \overset{d}{=} \int_0^{t_{j+1}-t_j} \int_\tau^{\tau+l} \alpha(v + t_{j+1} - t_j - s, x - t_{j+1} + t_j + s) \, dv \, ds
$$

$$
+ \int_0^{t_{j+1}-t_j} \int_\tau^{\tau+l} C(x + v) \exp \{ M (v + t_{j+1} - t_j) \} \, N \, dv \, dW_s
$$

$$
= \int_\tau^{\tau+l} C(x + v) \exp \{ Mv \} \, dv \times \int_0^{t_{j+1}-t_j} \exp \{ M (t_{j+1} - t_j - s) \} \, N \, dW_s
$$

(1.11)
is Normal distributed.

Furthermore, from Proposition 2 and Equation (1.11), for the model with volatility structure \( \sigma(\tau, x) = (\sigma_1(\tau, x), \ldots, \sigma_d(\tau, x)) \) as in (1.10), we obtain in analogy to Equation (1.3)

\[
\bar{F}_l(t_j, t_{j+1}) \approx \mathbb{E} \left[ \bar{F}_l(t_j, t_{j+1}) \right] + \sum_{\nu=1}^{d} (\omega(\tau_i, x_i) \times O_\nu(\tau_i, x_i))_{1 \leq i \leq K} \\
\times \frac{1}{\sqrt{t_{j+1} - t_j}} \int_{0}^{t_{j+1} - t_j} \exp \{ M_\nu(t_{j+1} - t_j - s) \} N_\nu \, dW^{(\nu)}_s, \tag{1.12}
\]

where \( O_\nu(\tau_i, x_i) = \int_{\tau_i}^{\tau_i + l} C_\nu(x_i + s) \exp \{ M_\nu s \} \, ds \) and \( \frac{1}{\sqrt{t_{j+1} - t_j}} \int_{0}^{t_{j+1} - t_j} \exp \{ M_\nu(t_{j+1} - t_j - s) \} N_\nu \, dW^{(\nu)}_s \) is an \( m_\nu \)-dimensional vector of Normal random variables, \( 1 \leq \nu \leq d \). While this vector may not necessarily consist of perfectly correlated random variables, they are all driven by the same (scalar) Brownian motion and will thus be strongly related. In particular, for a short time step \( (t_{j+1} - t_j) \), the standard Euler scheme yields an approximation by a perfectly correlated random vector

\[
\begin{align*}
&\frac{1}{\sqrt{t_{j+1} - t_j}} \int_{0}^{t_{j+1} - t_j} \exp \{ M_\nu(t_{j+1} - t_j - s) \} N_\nu \, dW^{(\nu)}_s \\
&\approx \frac{1}{\sqrt{t_{j+1} - t_j}} \exp \{ M_\nu(t_{j+1} - t_j) \} N_\nu \left( W^{(\nu)}_{t_{j+1} - t_j} - W^{(\nu)}_{0} \right) \\
&= \exp \{ M_\nu(t_{j+1} - t_j) \} N_\nu \times \tilde{Z}_{\nu,j},
\end{align*}
\tag{1.13}
\]

where \( \tilde{Z}_{\nu,j} \) again are standard Normal random variables and independent for different \( \nu, j \in \{1, \ldots, d\} \times \{1, \ldots, N - 1\} \). Hence, by matching Equations (1.12)/(1.13) to Equation (1.3)
and setting $\omega(\tau, x) \equiv 1$ as before, we obtain

$$
(\hat{O}_\nu(\tau_i, x_i))_{1 \leq i \leq K} \times \tilde{N}_\nu \approx u_\nu \sqrt{\lambda_\nu}, \ 1 \leq \nu \leq d.
$$

(1.14)

Now Proposition 2 implies that we may examine each component $\sigma_\nu(\tau, x)$ and, hence, each eigenvector $u_\nu$, separately, $\nu \in \{1, \ldots, d\}$. To simplify notation, we assume $(t_{j+1} - t_j) = \Delta$, although similar relationships hold for non-equidistant data. From Equations (1.13) and (1.14), we obtain for small $l$ (here $l=1$)

$$
u \sqrt{\lambda_\nu} \approx (\hat{O}_\nu(\tau_i, x_i))_{1 \leq i \leq K} \times \tilde{N}_{\nu,j} = \left( \int_{\tau_i}^{\tau_i+l} C_\nu(x_i + s) \exp\{M_\nu s\} \, ds \right)_{1 \leq i \leq K} \times \tilde{N}_{\nu,j}
$$

$$
\approx (C_\nu(x_i + \tau_i + l/2) \times \exp\{M_\nu(\tau_i + l/2)\} \cdot l)_{1 \leq i \leq K} \times \exp\{M_\nu \Delta\} \times N_\nu
$$

$$
= (C_\nu(x_i + \tau_i + 1/2) \times \exp\{M_\nu(\tau_i + 1/2 + \Delta)\} \times N_\nu)_{1 \leq i \leq K}.
$$

(1.15)

Based on Equation (1.15), we are now able to estimate $C_\nu(x)$, $M_\nu$, and $N_\nu$ via regression. Note, however, that in doing so, we are only utilizing the variance part of $\hat{F}_l(t_j, t_{j+1})$, with all information on $\mathbb{E}[\hat{F}_l(t_j, t_{j+1})]$ being neglected (cf. Equation (2.2)). Furthermore, the underlying approximations may lead to a slight bias in our estimation of the parameters. Therefore, we are not going to finalize the estimation of the parameter values here, but rather rely on the gained insights to determine suitable functional assumptions for $C_\nu(\cdot)$ as well as structures for $M_\nu$ and $N_\nu$. The actual estimation of the parameter values based on the maximum likelihood approach will be detailed in the following subsection.
Moreover, a direct (unconstrained) regression brings about problems. More specifically, choosing $m_i \equiv 1$ heavily constrains possible shapes since the matrix exponential is one-dimensional, whereas $m_i > 1$ leads to identification problems. Thus, here we take a two step identification procedure: In the first step, we investigate $M_\nu$ and $N_\nu$ without specifying any functional assumption on $C_\nu(x)$ by relying on examples from interest rate modeling (see, e.g. Björk and Gombani (1999)) that are able to capture the term shapes displayed by the eigenvectors (cf. Figures 1.2, 1.3, and 1.4); in the second step, in order to reduce the number of parameters to make the calibration procedure tractable, we determine appropriate functional assumptions for $C_\nu(x)$, which can then be employed in the maximum likelihood estimation.

The Slope Factor Since the slope factor is observed across all sample populations, we impose the same assumption throughout all cases. We choose $m_1 = 2$ and set

$$C_1(x + \tau) = f(x + \tau) \times \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$$M_1 = \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix}, \text{ and}$$

$$N_1 = \begin{bmatrix} 1 - ab \\ a \end{bmatrix},$$
which is a slight modification of Example 6.2 in Björk and Gombani (1999), and obtain

\[
\sigma_1(\tau, x) = f(x + \tau) \times \begin{bmatrix} 0 & 1 \end{bmatrix} \times \exp \begin{bmatrix} -2b\tau & -b^2\tau \\ \tau & 0 \end{bmatrix} \times \begin{bmatrix} 1 - ab \\ a \end{bmatrix} \\
= f(x + \tau)(a + \tau) \exp(-b\tau).
\]

This functional form is specifically chosen to capture the increasing, concave shape of the “diagonal curves” observed in the surfaces.

With above specifications, from Equation (1.15) we can approximate

\[
u_1 \sqrt{\lambda_1} = \int_\tau^{\tau + l} \sigma_1(s + \Delta, x - \Delta) \, ds \approx \sigma_1(\tau + \Delta + \frac{l}{2}, x - \Delta) \cdot l \\
= f\left(x + \tau + \frac{l}{2}\right)\left(a + \tau + \Delta + \frac{l}{2}\right) \exp\left\{-b\left(\tau + \Delta + \frac{l}{2}\right)\right\} \cdot l.
\]

Notice that even with \(m_1 = 2\), \(f(x + \tau)\) is one-dimensional, and there only exists one driving Brownian motion. A non-parametric regression yields the function \(f(\cdot)\) (see Figure 1.5).

We find that in all cases, a logistic-Gompertz function,

\[
f(x) = k \times \frac{\exp(cx + d)}{(1 + \exp(cx + d))}
\]

is an appropriate choice to fit \(f(\cdot)\). To illustrate, in addition to the non-parametric regression function, Figure 1.5 displays the logistic-Gompertz functional fit from the nonlinear least-squares estimation for USA data (similar figures are obtained for other countries and are thus omitted).
Figure 1.5. Fitting $C_1(x)$ with Logistic-Gompertz function, USA data
The Twist Factor  Since the weight of the second eigenvalue is small relative to the first one, we assume $m_2 = 1$ for the sake of parameter parsimony. That is,

$$
\sigma_2(\tau, x) = C_2(x + \tau) \times \exp(M_2 \tau) \times N_2,
$$

$M_2 \in \mathbb{R}, N_2 \in \mathbb{R}$, therefore, $N_2$ can be further integrated into $C_2(x + \tau)$. Similarly as in the analysis of the first factor, we can approximate

$$
\sqrt{\lambda_2} u_2 = \int_{\tau}^{\tau+l} \sigma_2(s + \Delta, x - \Delta) \, ds \approx \sigma_2(\tau + \Delta + \frac{l}{2}, x - \Delta) \cdot l
$$

$$
= C_2\left(x + \tau + \frac{l}{2}\right) \exp\left\{M_2\left(\tau + \Delta + \frac{l}{2}\right)\right\} \cdot l.
$$

Furthermore, from a corresponding non-parametric regression (cf. Figure 1.6), we observe that the decreasing-then-increasing shape of $\sigma_2(\tau, x)$ can be well captured by choosing $C_2(\cdot)$ as the difference between two logistic-Gompertz functions

$$
C_2(x) = k_1 \frac{\exp(c_1 x + d_1)}{1 + \exp(c_1 x + d_1)} - k_2 \frac{\exp(c_2 x + d_2)}{1 + \exp(c_2 x + d_2)}.
$$

To further reduce the number of parameters, we require $c_1 = c_2$, and $k_1 - k_2 = 1$. The latter requirement is also useful in setting a possible upper bound of the volatility, which is implied by the natural boundedness of mortality rates. In addition to the non-parametric regression, Figure 1.6 shows the functional fit of $C_2(\cdot)$ from a nonlinear least-squares estimation for England & Wales as well as for USA data as an example.
1.3.3 Maximum Likelihood Estimation

Similarly as in the factor analysis, we rely on the quantities $F_l(t_j, t_{j+1}, (\tau, x)), (\tau, x) \in \tilde{C}$ as the basis for our estimation. In particular, we can express Equation (1.11) as:

$$F_l(t_j, t_{j+1}, (\tau, x)) = -\log \left\{ \frac{(t_j + \tau + t_{j+1})}{t_j + \tau} \frac{p_x(t_{j+1}, t_{j+1} + \tau + l)}{p_x(t_{j+1}, t_{j+1} + \tau)} \right\}$$

$$= \int_{t_j}^{t_{j+1}} \int_0^l \alpha(v + \tau + t_{j+1} - s, x - t_{j+1} + s) dv ds$$

$$+ \int_{t_j}^{t_{j+1}} \int_0^l \sigma(v + \tau + t_{j+1} - s, x - t_{j+1} + s) dv dW_s.$$ 

Therefore, with Equation (2.2), $F_l(t_j, t_{j+1}, (\tau, x))$ is Normal distributed with expected value

$$\mathbb{E}[F_l(t_j, t_{j+1}, (\tau, x))] = \int_{t_j}^{t_{j+1}} \int_0^l \sigma(v + \tau + t_{j+1} - s, x - t_{j+1} - s) \int_0^{v + \tau + t_{j+1} - s} \sigma'(u, x - t_{j+1} - s) du dv ds$$
and covariance structure

\[
\text{Cov} \left[ F_l(t_j, t_{j+1}, (\tau_1, x_1)), F_l(t_k, t_{k+1}, (\tau_2, x_2)) \right] = \\
\delta_{jk} \times \int_{t_j}^{t_{j+1}} \int_0^l \sigma(v + \tau_1 + t_{j+1} - s, x_1 - t_{j+1} + s) \, dv \int_0^l \sigma'(v + \tau_2 + t_{j+1} - s, x_2 - t_{j+1} + s) \, dv \, ds
\]

by a simple application of Itô’s product formula, in which \( \delta_{jk} \) equals 1 if \( j = k \) and 0 otherwise.

In particular, for \( t_{j+1} - t_j = \Delta \),\(^{13}\) the vectors \( \bar{F}_l(t_j, t_{j+1}) = (\omega(\tau, x) \times F_L(t_j, t_{j+1}, (\tau, x)_i))_{1 \leq i \leq K} \) are i.i.d. Normal with expected values

\[
\bar{\mu} = \left( \omega(\tau, x) \int_0^\Delta \left\{ \frac{1}{2} \int_{\tau+\Delta-s}^{\tau+\Delta+s} \sigma(u, x - \Delta + s) \, du \int_{\tau+\Delta-s}^{\tau+\Delta+s} \sigma'(u, x - \Delta + s) \, du \\
+ \int_{\tau+\Delta-s}^{\tau+\Delta+s} \sigma(u, x - \Delta + s) \, du \int_0^{\tau+\Delta-s} \sigma'(u, x - \Delta + s) \, du \right\} \, ds \right)_{(\tau, x) \in \tilde{C}}
\]

and covariance matrix \( \Sigma = (\Sigma_{ij})_{1 \leq i,j \leq K} \), where

\[
\Sigma_{ij} = \omega(\tau_i, x_i) \omega(\tau_j, x_j) \times \int_0^\Delta \int_0^l \sigma(v+\tau_i+\Delta-s, x_i-\Delta+s) \, dv \int_0^l \sigma'(v+\tau_j+\Delta-s, x_j-\Delta+s) \, dv \, ds.
\]

Similar ideas were applied in Bauer et al. (2008a) and Bauer (2009) for their maximum-likelihood calibration algorithms. However, as pointed out in their contributions, such an approach only allows for the consideration of a (very) small number of term/age combinations \( (\tau_i, x_i) \) (i.e. a small value of \( K \)) since (non-systematic) deviations are not admissible. In order to overcome this problem, we allow for non-systematic deviations in the “observed” vectors

\(^{13}\)Similarly as above, we consider this special case for notational convenience, while analogous results also apply for non-equidistant data.
\( \bar{F}^{\text{obs}}(t_j, t_{j+1}) \) from our model-endogenous vectors \( \bar{F}^{\text{mod}}(t_j, t_{j+1}) \). More specifically, we assume

\[
\bar{F}^{\text{obs}}(t_j, t_{j+1}) = \bar{F}^{\text{mod}}(t_j, t_{j+1}) + \xi_j,
\]

(1.16)

where \( \xi_j \) are mutually independent and independent of \( \bar{F}^{\text{mod}}(t_j, t_{j+1}) \), \( \xi_j \sim N(0, \alpha \cdot \text{diag}\{\Sigma\}) \), \( j = 1, \ldots, N - 1 \), in which \( \alpha \) is the sum of the weights of all other eigenvalues that are not considered within our model specification. Intuitively, the \( \xi_j \) pick up the variation not accounted for by the considered first \( d \) factors. Thus, we obtain

\[
\bar{F}^{\text{obs}}(t_j, t_{j+1}) \sim N(\bar{\mu}, \tilde{\Sigma}),
\]

(1.17)

where \( \tilde{\Sigma} = \Sigma + \alpha \cdot \text{diag}\{\Sigma\} \) and the log-likelihood function is of the form

\[
L(F^{\text{obs}}(t_1, t_2), \ldots, F^{\text{obs}}(t_{N-1}, t_N); C, M, N, \sigma_e) = \log \exp \left\{ -\frac{1}{2} \left( \bar{F}^{\text{obs}}(t_j, t_j) - \bar{\mu} \right)^T \tilde{\Sigma}^{-1} \left( \bar{F}^{\text{obs}}(t_j, t_j) - \bar{\mu} \right) \right\}
\]

\[
= \frac{1}{2} \left[ \sum_{j=2}^{N} - \log \left( \det(\tilde{\Sigma}) \right) - (F^{\text{obs}}(t_{j-1}, t_j) - \bar{\mu})^T \tilde{\Sigma}^{-1} (F^{\text{obs}}(t_{j-1}, t_j) - \bar{\mu}) \right] + \text{const}(1.18)
\]

To determine maximum likelihood estimates for our model parameters, it now suffices to determine the maximum values for \( \bar{L} \), which can be carried out numerically for each case. Similarly as in the factor analysis, here we assume \( \omega(\tau, x) = 1, \forall \tau, x \).
Table 1.6. The MLE results for female population – Lee-Carter Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Parameters</th>
<th>$\tilde{L}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$c$</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td>0.0026</td>
<td>0.0718</td>
</tr>
<tr>
<td>France</td>
<td>0.0015</td>
<td>0.0711</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0012</td>
<td>0.0696</td>
</tr>
<tr>
<td>United States</td>
<td>0.2145</td>
<td>0.0652</td>
</tr>
<tr>
<td>West Germany</td>
<td>0.0002</td>
<td>0.0712</td>
</tr>
</tbody>
</table>

**The Lee-Carter Approach**  Recall that $d = 1$ for female populations and $d = 2$ for male populations under the Lee-Carter forecasts. Table 1.6 and 1.7 display the estimated parameter values together with values of $\tilde{L}$ for female populations and male populations, respectively.

**The CBD-Perks Model**  For the CBD-Perks model, recall that $d = 1$ for both female and male populations. The estimated parameter values together with values of $\tilde{L}$ are displayed in Table 1.8 and 1.9 for female and male populations, respectively.

**The P-spline Method**  For the $P$-spline method, similarly, $d = 1$ for both female and male populations. The estimated parameter values together with values of $\tilde{L}$ are displayed in Table 1.10 and 1.11 for female and male populations, respectively.
### Table 1.7. The MLE results for male population – Lee-Carter Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Parameters</th>
<th>$\tilde{L}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$c$</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td>0.0020</td>
<td>0.0680</td>
</tr>
<tr>
<td>France</td>
<td>0.0007</td>
<td>0.0622</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0018</td>
<td>0.0466</td>
</tr>
<tr>
<td>United States</td>
<td>0.0012</td>
<td>0.2213</td>
</tr>
<tr>
<td>West Germany</td>
<td>0.0064</td>
<td>0.0714</td>
</tr>
<tr>
<td></td>
<td>$k_1$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td>-0.2442</td>
<td>0.0765</td>
</tr>
<tr>
<td>France</td>
<td>102.8762</td>
<td>0.9945</td>
</tr>
<tr>
<td>Japan</td>
<td>128.5159</td>
<td>0.1181</td>
</tr>
<tr>
<td>United States</td>
<td>3.5431</td>
<td>0.0605</td>
</tr>
<tr>
<td>West Germany</td>
<td>38.5996</td>
<td>0.4616</td>
</tr>
</tbody>
</table>

### Table 1.8. The MLE results for female population – CBD-Perks Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Parameters</th>
<th>$\tilde{L}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$c$</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td>0.5979</td>
<td>0.1094</td>
</tr>
<tr>
<td>France</td>
<td>0.1275</td>
<td>0.0987</td>
</tr>
<tr>
<td>Japan</td>
<td>2.0325</td>
<td>0.1018</td>
</tr>
<tr>
<td>United States</td>
<td>0.0669</td>
<td>0.0885</td>
</tr>
<tr>
<td>West Germany</td>
<td>0.3768</td>
<td>0.1053</td>
</tr>
<tr>
<td>Country</td>
<td>Parameters</td>
<td>( \tilde{L} )</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>England &amp; Wales</td>
<td>0.3031 0.1040 -19.2400 1.5103 \times 10^3 -0.0101 -3.8873 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.1411 0.0858 -16.5756 825.8940 -0.0090 -4.5835 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.0454 0.0975 -16.7923 1.2296 \times 10^3 -0.0062 -4.5510 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>0.0026 0.0735 -10.1936 182.9198 -0.0064 -4.5211 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>West Germany</td>
<td>0.1905 0.0971 -15.6849 111.2805 -0.0064 -3.8650 \times 10^4</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.9. The MLE results for male population – CBD-Perks Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Parameters</th>
<th>( \tilde{L} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>England &amp; Wales</td>
<td>0.6613 0.1122 -17.6146 50.3170 -2.8369 \times 10^{-5} -5.9634 \times 10^3</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.1085 0.1066 -17.4736 437.0587 -0.0103 -4.0801 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.0769 0.0996 -16.2165 222.0080 0.0013 -5.3582 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>0.6614 0.0945 -16.1741 42.6901 -0.0050 -3.4813 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>West Germany</td>
<td>2.0612 0.1128 -18.3485 25.1029 0.0059 -4.3360 \times 10^4</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.10. The MLE results for female population – P-spline Model

<table>
<thead>
<tr>
<th>Country</th>
<th>Parameters</th>
<th>( \tilde{L} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>England &amp; Wales</td>
<td>0.2103 0.0997 -17.9921 632.1372 -0.0128 -4.0891 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.3584 0.0859 -17.3926 918.8431 -0.0088 -3.6422 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>21.1012 0.0965 -20.1085 75.6164 -0.0015 -4.0413 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>0.1352 0.0737 -16.1479 1.2766 \times 10^3 -0.0128 -4.5127 \times 10^4</td>
<td></td>
</tr>
<tr>
<td>West Germany</td>
<td>1.3299 0.0922 -16.6010 20.0354 0.0041 -5.3963 \times 10^4</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.11. The MLE results for male population – P-spline Model
1.3.4 Should the Self-consistency Condition be Imposed?

In the literature on modeling the term structure of interest rate, the necessity—and even the conduciveness—of imposing cross-sectional constraints is heavily debated. While some argue it will increase the efficiency of estimates, others uphold that models not satisfying cross-sectional constraints produce more accurate forecasts (see, among others, Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006), Christensen et al. (2011), and Piazzesi (2010)). Recent contributions by Joslin et al. (2011) and Duffee (2011) add some theoretical substance to this discussion. More specifically, Joslin et al. (2011) show that for Gaussian term structure models without any restrictions on risk premium dynamics, no-arbitrage restrictions are irrelevant to estimating factor dynamics. The intuition is that for arbitrary risk premium dynamics, the factor dynamics and cross-sectional equations are sufficiently unrelated. Duffee (2011), on the other hand, even asserts that no-arbitrage restrictions are unnecessary to estimate the cross-sectional mapping. The key idea here is that the cross-sectional equations can be estimated with very high accuracy, such that restrictions only bite for the cross-sectional estimation if they are inconsistent with the true cross-sectional patterns.

It is important to note that the argument from Joslin et al. (2011) does not apply in our context since there is no embedded measure change, i.e. expectations in the cross-sectional direction are taken under the physical, data-generating measure. Hence, the common intuition that cross-sectional restrictions will help to improve the efficiency of estimates as e.g.
explained in Piazzesi (2010) holds true in the current mortality setting.\(^{14}\) This warrants a positive answer to the question raised in the headline of this section, i.e. there are good reasons to consider self-consistent forward mortality models.

However, the insight from Duffee (2011) that imposing cross-sectional constraints should not invalidate estimates resulting in their absence as long as these constraints are satisfied in the data still holds true. In other words, if the estimates significantly change after the self-consistency condition is imposed, this is an indication that the underlying forecasting model is misspecified. In particular, this allows us to test the self-consistency of different forecasting approaches.

More precisely, we focus on the factor \(Y_\nu(j)\) from Equation (1.4), where for simplicity and without loss of generality, in what follows we only look at the one factor case and thus omit all subscripts. Under the functional assumption from this section, the \(Y(j)\) is Normal distributed with mean

\[
\mu_Y \approx \left( \int_{\tau}^{\tau+1} \sigma_1(s + \Delta, x - \Delta) \, ds \right)^T \times \mathbb{E}[\bar{F}]
\]

\[
= \left( \int_{\tau}^{\tau+1} \frac{k \exp(c(x + s) + d)}{1 + \exp(c(x + s) + d)} (a + s + \Delta) \exp(-b(s + \Delta)) \, ds \right)^T \times \mathbb{E}[\bar{F}] \quad \text{(1.19)}
\]

\(^{14}\)Piazzesi (2010) also lists a few other reasons why cross-equations restrictions should be taken into account, some of which also apply in the mortality case.
and standard error

\[ \sigma_Y \approx \left( \int_{\tau}^{\tau+1} \sigma_1(s + \Delta, x - \Delta) \, ds \right)^T \left( \int_{\tau}^{\tau+1} \sigma_1(s + \Delta, x - \Delta) \, ds \right) = \sum_{\tau,x} \left( \int_{\tau}^{\tau+1} \frac{k \exp(c(x + s) + d)}{1 + \exp(c(x + s) + d)} (a + s + \Delta) \exp(-b(s + \Delta)) \, ds \right)^2. \]  

(1.20)

It is important to note that unlike the MLE in the previous subsection, here we do not impose the self-consistency condition on \( \mu_Y \) and \( \sigma_Y \). The procedure then becomes clear: We have already obtained estimates with considering the self-consistency condition in Section 1.3.3. By estimating parameter values \( k, c, d, a, \) and \( b \) from maximum likelihood estimation using Equation (1.19) and (1.20) in the corresponding Gaussian distribution, we further obtain estimates without considering the self-consistency condition. Eventually, we are able to compare estimates for \( \mu_Y \) and \( \sigma_Y \) in three cases: (1) Directly calculated as the sample mean and standard error from the data set denoted by \( (\mu_s^Y, \sigma_s^Y) \); (2) estimated based on the specific functional assumption on \( \sigma(\tau, x) \) but without the self-consistency condition in place denoted by \( (\mu_u^Y, \sigma_u^Y) \); and (3) estimated via the MLE from the previous subsection under the specific functional assumption on \( \sigma(\tau, x) \) with the self-consistency condition in place denoted by \( (\mu_c^Y, \sigma_c^Y) \). For an illustration, similar to Section 1.2.4, we again focus on female US data.

Table 1.12 shows the estimated parameter values from the unconstrained MLE—i.e. the parameter estimates underlying \( (\mu_u^Y, \sigma_u^Y) \)—under all three underlying mortality projection methodologies. Table 1.13 shows the estimated mean and standard deviation for the \( Y(j) \) for all three estimation approaches, i.e. \( (\mu_u^Y, \sigma_u^Y) \), \( (\mu_r^Y, \sigma_r^Y) \), and \( (\mu_s^Y, \sigma_s^Y) \), as well as the corresponding confidence interval for \( (\mu_s^Y, \sigma_s^Y) \) under all three forecasting methodologies.
### Methodology

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$k$</th>
<th>$c$</th>
<th>$d$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lee-Carter</strong></td>
<td>0.0714</td>
<td>0.1855</td>
<td>−23.5672</td>
<td>8.0636</td>
<td>0.0053</td>
</tr>
<tr>
<td><strong>CBD-Perks</strong></td>
<td>0.0015</td>
<td>0.1067</td>
<td>−11.9913</td>
<td>16.9687</td>
<td>0.0097</td>
</tr>
<tr>
<td><strong>P-spline</strong></td>
<td>0.0200</td>
<td>0.1449</td>
<td>−17.1565</td>
<td>−0.8585</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Table 1.12. MLE without Self-Consistency Condition

<table>
<thead>
<tr>
<th>Methodology</th>
<th>$\mu_1^Y$</th>
<th>$\sigma_Y^Y$</th>
<th>$\mu_Y^\mu$</th>
<th>$\sigma_Y^\mu$</th>
<th>$\mu_Y^\xi$</th>
<th>$\sigma_Y^\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lee-Carter</strong></td>
<td>0.0157</td>
<td>0.0235</td>
<td>0.0156</td>
<td>0.0235</td>
<td>0.0170</td>
<td>0.0282</td>
</tr>
<tr>
<td></td>
<td>(0.0047, 0.0266)</td>
<td>(0.0184, 0.0347)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CBD-Perks</strong></td>
<td>0.0016</td>
<td>0.0204</td>
<td>0.0016</td>
<td>0.0204</td>
<td>0.0043</td>
<td>0.1418</td>
</tr>
<tr>
<td></td>
<td>(−0.0079, 0.0111)</td>
<td>(0.0160, 0.0302)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P-spline</strong></td>
<td>−0.0249</td>
<td>0.1698</td>
<td>−0.0249</td>
<td>0.1698</td>
<td>−0.0575</td>
<td>0.9968</td>
</tr>
<tr>
<td></td>
<td>(−0.1041, 0.0543)</td>
<td>(0.1331, 0.2512)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.13. Comparison of Mean and Standard Variance
Rather unsurprisingly, the two unconstrained estimates \((\mu^u_Y, \sigma^u_Y)\) and \((\mu^s_Y, \sigma^s_Y)\) are very close and the marginal differences may be attributed to numerical errors. However, this is not the case for the constrained estimates \((\mu^c_Y, \sigma^c_Y)\). Here the \((\mu^c_Y, \sigma^c_Y)\) are only close to the \((\mu^s_Y, \sigma^s_Y)\)—and comfortably fall into the 95% confidence interval—for the Lee-Carter method. For the other forecasting approaches, \(\sigma^c_Y\) considerably exceeds the upper bound of the corresponding confidence interval. This indicates that for the underlying data set, only the Lee-Carter method produces self-consistent forecasts. In particular, our results endorse the use of the Lee-Carter methodology for producing (deterministic) mortality forecasts.

1.4 A Non-negative Model Variant

While the deterministic assumption on \(\Sigma\) within our basic evolution Equation (1.1) and the resulting Gaussian distribution of \(\mu_t\) greatly facilitate the specification and estimation process, as indicated above it inevitably entails the theoretical shortcoming that survival probabilities may exceed unity. Hence, at least from a theoretical perspective, a non-negative model would be desirable, especially if it retains some of the tractability.

In order to devise such a non-negative model that still remains compatible with the data, we integrate a non-negative affine process into our model while preserving its main characteristics. This approach again shows similarities to interest rate modeling, where the transformation from the popular Vasicek model (Vasicek (1977)) to the equally popular CIR model (Cox et al. (1985)) can be plainly interpreted as a non-negative modification that preserves the main features. Again, we focus on the one-factor version of our model with the assumptions from Section 1.3.3—and particularly the self-consistency condition—in place.
Combining Equation (1.9) and the functional form from Section 1.3.2, the spot force of mortality can be written as

$$\mu_t(0, x) = \mu_0(t, x - t) + \int_0^t \alpha(t - s, x - t + s) \, ds + \frac{k \exp(cx + d)}{1 + \exp(cx + d)} \times Z_t^{(2)},$$

in which $Z_t^{(2)}$ is the second component of $Z_t$. Therefore, although being a two-dimensional vector, $Z_t$ affects the spot force of mortality only via its second component, so that we only need to consider a non-negative process related with $Z_t^{(2)}$.

The dynamics of $Z_t$ as defined in Equation (1.9) take the form

$$dZ_t = \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix} \times Z_t \, dt + \begin{bmatrix} 1 - ab \\ a \end{bmatrix} \, dW_t.$$

Now, with $b > 0$, $Z_t$ is a mean-reverting process with long-term mean at $[0, 0]'$, thus, it seems not appropriate to directly convert $Z_t^{(2)}$ to a non-negative process. Hence, we define a new state factor $\tilde{Z}_t = (Z_t^{(1)} + \xi_1, Z_t^{(2)} + \xi_2)'$ with both $\xi_1$ and $\xi_2 > 0$, so that the diffusion process of $\tilde{Z}_t$ can be written as

$$d\tilde{Z}_t = \begin{bmatrix} 2\xi_1 b + \xi_2 b^2 \\ -\xi_1 \end{bmatrix} + \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix} \times \tilde{Z}_t \, dt + \begin{bmatrix} 1 - ab \\ a \end{bmatrix} \, dW_t.$$

This in turn implies

$$\mu_t(0, x) = \mu_0(t, x - t) + \int_0^t \alpha(t - s, x - t + s) \, ds - \xi_2 \times \frac{k \exp(cx + d)}{1 + \exp(cx + d)} + \frac{k \exp(cx + d)}{1 + \exp(cx + d)} \times \tilde{Z}_t^{(2)}.$$
By rewriting the diffusion process of $\tilde{Z}_t$ to

$$d\tilde{Z}_t = \begin{cases} 
\begin{bmatrix} 2\xi_1 b + \xi_2 b^2 \\ -\xi_1 \end{bmatrix} + \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix} \times \tilde{Z}_t 
\end{cases} \begin{bmatrix} \frac{1 - ab}{a \times \sqrt{\tilde{Z}_t^{(2)}}} \end{bmatrix} dt + \begin{bmatrix} \frac{k \exp(c(x + \tau) + d)}{1 + \exp(c(x + \tau) + d)} \\
2b \xi_2 \end{bmatrix} f_{\tilde{Z}_t^{(2)}},
\end{cases}$$

we now change $\tilde{Z}_t^{(2)}$ to a non-negative process. Since the above formulation still satisfies the (generalized) definition of affine diffusion process (cf. Duffie et al. (2000)), the (forward) survival probabilities $s_{p_x}(t)$ can be formulated of the form

$$s_{p_x}(t) = \exp \left( A_x(s) + B_x(s)^T \tilde{Z}_t \right),$$

where the coefficients $A_x(s) \in \mathbb{R}$ and $B_x(s) \in \mathbb{R}^2$ solve the well-known Riccati ODEs

$$A_x'(\tau) = -\mu_0(\tau, x - \tau) - \int_0^\tau \alpha(\tau - v, x - \tau + v) dv + \xi_2 \times \frac{k \exp(c(x + \tau) + d)}{1 + \exp(c(x + \tau) + d)} + (2\xi_1 b + \xi_2 b^2) B_x^{(1)}(\tau) - \xi_1 B_x^{(2)}(\tau),$$

$$B_x'(\tau) = - \begin{bmatrix} \frac{k \exp(c(x + \tau) + d)}{1 + \exp(c(x + \tau) + d)} \\
2b \xi_2 \xi_1 \\
\frac{b^2}{1 + \exp(c(x + \tau) + d)} \end{bmatrix} - \begin{bmatrix} 0 & 2b & -1 \\
2b & -1 & 0 \\
0 & b^2 & 0 \end{bmatrix} B_x(\tau) + \begin{bmatrix} 0 \\
\frac{1}{2}((1 - ab) B_x^{(1)}(\tau) + a B_x^{(2)}(\tau))^2 \end{bmatrix},$$

with starting value $A_x(0) = 0$ and $B_x(0) = 0$.

Therefore, we are given a simply and—under restrictions on the initial mortality surface $\mu_0(\cdot, \cdot)$—non-negative modification of our model that preserves the main characteristics.

\(^{15}\)Note that this process does not fall into the affine class as defined by some authors such as Biffis (2005) or Piazzesi (2010).
Moreover, due to its affine-structure, it remains tractable. In particular, for the calculation of survival probabilities it is sufficient to solve simple ODEs.

In order to estimate parameter values, note that since $\tilde{Z}_t^{(2)}$ now enters the volatility component, $\{\tilde{Z}_{t_{j+1}} - \tilde{Z}_{t_j}\}_{j=1,...,N}$ are not time-homogeneous. Therefore, $\bar{F}_t(t_j, t_{j+1})$ are no longer i.i.d. and we cannot use the maximum likelihood approach as in the Gaussian case. This problem can be overcome by the use of an unscented Kalman filter (cf. Wan and Van Der Merwe (2000)), with $\tilde{Z}_t$ as the unobserved state vector. For illustration, Table 1.14 displays the estimated parameter values for female U.S. population under the Lee-Carter forecasts.

### 1.5 Application

To illustrate the results generated by our models—and particularly to compare these results to conventional mortality forecasting approaches—in this section we consider a simple example application in which the risk in mortality projections is important. More specifically, we calculate confidence intervals for the expected future life-time for different age cohorts of the US female population one year from the terminal point of the underlying mortality time series (2006, time zero). We compare confidence intervals generated based on four different

<table>
<thead>
<tr>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
</tr>
<tr>
<td>$c$</td>
</tr>
<tr>
<td>$d$</td>
</tr>
<tr>
<td>$a$</td>
</tr>
<tr>
<td>$b$</td>
</tr>
<tr>
<td>$\xi_1$</td>
</tr>
<tr>
<td>$\xi_2$</td>
</tr>
<tr>
<td>$\tilde{Z}_0^{(1)}$</td>
</tr>
<tr>
<td>$\tilde{Z}_0^{(2)}$</td>
</tr>
</tbody>
</table>

Table 1.14. Non-negative estimation: U.S. female & Lee-Carter
approaches: (1) The stochastic version of the underlying mortality projection methodology,\textsuperscript{16} (2) the unrestricted simple factor model as described in Section 1.2.4, (3) the self-consistent forward mortality factor model as described in Section 1.3.2, and (4) the non-negative spot force model as described in Section 1.4. We focus on the Lee-Carter method as the underlying forecasting approach as it is the only methodology that yields self-consistent forecasts as demonstrated in Section 1.3.4.

Figure 1.7 displays the confidence intervals of life expectancies at time one $e_{x+1}$ for ages 20, 30, 40, 50, 60, and 70, and all four approaches described above. First, we observe that in the first three cases, the length of the 95\% confidence intervals for the life expectancy one year from now is less than two-and-one half per-cent of the corresponding median value and even less than two per cent for ages under 50. This may not come as a surprise given the one year forecasting period. Also, the median projections are very close in all three cases, which again is not surprising since the Lee-Carter forecasts underlie all three approaches. On the other hand, the median projections are relatively higher for the non-negative model case in order to compensate the one-sided outliers of life expectancies, a property naturally induced by the capping of survival probabilities (at 1).

Furthermore, we find that the intervals differ considerably among the four different approaches. More precisely, while the confidence intervals behave similarly among the simple factor model, the forward mortality model, and the non-negative model at each representative age, they are considerably wider than the confidence intervals that are generated from the stochastic Lee-Carter approach. This indicates that the error terms in the Lee-Carter do not

\textsuperscript{16}For the first approach, for consistency the parameters are estimated from the last thirty-year time window, i.e. from year 1977-2006.
Figure 1.7. Simulated Life Expectancy After One Year
accurately reflect the persistence of the random sources. This interpretation is in line with the observation that the difference between the confidence intervals is particularly pronounced for young ages, where the long end of the mortality projections are important. For older ages, on the other hand, where trend risk plays a smaller role, the approaches come closer.

Hence, all in all, we find that conventional mortality forecasting approaches, which are designed to optimally fit past mortality experience in the short end, fail to identify the persistence of the error terms, and thus generally underestimate the risk in mortality projections.

1.6 Conclusion

Having a suitable understanding of the risk in mortality projections is important at the individual level for retirement planning; at the corporate level for pension liabilities; and the societal level in view of intergenerational risk sharing. However, conventional mortality forecasting approaches fail to capture this risk in a consistent manner. This is by design since these models ubiquitously aim at as good as possible modeling mortality experience and associated error terms. In contrast, in this paper we attempt to develop coherent models for the risk in mortality projections.

This is accomplished by analyzing time-series of mortality forecasts that are generated from a rolling window of annual mortality data and some fixed forecasting methodology rather than directly considering the underlying data for historical mortality experience. A factor analysis shows that one to two factors are sufficient to represent the majority of the variation in the data. Moreover, the factors exhibit very consistent shapes across populations
and underlying forecasting methodologies. Resulting factor models are simple and easy to estimate, but they do not necessarily adhere to self-consistency conditions that originate from the interpretation of the data as forecasts. In contrast, so-called forward mortality models satisfy these conditions. By relying on a semi-parametric representation that encompasses all coherent models under Gaussian random noise, we devise suitable models with few parameters and estimate them via a maximum likelihood approach. These models thus present coherent, parsimonious, and tractable workhorses when the appraisal of mortality risks within medium- to long-term projections plays a dominant role. We discuss extensions and example applications, which demonstrate the relative deficiencies of conventional approaches.

In the paper, we treat sample data with different genders and from different countries separately in the principal component analysis. An alternative approach could be the use of common principal component, with which we could deal with multiple populations in one single analysis. As the next step, we intend to study the consequences of the improved risk estimates. Specifically, we will consider an individual’s optimal annuitization choice in the presence of aggregate mortality risk based on a life-cycle model, where the agent maximizes her lifetime utility over consumption, asset portfolio, and annuitization choices.
1.7 Appendix: Proofs

Proof of Proposition 1

We have:

\[ F_l(t_j, t_{j+1}, (\tau, x)) = \int_\tau^{\tau+l} \mu_{t_{j+1}}(v, x) dv - \int_\tau^{\tau+l+t_{j+1}-t_j} \mu_{t_j}(v, x - t_{j+1} + t_j) dv \]
\[ = \int_\tau^{\tau+l} \mu_{t_j}(v + t_{j+1} - t_j, x - t_{j+1} + t_j) + \int_{t_j}^{t_{j+1}} \Lambda(v + t_{j+1} - s, x - t_{j+1} + s) ds \]
\[ + \int_{t_j}^{t_{j+1}} \Sigma(v + t_{j+1} - s, x - t_{j+1} + s) dW_s dv - \int_{\tau+t_{j+1}-t_j}^{\tau+l+t_{j+1}-t_j} \mu_{t_j}(v, x - t_{j+1} + t_j) dv \]
\[ = \int_{t_j}^{t_{j+1}} \int_\tau^{\tau+l} \Lambda(v + t_{j+1} - s, x - t_{j+1} + s) dv ds + \int_{t_j}^{t_{j+1}} \int_{t_j}^{\tau+l} \Sigma(v + t_{j+1} - s, x - t_{j+1} + s) dv dW_s \]
\[ \overset{d}{=} \int_0^{t_{j+1}-t_j} \int_\tau^{\tau+l} \Lambda(v + t_{j+1} - t_j - s, x - (t_{j+1} - t_j) + s) dv ds \]
\[ + \int_0^{t_{j+1}-t_j} \int_{t_j}^{\tau+l} \Sigma(v + t_{j+1} - t_j - s, x - (t_{j+1} - t_j) + s) dv dW_s, \]

so that obviously \( F_l(t_j, t_{j+1}) \) is independent for different \( j \), Gaussian and i.i.d. if \( t_{j+1} - t_j = \Delta \).
Proof of Proposition 2

\[ \begin{align*}
[\sigma_1(\tau,x), \ldots, \sigma_d(\tau,x)] &= [C_1(x+\tau) \times \exp\{M_1 \tau\} \times N_1, \ldots, C_d(x+\tau) \times \exp\{M_d \tau\} \times N_d] \\
&= \left[ C_1(x+\tau), \ldots, C_d(x+\tau) \right] \times \\
&\quad \begin{pmatrix}
\exp\{M_1 \tau\} & 0 & \cdots & 0 \\
0 & \exp\{M_2 \tau\} & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \exp\{M_d \tau\}
\end{pmatrix} \times \\
&\quad \begin{pmatrix}
N_1 & 0 & \cdots & 0 \\
0 & N_2 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & N_d
\end{pmatrix}_N \\
&= C(x+\tau) \times \exp\{\text{diag}\{M_1, \ldots, M_d\} \tau\} \times N.
\end{align*} \]

\[\square\]
APPLICATIONS OF FORWARD MORTALITY FACTOR MODELS IN LIFE INSURANCE PRACTICE

2.1 Introduction

Two of the most important challenges for the application of stochastic mortality models in life insurance practice are the apparent incompatibility of most stochastic methods with classical life contingencies theory, which presents the backbone of insurers’ Electronic Data Processing (EDP) systems, and the complexity of many of the proposed approaches. These obstacles have not only led to an increasing discrepancy between life insurance research and actuarial practice and education in some parts of the world, but the reluctance of practitioners to rely on stochastic mortality models may also be a primary reason for the sluggish development of the mortality-linked capital market. Specifically, stochastic methods are necessary to assess a company’s capital relief when hedging part of its mortality risk exposure, which should be one of the key drivers of the demand for mortality-linked securities.

One model class that overcomes these problems are so-called forward mortality models, which infer dynamics on the entire age/term-structure of mortality. As already pointed out by Milevsky and Promislow (2001), the “traditional rates used by actuaries” really are forward rates so that the forward approach can be viewed as the natural extension of traditional actuarial theory. In particular, the actuarial present values for traditional insurance
products such as term-life insurance, endowment insurance, or life annuity contracts are of the same form as in classical actuarial theory, where the “survival probabilities” now are to be interpreted as expected values of realized survival probabilities (cf. Bauer et al. (2012)). Hence, the inclusion of such models in the operations of a life insurer or a pension fund will not require alterations in the management of traditional product lines, but nonetheless present a coherent way to take mortality risk into account when necessary. Examples of such situations include the calculation of economic capital based on internal models or the pricing and risk management of mortality-linked guarantees in life insurance or pension products.

However, only few forward mortality models have been proposed so far, and most authors have relied on “qualitative” insights and/or modeling convenience for determining suitable specifications (cf. Bauer et al. (2008a), Dawson et al. (2010), or Schrager (2006)). Moreover, some of the presented models entail a high degree of complexity, which may lead to problems in their calibration (see e.g. Bauer et al. (2008a)).

In the companion paper Zhu and Bauer (2011b), we present an alternative, data-driven approach by relying on forward mortality factor models with Normal-distributed transition factors, the (necessary) explicit functional form for which has been identified by Bauer et al. (2012). More specifically, we use principal component analyses of time series of mortality forecasts generated based on rolling windows of annual mortality data to derive a suitable number of stochastic factors and their functional forms. The resulting specifications are then (re)calibrated based on maximum likelihood estimation (the main ideas and relevant results are summarized in Section 2.2 and Section 2.6 below). In this paper, we demonstrate the technical advantages of this model class by discussing and implementing several important
example applications. Furthermore, our numerical results based on a simple model version calibrated to British population mortality data illustrate the economic significance of systematic mortality risk.

The first application concerns the calculation of economic capital for life insurance companies. After providing a framework for this problem similar to that from Bauer et al. (2010a), we explicitly demonstrate how to derive the economic capital for a stylized life insurance company offering traditional life insurance products in our setting. Our implementation highlights the tractability of forward mortality factor models as well as the important advantage of this model class in that it avoids the necessity of nested simulations.\(^1\) Furthermore, our numerical results display that in addition to financial risk, (systematic) mortality risk has a considerable impact on the results and thus plays an important role for a life insurance company’s solvency.

In the second application, we discuss the valuation of different mortality-contingent embedded options within life insurance contracts. Specifically, we derive a closed-form valuation formula for simple Guaranteed Annuity Options (GAOs) within traditional endowment policies in the considered forward mortality model framework. Moreover, we demonstrate how to derive the fair option fee for Guaranteed Minimum Income Benefits (GMIBs) within Variable Annuity contracts based on Monte Carlo simulations; here, akin to the first application, forward mortality models bear the profound advantage that no nested simulations are necessary. Our numerical results again emphasize the economic significance of systematic mortality risk.

\(^1\)For an overview on Monte Carlo methods in financial modeling, see e.g. Glasserman (2004).
The remainder of the paper is organized as follows: Section 2.2 provides a short summary of the considered mortality modeling framework from Zhu and Bauer (2011b) and introduces the model underlying our implementations. Section 2.3 and 2.4 present our application to economic capital modeling and the valuation of annuitization options, respectively. Finally, Section 2.5 concludes.

2.2 Forward Mortality Factor Models

In a best estimate generation life table at time \( t \geq 0 \), forward survival probabilities \( \tau p_x(t, t+\tau) \) are listed for a (large) collection of ages \( x \geq 0 \) and terms \( \tau \geq 0 \), where \( \tau p_x(t, T+\tau) \) is the \( F_t \)-measurable random variable satisfying \(^2\)

\[
\tau p_x(t, T + \tau) \mathbb{1}_{\{\Upsilon_{x-T} > T\}} = \mathbb{E}_P^\mathcal{F}_t \left[ \mathbb{1}_{\{\Upsilon_{x-T} > T + \tau\}} \mathcal{F}_t \lor \{\Upsilon_{x-T} > T\} \right], \quad 0 \leq T \leq t \leq T + \tau,
\]

and \( \Upsilon_{x_0} \) is the (random) time of death or future lifetime of an \( x_0 \)-year old at time zero. Hence, \( \tau p_x(t, T + \tau) \) denotes the—at time \( t \)—expected \( \tau \)-year survival probability for an \( x \)-year old at time \( T \), whereas \( \tau p_x(T + \tau, T + \tau) \) denotes the corresponding realized survival probability.

We introduce the so-called forward force of mortality

\[
\mu_t(\tau, x) = -\frac{\partial}{\partial \tau} \log \{\tau p_x(t, t + \tau)\}
\]

\(^2\)As usually in this context, underlying our considerations is a filtered probability space \( (\Omega, \mathcal{H}, \mathbf{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}) \). Here, the filtration \( \mathbf{F} \) satisfies the usual conditions and models the information flow of aggregate population dynamics, whereas the sigma algebra \( \mathcal{H} \) also contains information about individual deaths within the population. We refer to Bauer et al. (2012) for technical details.
as a—from a modeling perspective—convenient representation of the forward survival probabilities constituting the generation life table at \( t \), so that we have

\[
\tau p_x(t, t + \tau) = \exp \left\{ - \int_0^\tau \mu_t(s, x) \, ds \right\}.
\]

A (forward) mortality model now specifies the evolution of the generation life tables

\[
(\tau p_x(t, t + \tau), x, \tau \geq 0)_{t \geq 0},
\]

or equivalently \((\mu_t(\tau, x), x, \tau \geq 0)_{t \geq 0}\), over time, which can be formulated as a stochastic (differential) equation of the form (cf. Bauer et al. (2012))

\[
d\mu_t = (A \mu_t + \alpha_t) \, dt + \sigma_t \, dW_t, \quad \mu_0(\cdot, \cdot) > 0, \quad (2.1)
\]

where \( \alpha_t \) and \( \sigma_t \) are adequate, function-valued stochastic processes, \( A = \frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \), and \((W_t)\) is a \( d \)-dimensional Brownian motion. Furthermore, if the dynamics (2.1) are formulated under \( \mathbb{P} \), we have the drift condition (cf. Cairns et al. (2006a), Bauer et al. (2012))

\[
\alpha_t(\tau, x) = \sigma_t(\tau, x) \times \int_0^\tau \sigma'_t(s, x) \, ds. \quad (2.2)
\]

Hence, to specify a model, it is sufficient to specify a suitable volatility structure \((\sigma_t(\tau, x), x, \tau \geq 0)_{t \geq 0}\) in \( \mathbb{R}^d \).
In Zhu and Bauer (2011b), we consider time-homogenous models of this type for which transitions can be realized by Normal-distributed random vectors, i.e. models of the form

$$\mu_t(\tau, x) = G(\tau, x, Z_t),$$

for some Normal-distributed random vector $Z_t$, which have been studied in detail in Bauer et al. (2012). In particular, the authors show that the volatility structure must necessarily be of the form

$$\sigma(\tau, x) = C(x + \tau) \times \exp \{ M \tau \} \times N,$$

where $N \in \mathbb{R}^{m \times d}$, $M \in \mathbb{R}^{m \times m}$, and $C' \in C^1 ([0, \infty), \mathbb{R}^m)$, so that

$$\mu_t(\tau, x) = \mu_0(\tau + t, x - t) + \int_0^t \alpha(\tau + t - s, x - t + s) \, ds$$

$$+ C(x + \tau) \exp \{ M \tau \} \int_0^t \exp \{ M (t - s) \} \, N dW_s.$$ 

(2.4)

While the confinement to Normal distributions implies the theoretical shortcoming that realizations of survival probabilities—with a small probability—may exceed unity, it allows us to derive adequate specifications using a principal component analysis. Furthermore, as also shown in Zhu and Bauer (2011b), it is easy to devise a maximum likelihood approach for the (re)calibration of the resulting specifications that explicitly takes the drift condition (2.2) into account. To cast suitable specifications, we rely on mortality forecasts generated using different forecasting methods and rolling windows of annual mortality data for various countries. Our analyses show that the first one to two principal components capture the
great majority of all the variation in the data, and the corresponding error terms exhibit
systematic shapes that can be captured by few parameters.

In this paper, we also limit our focus to these type of models satisfying (2.3) since
suitable specifications are immediately available from Zhu and Bauer (2011b), but also due
to important advantages in applications that will be illustrated throughout this text. For our
implementations, we adopt a parsimonious (one-factor) model version, where the underlying
mortality forecasts were generated using the Lee and Carter (1992) methodology\(^3\) based on
male mortality data from England and Wales for the years 1947 to 2006 as available from
the *Human Mortality Database*.\(^4\) More precisely, we use observations for ages between 20
and 95 years to compile thirty-one consecutive generation life tables (1977-2007) each relying
on the mortality experience of the previous 30 years with the Lee-Carter parameters \(\{\alpha_x\},
\{\beta_x\}, \text{ and } \{\kappa_t\}\) calibrated independently and the random-walk drift for \(\kappa_t\) re-estimated for
each 30-year window. Hence, the first table uses mortality data from 1947-1976, the second
one uses 1948-1977, and so forth. As described in Zhu and Bauer (2011b), the resulting
time series of tables can then be used to derive suitable specifications based on a principle
component analysis. The results are displayed in Figure 2.1; more details on the underlying
procedure are provided in Section 2.6.

The top left panel (Fig. 2.1(a)) shows the volatility \(\sigma\) associated with the first principal
component as a function of \(\tau\) and \(x\), which explains more than 91% of all the variation in
the data. The (non-parametric) least-squares fit of \(C\) is shown in the top right panel (Fig.
\(^3\)More specifically, \(\{\alpha_x\}\) and \(\{\beta_x\}\) in the Lee-Carter model are calibrated via the weighted least-squares
algorithm, and \(\{\kappa_t\}\) is further adjusted by fitting a Poisson regression model (cf. Booth et al. (2002)).
\(^4\)Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for
Demographic Research (Germany). Available at [www.mortality.org](http://www.mortality.org) or
[www.humanmortality.de](http://www.humanmortality.de).
2.1(b)). The shape can be captured well by the specification (see also Björk and Gombani (1999))

\[
\sigma(\tau, x) = \frac{k \exp\{c(x + \tau) + d\}}{1 + \exp\{c(x + \tau) + d\}} (a + \tau) e^{-b\tau}
\]

(2.5)

The parameters are subsequently (re)calibrated based on a Maximum Likelihood Estimation (MLE) method (see also Appendix 2.6 for more details), with the resulting estimates displayed in Table 2.1. Figure 2.1(c) shows the projection of the first principal component based on the MLE, while Figure 2.1(d) depicts \( C(x + \tau) \) under the logistic-Gompertz functional form assumption after the MLE. The differences between Figures 2.1(a) and 2.1(c) arise since the MLE method explicitly takes the drift condition (2.2) into account.

\[
\begin{array}{cccccc}
  k & c & d & a & b \\
  0.0025 & 0.0840 & -10.4692 & 62.9958 & -0.0052 \\
\end{array}
\]

Table 2.1. Calibrated parameters from the MLE method
Figure 2.1. Principal Component Analysis, Estimation, and Projection
2.3 Application I: Economic Capital in Internal Models

As indicated above, one potential reason for the sluggish development of the mortality-linked capital market may be the struggle of insurers with the assessment of their capital relief when hedging part of their mortality risk exposure. In this section, we demonstrate that due to their compatibility with classical actuarial methods and their tractability, forward mortality factor models present a pertinent and simple way for incorporating mortality risk into insurers’ economic capital calculations (see Börger (2010) and Stevens et al. (2010) for alternative approaches).

We first introduce a mathematical framework similar to that from Bauer et al. (2010a) for determining the economic capital in a one-year mark-to-market approach as required by the dawning Solvency II regulation. Subsequently, to describe in detail the merits of forward mortality factor models in this context but also to analyze the influence of systematic mortality risk on economic capital in a reasonably realistic setting, we carry out example calculations for a stylized insurance company offering traditional life insurance products.

2.3.1 Model Framework

Assume the uncertainty with respect to a life insurer’s future profits arises from the uncertain development of a number of financial/economic and demographic factors, which are modeled with the help of the \(d_Y\)-dimensional, sufficiently regular Markov process \(Y = (Y_t)_{t \geq 0} = (Y_t^{(1)}, \ldots, Y_t^{(d_Y)})_{t \geq 0}\), the so-called state process. More specifically, we assume that the prices of all risky assets in the market—including zero coupon bonds—can be expressed in terms of \(Y_t\), and that there exists a locally risk-free asset (bank account) \(B = (B_t)_{t \geq 0}\) with
\[ B_t = \exp\{\int_0^t r_s \, ds\}, \] where \( r_t = r(Y_t) \) is the instantaneous risk-free rate at time \( t \). Similarly, we assume that survival probabilities as denoted in the relevant life table at time \( t \) can be formulated in terms of \( Y_t \), i.e. \( \tau p_x(t, t + \tau) = \tau p_x(Y_t) \). In this market environment, we take for granted the existence of a risk-neutral probability measure \( \mathbb{Q} \) equivalent to \( \mathbb{P} \) under which payment streams can be valued via their expected discounted values with respect to the numéraire \( B \).\(^5\)

Based on this economic/demographic environment, we assume that there exists a cash flow projection model, i.e. there exist functionals \( f_1, \ldots, f_T \) that derive the cash flows at time \( t \) from the state process up to time \( t \), where \( T \) is the time horizon. For instance, within the direct method for determining insurance liabilities (cf. Girard (2002)), the random variable \( X_t = f_t(Y_s, s \in [0, t]) \) corresponds to the benefits paid minus the premiums earned at time \( t \), \( t = 1, \ldots, T \), and the value of the liabilities can be determined via

\[
V_0 = \mathbb{E}^\mathbb{Q} \left[ \sum_{k=0}^T \frac{1}{B_k} X_k \right].
\]

Thus, the Available Capital (AC) at time zero can be derived from \( V_0 \) and the value of assets \( A_0 = A(Y_0) \) as

\[
AC_0 = A_0 - V_0.
\]

However, within the one-year mark-to-market approach for calculating economic capital, it is not sufficient to determine the available capital at time zero, but it is also necessary to

\(^5\)According to the Fundamental Theorem of Asset Pricing, this assumption is essentially equivalent with the absence of arbitrage in the market.
assess AC at time 1, $AC_1 = A_1 - V_1$, where

$$V_1 = X_1 + B_1 \mathbb{E}^Q \left[ \sum_{k=2}^{T} \frac{1}{B_k} X_k \bigg| Y_s, 0 \leq s \leq 1 \right].$$

More specifically, the one-year loss is defined as the $\mathcal{F}_1$-measurable random variable

$$L = AC_0 - p(0, 1) AC_1,$$

where $p(t, \tau)$ denotes the time $t$ price of a zero coupon bond with maturity $t + \tau$. The economical capital is then defined with the help of a monetary risk measure $\rho$: $L^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ as $\rho(L)$ (see e.g. Artzner et al. (1999)). For instance, if the economic capital (EC) is defined based on the Value-at-Risk (VaR) such as the Solvency Capital Requirement (SCR) within the Solvency II framework, we have

$$EC = SCR = \text{VaR}_\alpha(L) = \text{arg min}_x \{ \mathbb{P}(L > x) \leq 1 - \alpha \},$$

where $\alpha$ is a given threshold (99.5% within Solvency II). If economic capital is defined based on the Conditional Tail Expectation (CTE), on the other hand, we obtain

$$EC = CTE_\alpha = \mathbb{E}[L|L \geq \text{VaR}_\alpha(L)].$$
2.3.2 A Stylized Life Insurance Company

Consider now a (stylized) newly founded life insurance company only selling traditional life insurance products to male individuals, who form a representative sample of the England and Wales general male population. More specifically, let us assume that the company’s portfolio of policies consists of \( n_{x,i}^{\text{term}} \) \( i \)-year term-life policies for individuals aged \( x \) with face value \( B_{\text{term}} \), \( n_{x,i}^{\text{end}} \) \( i \)-year endowment policies for individuals aged \( x \) with face value \( B_{\text{end}} \), and \( n_{x,i}^{\text{ann}} \) single premium life annuities with an annual benefit of \( B_{\text{ann}} \) paid in arrears, \( x \in \mathcal{X}, i \in \mathcal{I} \). Furthermore, we assume that the (for term and endowment policies annual) benefit premium is calculated by the Equivalence Principle disregarding profits as well as expenses, and using the concurrent yield curve and the concurrent best-estimate generation table. Here and for the remainder of the text, we implicitly assume that the insurer is risk-neutral with respect to mortality risk, i.e. that the valuation measure \( \mathbb{Q} \) is the product measure of the risk-neutral measure for financial and the physical measure for (independent) biometric events. This is without much loss of generality for the deterministic volatility forward mortality models satisfying (2.3) since, under the assumption of a deterministic market price of systematic mortality risk, a risk-adjusted generation table can be derived from the best estimate generation table via a deterministic transformation (see Bauer et al. (2012) and Bauer et al. (2010b) for details). This is another important advantage of the considered mortality model class in the present context.
Hence, the insurer’s available capital at time zero $AC_0$ amounts to its equity capital $E$.

For the available capital at time 1, on the other hand, we have $AC_1 = A_1 - V_1$, where

\[
A_1 = \left( E + B_{\text{ann}} \sum_{x \in \mathcal{X}} a_x(0) n_x^{\text{ann}} + B_{\text{term}} \sum_{x \in \mathcal{X}, i \in I} \frac{A_{x; i}^{\text{term}}(0)}{\bar{a}_{x; i}(0)} n_{x, i}^{\text{term}} + B_{\text{end}} \sum_{x \in \mathcal{X}, i \in I} \frac{A_{x; i}^{\text{end}}(0)}{\bar{a}_{x; i}(0)} n_{x, i}^{\text{end}} \right) \times R_1,
\]

\[
V_1 = B_{\text{ann}} \sum_{x \in \mathcal{X}} \bar{a}_{x+1}(1) (n_x^{\text{ann}} - \mathcal{D}_x^{\text{ann}}(0, 1)) + B_{\text{term}} \sum_{x \in \mathcal{X}, i \in I} \mathcal{D}_{x, i}^{\text{term}}(0, 1) + B_{\text{end}} \sum_{x \in \mathcal{X}, i \in I} \mathcal{D}_{x, i}^{\text{end}}(0, 1)
\]

\[
\quad + B_{\text{term}} \sum_{x \in \mathcal{X}, i \in I} \left[ A_{x+1; i-1}^{\text{term}}(0) - A_{x; i}^{\text{term}}(0) \bar{a}_{x+1; i-1}(0) \frac{\bar{a}_{x; i}(1)}{\bar{a}_{x; i}(0)} \right] \times (n_{x, i}^{\text{term}} - \mathcal{D}_{x, i}^{\text{term}}(0, 1))
\]

\[
\quad + B_{\text{end}} \sum_{x \in \mathcal{X}, i \in I} \left[ A_{x+1; i-1}^{\text{end}}(0) - A_{x; i}^{\text{end}}(0) \bar{a}_{x+1; i-1}(0) \right] \times (n_{x, i}^{\text{end}} - \mathcal{D}_{x, i}^{\text{end}}(0, 1)).
\]

Here, $(R_1 - 1)$ is the return on the insurer’s asset portfolio, $\mathcal{D}_{x, i}^{\text{con}}(0, 1)$ is the number of deaths in the cohort of $x$-year old insureds with policies of term $i$ and of type $\text{con} \in \{\text{ann}, \text{term}, \text{end}\}$, and $\bar{a}_x(t)$, $A_{x; i}(t)$, etc. are the values of the contracts corresponding to the actuarial symbols calculated at time $t$ based on the yield curve and the generation table at time $t$. For instance,

\[
\bar{a}_x(t) = \sum_{k=0}^{\infty} kp_x(t, t + k) p(t, k).
\]

The economic capital of this insurer can then be determined as

\[
EC = \rho \left( E - (A_1 - V_1) p(0, 1) \right),
\]

where $\rho(\cdot)$ is a monetary risk measure as described above.
2.3.3 Implementation

We assume that our UK insurer only invests in 1, 3, 5, and 10-year government bonds as well as an equity index $S = (S_t)_{t \geq 0}$ (FTSE) at predetermined proportions. For the evolution of these assets, we assume a generalized Black-Scholes model with stochastic interest rates (Vasicek model), that is, under $\mathbb{P}$

$$
\begin{align*}
    dS_t &= S_t (\mu \, dt + \rho \sigma_A \, dB_t^{(1)} + \sqrt{1 - \rho^2} \sigma_A \, dB_t^{(2)}), \quad S_0 > 0, \\
    dr_t &= \kappa (\gamma - r_t) \, dt + \sigma_r \, dB_t^{(1)}, \quad r_0 > 0,
\end{align*}
$$

where $\mu, \sigma_A, \kappa, \gamma, \sigma_r > 0$, $\rho \in [-1, 1]$, and $\left( B_t^{(1)} \right)$ and $\left( B_t^{(2)} \right)$ are independent Brownian motions under $\mathbb{P}$ that are independent of $(W_t)$. Moreover, we assume that the market price of interest rate risk is constant and denote it by $\lambda$, i.e. we replace $\mu$ by $r_t$ and $\gamma$ by $\gamma - \frac{\lambda \sigma_r}{\kappa}$ for the dynamics under the risk-neutral measure $\mathbb{Q}$ (we assume $\lambda < \frac{\kappa}{\sigma_r} \gamma$ to ensure $r_t$ is mean-reverting under $\mathbb{Q}$).

We estimate the parameters based on UK data from June 1988 to June 2008 using a Kalman filter. More precisely, we use monthly data for the FTSE 100 total return index,\(^6\) treasury bills (3 months),\(^7\) and government bonds with maturities of 1 year, 3 years, 5 years, and 10 years.\(^8\) The resulting parameter estimates are displayed in Table 2.2.\(^9\)

---

\(^6\)Downloaded on 05/11/2011 from Bloomberg (code: TUKXG).
\(^7\)Downloaded on 05/11/2011 from the Bank of England’s website, [http://www.bankofengland.co.uk/mfsd/iadb/](http://www.bankofengland.co.uk/mfsd/iadb/).
\(^9\)Some of the parameters from our estimation procedure—particularly $\rho$ and $\lambda$—slightly deviate from values used in other studies, which may be due to idiosyncrasies of the considered time period. To ensure that our results are not specific to the considered parameters, we conducted detailed sensitivity analyses.
Table 2.2. Estimated capital market parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.1005</td>
</tr>
<tr>
<td>$\sigma_A$</td>
<td>0.1429</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.2502</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.0998</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0509</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>0.0090</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.1441</td>
</tr>
<tr>
<td>$r_0$</td>
<td>0.0473</td>
</tr>
</tbody>
</table>

Hence, based on realizations of the asset process and the instantaneous risk-free rate at time 1, $R_1$ can be determined as

$$R_1 = \frac{\omega_1 S_1}{S_0} + \omega_2 \frac{1}{p(0,1)} + \omega_3 \frac{p(1,2)}{p(0,3)} + \omega_4 \frac{p(1,4)}{p(0,5)} + \omega_5 \frac{p(1,9)}{p(0,10)},$$

in which $\omega_i$, $i = 1, \ldots, 5$, are predetermined proportions invested in each asset (for our numerical analyses, we impose equal proportions, i.e. $\omega_i = 20\%$, $i = 1, \ldots, 5$).

To generate realizations, $r_1$ and $S_1$ are simulated from a joint Normal distribution. More specifically, we have:

$$r_1 = e^{-\kappa}r_0 + \gamma(1-e^{-\kappa}) + \int_0^1 \sigma_r e^{-\kappa(1-s)} dB_s^{(1)},$$
$$S_1 = S_0 \times \exp \left\{ \mu - \frac{\sigma_A^2}{2} + \rho \sigma_A B_1^{(1)} + \sqrt{1-\rho^2} \sigma_A B_1^{(2)} \right\},$$

which can be conveniently used in Monte Carlo algorithms (see e.g. Zaglauer and Bauer (2008)). Bond prices can then be calculated in a straight-forward manner due to their

for all of our applications. Since the analyses do not reveal new insights, in order to keep the presentation concise, these results are not included in the paper.
exponential-affine structure, i.e. we have \( p(t, \tau) = \exp(-A(\tau)r_t + C(\tau)) \), where

\[
A(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa \tau}), \text{ and } C(\tau) = (\gamma - \frac{\lambda \sigma_x}{\kappa} - \frac{\sigma_r^2}{2\kappa^2})(\frac{1 - e^{-\kappa \tau}}{\kappa} - \tau) - \frac{\sigma_r^2(1 - e^{-\kappa \tau})^2}{4\kappa^3}.
\]

With respect to mortality risk, two different approaches are considered. In the first approach, we assume that mortality rates evolve deterministically, i.e. we use the last generation life table from our estimation procedure outlined in Section 2.2 and ignore systematic mortality risk.\(^{10}\) The realized deaths \(( \mathcal{D}^{\text{con}}_{x,i}, x \in \mathcal{X}, i \in \mathcal{I}, \text{con} \in \{\text{ann}, \text{term}, \text{end}\}) \) are then simulated from a Binomial distribution using the corresponding one-year mortality rates denoted in the life table.

In the second approach, systematic mortality risk is considered by relying on the model introduced in Section 2.2. More specifically, from Equation (2.4) we obtain

\[
\tau p_x(1, 1+\tau) = \frac{\tau_1 p_x - 1(0, 1 + \tau)}{1 p_x - 1(0, 1)} \times \exp \left\{- \int_0^1 \int_0^\tau \alpha(v + 1 - s, x - 1 + s) \, dv \, ds - \int_0^\tau C(x + v) e^{Mv} \, dv \times Z_1 \right\},
\]

in which with Equation (2.2)

\[
\int_0^1 \int_0^\tau \alpha(v + 1 - s, x - 1 + s) \, dv \, ds = \int_0^1 \left\{ \frac{1}{2} \int_{1-s}^{\tau+1-s} \sigma(u, x - 1 + s) \, du \int_{1-s}^{\tau+1-s} \sigma'(u, x - 1 + s) \, du \\
+ \int_{1-s}^{\tau+1-s} \sigma(u, x - 1 + s) \, du \int_0^{1-s} \sigma'(u, x - 1 + s) \, du \right\} \, ds.
\]

\(^{10}\)For simplicity, we use a maximal age of \( \bar{\omega} = 95 \) throughout this paper since we only rely on mortality data up to the age of 95 in the estimation procedure. This is without much loss of generality since probabilities for ages beyond 95 could be e.g. easily determined by extrapolation (see Börger (2010)).
From Equation (2.5), on the other hand,

$$e^{Mv} = \begin{pmatrix}
(1 - b_1 v) e^{-b_1 v} & -b_1^2 v e^{-b_1 v} \\
v e^{-b_1 v} & (1 + b_1 v) e^{-b_1 v}
\end{pmatrix},$$

so that

$$C(x + v) e^{Mv} = \begin{pmatrix}
\frac{k_1 e^{c_1(x+v)+d_1}}{1+e^{c_1(x+v)+d_1}} v e^{-b_1 v} & \frac{k_1 e^{c_1(x+v)+d_1}}{1+e^{c_1(x+v)+d_1}} (1 + b_1 v) e^{-b_1 v}
\end{pmatrix}.$$

Therefore, for each $Z_1$ simulated according to (2.4), Equation (2.6) immediately yields a generation life table $\{\tau p_x(1; 1 + \tau), x, \tau \geq 0\}$ at time 1.

However, it is important to notice that solely relying on the simulated life tables at time one and otherwise proceeding as in the deterministic mortality approach will lead to a slight bias. More precisely, to obtain estimates that—on average—correspond to the deterministic mortality case, we also need to consider the stochastic evolution of mortality within the simulation of the numbers of deaths $D_{x,i}^{\text{con}}, x \in \mathcal{X}, i \in \mathcal{I}, \text{con} \in \{\text{ann, term, end}\}$, which for each cohort now follow Binomial distributions with a random mortality probability $1q_x(1, 1) = 1 - 1p_x(1, 1)$, where

$$1p_x(1, 1) = \exp \left\{ - \int_0^1 \mu_s(0, x + s) ds \right\} = \exp \left\{ - \int_0^1 \left( \mu_0(s, x) + \int_0^s \alpha(s - u, x + u) du + C(x + s) Z_s \right) ds \right\} = 1p_x(0, 1) \times \exp \left\{ - \int_0^1 \int_0^s \alpha(s - u, x + u) du ds - \int_0^1 C(x + s) Z_s ds \right\} (2.7)$$
Thus, it is not sufficient to sample $Z_1$ only as for the simulation of $\tau p_x(1, 1 + \tau)$ according to Equation (2.6), but it is necessary to simulate the first year path $(Z_s)_{0 \leq s \leq 1}$. Here, we can rely on an exact simulation of the increments since for $t > u$, we have

$$Z_t = \int_0^t e^{M(t-s)} N\,dW_s$$

$$= e^{Mt} \times \left( e^{-M u} e^{M u} \int_0^u e^{-M s} N\,dW_s + \int_u^t e^{-M s} N\,dW_s \right)$$

$$= e^{M(t-u)} Z_u + \int_0^{t-u} e^{M(t-u-v)} N\,dW_v,$$

where $Y$ is Normal-distributed with mean 0 and $\text{Cov}(Y) = \int_0^{t-u} e^{M(t-u-v)} N N' e^{M(t-u-v)'}dv$. Then, (2.7) can be immediately simulated by approximating the integral via its left sum, i.e. for large $n^{11}$

$$\int_0^1 C(x + s) Z_s\,ds \approx \frac{1}{n} \sum_{i=0}^{n-1} C(x + \frac{i}{n}) Z_{i/n}.$$

The above illustrates the distinct advantages of forward mortality factor models in this context: On the one hand, since the actuarial present values at time 1 are given by the simulated generation life tables, it is possible to avoid a nested simulation structure as it would arise when relying on popular mortality models such as the Lee-Carter model. On the other hand, in contrast to the reliance on other forward mortality models when calculating economic capital, where the entire mortality surface needs to be simulated (see e.g. Börger (2010)), here it is sufficient to simulate the finite-dimensional diffusion process $(Z_s)_{0 \leq s \leq 1}$, which can be carried out in a straight-forward manner. Therefore, these models present a

---

11 We choose $n = 100$ within our calculations.
coherent and feasible method for incorporating stochastic mortality into internal economic capital models.

2.3.4 Results

We consider a stylized insurer with portfolio parameters as displayed in Table 2.3, and we use 25,000 simulations of the loss at time one generated according to the algorithm outlined in the previous subsections to calculate the monetary risk measures of interest. Furthermore, in order to obtain estimates of simulation errors, we repeat the above procedure 100 times. The empirical cumulative distribution functions of the portfolio loss $L$ for both approaches, i.e. without and with the consideration of systematic mortality risk, are displayed in Figure 2.2. Table 2.4 shows the results of economic capital calculations based on the Value-at-Risk (VaR) and the Conditional Tail Expectation (CTE).

By comparing the results, we observe that the economic capital increases considerably when including systematic mortality risk: For the 90% (99%) threshold, the Value-at-Risk increases by around 12% (13%) and the Conditional Tail Expectation increases by around 13% (14%). While the absolute levels have to be interpreted with care since we neither take expenses nor profits into account and since we do not consider a mortality risk premium associated with systematic mortality risk, our results demonstrate that mortality risk is an important risk factor and should be incorporated in risk-based capital calculations.

To investigate the effect of systematic mortality risk on the different types of insurance products, we calculate the 99% Value-at-Risk with respect to the three products separately
<table>
<thead>
<tr>
<th>$x$</th>
<th>$i$</th>
<th>$n_{x,i}^{\text{term/end/ann}}$</th>
<th>$B_{\text{term/end/ann}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>20</td>
<td>250</td>
<td>100,000</td>
</tr>
<tr>
<td>35</td>
<td>15</td>
<td>250</td>
<td>100,000</td>
</tr>
<tr>
<td>40</td>
<td>10</td>
<td>250</td>
<td>100,000</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>250</td>
<td>100,000</td>
</tr>
</tbody>
</table>

**Term Life**

**Endowment**

<table>
<thead>
<tr>
<th>$x$</th>
<th>$i$</th>
<th>$n_{x,i}^{\text{term/end/ann}}$</th>
<th>$B_{\text{term/end/ann}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>20</td>
<td>500</td>
<td>50,000</td>
</tr>
<tr>
<td>45</td>
<td>15</td>
<td>500</td>
<td>50,000</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>500</td>
<td>50,000</td>
</tr>
</tbody>
</table>

**Annuities**

<table>
<thead>
<tr>
<th>$x$</th>
<th>$i$</th>
<th>$n_{x,i}^{\text{term/end/ann}}$</th>
<th>$B_{\text{term/end/ann}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>(35)</td>
<td>250</td>
<td>18,000</td>
</tr>
<tr>
<td>70</td>
<td>(25)</td>
<td>250</td>
<td>18,000</td>
</tr>
</tbody>
</table>

Table 2.3. Portfolio for the company, $E = 2,000,000$

Figure 2.2. Empirical CDF of Economic Capital
Table 2.4. Economic capital for the stylized company based on different risk measures

<table>
<thead>
<tr>
<th>Confidence level α</th>
<th>VaR</th>
<th>standard error</th>
<th>CTE</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deterministic mortality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>3,995,515</td>
<td>39,596</td>
<td>5,712,494</td>
<td>41,787</td>
</tr>
<tr>
<td>99%</td>
<td>7,804,053</td>
<td>93,629</td>
<td>9,047,340</td>
<td>106,028</td>
</tr>
<tr>
<td><strong>Stochastic mortality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>4,494,682</td>
<td>41,753</td>
<td>6,457,341</td>
<td>45,018</td>
</tr>
<tr>
<td>99%</td>
<td>8,844,264</td>
<td>98,542</td>
<td>10,306,897</td>
<td>118,670</td>
</tr>
</tbody>
</table>

under a stand alone perspective and the capital allocated to each product line based on the covariance capital allocation technique for VaR described in Kalkbrener (2005). The results are displayed in Table 2.5. While it is not surprising that the required capital of the total portfolio is smaller than the sum of the Values-at-Risk of the individual product lines, we find that the different types of products are affected dissimilarly under systematic mortality risk. More precisely, the capital allocated to the term business within the portfolio is negative, implying a capital relief by exploiting natural hedging opportunities between the different lines (see also Cox and Lin (2007)). Nevertheless, the company’s exposure to longevity risk still is considerable, and the capital relief by participating in the market for mortality-linked securities may be substantial.

\^12We assume the equity capital E is distributed among different products according to their expected total premiums.
<table>
<thead>
<tr>
<th>Evaluation technique</th>
<th>stand alone</th>
<th>capital allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deterministic mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Life</td>
<td>338,711</td>
<td>27,343</td>
</tr>
<tr>
<td>Endowment</td>
<td>2,624,822</td>
<td>1,697,119</td>
</tr>
<tr>
<td>Annuities</td>
<td>5,961,906</td>
<td>6,079,590</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8,925,439</td>
<td>7,804,053</td>
</tr>
<tr>
<td><strong>Stochastic mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Life</td>
<td>473,474</td>
<td>−120,123</td>
</tr>
<tr>
<td>Endowment</td>
<td>2,652,375</td>
<td>1,458,029</td>
</tr>
<tr>
<td>Annuities</td>
<td>7,670,600</td>
<td>7,506,357</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10,796,450</td>
<td>8,844,264</td>
</tr>
</tbody>
</table>

Table 2.5. Calculations of VaR_{99%}
2.4 Application II: Valuation of Annuitzation Options

As demonstrated in the previous section, traditional life insurance and pension products such as whole life, term life, or endowment insurances as well as life annuities may be evaluated directly using the “concurrent” (time $t$) mortality surface $\mu_t(\tau, x)$. For more complex life insurance and annuity contracts containing mortality-contingent embedded options, however, the stochasticity has to be taken into account explicitly. Common types of longevity-contingent options are so-called Guaranteed Annuity Options (GAOs) within traditional or participating life insurance contracts and Guaranteed Minimum Income Benefits (GMIBs) within Variable Annuities. A GAO provides the policyholder with the option to choose, at retirement, between a lump-sum payment or a life-long immediate annuity, which is calculated based on a guaranteed annuity rate. In contrast, a GMIB gives the insured the possibility to annuitize a guaranteed amount at a pre-specified rate. In this section, we provide a closed-form pricing formula for simple GAOs and an efficient numerical method for pricing GMIBs in our forward mortality modeling framework.

2.4.1 Valuation of Guaranteed Annuity Options

Several authors have studied the problem of evaluating GAOs without taking systematic mortality risk into account (see Boyle and Hardy (2003), Pelsser (2003), and references therein). In contrast, Milevsky and Promislow (2001) provide a discrete- and continuous-time pricing framework for simple annuitization options that takes the stochasticity of mortality rates into account. Similarly, Ballotta and Haberman (2006) present a pricing approach for GAOs, which accounts for both interest rate and mortality risk. Since the solution for the
price is not in closed form, they rely on Monte Carlo simulations for the derivation of their numerical results. Using affine processes for modeling the financial market as well as the (spot) mortality evolution, Biffis and Millossovich (2006) also present a pricing framework for GAOs. Under some structural assumptions, they derive analytical solutions up to the computation of Fourier transforms and/or numerical integrals for various contract designs.

Following ideas by Cairns et al. (2006a) in their annuity market model, we consider the valuation of simple GAOs in our forward mortality modeling framework. Specifically, we focus on contracts providing payoffs of the following form at time $T$ if the policyholder is alive:

$$V^\text{GAO}_T = \max \left\{ 1, g^\text{GAO}_{x_0} \ddot{a}_{x_0} + T \right\}.$$

Here, $g^\text{GAO}_{x_0}$ is the guaranteed annuity rate under the GAO for annuitization at time $T$ contracted at time 0 for an $x_0$-year-old. Thus, the contract may be interpreted as a $T$-year pure endowment policy with the additional option to annuitize at the fixed rate $g^\text{GAO}_{x_0}$ at maturity.\(^{13}\)

We have

$$V^\text{GAO}_T = 1 + \ddot{a}_{x_0} + T (T) \max \left\{ g^\text{GAO}_{x_0} - \frac{1}{\dot{a}_{x_0} + T (T)}, 0 \right\} = C^\text{GAO}_T$$

\(^{13}\)Of course, this pure endowment part may easily be combined with a term insurance contract to obtain an endowment contract including a GAO, which are common contracts in many markets.
and thus,\(^{14}\)

\[
V_0^{GAO} = \mathbb{E}^Q \left[ 1_{\{x_0 > T\}} e^{-\int_0^T r_s \, ds} \, V_T^{\text{GAO}} \right] = p(0, T) \, TP_{x_0} (0, T) + \mathbb{E}^Q \left[ e^{-\int_0^T \mu_s (0, x_0 + s) \, ds} e^{-\int_0^T r_s \, ds} \, C_T^{\text{GAO}} \right] = C_0^{\text{GAO}}.
\]

Define \((X_t)_{0 \leq t \leq T}\) via

\[
X_t = \sum_{k=T}^{\infty} p(t, k - t) \left( e^{-\int_0^k \mu_s (0, x_0 + s) \, ds} \bigg| \mathcal{F}_t \right).
\]

Then, for the value of the GAO,

\[
C_0^{\text{GAO}} = \mathbb{E}^Q \left[ \frac{e^{-\int_0^T \mu_s (0, x_0 + s) \, ds}}{B_T} \sum_{k=T}^{\infty} p(T, k - T) k - TP_{x_0 + T}(T, k) \times \left( \frac{g_{x_0, T}^{\text{GAO}} - \frac{e^{-\int_0^T \mu_s (0, x_0 + s) \, ds}}{X_T} \, \tilde{a}_{x_0 + T}(T)}{X_T} \right)^+ \right]
\]

\[
= \mathbb{E}^Q \left[ \frac{X_T}{B_T} \left( g_{x_0, T}^{\text{GAO}} - \frac{e^{-\int_0^T \mu_s (0, x_0 + s) \, ds}}{X_T} \right)^+ \right] = X_0 \mathbb{E}^{Q_X} \left( g_{x_0, T}^{\text{GAO}} - \frac{TP_{x_0} (T, T)}{X_T} \right)^+ , \quad (2.8)
\]

where \(Q_X\) is the equivalent martingale measure associated with the numéraire process \((X_t)\) (see e.g. Björk (1999) for details on the change of numéraire technique). In particular, the price process of a security with payoff \(TP_{x_0} (T, T)\) at time \(T\) discounted by \((X_t)\) will be a martingale under \(Q_X\). Hence, in order to evaluate the expectation in (2.8), we solely need to assess the volatility term of \(\frac{p(t, T-t) \, TP_{x_0} (t, T)}{X_t}\) and, thus, the volatility term of \((X_t)\). From Proposition 2.1 in Bauer et al. (2011) and Proposition 20.5 in Björk (1999), we obtain that

\(^{14}\)Again, akin to the previous section, for simplicity and without much loss of generality, we assume that the insurer is risk-neutral with respect to mortality risk.
under \( \mathbb{Q}^{15} \)

\[ dX_t = \sum_{k=T}^{\infty} p(t, k - t) k p_{x_0}(t, k) (S(t, k; x_0), v(t, k - t)) d \begin{pmatrix} W_t \\ B^{(1)}_t \end{pmatrix}, \]

where \( v(t, k) \) is the time \( t \) volatility of a zero coupon bond maturing at time \( t + k \) and

\[ S(t, k; x_0) = -\int_0^{k-t} \sigma(s, x_0 + t) ds. \]

Therefore, the volatility of \((X_t)\) is given by

\[
\sum_{k=T}^{\infty} p(t, k - t) k p_{x_0}(t, k) (S(t, k; x_0), v(t, k - t)) = X_t \sum_{k=T}^{\infty} \frac{p(t, k - t) k p_{x_0}(t, k)}{\sum_{l=T}^{\infty} p(t, l - t) l p_{x_0}(t, l)} (S(t, k; x_0), v(t, k - t)).
\]

By an application of Itô’s Lemma, the volatility of \( \frac{p(t, T-t) T p_{x_0}(t, T)}{X_t} \) is then given by

\[
\left( \frac{p(t, T-t) T p_{x_0}(t, T)}{X_t} \right) \left( (S(t, T; x_0), v(t, T-t)) - \sum_{k=T}^{\infty} w_t(k) (S(t, k; x_0), v(t, k - t)) \right)
\]

\[
= \left( \frac{p(t, T-t) T p_{x_0}(t, T)}{X_t} \right) \left( \sum_{k=T}^{\infty} \gamma_t(k) ((S(t, T; x_0), v(t, T-t)) - (S(t, k; x_0), v(t, k - t))) \right).
\]

If now \( \gamma(\cdot, \cdot, \cdot) \) were deterministic, we would be able to derive an analytical expression for (2.8) via a Black-type formula (this is basically the idea of Cairns et al. (2006a), who propose

\[ \text{If not noted otherwise, we adopt the capital market model introduced in Section 2.3.} \]
to directly model the forward annuity rate \( \left( \frac{p(t, T-t) \tau_p(t, T)}{X_t} \right) \). However, in our framework, the weights \( w_t(k) \) are in fact stochastic. Nevertheless, such an approach may be understood as an approximation, and Pelsser (2003) points out that for the deterministic mortality case, one may infer \( \gamma(t, T, x_0) \) “by “freezing” the stochastic weights at their current values”. This is similar to a common approximation in the popular LIBOR market models (see Theorem 3.2 in Brace et al. (1997)). In particular, since \( v(\cdot, \cdot) \) is deterministic, in our problem we may fix

\[
\gamma(t, T, x_0) \approx \sum_{k=T}^{\infty} w_0(k) \left( (S(t, T, x_0), v(t, T-t)) - (S(t, k, x_0), v(t, k-t)) \right),
\]

which then yields:

\[
\begin{align*}
C_{0}^{GAO} & \approx X_0 \left( g_{x_0, T}^{GAO} \Phi \left( -d_2^{GAO} \right) - \frac{p(0, T) \tau_p x_0(0, T)}{X_0} \Phi \left( -d_1^{GAO} \right) \right),
\end{align*}
\]

(2.9)

where

\[
\begin{align*}
d_1^{GAO} &= \log \left\{ \frac{p(0, T) \tau_p x_0(0, T)}{X_0 g_{x_0, T}^{GAO}} \right\} + \frac{1}{2} \sigma^2_{GAO}, \\
d_2^{GAO} &= d_1^{GAO} - \sigma_{GAO}, \\
\sigma^2_{GAO} &= \int_0^T \| \gamma(u, T, x_0) \|^2 du,
\end{align*}
\]

and \( \Phi(\cdot) \) denotes the cumulative distribution function of the standard Normal distribution.

It is important to note that we restricted ourselves to constant payoffs. In particular, this means that the GAO considered here provides a constant annuity guarantee, and thus, in this special case, it coincides with a GMIB within a “non-variable” VA-contract as introduced above. In practice, GAOs are also often attached to unit-linked or participating policies, and GMIBs are usually granted within “truly variable” VA contracts; in these cases, the lump-
sum payment will be stochastic, and hence, formula (2.9) does not apply.\footnote{Ballotta and Haberman (2003) provide pricing formulas for GAOs within unit-linked policies for a deterministic mortality evolution by applying the ideas of Jamshidian (1989). However, their results cannot be easily carried over to the stochastic mortality environment unless interest rates and the mortality evolution are driven by the same one-dimensional Brownian motion $W$, which seems very unrealistic.} Thus, one may have to resort to numerical methods for the valuation.

### 2.4.2 Valuation of Guaranteed Minimum Income Benefits

For the time zero value of a Variable Annuity contract including a GMIB with guaranteed annuity payment $g_{x_0,T}^{\text{GMIB}}$ per invested unit for an $x_0$-year old insured when annuitizing at time $T$ and no death benefit guarantee, i.e. in the case of death only the current account value is paid out at the end of the year of death, we have:

\[
V_0^{\text{GMIB}} = \mathbb{E}^Q \left[ 1\{\Upsilon_{x_0} > T\} e^{-\int_0^T r_s \, ds} \max \left\{ g_{x_0,T}^{\text{GMIB}} A_0 \bar{a}_{x_0+T}(T), A_T \right\} \right] \\
+ \sum_{k=0}^{T-1} \mathbb{E}^Q \left[ 1\{\Upsilon_{x_0} \in [k,k+1)\} e^{-\int_k^{k+1} r_s \, ds} A_{k+1} \right] \\
= \mathbb{E}^Q \left[ e^{-\int_0^T r_s + \mu_s(0,x_0+s) \, ds} \max \left\{ g_{x_0,T}^{\text{GMIB}} A_0 \sum_{k=T}^{\infty} p(T, k - T) k - T p_{x_0+T}(T, k), A_T \right\} \right] \\
+ A_0 \sum_{k=0}^{T-1} (k p_{x_0}(0, k) - (k+1) p_{x_0}(0, k+1)) e^{-\phi (k+1)}. \tag{2.10}
\]

Here $(A_t)$ denotes the insured’s account value at time $t$, which—in our generalized Black-Scholes framework—is assumed to evolve according to the stochastic differential equation (cf. Bauer et al. (2008b))

\[
dA_t = A_t \left( (r_t - \phi) dt + \rho \sigma_A \, dB_t^{(1)} + \sqrt{1 - \rho^2} \sigma_A \, dB_t^{(2)} \right), \ A_0 > 0,
\]
where \( \tilde{B}_t^{(1)} \) and \( \tilde{B}_t^{(2)} \) are independent Brownian motions under \( \mathbb{Q} \) independent of \( (W_t) \), and \( \phi \) denotes the continuously deducted option fee. Now, methods similar to those proposed in Bauer et al. (2008b), where a deterministic evolution of mortality is assumed, can be employed to determine the value. Similarly as in the first application, Equation (2.10) reconfirms the advantage of forward mortality models and illustrates why Cairns (2007) considers them to be “ideal for pricing contracts with embedded options”: Within other mortality models, \( \ddot{a}_{x_0+T}(T) \) is typically not available, but each “simulation” will require a “further bundle of simulations from time \( T \) to evaluate forward survival probabilities [i.e. a nested simulation structure]. In contrast, forward survival probabilities are a standard part of the output [of forward force models] at time \( T \).”

2.4.3 Implementation and Results

Guaranteed Annuity Options For guaranteed annuity options, we consider the guaranteed annuity rate

\[
g_{x_0,T}^{GAO} = \frac{p(0,T)T p_{x_0}(0,T)}{T \ddot{a}_{x_0}(0)}.
\]  

in which \( T \ddot{a}_{x_0}(0) = \ddot{a}_{x_0}(0) - \ddot{a}_{x_0:LT}(0) \) is the present value of a \( T \)-year deferred annuity at time zero. Note that this is exactly the rate policyholders would obtain if they committed to annuitizing, i.e. our GAO is “at the money”. Therefore,

\[
\frac{p(0,T)T p_{x_0}(0,T)}{X_0 g_{x_0,T}^{GAO}} = 1
\]
and
\[ d_1^{GAO} = -\sigma_2^{GAO} = \frac{1}{2}\sigma_{GAO}^2, \]
so that
\[ C_0^{GAO} \approx X_0 \left( g_{x_0,T}^{GAO} \Phi \left( \frac{1}{2}\sigma_{GAO}^2 \right) - g_{x_0,T}^{GAO} \Phi \left( -\frac{1}{2}\sigma_{GAO}^2 \right) \right) \]
\[ = p(0,T) \tau p_{x_0}(0,T) \left( 2\Phi \left( \frac{1}{2}\sigma_{GAO}^2 \right) - 1 \right). \] (2.12)

Furthermore, notice that (cf. Pelsser (2003))
\[ v(t, \tau) = -\sigma_r \int_0^\tau e^{-\kappa s} ds. \]

Hence, from Equation (2.11) and (2.12), we can calculate guaranteed annuity rates and prices of guaranteed annuity options for three representative age/term combinations (note that for a deterministic evolution of mortality, our approach coincides with that from Pelsser (2003)). The results for the case without and with considering systematic mortality risk are displayed in Table 2.6.

We find that systematic mortality risk has a significant impact on the valuation of simple Guaranteed Annuity Options. We also observe different patterns of \( C_0^{GAO} \) with respect to maturity \( T \) in the two cases: While \( C_0^{GAO} \) is decreasing in \( T \) in either situation, the marginal effect of \( T \) on \( C_0^{GAO} \) is relatively higher when there is no mortality risk. This appears to be a consequence of the exponential decrement of \( \nu(t, \tau) \) with respect to \( \tau \) for the volatility of the zero coupon bond against an increasing pattern of \( S(t, t + \tau, x_0) \). More precisely, for an
Increasing maturity, the first two terms of (2.12) decrease while the last term is increasing in volatility and, thus, in $T$. It now seems that under a deterministic evolution of mortality, the first effect is more pronounced while under stochastic mortality, the increasing pattern of $S(t, t + \tau, x_0)$ implies a stronger impact of the latter effect.\footnote{It is necessary to note that of course our results are affected by our choice of the interest rate model. Specifically, the volatilities of long-term interest rates implied by a one-factor Vasicek model may be too small to accurately capture the volatility of an annuity. We accept this shortcoming since our focus is on mortality risk but note that our quantitative results have to be interpreted with care.}

Thus, aside from considerable quantitative effects, stochastic mortality also affects the qualitative patterns of GAO values.

**Guaranteed Minimum Income Benefits** For the GMIB, similarly to the GAO, the guaranteed annuity payment rate is set as

$$g^\text{GMIB}_{x_0,T} = \frac{\tau p_{x_0}(0,T)}{T|a_{x_0}(0)|^\gamma},$$

<table>
<thead>
<tr>
<th>$(x_0, T)$</th>
<th>$g^\text{GAO}_{x_0,T}$</th>
<th>$V^\text{GAO}_0$</th>
<th>$C^\text{GAO}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(30, 35)</td>
<td>0.0803</td>
<td>0.1315</td>
<td>0.0050</td>
</tr>
<tr>
<td>(40, 25)</td>
<td>0.0825</td>
<td>0.2366</td>
<td>0.0088</td>
</tr>
<tr>
<td>(50, 15)</td>
<td>0.0850</td>
<td>0.4242</td>
<td>0.0152</td>
</tr>
</tbody>
</table>

Table 2.6. Valuation of GAOs
i.e. the GMIB is also “at the money”. Since a closed-form solution for the value $V_{0}^{\text{GMIB}}$ is not available in this case, we have to rely on simulations under the risk-neutral measure $Q$.

With respect to financial/economic risk, in contrast to Section 2.3.3 where only simulations of $S_1$ and $r_1$ are necessary, here we additionally need to simulate the discount factor $\exp\left\{-\int_{0}^{T} r_s ds\right\}$. This can be achieved by sampling $r_T$, $\log\{A_T\}$, and $\int_{0}^{T} r_s ds$ simultaneously from their joint Normal distribution under $Q$. We refer to Zaglauer and Bauer (2008) for details.

With respect to future mortality risk, we need simulations of $e^{-\int_{0}^{T} \mu_s(0,x_0+s) ds}$ and $k-Tp_{x_0+T}(T,k)$, $k = T, \ldots, \infty$, which analogously to Section 2.3.3 can be expressed as

$$T_{p_{x_0}}(T,T) = e^{-\int_{0}^{T} \mu_s(0,x_0+s) ds}$$

$$= T_{p_{x_0}}(0,T) \times \exp\left\{-\int_{0}^{T} \int_{0}^{s} \alpha(s-u,x_0+u) du \, ds - \int_{0}^{T} C(x_0+s) Z_s \, ds\right\},$$

$$k-Tp_{x_0+T}(T,k)$$

$$= \frac{k_{p_{x_0}}(0,k)}{T_{p_{x_0}}(0,T)} \times \exp\left\{-\int_{0}^{T} \int_{0}^{k-T} \alpha(v + T - s, x_0 + s) \, dv \, ds - \int_{0}^{k-T} C(x_0 + T + v)e^{Mv} \, dv \, Z_T\right\}.$$

Just as in Section 3.3 above, the latter equations can be conveniently used in Monte Carlo simulations.

With the above simulation procedures, we can calculate the contract value $V_{0}^{\text{GMIB}}(x_0,T)$ for given $\phi$. In particular, we are interested in the fair fee, i.e. the continuously deducted option fee $\phi^*$ such that $V_{0}^{\text{GMIB}} = A_0$. Here we use the same age/term combinations as in Section 4.3.1. The results are shown in Table 2.7.
We find that $\phi^*$ is greater when taking mortality risk into account, which appears intuitive since additional volatility naturally implies a relatively higher option price. However, we observe that the increase of $\phi^*$ is relatively modest in comparison to the results for the guaranteed annuity options. Thus, it seems that while the effect of mortality risk is considerable relative to interest rate risk as shown in our analysis of GAOs, the influence is moderate relative to equity risk—the main risk driver for the VA/GMIB. In particular, our analyses show that mortality risk affects different mortality-contingent options dissimilarly.

2.5 Conclusion

Due to their tractability and their compatibility with classical actuarial theory, *Forward Mortality Factor Models* present a convenient way of introducing systematic mortality risk to actuarial practice. This not only improves the accuracy of common actuarial calculations, but also helps to provide a more coherent “risk picture” of a life insurance company’s operations.

The current paper documents the advantages of this model class by discussing in detail important example applications. In the first application, we derive the economic capital for a stylized life insurance company based on different monetary risk measures. Our results

<table>
<thead>
<tr>
<th>$(x_0, T)$</th>
<th>Deterministic mortality</th>
<th>Stochastic mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(30, 35)$</td>
<td>0.0278</td>
<td>0.0319</td>
</tr>
<tr>
<td>$(40, 25)$</td>
<td>0.0321</td>
<td>0.0349</td>
</tr>
<tr>
<td>$(50, 15)$</td>
<td>0.0423</td>
<td>0.0445</td>
</tr>
</tbody>
</table>

Table 2.7. Fair option fee $\phi^*$
suggest that mortality risk has a considerable effect on the solvency of the company and thus should not be disregarded when assessing its financial stability. In the second application, we discuss the valuation of mortality-contingent options, specifically Guaranteed Annuity Options and Guaranteed Minimum Income Benefits within Variable Annuity contracts. Our analyses indicate that mortality risk affects the two options differently: While for simple Guaranteed Annuity Options, the values increase greatly with the inclusion of mortality risk, for the fair option fees for Variable Annuity contracts with Guaranteed Minimum Income Benefits market risks dominate. However, it is important to note that the latter risk—at least to a large extent—is hedgeable whereas markets for mortality risk are only slowly maturing. Thus, while the influence of mortality risk on pricing may be relatively moderate in this case, the same may not necessarily be true for the associated profit and loss distribution. Moreover, here we assume a “financially optimal” exercise behavior, which may not be true in practice; in particular, it is conceivable that mortality risk has an impact on the exercise behavior.

2.6 Appendix: Principle Component Analysis

In this appendix, we briefly outline the approach introduced in Zhu and Bauer (2011b) for the estimation of the forward volatility function using principle component analysis.

Assume we are given a time series of generational mortality data \( (\tau \rho_x(t_j, t_{j+1} + \tau))_{(\tau, x) \in \mathcal{C}}, j = 1, \ldots, N, t_{j+1} - t_j = \Delta, \) where \( \mathcal{C} \) denotes a (large) collection of term/age combinations,
For each \((\tau, x) \in C, j \in \{1, 2, \ldots, N - 1\}\), define

\[
F(j, \tau, x) = -\log \left\{ \frac{\tau + 1 + \Delta p_x(t_{j+1}, t_{j+1} + \tau + 1)}{\tau p_x(t_{j+1}, t_{j+1} + \tau)} \right\}
\]

Intuitively, \(F(j, \tau, x)\) measures the shift (log change) of the one-year marginal survival probability for the period \([t_{j+1} + \tau, t_{j+1} + \tau + 1]\) from projection at time \(t_{j+1}\) relative to time \(t_j\), for an—at time \(t_{j+1}\)—\(x\)-year old. With some basic manipulations, we obtain

\[
F(j, \tau, x) \overset{d}{=} \int_0^\Delta \int_{\tau}^{\tau+1} \alpha(v + \Delta - s, x - \Delta + s) \, dv \, ds
+ \int_0^\Delta \int_{\tau}^{\tau+1} C(x + v) \exp\{M(v + \Delta - s)\} \, N \, dv \, dW_s
\overset{d}{=} \int_0^\Delta \int_{\tau}^{\tau+1} \alpha(v + \Delta - s, x - \Delta + s) \, dv \, ds
+ \int_{\tau}^{\tau+1} C(x + v) \exp\{Mv\} \, dv \times \int_0^\Delta \exp\{M(\Delta - s)\} \, N \, dW_s
\]

is Normal distributed. As is usual within a principle component analysis, we then compute the empirical covariance matrix \(\hat{\Sigma}\) of the i.i.d. vectors \(\tilde{F}_j = (F(j, \tau, x))_{(\tau, x) \in C}\) and decompose it as

\[
\hat{\Sigma} = U \times \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & 0 & \\
\vdots & \ddots & \ddots & \\
0 & 0 & \cdots & \lambda_K
\end{pmatrix} \times U' = \sum_{\nu=1}^K \lambda_\nu u_\nu u_\nu',
\]

\(^{18}\)For our dataset described in Section 2, we have \(t_1 = 1977\), \(\Delta = 1\), \(N = 31\), and \(K = 2850\) since we rely on term/age combinations with \(0 \leq \tau \leq 75\), \(20 \leq x \leq 95\), and \(20 \leq x + \tau \leq 95\).
where \( U = (u_1, u_2, \ldots, u_K) \) is an (orthogonal) matrix consisting of the eigenvectors of \( \hat{\Sigma} \) and \( \lambda_\nu, \nu = 1, 2, \ldots, K \), are the corresponding (ordered) eigenvalues. Picking the \( d \) greatest eigenvalues that explain the majority of the data, we obtain the approximation

\[
\bar{F}_j \approx E[\bar{F}_j] + \sum_{\nu=1}^{d} u_{\nu} \sqrt{\lambda_\nu} Z_{\nu,j},
\]

where \( Z_{\nu,j} \) are i.i.d. (scalar) standard Normal random variables, \( 1 \leq j \leq N - 1 \). In particular, when choosing \( d = 1 \), we obtain

\[
\bar{F}_j \approx E[\bar{F}_j] + u_1 \sqrt{\lambda_1} Z_{1,j},
\]

and by evaluating or approximating the integrals in Equation (2.13) we can find suitable candidates for the components \( C(x) \), \( M \), and \( N \)—and thus for the volatility function—from the eigenvectors. For instance, when relying on a crude midpoint approximation for the first integral and an Euler approximation of the stochastic integral, we have

\[
u_1 \sqrt{\lambda_1} \approx \left( C \left( x + \tau + \frac{1}{2} \right) \times \exp \left\{ M \left( \tau + \frac{1}{2} + \Delta \right) \right\} \times N \right) \left( \tau, x \right) \in C.
\]

For the (re)estimation of parametric models that are developed based on this procedure, we can now take advantage of the known distribution of the vectors \( \bar{F}_j \). In particular, we can explicitly take the drift condition (2.2) into account. We refer to Zhu and Bauer (2011b) for more details.
Chapter 3

COHERENT PRICING OF LIFE SETTLEMENTS UNDER ASYMMETRIC INFORMATION

3.1 Introduction

While markets for mortality-linked securities are only slowly emerging in recent years, one segment has established itself as an abiding investment opportunity: The life settlement market. Evolving from so-called “viatical settlements” popular in the late 1980s that only targeted severely ill life insurance policyholders, life settlements generally involve senior insureds with below average life expectancies. Within such a transaction, both the liability of future contingent premiums and the benefits of a life insurance contract are transferred from the policyholder to a life settlement company, which may further securitize a bundle of contracts in the capital market (cf. Chen et al. (2011), Stone and Zissu (2006)). However, typical investments only involve a limited number of contracts from rather specific population strata. Thus, in contrast to other mortality derivatives, the key risk factors for a life settlement are idiosyncratic in nature, i.e. they derive from uncertainties in the individual rather than population mortality trends. Nevertheless, a sizable life settlement market may help promote other mortality-linked capital markets by increasing investors’ familiarity and awareness. In particular, all these securities are jointly advertised to deliver profitable in-
vestment opportunities with a low correlation to market systematic risk (cf. Cowley and Cummins (2005)).

However, at least for the life settlement market, recent investigations reveal a discrepancy of expected and realized returns (see Beyerle (2007) or Gatzert (2010)). More specifically, while expected returns calculated on a policy-by-policy basis range from 8-12% annually (cf. Gatzert (2010)), according to Braun et al. (2011) open-end life settlement funds between 2004 and 2010 on average returned approximately 4.8%, which considering substantial lock-up periods and redemption fees only slightly exceeds risk-free rates. While thus far this gap between anticipated and realized returns has been attributed to the bad quality of the underlying life expectancy estimates (see Gatzert (2010) and references therein), potential systematic biases should have been swiftly corrected whereas unsystematically erroneous estimates by themselves do not provide a coherent explanation for an aggregate underperformance. It is exactly this difficulty in assessing the expected return of a life settlement due to “unique risks” that led rating agencies to decline rating these “death-bet securities”.¹ This, on the other hand, could impede the further development of this market, as ratings are usually essential to a broad investor interest.

In this paper, we propose a different viewpoint on the seemingly high reported excess returns based on adverse selection. We start by presenting a simple one-period expected utility model to derive the offer price in a competitive life settlement market. Specifically, by analyzing a representative policyholder’s decision to settle her policy, we derive the per-policy expected profit for a representative life settlement company in two cases: (1) When the

policyholder’s lifetime distribution is public information, i.e. if the life settlement company and the policyholder have the same (symmetric) information on the policyholder’s condition; and (2) when the policyholder has private insights that improve her assessment of her lifetime distribution, i.e. when there exists hidden (asymmetric) information. In the former case (1), the life settlement company relies on its best estimate of the policyholder’s expected lifetime to derive the actuarially fair offer price. In contrast, if information is asymmetric, a rational life settlement company will not directly use the (unconditional) expected lifetime for pricing, even if the estimation itself is unbiased and the company is risk-neutral. Rather, it will adjust the pricing scheme to cover possible one-sided losses because profitable policyholders may walk away from the transaction whereas unprofitable offers are accepted. In particular, this renders an offer price lower than the actuarially fair price—which is typically the benchmark used in practice where the hurdle rate is set according to investors’ return expectations.

Building on these insights, we then derive applicable pricing formulas for life settlement transactions within a simple lifetime utility framework. More precisely, akin to Vaupel et al. (1979), we propose a frailty model for generation life tables in order to introduce heterogeneity. By evaluating the policyholder’s (optimal) lifetime utility of consumption and bequests based on the private information regarding her lifetime distribution in a simple model setup, we then derive the threshold set for accepting the settlement offer—and, thus, the offer price. Moreover, we discuss various generalizations, particularly when the policyholder has the option to settle at various dates. Here, the derivation of the offer price requires the solution

---

2This observation resembles so-called clientele effects identified in the insurance literature (see e.g. Hoy and Polborn (2000) and Villeneuve (2003)).
of an optimal stopping problem and uncertainties in the population mortality trend become material.

Our numerical examples reveal that—depending on various parameter choices—the impact of asymmetric information on the offer price varies. If policyholders receive little utility from their life insurance contracts and the settlement is highly priced, private information may be irrelevant. However, if the policyholder is on the verge of keeping her policy, if pricing is not beneficial, or if she can settle at various dates against a small cost, the information asymmetry may considerably affect the equilibrium offer price and—in extreme cases—may even yield a breakdown of the market (cf. Akerlof (1970)).

The remainder of this paper is organized as follows: The simple one-period model is introduced in Section 3.2. The generalized lifetime utility framework to derive the offer price as well as extensions are presented in Section 3.3. Section 3.4 contains our numerical results, and finally Section 3.5 concludes.

3.2 One-Period Model for Life Settlements

In order to obtain our key implication while keeping the setup as succinct as possible, we commence by looking at a simple one-period expected utility model. More precisely, we consider a representative policyholder endowed with a one-period term-life insurance policy with face value $F$, no future contingent premiums, and zero cash surrender value. We identify the policyholder’s condition by her survival probability to the end of the period, $p$. Moreover, we denote by $u(x)$ the policyholder’s utility function when surviving until the end of the period and receiving $x$ as a payment related to her life insurance policy,
whereas $v(y)$ denotes the policyholder’s bequest function when she dies within the period and receives $y$ from the policy.\footnote{Note that here the policyholder’s wealth levels in different states are implicitly incorporated in $u(\cdot)$ and $v(\cdot)$. Therefore, here $u(\cdot)$ and $v(\cdot)$ can be simply interpreted as “utilities from life insurance benefits”.
} Assume both $u(\cdot)$ and $v(\cdot)$ satisfy common conditions for utility functions such as increasingness and concavity. Finally, assume that the life insurance policy is nonseparable, i.e. the policyholder cannot partially settle her policy.\footnote{This assumption may not be innocuous as partial settlement may serve as a screening mechanism to reveal private information. However, since we are interested in pricing and actual transactions invariably entail “entire” policies, we refrain from a detailed analysis of this possibility.}

Without a secondary life insurance market, the policyholder has no other choice but to simply retain her policy. Her expected utility, $U^r$, can then be written as

$$U^r = p \times u(0) + (1 - p) \times v(F).$$

(3.1)

If there exists a secondary life market, on the other hand, the policyholder will have the additional option of settling the policy with a representative life settlement company. Here we assume the secondary life market is competitive and the life settlement company is risk-neutral. In the following two subsections, we derive the equilibrium offer price for a life settlement transaction with symmetric and asymmetric information on the policyholder’s condition $p$, respectively.

### 3.2.1 Symmetric Information

First, assume that $p$ is publicly observable by both the policyholder and the life settlement company. Then the expected profit of the life settlement company when offering a
price $x$ to the policyholder can be written as

$$\pi = (1 - p) \times F - x(1 + R),$$

where $R$ is the risk-free interest rate. Since a competitive equilibrium will yield zero expected profits, the equilibrium offer price, $OP^{\text{sym}}(p)$, will trivially be the *actuarially fair price*:

$$OP^{\text{sym}}(p) = \frac{(1 - p) \times F}{1 + R}.$$

Thus, the expected utility of the policyholder when settling, $U^s$, can be written as

$$U^s = p \times u((1 - p)F) + (1 - p) \times v((1 - p)F).$$  \hspace{1cm} (3.2)

The policyholder will choose to settle her policy if and only if $U^s \geq U^r$. Note that this condition is automatically valid when $u(\cdot) = v(\cdot)$, but it will not always be met under more general utility and bequest assumptions.

### 3.2.2 Asymmetric Information

Now consider the case in which the information on $p$ is asymmetric. More specifically, we assume that $p$ is only observable by the policyholder, whereas the life settlement company has to turn to an expert third party (a life evaluation company) to compile an estimate denoted by $\bar{p}$. Let $f(p|\bar{p})$ denote the estimated probability density of $p$ given $\bar{p}$. In order to
price the transaction, the life settlement company now has to rely on information from \( \bar{p} \), or more precisely, \( f(p|\bar{p}) \).

We first consider the case in which the offer price is solely determined from \( f(p|\bar{p}) \) without taking into account the policyholder’s behavior. The associated solution, \( OP^a(\bar{p}) \), is then a simple extension of \( OP^{sym}(p) \) as defined in Section 2.1:

\[
OP^a(\bar{p}) = \int_0^1 (1-p)f(p|\bar{p}) dp \times \frac{F}{1+R} = \frac{\mathbb{E}[(1-p)|\bar{p}] \times F}{1+R},
\]  

(3.3)

which simply results in \( OP^a(\bar{p}) = \frac{(1-\bar{p}) \times F}{1+R} \) if the estimation is unbiased. While \( OP^a(\bar{p}) \) is the benchmark used in practice, where the hurdle rate \( R \) is set according to investors’ return expectations, in what follows we show that \( OP^a(\bar{p}) \) is not economically correct in the sense that it does not consider the policyholder’s decision process. Instead, it comes with the strong—and possibly untenable—assumption that policyholders will always choose to settle.

The expected utility for a policyholder with condition \( p \) when retaining the policy, \( U^r(p) \), is given by Equation (3.1). Her expected utility when settling, \( U^s(p, OP) \), for an exogenously given offer price \( OP \), on the other hand, can be written analogously to Equation (3.2) as:

\[
U^s(p, OP) = p \times u(OP \times (1+R)) + (1-p) \times v(OP \times (1+R)).
\]
Therefore, given $OP$, a rational policyholder will settle her term-life policy if and only if

$$U^s(p, OP) - U^r(p) = p \times u(OP \times (1 + R)) + (1 - p) \times v(OP \times (1 + R)) - p \times u(0) - (1 - p) \times v(F)$$

$$= p \times (u(OP \times (1 + R)) - u(0) + v(F) - v(OP \times (1 + R))) + v(OP \times (1 + R)) - v(F) \geq 0.$$ 

Since $OP$ is naturally bounded between 0 and $\frac{F}{1+R}$, it is easy to verify that $u(OP \times (1 + R)) - u(0) + v(F) - v(OP \times (1 + R)) > 0$, and $v(OP \times (1 + R)) - v(F) \leq 0$, $\forall OP$. Hence, the above condition for settling can be equivalently written as:

$$p \geq \frac{v(F) - v(OP \times (1 + R))}{u(OP \times (1 + R)) - u(0) + v(F) - v(OP \times (1 + R))} \triangleq p^*(OP) \in [0,1]. \quad (3.4)$$

With Equation (3.4), the competitive equilibrium offer price under rational expectations, $OP^e$, can then be determined from the following equation:

$$OP^e(\bar{p}) \triangleq \arg \max_x \left\{ \int_0^1 [(1 - p)F - x(1 + R)] f(p|\bar{p}) dp = 0 \right\}, \quad (3.5)$$

where the integration on the right-hand side calculates the expected profit of the life settlement company for given offer price $x$ by only (rationally) taking into account health states above the threshold $p^*(x)$ as defined in Equation (3.4). The equilibrium offer price, $OP^e(\bar{p})$, is then the highest price that maintains a zero expected profit for the company.

The argument here resembles the so-called “Average Clientele Risk” (ACR) identified in the insurance literature, which suggests accounting for policyholders’ decision making under information asymmetry when pricing life insurance products (see e.g. Hoy and Polborn
(2000) and Villeneuve (2003)). However, there are at least two differences between our special case and the common ACR: First, from the policyholder’s perspective, in our case we have an information asymmetry at the time point of selling (settling) the insurance policy, whereas the common ACR investigates asymmetric information when purchasing life insurance policies. Second, in our case various equilibrium offer prices are derived independently for policyholders with different health states, whereas the “usual” ACR problem entails the identification of a level premium price for all prospective policyholders with different characteristics. Therefore, our findings can be interpreted as a supplement to the ACR discussion in the insurance literature.

The following proposition now provides a comparison result that gives rise to an alternative explanation for the discrepancy between expected and realized returns in the life settlement market.

**Proposition 3**

In the presence of asymmetric information with respect to \( p \), the rational expectation offer price, \( OP^e(\bar{p}) \), will be smaller than \( OP^a(\bar{p}) \), for all estimates \( \bar{p} \).

**Proof**

In order to prove \( OP^e(\bar{p}) \leq OP^a(\bar{p}) \), it is sufficient to show that

\[
\int_{p^*(OP^a)}^{1} ((1 - p) F - OP^a(1 + R)) f(p|\bar{p}) \, dp \leq 0
\]

\[
\iff F \times \int_{p^*(OP^a)}^{1} ((1 - p) - \mathbb{E}[(1 - p)|\bar{p}]) f(p|\bar{p}) \, dp \leq 0
\]

\[
\iff F \times \int_{p^*(OP^a)}^{1} f(p|\bar{p}) \, dp \times (\mathbb{E}[(1 - p)|\bar{p}, p \geq p^*(OP^a)] - \mathbb{E}[(1 - p)|\bar{p}]) \leq 0
\]

\[
\iff \mathbb{E}[p|\bar{p}, p \geq p^*(OP^a)] - \mathbb{E}[p|\bar{p}] \geq 0,
\]
which is trivially satisfied.

The key insight here is that a policyholder looking to settle her policy will reject the offer when the value of the policy is far underestimated. On the other hand, she will be happy to settle the policy for the offered price if it considerably exceeds the intrinsic value. Since the life settlement company cannot observe the “true” condition \( p \), it is not able to determine whether the offer price is too low or too high before the policyholder makes her choice. Therefore, as a way to balance expected profits in the competitive equilibrium, the life settlement company has to shift its pricing schedule to cover the possible tail loss. As a consequence, if the life settlement company determines its offer price based on Equation (3.3) and a given hurdle rate \( R \) set according to investors’ return expectations, the resulting returns will—on average—be lower than \( R \). Ascertaining a genuine return of \( R \) requires its incorporation into Equation (3.5) for \( OP^e(\bar{p}) \), where the decision to settle is taken into account. Thus, supposing life settlement companies or funds rely on Equation (3.3) for pricing their settlements,\(^5\) the reported discrepancies between “expected” and realized returns arise naturally in the presence of asymmetric information within the life settlement market.

Finally, it is worth pointing out that Equation (3.5) does not always have a solution. For instance, if the policyholder is risk-neutral, we have

\[
\int_{p^*(x)}^{1} \left( (1 - p)F - x(1 + R) \right) f(p|\bar{p}) \, dp < 0, \forall x > 0,
\]

\(^5\)See e.g. Erkmen (2011), Deloitte (2005), or www.lifesettlementguide.org.
so that the only admissible offer price is $x = 0$ for which clearly $p^*(0) = 1$. This is not surprising, since we are in the situation of a so-called “Lemon Market” as described by Akerlof (1970) in his seminal contribution. But even in the case of risk-averse policyholders, the market may break down, for instance, if $R$ is sufficiently high or if there is a significant transaction cost associated with eliciting the life expectancy estimation.

### 3.3 The Extended Framework

In Section 3.2, we derived equilibrium offer prices for life settlement transactions under asymmetric information with respect to the policyholder’s lifetime distribution in a simple one-period model. Building on the gained insights, in this section, we extend the simple setup to arrive at more coherent pricing formulas in a more realistic environment. Specifically, we consider a whole-life insurance policy with level annual premium $P$ and death benefit $F$. We start by introducing heterogeneity in the expected lifetime distribution via a frailty model for generation life tables. Subsequently, we describe how to evaluate the policyholder’s decision process in a lifetime utility framework and derive corresponding pricing formulas. Finally, we discuss possible generalizations.

#### 3.3.1 Heterogenous Generation Life Tables

Heterogeneity with respect to individual mortality rates has been long studied in the demographic as well as in the life insurance literature (see e.g. Vaupel et al. (1979) and Hoermann and Russ (2008)). A common approach to capture individual heterogeneity are so-called frailty models, i.e. to use a frailty factor by which the actual survival probabilities
of each individual deviate from the population survival probabilities. However, existing frailty models are commonly constructed based on a multiplicative structure that fails to directly connect the average of the heterogeneous individual life tables to the actually observed population life table, which is important in our case.

Thus, in this paper, we rely on an alternative assumption regarding the heterogeneity among individual life tables: Define $\tau p_x(T)$ as the $\tau$-year survival probability for the age-$x$ population at time $T$, and $\tau p_x^j(T)$ as the corresponding $\tau$-year survival probability for individual $j$. We are looking for a frailty model that satisfies

$$\mathbb{E}[\tau p_x^j(T)] = \tau p_x(T), \forall \tau,$$  \hspace{1cm} (3.6)

and

$$\tau p_x^j(T) \in [0, 1], \forall \tau, j.$$  \hspace{1cm} (3.7)

Based on these requirements, we propose:

$$\tau p_x^j(T) = \tau p_x(T) + A_j \times \min\{\tau p_x(T), 1 - \tau p_x(T)\} e^{-\gamma(\tau-1)},$$  \hspace{1cm} (3.8)

in which $A_j$ is an individual-specific (frailty) factor characterizing the heterogeneity, $A_j \in [-1, 1]$, and $\mathbb{E}[A_j] = 0$. Moreover, $\gamma \geq 0$ is used to allow for a time-diminishing effect of the frailty factor. It is easy to verify that by using the model in Equation (3.8), both conditions in Equation (3.6) and (3.7) are satisfied.
3.3.2 Evaluation of the Policyholder’s Decision

Similarly to Section 3.2.2, for a given offer price, the policyholder’s rational behavior is characterized by a threshold set, which may be derived from a lifetime utility model (see e.g. Chai et al. (2011) and references therein for more details on lifetime utility models). For simplicity, we assume that the policyholder can only settle her policy at time $T$. Extensions are discussed in Section 3.3.4.

Within the typical lifetime utility framework with time-separable preferences, the value function for the $x$-year old policyholder $j$ with initial wealth $W_0$ and life table $\tau p_x^j(T)$, $\tau = 1, \ldots, \omega - x$ when retaining the policy at time $T$, $V_T^r(W_0, j)$, is defined as

$$V_T^r(W_0, j) = \max_{c_T} \sum_{\tau=1}^{\omega-x} \tau^{-1} p_x^j(T) \beta^{\tau-1} \times u(c_T - P) + \sum_{\tau=1}^{\omega-x} (\tau^{-1} p_x^j(T) - \tau p_x^j(T)) \beta^{\tau} \times v(W_{\tau} + F),$$

(3.9)

s.t.

$$W_{\tau} = (W_{\tau-1} - c_T) \times \frac{1}{p(\tau-1, 1)}, \tau = 1, \ldots, \omega - x.$$ 

Here $F$ is the death benefit, $P$ is the periodic contingent premium, $\beta$ is the time discount factor of the policyholder, $p(t, \tau)$ is the time $t$ price of a zero coupon bond with maturity $t + \tau$, $u(\cdot)$ is the utility function of period consumption $c_T$ net of premium payment, and $v(\cdot)$ is the bequest function. The value function when settling her policy at offer price $OP$,

---

$^6$Here $\omega$ denotes the limiting age, i.e. $\tau p_x(T) = 0, \forall \tau \geq \omega - x$. 

$V^s_T(W_0, OP, j)$, on the other hand, is analogously defined as:

$$V^s_T(W_0, OP, j) = \max_{c_T} \sum_{\tau=1}^{\omega-x} \tau^{-1} \times u(c_T) + \sum_{\tau=1}^{\omega-x} (\tau^{-1} p_x^j(T) - \tau^{-1} p_x^j(T)) \beta^\tau \times v(W_\tau),$$

(3.10)

s.t.

$$W_1 = (W_0 - c_1 + OP) \times \frac{1}{p(0, 1)},$$

and

$$W_\tau = (W_{\tau-1} - c_\tau) \times \frac{1}{p(\tau-1, 1)}, \tau = 2, \ldots, \omega-x.$$

Therefore, for each given offer price $OP$, policyholder $j$ makes her decision by comparing the value functions and will only choose to settle if $V^s_T(W_0, OP, j) \geq V^r_T(W_0, j)$. Note that both $V^r_T(W_0, j)$ and $V^s_T(W_0, OP, j)$ depend on the individual life table $\tau p_x^j(T)$, or more precisely, the frailty factor $A_j$. Define

$$\Omega(OP) = \{ A_j : V^s_T(W_0, OP, j) \geq V^r_T(W_0, j) \}.$$

Thus, for a given offer price $OP$, $\Omega(OP)$ is the threshold set of $A_j$ in which settling is preferred to retaining.

### 3.3.3 Coherent Pricing Formulas

If there is no private information on the policyholder’s expected lifetime distribution, in analogy to Section 3.2.1, the offer price will be calculated following the actuarial equivalence principle. More precisely, for the $x$-year old individual $j$ with publicly observed future
survival probabilities \( \tau p_x^j(T) \), \( \tau = 1, \ldots, \omega - x \), and a whole-life policy with face value \( F \) and premium \( P \) at time \( T \), the equilibrium offer price, \( OP_{\text{sym}}(j) \), can be written as:

\[
OP_{\text{sym}}(j) = \sum_{\tau=1}^{\omega-x} \left[ (\tau - 1)p_x^j(T) - \tau p_x^j(T) \right] \times \frac{F}{(1+R)^\tau} - \tau - 1 \times \frac{P}{(1+R)^{\tau-1}}, \quad (3.11)
\]

in which \( R \) is the hurdle rate set by the life settlement company.

With asymmetric information introduced in the form of \( A_j \), on the other hand, akin to the ideas presented in Section 3.2.2, a coherent pricing formula needs to explicitly take into account the policyholder’s decision process. Specifically, denote by \( \bar{A} \) the estimate of \( A_j \) provided by the life evaluation company, and by \( f(A_j|\bar{A}) \) the corresponding probability density of \( A_j \) given \( \bar{A} \). The offer price for the \( x \)-year old individual with policy face value \( F \) and estimate \( \bar{A} \) at time \( T \), \( OP_{e}(\bar{A}) \), can then be derived via the following equation:

\[
OP_{e}(\bar{A}) \triangleq \arg \max_z \left\{ \int_{\Omega(z)} \left[ \sum_{\tau=1}^{\omega-x} \left[ (\tau - 1)p_x^j(T) - \tau p_x^j(T) \right] \times \frac{F}{(1+R)^\tau} - \tau - 1 \times \frac{P}{(1+R)^{\tau-1}} \right] - z \right\} f(A_j|\bar{A}) dA_j = 0 \}. \quad (3.12)
\]

Similarly to Section 3.2.2, the integration on the right-hand side calculates the expected profit of the life settlement company when offering \( z \) for the whole-life policy with the hurdle rate set at \( R \).

### 3.3.4 Generalization

In Section 3.3.2, we require that the policyholder can only settle her policy at time \( T \). However, this restriction may be problematic since the policyholder might also have the
option to settle in future periods. If we allow the policyholder to settle in any period, the
value function when settling, $V_s^T(W_0, OP(\bar{A}_T), j)$ remains essentially unchanged, and depends
on the life expectancy estimate at time $T$, $\bar{A}_T$. The value function when retaining at time
$T$, $\hat{V}_T^r(W_0, j)$, on the other hand, will be modified, since it now includes the possibility to
settle in future periods.

More specifically, the policyholder now solves a dynamic program:

$$\hat{V}_T^r(W_0, j) = \max_{c_1} u(c_1 - P) + \beta \times \mathbb{E} \left[ (1 - p_x^j(T)) \times u(W_1 + F) + p_x^j(T) \times \hat{V}_{T+1}^r(W_1, j) \right],$$

(3.13)

s.t.

$$W_1 = (W_0 - c_1) \times \frac{1}{p(0, 1)},$$

and

$$\hat{V}_{T+1}^r(W_1, j) = \max( V_{T+1}^r(W_1, j), V_{T+1}^s(W_1, OP(\bar{A}_{T+1}), j) ),$$

where $V_{T+1}^r(W_1, j)$ and $V_{T+1}^s(W_1, OP, j)$ are again defined as in Equation (3.13) and (3.10),
respectively, and $\hat{V}_{T+\tau}^r = 0, \forall \tau > \omega - x$.

Note that the new value function $\hat{V}_T^r(W_0, j)$ will generally increase since it now also
incorporates the possibility of settling in future periods. Therefore, when evaluating the
benefit of settling at time $T$, the policyholder now will also take into account the opportunity
cost of settling later. This will obviously lessen the willingness of the policyholder to settle
in the current period, and will thus truncate the threshold set $\Omega(OP)$. Hence, the effect
of asymmetric information and adverse selection on the pricing formula will be even more significant in this case.

Furthermore, note that the value function $\hat{V}_T^r$ now is nonlinear in the future survival probabilities $\tau p_{x+1}(T + 1)$. Thus, while in the case of one settlement date as in the previous subsections only concurrent life expectations in the form of the concurrent generation life table matter, when settling is possible in multiple periods systematic mortality risk at the population level becomes material.

### 3.4 Application

In this section, we evaluate exemplary life settlement transactions under asymmetric information to investigate the effect of adverse selection on the secondary life market. More precisely, we first generate projections of population generation life tables based on historical data, and then introduce our assumptions on the frailty model as well as on the policyholder’s preferences to derive offer prices as outlined in the previous sections. In doing so, we consider a representative female policyholder who purchased her whole-life policy in year $t = 1978$ at age 50, and in year $T = 2008$ is looking to settle the policy at age $x = 80$. For simplicity, we only consider the case of a single possible settlement date.

#### 3.4.1 Projections of Population/Individual Generation Life Tables

Generation life tables for the female population are derived using the Lee and Carter (1992) methodology\footnote{Here $\{\alpha_x\}$ and $\{\beta_x\}$ in the Lee-Carter model are estimated via the weighted least-squares algorithm, and $\{\kappa_t\}$ is further adjusted by fitting a Poisson regression model (cf. Booth et al. (2002)).} based on U.S. mortality data as available from the *Human Mortality*
Figure 3.1. Population Survival Probabilities Projected from the Lee-Carter Method

Database. More precisely, two series of forecasts are generated: First, in year 1978, the life insurer uses historical data from year 1958 to 1977 to generate forecasts of future survival probabilities when setting the fair level annual premium amount. Then, in year 2008, the life settlement company additionally incorporates mortality experience from year 1978 to 2007 to update the generation life table for pricing the life settlement transaction. Figure 3.1(a) and 3.1(b) display the projected population survival probabilities for an—at year 1978—50-year old female and an—at year 2008—80-year old female, respectively.

In deriving the annual premium in year 1978, for simplicity we disregard expenses and profit margins, and assume a constant risk-free rate \( r = 4\% \). The equivalence principle then yields a level annual premium of $16,245 per $1,000 death benefit.

In order to generate heterogeneous lifetime distributions from Equation (3.8) for a given population generation life table, the individual generation life tables are fully characterized

\[ \tau p_{50}^{1978}, \tau = 1, \ldots, 60 \]

\[ \tau p_{80}^{2008}, \tau = 1, \ldots, 30 \]
Figure 3.2. Empirical CDFs of Individual Survival Probabilities

by the frailty factor $A_j$ and $\gamma$. Assume that $\gamma = 0.1$ and that $\frac{A_j+1}{2}$ follows a Beta distribution with parameters $\alpha = \beta = 2$.\footnote{It is easy to verify that under this assumption, $A_j \in [-1, 1]$, $E[A_j] = 0.$} To illustrate the effect of these assumptions, Figure 3.2 shows the empirical cumulative distribution functions for $\tau p_{80}$ based on 1,000,000 simulations with $\tau$ at 1, 11, and 21, respectively.
3.4.2 Settlement Decision and Derivation of Offer Prices

In our numerical application, for the utility function $u(\cdot)$, we use a standard constant relative risk aversion (CRRA) form: $u(c) = \frac{c^{1-\rho}}{1-\rho}$. The bequest function, $v(\cdot)$, is assumed to be of the form

$$v(W) = \frac{1 + r}{r} \times \frac{(r + 1)W^{1-\rho}}{1 - \rho}.$$ 

This assumption entails that the bequest $W$ is transferred into a perpetuity with periodic payment $\frac{r}{1+r}W$ to the beneficiary, and thus we define the bequest function as the sum of the series of corresponding utility functions with the time discount factor $\beta = \frac{1}{1+r}$. The relative risk-aversion parameter $\rho$ is chosen at 1.584 as calibrated in Hall and Jones (2007). Moreover, we assume the initial wealth of the policyholder at the beginning of year 2008, $W_0$, is $500,000, whereas the death benefit of her whole-life policy, $F$, is $1,000,000. From Section 3.4.1, the annual contingent premium is then calculated at $16,245. Finally, with the risk-free interest rate fixed at 4%, the time discount factor $\beta$ is set at $\frac{1}{1.04}$.

We first calculate the offer price under symmetric information, $OP^{sym}$, for different choices of the hurdle rate $R$ based on Equation (3.11). For illustrative purposes, we simply assume that the estimated expected lifetime distribution is given by the population life table, i.e. $A_j = \bar{A} = 0$. Furthermore, contingent on the obtained $OP^{sym}$, the policyholder makes her optimal decision by comparing value functions under retaining and settling. Figure 3.3 shows the calculated $OP^{sym}$ as well as the associated value functions $V_T^r$ and $V_T^s$ for $R$ varying from 4% to 12%. In particular, we observe from Panel (b) that the secondary market transaction will only be executed when $OP^{sym} \geq 323,370$, i.e. when $R \leq 9.95\%$. 
In the case with asymmetric information introduced via $A_j$, we first need to evaluate the policyholder’s threshold set. For simplicity, we assume that the estimation function $f(A_j|\bar{A})$ is identical to $f(A_j)$, i.e. there is no prior information available.\textsuperscript{10} Thus, the actuarially fair offer price $OP^a$ is identical to $OP^{sym}$ calculated above.

By comparing the optimal value functions from Equations (3.9) and (3.10), we can calculate the associated reservation price for each policyholder type, i.e. the minimum offer price so that the policyholder will accept the settlement offer under $A_j$. The result is shown in Figure 3.4(a). In addition, Figure 3.4(a) displays $OP^{sym}$ with hurdle rates $R$ at 4%, 8%, and 12%, respectively. From the figure, we observe that when $R = 4\%$, $OP^{sym} = \$544,260$ exceeds the reservation price for all $A_j \in [-1, 1]$. Therefore, all policyholders will accept the offer and the equilibrium offer price $OP^e$ will be identical to $OP^{sym}$. When $R = 8\%$ or 12\%, on the other hand, the associated $OP^{sym}$ crosses with the reservation price curve. Hence, only

\textsuperscript{10}This assumption is justified if the generation life table in our case is interpreted as the one supplied by the life expectancy evaluator.
policyholders with relatively better private lifetime estimates will actually choose to settle their policies at $OP^{\text{sym}}$, whereas the rest will choose to retain their policies. In particular, the effect of adverse selection increases with $R$. Figure 3.4(b) now displays the threshold set $\Omega(OP)$ for $OP$ varying from $200,000$ to $450,000$. We observe that for given $OP$, only policyholders with $A_j$ greater than a threshold $A^*(OP)$ will settle, i.e. $\Omega(OP) = [A^*(OP), 1]$.

For given $\Omega(OP)$, we can continue to calculate $OP^e$ for our representative policyholder under different hurdle rates $R$ from Equation (3.12). Figure 3.5 shows the expected profits for the life settlement company under different scenarios, where we obtain a different pattern for each choice of $R \in \{0.07, 0.08, 0.09\}$. Specifically, Figure 3.5(a) displays the expected profits with $R = 0.07$ as a function of the offer price. In this case, we obtain $OP^e = OP^{\text{sym}} = 412,680$, i.e. there is no effect of adverse selection. In contrast, Figure 3.5(b) displays the expected profits when $R = 0.08$. Here, $OP^e$ is calculated at $367,930$, which is smaller than $OP^{\text{sym}} = 378,810$ and implies a modest effect due to asymmetric information. Finally,
Figure 3.5(c) displays the case with $R = 0.09$, and we find that in this case the expected profits will always be below zero no matter how much—or how little—the life settlement company is willing to offer. Therefore, in this case adverse selection from policyholders has a fatal impact as the market breaks down completely.

Hence, asymmetric information may or may not affect the offer price, where the eventual outcome depends on the underlying assumptions. For instance, by changing the distribution assumption of $A_j$ to a uniform distribution between $[-1, 1]$ and keeping all other parameters unchanged, we increase the heterogeneity among individual generation life tables. Figure 3.5(d) displays the expected profits of the life settlement company with $R = 0.08$ under the modified distributional assumption, and the equilibrium offer price $OP^e$ is now calculated at $339,100 < 367,930$, i.e. the effect of adverse selection is much stronger in this case. We conducted other sensitivity tests with similar outcomes. For instance, when reducing the risk-aversion parameter $\rho$ to 1.20, adverse selection becomes relevant at smaller values of the hurdle rate: With $R = 6\%$, the offer price is calculated as $OP^e = 440,050$ in contrast to $OP^{sym} = 451,010$.

Moreover, as discussed in Section 3.3.4, the impact of adverse selection will even become more pronounced when including the option to settle in future periods. In particular, the reservation function displayed in Figure 3.4(a) should then be steeper so that even for relatively small values of $R$, information asymmetries may affect the equilibrium offer prices.
Figure 3.5. Equilibrium Offer Prices

(a) Beta Distribution, $R = 0.07$

(b) Beta Distribution, $R = 0.08$

(c) Beta Distribution, $R = 0.09$

(d) Uniform Distribution, $R = 0.08$
3.5 Conclusion

In this paper, we discuss the effect of asymmetric information with respect to policyholders’ lifetime distribution on the pricing of life settlements in a competitive equilibrium. The key insight is that the possibility for the policyholder to walk away from “bad” offers while gladly agreeing to “good” offers will result in a shifted pricing schedule. In particular, the resulting equilibrium offer price will be lower than the “actuarially fair” price calculated on the basis of best estimate survival probabilities and a hurdle rate set according to the investors’ return expectations—which is the benchmark used in practice. Thus, information asymmetries may explain the discrepancy between expected and realized returns within the life settlements market, an observation that previously was simply attributed to “errors” in life expectancy estimates.

We derive coherent pricing formulas that account for this effect in a simple lifetime utility framework, and we conduct several example calculations. Our results suggest that information asymmetries may or may not have a crucial impact on pricing secondary life insurance market transactions, depending on the underlying model specification and the parameter assumptions. Thus, future research is necessary to quantify these effects. Nonetheless, our considerations provide a new angle on the financial analysis of life settlements, and therefore shed light on the nature of the “unique risks” within life settlements as recently discussed in the financial press.
REFERENCES


