Three Essays on Econometric Modelling of Financial Time Series

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximillians-Universität München

vorgelegt von

Christoph Berninger

am 16ten Juli, 2021

1. Gutachter: Prof. Stefan Mittnik, PhD
2. Gutachter: Prof. Dr. Martin Missong
3. Gutachter: Prof. Dr. Kai Carstensen

Tag der mündlichen Prüfung: 23.11.2021

Summary

Repeated severe market downturns since the turn of the century and the current low interest rate environment represent two main challenges for banks and insurance companies in the past decades. This dissertation covers precisely these two topics. The first two essays apply financial mathematical and statistical methods to improve the modelling of interest rates. Essay 3 examines the practical application of risk-based investment strategies.

The first essay proposes a new modelling framework for the *2-Additive-Factor Gaussian (Gauss2++) model*. This interest rate model is also known in a different representation as the *2-Factor Hull-White model*. It is commonly used in the insurance industry to calculate a company's liabilities as well as for risk management and forecasting purposes. Our framework allows us to apply the model under the risk neutral and the real world measure in a consistent manner. It further accounts for stable and realistic long-run interest rate forecasts under the real world measure. In this context "stable" means that long-run forecasts (e.g. 30 years forecasts) do not fluctuate much if the model is calibrated on, e.g., a yearly basis. Large fluctuations of long-run forecasts is a common problem of the Gauss2++ model in the insurance industry and results from the fact that a constant function calibrated on short-term interest rate forecasts is used to determine the transition from the risk neutral to the real world measure. We introduce a time dependent function, which allows us to regularize the interest rates in the long forecasting horizon without loosing the analytic tractability of zero-coupon bond prices or increasing computational effort much.

In Essay 2 we propose a time-varying autoregressive model of order one for short- and long-term predictions of interest rates. Our model improves the forecasting performance in the short horizon compared to the Gauss2++ model, while still producing realistic values in the long-run. Interest rates as well as other economic variables

show a (close to) random walk behaviour although economic theory says that they are mean reverting. We assume that this behaviour originates from a time-varying mean reversion level. By modelling the constant of an autoregressive model of order one (AR(1) model) by a stationary latent process we account for these changes in our model. We use a Bayesian formulation to incorporate prior assumptions on the mean reverting process in the model and thereby regularize predictions in the far future. We use MCMC-based inference by deriving all full conditional distributions and employ a Metropolis-Hastings within Gibbs sampler approach to sample from the posterior (predictive) distribution. In combining data-driven short-term predictions with long-term distribution assumptions our model is competitive to existing methods in the short horizon while yielding reasonable predictions in the long-run.

The third essay investigates the application of risk-based investment strategies. Having experienced a sequence of dramatic market downturns in the recent past, such as the dotcom crisis starting in 2000, the financial crisis around 2008, the European sovereign debt crisis unfolding in 2010, and most recently the Corona crash in 2020, there has been an increased interest in risk-based investment strategies. In the literature it is shown that risk-based investment strategies can produce large alphas and increase Sharpe ratios. However, such strategies have not yet made major inroads into many practical applications. In this essay we fill this gap and apply a risk-based investment strategy on a framework for a pension scheme, which implements an intergenerational risk transfer by establishing a collective reserve. Combining this pension scheme with the risk-based investment strategy improves the performance of the pension fund and reduces the risk of a negative reserve in times of a market crisis. We furthermore investigate the implications of imposing varying degrees of diversification across assets in such a scheme.

Zusammenfassung

Wiederholt schwere Markteinbrüche seit der Jahrtausendwende und das aktuelle Niedrigzinsumfeld stellten Banken und Versicherungen in den letzten Jahrzehnten vor große Herausforderungen. Diese Dissertation befasst sich genau mit diesen beiden Problemen. Die ersten beiden Essays wenden finanzmathematische und statistische Methoden an, um die Modellierung von Zinsen zu verbessern. In Essay 3 werden praktische Anwendungen risikobasierter Anlagestrategien untersucht.

Das erste Essay schlägt einen neuen Modellierungsrahmen für das *2-Additive-Factor-Gauß-Modell (Gauss2++ Modell)* vor. Dieses Modell ist in einer anderen Darstellung auch bekannt als das *2-Factor Hull-White-Modell*. Es wird in der Versicherungsbranche häufig zur Bewertung von Verbindlichkeiten des Unternehmens sowie für Risikomanagement- und Prognosezwecke verwendet. Unser Modellierungsrahmen ermöglicht es, das Modell unter dem risikoneutralen und dem sogenannten real-world Maß in konsistenter Weise anzuwenden. Darüber hinaus trägt es zu stabilen und realistischen langfristigen Zinsprognosen unter dem real-world Maß bei. In diesem Zusammenhang bedeutet "stabil", dass langfristige Prognosen (z.B. 30-Jahres-Vorhersagen) nicht stark schwanken, wenn das Modell z.B. auf Jahresbasis kalibriert wird. Große Schwankungen von langfristigen Prognosen sind ein bekanntes Problem des Gauss2++ Modells in der Versicherungswirtschaft und resultieren daher, dass eine an kurzfristigen Zinsprognosen kalibrierte konstante Funktion verwendet wird, um den Übergang von der risikoneutralen in die reale Welt zu modellieren. Wir führen eine zeitabhängige Funktion ein, die es uns ermöglicht, langfristige Zinsprognosen zu regulieren, ohne die analytische Berechenbarkeit von Nullkupon-Anleihen zu verlieren oder den Rechenaufwand stark zu erhöhen.

In Essay 2 schlagen wir ein zeitvariables autoregressives Modell erster Ordnung für kurz- und langfristige Zinsvorhersagen vor. Unser Modell hat eine verbesserte Prog-

nosegüte für kurzfristige Zinsvorhersagen im Vergleich zum Gauss2++ Modell und liefert gleichzeitig langfristig realistische Werte. Zinsen sowie auch andere ökonomische Zeitreihen verhalten sich (nahezu) wie ein Random Walk, obwohl sie nach der Theorie immer wieder zu einem Mittelwert zurückkehren sollten. Wir vermuten, dass dieses Verhalten von einem zeitvariablen Mean-Reversion-Niveau herrührt. Indem wir die Konstante eines AR(1)-Modells durch einen stationären latenten Prozess modellieren, berücksichtigen wir diese Eigenschaft in unserem Modell. Wir verwenden einen Bayesianischen Ansatz, um a-priori Annahmen über den Mittelwertprozesses in das Modell aufzunehmen und damit die langfristigen Vorhersagen zu regulieren. Wir verwenden MCMC-basierte Inferenz, indem wir alle vollständig bedingten Verteilungen herleiten und einen Metropolis-Hastings-Ansatz im Rahmen eines Gibbs-Samplers verwenden, um eine Stichprobe aus der posteriori (prädiktiven) Verteilung zu erhalten. Durch die Kombination datengetriebener kurzfristiger Vorhersagen mit langfristigen Verteilungsannahmen hat unser Modell für kurzfristige Vorhersagen eine ähnliche Güte wie existierende statistische Modelle, liefert aber zusätzlich langfristig realistische Vorhersagen.

Das dritte Essay untersucht die Anwendung risikobasierter Anlagestrategien. Nach einer Reihe dramatischer Markteinbrüche in der jüngeren Vergangenheit, wie der Dotcom-Krise ab 2000, der Finanzkrise um 2008, der europäischen Staatsschuldenkrise im Jahr 2010 und zuletzt dem Corona-Crash im Jahr 2020, gab es ein erhöhtes Interesse an risikobasierten Anlagestrategien. In der Literatur wird gezeigt, dass risikobasierte Anlagestrategien große Alphas erzeugen und Sharpe-Ratios erhöhen können. Allerdings werden solche Strategien noch selten in praktischen Anwendungen eingesetzt. In diesem Essay schließen wir diese Lücke und wenden eine risikobasierte Anlagestrategie im Rahmen eines Rentensystems an, welches einen generationsübergreifenden Risikotransfer mit Hilfe einer kollektiven Reserve umsetzt. Die Kombination eines solchen Rentensystems mit einer risikobasierten Anlagestrategie verbessert die Performance des Pensionsfonds und verringert das Risiko einer negativen Reserve in Zeiten einer Marktkrise. Darüber hinaus untersuchen wir die Auswirkungen unterschiedlicher Diversifizierungsvorgaben in einem solchen System.

Acknowledgements

I would like to thank a few people without whose help and support I would not have been able to write my doctoral thesis.

First and foremost, I would like to thank Stefan for giving me the opportunity to do my doctorate with him. With his charismatic nature, he always managed to keep you working more motivated and in a better mood, even if he just walked past your office and only exchanged a few words. After discussing with him your doctoral thesis, you came out of the conversation more optimistic and full of energy, even if you came to it disillusioned and almost desperate.
I also would like to thank Martin Missong and Kai Carstensen for joining my PhD committee as external reviewers and Volker Schmid and Christian Heumann for agreeing to serve as internal members in the committee.

Furthermore, I would like to thank all employees at the institute and at the chair, who let me remember my PhD time as a wonderful time. Above all Henry and Maria, with whom I worked the longest door to door and in the same office. Without Henry I would have had only half as much fun here and desperate much more often. Without Maria's musical support and constant reminder („You won't get your PhD with this attitude!“) I would probably have needed a lot longer to finish this thesis. I would also like to thank Martina. She is a real godsend for this chair. You could always come to her when you had a problem, and she almost always had a solution ready.
I would also like to thank my colleagues at ROKOCO, especially Julian Pfeiffer, with whom I collaborated on a project for this doctoral thesis.

Last but not least I would like to thank my family for their support and encouragement. I am very happy and grateful to have such a wonderful support in my life.

Contents

I. Introduction and Methods

The application of financial econometric models combines statistical and mathematical theory and methods to understand and solve problems in financial economics. Especially the modelling of financial times series such as prices, returns, interest rates, etc. as well as portfolio optimization and risk measures are important fields in financial econometrics. This thesis contributes to two broad challenges: Modelling and forecasting of interest rates and investigating risk-based investement strategies in practical applications.

The following three subsections successively introduce the theory and methods for the three essays, which are presented in chapters 2 to 4.

1. Introduction to Essay 1

Essay 1 covers the Gauss2++ interest rate model and is titled:

The Gauss2++ model – A Comparison of Different Measure Change Specifications for a Consistent Risk-Neutral and Real World Calibration

The challenge of modelling interest rates consists in the multivariate setting since each interest rate with a specific maturity represents a modelling dimension. The Gauss2++ model belongs to the class of short-rate models, which only model the (one-dimensional) short-rate – the instantaneous interest rate – and derive all fundamental quantities (like spot interest rates or bonds) from it by using the financial mathematical method of risk-neutral valuation. In the following important definitions as well as mathematical concepts are introduced mainly using Brigo and Mercurio (2007) as a reference.

1.1. Definitions

The main building blocks of fixed income quantities are zero-coupon bonds. They represent the present value of one amount of currency, which is to be paid at a future time T. They can not be observed in the market, but are just auxiliary quantities from which other financial quantities – e.g., interest rates, swaps, swaptions, etc. – can be derived. A detailed definition is given in Brigo and Mercurio (2007):

1.1 Definition. *A* ***zero-coupon bond*** *is a financial contract, which guarantees the holder the payment of one amount of currency at a time* T*, which is the maturity of the bond. The contract value at time* $t < T$ *is denoted by* $P(t,T)$*.*

Another important building block in the fixed income market are spot interest rates. We distinguish between three types, namely: continuously-compounded, annually-compounded and simply-compounded spot interest rates. Interest rates with other compounding frequencies can be defined, but they are not relevant for the essays in this thesis. Interest rates can be calculated from zero-coupon bond prices and vice versa. A detailed definition can be found in Brigo and Mercurio (2007):

1.2 Definition. *The* ***continuously-compounded spot interest rate*** *is the constant rate, at which an investment of* $P(t,T)$ *at time* t *accrues continuously to yield one amount of currency at maturity* T*. It is denoted by* $r(t,T)$ *and is given by*

$$r(t,T) = \frac{-\ln(P(t,T))}{T-t}. \tag{I.1}$$

Therefore, a zero-coupon bond price at time t with maturity T in terms of continuously-compounded interest rates is

$$P(t,T) = e^{-r(t,T)(T-t)}.$$

1.3 Definition. *The* ***annually-compounded spot interest rate*** *is the constant rate, at which an investment of* $P(t,T)$ *at time* t *yields one amount of currency if reinvested once a year at this rate. It is denoted by* $R(t,T)$ *and is given by*

$$R(t,T) = \frac{1}{P(t,T)^{\frac{1}{T-t}}}. \tag{I.2}$$

Therefore, a zero-coupon bond at time t with maturity T in terms of annually-compounded interest rates is

$$P(t,T) = \frac{1}{(1 + R(t,T))^{(T-t)}}.$$

1.4 Definition. *The **simply-compounded spot interest rate** is the constant rate, at which an investment has to be made to produce an amount of one unit of currency at maturity, starting from $P(t,T)$ units of currency at time t, when accruing occurs proportionally to the investment time. It is denoted by $L(t,T)$ and is given by*

$$L(t,T) = \frac{1 - P(t,T)}{(T-t)P(t,T)}.$$

Therefore, a zero-coupon bond at time t with maturity T in terms of simply-compounded interest rates is

$$P(t,T) = \frac{1}{1 + L(t,T)(T-t)}.$$

From interest rates and zero-coupon bond prices derivatives can be constructed. In Essay 1 swaptions are used to calibrate the risk-neutral dynamics of the Gauss2++ model. A swaption is an option on an interest rate swap.

1.5 Definition. *An **interest rate swap** is a contract that exchanges interest rate payments between two parties starting from a future time instant.*

Often swaps involve the exchange of a fixed interest rate for a floating rate, or vice versa, to reduce or increase exposure to fluctuations in interest rates. The interest rate swap is called a payer swap (PS) if the fixed rate is paid and the floating leg is received. If it is the other way around it is called a receiver swap (RS). We assume that payments of the fixed and the floating rates occur at the same dates and with the same year fraction. The floating rate is reset at points in time $\{T_\alpha, T_{\alpha+1}, \ldots, T_{\beta-1}\}$ and payments occure at $\{T_{\alpha+1}, \ldots, T_\beta\}$. We further set $\mathcal{T} = \{T_\alpha, \ldots, T_\beta\}$ and $\tau = \{\tau_{\alpha+1}, \ldots, \tau_\beta\}$, where $\tau_i = T_i - T_{i-1}$. In this case the value at time t of a RS is given by

$$RS(t, \mathcal{T}, \tau, N, K) = N \sum_{i=\alpha+1}^{\beta} \tau_i P(t,T_i) \left(K - F(t,T_{i-1},T_i)\right), \tag{I.3}$$

where N is the nominal value of the contract, K is the fixed rate and $F(t,T_{i-1},T_i)$ is the forward interest rate. A suitable definition for the forward interest rate is given in Brigo and Mercurio (2007):

1.6 Definition. *The* ***simply-compounded forward interest rate*** *prevailing at time t for the expiry $T > t$ and maturity $S > T$ is denoted by $F(t, T, S)$ and is defined by*

$$F(t,T,S) = \frac{1}{S-T}\left(\frac{P(t,T)}{P(t,S)} - 1\right).$$

It is the fair value at time t for the fixed rate, K, in a forward rate agreement with expiry T and maturity S.

1.7 Definition. *A* ***forward rate agreement*** *is a contract that pays the holder a fixed interest rate for the period between T and S. Its value is given by*

$$N(S-T)(K - L(T,S)).$$

With Definition 1.5 and equation (I.3) we can now give a definition for a swaption, an interest rate derivative, which is often used in the insurance industry to calibrate interest rate models.

1.8 Definition. *A* ***swaption*** *is an option on an interest rate swap. A european payer (receiver) swaption gives the holder the right, but not the obligation, to enter a payer (receiver) swap, at a given future time, the swaption maturity.*

Usually the swaption maturity coincides with the first reset date of the underlying interest rate swap. If the value of the swap at maturity is positive the option will be exercised. Therefore, the payoff of a receiver swaption at maturity is given by applying the positive part function on the value of the receiver swap at maturity (c.f equation (I.3))

$$N\left(\sum_{i=\alpha+1}^{\beta} \tau_i P(T_\alpha,T_i)\left(K - F(T_\alpha,T_{i-1},T_i)\right)\right)^+.$$

Discounting this value to the current time yields the value of the swaption at time t. The value of a payer swaption can be derived analogously.

The value of the swaption entails information about what the market participants expect today at time t for the distribution of future interest rates at time T_α. For example, if the price of a receiver swaption decreases, this can mean that the expectation of market participants about future forward rates $F(T_\alpha,T_{i-1},T_i)$ increases and vice versa. If we want to determine the parameters of an interest rate model, we can use this forward-looking information by choosing the parameters such that the model reproduces the market price of the swaption. This approach is different to the approach using historic data. In this case the historic data determines distributional characteristics and with it the parameters of the model. For example, the maximum likelihood method chooses the parameters, with which the model would have most likely produced the historic data. Of course, both approaches have their advantages and disadvantages. Using historic data as a random sample is an intuitive approach. Neglecting this information could be criticised for the forward-looking approach. At the same time the time series must not behave the same in the future as it behaved in the past. It lies in the responsibility of the user, which approach is appropriate for the given application.

1.2. The Mathematical Concept of Risk-Neutral Valuation

The price of a financial product highly depends on its riskiness. Investors are typically risk-averse and therefore demand for an additional compensation to invest in risky assets. For example, to calculate the fair price of a claim on a risky amount paid in the future the expected payoff needs to be adjusted according to the inherent risk. This must be done for each claim individually.

The concept of risk-neutral valuation gives an alternative approach to calculate the fair price. Its basic idea is that a claim can be hedged, i.e., the payoff can be replicated by a self-financing strategy (c.f. Definitions 1.9 and 1.10). If the market is arbitrage free and complete, this strategy must have the same unique price as the claim. But if the claim can be replicated the price does not depend on the risk appetite of the investors. Therefore, one can build a theoretical risk-neutral world, in which all investors are risk-neutral, and today's price of the claim in this world should be the same. This theoretical world simplifies the calculation of the expected payoff since all assets behave the same in expectation, because a riskier asset does not need to have a higher expected return as the investors are risk-neutral. Therefore, the discounted payoff of any claim does not need to be adjusted for claim specific risk characteristics.

Harrison and Kreps (1979) and Harrison and Pliska (1981, 1983) were the first, who formulated this concept in a sound mathematical framework. We use a similar notation as Brigo and Mercurio (2007). Consider a time horizon $T > 0$, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a right-continuous filtration $\mathbb{F} = \{\mathcal{F}_t : 0 \le t \le T\}$. In the considered economy $K+1$ assets are traded continuously from time 0 until time T. Their price processes are modelled by a $(K+1)$-dimensional Itô process $S = (S_t)_{t\in[0,T]}$ with components $S^0, S^1, \dots, S^K$, where $S^i = (S^i_t)_{t\in[0,T]}$ for $i = 0, \dots, K$. The first asset S^0 is the bank account and defined by

$$dS^0_t = r_t S^0_t dt, \qquad S^0_0 = 1,$$

where r_t is the short-rate at time t. On this mathematical model of an economy Harrison and Kreps (1979) and Harrison and Pliska (1981, 1983) established a link between the economic concept of an arbitrage free market (c.f. Definition 1.11) and the existence of an equivalent martingale measure (or risk-neutral measure) (c.f. Definition 1.12). On this basis, they proved that the unique arbitrage free price of any attainable contingent claim is given by the expectation of the discounted claim payoff under such an equivalent martingale measure. This last result represents the main formula that explains how spot interest rates can be modelled by using a short-rate model. The following shall give more insight into this concept by first defining both: an arbitrage free market and an equivalent martingale measure.

For a formal description of an arbitrage free market we first have to provide a definition for a trading strategy as well as a self financing trading strategy (see Brigo and Mercurio (2007)):

1.9 Definition. *A **trading strategy** is a $(K+1)$-dimensional process $\phi = (\phi_t)_{t\le T}$, whose components $\phi^0, \phi^1, \dots, \phi^K$ are locally bounded and predictable. They represent the number of units of the assets $S^0, \dots, S^K$, respectively, held by an investor. The value process associated with a strategy ϕ is defined by*

$$V_t(\phi) = \phi_t S_t = \sum_{k=0}^{K} \phi^k_t S^K_t, \qquad 0 \le t \le T$$

1.10 Definition. *A strategy is **self-financing** if its value changes only due to changes in the asset prices.*

1.11 Definition. *An **arbitrage free market** is characterized by the non-existence of a self-financing strategy ϕ, such that $V_0(\phi) = 0$, but $\mathbb{P}(V_T(\phi) > 0) > 0$.*

The definition of an equivalent martingale measure is as follows (see Brigo and Mercurio (2007)):

1.12 Definition. *An **equivalent martingale measure** $\mathbb{Q}$ with respect to the bank account is a probability measure on the space $(\Omega, \mathcal{F})$, such that*

- *$\mathbb{Q}$ and $\mathbb{P}$ are equivalent measures, i.e. $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$ for every $A \in \mathcal{F}$.*
- *the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$ belongs to $L^2(\Omega,\mathcal{F},\mathbb{P})$, i.e. it is square integrable with respect to $\mathbb{P}$.*
- *The discounted asset price process $D(0,t)S_t$ is an $(\mathcal{F}, \mathbb{Q})$-martingale, i.e.,*

  $$E^{\mathbb{Q}}[D(0,t)S_t^k \mid \mathcal{F}_u] = D(0,u)S_u^k,$$

  for all $k = 1, \dots, K$ and all u, t with $0 \le u \le t \le T$.

$E^{\mathbb{Q}}[\cdot]$ denotes the expectation under $\mathbb{Q}$. $D(t,T)$ is the discount factor and is given by $D(t,T) = \frac{S_t^0}{S_T^0}$ since S^0 is the bank account.

Harrison and Pliska (1981) then proved that the market is free of arbitrage if (and only if) there exists an equivalent martingale measure. This link builts the basis for another fundamental result concerning attainable contingent claims (see Brigo and Mercurio (2007)):

1.13 Proposition. *Assume there exists an equivalent martingale measure $\mathbb{Q}$ and let H be an attainable contingent claim. Then, for each time t, $0 \le t \le T$, there exists a unique price π_t associated with H, i.e.,*

$$\pi_t = E^{\mathbb{Q}}[D(t,T)H|\mathcal{F}_t]. \tag{I.4}$$

1.14 Definition. *A* ***contingent claim*** *is a square-integrable and positive random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$.

1.15 Definition. *A contingent claim* H *is called* ***attainable*** *if there exists some self-financing* ϕ *such that* $V_T(\phi) = H$. *Such a* ϕ *is said to generate* H, *and* $\pi_t = V_t(\phi)$ *is the price at time* t *associated with* H.

According to Proposition 1.13 the price of an attainable contingent claim is given by the discounted expected payoff under the equivalent martingale measure. The point is that this holds for any attainable contingent claim and therefore generalizes the result of Black and Scholes (1973). Also a zero-coupon bond price can be seen as a contingent claim that has a certain payoff of one amount of currency at maturity. Short-rate models use equation (I.4) to price zero-coupon bonds, from which spot interest rates can be derived. This shows the importance of this result for short-rate models.

The second financial mathematical result, which we want to introduce in this chapter, is the Girsanov theorem (see Girsanov (1960)). Changing the probability measure from $\mathbb{P}$ to $\mathbb{Q}$ or vice versa also changes the dynamics of the underlying assets $S^0, \ldots, S^K$. If we change the measure according to Girsanov, the drift of the SDE changes but the diffusion coefficient remains the same.

1.16 Proposition. *Let* $(W_t)_{t\in[0,T]}$ *be an* $(\mathcal{F}_t)_{t\in[0,T]}$*-Brownian motion on* $[0,T]$ *under* $\mathbb{P}$ *and* $\Phi = (\Phi_t)_{t\in[0,T]}$ *a progressive process such that* $\int_0^T \|\Phi_s\|^2 ds < \infty$.
Let $Z_T = e^{(\int_0^T \Phi_s^{tr} dW_s - \frac{1}{2}\int_0^T \|\Phi_s\|^2 ds)}$ *and suppose* $E[Z_T] = 1$. *Then* $(\widetilde{W}_t)_{t\in[0,T]}$ *given by*

$$\widetilde{W}_t = W_t - \int_0^t \Phi_s ds$$

is an $(\mathcal{F}_t)_{t\in[0,T]}$*-Brownian motion on* $[0,T]$ *under* $\mathbb{Q}$, *where*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T.$$

Assuming the dynamics of the asset price process $(S_t^k)_{t\in[0,T]}$ is under $\mathbb{P}$ as follows

$$dS_t^k = \mu(t, S_t^k)dt + \sigma(t, S_t^k)dW_t,$$

where $\mu(t,S_t^k)$ and $\sigma(t,S_t^k)$ is the drift and the diffusion coefficient, respectively, depending on time t and the asset S^k itself. If we perform a measure change according to Girsanov the dynamics changes as follows:

$$dS_t^k = \left[\mu(t,S_t^k) + \sigma(t,S_t^k)\Phi_t\right]dt + \sigma(t,S_t^k)d\widetilde{W}_t,$$

where $(\Phi_t)_{t\in[0,T]}$ is the progressive process of the above proposition and $(\widetilde{W}_t)_{t\in[0,T]}$ is a Brownian motion under the new measure $\mathbb{Q}$.

The dynamics under the risk-neutral measure are needed if we are interested in pricing contingent claims. The dynamics under the real world measure are important if we are interested, e.g., in forecasting or calculating risk measures. Since we are interested in a consistent model for both worlds in Essay 1, the process Φ needs to be specified. If we have a complete market, i.e., all contingent claims are attainable, and we start from the real world measure, Φ is unique. But if we start the other way around from the risk-neutral dynamic, any function that fullfills the Girsanov conditions can be chosen. It is the responsibility of the user to choose Φ appropriately. In Essay 1 we suggest a time-dependent function for Φ, while the standard model in the insurance industry assumes a constant. This standard model suffers form the inflexibility of the former especially regarding predictions in the long horizon. We show that this can be avoided by employing a time-varying function.

1.3. Short-rate models

There exist two broad classes of short-rate models: equilibrium and no-arbitrage models. Most equilibrium short-rate models concentrate on the dynamic of the short-rate and derive interest rates with longer maturities from it. Prominent candidates of this model class include Vasicek (1977), Cox et al. (1985) and Duffie and Kan (1996). No-arbitrage models focus on exactly fitting the term structure at a specific point in time to prevent arbitrage opportunities. Representatives of this class are introduced by Hull and White (1990) and Heath et al. (1992).
Short-rate models are mainly used for pricing interest rate derivatives. They are, therefore, directly defined under the risk-neutral measure such that the method of risk-neutral valuation can be applied. Consider a one-factor short-rate model with

the following general form of a SDE:

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t,$$

where $(W_t)_{t\in[0,T]}$ is a Brownian motion under the risk neural measure $\mathbb{Q}$ and and μ and σ are two functions of time t and the short-rate r_t. Since a zero-coupon bond can be seen as a contingent claim, we can use equation (I.4) to calculate its price. As the price of a zero-coupon bond amounts to 1 at maturity T and the discount factor $D(t,T)$ is given by $\frac{S_T^0}{S_t^0} = e^{-\int_t^T r_s ds}$, this formula reduces to

$$\begin{aligned} P(t,T) &\overset{(I.4)}{=} E^{\mathbb{Q}}[D(t,T)P(T,T)|\mathcal{F}_t] \\ &= E^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds}\middle|\mathcal{F}_t\right]. \end{aligned} \tag{I.5}$$

From zero-coupon bond prices spot interest rates are readily defined, e.g., via (I.1) or (I.2). Therefore, spot interest rate scenarios can be generated and, e.g., used to price (complex) interest rate derivatives with a Monte Carlo approach. Equation (I.5), therefore, represents the core formula for short-rate models.

Working with interest rate forecasts or risk measures, the short-rate model needs to be regarded under the real world measure. If we perform a measure change according to Girsanov (c.f. proposition 1.16) the corresponding one-factor short-rate model has the following dynamic

$$dr_t = \Big[\mu(t,r_t) + \Phi(t,r_t)\sigma(t,r_t)\Big]dt + \sigma(t,r_t)d\widetilde{W}_t,$$

where $(\widetilde{W}_t)_{t\in[0,T]}$ is now a Brownian motion under the real world measure and Φ is a progressive process satisfying the conditions in the Girsanov proposition. It can be interpreted as the market price of risk, as it adds multiplied with the diffusion coefficient $\sigma(t,r_t)$ an additional amount to the risk-neutral drift $\mu(t,r_t)$.
To derive the zero-coupon bond price under the real world measure, we can not use formula (I.4) as the price process discounted with the bank account is no longer a martingale. To get a martingale under the real world measure we have to discount

the zero-coupon bond price by a cash flow $X_{P(t,T)}(t)$, which has the same drift as the zero-coupon bond under the real world measure. In general $X_{P(t,T)}(t)$ is not known and differs for each product, which is the reason why one switches to the risk-neutral measure if interested in pricing contingent claims, as one can use the bank-account for any product, here. But by specifying the change of measure according to Girsanov, i.e., by specifying Φ, we have implicitly defined this cash flow for every asset in the market. Therefore, we can calculate the price of a zero-coupon bond analogously to the risk-neutral case by using the martingale property and that $P(T,T) = 1$:

$$P(t,T) = E^{\mathbb{P}}\left[\frac{X_{P(t,T)}(t)}{X_{P(t,T)}(T)}\middle|\mathcal{F}_t\right]. \tag{I.6}$$

Note that we take the expectation under the real world measure $\mathbb{P}$. From formula I.6 interest rate scenarios under the real world measure can be generated and used for forecasts as well as to calculate risk measures.

Short-rate models differ in the underlying process of the short-rate. In Essay 1 of this thesis we focus on the Gauss2++ model – in a different representation also known as the 2-Factor Hull-White model. For this short-rate model there exists an analytic solution for the risk-neutral price of a zero-coupon bond (c.f. equation (I.5)). We show that for this model any time-dependent function for Φ can be used without loosing the analytic tractability of a zero-coupon bond price in the real world. We emphasise that a time-dependent function is necessary to regularize long-run predictions by referring to an analysis in Hull et al. (2014), where it is argued that otherwise unrealistic interest rates in the long horizon can be reached. This is a common problem in the insurance industry, where a constant function for Φ is used. This issue will be tackled in Essay 2. We suggest two time-dependent candidates for Φ, which are still easy to calibrate: a step function and a linear function. In our application we test all three approaches (”constant”, ”step” and ”linear”) over a time horizon of 3 years and indeed observe more realistic and stable interest rates in the long forecasting horizon for the time-dependent candidates.

2. Introduction to Essay 2

Essay 2 of this thesis concentrates on the forecasting performance of interest rate models and is titled:

A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction

Short-rate models perform poorly in forecasting (see, e.g., Duffee (2002)). Factor models – like the dynamic Nelson-Siegel model developed by Diebold and Li (2006) – often concentrate on short-term predictions (up to 1 or 2 years) and neglect the performance of the model in the long-run (e.g., up to 40 years). The goal of our research is to propose a model, which accounts for both: short- and long-term predictions. By accounting for long-term predictions we mean that we want a model that is able to regularize long-term forecasts to prohibit unrealistic results. We suggest a Bayesian time-varying autoregressive model, which is competitive to the dynamic Nelson-Siegel model in the short-horizon, but also generates realistic long-term predictions. This is especially of interest for insurance companies, as they use interest rate scenarios for up to 40 years to calculate risk measures for specific insurance products.

In Essay 2 we compare the results of our Bayesian time-varying autoregresseive model to the dynamic Nelson-Siegel model. We therefore introduce both models in the following.

2.1. The dynamic Nelson-Siegel Model

In practice interest rates are not directly observed but need to be estimated from bond prices. There are different estimation procedures:
McCulloch (1975) and McCulloch and Kwon (1993) estimate a smooth discount curve using cubic splines.

2.1 Definition. *The* ***discount curve*** *at time t is defined by the assignment*

$$T \to P(t,T), \qquad T > t.$$

The smooth discount curve is then converted to interest rates (c.f. equation (I.1)) for the relevant maturities. Shea (1983) shows that the resulting discount curve diverges at long maturities, which results in unlikely interest rate curve shapes. Vasicek and Fong (1982) apply a similar approach using exponential splines and a negative transformation of the maturity instead of the maturity itself. This ensures that the interest rates converge to a fixed limit with increasing maturity. A third and very popular approach is given by Fama and Bliss (1987), who construct forward rates by calculating the forward rate step by step that is necessary to meet the bond price with the next longer maturity. From these forward rates the corresponding interest rates are derived.
The development of a parsimonious model for the interest rate curve was also the concern of Nelson and Siegel (1987), who suggest a parsimonious exponential approximation for the instantanious forward rate.

2.2 Definition. *The* ***instantanious forward rate*** *at time t for maturity T, $T > t$ is denoted by $f(t,T)$ and is defined by*

$$f(t,T) = \lim_{S \to T^+} F(t,T,S) = \frac{\partial P(t,T)}{\partial T}.$$

$F(t,T,S)$ is the forward rate (c.f. Definition 1.6).

The "Nelson-Siegel forward rate" can be viewed as the sum of a constant and a Laguerre function, which is the product of a polynomial and an exponential decay term and is used for mathematical approximations. It has the following functional form:

$$f(t,T) = \beta_1 + \beta_2 e^{-\lambda(T-t)} + \beta_3 \lambda e^{-\lambda(T-t)},$$

where λ is a constant, which governs the exponential decay, and $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$. Spot interest rates are then given by the average of the instantaneous forward rates

$$r(t,T) = \frac{1}{T-t} \int_t^T f(t,m) dm,$$

which results in the following functional form for spot interest rates:

$$r(t,T) = \beta_1 + \beta_2 \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} + \beta_3 \left(\frac{1 - e^{\lambda(T-t)}}{\lambda(T-t)} - e^{\lambda(T-t)} \right).$$

This functional form is flexible enough to represent the common interest rate curve shapes: monotonic, humped and S-shaped. While the Nelson and Siegel model (see Nelson and Siegel (1987)) is a static model, Diebold and Li dynamize it by making the parameters β_1, β_2 and β_3 time-varying (see Diebold and Li (2006)), i.e.,

$$r(t,T) = \beta_{1,t} + \beta_{2,t} \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} + \beta_{3,t} \left(\frac{1 - e^{\lambda(T-t)}}{\lambda(T-t)} - e^{\lambda(T-t)} \right)$$

and interpret them as three latent dynamic factors (see Diebold and Li (2006)). The exponential terms are regarded as the factor loadings.

The loading on $\beta_{1,t}$ is one. Therefore, this factor can be interpreted as the level of the interest rate curve, as it loads on each maturity with the same amount. The loading on $\beta_{2,t}$ is $\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}$, which represents a function (of the time to maturity $T-t$) that starts in 1 and slowly decays for longer maturities. Loading a higher amount for the short than for the long maturities influences the steepness of the interest rate curve. Therefore, Diebold and Li interpret the second factor as the slope of the interest rate curve. And finally the loading of β_3 is $\left(\frac{1-e^{\lambda(T-t)}}{\lambda(T-t)} - e^{\lambda(T-t)}\right)$, which starts at 0, then increases and decays back to zero. It, therefore, influences the medium long maturities and $\beta_{3,t}$ can be interpreted as the curvature.

Diebold and Li estimate the three factors β_1, β_2 and β_3 from historic interest rate data and apply time series models on the extracted factor data for forecasting purposes. From the forecasted factors, forecasts of the whole interest rate curve can be derived. They focused their analysis on one, six and twelve months predictions, neglecting the results of long-term forecasts of, e.g., up to 40 years. This is in line with economic theory as today's information hardly has any predictive power for interest rates in such a long horizon. But one can anticipate that the model still produces forecasts that lie in an economically reasonable range. In Essay 2 we show that this is not the case for a standard AR(1) model, which is also applied by Diebold and Li (2006). As today's data can not be used to predict interest rates in the long horizon we suggest a Bayesian approach that allows to incorporate prior assumptions about the long-term behaviour in a sound mathematical way and without influencing short-term predictions much.

2.2. Stationary processes

A key factor for long-term forecasts is stationarity of a (possibly trend and seasonal adjusted) time series.

2.3 Definition. *A stochastic process* $(x_t)_{t\geq 0}$ *is* ***weakly stationary*** *if the following conditions hold:*

1. *the unconditional mean is constant, i.e.* $E[x_t] = \mu < \infty$ *for all* $t \geq 0$
2. *the unconditional variance is finite, i.e.* $Var(x_t) = \sigma < \infty$ *for all* $t \geq 0$
3. *the autocovariance* $Cov(x_{t_k}, x_{t_l})$ *with* $k < l$ *depends only on the time difference* $t_l - t_k$*, i.e.* $Cov(x_{t_k}, x_{t_l}) = Cov(x_{t_k+s}, x_{t_l+s})$*, where* $s \in \mathbb{N}$.

If the unconditional mean and unconditional variance lie in a reasonable range, the model produces realistic long-term predictions.

In Essay 2 we work with an AR(1) process, which is defined as follows

$$x_t = \alpha + \beta x_{t-1} + \epsilon_t, \tag{I.7}$$

where x_t represents the observed variable at time t and α and β are real valued constants. If $|\beta| < 1$ the process is stationary. The innovation process ϵ_t can be, e.g., a Gaussian white noise process.

2.4 Definition. *A* ***weak white noise process*** *is a stochastic process* $(x_t)_{t\geq 0}$ *if for all* t*,* t_k *and* t_l *with* $k \neq l$ *it holds*

1. $E[x_t] = 0$
2. $Var(x_t) = \sigma < \infty$
3. $Cov(x_{t_k}, x_{t_l}) = 0$

2.5 Definition. *A* ***Gaussian white noise process*** $(x_t)_{t\geq 0}$ *is a weak white noise process, where* x_t *is independent and identically (i.i.d) normally distributed for all* $t > 0$, *i.e.* $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$

If β approaches 1 the unconditional expectation of the process given by

$$E[x_t] = \frac{\alpha}{1-\beta}$$

and the unconditional variance

$$Var(x_t) = \frac{\sigma^2}{1-\beta^2}$$

approach infinity (for $\alpha > 0$). Note that if β is 1 the process is no longer stationary. Lanne and Saikkonen (2002) state that applying a linear model to economic time series that exhibit an almost non-stationary behaviour can lead to implications not in line with economic theory. For example, the almost non-stationary behaviour can only be captured by the model by a large variance, which might lead to values that are not realistic for the relevant time series. Lanne and Saikkonen (2002) argue that it might be an indication of factors not accounted for by the employed linear model if the time series shows an almost non-stationary behaviour. This is the point of view we take in Essay 2. We assume that unobserved factors influence the current mean reversion level, which results in longer deviations from the long-term mean giving the impression of non-stationarity. We account for these latent factors by introducing a latent continuous stochastic process for the constant α into an AR(1) model. In contrast to the Markov switching regression model by Hamilton (1989) or the threshold autoregressive model (TAR) and the smooth transition autoregressive model (STAR) introduced by Lim and Tong (1980) and Chan and Tong (1986), we do not assume that changes to the parameter occur on a discrete basis and stay piecewise constant, but arise continuously. This leads to a time-varying mean reversion level. As we assume that the latent process is weakly stationary itself the original process is still weakly stationary.

The idea is further visualized in Figure I.1. We simulated an almost non-stationary time series with an unconditional mean of zero. In Figure a) the simulated series is visualized as well as the expected future development of three fitted models: An AR(1)

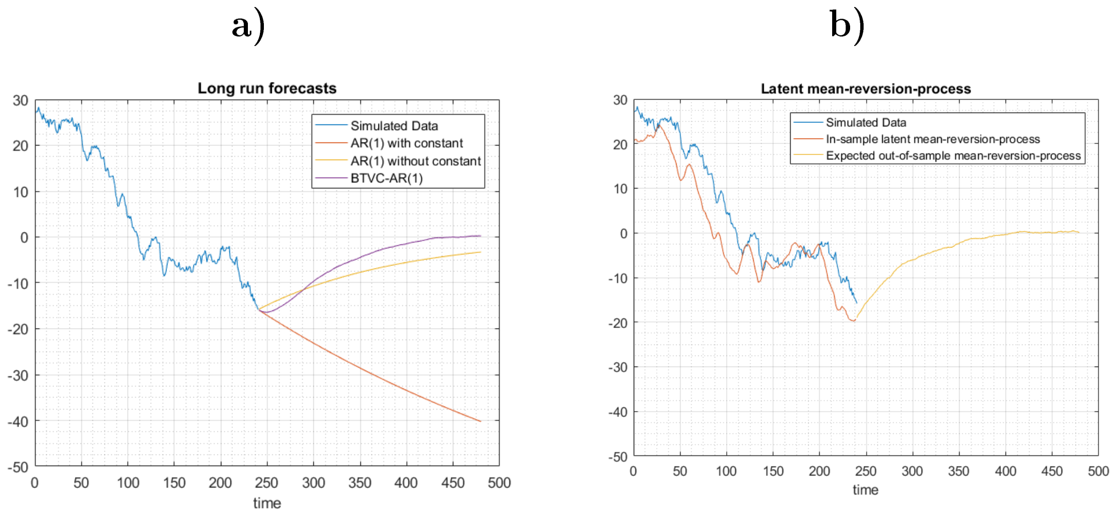


Figure I.1.: *A comparison of a linear AR(1) model with no restrictions for the constant parameter, a linear AR(1) model restricting the constant parameter to* 0 *and a BTVC-AR(1) model applied on a simulated time series.*

model with a constant, an AR(1) model with a constant restricted to 0 and our new model, the Bayesian time-varying constant autoregressive model of order 1 (BTVC-AR(1)). The AR(1) model with a constant takes an unconditional mean which is highly negative and even far away from the image of the simulated time series. The second AR(1) model restricts the constant to 0 to regularize the long-run mean to 0, but at the same time the expected values in the short-horizon are pulled to this long-run mean, which can lead to inferior short-term predictions. The BTVC-AR(1) model assumes that the almost non-stationary behaviour stems from fundamental changes in the mean reversion level due to unobserved factors. The time-varying constant leads to a time-varying mean reversion level, which is visualized in Figure b) of Figure I.1. In the short-horizon the model follows the current trend in the time series as the mean reversion level lies below the last observation. Since the latent process mean reverts to zero also the original time series tends to this value in expectation in the long-run. The Bayesian formulation further allows us to impose prior assumptions about the long-term mean and long-term variance via the latent process for parameter α. These prior assumptions do not influence the short-term predictions much but strike through for the longer forecasting horizon. In the short-horizon the model accounts for the current data and, therefore, for the current market situation, but in the long-run the distribution tends to the prior assumptions. Therefore, it allows to regularize the long-term predictions.

Extracting the level factor of the Nelson-Siegel model from historic interest rate data yields that this factor shows an (almost) non mean reverting behaviour, which implies an (almost) non-stationary behaviour. Economic theory predominantly assumes that interest rates are (in the long-run) mean reverting but as statistical tests would reject the stationarity assumption for interest rates, this theory lacks statistical evidence. We assume that unobserved macroeconomic factors and political interest rate decisions by central banks influence the temporary mean reversion level of interest rates. Therefore, the BTVC-AR(1) model can be used to account for that if distributional properties in the short- as well as in the long horizon are of interest. In the short-horizon the model is competitive to linear time series models, appropriate prior assumptions are used to regularize the interest rates in the long run.

3. Introduction to Essay 3

Essay 3 of this thesis investigates the application of risk-based investment strategies and is titled:

Risk-managed Collective Pension Schemes with Intergenerational Benefit Smoothing

The interest for risk-based investment strategies has grown in the last years, especially since the dotcom crisis starting in 2000, the financial crisis in 2008, the European sovereign debt crisis unfolding in 2010 and the market crash in 2020 because of the corona pandemic. These strategies, however, have not found their way into many practical applications. In Essay 3 we fill this gap and show the benefit of a risk-based investment strategy applied to a pension scheme framework. This pension scheme implements an intergenerational transfer of market risk by establishing a collective reserve. Combining this pension scheme with a risk-based investment strategy improves the performance of the pension investments and decreases the risk of a negative reserve in times of a market crisis.

In the following risk-based investement strategies are introduced and the idea of risk sharing penision schemes is explained.

3.1. Risk-based investment strategies

Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) showed that in context of a (zero-investment and self-financing) momentum strategy it is beneficial to scale the investment according to a risk measure. Popular choices for a risk measure are the variance or the volatility. Scaling corresponds to having a weight in the long and short position of the zero-investment strategy that is different from one and varies over time. Moreira and Muir (2017) generalized this result to additional investment factors: the market, size, value, profitability and investment factors from the Fama and French five-factor model (see Fama and French (1993)), the profitability and investment factors from the q-factor model of Hou et al. (2015), and the betting-against-beta factor of Frazzini and Pedersen (2014).
There are several ways to construct risk-managed portfolios, but they are all similar in spirit. Let r_t be the return of a portfolio. Scaling r_t by a function of a risk measure, ϑ, results in the return of the managed portfolio:

$$r_{\vartheta,t} = g(\vartheta_t) r_t \tag{I.8}$$

ϑ_t represents the conditional risk measure. The function $g(\cdot)$ scales the investment to meet a given risk target (e.g. the unconditional risk level). As we just scale the long and short position of the investment the managed portfolio is still self-financing. In the literature one can find the application of various risk measures. For example Barroso and Santa-Clara (2015) and Barroso et al. (2017) use the volatility, while Cederburg et al. (2020) and Moreira and Muir (2017) use the variance. Also the estimation of the risk measure may vary. Cederburg et al. (2020) use nonparametric sample estimates of realized variance, whereas Daniel and Moskowitz (2016) and Moreira and Muir (2017) use a parametric model. All these studies specify $g(\cdot)$ to be proporational to the inverse of the used risk measure, i.e.,

$$g(\vartheta) = \frac{c}{\vartheta},$$

where c is a constant to meet a given risk target. In Essay 3 we follow the approach of Moreira and Muir (2017) and obtain the risk-based portfolio return, $r_{RA,t}$, by setting

$$r_{RA,t} = \frac{c_t}{\sigma_t^2},$$

where σ_t^2 is the conditional variance estimated by a GARCH(1,1) model and c_t is a time-varying c-factor such that the unconditional variance matches a given target level conditioned only on past return observations.

3.2. Risk sharing pension schemes

In Essay 3 we apply a risk-based investment strategy to a pension scheme that balances the market risk between different generations of investors.

The best known pension schemes are the defined benefit (DB) and defined contribution (DC) pensions schemes. They represent the two extreme versions regarding the amount of market risk the two parties – the employer and the employee (or investor) – are exposed to.
A DB pension plan promises the employee a defined amount at their retirement. The employee is therefore not exposed to market risk as this specific amount does not depend on the performance of the capital markets. The employer on the other hand bears all the market risk, as he or she has to compensate for the amount if the capital markets perform bad such that the defined benefit amount is not reached at retirement. Because of low interest rates in recent years many DB plans were underfunded (see, e.g., Donnelly (2017)). Therefore, employers are closing down their DB plans and replacing them by DC plans instead (see, e.g., Donnelly (2017)).
A DC pension plan defines a certain contribution provided by the employer on a regular basis. The money is invested in the capital markets and paid off at retirement. The final amount depends on the performance of the capital markets. In a time of crisis this can result in huge differences for employees, whose retirement dates lie only a few months apart. Therefore, in this case the employee is fully exposed to market risk, while the employer is not.
The pension scheme we are working with in Essay 3 is a collective defined contri-

bution (CDC) scheme. It is a special type of risk sharing pension plans that allows different generations of savers to share the market risk. Risk sharing pension schemes in general have been shown to be welfare-improving, compared to the individually optimal lifecyclestrategy (see, e.g., Gollier (2008); Cui et al. (2011); Donnelly (2017)) and are discussed, e.g., in Pugh and Yermo (2008) and Blommestein et al. (2009). In our framework, we conduct the risk sharing via a collective reserve that belongs not to an individual investor but to the collective. In times, in which the market decreases, the negative return for the investors is compensated by releases of the reserve. If the capital markets perform better than expected part of the return is used to replenish the collective reserve. This approach smoothes the volatile market returns and achieves that the investment of all generations of investors performs similarly. In other words, the market risk is shared between them.

In Essay 3 we examine the application of a risk-based investment strategy to such a risk sharing pension scheme framework. We compare performance and risk measures of the pension fund with respect to a static weight strategy, which assumes constant weights over the investment horizon. Furthermore, we investigate the implications of imposing varying degrees of minimum diversifiation requirements across the assets under investigation.

Bibliography

Barroso, P. and Santa-Clara, P. Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120, 2015.

Barroso, P., Detzel, A. L., and Maio, P. F. Managing the risk of the low risk anomaly. In *30th Australasian Finance and Banking Conference*, 2017.

Black, F. and Scholes, M. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Blommestein, H. J., Kortleve, N., and Yermo, J. Evaluating the design of private pension plans: costs and benefits of risk-sharing. 2009.

Brigo, D. and Mercurio, F. *Interest rate models – theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.

Cederburg, S., O'Doherty, M. S., Wang, F., and Yan, X. S. On the performance of volatility-managed portfolios. *Journal of Financial Economics*, 2020.

Chan, K. S. and Tong, H. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7(3):179–190, 1986.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, pages 363–384, 1985.

Cui, J., De Jong, F., and Ponds, E. Intergenerational risk sharing within funded pension schemes. *Journal of Pension Economics & Finance*, 10(1):1–29, 2011.

Daniel, K. and Moskowitz, T. J. Momentum crashes. *Journal of Financial Economics*, 122(2):221–247, 2016.

Diebold, F. X. and Li, C. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

Donnelly, C. A discussion of a risk-sharing pension plan. *Risks*, 5(1):12, 2017.

Duffee, G. R. Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443, 2002.

Duffie, D. and Kan, R. A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406, 1996.

Fama, E. F. and Bliss, R. R. The information in long-maturity forward rates. *The American Economic Review*, pages 680–692, 1987.

Fama, E. F. and French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 1993.

Frazzini, A. and Pedersen, L. H. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

Girsanov, I. V. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5 (3):285–301, 1960.

Gollier, C. Intergenerational risk-sharing and risk-taking of a pension fund. *Journal of Public Economics*, 92(5-6):1463–1485, 2008.

Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

Harrison, J. M. and Kreps, D. M. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408, 1979.

Harrison, J. M. and Pliska, S. R. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11(3):215–260, 1981.

Harrison, J. M. and Pliska, S. R. A stochastic calculus model of continuous trading: complete markets. *Stochastic Processes and Their Applications*, 15(3):313–316, 1983.

Heath, D., Jarrow, R., and Morton, A. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105, 1992.

Hou, K., Xue, C., and Zhang, L. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

Hull, J. and White, A. Pricing interest rate derivative securities. *The Review of Financial Studies*, 3(4):573–592, 1990.

Hull, J., Sokol, A., and White, A. Short rate joint measure models. *Risk*, 10:59–63, 2014.

Lanne, M. and Saikkonen, P. Threshold autoregressions for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 20(2):282–289, 2002.

Lim, K. and Tong, H. Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Sociaty, B*, 42:245–92, 1980.

McCulloch, J. H. An estimate of the liquidity premium. *Journal of Political Economy*, 83(1):95–119, 1975.

McCulloch, J. H. and Kwon, H.-C. *US term structure data, 1947-1991.* Department of Economics, Ohio State University, 1993.

Moreira, A. and Muir, T. Volatility-managed portfolios. *The Journal of Finance*, 72 (4):1611–1644, 2017.

Nelson, C. R. and Siegel, A. F. Parsimonious modeling of yield curves. *Journal of Business*, pages 473–489, 1987.

Pugh, C. and Yermo, J. Funding regulations and risk sharing. 2008.

Shea, G. S. The japanese term structure of interest rates. 1983.

Vasicek, O. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

Vasicek, O. A. and Fong, H. G. Term structure modeling using exponential splines. *The Journal of Finance*, 37(2):339–348, 1982.

II. The Gauss2++ Model – A Comparison of Different Measure Change Specifications for a Consistent Risk Neutral and Real World Calibration

This chapter is a reprint of:
C. Berninger and J. Pfeiffer. 2021. The Gauss2++ Model – A Comparison of Different Measure Change Specifications for a Consistent Risk Neutral and Real World Calibration. https://doi.org/10.1007/s13385-021-00260-7. *European Actuarial Journal.*

Author Contributions

The project was initiated jointly by Christoph Berninger and Julian Pfeiffer. All proofs have been developed in close collaboration and have been improved and simplified in several rounds of revision. The writing as well as implementations in Python, figures and simulations have been done by Christoph Berninger.

European Actuarial Journal
https://doi.org/10.1007/s13385-021-00260-7

ORIGINAL RESEARCH PAPER

The Gauss2++ model: a comparison of different measure change specifications for a consistent risk neutral and real world calibration

Christoph Berninger[1,2] · Julian Pfeiffer[2]

Received: 1 July 2020 / Revised: 30 October 2020 / Accepted: 8 January 2021

Abstract

Especially in the insurance industry interest rate models play a crucial role, e.g. to calculate the insurance company's liabilities, performance scenarios or risk measures. A prominant candidate is the *2-Additive-Factor Gaussian Model (Gauss2++ model)*—in a different representation also known as the *2-Factor Hull-White model*. In this paper, we propose a framework to estimate the model such that it can be applied under the risk neutral and the real world measure in a consistent manner. We first show that any time-dependent function can be used to specify the change of measure without loosing the analytic tractability of, e.g. zero-coupon bond prices in both worlds. We further propose two candidates, which are easy to calibrate: a step and a linear function. They represent two variants of our framework and distinguish between a short and a long term risk premium, which allows to regularize the interest rates in the long horizon. We apply both variants to historical data and show that they indeed produce realistic and much more stable long term interest rate forecast than the usage of a constant function, which is a popular choice in the industry. This stability over time would translate to performance scenarios of, e.g. interest rate sensitive fonds and risk measures.

Keywords 2-Factor Hull-White model · Gauss2++ model · Risk neutral and real world · Change of measure · Time-varying market price of risk

The original version of this article was revised: Equations under Section 3 and Section 3.3 are updated.

✉ Christoph Berninger
christoph.berninger@stat.uni-muenchen.de

[1] Department of Statistics, LMU München, Munich, Germany

[2] ROKOCO GmbH, Ludwig-Ganghofer-Str. 6, 82031 Grünwald, Germany

Published online: 09 February 2021

1 Introduction

Two prominent approaches to model the term structure of interest rates are the classes of equilibrium and no-arbitrage models. Most equilibrium models concentrate on the dynamic of the short-rate—the instantaneous interest rate—and derive interest rates with longer maturities from it. Prominent candidates of this model class include the models of Cox et al. [3], Duffie and Kan [9] and Vasicek [18]. No-arbitrage models focus on exactly fitting the term structure at a specific point in time to prevent arbitrage possibilities. Representatives of this class are introduced by Heath et al. [11] and Hull and White [12].

Applications of these models often relate to pricing interest rate derivatives, which is the reason why they are directly defined under the risk neutral measure most of the time. A general form of a one-factor short-rate model under the risk neutral measure is, e.g. given by

$$dr(t) = \mu(t,r)dt + \sigma(t,r)dW(t),$$

where μ and σ are two functions, which can depend on time point t and the short-rate r, and W is a Brownian motion. A lot of advances in theoretic models and their estimation have been conducted in the last 30 years, but only in connection to pricing (see Diebold and Li [6]). Regarding these models little attention has been given to forecasting and risk management purposes (see Diebold and Li [6]). For these applications the corresponding model needs to be regarded under the real world measure. Under this measure the corresponding one factor short-rate model has the following dynamic

$$dr(t) = \left[\mu(t,r) + \lambda(t,r)\sigma(t,r)\right]dt + \sigma(t,r)d\widetilde{W}(t),$$

where λ is the market price of risk and can also depend on t and r. $\widetilde{W}$ is a Brownian motion under the real world measure. The exact functional choice for λ completes the model specification under the real world measure. Dai and Singleton [5] as well as Jong [14] use a fixed multiple of the model's variance for the market price of risk and investigate the in sample fit of specific short-rate models, but do not focus on forecasting. Duffee [8] concludes that the class of term structure models analysed in Dai and Singleton [5] fail in forecasting. He argues that a restriction for the market price of risk to be a fixed multiple of the variance reduces the flexibility of the model. Hull et al. [13] stress that the market price of risk for a model with few factors should be time-dependent. This results not from an economic interpretation but from a modelling issue because of an insufficient number of factors (see Hull et al. [13]). They estimated the market price of risk based on historical 3-month and 6-month interest rates and came to a similar result as Ahmad and Wilmott [1], Cox and Pedersen [4] and Stanton [17]. But they argue that this value is only valid in the short horizon. Keeping this market price of risk constant could lead to extreme risk premiums and interest rates in the long horizon.

In this paper we tackle exactly this problem for the Gauss2++ model. Instead of assuming a constant, we assume a time-varying function for the market price of risk.

In contrast to Hull et al. [13], who estimate the market price of risk for each forecasting horizon individually, we propose two parametric functions. The step function is the easiest non-constant function, which allows to model a market price of risk valid in the short and one valid in the long horizon. The linear function assumes that the market price of risk in the short horizon converges linearly to a long-term level. With these simplified time-dependent functions it is possible to account for the problem mentioned by Hull et al. [13] and the functions can still be easily estimated by historical data or calibrated in a forward looking manner to interest rate forecasts.

In our backtest we use a very similar calibration approach as described in Korn and Wagner [15]. The framework illustrated in this monograph has been developed by the Fraunhofer ITWM on behalf of the Produktinformationsstelle Altersvorsorge GmbH (PIA) and is the industry standard to classify packaged retail and insurance based investment products (PRIIPs) into chance-risk classes. For the interest rate model they use a Gauss2++ model with a presumed constant market price of risk. Following their calibration procedure allows us to compare our results to real applications in the insurance industry.

The structure of the paper is as follows. In Sect. 2 we introduce the Gauss2++ model under the risk neutral and the real world measure in a very general framework. In Sect. 3 we propose the constant function for comparison reason as well as the step and the linear function to specify the change of measure and explain how they can be estimated. All three variants of the Gauss2++ model are applied to data and backtested for the last 3 years in Sect. 4. In the final section the results are summarized and concluded.

2 The Gauss2++ model in the risk neutral and the real world

Throughout this section a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\mathcal{T}]}, \mathbb{M})$ is given, where $\mathbb{M}$ is either the risk neutral measure $\mathbb{Q}$ with respect to the bank account or the real world measure $\mathbb{P}$. $\mathcal{T}$ represents an appropriate modelling horizon. The bank account $(B(t))_{t\in[0,\mathcal{T}]}$ is given by

$$dB(t) = r(t)B(t)dt, \qquad B(0) = 1,$$

where $r(t)$ denotes the short-rate. We further adopted notations and descriptions of the Gauss2++ model from the relevant chapters in Brigo and Mercurio [2].

2.1 The Gauss2++ model under the risk neutral measure

Short-rate models differ in the underlying process for the short-rate. The Gauss2++ model assumes that the short-rate is given by a sum of two correlated normally distributed processes, $(x(t))_{t\in[0,\mathcal{T}]}$ and $(y(t))_{t\in[0,\mathcal{T}]}$, and a deterministic function φ, which is well defined on the time interval $[0, \mathcal{T}]$:

$$r(t) = x(t) + y(t) + \varphi(t), \qquad r(0) = r_0,$$

where r_0 is the short-rate at time point 0. The processes $(x(t))_{t\in[0,T]}$ and $(y(t))_{t\in[0,T]}$ satisfy under the risk neutral measure $\mathbb{Q}$ the following stochastic differential equations

$$\begin{aligned} dx(t) &= -ax(t)dt + \sigma dW^1(t), \qquad x(0) = 0, \\ dy(t) &= -by(t)dt + \eta dW^2(t), \qquad y(0) = 0, \\ \rho dt &= dW^1(t)dW^2(t), \end{aligned}$$

where a, b, σ, η are non-negative constants and $-1 \le \rho \le 1$ is the instantaneous correlation between the two Brownian motions W^1 and W^2.

Short-rate models derive spot rates via prices of zero-coupon bonds. As the short-rate in the Gauss2++ model is normally distributed, there exists an analytic solution for a zero-coupon bond price, $P(t, T)$, at time point t and maturity T:

$$P(t,T) = e^{-\int_t^T \varphi(s)ds - B(a,t,T)x(t) - B(b,t,T)y(t) + \frac{1}{2}V(t,T)}, \tag{1}$$

where

$$B(z,t,T) = \frac{1 - e^{-z(T-t)}}{z}$$

and

$$\begin{aligned} V(t,T) = &\frac{\sigma^2}{a^2}\left[(T-t) + \frac{2}{a}e^{-a(T-t)} - \frac{1}{2a}e^{-2a(T-t)} - \frac{3}{2a}\right] \\ &+ \frac{\eta^2}{b^2}\left[(T-t) + \frac{2}{b}e^{-b(T-t)} - \frac{1}{2b}e^{-2b(T-t)} - \frac{3}{2b}\right] \\ &+ 2\rho\frac{\sigma\eta}{ab}\left[(T-t) + \frac{e^{-a(T-t)} - 1}{a} + \frac{e^{-b(T-t)} - 1}{b} - \frac{e^{-(a+b)(T-t)} - 1}{a+b}\right]. \end{aligned}$$

A derivation can be found in Brigo and Mercurio [2]. With formula (1) for the zero-coupon bond price under the risk neutral measure spot rates can be directly derived via

$$r(t,T) = \frac{-ln(P(t,T))}{T-t}, \tag{2}$$

where $r(t, T)$ represents the spot rate at time point t and a maturity of T.

The financial market we actually model consists of a bank account and a set of zero-coupon bonds, $P(t, T)$, which differ in the maturity T. The dynamic of a zero-coupon bond price can be derived from the bond price formula in (1) by applying Ito's formula and is given by

$$dP(t,T) = P(t,T)\left[r(t)dt - \sigma B(a,t,T)dW^1(t) - \eta B(b,t,T)dW^2(t)\right].$$

A detailed derivation can be found in "Appendix 1". Note that all assets have the same drift as it is the case in the risk neutral world.

2.2 The Gauss2++ model under the real world measure

To calculate performance scenarios and risk indicators the Gauss2++ model must be regarded under the real world measure $\mathbb{P}$.

2.2.1 The change of measure

By specifying the Gauss2++ model under the risk neutral measure, we implicitly assume an arbitrage free market. Therefore, we can make the transition to a real world measure $\mathbb{P}$ by defining the change of measure according to Girsanov, who states that a progressive and square-integrable process $\mathbf{\Phi} = \left(\Phi^1(t), \Phi^2(t), \dots, \Phi^d(t)\right)_{t\in[0,T]}$ determines a new probability measure $\mathbb{P}$ such that if $(\widehat{W}(t))_{t\in[0,T]}$ is a standard d-dimensional $(\mathcal{F}_t)_{t\in[0,T]}$-Brownian motion under $\mathbb{Q}$, then

$$\breve{W}(t) := \widehat{W}(t) + \int_0^t \mathbf{\Phi}(s)ds$$

defines a standard d-dimensional $(\mathcal{F}_t)_{t\in[0,T]}$-Brownian motion under $\mathbb{P}$ (see Girsanov [10]).

The Gauss2++ model is a two-factor model and $\mathbf{\Phi}$ is therefore two-dimensional. Its components can be interpreted as the market price of risk for each factor in the model. We will represent $\mathbf{\Phi}$ such that the resulting processes $(x(t))_{t\in[0,T]}$ and $(y(t))_{t\in[0,T]}$ still belong to the class of Ornstein–Uhlenbeck processes under $\mathbb{P}$

$$\mathbf{\Phi}(t) = \begin{pmatrix} \Phi^1(t) \\ \Phi^2(t) \end{pmatrix} = \begin{pmatrix} -\frac{ad_x(t)}{\sigma} \\ -\frac{bd_y(t)}{\eta\sqrt{1-\rho^2}} + \frac{\rho a d_x(t)}{\sigma\sqrt{1-\rho^2}} \end{pmatrix}. \tag{3}$$

Note that we restrict $\mathbf{\Phi}$ to be a function of time. By this the change of measure only changes the mean reversion level. More general measure change specification can be applied. For example, Diez and Korn [7] introduce a measure change for the 1-Factor Vasicek model, which influences the mean reversion level as well as the mean reversion speed.

The conditions for the Girsanov theorem translate directly to the functions $d_x(t)$ and $d_y(t)$. In the following we will specify the change of measure via $d_x(t)$ and $d_y(t)$. An appropriate interpretation of these functions will be given in Sect. 2.2.2.

2.2.2 The dynamics under the real world measure $\mathbb{P}$

With the representation of $\mathbf{\Phi}$ as in (3) the dynamics of the processes x and y in the Gauss2++ model change according to Girsanov to

$$dx(t) = a(d_x(t) - x(t))dt + \sigma d\widetilde{W}^1(t), \qquad x(0) = 0, \tag{4}$$

$$dy(t) = b(d_y(t) - y(t))dt + \eta d\widetilde{W}^2(t), \qquad y(0) = 0, \tag{5}$$

where $\widetilde{W}^1$ and $\widetilde{W}^2$ are two correlated Brownian motions under $\mathbb{P}$. The derivation can be found in "Appendix 2". We observe that x and y are still Ornstein–Uhlenbeck processes with the solutions

$$x(t) = \int_0^t e^{-a(t-u)} a d_x(u) du + \sigma \int_0^t e^{-a(t-u)} d\widetilde{W}(u), \tag{6}$$

$$y(t) = \int_0^t e^{-b(t-u)} b d_y(u) du + \eta \int_0^t e^{-b(t-u)} d\widetilde{W}(u). \tag{7}$$

The mean reversion level of each process at time point t amounts to $d_x(t)$ and $d_y(t)$, respectively. Recall that the sum of $x(t)$ and $y(t)$ and a deterministic function $\varphi(t)$ under the risk neutral measure adds up to the instantaneous return rate $r(t)$ of a risk free investment. Changing the measure changes the mean reversion level at time point t from 0 to $d_x(t)$ for the process x and to $d_y(t)$ for the process y. Therefore, $d_x(t) + d_y(t)$ can be interpreted as the local long run risk premium of the short-rate—the amount, which is added in the real world to the risk neutral short-rate in the long run, if $d_x(t) + d_y(t)$ would stay constant over time. If this amount is negative, future bond prices increase in expectation compared to the risk neutral world and a risk averse investor, therefore, gets compensated for the risk of investing in a risky bond. This means in contrast to equity prices, in a market where investors are risk averse, future interest rates tend to be lower in the real world than in the risk neutral world (see, e.g. Hull et al. [13]). Therefore, $d_x(t)$ and $d_y(t)$ can be interpreted as the local long run risk premium the corresponding risk factor is mean reverting to at time point t.

In the following we will specify the change of measure by these two functions instead of the market prices of risk. The market price of risk of each risk factor is then directly defined by these two functions.

$$\text{Market price of risk of risk factor 1:} -\frac{a d_x(t)}{\sigma}$$

$$\text{Market price of risk of risk factor 2:} -\frac{b d_y(t)}{\eta\sqrt{1-\rho^2}} + \frac{\rho a d_x(t)}{\sigma\sqrt{1-\rho^2}}.$$

If we assume a step or a piecewise linear function for $d_x(t)$ and $d_y(t)$ the functional form of the individual market prices of risk are the same.

The dynamics of a zero-coupon bond with maturity T under $\mathbb{P}$ has the following form

$$\begin{aligned} dP(t,T) =& P(t,T)\left[r(t) - B(a,t,T) a d_x(t) - B(b,t,T) b d_y(t)\right] dt \\ & - P(t,T)B(a,t,T)\sigma d\widetilde{W}^1(t) - P(t,T)B(b,t,T)\eta d\widetilde{W}^2(t) \end{aligned} \tag{8}$$

The derivation can be found in "Appendix 3".

2.2.3 The bond price formula under the real world measure

The price of a zero-coupon bond under $\mathbb{P}$ is obtained by the same analytic formula as in (1). The only difference is that the x- and the y-process are now regarded under the real world measure (see Diez and Korn [7]). In the following we will shortly explain why the formula does not change under this new measure.

To calculate the price of a zero-coupon bond under the real world measure we use the following conditional expectation

$$\frac{P(t,T)}{X_{P(t,T)}(t)} = E^{\mathbb{P}}\left[\frac{P(T,T)}{X_{P(t,T)}(T)}\middle|\mathcal{F}_t\right],$$

where $X_{P(t,T)}$ represents the cash flow, with which we have to discount the zero-coupon bond such that the discounted price process is a martingale under $\mathbb{P}$. The dynamic of $X_{P(t,T)}$ coincides with the deterministic part of the zero-coupon bond price dynamic in (8) and is therefore specified by the change of measure:

$$dX_{P(t,T)}(t) = X_{P(t,T)}(t)\big[r(t) - B(a,t,T)ad_x(t) - B(b,t,T)bd_y(t)\big]dt, \qquad X_{P(t,T)}(0) = 1.$$

A short proof can be found in "Appendix 4". The solution of this dynamic is given by

$$X_{P(t,T)}(t) = e^{\int_0^t \left(r(u)-B(a,u,T)ad_x(u)-B(b,u,T)bd_y(u)\right)du}.$$

The price of a zero-coupon bond at time point t is therefore given by

$$P(t,T) = E^{\mathbb{P}}\left[\frac{X_{P(t,T)}(t)}{X_{P(t,T)}(T)}\middle|\mathcal{F}_t\right].$$

The ratio in the expectation amounts to

$$\frac{X_{P(t,T)}(t)}{X_{P(t,T)}(T)} = e^{-\int_t^T \left(r(u)-B(a,u,T)ad_x(u)-B(b,u,T)bd_y(u)\right)du}.$$

To determine the distribution of this ratio, we first derive the distribution of the integral in the exponent, i.e.

$$I(t,T) := \int_t^T \big(r(u) - B(a,u,T)ad_x(u) - B(b,u,T)bd_y(u)\big)du.$$

It can be shown that $I(t, T)$ is normally distributed with mean

$$M(t,T) = \int_t^T \varphi(u)du + \frac{1-e^{-a(T-t)}}{a}x(t) + \frac{1-e^{-b(T-t)}}{b}y(t) \tag{9}$$

and variance

$$V(t,T) = \frac{\sigma^2}{a^2}\left[(T-t) + \frac{2}{a}e^{-a(T-t)} - \frac{1}{2a}e^{-2a(T-t)} - \frac{3}{2a}\right] \\ + \frac{\eta^2}{b^2}\left[(T-t) + \frac{2}{b}e^{-b(T-t)} - \frac{1}{2b}e^{-2b(T-t)} - \frac{3}{2b}\right] \\ + 2\rho\frac{\sigma\eta}{ab}\left[(T-t) + \frac{e^{-a(T-t)}-1}{a} + \frac{e^{-b(T-t)}-1}{b} - \frac{e^{-(a+b)(T-t)}-1}{a+b}\right]. \tag{10}$$

The variance is the same as in the risk neutral world as the change of measure does not influence the variance of the processes. Note that also the mean has the same form as in the risk neutral case as the terms $B(a,u,T)ad_x(u)$ and $B(b,u,T)bd_y(u)$ in $I(t,T)$ cancel out in the calculations. The derivations can be found in "Appendix 5".

The expression $e^{-I(t,T)}$ is therefore log-normally distributed and the zero-coupon bond price under $\mathbb{P}$ is given by the same analytic formula as under $\mathbb{Q}$:

$$P(t,T) = E^{\mathbb{P}}\left[e^{-\int_t^T r(u)-B(a,u,T)ad_x(u)-B(b,u,T)bd_y(u)du} \mid \mathcal{F}_t\right] \\ = e^{-M(t,T)+\frac{1}{2}V(t,T)} \\ = e^{-\int_t^T \varphi(u)du - \frac{1-e^{-a(T-t)}}{a}x(t) - \frac{1-e^{-b(T-t)}}{b}y(t) + \frac{1}{2}V(t,T)}.$$

3 Local long run risk premium functions—specification and calibration

In the following three different types of functions for $d_x(t)$ and $d_y(t)$ are introduced: the constant, the step and the linear function. Following the interpretation in Sect. 2.2.2 these functions represent the long run risk premium for each risk factor at a specific time point t in the Gauss2++ model. The functional equations of the three types are

$$\begin{aligned} \text{Constant:}\quad & d_x(t) = d_x \\ & d_y(t) = d_y \\ \text{Step:}\quad & d_x(t) = \mathbb{1}_{t\le\tau}d_x + \mathbb{1}_{t>\tau}l_x \\ & d_y(t) = \mathbb{1}_{t\le\tau}d_y + \mathbb{1}_{t>\tau}l_y \\ \text{Linear:}\quad & d_x(t) = \mathbb{1}_{t\le\tau}(1-m_x t)d_x + \mathbb{1}_{t>\tau}l_x \\ & d_y(t) = \mathbb{1}_{t\le\tau}(1-m_y t)d_y + \mathbb{1}_{t>\tau}l_y \end{aligned}$$

where d_x, l_x, m_x and d_y, l_y, m_y are real valued constants and $\mathbb{1}_A$ represents the indicator function of a subset A.

The constant function assumes that the local long run risk premium is constant for the whole modelling horizon. The latter two functions distinguish between a local long run risk premium valid in the short and in the long horizon, seperated at time point τ. As mentioned in Sect. 2.2.2 the same holds for the market price of risk, respectively. Hull et al. [13] argue that a time-varying market price of risk is necessary to account for unobserved risk factors and to prevent unrealistic interest rate forecasts in the long horizon. They therefore estimate an individual market price of risk for each forecasting horizon. We use a more parsimonious function with regard to the number of parameters. The step function we propose is the simplest time-varying function that expects that the local long run risk premium differs in the short and the long horizon but is still constant in each period. The linear function implements the property that the local long run risk premium in the short horizon approaches the long term level linearly. The simplicity of these functions allows a straight forward calibration to interest rate forecasts.

Because of the distributional properties of the Gauss2++ model the expected values for interest rates under the real world measure $\mathbb{P}$ for any future time point can be calculated:

$$E^{\mathbb{P}}[r(t,T)] = E^{\mathbb{Q}}[r(t,T)] + \frac{B(a,t,T)}{T-t}RP_x(t) + \frac{B(b,t,T)}{T-t}RP_y(t), \tag{11}$$

where $RP_x(t)$ and $RP_y(t)$ represent the actual risk premium of the short-rate at time point t for each risk factor and are given by the first integral in (6) and (7)

$$RP_x(t) := \int_0^t e^{-a(t-u)} a d_x(u) du,$$
$$RP_y(t) := \int_0^t e^{-b(t-u)} b d_y(u) du.$$

For the constant, the step and the linear function these integrals can be easily calculated. To get the risk premium for longer maturities the functions $RP_x(t)$ and $RP_y(t)$ are weighted by a loading function, which accounts for the different riskiness of the corresponding zero-coupon bonds

$$\frac{B(a,t,T)}{T-t} \quad \text{and} \quad \frac{B(b,t,T)}{T-t}.$$

To calibrate the local long run risk premium functions, $d_x(t)$ and $d_y(t)$, the parameters of the functions are chosen in such a way that the model meets specific interest rate forecasts in expectation. For the constant type two interest rate forecasts are needed. For the other two types four interest rate forecasts are necessary—two short term and two long term forecasts. The time parameter τ, which determines the separation between the short and the long term local long run risk premium must lie between the forecasting horizons of the two short and the two long term forecasts.

In Fig. 1 the three types of local long run risk premium functions have been exemplary calibrated. τ has been set to 24 months, which is the forecasting horizon of the short term interest rate forecasts.

In the following subsections the calibration procedures for all three types of local long run risk premium functions, which are applied in this paper, are described.

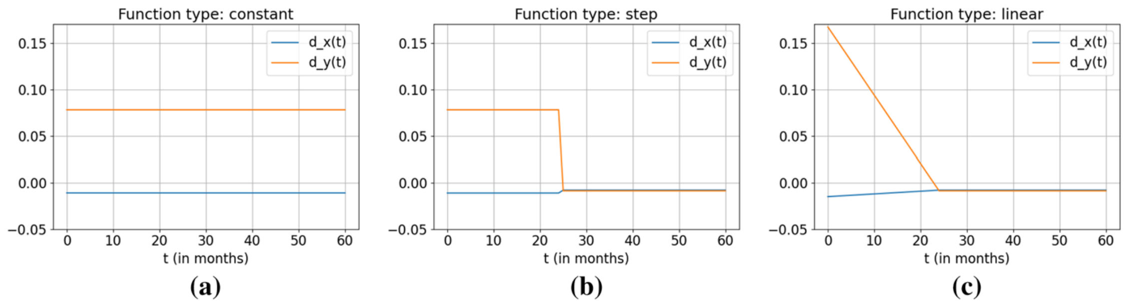


Fig. 1 Local long run risk premium functions

3.1 The constant function

The constant functions represented in Fig. 1a implement a constant local long run risk premium for the whole modelling horizon, which can amount to 40 years or more for actual applications in the insurance industry, e.g. to classify certified pension contracts into risk classes. The absolute risk premiums, $RP_x(t)$ and $RP_y(t)$, are given by:

$$RP_x(t) = (1 - e^{-at})d_x,$$
$$RP_y(t) = (1 - e^{-bt})d_y.$$

Note that if $t \to \infty$, $RP_x(t)$ and $RP_y(t)$ indeed converge to d_x and d_y, the long run risk premiums, respectively. To calibrate the parameters of the constant functions two interest rate forecasts, $\hat{r}(t_1, T_1)$ and $\hat{r}(t_2, T_2)$, are used. Plugging the absolute risk premium functions, $RP_x(t)$ and $RP_y(t)$, into (11) and setting the expectations equal to the interest rate forecasts results in the following two equations

$$\text{(I)}\ \ \hat{r}(t_1, T_1) \overset{!}{=} E^Q[r(t_1, T_1)] + \frac{B(a,t_1,T_1)}{(T_1-t_1)}(1 - e^{-at_1})d_x + \frac{B(b,t_1,T_1)}{(T_1-t_1)}(1 - e^{-bt_1})d_y,$$
$$\text{(II)}\ \ \hat{r}(t_2, T_2) \overset{!}{=} E^Q[r(t_2, T_2)] + \frac{B(a,t_2,T_2)}{(T_2-t_2)}(1 - e^{-at_2})d_x + \frac{B(b,t_2,T_2)}{(T_2-t_2)}(1 - e^{-bt_2})d_y.$$

As the expectations are linear functions in d_x and d_y, the two parameters can be easily determined.

The constant function for the local long run risk premium in the Gauss2++ model and this calibration procedure is a standard approach in the insurance industry. As the values for d_x and d_y determine the risk premium for the whole modelling horizon, their calibration is crucial for the model's interest rate distribution. Especially if the interest rate forecasts used for the calibration have a short forecasting horizon, the resulting distribution in the long horizon is very sensitive to these forecasts. For example if the interest rate forecasts and the forward rates—calculated from the current yield curve—are very different, to reach the forecasts a huge risk premium is necessary, which might be valid in the short horizon, but produces extreme interest rates in the long horizon. The next two functions account for this problem by representing a time-varying local long run risk premium.

3.2 The step function

The step functions represented in Fig. 1b take the same value as the corresponding constant function up to time τ as the same interest rate forecasts have been used for the short horizon, but then they jump to a different level to account for the risk premium in the long horizon. Similar to the constant function the absolute risk premium functions can easily be calculated and amount to

$$RP_x(t) = \left(e^{-a(t-\min(t,\tau))} - e^{-at}\right)d_x + \left(1 - e^{-a(t-\min(t,\tau))}\right)l_x,$$
$$RP_y(t) = \left(e^{-b(t-\min(t,\tau))} - e^{-bt}\right)d_y + \left(1 - e^{-b(t-\min(t,\tau))}\right)l_y.$$

Note that if $t \to \infty$, $RP_x(t)$ and $RP_y(t)$ now converge to l_x and l_y, respectively. To calibrate the four parameters of the step function two short term and two long term interest rate forecasts are used resulting in the following equations:

$$\begin{aligned}
&(\text{I}) \;\; \hat{r}(t_1,T_1) \stackrel{!}{=} E^Q[r(t_1,T_1)] + \tfrac{B(a,t_1,T_1)}{(T_1-t_1)}RP_x(t_1) + \tfrac{B(b,t_1,T_1)}{(T_1-t_1)}RP_y(t_1),\\
&(\text{II}) \;\; \hat{r}(t_2,T_2) \stackrel{!}{=} E^Q[r(t_2,T_2)] + \tfrac{B(a,t_2,T_2)}{(T_2-t_2)}RP_x(t_2) + \tfrac{B(b,t_2,T_2)}{(T_2-t_2)}RP_y(t_2),\\
&(\text{III}) \;\; \hat{r}(t_3,T_3) \stackrel{!}{=} E^Q[r(t_3,T_3)] + \tfrac{B(a,t_3,T_3)}{(T_3-t_3)}RP_x(t_3) + \tfrac{B(b,t_3,T_3)}{(T_3-t_3)}RP_y(t_3),\\
&(\text{IV}) \;\; \hat{r}(t_4,T_4) \stackrel{!}{=} E^Q[r(t_4,T_4)] + \tfrac{B(a,t_4,T_4)}{(T_4-t_4)}RP_x(t_4) + \tfrac{B(b,t_4,T_4)}{(T_4-t_4)}RP_y(t_4),
\end{aligned}$$

where $t_1 \le t_2 < t_3 \le t_4$. τ must lie between t_2 and t_3, i.e. $t_2 \le \tau < t_3$.

Instead of interest rate forecasts direct forecasts of the absolute risk premium of the short-rate can be used. This approach is applied by Hull et al. [13], who estimate risk premiums for each forecasting horizon from historical data, but they also scale their result to a long term short-rate forecast. Another possible approach is to take the ultimate forward rate (UFR) from Solvency II as a long term target, which is reached at a future time point with a specific percentage (e.g. 95% of the UFR in 40 years) and to 100% in the limit, i.e. $t \to \infty$.

3.3 The linear function

The linear functions represented in Fig. 1c avoid the sudden jump as it is the case in the step functions and converge in the short term linearly to a long term level. The absolute risk premiums at time point t can be calculated as before and amount to

$$\begin{aligned}
RP_x(t) = &\left(\left(e^{-a(t-\min(t,\tau))} - e^{-at}\right)\left(1 + \frac{m_x}{a}\right) - e^{-a(t-\min(t,\tau))}m_x \min(t,\tau)\right)d_x\\
&+ \left(1 - e^{-a(t-\min(t,\tau))}\right)l_x,\\
RP_y(t) = &\left(\left(e^{-b(t-\min(t,\tau))} - e^{-bt}\right)\left(1 + \frac{m_y}{b}\right) - e^{-b(t-\min(t,\tau))}m_y \min(t,\tau)\right)d_y\\
&+ \left(1 - e^{-b(t-\min(t,\tau))}\right)l_y.
\end{aligned}$$

Note again that if $t \to \infty$, $RP_x(t)$ and $RP_y(t)$ converge to l_x and l_y, the long term risk premiums, respectively. To calibrate d_x, l_x, d_y and l_y four interest rate forecasts as for the step function are used. By imposing that the absolute risk premium functions, $RP_x(t)$ and $RP_y(t)$, are differentiable at the forecasting horizon τ to prevent a kink in the absolute risk premium function, two further conditions are incorporated to specify m_x and m_y:

$$\text{(V)}\ \ (RP_x)'_-(\tau) = (RP_x)'_+(\tau),$$
$$\text{(VI)}\ \ (RP_y)'_-(\tau) = (RP_y)'_+(\tau),$$

where $(RP_\cdot)'_-(\tau)$ and $(RP_\cdot)'_+(\tau)$ denote the derivative from the left and from the right, respectively. Solving the equations for m_x and m_y leads to the following closed form solutions reducing the number of free parameters to four:

$$m_x = \frac{d_x - l_x}{d_x \tau},$$
$$m_y = \frac{d_y - l_y}{d_y \tau}.$$

Note that with this condition the same number of interest rate forecasts as for the step function are needed to calibrate $d_x(t)$ and $d_y(t)$.

4 Results

In this section the calibration results of three variants of our framework for the Gauss2++ model are presented. The variants differ in the assumption about the local long run risk premium functions, which determine the change from the risk neutral to the real world measure. Variant 1 assumes a constant, variant 2 a step and variant 3 a linear local long run risk premium function for the risk factors. In the first subsection the three variants of the Gauss2++ model are compared if calibrated at the same valuation date. In Sect. 4.2 we show with a backtest over the last three years that variant 2 and 3 produce much more stable interest rate scenarios for the long forecasting horizon over this time period. This stability would transfer to performance scenarios and risk measures of, e.g. an interest rate sensitive fonds.

4.1 Calibration at one valuation date

The calibration process of the Gauss2++ model can be split into two steps. In the first step the model is calibrated under the risk neutral measure. This step does not depend on the choice of the local long run risk premium function and is therefore the same for all modelling cases. In the second step the change of measure is calibrated. The choice of the local long run risk premium function plays an important role and leads to different interest rate scenarios, performance measures and risk indicators.

Table 1 Parameters of the Gauss2++ model calibrated at 31.12.2019

a	b	σ	η	ρ
0.2997	0.0407	0.0114	0.0114	– 0.9998

To calibrate the model at a specific valuation date under the risk neutral measure the term structure of interest rate swaps and swaption volatilities at this date are used. In the Gauss2++ model the deterministic and time-dependent function φ ensures the market consistency regarding the current term structure by being defined as follows:

$$\varphi(t) = f^M(0,t) + \frac{\sigma^2}{2a}(1-e^{-at})^2 + \frac{\eta^2}{2b}(1-e^{-bt})^2 + \rho\frac{\sigma\eta}{ab}(1-e^{-at})(1-e^{bt}).$$

$f^M(0,t)$ represents the instantaneous forward rate at time point 0 for a maturity t, i.e. $f^M(0,t) = \frac{\partial P^M(0,t)}{\partial T}$, where $\frac{\partial P^M}{\partial T}$ denotes the partial derivative with respect to the second argument and $P^M(0,t)$ is the market zero-coupon bond price. For the derivation and further information the reader is referred to Brigo and Mercurio [2]. The parameters a, b, σ, η and ρ of the model are chosen in such a way that the model prices of the considered swaptions coincide with the market prices. For this the downhill simplex algorithm[1] is used to minimize the root mean squared error ($RMSE$):

$$RMSE = \sqrt{\sum_{i=1}^{N}\left(C_{Model,i}(a,b,\sigma,\eta,\rho) - C_{Market,i}\right)^2},$$

where $C_{Model,i}$ represents the model price of swaption i of the Gauss2++ model and $C_{Market,i}$ is the market price of that swaption. The swaptions considered in the calibration process differ with respect to their tenor and maturity combination, which is denoted by the subscript i. N represents the number of considered swaptions. The analytic formula for the price of a swaption in the Gauss2++ model can be found in Brigo and Mercurio [2]. Table 1 shows the result of a calibration at the 31.12.2019. We used at-the-money receiver swaptions with a maturity and tenor combination of $\{5,7,10,12,15,20\} \times \{5,7,10,12,15,20\}$, i.e. in total $N = 36$ swaption prices. The $RMSE$ amounts to 0.0619. In the optimization we further restricted ρ to lie between – 1 and 1 as well as all other parameters to be > 0.

These parameters together with the current interest rate curve determine the dynamics of the Gauss2++ model under the risk neutral measure.

In the second step the local long run risk premium functions, which determine the change of measure, are calibrated to interest rate forecasts as described in Sects. 3.1–3.3. For the short term interest rate forecasts we use forecasts published by the OECD for a 3-month and a 10-year interest rate. To take the OECD forecasts

[1] For a detailed description of this algorithm—also known as the Nelder–Mead algorithm—the reader is referred to Nelder and Mead [16]. For the reflection coefficient, the expansion coefficient and the contraction coefficient of the algorithm we have chosen the values 1.0, 2.0 and 0.5, respectively.

Table 2 Parameters of the local long run risk premium functions

	d_x	d_y	l_x	l_y
Constant function	– 0.0112	0.0779		
Step function	– 0.0112	0.0779	– 0.0081	– 0.0088
Linear function	– 0.0151	0.1672	– 0.0081	– 0.0088

we have been inspired by the framework developed by the Frauenhofer ITWM on behalf of PIA to classify PRIIPs into chance-risk classes (see Korn and Wagner [15]). Their model represents the industry standard for PRIIP calculations. The latest OECD forecasts regarding the 31.12.2019 for the longest horizon, which is the fourth quarter of 2021, amount to – 0.4% and 0.4%, respectively.[2] For the long term interest rate forecasts, which are needed to calibrate the step and the linear function, we take the average of monthly 3-month and 10-year interest rates over the last 15 years also published by the OECD. This is a valid approach if interest rates follow a stationary process, because in this case historical data can be considered as a random sample from the corresponding interest rate distribution. Hull et al. [13] point out that this approach is questionable if monetary and fiscal policies are expected to be materially different from those in the past. Nevertheless any other model based on historical data would be questionable and the user of the model can alternatively provide personal estimates or an expert judgment. The historical average amounts to 1.08% for the 3-month and 1.84% for the 10-year interest rate and as we assume these forecasts to be a long run average we set the forecasting horizon to 40 years—the modelling horizon. We further set τ to 24 months, which is the forecasting horizon of the short term OECD forecasts.

Table 2 shows the calibration results for the three local long run risk premium function types.

The values of d_x and d_y coincide for the constant and the step function as the same interest rate forecasts have been used in the calibration process. But in contrast to the step function, which takes the values of l_x and l_y after 24 months, the constant function stays constant for the whole modelling horizon. It also appears that the step and the linear function take the same values for l_x and l_y. But there is a slight difference as their functional forms differ in the first two years, which influences the absolute risk premium in future time points. This influence decreases in time, such that the difference is negligible as we calibrated l_x and l_y to forecasts with an forecasting horizon of 40 years.

Figures 2, 3 and 4 visualize for the three calibrated variants of the Gauss2++ model the development of the expectation of the short-rate, the 10-year and the 20-year interest rate for forecasting horizons of up to 40 years. The solid line represents the expectation under the risk neutral measure, the dashed line shows the expected values under the real world measure.

[2] https://stats.oecd.org: The rounded numbers can be found, if one selects the data for the Economic Outlook N.106 of November 2019 in the section Economic Projections.

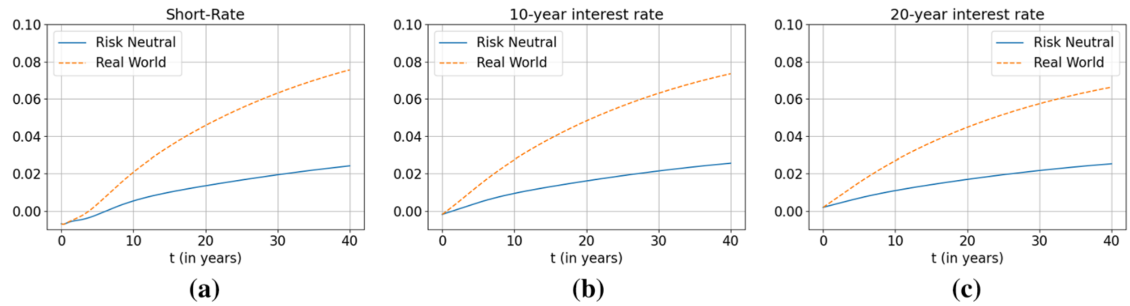


Fig. 2 Constant function: expected values of the short-rate, the 10-year and the 20-year interest rate under the risk neutral and the real world measure

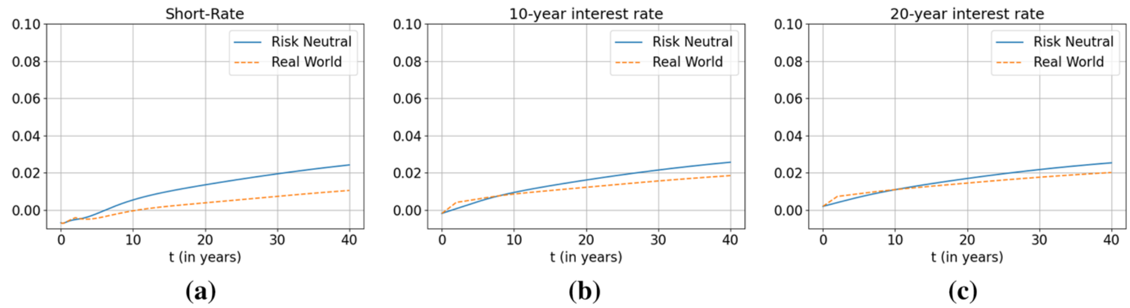


Fig. 3 Step function: expected values of the short-rate, the 10-year and the 20-year interest rate under the risk neutral and the real world measure

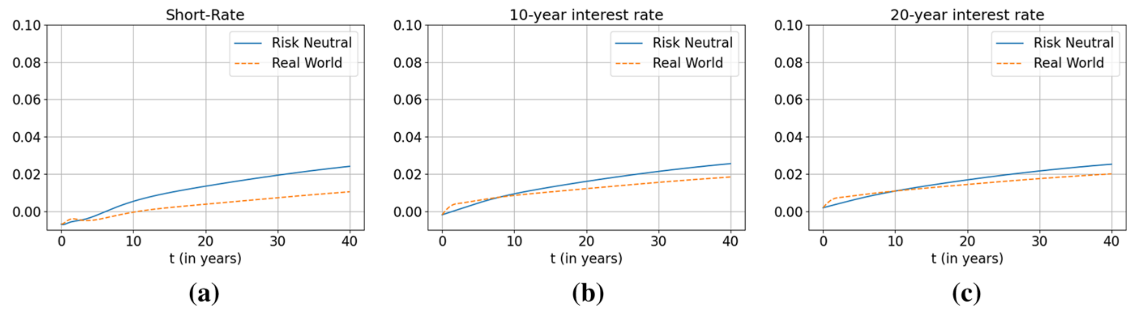


Fig. 4 Linear function: expected values of the short-rate, the 10-year and the 20-year interest rate under the risk neutral and the real world measure

For the variant of the Gauss2++ model, which uses the constant function as the local long run risk premium function, the expected real world interest rates lie above the risk neutral expectation. This means, that a risk seeking behaviour of the investors is assumed for the whole modelling period, because an investor accepts a lower expected return for a corresponding bond if the interest rates are expected to be higher in the real world compared to the risk neutral world. Ahmad and Wilmott [1] show that there have been time periods where investors seem to have historically behaved in this way. But in general investors are assumed to be risk averse and therefore interest rates should be lower in the real world than in the risk neutral world, which is an opposite behaviour to equity prices (see, e.g. Hull et al. [13]). For the other two variants of the Gauss2++ model the expected

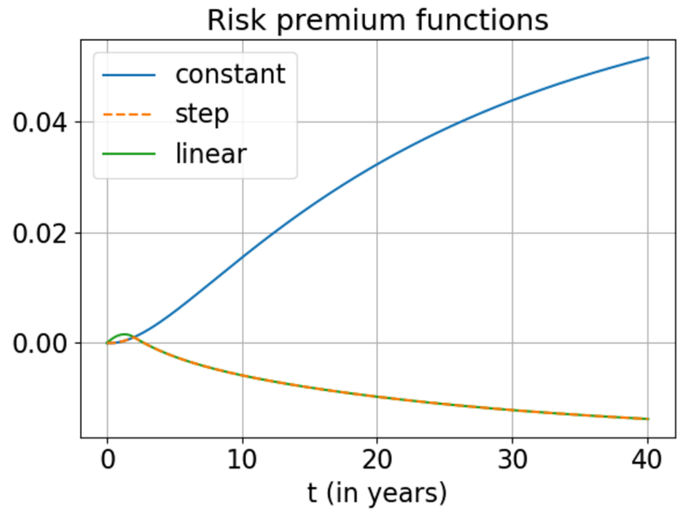


Fig. 5 Absolute risk premium function for the variants of the Gauss2++ model

real world interest rates lie also above the risk neutral interest rates in the short horizon but below in the long horizon. This assumption of risk seeking behaviour in the short horizon stems from the quite high forecasts of the OECD for the short horizon, but it might be valid in the current market situation. In contrast to the constant case, which keeps this risk seeking behaviour assumption for the whole modelling horizon, in the long run the other two variants of the Gauss2++ model assume in this calibration a risk averse behaviour. The difference between the step and the linear function is only visible in the short horizon. While the step function has a kink in the expectation after τ years, the linear function is smoother due to its condition that the derivative of the absolute risk premium function exists at this time point.

Furthermore, the absolute difference in the risk neutral and real world expectations decreases for interest rates with longer maturities. This results from the less variation of interest rates with longer maturities, which is an implicit model characteristic of the Gauss2++ model and is supported by historical data as well. A risk premium is therefore higher (less negative) for a risk averse and lower (less positive) for a risk seeking investor in an arbitrage free market.

Figure 5 shows the absolute risk premium functions of the short-rate for all three modelling types.

It can be observed that for the constant and the step function the absolute risk premium is the same up to year 2. After that year the Gauss2++ variant with the step function has a kink in the absolute risk premium as the local long run risk premium changes to a different level, while the modelling case with the constant function continuous to approach the long term risk premium determined by the short term interest rate forecasts. The modelling case with the linear function results in a different risk premium for the first 2 years, but approaches—without a kink—the same long term risk premium as the step function.

All three functions intersect after 2 years as this is the forecasting horizon of the short term interest rate forecasts, which were used for the calibration. The absolute risk premium at this time point must be the same for all modelling cases such that the expected interested rates of the model coincide with the forecasts.

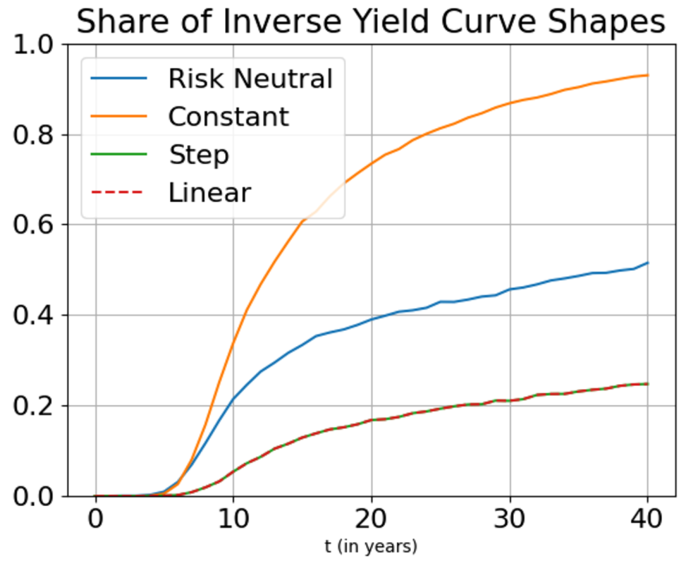


Fig. 6 The share of inverse yield curves for the Gauss2++ model under the risk neutral measure and under the real world measure using a constant, a step and a linear function for the market price of risk

We further investigated the resulting yield curve shapes of the three variants of the Gauss2++ model. The variant, which uses a constant function, represents the industry standard regarding PRIIP calculations (see Korn and Wagner [15]). An unpleasent feature of this model is the unrealistic high frequency of inverse yield curves with growing time (see Diez and Korn [7]). In their paper they show that for the 2-Factor Vasicek model the number of inverse yield curves can be reduced by assuming a negative risk premium. The share of inverse yield curves in our calibration of the three variants were investigated in a simulation study. We simulated 10,000 yield curve paths with each calibrated model and counted the number of yield curves, which have a higher 1-year interest rate than a 30-year interest rate. The result is visualized in Fig. 6. We can see a similar behaviour as described in the paper of Diez and Korn [7]. The variant with the constant function, which has a positive risk premium over the modelling horizon, shows an unrealistic high share of inverse yield curves. The other two variants have a negative risk premium and decrease the number of inverse yield curves in the long run compared to the risk neutral case. Using the step or the linear function for the risk premium results therefore not only in more realistic interest rates but also in more realistic yield curve shapes in the long horizon.

4.2 Backtest

In this subsection the different variants of the Gauss2++ model calibrated on a quarterly basis over the last 3 years are compared.

As in Sect. 4.1 interest rate swaps and swaption volatilities have been used for the risk neutral calibration of the Gauss2++ model. To calibrate the parameters of the local long run risk premium functions in the second calibration step short term interest rate forecasts published by the OECD and a long term average have been used. The forecasts are shown in Table 3. The calibration results of the parameters of the Gauss2++ model under the risk neutral measure and of the local long run risk

Table 3 Interest rate forecasts of the OECD and historical average of the 3-month and the 10-year interest rate

Date	Short term interest rate forecasts			Historical average	
	Forecasting horizon (in months)	3-m IR (in %)	10-y IR (in %)	3-m IR (in %)	10-y IR (in %)
30.09.2019	15	−0.3	1.0	1.13	1.91
30.06.2019	18	−0.3	1.0	1.18	1.98
31.03.2019	21	−0.2	1.6	1.22	2.04
31.12.2018	24	−0.2	1.6	1.26	2.10
30.09.2018	15	−0.2	1.3	1.31	2.16
30.06.2018	18	−0.2	1.3	1.35	2.23
31.03.2018	21	−0.3	1.4	1.39	2.30
31.12.2017	24	−0.3	1.4	1.44	2.36
30.09.2017	15	−0.3	1.6	1.48	2.43
30.06.2017	18	−0.3	1.6	1.52	2.50
31.03.2017	21	−0.3	1.6	1.57	2.56
31.12.2016	24	−0.3	1.6	1.63	2.63

https://stats.oecd.org: The rounded numbers can be found, if one selects the annual interest rate forecasts of the corresponding Economic Outlook in the section Economic Projections

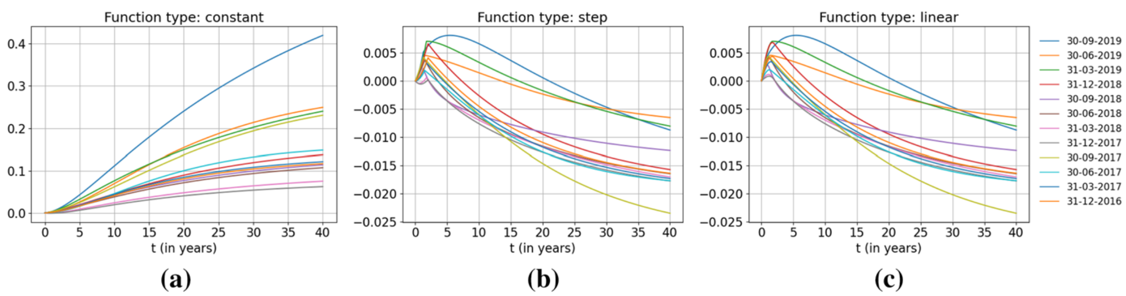


Fig. 7 Absolute risk premium functions

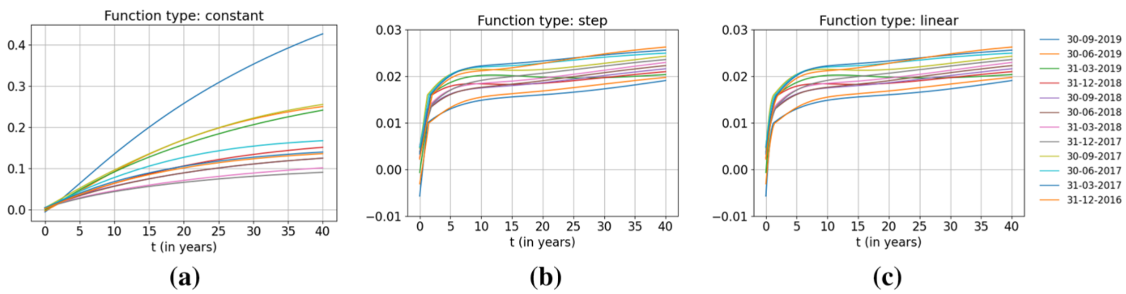


Fig. 8 Development of the expectation of the 10-year interest rate over the modelling horizon for all three variants of the Gauss2++ model

premium function for each variant of the Gauss2++ model can be found in Tables 4, 5, 6 and 7 in “Appendix 6”.

For each calibration the absolut risk premium function of the short-rate and the development of the expected 10-year interest rate have been calculated and visualised in Figs. 7 and 8.

The absolute risk premium function of the short-rate for the Gauss2++ model, which uses the constant function for the local long run risk premium, depends highly on the risk neutral calibration results and the forecasts of the OECD. An unfavorable combination of market data and interest rate forecasts can lead to a high value for the local long run risk premium. This value might be reasonable to meet the short term forecasts used for the calibration, but as it stays constant over time it is the value the absolute risk premium is converging to. Therefore, this problem can strike through if the modelling horizon is much longer than the forecasting horizon of the interest rates used for the calibration. In this case a time-varying local long run risk premium function, which can be calibrated to a short and a long term forecast, is more convenient to regularize the risk premium. As it can be seen in Fig. 7 the variants of the Gauss2++ model, which use the step or the linear function for the local long run risk premium, produce more stable risk premiums in the long horizon. In each calibration the absolute risk premium is positive in the first years, which presumes a risk seeking behaviour of the investors, but in the long horizon the absolute risk premium lies between -0.5 and -2.5% representing a risk averse market. Also the interest rate distribution in the long horizon is more stable. Figure 8b, c show that the expectation of the 10-year interest rate in the long horizon change only little in each calibration according to the historical average, which was used for the long term interest rate forecast.

5 Conclusion

As the Gauss2++ model is often used for pricing purposes, the focus in the literature lies on the evolution of interest rates under the risk neutral measure $\mathbb{Q}$. But regarding risk management and forecasting applications the model under the real world measure is needed. In this paper we introduced a framework to apply the model under both measures in a consistent manner. This framework first conducts a calibration under the risk neutral measure and then determines the change of measure such that it is possible to switch between the risk neutral and the real world. We showed that according to Girsanov this change of measure can be specified by any time-dependent function without loosing the analytic tractability of, e.g. zero-coupon bond prices. Hull et al. [13] argue that because of unobserved risk factors, which are not included in the model, a time-varying function should be used, because otherwise unrealistic interest rates in the long forecasting horizon could be reached. We therefore compared the industry standard, which uses a constant function to model the change of measure, with two variants, which use either a step or a linear function. These functions are the simplest extensions of the constant function to a time-varying function without increasing the computational effort much. By accounting for different risk premiums in the short and in the long horizon the time-varying functions result in much more stable interest rate forecasts in the long run if calibrated at different valuation dates. From a macroeconomical point of view it makes sense that current market fluctuations should not influence interest rate forecasts in the long horizon, e.g. in 40 years, much. This would also imply that risk measures calculated with the Gauss2++ model, which uses one of the time-varying functions for the change of measure, would be more consistent if estimated at different valuation time points.

We further investigated the yield curve shapes by conducting a simulation study. The result is in line with the findings of Diez and Korn [7] for the 2-Factor Vasicek model. Assuming a positive risk premium—as it was the case in our calibration for the constant function—the number of inverse yield curves increases compared to the risk neutral case. This also replicates the problem of too many inverse yield curves in the insurance industry for PRIIP calculations (see Diez and Korn [7]). The other two variants represented in this paper, which apply a time-varying function for the market price of risk, assume a negative risk premium in the long run and have a much lower amount of inverse yield curves. Using a step or a linear function for the market price of risk, therefore, not only leads to more realistic interest rates in the long run, but also creates more realistic yield curve shapes.

Appendix 1: Bond price dynamic under the risk neutral measure

By defining

$$A(t,T) = -\int_t^T \varphi(s)ds + \frac{1}{2}V(t,T),$$

the price of a zero-coupon bond $P(t, T)$ at time point t and maturity T can be calculated for the Gauss2++ model under the risk neutral measure $\mathbb{Q}$ by

$$P(t,T) = e^{A(t,T)-B(a,t,T)x(t)-B(b,t,T)y(t)}. \tag{12}$$

A proof of this formula can be found in Brigo and Mercurio [2]. The derivatives of $A(t, T)$ and $V(t, T)$ with respect to the first entry and of $B(z, t, T)$ with respect to the second entry are given by

$$\begin{aligned}
A'(t,T) &= \varphi(t) + \frac{1}{2}V'(t,T),\\
V'(t,T) &= -\sigma^2 B(a,t,T)^2 - \eta^2 B(b,t,T)^2 - 2\sigma\eta\rho B(a,t,T)B(b,t,T),\\
B'(z,t,T) &= -e^{-z(T-t)}.
\end{aligned}$$

Furthermore, it holds

$$B(z,t,T)z - B'(z,t,T) = 1.$$

To calculate the zero-coupon bond price dynamic, we apply Ito's formula to (12), i.e.

$$
\begin{aligned}
dP(t,T) &= P(t,T)\big[A(t,T) - B(a,t,T)x(t) - B(b,t,T)y(t)\big]' dt + P(t,T)(-B(a,t,T))dx(t) \\
&\quad + P(t,T)(-B(b,t,T))dy(t) \\
&\quad + \frac{1}{2}P(t,T)B(a,t,T)^2\sigma^2 dt \\
&\quad + \frac{1}{2}P(t,T)B(b,t,T)^2\eta^2 dt \\
&\quad + P(t,T)B(a,t,T)B(b,t,T)\sigma\eta\rho dt \\
&= P(t,T)\Bigg[A'(t,T) - B'(a,t,T)x(t) - B'(b,t,T)y(t) + B(a,t,T)ax(t) + B(b,t,T)by(t) \\
&\quad + \frac{1}{2}B(a,t,T)^2\sigma^2 + \frac{1}{2}B(b,t,T)^2\eta^2 \\
&\quad + B(a,t,T)B(b,t,T)\sigma\eta\rho\Bigg] dt \\
&\quad - B(a,t,T)P(t,T)\sigma dW^1(t) \\
&\quad - B(b,t,T)P(t,T)\eta dW^2(t) \\
&= P(t,T)[\varphi(t) + x(t) + y(t)]dt - B(a,t,T)P(t,T)\sigma dW^1(t) - B(b,t,T)P(t,T)\eta dW^2(t) \\
&= P(t,T)r(t)dt - B(a,t,T)P(t,T)\sigma dW^1(t) - B(b,t,T)P(t,T)\eta dW^2(t).
\end{aligned}
$$

Appendix 2: The dynamics of the Gauss2++ factors *x* and *y* under the real world measure

The dynamics of the two processes x and y under the risk neutral measure $\mathbb{Q}$ can be expressed in terms of two independent Brownian motions $\widehat{W}^1$ and $\widehat{W}^2$, i.e.

$$
\begin{aligned}
dx(t) &= -ax(t)dt + \sigma d\widehat{W}^1(t), \\
dy(t) &= -by(t)dt + \eta\rho d\widehat{W}^1(t) + \eta\sqrt{(1-\rho^2)}d\widehat{W}^2(t),
\end{aligned}
$$

where

$$
\begin{aligned}
dW^1(t) &= d\widehat{W}^1(t), \\
dW^2(t) &= \rho d\widehat{W}^1(t) + \sqrt{(1-\rho^2)}d\widehat{W}^2(t).
\end{aligned}
$$

According to Girsanov's theorem , as $\widehat{\boldsymbol{W}} = (\widehat{W}^1, \widehat{W}^2)$ is a standard 2-dimensional Brownian motion and let $(\boldsymbol{\Phi}(t))_{t\in[0,T]} = (\Phi^1(t), \Phi^2(t))_{t\in[0,T]}$ be a progressive and square-integrable process, the process $\breve{\boldsymbol{W}}$ defined by

$$
\breve{\boldsymbol{W}}(t) := \widehat{\boldsymbol{W}}(t) + \int_0^t \boldsymbol{\Phi}(s)ds
$$

is a standard 2-dimensional Brownian motion under a new measure, which we call $\mathbb{P}$ and declare to be the real world measure. This means that the dynamic of the two Brownian motion $\widehat{W}^1$ and $\widehat{W}^2$ under the real world measure $\mathbb{P}$ is given by

$$d\widehat{W}^1(t) = d\breve{W}^1(t) - \Phi^1(t)dt,$$
$$d\widehat{W}^2(t) = d\breve{W}^2(t) - \Phi^2(t)dt.$$

Therefore, the dynamics of the two processes x and y under the real world measure are then given by

$$\begin{aligned} dx(t) &= \Big[- \Phi^1(t)\sigma - ax(t) \Big] dt + \sigma d\breve{W}^1(t), \\ dy(t) &= \Big[- \Phi^1(t)\eta\rho - \Phi^2(t)\eta\sqrt{(1-\rho^2)} - by(t) \Big] dt + \eta\rho d\breve{W}^1(t) \\ &\quad + \eta\sqrt{(1-\rho^2)} d\breve{W}^2(t). \end{aligned}$$

If we specify $\Phi(t)$ as in (3) this simplifies to

$$\begin{aligned} dx(t) &= a(d_x(t) - x(t))dt + \sigma d\breve{W}^1(t), \\ dy(t) &= b(d_y(t) - y(t))dt + \eta\rho d\breve{W}^1(t) + \eta\sqrt{(1-\rho^2)} d\breve{W}^2(t). \end{aligned}$$

Representing the dynamics by two correlated Brownian motions $\widetilde{W}^1$ and $\widetilde{W}^2$ results in the equations given in (4) and (5).

Appendix 3: Bond price dynamic under the real world measure

The dynamic of a zero-coupon bond price $P(t, T)$ under the risk neutral measure $\mathbb{Q}$ expressed by the two independent Brownian motions $\widehat{W}^1$ and $\widehat{W}^2$ is given by

$$\begin{aligned} dP(t,T) &= P(t,T)r(t)dt - P(t,T)B_\tau(a)\sigma d\widehat{W}^1(t) - P(t,T)B_\tau(b)\eta\rho d\widehat{W}^1(t) \\ &\quad - P(t,T)B_\tau(b)\eta\sqrt{(1-\rho^2)} d\widehat{W}^2(t), \\ &= P(t,T)r(t)dt - \Big[P(t,T)B_\tau(a)\sigma + P(t,T)B_\tau(b)\eta\rho \Big] d\widehat{W}^1(t) \\ &\quad - P(t,T)B_\tau(b)\eta\sqrt{(1-\rho^2)} d\widehat{W}^2(t). \end{aligned}$$

Applying Girsanov's theorem as in "Appendix 2" the dynamic under the real world measure $\mathbb{P}$ amounts to

$$
\begin{aligned}
dP(t,T) =\,& P(t,T)r(t)dt - \Big[P(t,T)B_\tau(a)\sigma + P(t,T)B_\tau(b)\eta\rho\Big] d\widehat{W}^1(t) \\
& - P(t,T)B_\tau(b)\eta\sqrt{(1-\rho^2)}d\widehat{W}^2(t) \\
=\,& P(t,T)\Bigg[r(t) + \Big(B_\tau(a)\sigma + B_\tau(b)\eta\rho\Big)\left(-\frac{ad_x(t)}{\sigma}\right) \\
& + B_\tau(b)\eta\sqrt{(1-\rho^2)}\left(-\frac{bd_y(t)}{\eta\sqrt{(1-\rho^2)}} + \frac{\rho a d_x(t)}{\sigma\sqrt{(1-\rho^2)}}\right)\Bigg]dt \\
& - \Big[P(t,T)B_\tau(a)\sigma + P(t,T)B_\tau(b)\eta\rho\Big] d\breve{W}^1(t) \\
& - P(t,T)B_\tau(b)\eta\sqrt{(1-\rho^2)}d\breve{W}^2(t) \\
=\,& P(t,T)\Big[r(t) - B_\tau(a)ad_x(t) - B_\tau(b)bd_y(t)\Big]dt \\
& - \Big[P(t,T)B_\tau(a)\sigma + P(t,T)B_\tau(b)\eta\rho\Big] d\breve{W}^1(t) \\
& - P(t,T)B_\tau(b)\eta\sqrt{(1-\rho^2)}d\breve{W}^2(t).
\end{aligned}
$$

Representing the dynamic by two correlated Brownian motions $\widetilde{W}^1$ and $\widetilde{W}^2$ results in the equation given in (8).

Appendix 4: Individual discount rate for the zero-coupon bonds in the real world

Proof To proof that $\frac{P(t,T)}{X(t,T)}$ is indeed a martingale we calculate the dynamic of the discounted price process.

$$
\begin{aligned}
d\frac{P(t,T)}{X(t)} &= d\Big(\frac{1}{X(t)} \cdot P(t,T)\Big) \\
&= \frac{1}{X(t)} dP(t,T) + P(t,T) d\frac{1}{X(t)} + d\left\langle P(t,T), \frac{1}{X(t)} \right\rangle \\
&= \frac{1}{X(t)} dP(t,T) - \frac{P(t,T)}{X(t)} \big[r(t) - B(a,t,T)ad_x(t) - B(b,t,T)bd_y(t)\big] dt \\
&= \frac{P(t,T)}{X(t)} \big[r(t) - B(a,t,T)ad_x(t) - B(b,t,T)bd_y(t)\big] dt \\
&\quad - \frac{P(t,T)}{X(t)} B(a,t,T)\sigma d\widetilde{W}^1(t) - \frac{P(t,T)}{X(t)} B(b,t,T)\eta d\widetilde{W}^2(t) \\
&\quad - \frac{P(t,T)}{X(t)} \big[r(t) - B(a,t,T)ad_x(t) - B(b,t,T)bd_y(t)\big] dt \\
&= -\frac{P(t,T)}{X(t)} B(a,t,T)\sigma d\widetilde{W}^1(t) - \frac{P(t,T)}{X(t)} B(b,t,T)\eta d\widetilde{W}^2(t)
\end{aligned}
$$

Appendix 5: Bond price formula under the real world measure

To calculate the price of a zero-coupon bond under the real world measure $\mathbb{P}$, the distribution of

$$
exp\left(- \int_t^T \big(r(u) - B(a,u,T)ad_x(u) - B(b,u,T)bd_y(u) \big) du \right)
$$

has to be determined. In the following we show, that the integral in the exponent is normaly distributed and calculate the mean and the variance of

$$
I(t,T) := \int_t^T \big(r(u) - B(a,u,T)ad_x(u) - B(b,u,T)bd_y(u) \big) du. \tag{13}
$$

We first concentrate on the integral over the short-rate $r(s)$, which is a sum of the x- and the y-process and a deterministic function

$$
r(s) = x(s) + y(s) + \varphi(s).
$$

The integral over the process x is given by

$$\begin{aligned}\int_t^T x(u)du &= \int_t^T \left(x(t)e^{-a(u-t)} + \int_t^u ae^{-a(u-s)}d_x(s)ds \right.\\ &\quad \left. + \int_t^u \sigma e^{-a(u-s)}d\widetilde{W}^1(s) \right) du\\ &= \underbrace{\int_t^T x(t)e^{-a(u-t)}du}_{①} + \underbrace{\int_t^T \int_t^u ae^{-a(u-s)}d_x(s)dsdu}_{②}\\ &\quad + \underbrace{\int_t^T \int_t^u \sigma e^{-a(u-s)}d\widetilde{W}^1(s)du.}_{③}\end{aligned}$$

The first integral amounts to

$$① = x(t)\int_t^T e^{-a(u-t)}du = x(t)\Big[-\frac{1}{a}e^{-a(u-t)}\Big]_t^T = x(t)\frac{1-e^{-a(T-t)}}{a}.$$

For the second integral we use the integration by parts formula

$$\begin{aligned}② &= \int_t^T \left(\int_t^u e^{as}d_x(s)ds\right) ae^{-au}du\\ &= a\int_t^T \left(\int_t^u e^{as}d_x(s)ds\right) d_u\left(\int_t^u e^{-av}dv\right)\\ &= a\left[\left(\int_t^T e^{au}d_x(u)du\right)\left(\int_t^T e^{-av}dv\right) - \int_t^T \left(\int_t^u e^{-av}dv\right)e^{au}d_x(u)du\right]\\ &= a\left[\int_t^T \left(\int_u^T e^{-av}dv\right)e^{au}d_x(u)du\right]\\ &= \int_t^T \left(1-e^{-a(T-u)}\right)d_x(u)du\\ &= \int_t^T aB(a,u,T)d_x(u)du.\end{aligned}$$

For the third integral we again use the integration by parts formula

$$
\begin{aligned}
③ &= \sigma \int_t^T \left(\int_t^u e^{as} d\widetilde{W}^1(s) \right) a e^{-au} du \\
&= \sigma \int_t^T \left(\int_t^u e^{as} d\widetilde{W}^1(s) \right) d_u \left(\int_t^u e^{-av} dv \right) \\
&= \sigma \left[\left(\int_t^T e^{au} d\widetilde{W}^1(u) \right) \left(\int_t^T e^{-av} dv \right) - \int_t^T \left(\int_t^u e^{-av} dv \right) e^{au} d\widetilde{W}^1(u) \right] \\
&= \sigma \left[\int_t^T \left(\int_u^T e^{-av} dv \right) e^{au} d\widetilde{W}^1(u) \right] \\
&= \sigma \int_t^T \left[-\frac{e^{-av}}{a} \right]_u^T e^{au} d\widetilde{W}^1(u) \\
&= \frac{\sigma}{a} \int_t^T \left(1 - e^{-a(T-u)} \right) d\widetilde{W}^1(u)
\end{aligned}
$$

The corresponding expressions for $\int_t^T y(u)du$ can be obtained analogously. We observe that the results of integral ② for $\int_t^T x(u)du$ and $\int_t^T y(u)du$ cancel out with the last two terms in equation (13). Therefore it remains

$$
\begin{aligned}
I(t,T) = &\int_t^T \varphi(u)du + \frac{1-e^{-a(T-t)}}{a} x(t) + \frac{1-e^{-b(T-t)}}{b} y(t) \\
&+ \frac{\sigma}{a} \int_t^T \left(1 - e^{-a(T-u)}\right) d\widetilde{W}^1(u) + \frac{\eta}{b} \int_t^T \left(1 - e^{-b(T-u)}\right) d\widetilde{W}^2(u).
\end{aligned}
$$

As $\widetilde{\boldsymbol{W}} = (\widetilde{W}^1, \widetilde{W}^2)$ is a 2-dimensional Brownian motion under $\mathbb{P}$, $I(t, T)$ is normally distributed and the mean and the variance can be easily retrieved resulting in (9) and (10).

Appendix 6: Tables of backtest results

See Tables 4, 5, 6 and 7.

Table 4 Calibration results of the risk neutral calibration on a quarterly basis from 31.12.2016 to 30.09.2019

Date	a	b	σ	η	ρ
30.09.2019	0.2694	0.0269	0.0121	0.0089	− 0.8950
30.06.2019	0.1216	0.0628	0.0363	0.0283	− 0.9687
31.03.2019	0.3978	0.0331	0.0333	0.0091	− 0.8576
31.12.2018	0.1628	0.0521	0.0183	0.0154	− 0.8629
30.09.2018	0.6100	0.0429	0.0459	0.0104	− 0.8722
30.06.2018	0.2901	0.0459	0.0104	0.0112	− 0.9941
31.03.2018	0.5120	0.0386	0.0142	0.0097	− 1.0000
31.12.2017	0.3803	0.0471	0.0236	0.0120	− 0.8854
30.09.2017	0.0880	0.0655	0.0421	0.0460	− 0.9938
30.06.2017	0.1260	0.0890	0.0504	0.0517	− 0.9963
31.03.2017	0.2940	0.0581	0.0152	0.0146	− 0.9984
31.12.2016	0.2427	0.0606	0.0178	0.0173	− 1.0000

Table 5 Quarterly calibration results for the constant local long run risk premium functions from 31.12.2016 to 30.09.2019

Date	d_x	d_y
30.09.2019	− 0.0676	0.7400
30.06.2019	− 0.2848	0.5787
31.03.2019	− 0.0267	0.3636
31.12.2018	− 0.0539	0.2182
30.09.2018	− 0.0107	0.1518
30.06.2018	− 0.0173	0.1481
31.03.2018	− 0.0112	0.1099
31.12.2017	− 0.0150	0.0913
30.09.2017	− 0.7023	0.9836
30.06.2017	− 0.3883	0.5497
31.03.2017	− 0.0330	0.1710
31.12.2016	− 0.0405	0.1725

Table 6 Quarterly calibration results for the step local long run risk premium functions from 31.12.2016 to 30.09.2019

Date	d_x	d_y	l_x	l_y
30.09.2019	− 0.0676	0.7400	− 0.0090	− 0.0129
30.06.2019	− 0.2848	0.5787	− 0.0376	− 0.0292
31.03.2019	− 0.0267	0.3636	− 0.0114	− 0.0034
31.12.2018	− 0.0539	0.2182	− 0.0163	− 0.0029
30.09.2018	− 0.0107	0.1518	− 0.0101	− 0.0047
30.06.2018	− 0.0173	0.1481	− 0.0107	− 0.0090
31.03.2018	− 0.0112	0.1099	− 0.0087	− 0.0129
31.12.2017	− 0.0150	0.0913	− 0.0099	− 0.0111
30.09.2017	− 0.7023	0.9836	− 0.0364	− 0.0087
30.06.2017	− 0.3883	0.5497	− 0.0423	− 0.0233
31.03.2017	− 0.0330	0.1710	− 0.0131	− 0.0068
31.12.2016	− 0.0405	0.1725	− 0.0154	− 0.0033

Table 7 Quarterly calibration results for the linear local long run risk premium functions from 31.12.2016 to 30.09.2019

Date	d_x	d_y	l_x	l_y
30.09.0219	− 0.1332	1.5015	− 0.0090	− 0.0129
30.06.2019	− 0.5474	1.1457	− 0.0376	− 0.0292
31.03.2019	− 0.0461	0.7377	− 0.0114	− 0.0034
31.12.2018	− 0.0959	0.4471	− 0.0163	− 0.0029
30.09.2018	− 0.0114	0.3111	− 0.0101	− 0.0047
30.06.2018	− 0.0250	0.3087	− 0.0107	− 0.0090
31.03.2018	− 0.0144	0.2355	− 0.0087	− 0.0129
31.12.2017	− 0.0216	0.1970	− 0.0099	− 0.0111
30.09.2017	− 1.3930	1.9854	− 0.0364	− 0.0087
30.06.2017	− 0.7567	1.1001	− 0.0423	− 0.0233
31.03.2017	− 0.0567	0.3550	− 0.0131	− 0.0068
31.12.2016	− 0.0700	0.3556	− 0.0154	− 0.0033

Acknowledgements This research was supported by ROKOCO predictive analytics GmbH. We also thank the two anonymous reviewers for helpful comments and suggestions.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Ahmad R, Wilmott P (2006) The market price of interest-rate risk: measuring and modelling fear and greed in the fixed-income markets. Wilmott Mag 30:64–70
2. Brigo D, Mercurio F (2007) Interest rate models—theory and practice: with smile, inflation and credit. Springer, Berlin
3. Cox JC, Ingersoll JE Jr, Ross SA (1985) An intertemporal general equilibrium model of asset prices. Econom J Econom Soc 30:363–384
4. Cox SH, Pedersen HW (1999) Nonparameteric estimation of interest rate term structure and insurance applications. In: Proceedings of the 1999 ASTIN Colloquium, Tokyo, Japan **(to appear)**
5. Dai Q, Singleton Kenneth J (2000) Specification analysis of affine term structure models. J Financ 55(5):1943–1978
6. Diebold FX, Li C (2006) Forecasting the term structure of government bond yields. J Econom 130(2):337–364
7. Diez F, Korn R (2019) Yield curve shapes of vasicek interest rate models, measure transformations and an application for the simulation of pension products. Eur Actuar J 20:1–30
8. Duffee GR (2002) Term premia and interest rate forecasts in affine models. J Financ 57(1):405–443
9. Duffie D, Kan R (1996) A yield-factor model of interest rates. Math Financ 6(4):379–406
10. Girsanov IV (1960) On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory Probab Appl 5(3):285–301
11. Heath D, Jarrow R, Morton A (1992) Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econom J Econom Soc 20:77–105
12. Hull J, White A (1990) Pricing interest rate derivative securities. Rev Financ Stud 3(4):573–592
13. Hull J, Sokol A, White A (2014) Short rate joint measure models. Risk 10:59–63
14. de Jong F (2000) Time series and cross-section information in affine term-structure models. J Bus Econ Stat 18(3):300–314
15. Korn R, Wagner A (2019) Praxishandbuch Lebensversicherungsmathematik: Simulation und Klassifikation von Produktent. VVW GmbH, Karlsruhe
16. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7(4):308–313
17. Stanton R (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. J Financ 52(5):1973–2002
18. Vasicek O (1977) An equilibrium characterization of the term structure. J Financ Econ 5(2):177–188

III. A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction

This chapter is a reprint of:
C. Berninger, A. Stöcker and D. Rügamer. 2021. A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction. https://doi.org/10.1002/for.2802. *accepted by Journal of Forecasting.*

Author Contributions

The project was initiated by Christoph Berninger. The Bayesian Time-Varying Autoregressive Model framework was developed by all authors in close collaborations. The full conditional distributions of all model parameters have been derived by David Rügamer. The implementation and simulation in MATLAB have been done by Christoph Berninger. Most parts of the manuscript have been drafted by Christoph Berninger and all authors contributed to the manuscript.

A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction

Christoph Berninger[a], Almond Stöcker[b], David Rügamer[a]

[a]*Department of Statistics, LMU München*
[b]*School of Business and Economics, Humboldt University of Berlin*

Abstract

Motivated by the application to German interest rates, we propose a time-varying autoregressive model for short and long term prediction of time series that exhibit a temporary non-stationary behavior but are assumed to mean revert in the long run. We use a Bayesian formulation to incorporate prior assumptions on the mean reverting process in the model and thereby regularize predictions in the far future. We use MCMC-based inference by deriving relevant full conditional distributions and employ a Metropolis-Hastings within Gibbs sampler approach to sample from the posterior (predictive) distribution. In combining data-driven short term predictions with long term distribution assumptions our model is competitive to the existing methods in the short horizon while yielding reasonable predictions in the long run. We apply our model to interest rate data and contrast the forecasting performance to that of a 2-Additive-Factor Gaussian model as well as to the predictions of a dynamic Nelson-Siegel model.

Keywords: MCMC Metropolis-Hastings, Gibbs sampler, Bayesian time-varying autoregressive models, long run regluarization, interest rate models

1. Introduction

To forecast an univariate time series the first model of choice is often a linear model. An archetype of this model class in the context of time series analysis is the autoregressive model of order 1 (AR(1)), which is defined as follows:

$$x_t = \alpha + \beta x_{t-1} + \epsilon_t, \tag{1}$$

where x_t represents the variable that is defined on $t \in \mathbb{Z}$ and was observed at time points $t = 1, \ldots, T$. α and β are real valued constants, while $|\beta| < 1$ is assumed to ensure stationarity. The innovation process ϵ_t can be, e.g., a Gaussian white noise process, i.e., an independent and identically (i.i.d.) normal distributed $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ for all time points t. In this classical model, the one-step-ahead expectation $\mathrm{E}(x_t \mid x_{t-1}) = \alpha + \beta x_{t-1}$ and variance $\mathrm{Var}(x_t \mid x_{t-1}) = \sigma^2$ are closely linked to the marginal characteristics $\mathrm{E}(x_t) = \frac{\alpha}{1-\beta}$ and $\mathrm{Var}(x_t) = \frac{\sigma^2}{1-\beta^2}$ approached in the long run. In this sense, fitting the short-term behavior of a time series with a linear model has wide implications for its long-term behavior, and, conversely, controlling the long-term behavior of the model constraints its short-time fit. In practice, this close linkage may present an important limitation when short-term performance conflicts with long-term plausibility.

This gets evident in the macroeconomic literature and, more specifically, in contrasting modeling approaches for interest rates, which motivate this work: Among others, Diebold and Li (2006) focus on predictions in the short horizon. The authors apply an AR(1) process to extracted factors of the yield curve. While they do make the long-term assumption that interest rates are principally mean reverting, the process exhibits an almost integrated behavior. Estimating the model parameters of a near integrated but stationary AR(1) model might give large estimation errors and lead to unrealistic long run mean estimates far beyond the range of the data. Duffee (2011) even discard the mean reversion / stationarity assumption. In their proposed random walk model, the prediction variance linearly increases in time leading to extreme values in the long run. The strong focus on the short horizon might lead to questionable and potentially implausible long-term behavior. Also the model of Caldeira and Torrent (2017), who apply a nonparametric functional data approach, has the focus on the short horizon. In contrast, Korn and Wagner (2019) apply a linear model to model long horizon features. Their proposal is a Gauss2++ model with a forward looking estimation approach, i.e., they calibrate the parameters of their model without using historical data but amongst current prices of interest rate derivatives and long run survey forecasts. The Gauss2++ model is a standard model in the insurance industry, where plausible forecasts in the longer run are required. As demonstrated later in this paper, their model achieves a realistic long run distribution, however, at the expense of inferior short term predictions.

The aim of this paper is to bridge the gap between short and long horizons, generalizing above mentioned approaches to a model with the flexibility to a) sufficiently adapt to sample data to yield good short-term predictions and b) apply suitable regularization to achieve plausible long- and middle-term forecasts at the same time. This is particularly, yet not exclusively, relevant in applications where a stable stationary distribution is assumed in the long run – despite observing a 'temporary non-stationary behavior' in the available data where, e.g., unconstrained linear model fits or a Dickey-Fuller test (Dickey and Fuller, 1979) would indicate an integrated process.

As in general linearity is often a restrictive assumption in practice and many time series exhibit features that cannot be captured by a linear model (Hamilton, 1989), a lot of research has been conducted to introduce different types of nonlinear models in the last decades. In particular, nonlinear models offer more flexibility to account for both, short and long horizon.
A bi-linear model is an example of a nonlinear model, which assumes a nonlinear relationship between the covariates and response variable (Granger and Andersen, 1978, Rao and Gabr, 2012), although not often used in macroeconomic applications (Morley, 2009). A more immediate approach is to allow one (or more) parameters of a linear model to change over time. This comprises the regime switching and time-varying parameter models.

Regime switching models can allow for a different mean reversion level in the short and the long horizon. This feature can be used to regularize the long run mean of the model. The first approaches to regime switching models were conducted by Quandt (1958), who considered a switching regression model extending a linear regression model by allowing the parameters to switch between different states according to a random variable. Bacon and Watts (1971) introduced a smooth transition model, which implements a smooth transition from one regime to another without a sudden jump. Goldfeld and Quandt (1973) introduced the Markov switching regression model and use a discrete latent Markov process to determine the current regime. These models were adapted to time series models by Lim and Tong (1980) and Chan and Tong (1986) introducing the threshold autoregressive model (TAR) and the smooth transition autoregressive model (STAR), respectively. Hamilton (1989) introduced the Markov switching autoregressive model for applications in economics, which have been investigated thoroughly together with different variants in the literature (Haggan and Ozaki, 1981, Jansen and Teräsvirta, 1996, Teräsvirta, 1994). Lanne and Saikkonen (2002) used a TAR-model, which only allows regime changes for the constant parameter α, and applied it to strongly autocorrelated time series data – which is very related to temporary non-stationary behavior of the time series and, therefore, to the aim of the paper. When there is, however, no concrete indication for the process dynamics to result from switches between discrete underlying states, we consider it more natural and promising to assume a continuous latent process.

In contrast to regime switching models, time-varying parameter models allow one (or more) of the parameters in a linear model to be driven by its own continuous process (Morley, 2009). For example, if the parameter vector (α, β, σ^2) of the linear AR(1) model becomes a stochastic process, this results in a time-varying autoregressive model of order 1 (TV-AR(1))

$$x_t = \alpha_t + \beta_t x_{t-1} + \epsilon_t \tag{2}$$

with $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. Certain distribution assumptions for the underlying stochastic process of the parameter vector (α_t, β_t, σ_t) are made in practice to complete the TV-AR(1) model specification (Teräsvirta et al., 2010). Similar to the TAR model in Lanne and Saikkonen (2002) the time variation of the TV-AR(1) model can be restricted to the constant parameter α_t, resulting in a time-varying constant autoregressive model of order 1 (TVC-AR(1)):

$$x_t = \alpha_t + \beta x_{t-1} + \epsilon_t. \tag{3}$$

If $|\beta| < 1$ and the latent process of α_t is stationary, the process x is also stationary. But due to random shifts in the mean reversion level – because of the time-varying

constant parameter – realizations of the model can resemble those of a (close to) random walk process, when restricting to a limited time window.
Another time-varying parameter model is the shifting endpoint model introduced by Kozicki and Tinsley (2001). Similar to the TVC-AR(1) model their approach has a time-varying mean reversion level, referred to as shifting endpoints. In particular, Van Dijk et al. (2014) applied this model to interest rates presenting a slow moving trend using exponential smoothing or long survey forecasts. There is also a strand of literature, which connects the level of interest rates in the long run to the expected inflation dynamics, also referred to as trend inflation (Kozicki and Tinsley, 2001, Rudebusch and Wu, 2008, Bekaert et al., 2010, Cieslak and Povala, 2015). Associating the variable of interest with appropriate covariates might practically help in several scenarios, but does not directly address the core of the present problem.

In this paper, we propose a model approach competitive in terms of short horizon forecasts, yet controlled to obtain realistic predictive distributions for the long horizon. We propose a Bayesian TVC-AR(1) model, which is stationary but can resemble short-term properties of an integrated or nearly integrated linear process due to a stochastic mean reversion level. The model allows us to regularize the long run distribution of the time series without affecting short term distributions adversely. Different to Van Dijk et al. (2014) we do not use a deterministic shifting mean reversion level, but incorporate long run assumptions via prior information in a Bayesian approach, such that the latent coefficient process, and with it the mean reversion dynamics, are still estimated from the data.
In particular, the novelty of our approach lies in the model allowing us to (1) regularize long run predictions by using prior assumptions, (2) separate the modeling process into a data driven short horizon model-part and a long horizon model-part that is determined by prior (or expert) assumptions and (3) yield improved forecasting performance in the short horizon compared to the commonly used linear dynamic Nelson-Siegel model and Gauss2++ model, while retaining realistic long run distributions. Moreover, we place particular emphasis on the interpretability of the model structure and prior parameters, preserving a close link to the common linear models. This allows to include expert knowledge or assumptions in accordance with economic theory about the long run behavior of a time series into the model in a sound mathematical way. We here specifically focus on the application to interest rates. As our

model allows to regularize long run predictions, it is also of particular interest for insurance companies, where realistic long run interest rate forecasts are needed to evaluate the risk and performance of specific insurance products.

The remainder of this paper is arranged as follows. Section 2 specifies the Bayesian TVC-AR(1), including the derivation of required full conditional posterior distributions and the application of a Metropolis-Hastings within Gibbs sampling routine for statistical inference. In Section 3 we discuss an application of our model to interest rate data and compare the forecasting performance as well as the long run distribution of our nonlinear model with the dynamic Nelson-Siegel model (short-term focus) and the Gauss2++ model (long-term focus). We conclude with Section 4 and give a brief outlook on potential further research topics.

2. A Bayesian TVC-AR(1) Model for Long Run Regularization

In this Section we introduce the Bayesian TVC-AR(1) (BTVC-AR(1)) model. The model incorporates assumptions about the long-term behavior of the time series and thereby regularizes the process in the long horizon. At the same time, the model is mainly driven by the given data in the short run and thus fosters a good short-term prediction.

2.1. The BTVC-AR(1) Model

The BTVC-AR(1) model is defined as follows:

$$x_t = \alpha_t + \beta x_{t-1} + \epsilon_t, \qquad \text{for } t \in \mathbb{Z}, \tag{4}$$

where β represents the mean reversion speed and $|\beta| < 1$ to secure stationarity. ϵ_t is assumed to be a Gaussian white noise process, i.e., $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. We further specify α_t as a stationary Gaussian process, which is defined to have the unconditional expectation $\boldsymbol{\theta} := \vartheta \cdot \mathbf{1}$ and covariance $\boldsymbol{\Sigma}$ on the observed time points $t = 1, \ldots, T$, i.e.,

$$\boldsymbol{\alpha} := (\alpha_1, \alpha_2, ..., \alpha_T) \sim \mathcal{N}_T(\boldsymbol{\theta}, \boldsymbol{\Sigma}). \tag{5}$$

As this time frame is most relevant, we focus on $\boldsymbol{\alpha}$ for further investigations. The Bayesian approach considers the parameters of model (4) as random variables. For the conditional prior distribution of β conditional on σ^2 a truncated normal distribution with lower bound -1 and upper bound 1 is assumed as a prior, i.e.,

$$\beta|\sigma^2 \sim \mathcal{N}(\mu_\beta, \sigma^2 \cdot \sigma_\beta^2, -1, 1),$$

with conditional prior expectation μ_β and additional multiplicative variance parameter σ_β. The prior distribution for σ^2 is an inverse gamma distribution with shape and scale parameter, a and b, respectively,

$$\sigma^2 \sim \mathcal{IG}(a, b).$$

These two prior distributions are conjugate priors for model (4) if the respective other parameter is known and therefore allow for an analytical derivation of the corresponding full conditional distributions.

Using these priors the defined model can be seen as a Bayesian version of the TVC-AR(1) model. The mean $\boldsymbol{\theta}$ and covariance $\boldsymbol{\Sigma}$ might be assumed fixed or defined as random variables with further attached prior distributions. In the latter case (5) describes the distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. Placing priors on these parameters allows to incorporate assumptions about the long run distribution into the model as further elaborated in Section 2.2.

While this basic model setup is flexible in many ways and particular in terms of its covariance structure assumptions for $\boldsymbol{\alpha}$, further practical insights can be obtained from a more in-depth model characterization. In the following, we will shed light on useful properties of this framework when assuming an AR-covariance structure. This covariance structure has shown to provide good results in applications.

2.2. Arbitrating Between Short and Long Run Distribution

The goal of our work is to propose a new modeling framework, which can regularize the long run distribution of (nearly) integrated time series by keeping a good forecasting performance in the short horizon. Linear models often concentrate on the conditional distribution in the short horizon, but due to the near integration property of the time series this can lead to inappropriate long run distributions. For example, if the AR(1) model is estimated for a time series, which shows a (close to) random walk behavior, the parameter β of the model will take a value close to 1. This can lead to a large long run variance given by

$$\frac{\sigma^2}{1-\beta^2},$$

potentially yielding unrealistic values in the long run that have never been observed in the past. On the other hand, calibrating β to a given long run variance is not straightforward without deteriorating the short run prediction performance. Figure 1 depicts this undesired behavior by showing the long run mean of a linear AR(1) model which is driven by the conditional short run distribution at the expense of an unrealistic long-term distribution.

We address this issue by incorporating a time-varying mean reversion level, which locally preserves the good short term prediction and at the same time regularizes the long run distribution. The current mean reversion level valid in the short run can be different to the long run behavior accounting for the current market situation and therefore improving the short run prediction. We enable the model to stay in a reasonable range in the long run by assuming a stationary process for the time-varying mean reversion level and a stronger mean reversion to this time-varying level than a linear AR(1) model would induce to its constant mean reversion level. Such a behavior can be achieved by introducing a time-varying α parameter into a linear AR(1) model with additional prior assumptions. In particular, this does not change the (weak) stationarity property of the model if the assumed process for α is (weakly) stationary itself. This can be verified by calculating the unconditional mean, the unconditional variance and the unconditional covariance:

$$\begin{aligned}
\mathrm{E}(x_t) &= \frac{\vartheta}{1-\beta} \\
\mathrm{Var}(x_t) &= \frac{\sigma^2}{(1-\beta^2)} + \frac{\mathrm{Var}(\alpha_t) + 2\beta\,\mathrm{Cov}(\alpha_t, x_{t-1})}{(1-\beta^2)} \\
\mathrm{Cov}(x_{t-h}, x_t) &= \sum_{i=0}^{h-1} \beta^i\,\mathrm{Cov}(\alpha_{t-i}, x_{t-h}) + \beta^h\,\mathrm{Var}(x_{t-h})
\end{aligned}$$

As the α-process is stationary and

$$\mathrm{Cov}(\alpha_t, x_{t-h}) = \sum_{i=0}^{\infty} \beta^i\,\mathrm{Cov}(\alpha_t, \alpha_{t-h-i}),$$

the BTVC-AR(1) is (weakly) stationary.

The time-varying α increases the flexibility of our model to account for short and long run distributional properties. As current observations have almost no influence in the very long run, a reasonable way to include information about the long run mean and long run variance in a Bayesian setting is via prior assumptions for $\boldsymbol{\alpha}$. We will further elaborate this in Section 2.2.1 and 2.2.2. The time-varying α also increases the flexibility of the model such that the conditional distribution in the short run is consistent with the empirical data, i.e., $\mathrm{E}(x_{t+h}|x_t, x_{t-1}, \dots)$ and $\mathrm{Var}(x_{t+h}|x_t, x_{t-1}, \dots)$ still reflect the empirical distribution for a short horizon h.

Our BTVC-AR(1) model can therefore produce both a conditional short term distribution, which roughly corresponds to an unrestricted linear model, and a long run distribution with a reasonable range of values.

2.2.1. The Long Run Mean and Time-Varying Mean Reversion

The mean reversion level in a linear AR(1) model as specified in (1) amounts to

$$\frac{\alpha}{1-\beta}.$$

As the mean reversion level stays constant over time it is also the long run mean of the model. In contrast, the mean reversion level in the BTVC-AR(1) model changes over time and is given by

$$\frac{\alpha_t}{1-\beta}$$

for time point t. This local mean reversion level is in general different to the long-term mean and can even pull the process away from it in expectation, i.e,

$$\left|\ \mathrm{E}(x_{t+h}|x_t, x_{t-1}, \dots) - \mu_{LR}\right| \geq \left|\ x_t - \mu_{LR}\right|,$$

where μ_{LR} is the unconditional mean of the time series and x_t denotes a fixed realisation of the process. This helps fitting the model to a time series exhibiting a (close to) random walk behavior. The long run mean of the BTVC-AR(1) depends on the unconditional mean of $\boldsymbol{\alpha}$ and amounts to

$$\frac{\vartheta}{1-\beta}$$

in our model. We assume the data to be centered around a prior specified long run mean. By setting $\boldsymbol{\theta} = \mathbf{0}$, i.e., $\vartheta = 0$, this long run mean is reached in expectation after reshifting the simulated data.

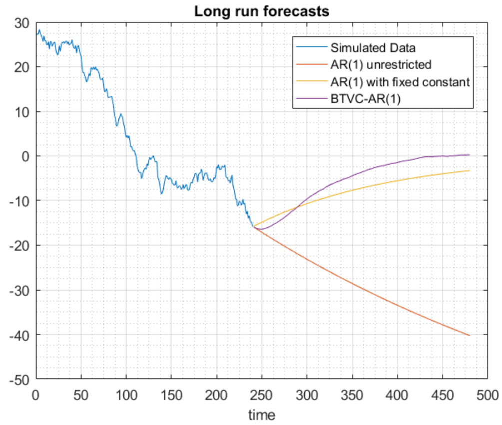
The implications of the time-varying mean reversion level of the BTVC-AR(1) model are visualized in Figure 1. Two AR(1) models (with unrestricted and restricted constant parameter) and the BTVC-AR(1) model have been exemplary fitted to a simulated stationary time series, which shows a (nearly) integrated behavior and is visualised in Figure 1.

In the left graphic the "historical" time series can be seen as well as the expected future development according to the three models. The AR(1) model with no restrictions has a long-term mean far away from the historical domain, as its focus lies on the conditional short term distribution. The restricted AR(1) model sets the α parameter to 0 to regularize the long run mean, but at the same time the expected values in the short horizon are pushed in the direction of the long run level leading to an inferior forecasting performance. The parameters of the estimated unrestricted and restricted AR(1) model are given in Table 1.

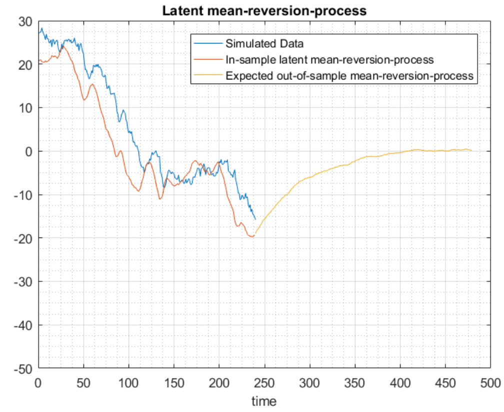
	constant	β	σ^2
restricted AR(1)	0	0.9935	0.7787
unrestricted AR(1)	-0.1645	0.9978	0.7554

Table 1: Estimated parameters for the restricted and unrestricted AR(1) model.

If we assume that the (close to) random walk behavior stems from changes in the mean reversion level determined by unobserved variables, the BTVC-AR(1) model has a more desired behavior. The time-varying constant parameter in the model leads to a time-varying mean reversion level and can therefore account for the changes induced by the unobserved variables. The long run mean can still be regularized to 0 while influencing the short term distribution less abruptly. This allows the time series to follow the current trend in expectation and veer away from the long run mean for a couple of time steps. The reason for this behavior is that the latent α-process induces a local mean reversion level that lies below the last observation, which can be seen in the right plot of Figure 1 showing the average latent mean reversion level extracted during the simulation process. In the long run the mean reversion level returns in expectation to the prespecified value of 0.



(a)



(b)

Figure 1: A comparison of a linear AR(1) model with no restrictions for the constant parameter, a linear AR(1) model with a fixed constant parameter ($\alpha = 0$) and a BTVC-AR(1) model applied on a simulated time series.

2.2.2. Long Run Variance

The long run variance of a linear AR(1) model is given by

$$\frac{\sigma^2}{1-\beta^2}.$$

The closer the model behaves like a random walk, i.e., the closer β approaches 1, the larger the long run variance gets under the assumption of a fixed conditional variance σ^2. In terms of the long run variance, the BTVC-AR(1) model is more flexible by incorporating two sources of variation, the residual term of the AR(1) model and variance of the latent α-process. The model's long run variance is given by

$$\mathrm{Var}(x_t) = \frac{\sigma^2}{1-\beta^2} + \frac{\mathrm{Var}(\alpha_t) + 2\beta\,\mathrm{Cov}(\alpha_t, x_{t-1})}{1-\beta^2}. \tag{6}$$

The first term has the same form as the long run variance of a linear AR(1) model and can be interpreted as the "unconditional" variance around the time-varying mean reversion level, i.e., the variance conditional on the α-process. The second term incorporates the part of the variance stemming from the α-process and depends on both its unconditional variance and unconditional covariances. This allows the BTVC-AR(1) model to be more flexible and to control the long run variance of x_t, while reducing the opposing effect on the conditional distribution in the short horizon. The model thus still produces short term distributions consistent with the given data. If α is a constant process, the second term is 0 and the BTVC-AR(1) model reduces to a linear AR(1) model.

Prior Assumptions. With the goal in mind to control the long-term variance based on prior information, a more refined specification of the BTVC-AR(1) model is helpful in order to translate this information into the model. We will use a centered α-process with an AR-covariance structure for demonstrative purposes. In this case, $\boldsymbol{\alpha}$ can be represented by a linear AR(1) model

$$\alpha_t = \rho\alpha_{t-1} + \eta_t, \qquad \text{for } t \in \mathbb{Z},$$

where ρ represents the correlation between two successive time steps and η_t is an i.i.d. Gaussian white noise process, i.e., $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \tau^2)$. The long run variance of the BTVC-AR(1) model is then given by

$$\mathrm{Var}(x_t) = \frac{\sigma^2}{(1-\beta^2)} + \frac{\tau^2(1+\rho\beta)}{(1-\rho\beta)(1-\beta^2)(1-\rho^2)}. \tag{7}$$

As indicated by the equation, τ^2 and σ^2 play an interrelated role in the model for the long run variance. Increasing one can be compensated by reducing the other one. To avoid identifiability issues, it is thus necessary to account for this interrelation through a meaningful prior parameter assumption. If the process x_t is supposed to reach a certain objective variance in the long run, the degrees-of-freedom in (7) reduce from four to three. For example, for given ρ, β and σ^2 and a prior value assumption for $\mathrm{Var}(x_t)$, the variance of x_t has a one-to-one relationship with τ^2 and it is straightforward to solve (7) for τ^2. Denote the solution by $\tilde{\tau}^2$. To ensure positivity the truncation limits for the prior distribution of β can be set to -1 and $\sqrt{\frac{\mathrm{Var}(x_t)-\sigma^2}{\mathrm{Var}(x_t)}}$. For this specific covariance structure, a possible prior distribution of τ^2 can thus be defined by the conditional distribution

$$\tau^2|\rho, \beta, \sigma^2 \sim \delta_{\tilde{\tau}^2}, \tag{8}$$

where δ denotes a degenerated distribution with point mass 1 at $\tilde{\tau}^2$. This definition forces the process to reach its prespecified long run variance $\mathrm{Var}(x_t)$ while controlling the speed of mean reversion of the α-process through ρ. A conjugate prior for ρ is a normal distribution truncated below by -1 and from above by 1, i.e.,

$$\rho \sim \mathcal{N}(\mu_\rho, \sigma_\rho^2, -1, 1), \tag{9}$$

with mean μ_ρ and variance σ_ρ^2.

The previous prior specifications allow to introduce prior information into the model in a straightforward manner while maintaining the properties of the BTVC-AR(1) model.

2.2.3. The Short Run Distribution

For the short run distribution of the BTVC-AR(1) model the goal is to balance between a consistent estimation with the observed data and the opposing effect of the prespecified long run distribution. For a linear AR(1) model with a restricted long run mean of 0 the conditional expectation and the conditional variance amount to

$$\begin{aligned} \mathrm{E}(x_{t+1}|x_t,\dots) &= \beta x_t, \\ \mathrm{Var}(x_{t+1}|x_t,\dots) &= \sigma^2. \end{aligned}$$

The model can get arbitrarily close to a centered random walk if β approaches 1, while the long run variance increases at the same time as shown in Section 2.2.2. For the BTVC-AR(1) model we get

$$\begin{aligned} \mathrm{E}(x_{t+1}|x_t,\dots) &= \mathrm{E}(\alpha_{t+1}|x_t,\dots) + \beta x_t, \\ \mathrm{Var}(x_{t+1}|x_t,\dots) &= \mathrm{Var}(\alpha_{t+1}|x_t,\dots) + \sigma^2. \end{aligned}$$

A random walk behavior, i.e., $\mathrm{E}(x_{t+1}|x_t) \approx x_t$, can be reached without β necessarily being close to 1 due to the conditional expectation of the α-process that supports the random walk behavior in the short horizon. This increases the flexibility of the BTVC-AR(1) model compared to a linear AR(1) model in combining short and long run distributional characteristics.

We can further decompose the conditional expectation to see the similarities of the BTVC-AR(1) model to a linear AR(T) process if we consider a time series observed on time points $t = 0, \dots, T$. Let $\breve{\boldsymbol{\alpha}} = (\alpha_1, ..., \alpha_{T+1})$ denote the time-varying constant extended to $T+1$ in a consistent manner with the BTVC-AR(1) model definition, i.e., the same covariance parameterization is assumed. For a given data set $\boldsymbol{x} = (x_0, ..., x_T)$, the conditional distribution of $\breve{\boldsymbol{\alpha}}|\boldsymbol{x}$ is multivariate normal (c.f. Appendix A.1), i.e.,

$$\breve{\boldsymbol{\alpha}}|\boldsymbol{x} \sim \mathcal{N}(\breve{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}).$$

As $\breve{\boldsymbol{\mu}} = \frac{1}{\sigma^2}\breve{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Delta}}$ with $\tilde{\boldsymbol{\Delta}} = (x_1 - \beta x_0, ..., x_T - \beta x_{T-1}, 0)^\top$, the conditional expectation of α_{T+1} is given by the last entry of $\breve{\boldsymbol{\mu}}$,

$$\mathrm{E}(\alpha_{T+1}|\boldsymbol{x}) = \frac{1}{\sigma^2}\boldsymbol{s_{T+1,.}}\tilde{\boldsymbol{\Delta}},$$

where $\boldsymbol{s_{T+1,.}} = (s_{T+1,1}, \ldots, s_{T+1,T+1})$ and $s_{i,j}$ represent the entries of $\breve{\boldsymbol{\Sigma}}$. The one step ahead conditional expectation of the model therefore amounts to

$$\mathrm{E}(x_{T+1}|x_T, \ldots) = \left(\frac{s_{T+1,T}}{\sigma^2} + \beta\right) x_T + \sum_{i=1}^{T-1} \frac{s_{T+1,T-i} - \beta s_{T+1,T-(i-1)}}{\sigma^2} x_{T-i} - \frac{s_{T+1,1}\beta}{\sigma^2} x_0.$$

This shows that the conditional expectation depends on all previous time points like in a linear AR(t) model, allowing the BTVC-AR(1) model to better account for current trends in the process. Due to the given covariance structure for $\boldsymbol{\alpha}$ the number of parameters are, however, much less than in an actual AR(t) process.

2.3. Bayesian Inference

The main parameters of interest in the BTVC-AR(1) model are $\widetilde{\boldsymbol{\alpha}}$, β and σ^2 with $\widetilde{\boldsymbol{\alpha}}$ extending $\boldsymbol{\alpha}$ by future time points up to the modeling horizon h, i.e.,

$$\widetilde{\boldsymbol{\alpha}} = (\alpha_1, ..., \alpha_T, ..., \alpha_{T+h}).$$

This extension is necessary to sample from the predictive posterior distribution of the parameters and to generate forecasts. The prior distribution of $\widetilde{\boldsymbol{\alpha}}$ incorporates the same assumptions as the prior distribution of $\boldsymbol{\alpha}$, i.e.,

$$\widetilde{\boldsymbol{\alpha}} \sim \mathcal{N}_{T+h}(\widetilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}})$$

where

$$\widetilde{\boldsymbol{\theta}} = \vartheta \cdot \mathbf{1} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Pi_{T+1}} & \dots & \boldsymbol{\Pi_{T+h}} \\ \boldsymbol{\Pi}_{\boldsymbol{T+1}}^{\top} & \sigma_\alpha^2 & & \\ \vdots & & \ddots & \\ \boldsymbol{\Pi}_{\boldsymbol{T+h}}^{\top} & & & \sigma_\alpha^2 \end{pmatrix}$$

with $\boldsymbol{\Pi_{T+j}} = \{\text{Cov}(\alpha_{T+j}, \alpha_1), \dots, \text{Cov}(\alpha_{T+j}, \alpha_{T+j-1})\}^{\top}$, i.e., the vector of covariances of α_{T+j} and all previous time points $1, \dots, (T + j - 1)$. For these time points the same (autoregressive) covariance parameterization as for $\boldsymbol{\alpha}$ is assumed for consistency reasons. σ_α^2 represents the unconditional variance of the latent α-process. In the case of an AR-covariance structure as assumed before $\sigma_\alpha^2 = \frac{\tau^2}{1-\rho^2}$ and $\text{Cov}(\alpha_s, \alpha_{s+k}) = \rho^k$ for any $s, k \in \mathbb{Z}$.

The goal of Bayesian inference is to find the joint posterior distribution, $p(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2 | \boldsymbol{x})$, conditional on the observed data $\boldsymbol{x} = (x_0, ..., x_T)$. If the full conditional distribution of all parameters is known, the Gibbs sampler (Gelman et al., 2013) can be used to draw samples from this joint posterior distribution and inference can be based on Monte Carlo approximation (Chib, 2001). By regularizing the long run variance under the assumption of an AR-covariance structure and choosing a degenerated prior distribution for τ^2 as in (8), the full conditional distributions of ρ, β and σ^2 depend on the prior of τ^2 and can not be derived analytically. We therefore apply a Metropolis-Hastings within Gibbs sampling routine (Millar and Meyer, 2000). We will state the algorithmic details in the following section and here only derive the necessary distributions.

As the model defined in Section 2.2.2 can be considered under a different parameterization where τ^2 is given by the function

$$\tau^2 = f(\rho, \beta, \sigma^2, \text{Var}(x_t))$$

and thus fixed for given ρ, β, σ^2 and a specified long run variance $\text{Var}(x_t)$, we will focus on deriving two conditional distributions in order to be able to employ a two-step Gibbs sampling procedure. The goal is to iteratively sample $\boldsymbol{\alpha}$ and the vector (ρ, β, σ^2) based on the respective other full conditional distribution. As it is not straightforward to derive the conditional distribution for the latter vector, we will here derive conditional distributions for all parameters involved as if the parameter τ^2 was fixed and later employ these distributions to derive a suitable proposal distribution in a Metropolis-Hastings procedure. In the following subsections we just state the (full) conditional distributions. A more detailed derivation can be found in Appendix A.1-Appendix A.4.

2.3.1. Full Conditional Distributions of $\boldsymbol{\alpha}$

In the following we derive the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$. It holds

$$p(\widetilde{\boldsymbol{\alpha}}|\beta,\sigma^2,\boldsymbol{x}) \propto p(\boldsymbol{x}|\widetilde{\boldsymbol{\alpha}},\beta,\sigma^2)\cdot p(\widetilde{\boldsymbol{\alpha}}) = \mathcal{L}(\widetilde{\boldsymbol{\alpha}},\beta,\sigma^2)\cdot p(\widetilde{\boldsymbol{\alpha}}). \tag{10}$$

Due to the conditional independence induced by the Markov assumption in the AR(1) model the likelihood of the parameters is given by

$$\mathcal{L}(\widetilde{\boldsymbol{\alpha}},\beta,\sigma^2) = p(\boldsymbol{x}|\alpha,\beta,\sigma^2) = \prod_{j=0}^{T-1}\phi(x_{T-j}|\alpha_{T-j}+\beta x_{T-j-1},\sigma^2), \tag{11}$$

where $\phi(\cdot|\mu,\widetilde{\sigma}^2)$ denotes the density function of a normal distribution with expectation μ and variance $\widetilde{\sigma}^2$. Note, that we have assumed a degenerated distribution with point mass 1 for the first entry in $\boldsymbol{x}$. An alternative option is to estimate the unconditional distribution. For increasing length of the time series the difference between these two approaches will however vanish.

With (10) and (11) and the prior distributions specified in Section 2.1 the full conditional distributions of $\widetilde{\boldsymbol{\alpha}}$, can be derived analytically. Under the assumption that $\widetilde{\boldsymbol{\theta}} = \boldsymbol{0}$ as specified in Section 2.2.1 to regularize the long run mean, the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$ is given by

$$\widetilde{\boldsymbol{\alpha}}|\beta,\sigma^2,\boldsymbol{x} \sim \mathcal{N}_T(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}},\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}).$$

with

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}\widetilde{\boldsymbol{\Delta}}\frac{1}{\sigma^2} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}} = \left(\tilde{\boldsymbol{\Sigma}}^{-1} + \frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{I_T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\right)^{-1}.$$

$\widetilde{\boldsymbol{\Delta}}$ in this case denotes

$$\widetilde{\boldsymbol{\Delta}} = (x_2-\beta x_1,\ldots,x_T-\beta x_{T-1},0,\ldots,0).$$

As $\widetilde{\boldsymbol{\Delta}}$ incorporates data information up to time point (vector entry) T, is 0 for time points $> T$ and $\mathrm{Cov}(\alpha_{T+j}, \alpha_T) \longrightarrow 0$ with increasing j, the mean of the full conditional distribution tends to 0, corresponding to the unconditional mean of the prior distribution. The covariance structure of the full conditional distribution behaves analogously. Therefore, the distribution of $\alpha_{T+j} \mid \boldsymbol{x}, \beta, \sigma^2$ in the long run tends to the prior distribution. This means that the prior distribution of $\boldsymbol{\alpha}$ effectively regularizes the distribution of $\boldsymbol{x}$ in the long horizon towards the prespecified long run mean and long run variance.

Note that the derivations are independent of the specific choice of $\tilde{\boldsymbol{\Sigma}}$. If prior distribution assumptions for the parameters in $\tilde{\boldsymbol{\Sigma}}$ are used, we need to further condition on the hyper-parameters for the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$. In the following, we assume an AR-covariance structure for $\widetilde{\boldsymbol{\alpha}}$ and therefore need to condition on the parameters ρ and τ^2.
We note that, in general, regarding the α process as a vector of parameters can lead to a computational burden when calculating the inverse of $\tilde{\boldsymbol{\Sigma}}$ for $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}$. In this case it might be preferable to represent the model in state space and to use a forward filtering backward sampling algorithm as proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). In our specific case, however, we initialize a specific parametric covariance structure and did not experience any computational problems.

2.3.2. Full Conditional Distributions of ρ, β, σ^2

If we assume an AR-covariance structure with prior distributions for its parameters as specified in Section 2.2.2, the conditional distribution of ρ is given by

$$\rho | \boldsymbol{\alpha}, \tau^2 \sim \mathcal{N}(\mu_{\rho,post}, \sigma^2_{\rho,post}, -1, 1)$$

where

$$\sigma^2_{\rho,post} = \left(\frac{\sum_{j=0}^{T-1} \alpha^2_{T-j-1}}{\tau^2} + \sigma_\rho^{-2} \right)^{-1}$$

$$\mu_{\rho,post} = \left(\frac{\sum_{j=0}^{T-1} \alpha_{T-j}\alpha_{T-j-1}}{\tau^2} + \frac{\mu_\rho}{\sigma_\rho^2} \right) \sigma^2_{\rho,post}.$$

The conditional distribution of β is given by

$$\beta|\boldsymbol{x}, \boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}\left(\mu_{\beta,post}, \sigma^2_{\beta,post}, -1, \sqrt{\frac{Var(x_t) - \sigma^2}{Var(x_t)}}\right)$$

where

$$\sigma^2_{\beta,post} = \left(\frac{\sum_{j=0}^{T-1} x^2_{T-j-1}}{\sigma^2} + (\sigma\sigma_\beta)^{-2}\right)^{-1}$$
$$\mu_{\beta,post} = \left(\frac{\sum_{j=0}^{T-1} \breve{d}_{T-j} x_{T-j-1}}{\sigma^2} + \frac{\mu_\beta}{\sigma^2\sigma^2_\beta}\right)\sigma^2_{\beta,post}.$$

$\breve{d}_{T-j}$ is defined by $\breve{d}_{T-j} := x_{T-j} - \alpha_{T-j}$.

The conditional distribution of σ^2 is given by an inverse gamma distribution with parameters

$$\tilde{a} = \frac{T+1}{2} + a \qquad \text{and} \qquad \tilde{b} = \frac{\sum_{j=0}^{T-1} \epsilon^2_{T-j}}{2} + b + \frac{(\beta - \mu_\beta)^2}{2\sigma_\beta},$$

where ϵ_t is the error term of the BTVC-AR(1) model (c.f. (4)). This means

$$\sigma^2|\tilde{\boldsymbol{\alpha}}, \beta, \boldsymbol{x} \sim \mathcal{IG}(\tilde{a}, \tilde{b}).$$

Note that this only holds if the prior of $\beta \mid \sigma^2$ is a normal distribution instead of a truncated normal distribution as assumed in Section 2.1. When using a truncated distribution assumption, the derivation of the full conditional of σ^2 is more intricate as the prior distribution of β also conditions on σ^2. Since our approach will make use of the full conditionals as proposal distributions in the Metropolis-Hastings part of our sampling routine, this simplification allows a more straightforward implementation while we observe that values outside the given truncation are highly unlikely and practically occur with zero probability in our application.

2.4. Markov Chain Monte Carlo Inference

In the following we assume again an AR-covariance structure for $\mathbf{\Sigma}$ determined by the parameters ρ and τ^2 with prior distributions as specified in Section 2.2.2.

To conduct inference, we use the Metropolis-Hastings within Gibbs sampler. More specifically, we generate samples from the posterior distribution by iteratively sampling from the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$ given a sample of (ρ, β, σ^2) and vice versa. Based on the derivation of the full conditional distribution for $\boldsymbol{\alpha}$ in the previous section we are able to directly sample from a multivariate normal distribution to generate values for $\boldsymbol{\alpha}$. To obtain a sample from $p(\rho, \beta, \sigma^2 \mid \boldsymbol{\alpha}, \boldsymbol{x})$ conditional on $\boldsymbol{\alpha}$, we apply the Metropolis-Hastings algorithm as neither the joint distribution of ρ, β, σ^2 nor each single full conditional distribution is available. A suitable and already available proposal distribution q for these parameters is given by

$$q(\rho, \beta, \sigma^2 \mid \boldsymbol{\alpha}, x) = q(\rho \mid \boldsymbol{\alpha}, x) q(\beta \mid \boldsymbol{\alpha}, \sigma^2, x) q(\sigma^2 \mid \boldsymbol{\alpha}, \beta, x). \tag{12}$$

In other words, we use the product of all full conditional distributions under the assumption of a fixed τ^2.

In the BTVC-AR(1) model we use this approach in a first step to draw from the joint posterior distribution $p(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2 \mid \boldsymbol{x})$. A detailed description of the sampling routine can be found in Appendix B. In a second and final step, we use these samples to generate paths of the x-process as follows:

$$x_{T+j}^{(m)} = \alpha_{T+j}^{(m)} + \beta^{(m)} x_{T+j-1}^{(m)} + \epsilon_{T+j}, \qquad j > 0,$$

where m denotes the simulated path. In Appendix C we give some insights about the performance and further details about the MCMC algorithm applied to our dataset.

3. Application To Interest Rate Data

We now apply the BTVC-AR(1) model to the first principal component (PC) of a principal component analysis (PCA) on interest rate data to predict the term structure of interest rates and compare it to the *2-Additive-Factor Gaussian (Gauss2++) model* (Brigo and Mercurio, 2007) and the *dynamic Nelson-Siegel model* (Diebold and Li, 2006) with respect to the forecasting performance and the long run distribution.

3.1. Motivation and Background

The Gauss2++ model is a popular short-rate model in the insurance industry, used, e.g., to classify certified pension contracts into risk classes. Because its mean reversion level is calibrated to external interest rate forecasts, it generates realistic interest rates in the long horizon, which is a necessary model feature for insurance companies, as they are obliged to calculate risk measures and performance scenarios for specific insurance contracts for up to 40 years (European Union, 2017). Nevertheless, Diebold and Li (2006) point out that short-rate models perform poorly in forecasting. Their dynamic Nelson-Siegel model shows a better forecasting performance than the Gauss2++ model in the short horizon, but can produce unrealistic interest rates in the very long horizon. Our model, which we call the *BTVC-AR(1)-Factor model* in the following as it applies the BTVC-AR(1) model to the first PC of a PCA, combines both: a good forecasting performance in the short horizon and realistic interest rates in the long horizon. It further accounts for the strong autocorrelation and the (close to) random walk behavior of interest rates.

3.2. Data

We use data of the German term structure of interest rates estimated by the Deutsche Bundesbank from prices of German government bonds. The exact estimation procedure can be found in Schich (1997). The time span ranges from September 1997 to August 2016. Figure 2 shows the monthly evolution of the interest rate curves.

In the last ten to fifteen years a decrease of the interest rates can be observed. Each maturity represents a dimension in the data set. We use PCA to reduce the dimension of the data set for the following reason. According to Litterman and Scheinkman (1991) a three factor model can explain for each interest rate with a specific maturity a minimum of 96% of the variability in the data. We here extract these (principle) factors but only use the first two to facilitate a fair comparison with the Gauss2++ model, which is a two factor model. Furthermore, the first two PCs already account for more than 99% of the variability in the given data. Figure 3 shows the loadings and the time series of the two extracted PCs.

The loadings of the first PC are similar for all 20 maturities, while the loadings of the second PC are positive for short and negative for long maturities. The first and the second PC are therefore often interpreted as level and slope of the term structure, respectively.

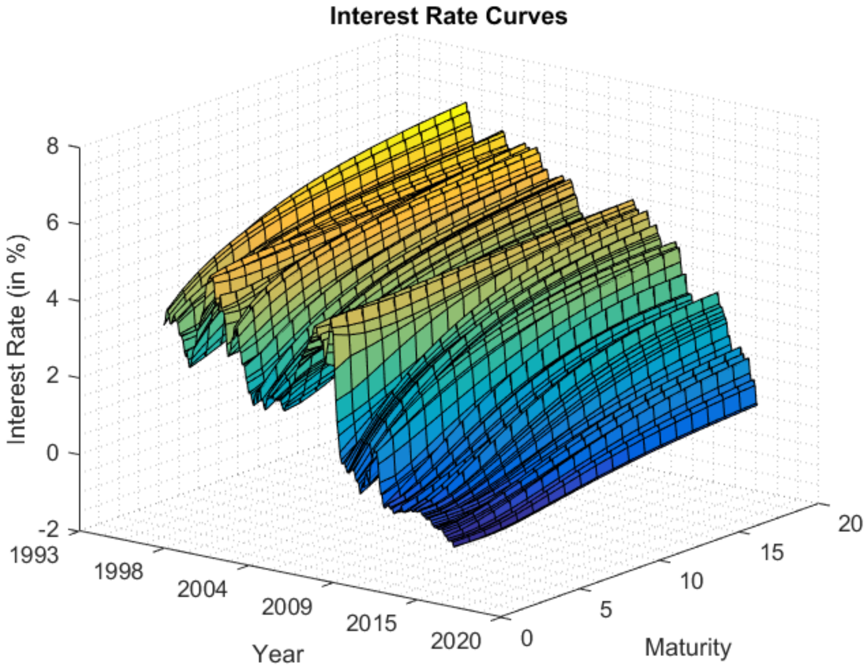


Figure 2: Time series of the term structure of German government bond yields.

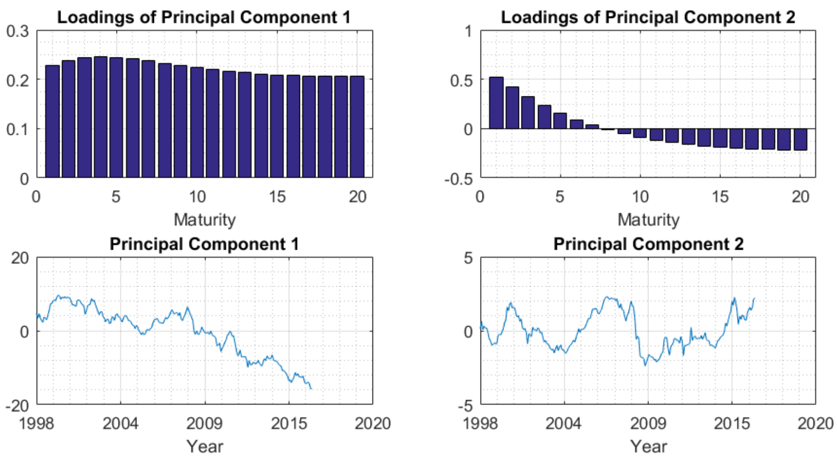


Figure 3: The scores and the loadings of the first two PCs.

The decrease of the interest rates in the last years is also visible in the level factor, showing a downward trend. There is an ongoing discussion in the literature about mean reversion of interest rates. Economic theory predominantly assumes that interest rates are (in the long run) mean reverting. But statistical evidence is not so clear (van den End, 2011). The mainstream literature says that unit roots can not be rejected, which would imply that interest rates are not mean reverting (Rose, 1988, Stock and Watson, 1988, Campbell and Shiller, 1991, Siklos and Wohar, 1997). More recent literature investigates the unit root hypothesis by fractional integrated techniques that apply differencing to time series by an order smaller than or greater than one (Baum et al., 2000, Gil-Alana, 2004). These studies find that shocks to interest rates have a long memory, which explains their (close to) random walk behavior.

3.3. Estimation of Model Parameters

In this subsection the estimation of the BTVC-AR(1)-Factor model and the two benchmark models is described.

3.3.1. Modeling Interest Rates with the BTVC-AR(1)-Factor Model

The factors of our BTVC-AR(1)-Factor model are the first two PCs extracted by a PCA and interpreted as level and slope of the interest rate curve. The level factor shows a (close to) random walk behavior, which can not be adequately captured by a stationary linear model. Following the economic theory view that interest rates (and therefore also the level) are mean reverting (in the long run) and assuming that the random walk behavior results from changes in the mean reversion level, we use therefore the BTVC-AR(1) model for this PC. It allows us to account for the (close to) random walk behavior as well as to regularize the level of the interest rate curve in the long horizon via prior assumptions. The slope factor is more stable over time. As an augmented Dickey Fuller test suggests that the existence of a unit root can be rejected, a linear AR(1) model is used for this factor. By modeling the level and the slope factor interest rate forecasts $\hat{r}_t(\tau)$ with maturity τ can be calculated via

$$\hat{r}_t(\tau) = \mu(\tau) + \xi_1(\tau)\hat{l}_t + \xi_2(\tau)\hat{s}_t, \tag{13}$$

where $\hat{l}_t$ and $\hat{s}_t$ denote the forecasts of the level and the slope factor, respectively. $\xi_1(\tau)$ and $\xi_2(\tau)$ denote the loading of the first and second PC for maturity τ. Before applying the PCA the data has been centered and therefore $\mu(\tau)$ is the mean interest rate of the data set for maturity τ. We now specify the prior assumptions of the BTVC-AR(1) model for the level factor and the estimation procedure of the AR(1) model for the slope factor.

The Level Factor

Latent AR1 parameter $\boldsymbol{\alpha}$. For this application we assume an AR-covariance structure for the α-process of the BTVC-AR(1) model with the parameters ρ and τ^2 representing the correlation of two successive time points and the conditional variance, respectively. The unconditional mean of the α-process is set to 0, which implies the assumption that the long run mean of the level factor is 0. Because we also assume that the slope factor is a centered process this means that the long run interest rate curve converges in expectation to the average interest rate curve of the dataset.
Autocorrelation parameter $\boldsymbol{\rho}$. As specified in Section 2.2.2 we assume for ρ a truncated normal distribution with the parameters $\mu_\rho = 0.98$ and $\sigma_\rho^2 = 0.001^2$ with lower truncation -1 and and upper truncation 1 as a hyper prior, i.e.,

$$\rho \sim \mathcal{N}(0.98, 0.001, -1, 1)$$

The truncation ensures the stationarity of the process. The parameters of this hyperprior rely on expert judgment and incorporate the assumption of a weak mean reverting α-process into the model and therefore allow the mean reversion level of the level factor to deviate from the long run mean for longer periods. This yields the (close to) random walk behavior present in (our) interest rate data.
Variance of the latent process. According to Section 2.2.2 the parameter τ^2 is set in each iteration of the sampling procedure such that the long run variance of the level factor amounts to a prespecified value. We here use the value 120, which is inferred from a quantile of the unconditional distribution. By giving consideration of the rather unusual market situation of extremely low interest rates we make the assumption that the last observation is equal to the 7.5%-quantile. Due to the model assumptions, the unconditional distribution is normal with mean 0 and the corresponding unconditional variance can be calculated easily.
Slope parameter of the AR(1) model. For β we assume that $\mu_\beta = 0.95$ and $\sigma_\beta^2 = 0.015^2$. This expert judgment represents a weak mean reversion to the time-

varying mean reversion level. The lower and upper truncation of the truncated normal distribution amount to -1 and $\sqrt{\frac{\text{Var}(x_t)-\sigma^2}{\text{Var}(x_t)}}$ to ensure the stationarity of the model as well as the positivity of τ^2, i.e.,

$$\beta|\sigma^2 \sim \mathcal{N}\left(0.95, \sigma^2 0.015^2, -1, \sqrt{\frac{\text{Var}(x_t) - \sigma^2}{\text{Var}(x_t)}}\right)$$

Residual variance. For the prior distribution of σ^2 the shape and scale parameter a and b are set to 0.5 and 2 respectively, representing an uninformative prior.

By specifying the parameters of the prior (and hyper-prior) distributions the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$ as well as the conditional distributions of the other parameters can be analytically derived as described in Section 2.3. Combining the Gibbs Sampler and the Metropolis-Hastings algorithm as explained in Section 2.4, paths of the level factor can be generated. Forecasts of the level factor are then represented by the average of the simulated paths.

The Slope Factor

The linear AR(1) model for the slope factor is given by

$$s_t = c + \gamma s_{t-1} + \eta_t,$$

where γ is a real valued constant between -1 and 1 and η_t is a Gaussian white noise process, i.e., $\eta_t \sim \mathcal{N}(0, \breve{\sigma}^2)$. The constant parameter c is set to 0. The other parameters are estimated by a standard ordinary least squares approach.

3.3.2. Modeling Interest Rates With the Gauss2++ Model

The Gauss2++ model – in a different representation also known as the 2-Factor-Hull-White model – is a popular interest rate model in the insurance industry used for pricing interest rate derivatives as well as for risk management and forecasting purposes. The model assumes that the short-rate $r(t)$, which is the interest rate with an infinitesimal small maturity, is given by the sum of two latent processes $(x(t))_{t\geq 0}$

and $(y(t))_{t\geq 0}$, and a deterministic function φ:

$$r(t) = x(t) + y(t) + \varphi(t).$$

The latent processes are modeled by dependent Ornstein-Uhlenbeck processes, which are the continuous version of a linear AR(1) process. Interest rates with longer maturities are then derived from the short-rate via pricing the corresponding zero-coupon bonds, which is analytically possible due to the model's distributional assumptions. The estimation process is materially different from the one of the other two models as it does not use historic data but calibrates the model to current future market assumptions (implicitly) provided by the current interest rate curve, interest rate derivatives as well as interest rate forecasts. By applying the downhill simplex algorithm the parameters of the model are chosen in such a way that forward rates – implicitly given by the current interest rate curve – and swaption prices are met in expectation. The relevant data has been extracted from Bloomberg. Additionally the mean reversion level of the two latent factors are analytically set such that two interest rate forecasts with a maturity of 3 months and 10 years, which are published by the Organisation for Economic Co-operation and Development (OECD), are met in expectation. This approach is in line with the standard calibration procedure in the insurance industry (Korn and Wagner, 2019).

3.3.3. Modeling Interest Rates With the Dynamic Nelson-Siegel Model

The dynamic Nelson-Siegel model of Diebold and Li applies specific time series models to extracted latent factors. Diebold and Li tested several time series models on the level, slope and curvature factors of the Nelson-Siegel interest rate curve and compared the forecasting performance Diebold and Li (2006). In this paper we follow one of their approaches, in which they apply a PCA on interest rate data and use an univariate linear AR(1) process for each of the first three PCs. Because of comparison reasons to the other two two-factor models in this paper, we just use the first two PCs. The parameters of the AR(1) model are estimated by the ordinary least squares method.

3.4. Backtest

We now compare the forecasting performance of the BTVC-AR(1)-Factor model, the Gauss2++ model and the dynamic Nelson-Siegel model and analyse their long run distributions of the 10-year interest rate.

3.4.1. Comparison of the Forecasting Performance

For the out-of-sample backtest we apply an expanding window approach. The data of the first 10 years of the observations are used to estimate the parameters of the BTVC-AR(1)-Factor model and the dynamic Nelson-Siegel model as described in the Section 3.3. The Gauss2++ model is calibrated to the current market data. We then forecast the interest rates for the maturities of 1, 3, 5 and 10 years (representing the yield curve) for the horizons of 1, 3, 6 and 12 months. We expand the training sample by one month and repeat the procedure again. This is done until 12 months before the last observation in the data set. To evaluate the forecasting performance the error between the predicted interest rate $\hat{r}_\tau(t)$ and the actual interest rate $r_\tau(t)$ with the maturity τ is calculated, i.e.,

$$error_\tau(t) = r_\tau(t) - \hat{r}_\tau(t).$$

Table (D.2)-(D.5) in the Appendix D show the mean and the standard deviation of this error for each model. In addition, the root mean squared error

$$RMSE(\tau) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(r_\tau(k) - \hat{r}_\tau(k))^2} \tag{14}$$

for the given deviation is calculated, where N is the number of forecasts conducted in the backtest.

The RMSE for the 1-month ahead forecasts is similar for all three models. For longer forecasting horizons the Gauss2++ model shows the highest RMSE. For example, the 6-month ahead forecast of the 10-year interest rate of the Gauss2++ model has a RMSE approximately twice as high as the RMSE of the other two models and more than three times as high for the 12-month ahead forecast. This supports the statement of Diebold and Li (2006) that short-rate models perform poorly in forecasting. However, it should be mentioned that the performance of the Gauss2++ model highly depends on the interest rate forecasts used in the calibration process. Regarding the predominant negative mean error suggests that the OECD forecasts have been too optimistic in the past.

The results of the BTVC-AR(1)-Factor model and the dynamic Nelson-Siegel model are more consistent. For the forecasting horizon of 1-month the BTVC-AR(1)-Factor model shows a slightly lower RMSE except for the 10-year interest rate. For the 3-months, 6-months and 12-months forecasting horizons the BTVC-AR(1) model shows a lower RMSE for the short maturities, but a higher RMSE for the longer maturities compared to the dynamic Nelson-Siegel model. Note that the dynamic Nelson-Siegel model anticipated the downward trend present in the last years, which might have been beneficial in terms of the forecasting performance in the past, but also produces unrealistic interest rates in the long horizon. In contrast the BTVC-AR(1)-Factor model forces the model to mean revert to a prespecified level to regularize the interest rates in the long horizon. It can therefore follow the current trend only for a couple of time steps, which might explain the slightly worse performance for the 6- and 12-months forecasting horizon. The fact that the RMSE error is still similar to the dynamic Nelson-Siegel model suggests that this does not affect the forecasting performance in the short horizon much.

3.4.2. Comparison of the Distribution in the Long Run

We further investigate the interest rate distribution in the long horizon. This is especially important for insurance companies as risk measures and performance scenarios for their products have to be calculated for up to 40 years (European Union, 2017). We therefore fit all three models on all data points up to the last observation date of the data set. We then simulate paths of the 10-year interest rate and visualize the distribution in 40 years. The median of the dynamic Nelson-Siegel model amounts to approximately -10%. A value that is not realistic for the 10-year interest rate. In comparison, the distribution of the BTVC-AR(1)-Factor model and the Gauss2++ model seem to be more realistic as the range of their distributions is (mainly) positive between 0% and 10%. It can be observed that the standard deviation of the Gauss2++ model is much smaller than of the BTVC-AR(1)-Factor model and as the median is quite high negative values are not reached by this model. This is due to the fact that the Gauss2++ model assumes a stronger mean reversion than historic data would suggest. The (close to) random walk behavior is better captured by the BTVC-AR(1)-Factor model leading to a prediction range which fits historical observations quite well. This is due to the regularization of the mean and the standard deviation of the BTVC-AR(1)-Factor model induced by appropriate prior assumptions, which

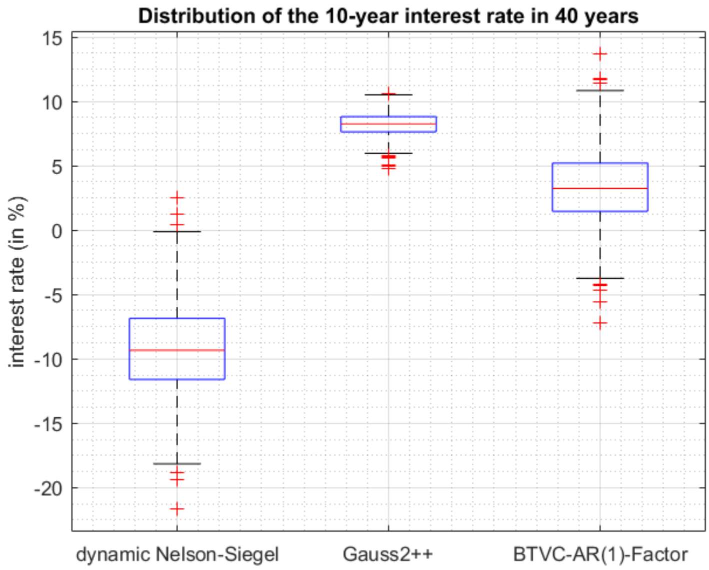


Figure 4: Comparison of the distributions of the 10-year interest rate in 40 years modeled by the dynamic Nelson-Siegel model, the Gauss2++ model and the BTVC-AR(1)-Factor model.

represents the main difference to other interest rate models.

4. Conclusion

In this paper we introduced a new Bayesian framework for the TVC-AR(1) model particularly suitable for nearly integrated time series which can not be estimated by a linear model consistent with economic theory or historical observations. In these cases a (close to) random walk behavior can be an indication for a missing variable, for which we account for by the usage of a non-linear model. The time-varying constant of the BTVC-AR(1) allows a stochastic mean reversion level leading to realizations, which exhibit a random walk behavior although being stationary and do not have an exploding long run variance. Additionally, with the Bayesian approach it is possible to incorporate prior assumption about the long run distribution into the model without affecting the short-term predictions adversely. This gives the possibility to include expert knowledge or well known economic facts about the long-term behavior of the time series into the model that is otherwise fully data-driven in the short term forecast.

We apply the proposed approach to interest rate data. We find that the BTVC-AR(1)-Factor model, which applies a BTVC-AR(1) model to the first PC of a PCA, shows a similar forecasting performance as the dynamic Nelson-Siegel model in the short horizon but in contrast produces realistic interest rates in the very long horizon and also yields better forecasts compared to the Gauss2++ model.

The presented framework allows for many different specifications and is, in particular, flexible in terms of the assumed covariance structure of the latent α process in the model. In this paper we propose an AR-covariance structure and explain how model parameters can be inferred in this special case. Investigating other covariance structures may further improve the forecasting performance in the short horizon while still regularizing the distribution in the long run.

Data Availability Statement

The data that support the findings of this study are available at `https://www.bundesbank.de/de/statistiken/zeitreihen-datenbanken`. These data were derived from the following resources available in the public domain: BBK01.WZ9801, BBK01.WZ9802, BBK01.WZ9803, BBK01.WZ9804, BBK01.WZ9805, BBK01.WZ9806.

References

David W. Bacon and Donald G. Watts. Estimating the transition between two intersecting straight lines. *Biometrika*, 58(3):525–534, 1971.

Christopher F. Baum, Basma Bekdache, et al. A re-evaluation of empirical tests of the fisher hypothesis. *Working Papers in Economics*, page 148, 2000.

Geert Bekaert, Seonghoon Cho, and Antonio Moreno. New keynesian macroeconomics and the term structure. *Journal of Money, Credit and Banking*, 42(1):33–62, 2010.

Damiano Brigo and Fabio Mercurio. *Interest rate models – theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.

João Caldeira and Hudson Torrent. Forecasting the us term structure of interest rates using nonparametric functional data analysis. *Journal of Forecasting*, 36(1):56–73, 2017.

John Y. Campbell and Robert J. Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Kung Sik Chan and Howell Tong. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7(3):179–190, 1986.

Siddhartha Chib. Markov chain monte carlo methods: computation and inference. In *Handbook of Econometrics*, volume 5, pages 3569–3649. Elsevier, 2001.

Anna Cieslak and Pavol Povala. Expected returns in treasury bonds. *The Review of Financial Studies*, 28(10):2859–2901, 2015.

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74 (366a):427–431, 1979.

Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

Gregory R. Duffee. Forecasting with the term structure: The role of no-arbitrage restrictions. Technical report, Working Paper, 2011.

European Union. Commission delegated regulation (eu) 2017/653 of 8 march 2017 supplementing regulation (eu) no 1286/2014 of the european parliament and of the council on key information documents for packaged retail and insurance-based investment products (priips) by laying down regulatory technical standards with regard to the presentation, content, review and revision of key information documents and the conditions for fulfilling the requirement to provide such documents. *Official Journal of the European Union*, 60, 2017.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Luis A. Gil-Alana. Long memory in the us interest rate. *International Review of Financial Analysis*, 13(3):265–276, 2004.

Stephen Goldfeld and Richard Quandt. The estimation of structural shifts by switching regressions. In *Annals of Economic and Social Measurement, Volume 2, number 4*, pages 475–485. NBER, 1973.

C. W. J. Granger and A. P. Andersen. *An introduction to bilinear time series models*. Vandenhoeck and Ruprecht: Göttingen, 1978.

Valérie Haggan and Tohru Ozaki. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68(1):189–196, 1981.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

Eilev S. Jansen and Timo Teräsvirta. Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, 58 (4):735–763, 1996.

Ralf Korn and Andreas Wagner. *Praxishandbuch Lebensversicherungsmathematik : Simulation und Klassifikation von Produktent*. VVW GmbH, 2019.

Sharon Kozicki and Peter A. Tinsley. Shifting endpoints in the term structure of interest rates. *Journal of Monetary Economics*, 47(3):613–652, 2001.

Markku Lanne and Pentti Saikkonen. Threshold autoregressions for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 20(2):282–289, 2002.

K. S. Lim and H. Tong. Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Sociaty, B*, 42:245–92, 1980.

Robert Litterman and Jose Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61, 1991.

Russell B. Millar and Renate Meyer. Non-linear state space modelling of fisheries biomass dynamics by using metropolis-hastings within-gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(3):327–342, 2000.

James C. Morley. Nonlinear time series in macroeconomics. *Encyclopedia of Complexity and System Science, forthcoming*, 2009.

Richard E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53 (284):873–880, 1958.

T. Subba Rao and M. M. Gabr. *An introduction to bispectral analysis and bilinear time series models*, volume 24. Springer Science & Business Media, 2012.

Andrew K. Rose. Is the real interest rate stable? *The Journal of Finance*, 43(5): 1095–1112, 1988.

Glenn D. Rudebusch and Tao Wu. A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal*, 118(530):906–926, 2008.

Sebastian Schich. Schätzung der deutschen zinsstrukturkurve. *Bundesbank Series 1 Discussion Paper*, 1997.

Pierre L. Siklos and Mark E. Wohar. Convergence in interest rates and inflation rates across countries and over time. *Review of International Economics*, 5(1):129–141, 1997.

James H. Stock and Mark W. Watson. Testing for common trends. *Journal of the American Statistical Association*, 83(404):1097–1107, 1988.

Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

Timo Teräsvirta, Dag Tjøstheim, C. W. J. Granger, et al. *Modelling nonlinear economic time series*. Oxford University Press Oxford, 2010.

Jan Willem van den End. Statistical evidence on the mean reversion of interest rates. *De Nederlandsche Bank Working Paper*, 2011.

Dick Van Dijk, Siem Jan Koopman, Michel Van der Wel, and Jonathan H. Wright. Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, 29(5):693–712, 2014.

Appendix A. Full Conditional Distributions

Appendix A.1. The Full Conditional Distribution of $\widetilde{\boldsymbol{\alpha}}$

The prior distribution of $\widetilde{\boldsymbol{\alpha}}$ is a centered Gaussian process with a specific covariance structure $\widetilde{\boldsymbol{\Sigma}}$, i.e.,

$$\widetilde{\boldsymbol{\alpha}} = (\alpha_1, ..., \alpha_T, ..., \alpha_{T+h}) \sim \mathcal{N}_T(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$$

The following derivations will be independent of the specific choice of $\widetilde{\boldsymbol{\Sigma}}$. By defining

$$\Delta_j = x_{j+1} - \beta x_j$$

as well as $\boldsymbol{\Delta} = (\Delta_0, \ldots, \Delta_{T-1})^\top$ and the fact that

$$\phi(x_t|\alpha_t + \beta x_{t-1}, \sigma^2) = \phi(\alpha_t|\Delta_{t-1}, \sigma^2)$$

allows a straightforward derivation of the full conditional of $\widetilde{\boldsymbol{\alpha}}$:

$$\begin{aligned}
p(\widetilde{\boldsymbol{\alpha}}|\beta, \sigma^2, \boldsymbol{x}) &\propto p(\boldsymbol{x}|\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2) p(\widetilde{\boldsymbol{\alpha}}|\beta, \sigma^2) \\
&\propto p(\boldsymbol{x}|\boldsymbol{\alpha}, \beta, \sigma^2) p(\widetilde{\boldsymbol{\alpha}}) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\Delta})^\top(\boldsymbol{\alpha} - \boldsymbol{\Delta})\right) \cdot \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{\alpha}}^\top \widetilde{\boldsymbol{\Sigma}}^{\mathbf{-1}} \widetilde{\boldsymbol{\alpha}}\right) \\
&\propto \exp\left(-\frac{1}{2}(\widetilde{\boldsymbol{\alpha}}^\top \widetilde{\boldsymbol{\Sigma}}^{\mathbf{-1}}_{\boldsymbol{post}} \widetilde{\boldsymbol{\alpha}} - 2\widetilde{\boldsymbol{\alpha}}^\top \widetilde{\boldsymbol{\Sigma}}^{\mathbf{-1}}_{\boldsymbol{post}} \underbrace{\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}} \widetilde{\boldsymbol{\Delta}}_{\mathbf{0}} \frac{1}{\sigma^2}}_{=:\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}})\right)
\end{aligned}$$

with $\widetilde{\boldsymbol{\Sigma}}^{\mathbf{-1}}_{\boldsymbol{post}} = \widetilde{\boldsymbol{\Sigma}}^{\mathbf{-1}} + \frac{1}{\sigma^2}\begin{pmatrix} \boldsymbol{I_t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ and $\widetilde{\boldsymbol{\Delta}}_{\mathbf{0}} = (\boldsymbol{\Delta}^\top, \mathbf{0})^\top$.

This is the kernel of a multivariate Gaussian distribution with covariance $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}$ and mean vector $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}$, i.e

$$\widetilde{\boldsymbol{\alpha}} \mid \beta, \sigma^2, \boldsymbol{x} \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}, \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}).$$

Appendix A.2. The Full Conditional Distribution of $\boldsymbol{\rho}$

If an AR-covariance structure is assumed for $\widetilde{\boldsymbol{\alpha}}$ the latent α-process can be written in the following form

$$\alpha_t = \rho\alpha_{t-1} + \eta_t,$$

where ρ determines the correlation between two successive time steps and η_t is a Gaussian white noise process, i.e., $\eta_t \sim \mathcal{N}(0, \tau^2)$.

The full conditional distribution of ρ can be therefore derived as follows:

$$p(\rho|\tau^2, \boldsymbol{\alpha}) \propto \mathcal{L}(\rho, \tau^2) \cdot p(\rho) = \prod_{j=0}^{T-1} \phi(\alpha_{T-j}|\rho\alpha_{T-j-1}, \tau^2) \cdot p(\rho). \tag{A.1}$$

The likelihood $\mathcal{L}(\cdot)$ in the above equation can be reformulated as

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\tau^2}\left\{-2\rho\left[\sum_{j=0}^{T-1}\alpha_{T-j}\alpha_{T-j-1}\right] + \rho^2\left[\sum_{j=0}^{T-1}\alpha^2_{T-j-1}\right]\right\}\right).$$

The calculation is similar to the one in appendix Appendix A.5. Defining the two terms in square brackets as η and χ, respectively, we get

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\tau^2}\left\{-2\rho\eta + \rho^2\chi\right\}\right).$$

Plugging this into (A.1) and using a normal prior with parameters μ_ρ, σ^2_ρ for ρ, we have

$$\begin{aligned} p(\rho|\tau^2, \boldsymbol{\alpha}) &\propto \exp\left(-\frac{1}{2}\left\{\frac{\rho^2\chi}{\tau^2} - 2\frac{\rho\eta}{\tau^2}\right\}\right)\exp\left(-\frac{1}{2}\left\{\frac{\rho^2}{\sigma^2_\rho} - 2\frac{\rho\mu_\rho}{\sigma^2_\rho}\right\}\right) \\ &\propto \exp\left(-\frac{1}{2}\left\{\rho^2\cdot\left(\frac{\chi}{\tau^2} + \sigma_\rho^{-2}\right) - 2\rho\left(\frac{\eta}{\tau^2} + \frac{\mu_\rho}{\sigma^2_\rho}\right)\right\}\right) \end{aligned}$$

and thus $\rho|\tau^2, \boldsymbol{\alpha} \sim \mathcal{N}(\mu_{\rho,post}, \sigma^2_{\rho,post})$ with

$$\sigma^2_{\rho,post} = \left(\frac{\chi}{\tau^2} + \sigma_\rho^{-2}\right)^{-1}$$

and

$$\mu_{\rho,post} = \left(\frac{\eta}{\tau^2} + \frac{\mu_\rho}{\sigma^2_\rho}\right)\sigma^2_{\rho,post}.$$

If a truncated normal prior is used, the truncation is transferred to the full conditional distribution.

Appendix A.3. The Full Conditional Distribution of $\boldsymbol{\beta}$

Analogously to (10) and (11) we have

$$p(\beta|\boldsymbol{\alpha},\sigma^2,\boldsymbol{x}) \propto \mathcal{L}(\beta,\boldsymbol{\alpha},\sigma^2)\cdot p(\beta) = \prod_{j=0}^{T-1}\phi(x_{T-j}|\alpha_{T-j}+\beta x_{T-j-1},\sigma^2)\cdot p(\beta). \quad \text{(A.2)}$$

By defining $\breve{d}_{t-j} := x_{t-j} - \alpha_{t-j}$ and as

$$\phi(x_{t-j}|\alpha_{t-j}+\beta x_{t-j-1},\sigma^2) = \phi(\beta x_{t-j-1}|\breve{d}_{t-j},\sigma^2)$$

the likelihood $\mathcal{L}(\cdot)$ in the above equation can be reformulated as

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta\left[\sum_{j=0}^{T-1}\breve{d}_{T-j}x_{T-j-1}\right]+\beta^2\left[\sum_{j=0}^{T-1}x_{T-j-1}^2\right]\right\}\right).$$

You can find a more detailed calculation in Appendix A.5. Defining the two terms in square brackets as η and χ, respectively, we get

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta\eta+\beta^2\chi\right\}\right).$$

Plugging this into (A.2) and using a normal prior with parameters $\mu_\beta, \sigma_\beta^2$ for β, we have

$$\begin{aligned} p(\beta|\boldsymbol{\alpha},\sigma^2,\boldsymbol{x}) &\propto \exp\left(-\frac{1}{2}\left\{\frac{\beta^2\chi}{\sigma^2}-2\frac{\beta\eta}{\sigma^2}\right\}\right)\exp\left(-\frac{1}{2}\left\{\frac{\beta^2}{\sigma_\beta^2}-2\frac{\beta\mu_\beta}{\sigma_\beta^2}\right\}\right) \\ &\propto \exp\left(-\frac{1}{2}\left\{\beta^2\cdot\left(\frac{\chi}{\sigma^2}+\sigma_\beta^{-2}\right)-2\beta\left(\frac{\eta}{\sigma^2}+\frac{\mu_\beta}{\sigma_\beta^2}\right)\right\}\right) \end{aligned}$$

and thus $\beta|\boldsymbol{x},\alpha,\sigma^2 \sim \mathcal{N}(\mu_{\beta,post},\sigma^2_{\beta,post})$ with

$$\sigma^2_{\beta,post} = \left(\frac{\chi}{\sigma^2}+\sigma_\beta^{-2}\right)^{-1}$$

and

$$\mu_{\beta,post} = \left(\frac{\eta}{\sigma^2} + \frac{\mu_\beta}{\sigma_\beta^2}\right)\sigma^2_{\beta,post}.$$

If a truncated normal prior is used, the truncation is transferred to the full conditional distribution.

Appendix A.4. The Full Conditional Distribution of $\boldsymbol{\sigma^2}$

In this Section we derive the full conditional distribution of σ^2. As before

$$p(\sigma^2|\boldsymbol{\alpha},\beta,\boldsymbol{x}) \propto \prod_{j=0}^{T-1}\phi(x_{T-j}|\alpha_{T-j}+\beta x_{T-j-1},\sigma^2)\cdot p(\sigma^2)\cdot p(\beta\mid\sigma^2),$$

which is equal to

$$(\sigma^2)^{-\frac{t}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{j=0}^{T-1}\epsilon^2_{T-j}\right)\cdot p(\sigma^2)\cdot p(\beta\mid\sigma^2) =: (\sigma^2)^{-\frac{t}{2}}\exp\left(-\frac{1}{2\sigma^2}\kappa\right)\cdot p(\sigma^2)\cdot p(\beta\mid\sigma^2).$$

By using an inverse gamma distribution with shape and scale parameters a, b, or short $\mathcal{IG}(a,b)$, for the prior of σ^2 we get

$$\begin{aligned}
&(\sigma^2)^{-(t/2)}\exp\left(-\frac{1}{2\sigma^2}\kappa\right)\cdot(\sigma^2)^{-(a+1)}\exp(-b/\sigma^2)\\
&\qquad\cdot(\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\sigma_\beta^2}(\beta-\mu_\beta)^2\right)\frac{1}{\Phi(\frac{1-\mu_\beta}{\sigma\sigma_\beta})-\Phi(\frac{-1-\mu_\beta}{\sigma\sigma_\beta})} =\\
&(\sigma^2)^{-(t/2)}\exp\left(-\frac{1}{2\sigma^2}\kappa\right)\cdot(\sigma^2)^{-\left(a+\frac{3}{2}\right)}\exp\left(-\frac{b+\frac{(\beta-\mu_\beta)^2}{2\sigma_\beta^2}}{\sigma^2}\right)\frac{1}{\Phi(\frac{1-\mu_\beta}{\sigma\sigma_\beta})-\Phi(\frac{-1-\mu_\beta}{\sigma\sigma_\beta})} \approx\\
&(\sigma^2)^{-\left(\frac{t+1}{2}+a+1\right)}\exp\left(-\frac{\frac{\kappa}{2}+b+\frac{(\beta-\mu_\beta)^2}{2\sigma_\beta^2}}{\sigma^2}\right).
\end{aligned}$$

In the last step we omitted the last term, which results from the truncation, as in our application the truncation is not very restrictive such that this term is close to 1. Thus the full conditional distribution is approximately also an inverse gamma distribution with parameters $\tilde{a} = \frac{t+1}{2} + a$ and $\tilde{b} = \frac{\kappa}{2} + b + \frac{(\beta-\mu_\beta)^2}{2\sigma_\beta^2}$, i.e.,

$$\sigma^2|\boldsymbol{\alpha},\beta,\boldsymbol{x} \sim \mathcal{IG}(\tilde{a},\tilde{b}).$$

Appendix A.5. Rewriting the Likelihood of the Parameters

By defining $\breve{d}_{t-j} := x_{t-j} - \alpha_{t-j}$, the likelihood of the parameters can be reformulated as follows:

$$
\begin{aligned}
\mathcal{L}(\beta, \boldsymbol{\alpha}, \sigma^2) &= \prod_{j=0}^{T-1} \phi(x_{T-j} | \alpha_{T-j} + \beta x_{T-j-1}, \sigma^2) \\
&= \prod_{j=0}^{T-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{T-j} - \alpha_{T-j} - \beta x_{T-j-1})^2}{2\sigma^2}\right) \\
&= \prod_{j=0}^{T-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\breve{d}_{T-j} - \beta x_{T-j-1})^2}{2\sigma^2}\right) \\
&= \prod_{j=0}^{T-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\breve{d}^2_{T-j} - 2\beta x_{T-j-1}\breve{d}_{T-j} + \beta^2 x^2_{T-j-1}}{2\sigma^2}\right) \\
&\propto \prod_{j=0}^{T-1} \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta x_{T-j-1}\breve{d}_{T-j} + \beta^2 x^2_{T-j-1}\right\}\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta \sum_{j=0}^{T-1} \breve{d}_{T-j} x_{T-j-1} + \beta^2 \sum_{j=0}^{T-1} x^2_{T-j-1}\right\}\right).
\end{aligned}
$$

Appendix B. Metropolis-Hastings within Gibbs Sampler Routine

Starting with an initial sample $(\alpha^{(0)}, \beta^{(0)}, (\sigma^2)^{(0)}, \rho^{(0)}, (\tau^2)^{(0)})$, where

$$(\tau^2)^{(0)} = f(\beta^{(0)}, (\sigma^2)^{(0)}, \rho^{(0)}, Var(x_t))$$

as specified in (8), we first draw a sample of $\widetilde{\boldsymbol{\alpha}}$ values from its full conditional distribution. We proceed with the Metropolis-Hastings algorithm step by drawing from the conditional distributions of ρ, σ^2 and β as derived in Section 2.3. Furthermore, τ^2 is set according to (6) such that a prior specified long run variance is met. We calculate the density value of the proposal distribution q specified in (12), i.e.,

$$
\begin{aligned}
&q(\rho^{(n+1)}, \beta^{(n+1)}, (\sigma^2)^{(n+1)} \mid (\tau^2)^{(n)}, \beta^{(n)}, (\sigma^2)^{(n)}, \boldsymbol{\alpha}, \boldsymbol{x}) = \\
&\qquad q(\rho^{(n+1)} \mid (\tau^2)^{(n)}, \boldsymbol{\alpha}, x) q((\sigma^2)^{(n+1)} \mid \beta^{(n)} \boldsymbol{\alpha}, \boldsymbol{x}) q(\beta^{(n+1)} \mid (\sigma^2)^{(n)}, \boldsymbol{\alpha}, \boldsymbol{x})
\end{aligned}
$$

We further calculate the density value of the proposal distribution for the parameters of the previous step conditional on the new drawn parameter, i.e.,

$$q(\rho^{(n)},\beta^{(n)},(\sigma^2)^{(n)} \mid (\tau^2)^{(n+1)},\beta^{(n+1)},(\sigma^2)^{(n+1)},\boldsymbol{\alpha},x) = \\ q(\rho^{(n)} \mid (\tau^2)^{(n+1)},\boldsymbol{\alpha},x)\cdot q((\sigma^2)^{(n)} \mid \beta^{(n+1)},\boldsymbol{\alpha},x)\cdot q(\beta^{(n)} \mid (\sigma^2)^{(n+1)},\boldsymbol{\alpha},x)$$

The true conditional posterior density is given by

$$p(\rho,\beta,\sigma^2 \mid \boldsymbol{\alpha},\boldsymbol{x}) \propto p(\boldsymbol{x} \mid \beta,\sigma^2,\boldsymbol{\alpha})p(\boldsymbol{\alpha} \mid \rho, f(\rho,\beta,\sigma^2))p(\rho)p(\beta \mid \sigma^2)p(\sigma^2)$$

The acceptance probability is calculated by

$$p_{accept.} = \min\left(1,\frac{p(\rho^{(n+1)},\beta^{(n+1)},(\sigma^2)^{(n+1)} \mid \boldsymbol{\alpha^{(n+1)}},\boldsymbol{x})}{p(\rho^{(n)},\beta^{(n)},(\sigma^2)^{(n)} \mid \boldsymbol{\alpha^{(n+1)}},\boldsymbol{x})}\cdot \right. \\ \left. \frac{q(\rho^{(n)},\beta^{(n)},(\sigma^2)^{(n)} \mid \rho^{(n+1)},\beta^{(n+1)},(\sigma^2)^{(n+1)},\boldsymbol{\alpha^{(n+1)}},\boldsymbol{x})}{q(\rho^{(n+1)},\beta^{(n+1)},(\sigma^2)^{(n+1)} \mid \rho^{(n)},\beta^{(n)},(\sigma^2)^{(n)},\boldsymbol{\alpha^{(n+1)}},\boldsymbol{x})}\right)$$

A new drawn sample is accepted if a uniform distributed random variable is smaller than the acceptance probability. Otherwise the sample from the previous step is taken. After a burn-in period the parameter set $(\widetilde{\boldsymbol{\alpha}}^{(m)},\beta^{(m)},(\sigma^2)^{(m)})$ is approximately distributed according to the joint posterior distribution $p(\widetilde{\boldsymbol{\alpha}},\beta,\sigma^2 \mid \boldsymbol{x})$.

Appendix C. Diagnostics of the MCMC algorithm

In this section we present diagnostics of the Metropolis-Hastings within Gibbs sampler (MHwGS) routine used for the dataset in our application.
To investigate the distribution of the 10-year interest rate at a forecasting horizon of 40 years, we use the prior distributions as specified in Section 3.3.1 and draw 10.000 samples by applying the MHwGS routine as described in Appendix B. The chosen proposal distribution results in an acceptance rate of 30.92%. After a burn-in phase of 100 samples, we apply thinning to reduce autocorrelation and take every 10th parameter set to simulated paths of the first principal component. The autocorrelation functions for selected parameters are are visualized in Figure C.5.

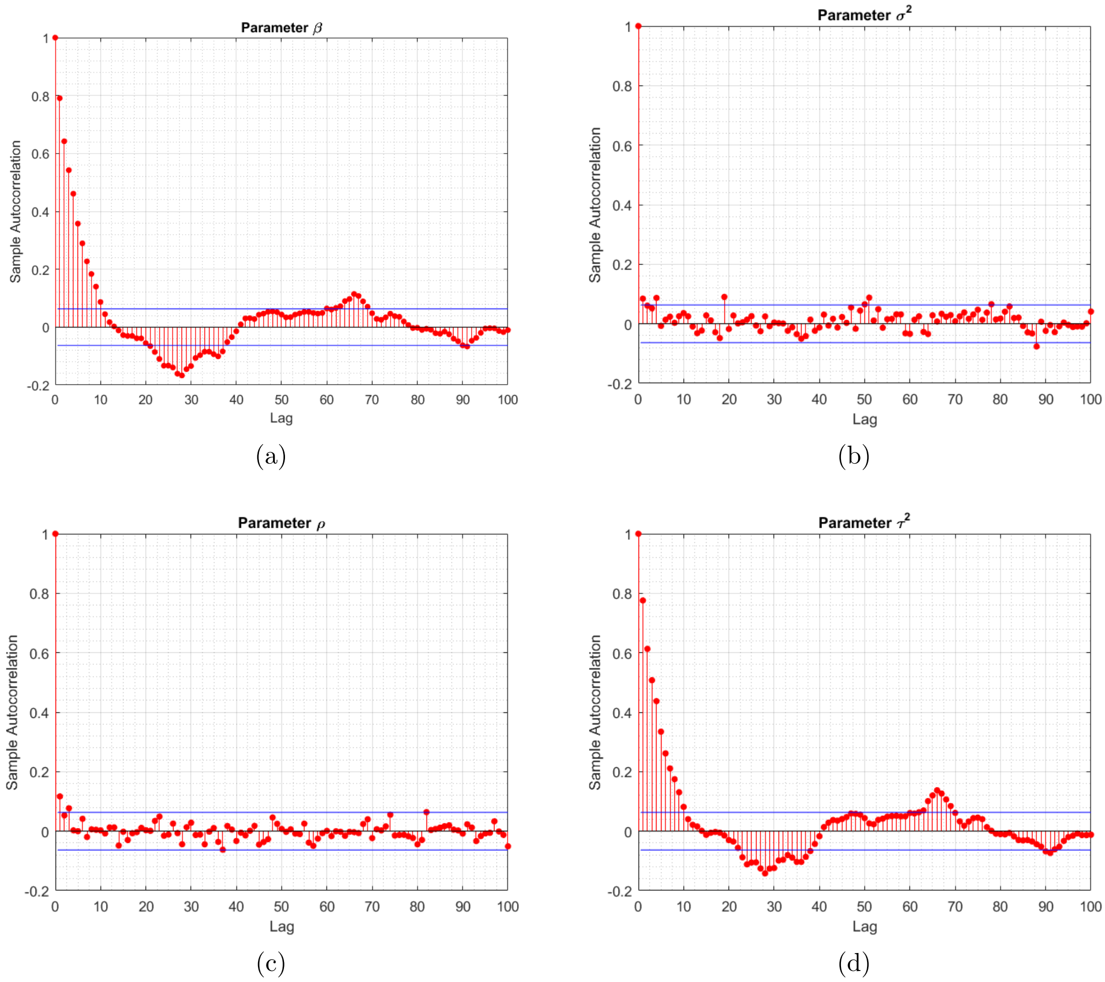


Figure C.5: Autocorrelation function of the paramters, which stay constant over time.

Appendix D. Backtest Results

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor model			
1 year	-0.0268	0.2566	0.0659
3 year	-0.0469	0.2289	0.0541
5 year	-0.0681	0.2402	0.0617
10 year	-0.0640	0.2346	0.0586
The Gauss2++ model			
1 year	-0.0808	0.2361	0.0618
3 year	-0.1037	0.2252	0.0610
5 year	-0.1203	0.2139	0.0598
10 year	-0.1429	0.2130	0.0654
The dynamic Nelson-Siegel model			
1 year	-0.0290	0.2615	0.0685
3 year	-0.0462	0.2311	0.0550
5 year	-0.0653	0.2410	0.0617
10 year	-0.0589	0.2340	0.0577

Table D.2: Results of the out-of-sample 1-month ahead forecasting.

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor model			
1 year	-0.1264	0.5064	0.2697
3 year	-0.1505	0.4640	0.2358
5 year	-0.1725	0.4354	0.2174
10 year	-0.1625	0.3875	0.1751
The Gauss2++ model			
1 year	-0.2057	0.5329	0.3236
3 year	-0.2707	0.4702	0.2923
5 year	-0.3098	0.4208	0.2714
10 year	-0.3435	0.3875	0.2667
The dynamic Nelson-Siegel model			
1 year	-0.1327	0.5152	0.2803
3 year	-0.1482	0.4665	0.2374
5 year	-0.1643	0.4343	0.2137
10 year	-0.1478	0.3827	0.1668

Table D.3: Results of the out-of-sample 3-month ahead forecasting.

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor model			
1 year	-0.2809	0.7683	0.6631
3 year	-0.3093	0.6941	0.5725
5 year	-0.3311	0.6330	0.5062
10 year	-0.3110	0.5462	0.3920
The Gauss2++ model			
1 year	-0.4094	0.8105	0.8184
3 year	-0.5402	0.6768	0.7457
5 year	-0.6098	0.6090	0.7393
10 year	-0.6545	0.5824	0.7693
The dynamic Nelson-Siegel model			
1 year	-0.2900	0.7857	0.6951
3 year	-0.3022	0.7045	0.5825
5 year	-0.3130	0.6380	0.5008
10 year	-0.2812	0.5446	0.3727

Table D.4: Results of the out-of-sample 6-month ahead forecasting.

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor model			
1 year	-0.5956	0.9591	1.2652
3 year	-0.6264	0.7861	1.0041
5 year	-0.6526	0.6834	0.8881
10 year	-0.6275	0.5986	0.7484
The Gauss2++ model			
1 year	-0.9047	1.0709	1.9546
3 year	-1.1531	0.7939	1.9541
5 year	-1.2745	0.7255	2.1458
10 year	-1.3345	0.8060	2.4246
The dynamic Nelson-Siegel model			
1 year	-0.6004	0.9961	1.3424
3 year	-0.6024	0.8218	1.0316
5 year	-0.6098	0.7096	0.8702
10 year	-0.5657	0.6024	0.6793

Table D.5: Results of the out-of-sample 12-month ahead forecasting.

IV. Risk-managed Collective Pension Schemes with Intergenerational Benefit Smoothing

This chapter is a reprint of:
C. Berninger and S. Mittnik. 2021. Risk-managed Collective Pension Schemes with Intergenerational Benefit Smoothing. *Available at SSRN:* http://ssrn.com/abstract=3885894:

Revision note: This chapter is a reprint of the revised manuscript available at SSRN since July 15, 2021.

Author Contributions

The origin of this research project was initialized by Stefan Mittnik, who gave the general idea and conceptualization. Christoph Berninger carried out the main work, including data curation, formal analysis, investigation, methodology, software, validation and visualization. The manuscript was drafted by Christoph Berninger and iteratively revised by Stefan Mittnik.

Risk–managed Collective Pension Schemes with Intergenerational Benefit Smoothing

Christoph Berninger[a], Stefan Mittnik[a]

[a]*Department of Statistics, LMU München*

Abstract

In view of the repeated severe market downturns since the turn of the century, the interest in risk–based investment strategies has grown in recent years. However, such strategies have not yet made major inroads into the design of pension programs. In this paper, we fill this gap by combining a risk–managed investment strategy with a pension scheme where benefits are smoothed across generations by establishing a collective reserve. We demonstrate that combining the two helps to improve the performance of the pension investments and decreases the risk of a negative reserve in times of a market crisis. We furthermore investigate the implications of imposing varying degrees of diversification across assets in such a scheme.

Keywords: collective defined–contribution plan, risk–sharing

---

1. Introduction

Having experienced a sequence of dramatic market downturns in the recent past, such as the dotcom crisis starting in 2000, the financial crisis around 2008, the European sovereign debt crisis unfolding in 2010, and most recently the Corona crash in 2020, there has been an increased interest in risk–based investment strategies. In these strategies, the asset allocation varies dynamically and is determined—at least in part—by the prevailing risk level of the assets. Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) showed that in the context of a momentum strategy, it is beneficial to invest less in an asset when its volatility is experiencing above normal levels and vice versa. Moreira and Muir (2017) demonstrated that a risk–managed investment strategy dominates a conventional buy–and–hold strategy.

There are various ways of constructing risk–managed portfolios, but they are all similar in spirit. Let r_t be the return of a portfolio representing a factor. Scaling r_t by a function of a risk measure, ϑ, results in the return of the managed portfolio

$$r_{\vartheta,t} = g(\vartheta_t) r_t, \tag{1}$$

where ϑ_t represents conditional risk. Function $g(\cdot)$ scales the investment such that a given risk target (e.g., the unconditional risk level) is met. Various risk measures have been applied in the literature. Common choices have been volatility (see, e.g., Barroso et al. (2017), Barroso and Santa-Clara (2015), Daniel and Moskowitz (2016), Eisdorfer and Misirli (2020)) or variance (see, e.g., Cederburg et al. (2020), Moreira and Muir (2017)). Also, the strategies for estimating the risk measure vary. Daniel and Moskowitz (2016) and Moreira and Muir (2017) use a parametric model, whereas Barroso and Santa-Clara (2015), Cederburg et al. (2020) and Daniel and Moskowitz (2016) use nonparametric sample estimates of realized volatility or realized variance. Regarding the choice of function $g(\cdot)$, all these studies specify $g(\cdot)$ to be proportional to the inverse of the risk measure employed, i.e.,

$$g(\vartheta) = \frac{c}{\vartheta},$$

where constant c is chosen such that the portfolio meets a given target risk.

In this paper, we investigate to what extent a collective pension scheme may benefit from a risk–managed investment strategy. Several concepts of collective pension schemes have been proposed in the literature and their outcomes compared to those of individual saving plans (see, for example, Gordon and Varian (1988), Krueger and Kubler (2006), Teulings and De Vries (2006), Ball and Mankiw (2007), Gollier (2008), Cui et al. (2011), Goecke (2013a), Bovenberg and Mehlkopf (2014) and Chen et al. (2016)). One feature of a collective pension scheme is the intergenerational risk transfer. Gordon and Varian (1988) introduce an optimal state–organized risk–sharing scheme that spans across future generations. Gollier (2008) proposes a model that smooths asset return volatility and allows for intergenerational risk sharing by introducing a collective reserve pool. He shows that such a scheme increases the expected utility of savers when compared to individual saving plans. See also Schumacher (2021) for a discussion of this scheme. Goecke (2013a) adopts this idea and proposes a specific decision mechanism that guides the dynamic fund management. Investment schemes involving intergenerational risk transfer have also been proposed in Grosen and Jørgensen (2000), Døskeland and Nordahl (2008), Hoevenaars and Ponds (2008), Baumann and Müller (2008), Westerhout (2011), and Bams et al. (2016).

In our analysis, we adopt the framework of a collective reserve, whose funds are not designated to any individual investor but rather serve as a collective buffer that smooths the individuals' returns across generations. In times where market returns fall below a given target return, investors are compensated by releases of the reserve. If they exceed the target return, the excess return is used to replenish the collective reserve. This approach smooths out volatile market returns and ensures that the investment of successive cohorts of investors performs similarly. In other words, the market risk is shared between generations of investors.

In this paper we investigate the properties of a risk–managed pension scheme with a collective reserve. In our framework, the dynamic asset allocation is directly tied to the projected market risk of the assets. This style of risk–based fund management contrasts with that in Goecke, which links risk exposure to the status of the reserve level. Focusing directly on risk has several advantages. First, it allows to directly control the risk of the fund—and it is precisely the goal to distribute risk fairly across generations. Second, due to their near–random–walk behavior, price movements on capital markets are hard to predict. Unlike capital market returns, market risks have a more or less clear dynamic structure that can be used for forecasting. Moreover, there is evidence of a dynamic interplay between risk and return. For example, at times when equity market risk is low or falling, returns tend to be high and vice versa. If this is the case, risk–based fund management can improve the the fund's risk–adjusted performance.

We also investigate the impact of imposing minimum–diversification requirements on the assets in the portfolio. We illustrate the proposed scheme by applying it to a portfolio consisting of two asset classes, namely equity, represented by the German DAX index, and government bonds, represented by the German REXP index. In doing so, we follow Goecke (2013a), which allows us to compare the reserve–based and the risk–based strategies.

Our results demonstrate that a risk–managed investment strategy can improve the absolute and the risk–adjusted performance of the pension fund. It also reduces the decline in the reserve when the market drops, as such declines are typically accompanied by higher market volatility—a phenomenon that our risk–based strategy explicitly takes into account.

The remainder of this paper is arranged as follows. In Section 2 we introduce the risk–managed investment strategy. Section 3 summarizes the reserve–driven approach in Goecke (2013a) and details our risk–driven framework. In Section 4 we empirically illustrate our approach using German market data covering the period 1967 to 2020 and compare the results of the two strategies. The final section concludes.

2. Risk–managed Investing

A risk–based investment strategy manages a portfolio dynamically according to the prevailing risk of the assets in the portfolio. Current risk can be measured in different ways. Conditional volatility or variance are commonly used candidates. Let r_t be the excess return of a portfolio and $\hat{\sigma}_t$ be an estimate of the portfolio's conditional volatility. We follow Moreira and Muir (2017) and obtain the risk–managed portfolio return, $r_{RA,t}$, by setting

$$r_{RA,t} = \frac{c_t}{\hat{\sigma}_t^2} r_t. \tag{2}$$

Moreira and Muir set $c_t = c$, where constant c is chosen such that the unconditional variance of the risk–managed portfolio matches a given target. For example, let σ^{*2}

be the target level of unconditional variance, the constant c is given as follows

$$\mathrm{Var}\left(\frac{c}{\hat{\sigma}_t^2} r_t\right) = \sigma^{*2} \quad \Leftrightarrow \quad c = \sqrt{\frac{\sigma^{*2}}{\mathrm{Var}\left(\frac{1}{\hat{\sigma}_t^2} r_t\right)}} \tag{3}$$

In their application, Moreira and Muir estimate $\mathrm{Var}\left(\frac{1}{\hat{\sigma}_t^2} r_t\right)$ using the full data sample, which gives the estimate a forward–looking character.

In our application, we specify a time–varying c-factor, c_t, which is conditioned only on past return observations. This avoids the use of an ex–post optimal c-factor and results in gradually varying unconditonal variance levels (c.f. Figure 3 below). Specifically, we choose c_t such that the risk–managed strategy has the same average weight as a compareable buy–and–hold strategy with a strategic weight β^*. In Equation (2) $\frac{c_t}{\hat{\sigma}_t^2}$ represents the weight at time t. We choose c_t for a given t as the value such that the average weight over a predefined time horizon equals β^*, i.e.,

$$\frac{1}{k} \sum_{i=1}^{k} \frac{c_t}{\hat{\sigma}_{t-i}^2} = \beta^* \quad \Leftrightarrow \quad c_t = \frac{\beta^*}{\frac{1}{k} \sum_{i=1}^{k} \frac{1}{\hat{\sigma}_{t-i}^2}}, \tag{4}$$

where k represents the number of past observations. $\hat{\sigma}_t^2$ is the estimate for the variance of the portfolio.

3. Collective Defined–Contribution Pension Scheme

A collective defined–contribution (CDC) pension scheme is a pension plan where the investor and the plan sponsor pay fixed contribution rates in terms of a defined–contribution (DC) plan. However, all the assets are pooled, which allows the implementation of risk–sharing mechanisms by which the investor is not exposed to the typical risks of a DC pension plan.

In this section we first describe the general outline of a CDC model with a reserve pool.[1] Next, then describe the reserve–based framework as adopted by Goecke and the risk–based framework we use for our risk–managed investment strategy.

[1] To make the differences between the reserve–based and the risk–based approaches clear, our notation largely follows that of Goecke (2013a).

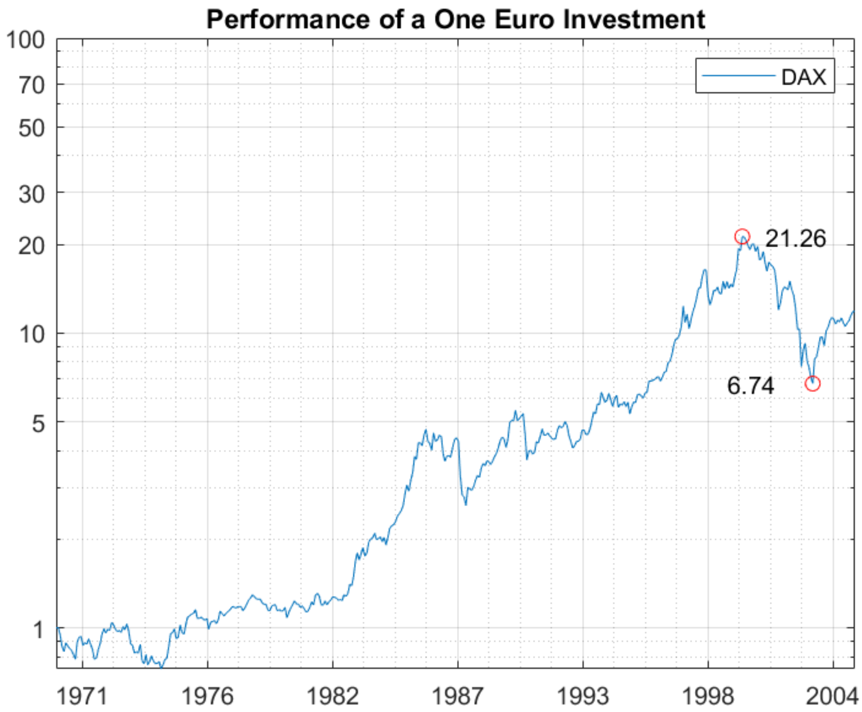


Figure 1: Illustration of potential variations in the returns of a CD investment plans

3.1. Motivation and General Outline

In a DC scheme the contributions made by the investor and the plan sponsor are attributed to the individual investor. The final amount received upon retirement depends entirely on the development of the capital market during the accumulation phase. This can lead to large variations in the returns different generations of investors receive. Figure 1 illustrates this problem. It shows the performance of an initial investment of one euro in 1970 in a market portfolio represented by the DAX. 30 years later the value grew to 21.26 euro. Three years later, at the end of the dotcom slump, a retiree would have received a return of just 6.74 euro on the initial investment.

The potential inequality of payouts received by different cohorts of retirees is a major criticism of proponents of pay–as–you–go plans. As we will see below, such intergenerational imbalances can be greatly reduced by building a risk–sharing mechanism into a DC pension plan.

In our framework, intergenerational risk–sharing is accomplished by introducing a collective reserve. The reserve is a fraction of the pooled total assets. It is not allocated to individual investors, but rather to the collective as a whole. The idea is that in times when investments underperform or decline, releases from the reserve will make up for the performance gaps in the individual accounts. And in times when investments exceed the target return, the reserve is replenished. This smooths volatility market returns and stabilizes performance across investor cohorts.

The basic idea of a CDC plan can be summarized by the balance–sheet equation

$$A_t = V_t + R_t, \tag{5}$$

where A_t represents the current value of the pension fund's assets, V_t is the total value of the individual accounts of the savers, and R_t is the reserve belonging to the collective. V_t and R_t represent the liabilities to the savers. The value of assets A_t evolves in line with the capital market and its value is therefore subject to significant fluctuations. By transferring most of these fluctuations to the reserve, the individual accounts V_t can grow steadily. And a steady growth rate of V_t reduces the risk of intergenerational inequality.

The task of the fund management is to decide on a, say, monthly basis on the portfolio weights and on the amounts allocated to individual accounts V_t and to or from the reserve R_t. The weights can be linked to the reserve status, via the reserve ratio, ρ_t, e.g.,[2]

$$\rho_t := \frac{R_t}{A_t}.$$

In the case the reserve ratio is high, i.e., the reserve accounts for a large fraction of the assets, A_t, the portfolio weights is shifted toward higher risk. Conversely, when the reserve ratio is below a certain target value, the weights in risky asset classes are reduced. Linking portfolio weights to reserve levels can result in a kind of momentum strategy, since reserve levels below a certain target value are often the consequence of market declines, whereas levels above the target occur in the wake of bull markets.

The amount distributed to the individual accounts and to the reserve is set by the profit participation, η_t, which determines the discrete or continuous monthly return

[2] Goecke uses the log reserve ratio, i.e., $\rho_t^{log} := \ln\left(\frac{A_t}{V_t}\right) = -\ln\left(1 - \frac{R_t}{A_t}\right)$, so that process ρ_t^{log} has a straightforward continuous–time representation.

on V_t, i.e.,

$$V_{t+1} = (1 + \eta_t)V_t, \tag{6}$$

or

$$V_{t+1} = e^{\eta_t} V_t. \tag{7}$$

Keeping the value of η_t close to constant lets V_t grow smoothly over time and, thus, reduces the risk of intergenerational inequality. By setting η_t at the beginning of a month, the growth of the individual accounts, V_{t+1}, is also automatically determined. However, at time t the end–of–month value of the total assets, A_{t+1}, is a stochastic quantity. If the return on A_t in month t deviates from η_t, the reserve pool R_{t+1} has to absorb the difference.

3.2. Reserve–based CDC Pension Scheme

The reserve–based scheme put forth by Goecke manages the pension fund by controlling the parameters σ_t and η_t based on the following asset liability management (ALM) adjustment rules:

$$\sigma_t = \sigma^* + a(\rho_t^{log} - \rho^{log*}) \tag{8}$$

$$\eta_t = \hat{\mu}_t(\sigma_t) + b(\rho_t^{log} - \rho^{log*}), \tag{9}$$

where a and b are nonnegative constants determining the speed of adjustment. σ^* and ρ^{log*} represent the target levels of the portfolio risk and the log reserve ratio, respectively. These levels are specified upfront by the fund manager.

The term $\rho_t^{log} - \rho^{log*}$ in (8) and (9) represents the link to the reserve status and balances the risk between generations of savers. If the pension fund follows a geometric Brownian motion, the expected portfolio return, $\hat{\mu}_t$, used to determine the declaration η_t, then depends on the volatility of the portfolio at time t via

$$\hat{\mu}_t(\sigma_t) = \mu_{1M,t} + r_{SR}\sigma_t - \frac{1}{2}\sigma_t^2,$$

where μ_{1M} is the one–month (risk–free) money market rate, and r_{SR} is the risk premium per unit of risk used in the drift of the geometric Brownian motion.

ALM-rules (8) and (9) determine the portfolio weights and the amounts allocated to the individual accounts V_t and the reserve R_t. Letting the fund portfolio consist of the two asset classes stocks and high–grade bonds and assuming that the volatility of the bond portfolio is negligible, then the fund's volatility is given by

$$\sigma_t = \beta_t \sigma_{stocks},$$

where β_t represents the weight in the stock portfolio, and σ_{stocks} is the volatility of the stock portfolio. The portfolio weights are implicitly given by σ_t. This implies the portfolio weight for stocks, i.e.,

$$\beta_t = \frac{\sigma_t}{\sigma_{stocks}}.$$

Unless leveraging is allowed, we have $0 \leq \sigma_t \leq \sigma_{stocks}$.

The declaration, η_t, is determined by ALM-rule (9). It determines V_{t+1}, the value of the individual accounts at the end of the month according to (7). Any imbalance between the realized growth of A_{t+1} and the already allotted V_{t+1} is then compensated by R_{t+1}. The declaration η_t is set so that savers participate fairly in capital market returns. If ρ_t is equal to the target ρ^*, the declaration corresponds to the return, $\hat{\mu}_t(\sigma_t)$, expected by the market model used. Thus, $b(\rho_t - \rho^*)$ balances the risk between different generations of savers. If $b > 0$ and $\rho_t < \rho^*$ the declaration is reduced and the reserve account is increased which benefits future savers. If, on the other hand, $b > 0$ and $\rho_t > \rho^*$, the current saver benefits from the good market performance in the past through a higher declaration.

3.3. Risk–Based CDC Pension Scheme

In our risk–based CDC scheme we tie the portfolio weights, β_t, directly to the market risk rather than first determining σ_t by ALM-rule (8) to implicitly obtain β_t. Moreover, we implement a different, more forward looking strategy for specifying $\hat{\mu}_t$ and, thus the declaration η_t.

According to (2), in the two-asset case with stocks and bonds, the risk–based weight of stocks, β_t^{RA}, is given by

$$\beta_t^{RA} = \frac{c_t}{\hat{\sigma}_t^2}. \tag{10}$$

Factor c_t can be chosen so that the average weight of the risk-managed strategy corresponds to some given value (c.f Equation (4)). $\hat{\sigma}_t^2$ is an estimate of the conditional variance of the equity returns. In the application below, it is obtained from a plain GARCH(1,1) model given by [3]

$$\begin{aligned} r_t &= \omega + \sigma_t \epsilon_t \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \sigma_{t-1}^2 + \gamma_2 r_t^2, \end{aligned} \tag{11}$$

where ϵ_t is an independent, identically and normally distributed random variable. The GARCH model is estimated recursively from rolling windows of daily log–returns. Then, conditional variance forecasts for the next 22 business days are generated and summed up to get a proxy for next month's variance, i.e.,

$$\hat{\sigma}_{temp,t}^2 = \sum_{j=1}^{22} \hat{\sigma}_{d,t+j}^2,$$

where $\hat{\sigma}_{d,j}^2$ is the daily conditional variance forecast given by

$$\begin{aligned} \hat{\sigma}_{d,t+1}^2 &= \gamma_0 + \gamma_1 \sigma_{d,t}^2 + \gamma_2 r_{stocks,t}^2, \\ \hat{\sigma}_{d,t+k}^2 &= \gamma_0 + (\gamma_1 + \gamma_2) \hat{\sigma}_{d,t+k-1}^2, \end{aligned}$$

with $r_{stocks,t}$ denoting the daily log–return of the stock portfolio at time t. In addition, we apply a smoothing procedure by taking exponential moving average over the preceding k_{smooth} months' estimates, i.e.,

$$\begin{aligned} \hat{\sigma}_1^2 &= \hat{\sigma}_{temp,1}^2 \\ \hat{\sigma}_t^2 &= \alpha \hat{\sigma}_{temp,t}^2 + (1-\alpha) \hat{\sigma}_{t-1}^2, \qquad \text{for } t > 2, \end{aligned}$$

where $\alpha = 2/(k_{smooth} + 1)$. Such a smoothing procedure avoids excessively volatile portfolio weights in risk–driven portfolio allocations.

Using (10), these weights are then linked to the reserve status via

$$\beta_t = \beta_t^{RA} \left(\frac{\rho_t}{\rho^*} \right)^m,$$

[3]Clearly, more elaborate GARCH models could be used in order to better capture possible fat–tailedness and skewness in the conditional distributions of the returns.

where m is a nonnegative constant determining the speed of adjustment with which the target level of reserve ratio is approached. If $m = 0$, $\beta_t = \beta_t^{RA}$ for all t. For $m > 0$ the equity ratio is linked to the reserve status, i.e., more than β_t^{RA} is invested in equity if the reserve ratio is above its target, ρ^*, and less if it is below. We restrict β_t to be between 0 and 1 to rule out any leveraging of equity investments by shorting bonds. Once the equity weights β_t are determined, the bond weights are $1 - \beta_t$, assuming that the pension fund remains fully invested in these two asset classes.

Finally, we need to determine the profit participation, η_t. Letting $\hat{\mu}_t$ be the weighted sum of estimates of the monthly returns of the two asset classes, i.e.,

$$\hat{\mu}_t = \beta_t \hat{\mu}_{stocks,t} + (1 - \beta_t)\hat{\mu}_{bonds,t},$$

we have

$$\eta_t = \hat{\mu}_t \left(\frac{\rho_t}{\rho^*} \right)^n ,$$

where n is again a nonnegative constant, which determines the speed of the adjustment toward the target reserve ratio. If $\hat{\mu}_{stocks,t}$ and $\hat{\mu}_{bonds,t}$ are both positive, then $\eta_t \geq 0$. To rule out the possibility of the reserve becoming negative, η_t may need to be adjusted to compensate for the losses that would result in a negative reserve.

4. Comparison of Reserve–based and Risk–based Schemes

In this section we present the results of a backtest using German equity and bond data. We construct CDC pension funds according to the two frameworks introduced in Section 3. These funds consist of two asset classes: stocks and bonds. The portfolio of stocks is represented by the German stock index (DAX) and the bond portfolio is represented by the German bond performance index (REXP). We compare performance and risk measures of the two frameworks. We further investigate the implication of imposing varying degrees of minimum diversification requirements accross the assets under investigation.

4.1. Data

We backtest the frameworks on a time series of the DAX and the REXP covering the period from 1967 to 2020. Monthly data of the REXP since 1967 and daily data of the DAX since 1987 is published by the Deutsche Bundesbank. For the time period before 1987 we use for the DAX a back-calculation of Stehle et al. (1996), which goes back until 1959. They amongst included dividend payments and their calculation, therefore, can be considered as a good representation of the performance of a well mixed German stock portfolio (see Stehle et al. (1996)). Figure 2 shows the performance of a one euro investment in the two indices. Furthermore, the figure includes the money market account representing a monthly revolving investment at the one-month money market rate. This index is included as it represents the risk–free rate of the applied market model in the reserve–based framework.

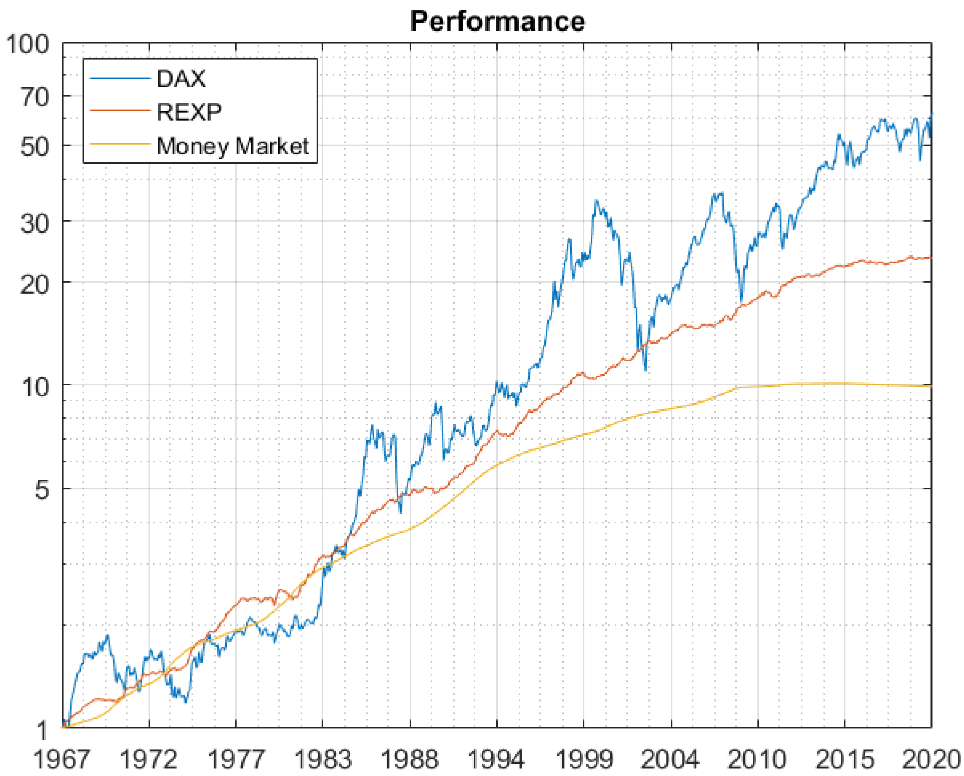


Figure 2: Performance of a one-euro investement from 1967 until 2020

4.2. Model Parameters for the Backtest

For the backtest we set the starting values of the pension funds as follows: $A_0 = 100$, $V_0 = 80$ and $R_0 = 20$. Therefore, the starting reserve ratio (log reserve ratio) is given by $\rho_0 = 20\%$ ($\rho_0^{log} = 18\%$).

For the reserve–based framework we use the same parameterization as in the reference model introduced in Goecke (2013b). The parameters a and b, which determine the speed of adjustment due to the reserve status, are set to 0.6 and 0.3, respectively. The target volatility σ^* of 10% corresponds to a target weight in the DAX of $\beta^* = 0.5$. The target log reserve ratio is set to the starting value of 18%. r_{SR} is the Sharpe ratio and corresponds to an annual risk premium of about 4.5% with respect to the risk–free money market rate. The unconditional annual standard deviation of the DAX is set to 20%. The values are summarized in Table 1.

a	b	σ^*	ρ_0^{log*}	r_{SR}	σ_{DAX}
0.6	0.3	10%	18%	0.225	20%

Table 1: Parameters set for the reserve–based framework.

For the risk–based framework with a risk–managed investment strategy we use the following parameterization. To determine the risk–managed portfolio weights β_t^{RA} the c-factor c_t and the conditional variance σ_t^2 have to be estimated. The c-factor c_t is determined using equation (4) by considering the previous 5 years, i.e., $k = 60$, and a target weight $\beta^* = 50\%$, which images the 10% for the target volatility σ^* in the reserve–based framework. To estimate the conditional variance σ_t^2 the GARCH(1,1) model is estimated on a rolling window of daily log-returns over the previous 3 years. For smoothing the variance estimates of the last 12 months are considered, i.e. $k_{smooth} = 12$. In Figure 3 the unsmoothed and smoothed monthly volatility estimates calculated according to the framework explained in Section 3.3 are shown.

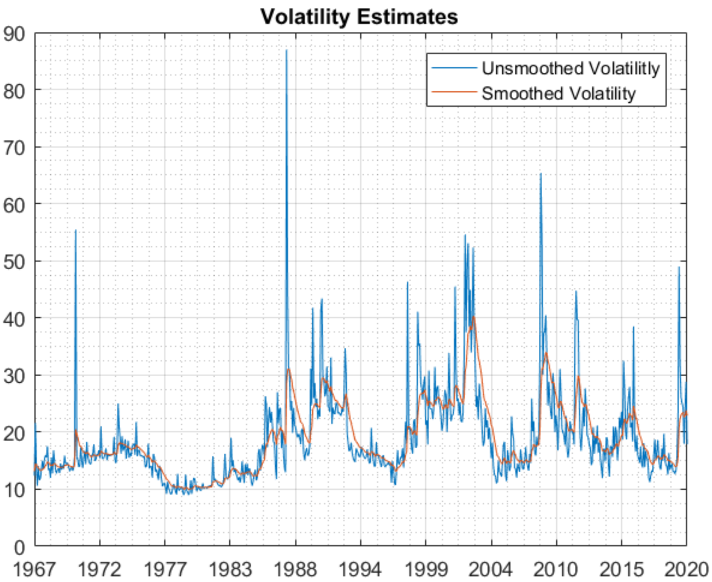


Figure 3: Volatility Esimates: Unsmoothed and smoothed with an exponential moving average approach.

To determine the expected return of the portfolio, $\hat{\mu}_t$, we set $\hat{\mu}_{DAX,t} = 0.67\%$ for all t, which corresponds to an expected annual return of approximately 8%. For $\hat{\mu}_{REXP,t}$ we use an average over a rolling window of monthly returns of the last 5 years as the return of the REXP is influenced amongst by political decisions of the central bank.

We further set the parameters m and n, which—similar to a and b in the reserve–based framework—control the speed of adjustment due to the reserve status, to 1.5 and 1.0, respectively. The target reserve ratio is set in accordance with the framework of Goecke to 20%. All parameters are summarized in Table 2.

m	n	k	β^*	k_{smooth}	ρ^*	$\hat{\mu}_{DAX}$
1.5	1.0	60	0.5	12	20%	0.67%

Table 2: Parameter set for the framework with a risk–managed investment strategy

With these values both frameworks are set and can be backtested on the given dataset.

4.3. Backtest

In Figure 4 we compare the performance of a DAX– and a REXP–investment with an investment in CDC pension funds using the two frameworks described in Section 3. We denote the two pension funds as *reserve–based CDC Plan* and *risk–based CDC Plan* according to the applied framework. The figure shows the development of a one euro investment in the beginning of the observation period. The very smooth course of the CDC plans is the result of the reserve, which smooths the volatile market returns over time. We observe that both CDC plans lie above the DAX most of the time although having only an average equity ratio of 59.8% and 51.9%, respectively. Furthermore, the risk–managed investment strategy indeed results in a higher performance.

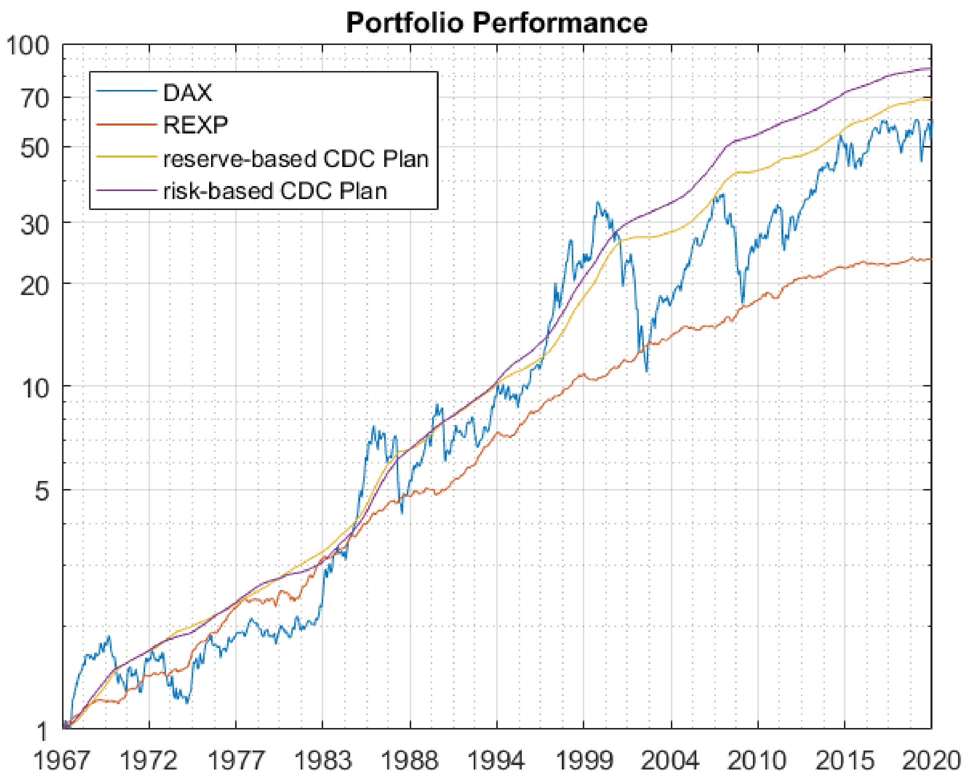
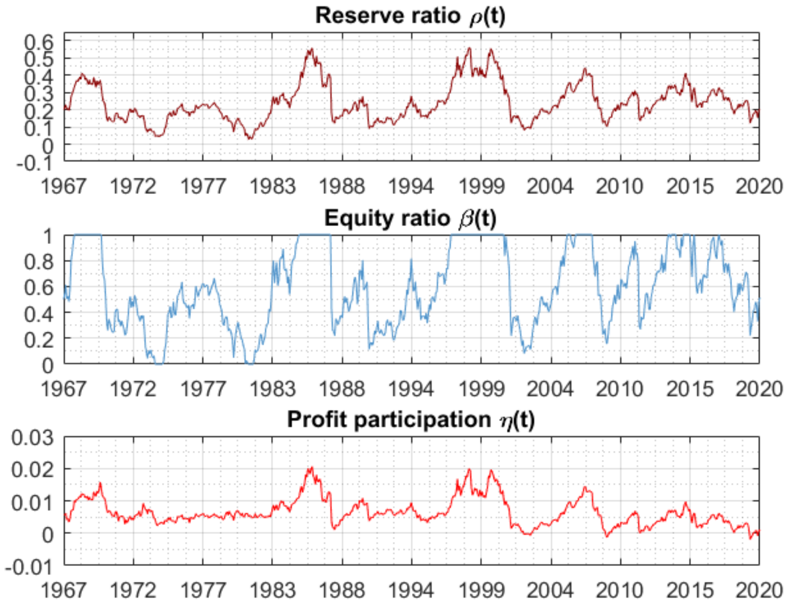
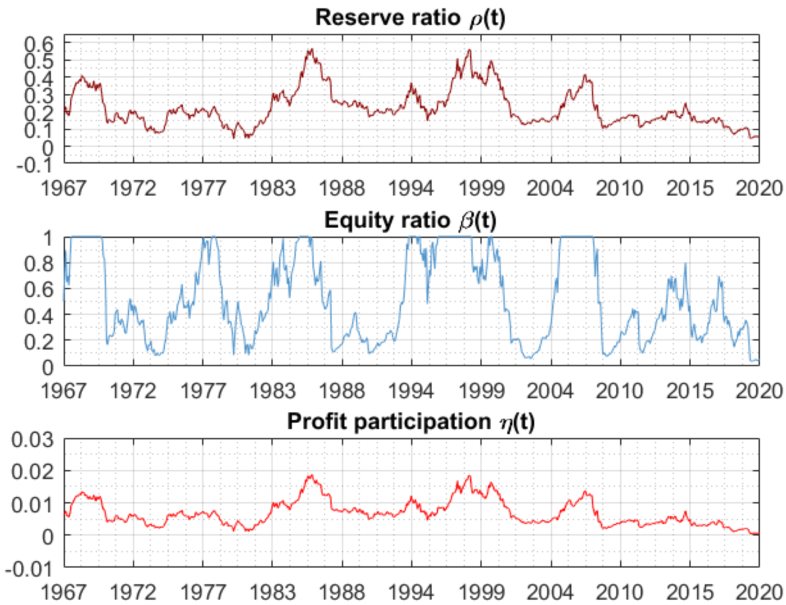


Figure 4: Performance of the reserve–based and the risk–based CDC Plan compared to the DAX and the REXP.

Figures 4a and 4b show the reserve ratio, ρ_t, the equity ratio, β_t, and the profit participation, η_t, for the reserve–based and the risk–based schemes over the backtest period.



(a) Reserve–based CDC Plan



(b) Risk–based CDC Plan

Figure 4: Development of the reserve ratio, the equity ratio and the profit participation of the CDC plans

The reserve–based CDC Plan has a positive reserve ratio over the investment horizon, which lies between 2.6% and 56.1%. The equity ratio takes values between 0% and 100% and is on average 59.8%. The profit participation may be negative.

The risk–based CDC Plan applies a risk–managed investment strategy, which leads to a smaller decline in the reserve ratio during a market crisis compared to the reserve–based CDC Plan. For example, in the dotcom crisis the equity ratio is decreased earlier than in the reserve–based CDC Plan due to the increase of the volatility in this time period. This reduces the losses and prohibits a large decline of the reserve ratio below its target value. The reserve ratio lies between 4.3% and 56.7%. The equity ratio takes values between 4.2% and 100% and is on average 51.9%. The profit participation is positive over the investment horizon.

We further investigate the performance and the risk of the CDC plans and a static saving plan. The static saving plan invests each month 50% in a stock portfolio represented by the DAX and 50% in a bond portfolio represented by the REXP. Table 3 presents the total performance, the annual performance, the annual standard deviation and annual semi standard deviation as well as the Sharpe ratio and Sortino ratio. Furthermore, the maximum drawdown and the average annual turnover of the applied investment strategy are given. To calculate the average annual turnover we first determine the monthly turnover by

$$\text{turnover}_{m,t} = \left| \beta_t - \frac{P_{DAX,t}}{A_t} \right|,$$

where $P_{DAX,t}$ is the value of the stock portfolio represented by the DAX at time t before adjusting the weights according to the applied framework or investment strategy. To get the annual turnover we sum up the monthly turnovers within a calendar year

$$\text{turnover}_{y,t_j} = \sum_{t=t_j-11}^{t_j} \text{turnover}_{m,t},$$

where t_j indicates the last month of a year. The average annual turnover is then given by the arithmetic mean over all annual turnovers.

	Static Plan	Reserve–based CDC Plan	Risk–based CDC Plan
Total Performance (%)	4881.01	6750.60	8354.48
Annual Performance (%)	7.51	8.16	8.58
Annual Std. Dev. (%)	9.93	1.47	1.36
Neg. Annual Semi-Vola (%)	6.17	0.04	0
Sharpe Ratio	0.76	5.55	6.30
Sortino Ratio	1.22	189.21	Inf
Maximum Drawdown (%)	34.52	0.38	0
Average Annual Turnover (%)	12.55	51.47	50.75

Table 3: Performance and Risk Measures

The static Plan has the lowest annual performance. As no risk–sharing mechanisms are implemented, the risk—measured by the annual standard deviation, the annual semi-vola or the maximum drawdown—is much higher compared to the pension plans. An unfavourable starting date for an investment in the static saving plan can therefore lead to large deviations from the average annual performance of 7.51%. The Sharpe ratio and Sortino ratio are therefore also small compared to the pension plans. The positive average turnover results from monthly adjustments to keep the weights constant over time.

In both CDC plans the collective reserve smooths the volatile market returns, which reduces the risk of the individual accounts V_t. As we consider the performance and risk not of A_t of a CDC plan but of V_t, whose fluctuations are mainly covered by the reserve R_t, this leads to lower risk measures and a higher Sharpe ratio and Sortino ratio. While the risk—measured by the annual standard deviation, the annual semi-vola or the maximum drawdown—are similar for both CDC plans, the annual performance of the risk–based CDC Plan is higher than of the reserve–based CDC Plan. This leads to an even higher Sharpe ratio and Sortino ratio. The Sortino ratio of the risk–based CDC Plan is infinity as the negative annual semi-vola is zero as the return of this strategy is always positive. The dynamic investment strategies lead to a higher average annual turnover than the static saving plan, but is for both plans with 51.47% for the reserve–based CDC Plan and 50.75% for the risk–based CDC Plan similar.

In practice, applied investment strategies often include minimum diversification requirements. We therefore investigate the implications of such requirements on the two CDC plans by imposing minimum investment constraints for the DAX and the

REXP. As we are invested in two asset classes imposing a minimum weight on one asset class results in a maximum weight constraint for the other. We impose the following varying degrees of minimum requirements for both asset classes: 0%, 10%, 20%, 30%, 40% and 50%. We calculate for all possible constraint combinations the same performance and risk measures as in Table 3 for the unconstrained case. Table 4 summarizes the results for the reserve–based CDC Plan and Table 5 for the risk–based CDC Plan.

For a fixed lower bound of the DAX increasing the lower bound of the REXP decreases the annual performance in both frameworks, but it also decreases the risk-measures such that the Sharpe ratio is increased. Only for the risk–based CDC Plan and a minimum DAX investment of 50% the Sharpe ratio decreases with the REXP constraint. The reason is that these strict constraints can lead to situations, in which according to the general allocation rules of the framework the reserve would get negative. This is especially the case in times of a market crisis. Compared to the reserve–based CDC Plan the risk–based CDC Plan prohibits a negative reserve by suspending the general rules of the framework to determine η_t in these situations. Instead a negative value for η_t is allowed, which exactly compensates the negative amount of the reserve. We can observe such situations if we impose a lower bound of 40% or higher on the DAX. A negative value for η_t increases the risk measures and can lead to smaller Sharpe ratios and Sortino ratios.

We observe that the risk–based CDC Plan has in general a higher annual performance, Sharpe ratios and Sortino ratios. Only in the case that the lower bound of the DAX is 40% or higher the Sharpe ratios and Sortino ratios for the reserve–based CDC Plan are higher. The reason is again, that while the reserve–based CDC Plan allows a negative reserve the risk–based CDC Plan prohibits this on the cost of higher risk measures.

Furthermore, we observe that the average annual turnover can be reduced by imposing higher lower bound constraints on the two assets.

Lower Bound DAX (%)		Lower Bound REXP (%)					
		0	10	20	30	40	50
0	Total Performance (%)	6750.60	6527.39	6256.91	6112.79	5687.66	5123.87
	Annual Performance (%)	8.16	8.09	8.01	7.96	7.82	7.61
	Annual Std. Dev. (%)	1.47	1.40	1.33	1.24	1.15	1.05
	Neg. Annual Semi-Vola (%)	0.04	0.04	0.04	0.04	0.04	0.03
	Sharpe Ratio	5.55	5.77	6.04	6.41	6.80	7.28
	Sortino Ratio	189.21	192.5	195.88	204.88	220.97	250.44
	Max. Drawdown (%)	0.38	0.36	0.34	0.31	0.28	0.29
	Avg. Annual Turnover (%)	51.47	47.88	43.77	37.42	31.71	25.82
10	Total Performance (%)	6734.14	6512.04	6241.47	6089.57	5668.27	5109.42
	Annual Performance (%)	8.15	8.08	8.0	7.95	7.81	7.61
	Annual Std. Dev. (%)	1.47	1.40	1.33	1.24	1.15	1.04
	Neg. Annual Semi-Vola (%)	0.04	0.04	0.04	0.04	0.04	0.03
	Sharpe Ratio	5.55	5.77	6.04	6.41	6.8	7.29
	Sortino Ratio	190.54	193.92	197.41	206.60	223.16	253.88
	Max. Drawdown (%)	0.38	0.36	0.34	0.31	0.28	0.29
	Avg. Annual Turnover (%)	50.02	46.43	42.34	36.10	30.36	24.43
20	Total Performance (%)	6759.16	6524.56	6248.57	6063.51	5647.45	5110.87
	Annual Performance (%)	8.16	8.09	8.00	7.94	7.80	7.61
	Annual Std. Dev. (%)	1.47	1.40	1.32	1.24	1.15	1.04
	Neg. Annual Semi-Vola (%)	0.04	0.04	0.04	0.04	0.04	0.03
	Sharpe Ratio	5.55	5.77	6.04	6.42	6.80	7.30
	Sortino Ratio	183.17	185.63	188.25	195.86	210.12	234.9
	Max. Drawdown (%)	0.38	0.36	0.34	0.31	0.29	0.30
	Avg. Annual Turnover (%)	47.99	44.55	40.3	34.54	28.82	22.50
30	Total Performance (%)	6314.17	6111.22	5844.02	5616.91	5250.46	4718.36
	Annual Performance (%)	8.02	7.96	7.87	7.79	7.66	7.45
	Annual Std. Dev. (%)	1.48	1.41	1.34	1.25	1.17	1.07
	Neg. Annual Semi-Vola (%)	0.08	0.08	0.08	0.08	0.07	0.07
	Sharpe Ratio	5.43	5.64	5.89	6.22	6.56	6.99
	Sortino Ratio	103.41	103.14	102.26	102.22	104.2	104.77
	Max. Drawdown (%)	1.13	1.13	1.13	1.12	1.07	1.04
	Avg. Annual Turnover (%)	44.36	40.83	36.75	31.4	25.52	19.61
40	Total Performance (%)	6176.84	5982.32	5737.15	5522.15	5138.20	4585.32
	Annual Performance (%)	7.98	7.92	7.83	7.76	7.62	7.40
	Annual Std. Dev. (%)	1.50	1.43	1.36	1.28	1.20	1.10
	Neg. Annual Semi-Vola (%)	0.13	0.13	0.13	0.13	0.13	0.13
	Sharpe Ratio	5.33	5.52	5.76	6.06	6.35	6.71
	Sortino Ratio	62.33	61.97	61.16	60.70	60.38	58.85
	Max. Drawdown (%)	2.41	2.41	2.41	2.41	2.36	2.32
	Avg. Annual Turnover (%)	39.58	35.90	31.70	26.77	20.75	15.25
50	Total Performance (%)	5997.36	5887.40	5698.83	5432.67	5032.40	4440.81
	Annual Performance (%)	7.92	7.89	7.82	7.73	7.58	7.33
	Annual Std. Dev. (%)	1.53	1.47	1.40	1.33	1.24	1.15
	Neg. Annual Semi-Vola (%)	0.19	0.19	0.19	0.19	0.18	0.18
	Sharpe Ratio	5.17	5.36	5.58	5.82	6.10	6.40
	Sortino Ratio	41.73	41.64	41.20	40.82	40.97	39.93
	Max. Drawdown (%)	3.95	3.94	3.94	3.91	3.75	3.63
	Avg. Annual Turnover (%)	34.94	31.64	26.75	21.96	15.86	12.55

Table 4: Performance and Risk Measures for the reserve–based CDC Plan

Lower Bound DAX (%)		Lower Bound REXP (%)					
		0	10	20	30	40	50
0	Total Performance (%)	8354.48	7802.53	7289.87	6672.46	6017.33	5248.36
	Annual Performance (%)	8.58	8.44	8.31	8.13	7.93	7.66
	Annual Std. Dev. (%)	1.36	1.31	1.26	1.19	1.11	1.02
	Neg. Annual Semi-Vola (%)	0	0	0	0	0	0
	Sharpe Ratio	6.30	6.45	6.62	6.81	7.12	7.54
	Sortino Ratio	Inf	Inf	Inf	Inf	Inf	Inf
	Max. Drawdown (%)	0	0	0	0	0	0
	Avg. Annual Turnover (%)	50.75	46.15	41.12	36.05	30.37	25.81
10	Total Performance (%)	8438.47	7880.26	7360.32	6735.12	6077.24	5296.18
	Annual Performance (%)	8.60	8.46	8.33	8.15	7.95	7.68
	Annual Std. Dev. (%)	1.36	1.31	1.25	1.19	1.11	1.01
	Neg. Annual Semi-Vola (%)	0	0	0	0	0	0
	Sharpe Ratio	6.32	6.48	6.65	6.84	7.15	7.58
	Sortino Ratio	Inf	Inf	Inf	Inf	Inf	Inf
	Max. Drawdown (%)	0	0	0	0	0	0
	Avg. Annual Turnover (%)	50.49	45.88	40.87	35.80	30.15	25.58
20	Total Performance (%)	8614.59	8020.40	7501.66	6879.63	6176.26	5361.66
	Annual Performance (%)	8.64	8.50	8.36	8.19	7.98	7.70
	Annual Std. Dev. (%)	1.35	1.30	1.24	1.18	1.10	1.00
	Neg. Annual Semi-Vola (%)	0	0	0	0	0	0
	Sharpe Ratio	6.39	6.56	6.73	6.94	7.25	7.68
	Sortino Ratio	Inf	Inf	Inf	Inf	Inf	Inf
	Max. Drawdown (%)	0	0	0	0	0	0
	Avg. Annual Turnover (%)	48.50	43.95	38.68	32.86	27.62	23.21
30	Total Performance (%)	8536.83	7896.88	7410.42	6890.92	6174.84	5335.89
	Annual Performance (%)	8.62	8.47	8.34	8.20	7.98	7.69
	Annual Std. Dev. (%)	1.34	1.28	1.23	1.17	1.09	1.00
	Neg. Annual Semi-Vola (%)	0	0	0	0	0	0
	Sharpe Ratio	6.43	6.60	6.80	7.03	7.31	7.71
	Sortino Ratio	Inf	Inf	Inf	Inf	Inf	Inf
	Max. Drawdown (%)	0	0	0	0	0	0
	Avg. Annual Turnover (%)	46.49	42.56	36.70	29.84	24.42	20.24
40	Total Performance (%)	8423.05	7792.20	7286.18	6799.91	6138.96	5267.38
	Annual Performance (%)	8.59	8.44	8.31	8.17	7.97	7.67
	Annual Std. Dev. (%)	1.42	1.37	1.32	1.27	1.21	1.17
	Neg. Annual Semi-Vola (%)	0.33	0.36	0.38	0.38	0.41	0.49
	Sharpe Ratio	6.06	6.16	6.30	6.45	6.58	6.57
	Sortino Ratio	25.75	23.63	21.86	21.53	19.60	15.76
	Max. Drawdown (%)	3.33	3.68	3.96	3.95	4.26	5.04
	Avg. Annual Turnover (%)	43.61	39.46	32.76	26.04	20.16	16.10
50	Total Performance (%)	8041.52	7529.90	6967.65	6485.76	5846.61	5093.13
	Annual Performance (%)	8.50	8.37	8.22	8.08	7.87	7.60
	Annual Std. Dev. (%)	1.98	1.95	1.94	1.90	1.89	1.90
	Neg. Annual Semi-Vola (%)	1.31	1.34	1.38	1.37	1.41	1.49
	Sharpe Ratio	4.29	4.29	4.23	4.26	4.17	4.01
	Sortino Ratio	6.49	6.26	5.96	5.89	5.57	5.10
	Max. Drawdown (%)	11.91	12.26	12.71	12.67	12.96	13.71
	Avg. Annual Turnover (%)	40.62	34.66	28.81	22.13	16.01	12.55

Table 5: Performance and Risk Measures for the risk–based CDC Plan

Regarding the total performance it seems to be optimal for both frameworks to choose a lower bound of 20% for the DAX and 0% for the REXP. Figure 5 shows the performance of a one-euro investment under these constraints. In Figure 6a and 6b the sequences of of the reserve ratio, the portfolio weight in the DAX and the profit participation η_t are shown as in the unconstraint case above.

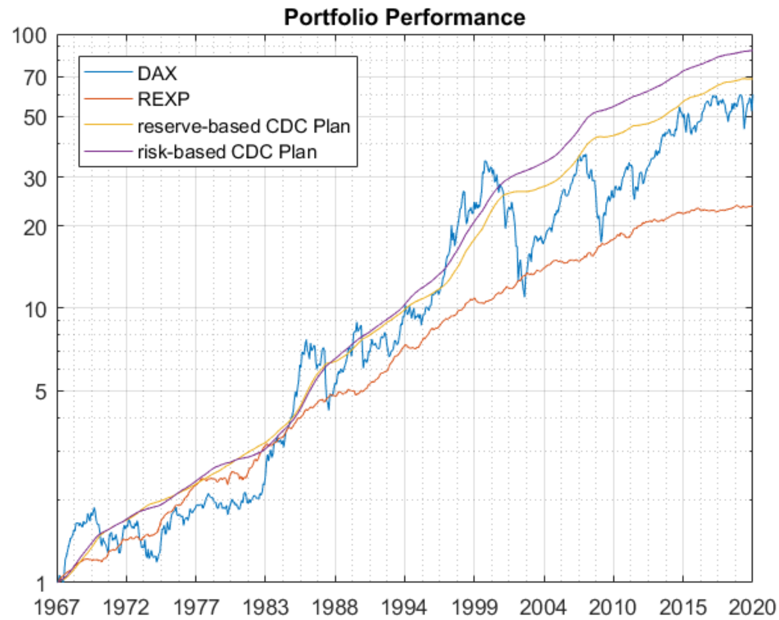
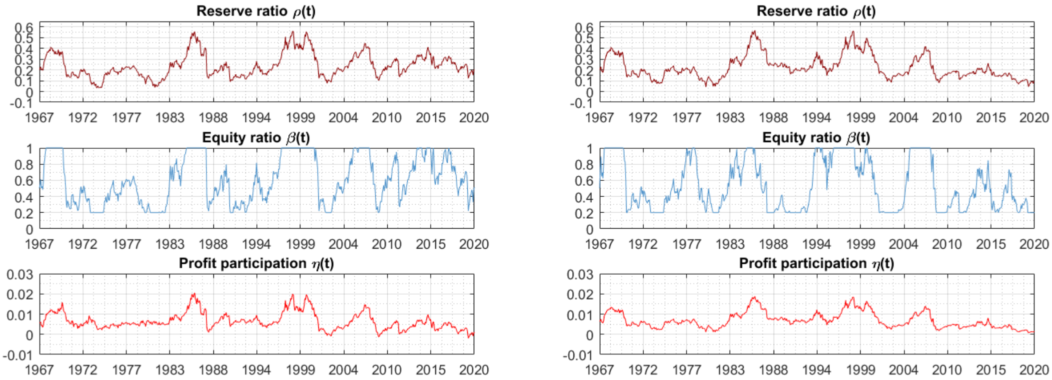


Figure 5: Performance of the reserve–based and the risk–based CDC Plan with lower bound constraints of 20% for the DAX investment and 0% for the REXP investment.



(a) Reserve–based CDC Plan

(b) Risk–based CDC Plan

Figure 6: Development of the reserve ratio, the equity ratio and the profit participation of the CDC plans with lower bound constraints of 20% for the DAX investment and 0% for the REXP investment.

In practical applications a minimum investment amount is imposed on all asset classes to guarantee a minimum of diversification. We therefore present in Figure 7 the performance of a one-euro investment with the constraint that at least 20% is invested in the DAX and 20% is invested in the REXP. Figure 8a and 8b show the developement of the reserve ratio, the portfolio weight in the DAX and the profit participation η_t as before. We can see that the weight in the DAX is not only bounded below by 20% but has also a maximum weight of 80%. This results from the lower bound constraint of the REXP investment as the pension fund is fully invested in these two asset classes.

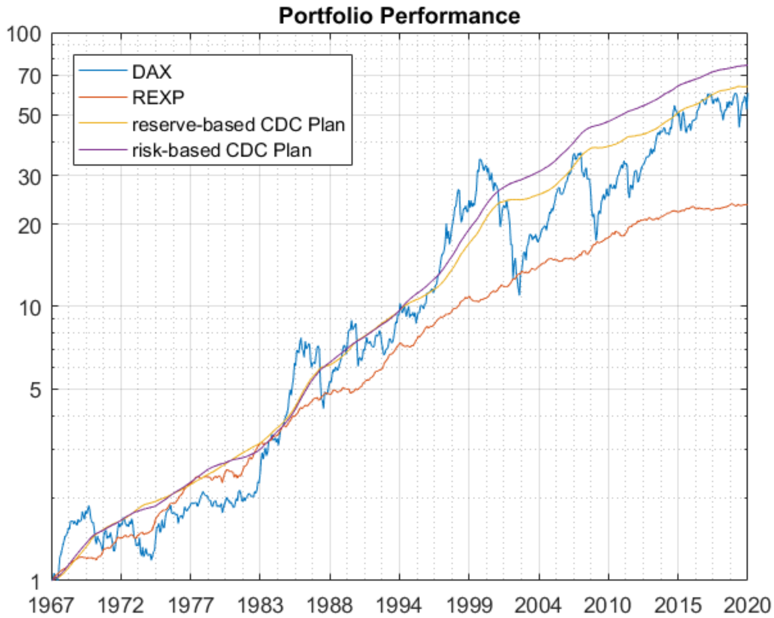
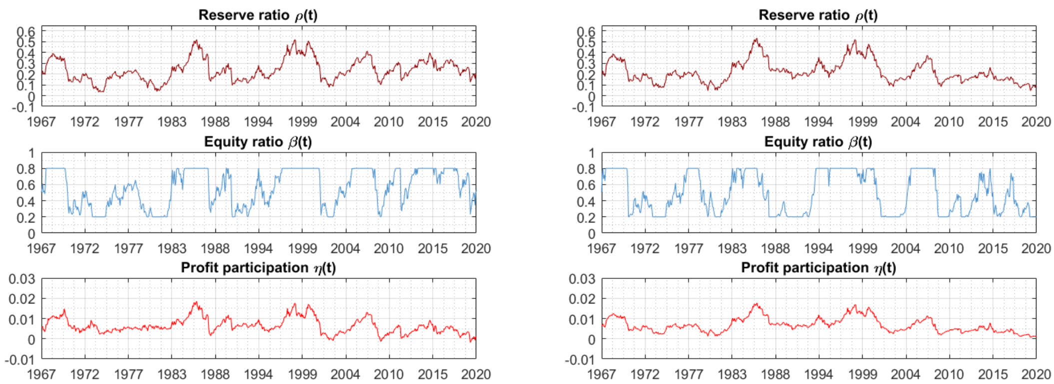


Figure 7: Performance of the reserve–based and the risk–based CDC Plan with lower bound constraint of 20% for the DAX investment and 20% for the REXP investment.



(a) Reserve–based CDC Plan

(b) Risk–based CDC Plan

Figure 8: Development of the reserve ratio, the equity ratio and the profit participation of the CDC plans with lower bound constraint of 20% for the DAX investment and 20% for the REXP investment.

5. Conclusion

We have applied a risk–managed investment strategy to a CDC pension framework that implements a collective reserve to smooth risk and benefits across generations.

We compared the results to a reserve–based CDC pension framework as put forth by Gollier (2008) and implemented by Goecke (2013a).

By smoothing the volatile market returns we achieve an intergenerational risk transfer as the investment of each generation of investors performs similar. This works well as we showed by comparing the performance of the pension scheme frameworks to a static saving plan. A risk–managed investment strategy accounts for the current risk in the capital markets and prevents serious declines of the collective reserve in times of a market crisis. We further showed that a risk–managed investment strategy applied in a CDC pension framework can increase the annual return, Sharpe ratios and other performance measures. We also investigate implications of imposing varying degrees of minimum–diversification requirements across assets. The results with constraints are more representative for real world applications. We showed that with these constraints we still give rise to an attractive absolute and risk–adjusted performance. Also the average annual turnover can be reduced by imposing suitable constraints.

In this paper we have applied the risk–managed investment strategy in the spirit of Moreira and Muir (2017) to investigate the implications of such a strategy on a given application. In our proposed framework any investment strategy can be easily implemented. Therefore, further approaches could add a momentum strategy which might further improve performance and risk measures.

References

Ball, L. and Mankiw, N. G. Intergenerational risk sharing in the spirit of arrow, debreu, and rawls, with applications to social security design. Journal of Political Economy, 115(4):523–547, 2007.

Bams, D., Schotman, P. C., and Tyagi, M. Asset allocation dynamics of pension funds. Netspar Discussion Paper, 2016.

Barroso, P. and Santa-Clara, P. Momentum has its moments. Journal of Financial Economics, 116(1):111–120, 2015.

Barroso, P., Detzel, A. L., and Maio, P. F. Managing the risk of the low risk anomaly. In 30th Australasian Finance and Banking Conference, 2017.

Baumann, R. T. and Müller, H. H. Pension funds as institutions for intertemporal risk transfer. Insurance: Mathematics and Economics, 42(3):1000–1012, 2008.

Bovenberg, L. and Mehlkopf, R. Optimal design of funded pension schemes. Annual Review of Economics, 6(1):445–474, 2014.

Cederburg, S., O'Doherty, M. S., Wang, F., and Yan, X. S. On the performance of volatility-managed portfolios. Journal of Financial Economics, 2020.

Chen, D. H., Beetsma, R. M., Ponds, E. H., and Romp, W. E. Intergenerational risk-sharing through funded pensions and public debt. Journal of Pension Economics & Finance, 15(2):127–159, 2016.

Cui, J., De Jong, F., and Ponds, E. Intergenerational risk sharing within funded pension schemes. Journal of Pension Economics & Finance, 10(1):1–29, 2011.

Daniel, K. and Moskowitz, T. J. Momentum crashes. Journal of Financial Economics, 122(2):221–247, 2016.

Døskeland, T. M. and Nordahl, H. A. Intergenerational effects of guaranteed pension contracts. The Geneva Risk and Insurance Review, 33(1):19–46, 2008.

Eisdorfer, A. and Misirli, E. U. Distressed stocks in distressed times. Management Science, 66(6):2452–2473, 2020.

Goecke, O. Pension saving schemes with return smoothing mechanism. Insurance: Mathematics and Economics, 53(3):678–689, 2013a.

Goecke, O. Sparprozesse mit kollektivem risikoausgleich – backtesting. Technical report, Forschung am ivwKöln, 2013b.

Gollier, C. Intergenerational risk-sharing and risk-taking of a pension fund. Journal of Public Economics, 92(5-6):1463–1485, 2008.

Gordon, R. H. and Varian, H. R. Intergenerational risk sharing. Journal of Public Economics, 37(2):185–202, 1988.

Grosen, A. and Jørgensen, P. L. Fair valuation of life insurance liabilities: the impact of interest rate guarantees, surrender options, and bonus policies. Insurance: Mathematics and Economics, 26(1):37–57, 2000.

Hoevenaars, R. P. and Ponds, E. H. Valuation of intergenerational transfers in funded collective pension schemes. Insurance: Mathematics and Economics, 42(2):578–593, 2008.

Krueger, D. and Kubler, F. Pareto-improving social security reform when financial markets are incomplete!? American Economic Review, 96(3):737–755, 2006.

Moreira, A. and Muir, T. Volatility-managed portfolios. The Journal of Finance, 72 (4):1611–1644, 2017.

Schumacher, J. M. A note on gollier's model for a collective pension scheme. Journal of Pension Economics & Finance, 20(2):187–211, 2021.

Stehle, R., Maier, J., and Huber, R. Rückberechnung des dax für die jahre 1955 bis 1987. Technical report, SFB 373 Discussion Paper, 1996.

Teulings, C. N. and De Vries, C. G. Generational accounting, solidarity and pension losses. De Economist, 154(1):63–83, 2006.

Westerhout, E. Intergenerational risk sharing in time-consistent funded pension schemes. Netspar Discussion Paper, 2011.

Eidesstattliche Erklärung

(Siehe Promotionsordnung vom 12. Juli 2011, §8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 15.07.2021

Christoph Berninger