DISSERTATION

Yield Curves and Chance-Risk Classification: Modeling, Forecasting, and Pension Product Portfolios

FRANZISKA SIEGLINDE DIEZ

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. Dr. An Chen

Datum der Disputation: 11. Dezember 2020

D386

Technische Universität Kaiserslautern
Fachbereich Mathematik
Gottlieb-Daimler-Straße 47
67663 Kaiserslautern

Fraunhofer ITWM
Abteilung Finanzmathematik
Fraunhofer-Platz 1
67663 Kaiserslautern

Abstract

This dissertation consists of three independent parts: The yield curve shapes generated by interest rate models, the yield curve forecasting, and the application of the chance-risk classification to a portfolio of pension products. As a component of the capital market model, the yield curve influences the chance-risk classification which was introduced to improve the comparability of pension products and strengthen consumer protection. Consequently, all three topics have a major impact on this essential safeguard.

Firstly, we focus on the obtained yield curve shapes of the Vasicek interest rate models. We extend the existing studies on the attainable yield curve shapes in the one-factor Vasicek model by analysis of the curvature. Further, we show that the two-factor Vasicek model can explain significantly more effects that are observed at the market than its one-factor variant. Among them is the occurrence of dipped yield curves. We further introduce a general change of measure framework for the Monte Carlo simulation of the Vasicek model under a subjective measure. This can be used to avoid the occurrence of a far too high frequency of inverse yield curves with growing time.

Secondly, we examine different time series models including machine learning algorithms forecasting the yield curve. For this, we consider statistical time series models such as autoregression and vector autoregression. Their performances are compared with the performance of a multilayer perceptron, a fully connected feed-forward neural network. For this purpose, we develop an extended approach for the hyperparameter optimization of the perceptron which is based on standard procedures like Grid and Random Search but allows to search a larger hyperparameter space. Our investigation shows that multilayer perceptrons outperform statistical models for long forecast horizons.

The third part deals with the chance-risk classification of state-subsidized pension products in Germany as well as its relevance for customer consulting. To optimize the use of the chance-risk classes assigned by Produktinformationsstelle Altersvorsorge gGmbH, we develop a procedure for determining the chance-risk class of different portfolios of state-subsidized pension products under the constraint that the portfolio chance-risk class does not exceed the customer's risk preference. For this, we consider a portfolio consisting of two new pension products as well as a second one containing a product already owned by the customer as well as the offer of a new one. This is of particular interest for customer consulting and can include other assets of the customer. We examine the properties of various chance and risk parameters as well as their corresponding mappings and show that a diversification effect exists. Based on the properties, we conclude that the average final contract values have to be used to obtain the upper bound of the portfolio chance-risk class. Furthermore, we develop an approach for determining the chance-risk class over the contract term since the chance-risk class is only assigned at the beginning of the accumulation phase. On the one hand, we apply the current legal situation, but on the other hand, we suggest

an approach that requires further simulations. Finally, we translate our results into recommendations for customer consultation.

Zusammenfassung

Diese Dissertation besteht aus drei voneinander unabhängigen Teilen: Die Zinsmodellierung und die daraus resultierenden Zinsstrukturkurvenformen, die Zinsstrukturkurvenvorhersage sowie die Bestimmung der Chancen-Risiko Klasse eines Portfolios aus Altersvorsorgeprodukten. Die Zinsstrukturkurvenform beeinflusst als Bestandteil des Kapitalmarktmodelles die Chancen-Risiko Klassifizierung, die zur besseren Vergleichbarkeit von Altersvorsorgeprodukten und Stärkung des Verbraucherschutzes eingeführt wurde. Alle drei Themen haben somit eine große Bedeutung für den Versicherungsnehmer.

Der erste Teil befasst sich mit den Zinsstrukturkurvenformen, die das Vasicek Zinsmodell generiert. Dabei erweitern wir die aktuellen Studien bezüglich der Zinsstrukturkurvenformen im Einfaktor Vasicek-Modell um die Analyse deren Krümmungsverhaltens. Weiter zeigen wir, dass das Zweifaktor Vasicek-Modell deutlich mehr am Markt beobachtbare Effekte erklären kann als seine Einfaktor-Variante. Dazu gehört das Auftreten von invers-gewölbte Zinsstrukturkurven. Darüber hinaus führen wir einen allgemeinen Rahmen eines Maßwechsels für die Monte-Carlo-Simulation des Vasicek-Modells unter einem subjektiven Maß ein. Dadurch kann das Auftreten einer – empirisch – viel zu hohen Anzahl inverser Zinsstrukturkurven mit zunehmender Zeit vermieden werden.

Im zweiten Teil untersuchen wir verschiedene Zeitreihenmodelle, inklusive Algorithmen maschinellen Lernens, zur Vorhersage der zukünftigen Zinsstrukturkurve. Dafür betrachten wir statistische Zeitreihenmodelle wie Auto- und Vektorautoregression. Deren Performance wird mit der eines mehrlagigen Perzeptrons, einem feedforward neuronalen Netz, verglichen. Hierzu entwickeln wir für die Hyperparameteroptimierung des Perzeptrons einen erweiterten Ansatz, der auf Standardverfahren wie Grid und Random Search basiert, aber es erlaubt einen größeren Hyperparameterraum zu durchsuchen. Unsere Untersuchung zeigt, dass mehrlagige Perzeptrone vor allem bei langen Vorhersagehorizonten die statistischen Modelle übertreffen.

Der dritte Teil beschäftigt sich mit der Chancen-Risiko Klassifizierung staatlich geförderter Altersvorsorgeprodukte auf dem deutschen Markt und ihrer Bedeutung für die Beratung. Um die Möglichkeiten der Nutzung der Chancen-Risiko Klassen der Produktinformationsstelle Altersvorsorge gGmbH bei der Kundenberatung zu maximieren, entwickeln wir ein Verfahren zur Bestimmung der Chancen-Risiko Klasse verschiedener Portfolios staatlich geförderter Altersvorsorgeprodukte. Hierbei soll die Chancen-Risiko Klasse des Portfolios nicht größer als die Risikopräferenzklasse des zu beratenden Kunden sein. Andere Anlagearten können eingeschlossen werden. Wir untersuchen die Eigenschaften der verschiedenen Chancen- und Risikoparameter samt zugehöriger Abbildungen und zeigen, dass ein Diversifikationseffekt bei der Klassifizierung vorliegt. Aufbauend auf den Eigenschaften erhalten wir als Ergebnis, dass als Obergrenze der Chancen-Risiko Klasse des Portfolios die gemittelten Endvermögen heranzuziehen sind und übersetzen dies in Empfehlungen für die Kundenberatung. Darüber hinaus entwickeln wir ein Verfahren zur Bestimmung der Chancen-Risiko Klasse über die Vertragslaufzeit eines Altersvorsorgeproduktes,

da diese nur zu Beginn der Vertragslaufzeit zugewiesen wird. Zum einen wenden wir die aktuelle Rechtslage an, zum anderen schlagen wir ein Verfahren vor, das weitere Simulationen erfordert.

Danksagung

Zuallererst möchte ich mich bei Prof. Dr. *Ralf Korn* für seine Unterstützung und Betreuung bedanken. Er stand immer für Fragen und Diskussion zur Verfügung und gab hilfreiche Anmerkungen und Ideen. Seine Betreuung trug entscheidend zu der Vollendung dieser Promotion bei und ich hätte mir keine bessere wünschen können.

Bei der Abteilung *Finanzmathematik* bedanke ich mich für die finanzielle Unterstützung der Promotion durch ein Stipendium des Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM und die fördernde sowie kollegiale Arbeitsumgebung. Namentlich gilt mein Dank Prof. Dr. *Andreas Wagner*, der mir als Abteilungsleiter die Möglichkeit der Promotion eröffnete, Dr. *Robert Knobloch*, mit dessen Hilfe und Unterstützung das Thema zur Vorhersage der Zinsstrukturkurve entstand, Dr. *Roman Horsky*, der mich in dem letzten Jahr meiner Promotion seitens der Abteilung betreute und mir mit konstruktivem Feedback und Rat zur Seite stand. Bei allen meinen Doktoranden- und Arbeitskollegen bedanke ich mich für die interessanten und ideenreichen Gespräche und den Spaß, den wir während und außerhalb der Arbeit hatten.

Bei meinen Großeltern bedanke ich mich für ihre liebevolle und fördernde Art und die vielen Momente, die ich mit ihnen erleben durfte. Leider durften sie meine Promotion nicht mehr miterleben, dennoch weiß ich, dass sie stolz auf mich wären, vor allem mein Großvater, *Raimund*, der schon während meiner Zeit als wissenschaftlicher Mitarbeiter in der Abteilung Finanzmathematik mich ermutigte, die Chance zu promovieren wahrzunehmen.

Zu guter Letzt gilt mein Dank meinen Eltern, *Claudia* und *Theo*, dafür, dass sie mir mein Studium und damit meinen weiteren beruflichen Weg ermöglicht haben und immer für mich da sind, meinen Geschwistern, *Sebastian*, *Veronika* und *Matthias*, für ihre Unterstützung in jeglicher Form und den Spaß, den wir immer zusammen haben, und meiner Patentante, *Sieglinde*, für ihren Zuspruch und Interesse an meiner Arbeit. Mein besonderer Dank gilt meinem Freund *Christian* für seine Geduld, Unterstützung sowie Verständnis in der Endphase.

Contents

List of Figures

List of Tables

List of Symbols

Mathematical symbols

$\sharp$	Cardinality of a set
Corr	Correlation
Cov	Covariance
$\mathbb{E}$	Expected value
$\mathbb{E}_{\mathbb{P}}$	Expected value under $\mathbb{P}$
$\mathcal{F}_t$	Natural filtration at time t
$\mathcal{N}$	Normal distribution
$\mathbb{N}$	Set of natural numbers
Ω	Set of scenarios
$\mathbb{P}$	Real-world or physical measure
$P\left(X \leqslant p\right)$	Probability that the random variable X is less than or equal to p
Φ	Cumulative distribution function of the standard normal distribution
$\mathbb{Q}$	Risk-neutral measure
Q_p	p quantile
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_0^+$	Set of non-negative real numbers
$\mathbb{R}^n$	n-dimensional set of real numbers
$\mathbb{R}^{m\times n}$	Cartesian product of $\mathbb{R}^m$ and $\mathbb{R}^n$
Θ	Euclidean vector space
Var	Variance

Yield curve shapes of Vasicek interest rate models

a	Mean-reversion speed in the one-factor Vasicek model and of the first stochastic process in the two-factor Vasicek short rate model
$\widetilde{a}$	Mean-reversion speed under $\mathbb{P}$ in the one-factor Vasicek short rate model
b	Mean-reversion speed of the second stochastic process in the two-factor Vasicek short rate model
b_{inv}	Lower bound for inverse yield curves in the one-factor Vasicek model
b_{norm}	Upper bound for normal yield curves in the one-factor Vasicek model
d_a	Additional drift of a
d_θ	Additional drift of θ
η	Volatility of the second stochastic process in the two-factor Vasicek short rate model
$f^M(0,T)$	Market instantaneous forward rate for maturity T
$\bar{p}$	Given probability to observe normal yield curves at time $\bar{s}$
$P(t,T)$	Price of a zero-coupon bond with maturity T at time t
$P^M(0,T)$	Market price of a zero-coupon bond with maturity T
$r(t)$	Short rate at time t
r_0	Short rate at time 0
$\bar{r}$	Given expected short rate at time $\bar{t}$
ρ	Correlation of the Brownian motions in the two-factor Vasicek short rate model
$\bar{s}$	Time point at which a specific probability to observe normal yield curves should be reached
σ	Volatility in the one-factor Vasicek short rate model and of the first stochastic process in the two-factor Vasicek short rate model
T	Maturity
t	Time point
$\bar{t}$	Time point at which the mean of the short rate should reach a specific value

θ	Mean-reversion level in the one-factor Vasicek short rate model
$\widetilde{\theta}$	Mean-reversion level under $\mathbb{P}$ in the one-factor Vasicek short rate model
$W(t)$	One dimensional Brownian motion under $\mathbb{Q}$
$\widetilde{W}(t)$	One dimensional Brownian motion under $\mathbb{P}$
x	Time to maturity
$y(t,x)$	Yield of a zero-coupon bond with price $P(t,t+x)$
$\bar{y}$	Asymptotic level of the yield curve

Multi-step yield curve forecasting

b	Bias
$b^{(i)}$	Bias of the variable $x^{(i)}$ in the estimation of $x^{(i)}$
b_l	Bias of the l^{th} layer
c_i	i^{th} Criterion bound of the hyperparameter optimization
$\triangle_h x_{t_k}$	Increment between x_{t_k} and $x_{t_{k-h}}$
$\widehat{\triangle}_h x_{t_k}$	Forecast increment between x_{t_k} and $x_{t_{k-h}}$
ϵ_{t_k}	Realizations of independent random variables with zero expectation and equal variance at time t_k
f	True functional relation
$\widehat{f}$	Estimation of the true functional relation
f_h	True functional relation of the forecast horizon h
$\widehat{f}_h$	Estimation of the true functional relation of the forecast horizon h
$f^{(i)}$	True functional relation of the variable $x^{(i)}$
$\widehat{f}^{(i)}$	Estimation of the true functional relation of the variable $x^{(i)}$
h	Forecast horizon
$\mathscr{H}$	Set of forecast horizons
H	Size of $\mathscr{H}$
L	Number of hidden MLP layers

m	Considered model
$\mathscr{M}$	Set of models
m_μ^{cv}	Best model according to performance measure μ and cross validation
m_μ^{ew}	Best model according to performance measure μ and expanding window approach
μ	Performance measure
μ_l^m	Performance measure of the l^{th} validation set of model m
$\overline{\mu}^m$	Mean of the performance measures over the validation sets of model m
$\overline{\mu}^{m,p}$	Mean of the performance measures over the different values of hyperparameter p of model m
$\mu_l^{m,p,v}$	Performance measure of the l^{th} validation set of hyperparameter p with value v and model m
$\overline{\mu}^{m,p,v}$	Mean of the performance measures over the validation sets of hyperparameter p with value v and model m
n	Number of observations
N	Number of forecasting variables
N_0	Input dimension of the MLP
N_L	Output dimension of the MLP
N_l	Dimension of the l^{th} hidden layer of the MLP
p	Hyperparameter
$\mathscr{P}$	Set of hyperparameters
ϕ	Activation function
$\sigma_{\mu^{m,p,v}}$	Empiric standard deviation of the performance measures over the validation sets of hyperparameter p with value v and the model m
$\sigma_{\overline{\mu}^{m,p}}$	Empiric standard deviation of $\overline{\mu}^{m,p,v}$ over v
t_k	Time point
v	Hyperparameter value
$\mathscr{V}$	Set of hyperparameter combinations

$\mathscr{V}^*$	Set of selected hyperparameter combinations
$\mathscr{V}_p$	Set of hyperparameter p values
$\mathscr{V}^*_{p,i}$	Set of selected hyperparameter p values according to the i^{th} Criterion
W	Coefficient matrix
$w^{(i)}$	Coefficient of the variable $x^{(i)}$ in the estimation of $x^{(i)}$
$w_t^{(i)}$	Coefficient of the variable time in the estimation of $x^{(i)}$
$w_j^{(i)}$	Coefficient of the variable $x^{(j)}$ in the estimation of $x^{(i)}$
W_l	Coefficient matrix of the l^{th} layer
$x^{(i)}$	Observations over time of the variable $x^{(i)}$
x_{t_k}	Observations at time t_k
$x_{t_k}^{(i)}$	Observation at time t_k of the variable $x^{(i)}$
$\widehat{x}_{t_k}$	Forecast values at time t_k
$\overline{x}_{t_k}$	Mean value of the observations at time point t_k
x_l^{tr}	l^{th} training set
x_l^{val}	l^{th} validation set

Chance-risk classification of a portfolio

α	Proportion of the entire capital invested in one pension product
α_i	Proportion of the entire capital invested in the i^{th} pension product
α^*	Maximum proportion of the entire capital invested in one pension product to obtain a portfolio chance-risk class not larger than given
α_j^*	Maximum proportion of the entire capital invested in one pension product to obtain a portfolio chance-risk class not larger than j
α_{incr}^*	Maximum proportion of the entire capital invested in increasing the existing pension product to obtain a portfolio chance-risk class not larger than given
b_T^j	μ^c-intercept of the chance-risk class boundary between chance-risk class j and $j+1$ for the accumulation phase T

β	Proportion of the new investment invested in the existing pension product
β^*	Maximum proportion of the new investment invested in the existing pension product to obtain a portfolio chance-risk class not larger than given
β_j^*	Maximum proportion of the new investment invested in the existing pension product to obtain a portfolio chance-risk class not larger than j
CRC	Chance-risk class
CRC^i	Chance-risk class of the i^{th} pension product
CRC^{cust}	Risk profile of the customer
CRC_t	Chance-risk class at time t
$\mathcal{CRC}$	Chance-risk class resulting from interpolation of the averaged final contract values
$\mathcal{CRC}^{ptf(\alpha)}$	Chance-risk class resulting from interpolation of the averaged final contract values of the portfolio defined by $\alpha V^1 + (1-\alpha)\, V^2$
$\mathcal{CRC}^{ptf(\alpha_1,\alpha_2)}$	Chance-risk class resulting from interpolation of the averaged final contract values of the portfolio defined by $\alpha_1 V^1 + \alpha_2 V^2 + (1-\alpha_1-\alpha_2)\, V^3$
f	Function for calculating the averaged final contract values
g	Function for calculating the measures from the averaged final contract values
g_T	Function for calculating the measures from the averaged final contract values depending on T
g_T^{-1}	Function for calculating the averaged final contract values from the measures depending on T
I	New investment
t	Passed time in months
μ	Chance or risk measure
μ^c	Chance measure
μ^r	Risk measure
$\widehat{\mu}^c$	μ^c for a fixed value μ^r equal zero

$\mu^{c,i}$	Chance measure of the i^{th} pension product
$\mu^{r,i}$	Risk measure of the i^{th} pension product
$\mu^{c,int(\alpha)}$	With proportion α interpolated chance measure of two pension products
$\mu^{r,int(\alpha)}$	With proportion α interpolated risk measure of two pension products
$\mu^{c,ptf(\alpha)}$	Chance measure of the portfolio defined by $\alpha V^1 + (1-\alpha)\, V^2$
$\mu^{r,ptf(\alpha)}$	Risk measure of the portfolio defined by $\alpha V^1 + (1-\alpha)\, V^2$
μ_t	Chance or risk measure at time t
μ_t^c	Chance measure at time t
μ_t^r	Risk measure at time t
$\mu_t^{c,i}$	Chance measure of the i^{th} pension product at time t
$\mu_t^{r,i}$	Risk measure of the i^{th} pension product at time t
$\mathcal{P}$	Premium of the contract
$\mathcal{P}^i$	Premium of the i^{th} pension product
$\mathcal{P}^{*,i}$	Maximum premium invested in the i^{th} pension product
P_T	Idealized customer's premium with accumulation phase T
T	Accumulation phase of PIA classification
$\mathcal{T}$	Accumulation phase of the contract
T^i	Accumulation phase of the i^{th} pension product in PIA classification
$\mathcal{T}^i$	Accumulation phase of the i^{th} pension product
T_t	Simulated accumulation phase at time t
$\mathcal{T}_t$	Remaining accumulation phase of the contract at time t
$\mathcal{V}_T$	Set of pension products with simulated accumulation phase T
v_t	Contract value after t months (before premium payment)
v_t^i	Contract value of the i^{th} pension product after t months (before premium payment)
$\widetilde{v}_t$	Scaled contract value after t months (before premium payment)

$\widetilde{v}_t^i$	Scaled contract value of the i^{th} pension product after t months (before premium payment)
V	Pension product
V^i	i^{th} pension product
v^k	Final contract value of the k^{th} simulation
v_t^k	Final contract value of the k^{th} simulation at time t
$\overline{V}$	Mean of all or of the 2,000 lowest final contract values
$\overline{V}^c$	Mean of the final contract values
$\overline{V}^r$	Mean of the 2,000 lowest final contract values
$\overline{V}^{c,i}$	Mean of the i^{th} pension product final contract values
$\overline{V}^{r,i}$	Mean of the i^{th} pension product 2,000 lowest final contract values of
$\overline{V}^{c,int(\alpha)}$	With α interpolated mean of the final contract values of two pension products
$\overline{V}^{r,int(\alpha)}$	With α interpolated mean of the 2,000 lowest final contract values of two pension products
$\overline{V}^{c,int(\alpha_1,\alpha_2)}$	With α_1 and α_2 interpolated mean of the final contract values of three pension products
$\overline{V}^{r,int(\alpha_1,\alpha_2)}$	With α_1 and α_2 interpolated mean of the 2,000 lowest final contract values of three pension products
$\overline{V}^{c,ptf(\alpha)}$	Mean of the final contract values of the portfolio defined by $\alpha V^1 + (1-\alpha)\, V^2$
$\overline{V}^{r,ptf(\alpha)}$	Mean of the 2,000 lowest final contract values of the portfolio defined by $\alpha V^1 + (1-\alpha)\, V^2$
$\overline{V}_t$	Mean of all or the 2,000 lowest final contract values at time t
$\overline{V}_t^c$	Mean of the final contract values at time t
$\overline{V}_t^r$	Mean of the 2,000 lowest final contract values at time t
$\overline{V}_t^{c,i}$	Mean of the i^{th} pension product final contract values of at time t
$\overline{V}_t^{r,i}$	Mean of the i^{th} pension product 2,000 lowest final contract values of at time t
x	Customer age at the contract start

List of Acronyms

ADF	augmented Dickey-Fuller
AltvPIBV	Altersvorsorge-Produktinformationsblattverordnung
AR	autoregression
cf.	compare
CRC	chance-risk class
CRK	Chancen-Risiko Klasse
e.g.	for example
ECB	European Central Bank
EI-QFM	European Institute for Quality Management for Financial Products and Methods
i.e.	id est
LIBOR	London Interbank Bank Offered Rate
MAE	mean absolute error
MAXE	maximum error
MLP	multilayer perceptron
MSE	mean squared error
OECD	Organisation for Economic Co-operation and Development
PIA	Produktinformationsstelle Altersvorsorge gGmbH
PIB	Produktinformationsblatt
VAR	vector autoregression

1 Introduction

The yield curve is of great relevance in the financial and economic world. Particularly, it contains all information on the attainable riskless returns for different maturity dates at the bond market and displays the spot rates or yields for all times to maturity (i.e. equivalent constant interest rates if a zero bond with remaining time to maturity is bought now and held until maturity). Every bond market has its own yield curves which represent its specific current market situation. For example, the U.S. yield curve differs from the European and the German one.

The yield curve shape is an indicator of both the demand for money and the estimation of the future interest rate development. A so-called normal yield curve increases monotonically in the time to maturity. This means that investors intend to be compensated with a higher spot rate for providing their money for a longer period of time. However, if there is a lot of demand for money now an inverse yield curve can appear. In this scenario, the market participants are not interested in getting money for a longer time period and thus the yield curve decreases monotonically in the time to maturity. Liquidity aspects will often lead to an initially normal curve that switches to an inverse behavior for large times to maturity. The main reason being that there is no liquid market for bonds with a maturity of say 40 years and longer. This is called a humped yield curve. While these are the main conceptual shapes, the yield curve shapes appearing in reality are not limited to them.

Due to its importance, a vast literature on the yield curve has been developed. One emphasis of the literature is put on the estimation of the yield curve by fitting the curve to the data using statistical techniques. The best-known examples are Nelson and Siegel (1987) and Svensson (1994) who introduce a parsimonious model with exponential decay terms. The model of Svensson (1994) is used by the European Central Bank (ECB) or the Deutsche Bundesbank to estimate the European or German yield curve. There are other approaches like Fama and Bliss (1987) who apply splines for the estimation. Another emphasis of the literature is the modeling of the yield curve. Three concepts can be distinguished: forward rate, short rate, and LIBOR market models. The name indicates which parameter is modeled. Ho and Lee (1986) and Heath et al. (1992) are contributions in the forward-rate models. Popular short rate models are Vasicek (1977), Cox and Ross (1985), and Hull and White (1990). The LIBOR market models are introduced by Miltersen et al. (1997) and Brace et al. (1997). In Ho and Lee (1986), Heath et al. (1992), and Hull and White (1990), the initial yield curve is perfectly fitted to the yield curve which is observed at the market. Therefore, these models are also called no-arbitrage models. In Chapter 2, we discuss the Vasicek model in more detail, especially with respect to the yield curve shapes. Diebold and Li (2006) began to focus on the forecast of the yield curve to which little

attention was paid in the literature of yield curve modeling. They use statistical time series models and evaluate the forecast out-of-sample. Based on this, further studies followed as outlined in Section 3.1.

The modeling of the yield curve is the basis of different applications and risk indicators. One example is the chance-risk classification by Produktinformationsstelle Altersvorsorge gGmbH (PIA). Every German state-subsidized pension product being sold has to be classified into one of five so-called chance-risk classes (CRCs). By law, the classification has to be based on the simulated evolution of the contract value of the pension products. As this contract value depends on the evolution of the capital market, models for the simulation of interest rates and stock prices had to be chosen. A two-factor Hull-White version of the Vasicek model for the short rate is used on the interest rate side together with a variant of the Black-Scholes model on the stock side (see Korn and Wagner (2018) for more details). These models are close to the industry standard and therefore enable a reproduction of the CRC by the industry. Particular arguments in favor of the two-factor variant in comparison with the one-factor Hull-White model are the possibilities

- to model different effects for near and far forward rates,
- to generate other yield curve shapes than the normal, humped, and inverse one,
- and a much better fit for the prices of interest rate derivatives (see e.g. Acar et al. (2011)).

In addition to the CRC, a standardized and mandatory Produktinformationsblatt (PIB) was also introduced for German state-subsidized pension products. This is intended to improve the transparency and comparability of the products which are particularly important for customer consulting. A good consulting is characterized by the fact that only products are recommended to the customer that correspond to his risk preference. To ensure this, algorithms to assign a customer a CRC are necessary. On this basis, only pension products with the same or a lower CRC are supposed to be recommended. This approach is extended by including already bought pension products in Chapter 4.

This thesis consists of three independent chapters. The main parts of Chapter 2 has been published in the European Actuarial Journal (see Diez and Korn (2020)), while parts of Chapter 4 has been submitted at submission of the thesis (see Diez et al. (2020)).

Chapter 2 is mainly concerned with the attainable yield curve shapes of the Vasicek interest rate model. We focus on the two-factor Vasicek model since the possible yield curve shapes of the one-factor Vasicek model are known in literature. In the one-factor model, we additionally analyze the curvature of the yield curve shapes. It can be shown that every normal yield curve is strictly convex, but not all inverse ones are strictly concave. The two-factor Vasicek model can generate more yield curve shapes than its one-factor variant and explain more observable effects of the market like the occurrence of dipped yield curves. Furthermore, we introduce a general change

of measure framework in the one-factor Vasicek model reducing the frequency of inverse yield curves to occur over time.

In Chapter 3, we deal with forecasting the yield curve. For this purpose, statistical time series models such as autoregression and vector autoregression are compared with multilayer perceptrons (MLPs), a machine learning algorithm. For the hyperparameter optimization of the MLP, we develop an approach that includes standard procedures such as Grid or Random Search but selects promising hyperparameters in advance. This allows the search of larger hyperparameter spaces. In addition to forecasting the yield curve of the next day, we consider a forecast horizon of one month, three months, half a year, and one year. The neural networks outperform the standard time series significantly for longer forecast horizons.

Chapter 4 focuses on the chance-risk classification of German state-subsidized pension products by PIA and its relevance for customer consulting. In this context, we develop an approach for determining the CRC of a portfolio. In doing so, we consider a portfolio consisting of two new products as well as one consisting of an existing one and of a new product. For this purpose, the chance and risk parameters of the classification are examined first. Based on this, the existence of a diversification effect is shown favoring the inclusion of the customer's portfolio into customer consulting. Furthermore, we develop an approach for determining the CRC of the pension product over the contract term based on the classification by PIA. In our approach, further simulations have to be performed. But the pension product and the evolution of its CRC of the contract is better represented.

All computations of Chapter 2 and 4 were done with the open source programming language R, which can be downloaded at `www.r-project.org`. For the computations and data manipulations of Chapter 3, the open source programming language Python which is available at `www.python.org` were employed. In particular, the package NumPy (`numpy.org`) and pandas (`pandas.pydata.org`) were used for data analysis and scikit-learn (`scikit-learn.org`) for the different models which include fitting and validation. The figures derived from our calculations are created with the package ggplot2 (`ggplot2.tidyverse.org`) under R and with matplotlib (`matplotlib.org`) under Python.

2 Yield Curve Shapes of Vasicek Interest Rate Models and Measure Transformations[1]

In this chapter, we investigate the possible yield curve shapes of the Vasicek short rate model. Keller-Ressel and Steiner (2008) and Keller-Ressel (2018) are the main academic references for the yield curve shapes that affine one-factor short rate models can generate. On the one hand, the results are drawn from those papers. On the other hand, they are complemented by adding results of the yield curve curvature in the one-factor Vasicek model and of the yield curve shapes in the two-factor Vasicek model. This model is particularly popular in connection with the chance-risk classification of German state-subsidized pension products. We show that the two-factor model is more adept to explain more empirically observed effects than the one-factor model. Furthermore, we consider the distribution of the yield curve shapes under the risk-neutral and the real-world measure in the one-factor Vasicek model and introduce a change of measure approach to avoid the occurrence of a far too high frequency of inverse yield curves with growing time.

Our main contributions in this chapter are

- the proof of the existence of dipped yield curves in the two-factor Vasicek model,
- an analysis of the yield curve shapes of the two-factor Vasicek model highlighting empirically relevant effects that one-factor models cannot explain,
- a general change of measure approach to model the evolution of the yield curve in Vasicek models over time,
- the development of a simulation framework for bonds and yields that respects empirical findings with regard to the frequency of yield curve structures even for long horizons.

We will illustrate all our findings with explicit examples. The essential parts and results of this chapter are summarized in Diez and Korn (2020) which is published in the European Actuarial Journal.

In the first section, we introduce definitions and notations in the context of yield curves. Afterwards, we examine the observed European yield curves from 2005 until

[1] Based on published work: Diez and Korn (2020).

2019 regarding their shapes and frequency of occurrence. These empirical results are presented in Section 2.2. Section 2.3 deals with the Vasicek short rate models. The one-factor Vasicek model is considered in Section 2.3.1 and the two-factor one in Section 2.3.2. In both sections, the short rate model and the calculation of the bond prices and the yields are described. Furthermore, we analyze which yield curve shapes are generated by the individual models. Additionally, we investigate the distribution of the yield curve shapes for the one-factor Vasicek model since explicit shapes of the distribution for the occurrence of the possible yield curve shapes are given here. Section 2.4 considers a measure change and its effect on the yield curve shapes in the one-factor Vasicek model. In Section 2.4.1, the use of the risk-neutral measure $\mathbb{Q}$ and the real-world measure $\mathbb{P}$ in simulations are discussed. Subsequently, Section 2.4.2 deals with the evolution of the short rate and yield curve shapes under the real-world measure. Finally, we suggest a different approach for determining the real-world measure which uses yield forecasts as well as conditions on the probability to observe normal yield curves in the last Section 2.4.3.

2.1 Definitions and Notations

This section introduces the basic instruments, notations, and definitions in the context of interest rates and their modeling. We refer to the corresponding chapters of the monographs Brigo and Mecurio (2007) and Desmettre and Korn (2018).

Definition 2.1 (Zero-coupon bond)**.** A *zero-coupon bond* with maturity T (for short a T-bond) pays its owner one unit of money at time T. No further payments occur before maturity. We denote its price at time $t \leqslant T$ by $P(t,T)$.

Note that $P(t,T)$ also represents the present value at time t of one unit of currency to be paid at time T. Clearly, $P(t,t) = 1$ for all t. Further, we assume all bonds occurring to be riskless, i.e. they do not admit default risk.

Buying a T-bond and holding it until maturity is a fully determined investment at the time of purchase. To judge its performance on the remaining time to maturity $T-t$, we introduce the notion of $x = T-t$ and write $P(t,t+x)$ instead of $P(t,T)$. This gives rise to the definition of the yield of this bond as the appropriate performance measure.

Definition 2.2 (Yield)**.** The *yield* $y(t,x)$ of a zero-coupon bond with price $P(t,t+x)$ is defined as

$$y(t,x) = -\frac{ln(P(t,t+x))}{x}$$

or equivalently as

$$P(t,t+x) = e^{-y(t,x)\cdot x}.$$

Thus, the yield is the continuously compounded interest rate that a money market account must perform as the zero bond in the remaining time span x to maturity. For an infinitesimally small time to maturity x $y(t,x)$ is also called the short rate.

Definition 2.3 (Short rate)**.** The *short rate* $r(t)$ is defined as

$$r(t) = \lim_{x \downarrow 0} y(t, x). \tag{2.1}$$

$y(t, x)$ depends on both the current time t and the time to maturity x. A fundamental curve describing the interest rate market is obtained by holding t fixed and mapping the zero bonds of different times to maturity into yields. This curve is called the yield curve at time t.

Definition 2.4 (Yield curve)**.** For $t \geqslant 0$ fixed the graph of the function

$$x \mapsto y(t, x), \quad x > 0$$

is called the *yield curve*.

The yield curve makes bonds of different times to maturity comparable with regard to their yields. Generally, a larger yield is expected for a longer investment horizon. However, this is not always the case. Regarding the shape of yield curves, we distinguish the following basic shapes.

Definition 2.5 (Yield curve shapes)**.** The yield curve as a function of x (for fixed t) is called

- *flat* if it is constant on $(0, \infty)$,
- *normal* if it is a strictly monotonically increasing and bounded from above function on $(0, \infty)$,
- *inverse* if it is a strictly monotonically decreasing and bounded from below function on $(0, \infty)$,
- *humped* if it has exactly one local maximum and no minimum on $(0, \infty)$,
- *dipped* if it has exactly one local minimum and no maximum on $(0, \infty)$.

The five basic yield curve shapes are illustrated in Figure 2.1. There are more shapes possible such as combinations of the ones given above that can e.g. have exactly one local maximum and one local minimum. Also, in reality, there can (and have) occur(ed) more irregular shapes than the ones mentioned in the definition. However, as only the above ones are relevant for our theoretical findings in Section 2.3, we will not consider any other shapes below.

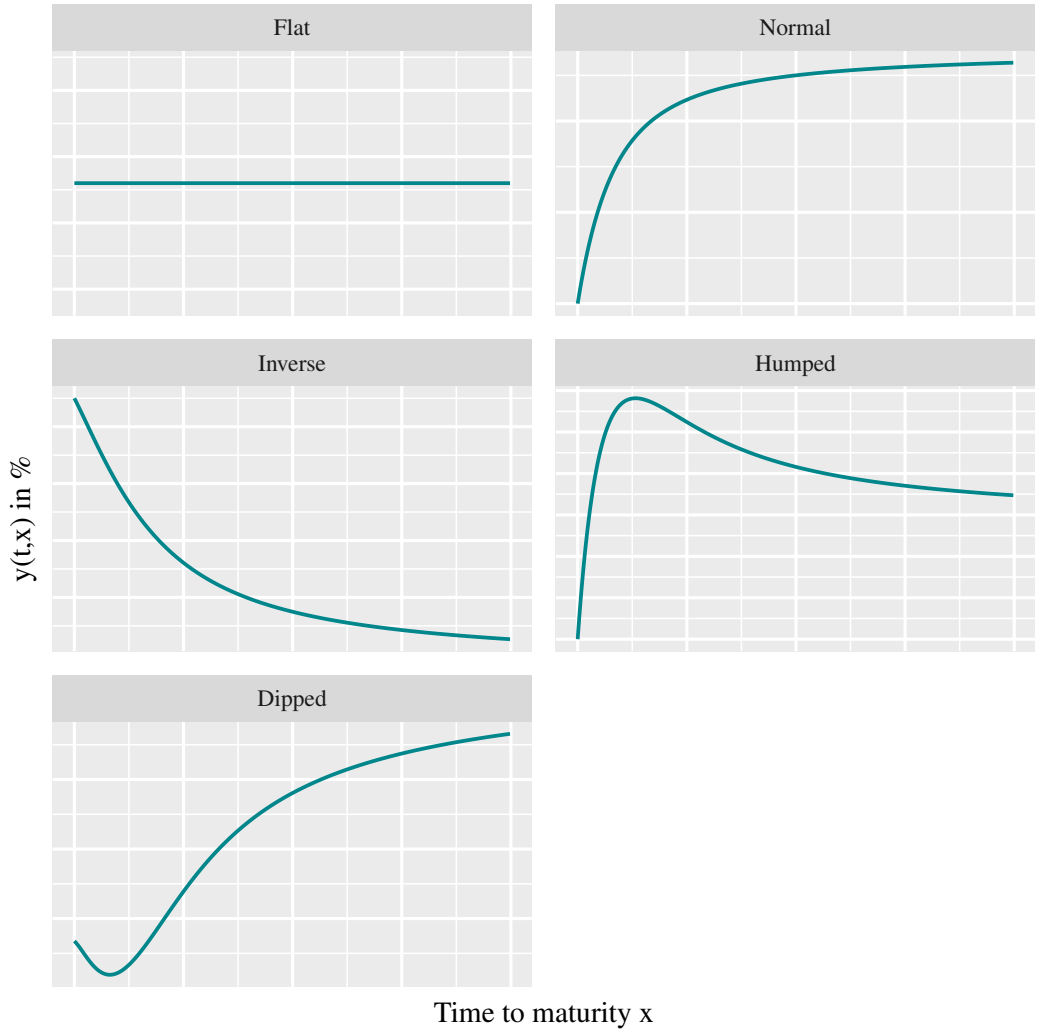


Figure 2.1: Basic yield curve shapes

2.2 Empirical Facts on the Frequency of Yield Curve Shapes

To give an impression of the occurrence frequency of different yield curve shapes, we present some facts that are based on publicly available interest rate data of triple-A-rated government bonds from the ECB[2]. Figure 2.2 shows the yield curve from 2005 until 2019. The calendar time is on the x-axis and the time to maturity x on the z-axis. The y-axis shows the yield $y(t,x)$ in $\%$. Different yield curve shapes can be observed: normal, humped, dipped, and humped-dipped. The frequency of the respective shapes is seen in Table 2.3. We consider daily data (i.e. 3,832 days) and bond maturities up to 20 years for reason of market illiquidity.

[2] `https://sdw.ecb.europa.eu/browse.do?node=9691417`, accessed on 2020 July 14th.

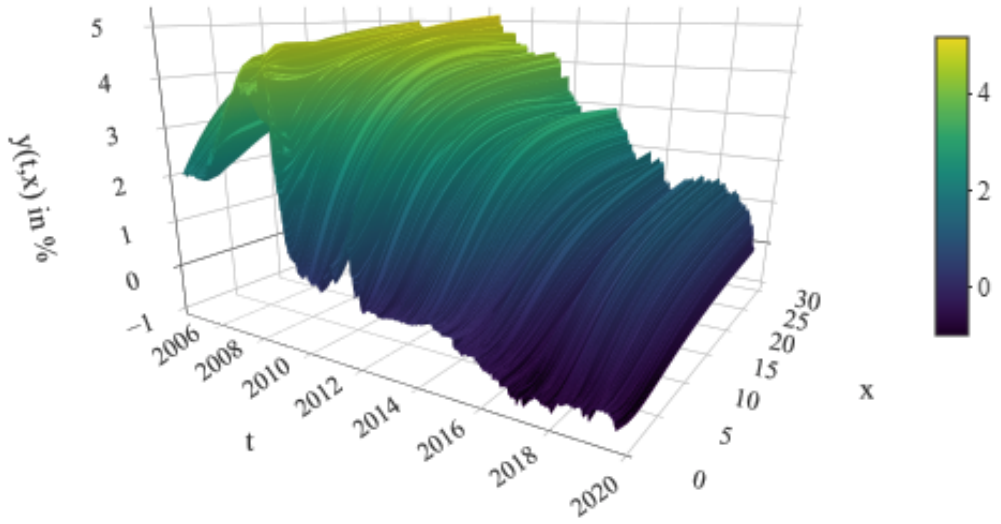


Figure 2.2: Yield curves at the European bond market of triple-A-rated government bonds from 2005 to 2019

Yield curve shapes	2005-2009	2010-2019	2005-2019
Normal	64.19 %	49.59 %	54.46 %
Dipped	15.25 %	44.89 %	34.99 %
Humped-dipped	17.98 %	2.04 %	7.36 %
Humped	2.58 %	3.49 %	3.18 %

Table 2.3: Yield curve shapes frequency occurring on (subsets of) January 1st 2005 to December 31th 2019 calculated from daily data of the ECB restricted to bond maturities up to 20 years

When considering the last 15 years of the yield curve, we can see that the normal one clearly occurs the most. No inverse yield curves can be observed in this data set. Before and during the credit crisis (2005-2009), the normal yield curves clearly dominate and the dipped ones are nearly catching up after the crisis.[3] Humped-dipped yield curves occurred about as frequently as dipped curves from 2005 until 2009 after which they have significantly decreased. The small proportion of humped ones remained approximately the same over the period.

To illustrate four examples of dipped yield curves, we present Figure 2.4 which contains four dipped yield curves at the dates indicated in the subfigures. Note in particular that for the occurrence of dipped yield curves neither the credit crisis nor a very low interest rate is necessary. For an overview of the development of the yields themselves, we refer to Chapter 3, Section 3.4.1.

[3]The same behavior can be observed for German yield curves, see Appendix B of Diez and Korn (2020).

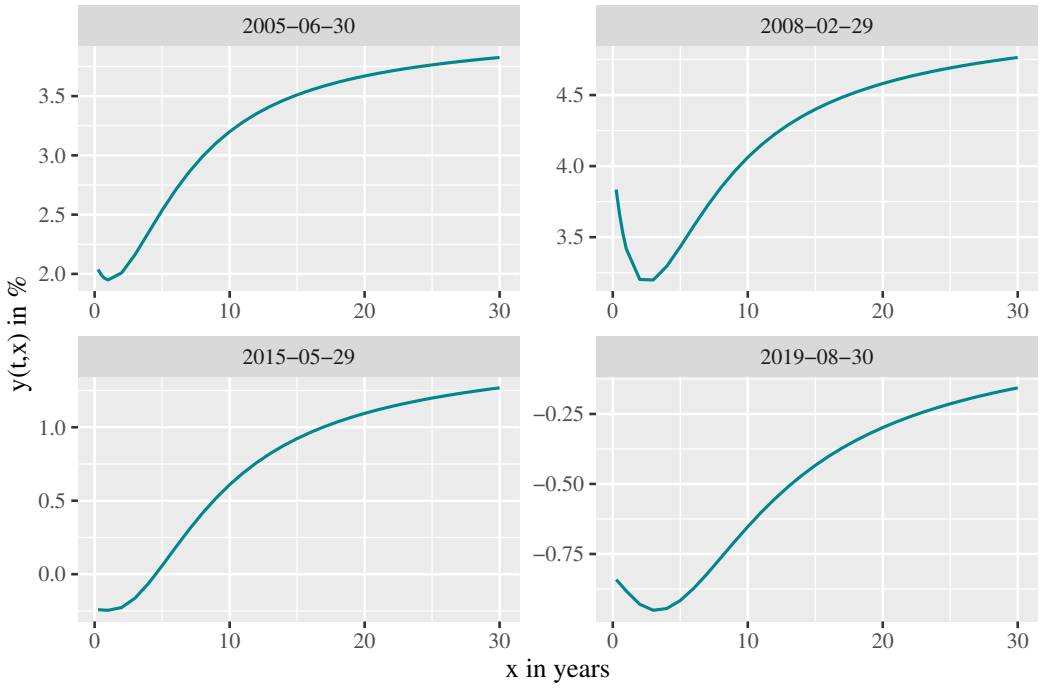


Figure 2.4: Examples of dipped yield curves at the European bond market of triple-A-rated government bonds for four different dates

2.3 Yield Curve Shapes in Vasicek Models

The Vasicek model is the interest rate model that allows the most explicit analytical analysis. We have therefore chosen to consider the one- and two-factor versions of it in this section.

2.3.1 The One-Factor Vasicek Model

The Vasicek model introduced in Vasicek (1977) belongs to the one-factor affine linear short rate models. Due to its analytic tractability, it is still used nowadays even though it can only reproduce normal, inverse, and humped yield curves. In the following, we introduce it together with its properties. Furthermore, the distribution of the different yield curve shapes is analyzed.

Short Rate Dynamics and Bond Prices

In the one-factor Vasicek model (see Brigo and Mecurio (2007), Desmettre and Korn (2018)) the short rate $r(t)$ follows an Ornstein-Uhlenbeck process with constant coefficients. We assume that under the risk-neutral measure $\mathbb{Q}$ the dynamics are given by

$$\mathrm{d}r(t) = a(\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \quad r(0) = r_0 \tag{2.2}$$

where r_0 and θ are constants, a and σ are positive constants, and $W(t)$ is a one dimensional Brownian motion. Consequently, the short rate $r(t)$ is mean-reverting at θ (the *mean-reversion level*) with *mean-reversion speed* a. The drift of the short rate process is negative if the short rate is larger than θ and positive if the short rate is lower than θ.

Integrating the differential equation yields the unique solution

$$r(t) = r_0 e^{-at} + \theta(1 - e^{-at}) + \sigma \int_0^t e^{-a(t-u)} \, \mathrm{d}W(u).$$

Thus, the short rate $r(t)$ is normally distributed with mean and variance given by

$$\begin{aligned} \mathbb{E}\left(r(t)\right) &= r_0 e^{-at} + \theta(1 - e^{-at}), \\ \mathrm{Var}\left(r(t)\right) &= \frac{\sigma^2}{2a}(1 - e^{-2at}). \end{aligned}$$

The limit of the mean and variance for t going to infinity leads to

$$\begin{aligned} \lim_{t\to\infty} \mathbb{E}\left(r(t)\right) &= \theta, \\ \lim_{t\to\infty} \mathrm{Var}\left(r(t)\right) &= \frac{\sigma^2}{2a}. \end{aligned}$$

As a consequence, θ is the long-term average short rate. For $r_0 < \theta$ the mean short rate converges towards θ in a monotonically increasing way and in a monotonically decreasing way for $r_0 > \theta$. The short rate can attain negative values. In the current low interest rate environment, this is not necessarily a drawback of the Vasicek model.

The price $P(t, t+x)$ is given under no arbitrage by

$$P(t, t+x) = \mathbb{E}\left(e^{-\int_t^{t+x} r(u)\,\mathrm{d}u} | \mathcal{F}_t\right)$$

where $\mathcal{F}_t$ is the natural filtration of the process at time t. In the Vasicek model, the price $P(t, t+x)$ has an analytical representation given by

$$\begin{aligned} P(t, t+x) &= e^{A(x) - B(x) r(t)}, \\ A(x) &= \left(\theta - \frac{\sigma^2}{2a^2}\right)(B(x) - x) - \frac{\sigma^2}{4a} B(x)^2, \\ B(x) &= \frac{1}{a}(1 - e^{-ax}). \end{aligned} \tag{2.3}$$

This yields the following formula of the yield $y(t, x)$ as

$$y(t, x) = -\frac{A(x)}{x} + \frac{B(x)}{x} r(t). \tag{2.4}$$

As $A(x)$ and $B(x)$ are deterministic, the yield is a linear function of $r(t)$. Thus, the Vasicek model counts among the affine linear models as already mentioned above.

The entire yield curve is determined by the short rate $r(t)$ and is monotonically increasing in $r(t)$. This means that for a larger value of $r(t)$ the entire yield curve lies above the one with a lower $r(t)$. Due to the explicit affine linear structure, a shock to the short rate $r(t)$ affects the entire yield curve. At any time t, two yields $y(t, x_1)$ and $y(t, x_2)$ are perfectly correlated since

$$\text{Corr}\left(y(t,x_1), y(t,x_2)\right) = \frac{\text{Cov}\left(\frac{B(x_1)}{x_1} r(t), \frac{B(x_2)}{x_2} r(t)\right)}{\sqrt{\text{Var}\left(\frac{B(x_1)}{x_1} r(t)\right)}\sqrt{\text{Var}\left(\frac{B(x_2)}{x_2} r(t)\right)}} = 1.$$

Clearly, these properties are not realistic and a crude simplification.

Hull-White Approach An additional disadvantage of the Vasicek model is that the zero-coupon bond prices which are currently observed in the market, denoted by $P^M(0,T)$, cannot be exactly fitted due to the few free parameters. Hull and White (1990) developed the approach of replacing the free parameters with time-dependent parameters generating more degrees of freedom for an exact fit of $P^M(0,T)$. In the extension of the Vasicek model by the Hull-White approach the short rate dynamic under the risk-neutral measure $\mathbb{Q}$ is given by

$$\mathrm{d}r(t) = \left(\theta(t) - ar(t)\right)\mathrm{d}t + \sigma \mathrm{d}W(t), \quad r(0) = r_0$$

where a and σ are positive constants, r_0 is a constant, and $\theta(t)$ is a time-dependent function chosen in such a manner that the model zero-coupon bond prices $P(0,T)$ correspond to the market zero-coupon bond prices $P^M(0,T)$. This model is often called the one-factor Hull-White model. The explicit solution of the short rate equation is obtained by

$$r(t) = r_0 e^{-at} + \int_0^t e^{-a(t-s)}\theta(s)\,\mathrm{d}s + \sigma \int_0^t e^{-a(t-u)}\,\mathrm{d}W(u).$$

For determining $\theta(t)$, we introduce the market instantaneous forward rate at time zero for the maturity T denoted by $f^M(0,T)$ and calculated by

$$f^M(0,T) := -\frac{\partial ln P^M(0,T)}{\partial T}.$$

It can be shown that with the choice of $\theta(t)$ by

$$\theta(t) = \frac{\partial f^M(0,T)}{\partial T}\Big|_{T=t} + af^M(0,t) + \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)$$

the model zero-coupon bond prices fit the market zero-coupon bond prices.

The zero-coupon bond price is then given by

$$\begin{aligned} P(t,t+x) &= e^{A(t,t+x)-B(x)r(t)}, \\ A(t,t+x) &= ln\left(\frac{P^M(0,t+x)}{P^M(0,t)}\right) + f^M(0,t)B(x) - \frac{\sigma^2}{4a}(1-e^{-2at})B(x)^2, \\ B(x) &= \frac{1}{a}(1-e^{-ax}). \end{aligned}$$

This extension of the Vasicek model is also an affine linear model and analytically tractable. However, $\theta(t)$ is determined by the market instantaneous forward rates which are only observed in principle on the market as well as their derivatives. For it, a continuum of bond prices is required. Since this is not a given on the market, additional assumptions on the shape of the forward rate and yield curves have to be made. Furthermore, the mean-reversion property of the short rate at a long-term level is no longer given since the mean-reversion level is replaced by the non-constant function $\frac{\theta(t)}{a}$. Due to these disadvantages and the missing general analysis of the yield curve shapes, this Hull-White type approach will not be discussed further and we return to the one-factor Vasicek model.

Yield Curve and Yield Curve Shapes

We now take a closer look at the possible yield curve shapes the one-factor Vasicek model can produce. This has already been pointed out in the original paper by Vasicek (1977). Proofs of the claims made there can be found in Keller-Ressel (2018) and Desmettre and Korn (2018) (see also Keller-Ressel and Steiner (2008) who quote contradicting statements from the literature which show that the possible shapes of the yield curves – even for such simple models as the one-factor Vasicek model – are commonly not completely understood). This section is based on the results given in the above two sources. We also consider the case where the short rate is deterministic. This is obtained by σ equal zero.

Applying de L'Hospital's rule provides the limits of the yield curve on the short and long end which are

$$\lim_{x\downarrow 0} y(t,x) = r(t), \tag{2.5}$$

$$\lim_{x\uparrow\infty} y(t,x) = \theta - \frac{\sigma^2}{2a^2} =: \bar{y}. \tag{2.6}$$

Equation (2.1) shows that the yield curve is continuous at $x = 0$. Furthermore, the long-term yield $\bar{y}$ is situated below the long-term average θ of the short rate $r(t)$. Thus, under the risk-neutral measure more yield curves with a larger short-term yield than long-term yield are expected for large t. We will get back to this issue later in Section 2.4.

Although the behavior of the yield curve on the short and long end is known by Equation (2.5) and (2.6), it is not sufficient to determine the shape of the yield curve.

For one-factor time-homogeneous and affine linear short rate models, Keller-Ressel (2018) shows that only inverse, normal, and humped yield curves exist. Each attainable yield curve shape is determined by the level of the current short rate $r(t)$. Keller-Ressel (2018) and Keller-Ressel and Steiner (2008) prove this by a detailed analysis of generalized Riccati equations. Desmettre and Korn (2018) only consider the Vasicek model and derive the yield curve shapes depending on $r(t)$ by an elementary analysis of the explicit yield curve shape.

Theorem 2.6. *Let the risk-neutral short rate process be given by Equation* (2.2)*. Set*

$$b_{norm} := \theta - \frac{3\sigma^2}{4a^2} \quad \textit{and} \quad b_{inv} := \theta.$$

a) For $\sigma > 0$ we have

- *the yield curve can only be normal, inverse, or humped,*
- *the yield curve is normal if $r(t) \leqslant b_{norm}$,*
- *the yield curve is humped if $b_{norm} < r(t) < b_{inv}$,*
- *the yield curve is inverse if $r(t) \geqslant b_{inv}$.*

b) For $\sigma = 0$ we have

- *the yield curve can only be flat, normal, or inverse,*
- *the yield curve is normal if $r(t) < \theta$,*
- *the yield curve is flat if $r(t) = \theta$,*
- *the yield curve is inverse if $r(t) > \theta$.*

Proof. a) We refer to Desmettre and Korn (2018), proof of Theorem 2.53, p. 163 ff.
b) In the case of $\sigma = 0$, the yield curve $y(t,x)$ is deterministic and is calculated as

$$y(t,x) = \theta + \frac{B(x)}{x}\left(r(t) - \theta\right).$$

We obtain the first derivative as

$$\frac{\partial y(t,x)}{\partial x} = \frac{\partial B(x)/x}{\partial x}\left(r(t) - \theta\right).$$

Derivating $\frac{B(x)}{x}$ results in

$$\frac{\partial B(x)/x}{\partial x} = \frac{(1+ax)e^{-ax} - 1}{ax^2} < 0$$

for $a, x > 0$ (see also Lemma A.1 in Appendix A). Consequently, it holds

$$\frac{\partial y(t,x)}{\partial x} \begin{cases} < 0, & r(t) > \theta \\ = 0, & r(t) = \theta \\ > 0, & r(t) < \theta. \end{cases}$$

As $\frac{B(x)}{x}$ is bounded, $y(t,x)$ is also bounded. It follows that the yield curve is normal for $r(t) < \theta$, flat for $r(t) = \theta$, and inverse for $r(t) > \theta$. □

Obviously, under the conditions of Theorem 2.6 we have

$$b_{norm} < \bar{y} < b_{inv}.$$

Note that for σ equal zero the short rate process $r(t)$ is deterministic. In this case, the calculation of the zero-coupon bond price simplifies to

$$P(t, t+x) = e^{-\int_t^{t+x} r(s)\,\mathrm{d}s}.$$

Further, b_{norm}, b_{inv}, and $\bar{y}$ coincide with θ.

Figure 2.5 illustrates Theorem 2.6. Both can in principle be found in Keller-Ressel and Steiner (2008) and Desmettre and Korn (2018). In Figure 2.5a the short rate process implies a random component. Note that the quantities b_{norm}, b_{inv}, and $\bar{y}$ are also shown. The yield curves are normal for $r(t) \leqslant b_{norm}$, humped for $b_{norm} < r(t) < b_{inv}$, and inverse for $r(t) \geqslant b_{inv}$. All yield curves asymptotically tend to $\bar{y}$. Figure 2.5b depicts the attainable yield curve shapes for a deterministic short rate process. Humped yield curves do not exist but a flat one for $r(t)$ equal θ.

(a) $\sigma > 0$. (b) $\sigma = 0$.

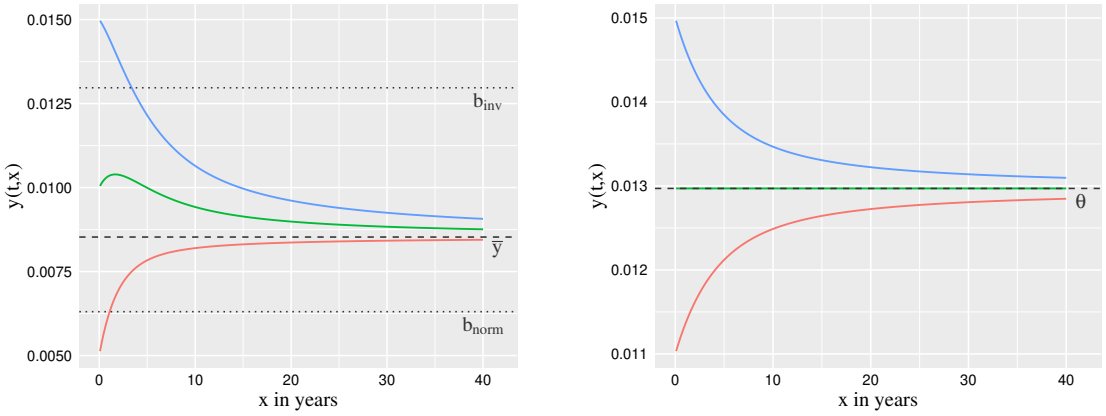


Figure 2.5: Different yield curve shapes in the one-factor Vasicek model

Besides fully classifying the attainable yield curve shapes, it is an important result of Theorem 2.6 that in the one-factor Vasicek model the yield curve shape depends only on the level of the current short rate $r(t)$. Therefore, a global upwards or downwards shift of the entire yield curve is not possible in the Vasicek model. However, such global shifts are often elements of stress tests.

The curvature of the one-factor Vasicek model can also be classified.

Theorem 2.7. *Let the risk-neutral short rate process be given by Equation* (2.2).

a) For $\sigma > 0$ the following assertions hold:

- *Every normal yield curve is strictly concave.*
- *The inverse yield curve is strictly convex if* $r(t) \geqslant b_{inv} + \frac{\sigma^2}{a^2}$.

b) For $\sigma = 0$ *the following assertions hold:*

- *Every normal yield curve is strictly concave.*
- *Every inverse yield curve is strictly convex.*

Proof. a) We consider the following representation of $y(t,x)$:

$$y(t,x) = \theta - \frac{\sigma^2}{2a^2} - \frac{\sigma^2}{4a^2}\frac{B(x)}{x}\left(c - \left(1 - e^{-ax}\right)\right)$$

where

$$c := (\theta - r(t))\,\frac{4a^2}{\sigma^2} - 2.$$

First, we will prove that every normal yield curve is strictly concave. By the proof of Theorem 2.53 in Desmettre and Korn (2018), p. 163 ff, we have $c \geqslant 1$ for normal yield curves. The second derivative of $y(t,x)$ with respect to x is

$$\frac{\partial^2 y\,(t,x)}{\partial x^2} = -\frac{\sigma^2}{4a^2}\left(\frac{\partial^2 B(x)/x}{\partial x^2}\left(c - 1 + e^{-ax}\right) - 2a\frac{\partial B(x)/x}{\partial x}e^{-ax} + a^2\frac{B(x)}{x}e^{-ax}\right).$$

Inspection of $\frac{B(x)}{x}$ delivers the following properties for $a, x > 0$ according to Lemma A.1 in Appendix A:

$$\begin{aligned}
\frac{B(x)}{x} &= \frac{1 - e^{-ax}}{ax} > 0,\\
\frac{\partial B(x)/x}{\partial x} &= \frac{(1+ax)e^{-ax} - 1}{ax^2} < 0,\\
\frac{\partial^2 B(x)/x}{\partial x^2} &= \frac{2\left(1 - (1+ax)e^{-ax}\right) - a^2x^2e^{-ax}}{ax^3} > 0.
\end{aligned}$$

With $a, \sigma, x > 0$ and $e^{-ax} > 0$, we obtain for $c \geqslant 1$

$$\frac{\partial^2 y\,(t,x)}{\partial x^2} \leqslant -\frac{\sigma^2}{4a^2}\left(\frac{\partial^2 B(x)/x}{\partial x^2}e^{-ax} - 2a\frac{\partial B(x)/x}{\partial x}e^{-ax} + a^2\frac{B(x)}{x}e^{-ax}\right) < 0.$$

Therefore, the yield curve is strictly concave for $r(t) \leqslant b_{norm}$.

Next, we will prove that an inverse yield curve is strictly convex if $r(t) \geqslant b_{inv} + \frac{\sigma^2}{a^2}$. We obtain the second derivative of the yield curve $y(t,x)$ as

$$\begin{aligned}
\frac{\partial^2 y(t,x)}{\partial x^2} &= -\frac{\sigma^2}{4a^2}\left(\frac{2(c-1) + (2-c)e^{-ax}\left(a^2x^2 + 2ax + 2\right)}{ax^3}\right.\\
&\qquad\left. - \frac{e^{-2ax}\left(4a^2x^2 + 4ax + 2\right)}{ax^3}\right)\\
&=: -\frac{\sigma^2}{4a^2}\frac{2(c-1) + (2-c)z(x) - z(2x)}{ax^3} =: -\frac{\sigma^2}{4a^2}\frac{k_c(x)}{ax^3}.
\end{aligned}$$

Since $\sigma^2 > 0$ and $4a^3x^3 > 0$, $\frac{\partial^2 y(t,x)}{\partial x^2}$ has the same sign as $k_c(x)$. Therefore, we consider $k_c(x)$. It attains the following limits of

$$\lim_{x \downarrow 0} k_c(x) = 0, \quad \lim_{x \uparrow \infty} k_c(x) = 2(c-1)$$

and has the first derivative of

$$\frac{\partial k_c(x)}{\partial x} = a^3 x^2 e^{-ax} \left(8e^{-ax} - (2-c)\right).$$

As $a^3x^2e^{-ax} > 0$ and e^{-ax} is strictly monotonically decreasing from 1 to 0 for $a, x > 0$, $\frac{\partial k_c(x)}{\partial x}$ has at most one zero for $x > 0$ depending on c and it is given as

$$\overline{x}_c = -\frac{1}{a} ln \left(\frac{2-c}{8}\right).$$

$\frac{\partial k_c(x)}{\partial x}$ does not have any zero in $x > 0$ for $c \leqslant -6$ and we have for $a, x > 0$

$$\frac{\partial k_c(x)}{\partial x} \leqslant 8a^3x^2e^{-ax} \left(e^{-ax} - 1\right) < 0.$$

Consequently, $k_c(x)$ is negative since it strictly monotonically decreases from zero on and thus $\frac{\partial^2 y_c(t,x)}{\partial x^2}$ is positive for all $x > 0$. Therefore, the yield curve $y(t,x)$ is strictly convex if $r(t) \geqslant \theta + \frac{\sigma^2}{a^2}$. Since we have

$$r(t) \geqslant \theta + \frac{\sigma^2}{a^2} > \theta,$$

the yield curve is also inverse due to Theorem 2.6.

b) In the case of $\sigma = 0$, the yield curve $y(t,x)$ is deterministic and is calculated as

$$y(t,x) = \theta + \frac{B(x)}{x} \left(r(t) - \theta\right).$$

We obtain the derivatives as

$$\frac{\partial y(t,x)}{\partial x} = \frac{\partial B(x)/x}{\partial x} \left(r(t) - \theta\right),$$
$$\frac{\partial^2 y(t,x)}{\partial x^2} = \frac{\partial^2 B(x)/x}{\partial x^2} \left(r(t) - \theta\right).$$

As $\frac{\partial^2 B(x)/x}{\partial x^2} > 0$ for $a, x > 0$ as in a) shown, we have

$$\frac{\partial^2 y(t,x)}{\partial x^2} = \begin{cases} < 0, & r(t) < \theta \\ > 0, & r(t) > \theta. \end{cases}$$

Consequently, every normal yield curve is strictly concave and every inverse one is strictly convex for $\sigma = 0$. □

Note that for $\sigma > 0$ the strict convexity condition for the inverse yield curves stated in the theorem is equivalent to the condition $c \leqslant -6$. In particular, one can verify that not all inverse yield curves are strictly convex. According to the above proof, the second derivative of the yield curve has exactly one zero for $-6 < c \leqslant -2$. We can also determine values of c for which even the derivative of $y(t, x)$ is strictly convex or strictly concave (see Lemma 2.12).

Theorem 2.7 further restricts the possible yield curve shapes in the one-factor Vasicek model. Only strictly concave normal yield curves can be generated.

Yet, there are further consequences of the above-mentioned theorems regarding the occurring yield curve shapes. According to Theorem 2.6 humped yield curves occur only in the case of randomness ($\sigma > 0$). Thus, they are a sufficient sign of randomness. Equally, Theorem 2.7 implies that inverse yield curves which are not strictly convex only exist in the case of randomness.

Corollary 2.8. *Let the risk-neutral short rate process be given by Equation* (2.2)*. In that case, the following assertions are equivalent:*

(i) The short rate process $r(t)$ is deterministic.

(ii) There is a value of $r(t)$ such that the yield curve is flat.

(iii) There is no value of $r(t)$ such that the yield curve is humped.

(iv) There is no value of $r(t)$ such that an inverse yield curve is not strictly convex.

Distribution of Yield Curve Shapes under $\mathbb{Q}$

The one-factor Vasicek model generates three shapes of yield curves. However, how likely is the observation of one particular shape in future times? In this section, we give an answer by considering the distribution of the yield curve shapes under the risk-neutral measure $\mathbb{Q}$ for $t > 0$.

As a result of Section 2.3.1, the entire yield curve at time t depends only on the short rate $r(t)$. Furthermore, it is monotonic in $r(t)$ due to the affine linear structure of $y(t, x)$. Thus, the distribution of $r(t)$ determines the distribution of the yield curve shapes. The short rate $r(t)$ is normally distributed under $\mathbb{Q}$ with

$$r(t) \sim \mathcal{N}\left(r_0 e^{-at} + \theta(1 - e^{-at}), \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)\right).$$

Consequently, the proportions at time t of each yield curve shape can be calculated.

Proposition 2.9. *Let the risk-neutral short rate process be given by Equation* (2.2) *with $\sigma > 0$. Let $\mathbb{E}(r(t))$ be the mean of $r(t)$ and $\mathrm{Var}(r(t))$ its variance under the risk-neutral measure $\mathbb{Q}$. For $t > 0$ being fixed, the probabilities of each yield curve shape to appear under $\mathbb{Q}$ are given by:*

$$q_{norm} = P\left(r(t) \leqslant b_{norm}\right) = \Phi\left(\frac{\theta - \frac{3\sigma^2}{4a^2} - \mathbb{E}(r(t))}{\sqrt{\mathrm{Var}(r(t))}}\right),$$

$$q_{hump} = P\left(b_{norm} < r(t) < b_{inv}\right) = \Phi\left(\frac{\theta - \mathbb{E}(r(t))}{\sqrt{\mathrm{Var}(r(t))}}\right) - \Phi\left(\frac{\theta - \frac{3\sigma^2}{4a^2} - \mathbb{E}(r(t))}{\sqrt{\mathrm{Var}(r(t))}}\right),$$

$$q_{inv} = P\left(r(t) \geqslant b_{inv}\right) = 1 - \Phi\left(\frac{\theta - \mathbb{E}(r(t))}{\sqrt{\mathrm{Var}(r(t))}}\right)$$

where Φ is the cumulative distribution function of the standard normal distribution.

The mean of the short rate $r(t)$ converges monotonically increasing towards its limit θ for $r_0 < \theta$. Thus, the proportion of humped and inverse yield curves grows with the passing of time in the case of $r_0 < \theta$. Hence, there is a time t^* when at least $50\ \%$ humped and inverse yield curves are expected. t^* is given by

$$t^* = -\frac{1}{a} ln\left(\frac{b_{norm} - \theta}{r_0 - \theta}\right).$$

For $r_0 > \theta$ the mean of $r(t)$ decreases monotonically to θ. Hence, the proportion of inverse yield curves declines but stays above $50\ \%$ for all times $t > 0$.

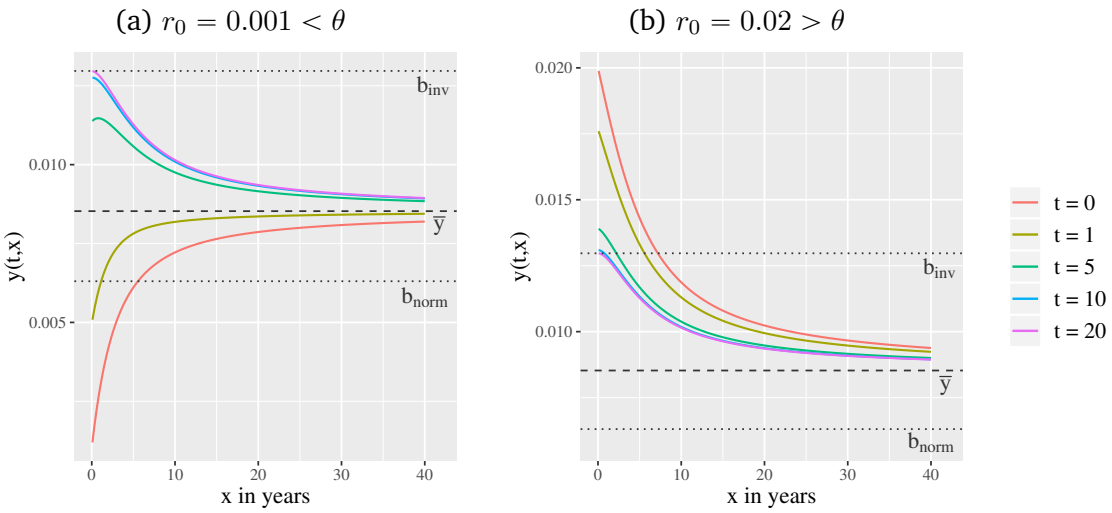


Figure 2.6: Median yield curve at different times t. Parameters: $a = 0.401$, $\theta = 0.01297$, $\sigma = 0.0378$

The evolution of the yield curve is illustrated by a numerical example with the parameter $a = 0.401$, $\theta = 0.01297$, and $\sigma = 0.0378$ (the choice of the parameter, similarly to the remaining examples below, is close to the parameters used for chance-risk

classification of German pension products). We denote the yield curve corresponding to the mean of the short rate $r(t)$ as *median yield curve*. Figures 2.6 shows the median yield curves at different future times t. In Figure 2.6a we choose $r_0 = 0.001$. The initial yield curve is normal due to Theorem 2.6. The mean of $r(1)$ lies below b_{norm}. Thus, at least $50\ \%$ normal yield curves are expected. At $t = 5$ the mean of $r(t)$ is larger than b_{norm}. Hence, more than half of the yield curves are humped or inverse. The same applies for t equal 10 and 20 years. Figure 2.6b illustrates the case for $r_0 > \theta$ with $r_0 = 0.02$. For $\mathbb{E}(r(t)) > b_{inv}$, $t > 0$, $50\ \%$ of the yield curves lie above the plotted yield curves and are also inverse.

For $t \to \infty$ we obtain the following limit proportions:

$$
\begin{aligned}
q_{norm} &\xrightarrow{t\to\infty} \Phi\left(-\frac{3\sigma}{\sqrt{8a^3}}\right), \\
q_{hump} &\xrightarrow{t\to\infty} 0.5 - \Phi\left(-\frac{3\sigma}{\sqrt{8a^3}}\right), \\
q_{inv} &\xrightarrow{t\to\infty} 0.5.
\end{aligned}
$$

Asymptotically, $50\ \%$ inverse yield curves are expected independent of r_0. The other $50\ \%$ are shared by humped and normal yield curves. For $\sigma > 0$ the proportion of humped ones is not zero. Thus, less than $50\ \%$ of normal yield curves are expected in the far future, a behavior that is not in line with empirical observations in Section 2.2.

2.3.2 The Two-Factor Vasicek Model

The two-factor Vasicek model is able to overcome some disadvantages of the one-factor version such as perfectly correlated yields for different times to maturity or the full dependence on $r(t)$ of the yield curve shapes. Further, its analytical tractability is comparable to the one-factor Vasicek model. We will look at its properties and in particular at its corresponding yield curve shapes below. Section 2.3.2 is based on Brigo and Mecurio (2007).

Short Rate Dynamics and Bond Prices

In the two-factor Vasicek model, the short rate is given as the sum of a deterministic function and two correlated Ornstein-Uhlenbeck processes with constant coefficients. Under the risk-neutral measure $\mathbb{Q}$, the short rate $r(t)$ follows

$$
r(t) = \psi(t) + X(t) + Y(t), \quad r(0) = r_0 \tag{2.7}
$$

where $X(t)$ and $Y(t)$ are stochastic processes with dynamics

$$
\mathrm{d}X(t) = -aX(t)\mathrm{d}t + \sigma \mathrm{d}W_1(t), \quad X(0) = 0, \tag{2.8}
$$

$$
\mathrm{d}Y(t) = -bY(t)\mathrm{d}t + \eta \mathrm{d}W_2(t), \quad Y(0) = 0. \tag{2.9}
$$

Here, r_0 is a constant and a, b, σ, and η are positive constants. $W_1(t)$ and $W_2(t)$ are two Brownian motions with correlation $\rho \in (-1, 1)$. The function $\psi(t)$ is deterministic and defined as

$$\psi(t) = r_0 e^{-at} + \theta(1 - e^{-at}) \tag{2.10}$$

where θ is a positive constant and $\psi(0) = r_0$. Integrating of Equations (2.8) and (2.9) leads to

$$X(t) = \sigma \int_0^t e^{-a(t-u)} \, \mathrm{d}W_1(u),$$
$$Y(t) = \eta \int_0^t e^{-b(t-u)} \, \mathrm{d}W_2(u).$$

$X(t)$ and $Y(t)$ are normally distributed with

$$X(t) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)\right),$$
$$Y(t) \sim \mathcal{N}\left(0, \frac{\eta^2}{2b}\left(1 - e^{-2bt}\right)\right).$$

Thus, the short rate $r(t)$ is given by

$$r(t) = r_0 e^{-at} + \theta(1 - e^{-at}) + \sigma \int_0^t e^{-a(t-u)} \, \mathrm{d}W_1(u) + \eta \int_0^t e^{-b(t-u)} \, \mathrm{d}W_2(u).$$

Hence, the short rate $r(t)$ is again normally distributed with

$$\mathbb{E}\left(r(t)\right) = r_0 e^{-at} + \theta(1 - e^{-at}),$$
$$\mathrm{Var}\left(r(t)\right) = \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right) + \frac{\eta^2}{2b}\left(1 - e^{-2bt}\right) + 2\rho\frac{\sigma\eta}{a+b}\left(1 - e^{-(a+b)t}\right).$$

Again, the short rate can also be negative. Its mean is the same as in the one-factor Vasicek model, while the variance contains additional terms due to the second source of randomness. If t goes to infinity, the mean and variance tend to

$$\lim_{t\to\infty} \mathbb{E}\left(r(t)\right) = \theta,$$
$$\lim_{t\to\infty} \mathrm{Var}\left(r(t)\right) = \frac{\sigma^2}{2a} + \frac{\eta^2}{2b} + 2\rho\frac{\sigma\eta}{a+b}.$$

θ is the long-term average rate.

The zero-coupon bond price $P(t, t+x)$ is given by

$$P(t, t+x) = \mathbb{E}\left(e^{-\int_t^{t+x} r(u)\,\mathrm{d}u} | \mathcal{F}_t\right)$$

where the expectation is computed under the risk-neutral measure $\mathbb{Q}$. Noting that conditionally on the natural filtration $\mathcal{F}_t$ we have

$$\int_t^{t+x} (X(u) + Y(u))\,\mathrm{d}u \;\sim\; \mathcal{N}\left(M(t,x), V(x)\right)$$

with mean $M(t,x)$ and variance $V(x)$ given by (see the relevant section in Brigo and Mecurio (2007))

$$M(t,x) = \frac{1-e^{-ax}}{a}X(t) + \frac{1-e^{-bx}}{b}Y(t)$$

and

$$\begin{aligned} V(x) = \frac{\sigma^2}{a^2}\left(x + \frac{2}{a}e^{-ax} - \frac{1}{2a}e^{-2ax} - \frac{3}{2a}\right) + \frac{\eta^2}{b^2}\left(x + \frac{2}{b}e^{-bx} - \frac{1}{2b}e^{-2bx} - \frac{3}{2b}\right) \\ + 2\rho\frac{\sigma\eta}{ab}\left(x - \frac{1-e^{-ax}}{a} - \frac{1-e^{-bx}}{b} + \frac{1-e^{-(a+b)x}}{a+b}\right). \end{aligned}$$

We use these equations and the fact that $\mathbb{E}\left(exp(Z)\right) = exp\left(\mathbb{E}(Z) + \frac{1}{2}\mathrm{Var}(Z)\right)$ for a normally distributed random variable Z with mean $\mathbb{E}(Z)$ and variance $\mathrm{Var}(Z)$ to compute the price at time t of a zero-coupon bond with time to maturity x. Since we have in addition

$$\int_t^{t+x} \psi(u)\,\mathrm{d}u = \theta x + (\psi(x) - \theta)\frac{1-e^{-ax}}{a},$$

we finally obtain the explicit form of the zero bond price as

$$P(t,t+x) = exp\left(-\theta x - \frac{1-e^{-ax}}{a}\left(\psi(t) + X(t) - \theta\right) - \frac{1-e^{-bx}}{b}Y(t) + \frac{1}{2}V(x)\right).$$

The yield curve $y(t,x)$ results directly as

$$y(t,x) = \theta + \frac{B_a(x)}{x}(\psi(t) + X(t) - \theta) + \frac{B_b(x)}{x}Y(t) - \frac{1}{2}\frac{V(x)}{x} \tag{2.11}$$

where we have used the notation

$$B_z(x) = \frac{1}{z}(1 - e^{-zx}),\; z \in \{a,b\}.$$

Due to the linear structure of $y(t,x)$, the two-factor Vasicek model also counts among the affine linear models – even though we are in a multidimensional setting.

Note that with the choice of $\psi(t)$ as in Equation (2.10) the model does in general not fit the zero-coupon bond prices $P^M(0,T)$ which are currently observed in the market. For an exact fit of the market prices $P^M(0,T)$, the deterministic function $\psi(x)$ has to be chosen so that

$$exp\left(-\int_t^{t+x}\psi(u)\,\mathrm{d}u\right) = \frac{P^M(0,t+x)}{P^M(0,t)}exp\left(-\frac{1}{2}(V(t+x) - V(t))\right).$$

This Hull-White type approach is not discussed further as it does not allow a general analysis of the yield curve shapes.

Yield Curve and Yield Curve Shapes

In this section, we analyze the yield curves with regard to their possible shapes. If $r(t)$ is deterministic ($\sigma = 0$, $\eta = 0$), we obtain the same equations and results as in the deterministic case of the one-factor model. Therefore, we consider $\sigma > 0$ and $\eta > 0$ in the following. Equation (2.11) can be transformed in terms of the short rate $r(t)$ to

$$y(t,x) = \theta\left(1 - \frac{B_a(x)}{x}\right) + \frac{B_a(x)}{x}r(t) + \left(\frac{B_b(x)}{x} - \frac{B_a(x)}{x}\right)Y(t) - \frac{1}{2}\frac{V(x)}{x}.$$

Obviously, the yield curve depends not only on the short rate $r(t)$ as in the one-factor Vasicek model but also on the stochastic process $Y(t)$. Therefore, different yield curves can exist for one and the same short rate level $r(t)$.

Considering all terms which are forming $y(t,x)$ separately, we obtain the limits of the yield curve on the short and long end as

$$\begin{aligned}\lim_{x\downarrow 0} y(t,x) &= r(t),\\ \lim_{x\uparrow\infty} y(t,x) &= \theta - \frac{\sigma^2}{2a^2} - \frac{\eta^2}{2b^2} - \rho\frac{\sigma\eta}{ab} =: \bar{y}.\end{aligned}$$

As in the one-factor Vasicek model, the yield curve is continuous at x equal to zero. Since the mean of the short rate $r(t)$ tends to θ, more yield curves with a larger short-term yield than long-term yield are expected for large t. This observation is identical to the one in the one-factor Vasicek model. Note that this also holds true for $\rho < 0$. This follows from $|\rho| < 1$ and the binomial formula.

For presenting results that lead to different behavior of the two-factor Vasicek yield curve, we analyze its derivative at x equal to zero.

Theorem 2.10. *Let the risk-neutral short rate process be given by Equation* (2.7)*. Under the assumption of $a \neq b$, pairs of $(X(t), Y(t))$ exist with $\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} > 0$ and $\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} < 0$ for all values of the short rate $r(t)$.*

Proof. Differentiating the yield as given by Equation (2.11) with respect to x delivers

$$\begin{aligned}\frac{\partial y(t,x)}{\partial x} =& \frac{(1+ax)e^{-ax}-1}{ax^2}(\psi(t) + X(t) - \theta) + \frac{(1+bx)e^{-bx}-1}{bx^2}Y(t)\\ &+ \frac{\sigma^2}{2a^2}\left(2\frac{(1+ax)e^{-ax}-1}{ax^2} - \frac{(1+2ax)e^{-2ax}-1}{2ax^2}\right)\\ &+ \frac{\eta^2}{2b^2}\left(2\frac{(1+bx)e^{-bx}-1}{bx^2} - \frac{(1+2bx)e^{-2bx}-1}{2bx^2}\right)\\ &+ \rho\frac{\sigma\eta}{ab}\left(\frac{(1+ax)e^{-ax}-1}{ax^2} + \frac{(1+bx)e^{-bx}-1}{bx^2}\right.\\ &\qquad\left. - \frac{(1+(a+b)x)e^{-(a+b)x}-1}{(a+b)x^2}\right).\end{aligned}$$

Applying de L'Hospital's rule twice at $x = 0$, we get

$$\lim_{x \downarrow 0} \frac{(1+zx)e^{-zx} - 1}{zx^2} = -\frac{1}{2}z$$

for an arbitrary non-zero constant z. Using this for the relevant terms in the derivative above, we obtain

$$\lim_{x \downarrow 0} \frac{\partial y(t,x)}{\partial x} = -\frac{1}{2}\left(a\left(r(t) - \theta\right) - (a-b)Y(t)\right).$$

For a fixed value of $r(t)$, we get different values of the yield curve derivative at $x = 0$ by choosing different values of $Y(t)$. The corresponding value of $X(t)$ is obtained via $X(t) = r(t) - \psi(t) - Y(t)$. □

Theorem 2.10 implies that different derivatives of the yield curve at the short end exist for one and the same short rate $r(t)$. Since the asymptotic long-term yield is known, at least four yield curve shapes can occur in the two-factor Vasicek model.[4] Note that the derivative of the yield curve at the short end cannot vary for $a = b$ and depends on the relation between $r(t)$ and θ.

Before we specify the yield curve shapes, the following useful lemmas are introduced.

Lemma 2.11. *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$, be strictly concave (convex) and $f(x) \neq g(x)$. Then, both functions have a maximum of two intersections.*

Proof. We prove this lemma for f and g being strictly concave. For f and g being strictly convex, the proof follows analogously.

We assume that f and g have three intersections: x_1, x_2, and x_3 with $x_1 \neq x_2 \neq x_3$. Furthermore, we assume $x_1 < x_2 < x_3$ without loss of generality. Since f and g are strictly concave, it holds for $\alpha \in (0;1)$

$$f(\alpha x_1 + (1-\alpha)x_3) > \alpha f(x_1) + (1-\alpha)f(x_3),$$
$$g(\alpha x_1 + (1-\alpha)x_3) > \alpha g(x_1) + (1-\alpha)g(x_3).$$

Taking the difference between the two inequalities, we obtain for $\alpha \in (0;1)$

$$f(\alpha x_1 + (1-\alpha)x_3) - g(\alpha x_1 + (1-\alpha)x_3) > 0$$

because $f(x_1) = g(x_1)$ and $f(x_3) = g(x_3)$. Since $x_1 < x_2 < x_3$, there is an $\bar{\alpha} \in (0;1)$ so that $x_2 = \bar{\alpha}x_1 + (1-\bar{\alpha})x_3$ and the previous inequality holds for x_2

$$f(x_2) - g(x_2) = f(\bar{\alpha}x_1 + (1-\bar{\alpha})x_3) - g(\bar{\alpha}x_1 + (1-\bar{\alpha})x_3) > 0.$$

However, this is a contradiction to the assumption that x_2 is an intersection of both functions for which $f(x_2) = g(x_2)$. Consequently, f and g can have a maximum of two intersections. □

[4] In the working paper Keller-Ressel (2019) which appeared during the revision process of the European Actuarial Journal the results of Theorem 2.13 are generalized using the concept of total positivity.

Lemma 2.12. *Let $a, \sigma, x > 0$ and $K_c(x)$ be defined as*

$$K_c(x) := -\frac{\sigma^2}{4a^2}\frac{B_a(x)}{x}\left(c - \left(1 - e^{-ax}\right)\right).$$

The first derivative of $K_c(x)$ with respect to x is

- *strictly convex for $c \geqslant 1$,*
- *strictly concave for $c \leqslant -14$.*

Proof. First, we will prove that $\frac{\partial K_c(x)}{\partial x}$ is strictly convex for $c \geqslant 1$. The third derivative of $K_c(x)$ with respect to x is

$$\frac{\partial^3 K_c(x)}{\partial x^3} = -\frac{\sigma^2}{4a^2}\left(\frac{\partial^3 B_a(x)/x}{\partial x^3}\left(c - (1 - e^{-ax})\right) - 3a\frac{\partial^2 B_a(x)/x}{\partial x^2}e^{-ax} + 3a^2\frac{\partial B_a(x)/x}{\partial x}e^{-ax} - a^3\frac{B_a(x)}{x}e^{-ax}\right).$$

Inspection of $\frac{B_a(x)}{x}$ delivers the following properties for $a, x > 0$ according to Lemma A.1 in Appendix A:

$$\begin{aligned}
\frac{B_a(x)}{x} &= \frac{1 - e^{-ax}}{ax} > 0,\\
\frac{\partial B_a(x)/x}{\partial x} &= \frac{(1 + ax)e^{-ax} - 1}{ax^2} < 0,\\
\frac{\partial^2 B_a(x)/x}{\partial x^2} &= \frac{2\left(1 - (1 + ax)e^{-ax}\right) - a^2x^2e^{-ax}}{ax^3} > 0,\\
\frac{\partial^3 B_a(x)/x}{\partial x^3} &= \frac{6\left((1 + ax)\,e^{-ax} - 1\right) + 3a^2x^2e^{-ax} + a^3x^3e^{-ax}}{ax^4} < 0.
\end{aligned}$$

With $e^{-ax} > 0$ we obtain

$$\frac{\partial^3 K_c(x)}{\partial x^3} > 0$$

for $c \geqslant 1$. Therefore, the first derivative of $K_c(x)$ is strictly convex for $c \geqslant 1$.

Next, we will prove that $\frac{\partial K_c(x)}{\partial x}$ is strictly concave for $c \leqslant -14$. The third derivative of $K_c(x)$ can be rewritten as

$$\begin{aligned}
\frac{\partial^3 K_c(x)}{\partial x^3} &= \frac{\sigma^2}{4a^2}\left(\frac{6(c-1) + (2-c)e^{-ax}\left(a^3x^3 + 3a^2x^2 + 6ax + 6\right)}{ax^4} - \frac{e^{-2ax}\left(8a^3x^3 + 12a^2x^2 + 12ax + 6\right)}{ax^4}\right)\\
&=: \frac{\sigma^2}{4a^2}\frac{6(c-1) + (2-c)z(x) - z(2x)}{ax^4} =: \frac{\sigma^2}{4a^2}\frac{k_c(x)}{ax^4}.
\end{aligned}$$

Since $\sigma^2 > 0$ and $4a^3x^4 > 0$, $\frac{\partial^3 K_c(x)}{\partial x^3}$ has the same sign as $k_c(x)$. Therefore, we consider $k_c(x)$. It attains the following limits

$$\lim_{x \downarrow 0} k_c(x) = 0, \quad \lim_{x \uparrow \infty} k_c(x) = 6(c-1)$$

and has the first derivative of

$$\frac{\partial k_c(x)}{\partial x} = a^4 x^3 e^{-ax} \left(16e^{-ax} - (2-c)\right).$$

As $a^4x^3e^{-ax} > 0$ as well as e^{-ax} is strictly monotonically decreasing from 1 to 0 for $a, x > 0$, $\frac{\partial k_c(x)}{\partial x}$ has at most one zero in $x > 0$ depending on c and it is given by

$$\overline{x}_c = -\frac{1}{a} ln \left(\frac{2-c}{16}\right).$$

$\frac{\partial k_c(x)}{\partial x}$ does not have any zero in $x > 0$ for $c \leqslant -14$ and we have

$$\frac{\partial k_c(x)}{\partial x} \leqslant 16a^4 x^3 e^{-ax} \left(e^{-ax} - 1\right) < 0.$$

Consequently, $k_c(x)$ is negative since it strictly monotonically decreases from zero on and thus $\frac{\partial^3 K_c(x)}{\partial x^3}$ is negative for $x > 0$. Therefore, the first derivative of $K_c(x)$ is strictly concave if $c \leqslant -14$. □

We can now specify yield curve shapes generated by the two-factor Vasicek model summarized in the following theorem.

Theorem 2.13. *For the risk-neutral short rate process given by Equation* (2.7)*, the following assertions hold under the assumption of* $a \neq b$*:*[5]

(i) There are normal yield curves but only for $r(t) < \bar{y}$*,*

(ii) there are inverse yield curves but only for $r(t) > \bar{y}$*,*

(iii) there are humped yield curves for $r(t) \geqslant \bar{y}$*,*

(iv) there are dipped yield curves for $r(t) \leqslant \bar{y}$*.*

Proof. Note first that although the one-factor Vasicek model is a special case of the two-factor one, we cannot simply infer the existence of the three classical yield curve shapes by setting $Y(t)$ equal zero. This is due to the fact that the remaining ingredients related to the second factor still remain in the yield representation. On top of that, even for $Y(t)$ equal zero their contribution to the yield curve depends on the correlation coefficient ρ. Thus, we also have to prove the above-mentioned assertions for normal, inverse, and humped yield curves.

[5] In contrast to the published version Diez and Korn (2020), the assertions *(iii)* and *(iv)* have been modified leading to a different proof.

Depending on ρ two transformations of the yield curve $y(t,x)$ are used:
Transformation 1:

$$\begin{aligned} y(t,x) =& -\frac{\sigma^2}{4a^2}\frac{B_a(x)}{x}\left(c_A - \left(1-e^{-ax}\right)\right) - \frac{\eta^2}{4b^2}\frac{B_b(x)}{x}\left(c_B - \left(1-e^{-bx}\right)\right) \\ & - \rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} + \theta - \frac{\sigma^2}{2a^2} - \frac{\eta^2}{2b^2} - \rho\frac{\sigma\eta}{ab} \\ =:& K_{c_A}(x) + K_{c_B}(x) + R_1(x) + \bar{y} \end{aligned}$$

where

$$\begin{aligned} c_A &:= -\frac{4a^2}{\sigma^2}\left(X(t) + \psi(t) - \theta + \frac{\sigma^2}{2a^2} + \rho\frac{\sigma\eta}{ab}\right), \\ c_B &:= -\frac{4b^2}{\eta^2}\left(Y(t) + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right). \end{aligned}$$

Transformation 2:

$$\begin{aligned} y(t,x) =& -\frac{\sigma^2}{4a^2}\frac{B_a(x)}{x}\left(c_A - \left(1-e^{-ax}\right)\right) - \frac{\eta^2}{4b^2}\frac{B_b(x)}{x}\left(c_B - \left(1-e^{-bx}\right)\right) \\ & + \rho\frac{\sigma\eta}{ab}\left(\frac{B_a(x)}{x} + \frac{B_b(x)}{x} - \frac{B_{a+b}(x)}{x}\right) + \theta - \frac{\sigma^2}{2a^2} - \frac{\eta^2}{2b^2} - \rho\frac{\sigma\eta}{ab} \\ =:& K_{c_A}(x) + K_{c_B}(x) + R_2(x) + \bar{y} \end{aligned}$$

where

$$\begin{aligned} c_A &:= -\frac{4a^2}{\sigma^2}\left(X(t) + \psi(t) - \theta + \frac{\sigma^2}{2a^2}\right), \\ c_B &:= -\frac{4b^2}{\eta^2}\left(Y(t) + \frac{\eta^2}{2b^2}\right). \end{aligned}$$

The function $K_c(x)$, $c \in \{c_A, c_B\}$, has the following specific properties depending on c for $x > 0$ as shown in Theorem 2.6 (see also Desmettre and Korn (2018), proof of Theorem 2.53, p. 163 ff), Theorem 2.7, and Lemma 2.12:

- $K_c(x)$ is strictly monotonically increasing and strictly concave for $c \geqslant 1$,
- $K_c(x)$ is strictly monotonically decreasing for $c \leqslant -2$,
- $K_c(x)$ is strictly monotonically decreasing and strictly convex for $c \leqslant -6$,
- $\frac{\partial K_c(x)}{\partial x}$ is strictly monotonically decreasing and strictly convex for $c \geqslant 1$,
- $\frac{\partial K_c(x)}{\partial x}$ is strictly monotonically increasing and strictly concave for $c \leqslant -14$.

Further, it can be shown in a basic way that the functions $R_1(x)$ and $R_2(x)$ have the following properties:

- $R_1(x)$ is strictly monotonically increasing and strictly concave for $\rho > 0$,
- $R_1(x)$ is constant zero for $\rho = 0$,
- $R_1(x)$ is strictly monotonically decreasing and strictly convex for $\rho < 0$,
- $\frac{\partial R_1(x)}{\partial x}$ is strictly monotonically decreasing and strictly convex for $\rho > 0$,
- $\frac{\partial R_1(x)}{\partial x}$ is constant zero for $\rho = 0$,
- $\frac{\partial R_1(x)}{\partial x}$ is strictly monotonically increasing and strictly concave for $\rho < 0$,
- $R_2(x)$ is strictly monotonically decreasing for $\rho > 0$,
- $R_2(x)$ is constant zero for $\rho = 0$,
- $R_2(x)$ is strictly monotonically increasing for $\rho < 0$.

(i) As all yield curves converge towards the long-term yield $\bar{y}$, there can only be normal yield curves if we have $r(t) < \bar{y}$. Otherwise the necessary condition for a normal yield curve of

$$\lim_{x \downarrow 0} \frac{\partial y(t,x)}{\partial x} > 0$$

together with the convergence of $y(t,x)$ towards $\bar{y}$ leads to the existence of a local maximum $y(t,x^*)$ for $x^* > 0$ in the case of $r(t) > \bar{y}$ which contradicts the normality of the yield curve.

Next, we will show the existence of normal yield curves. In the case of $\rho \geqslant 0$, we use Transformation 1 of $y(t,x)$ as mentioned above. $X(t)$ and $Y(t)$ are chosen in such a way that $K_{c_A}(x)$ and $K_{c_B}(x)$ are both strictly monotonically increasing. This holds valid for $c_A \geqslant 1$ and $c_B \geqslant 1$ and hence $X(t)+\psi(t) \leqslant \theta - \frac{3\sigma^2}{4a^2} - \rho\frac{\sigma\eta}{ab}$ and $Y(t) \leqslant -\frac{3\eta^2}{4b^2} - \rho\frac{\sigma\eta}{ab}$. $R_1(x)$ is also strictly monotonically increasing or constant zero. As the sum of strictly monotonically increasing functions, $y(t,x)$ is strictly monotonically increasing. Since $y(t,x)$ converges towards the long-term yield $\bar{y}$, it is also bounded from above. Consequently, the yield curve is normal for $X(t)$ and $Y(t)$ above. Further, the corresponding short rate is below $\bar{y}$.

For $\rho < 0$ we consider the Transformation 2 of the yield curve. $K_{c_A}(x)$ and $K_{c_B}(x)$ are chosen again to be strictly monotonically increasing via $c_A \geqslant 1$ and $c_B \geqslant 1$. This leads to $X(t)+\psi(t) \leqslant \theta - \frac{3\sigma^2}{4a^2}$ and $Y(t) \leqslant -\frac{3\eta^2}{4b^2}$. $R_2(x)$ is also strictly monotonically increasing. Thus, $y(t,x)$ is also strictly monotonically increasing and bounded from above for $X(t)$ and $Y(t)$ as seen above. Again, the related short rate $r(t)$ is below $\bar{y}$.

(ii) The case for inverse yield curves follows with obvious modifications if we now consider that necessary conditions for the relevant forms of $K_{c_A}(x)$ and $K_{c_B}(x)$ to be strictly monotonically decreasing are $c_A \leqslant -2$ and $c_B \leqslant -2$.

(iii) We consider Transformation 1 of $y(t,x)$ and a short rate $r(t) \geqslant \bar{y}$.

We start with the case $a < b$. We choose c_B as

$$c_B > max\left\{1; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right) + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right)\right\}$$

and c_A for $\rho \leqslant 0$ as

$$c_A \leqslant min\left\{-14; -\frac{a^2\eta^2}{b^2\sigma^2}c_B\right\}$$

and for $\rho > 0$ as

$$c_A \leqslant min\left\{-14; -\frac{4a^2}{\sigma^2}\left(\frac{\eta^2}{4b^2}c_B + \rho\frac{\sigma\eta}{ab}\right)\right\}.$$

The first derivative of $y(t,x)$ with respect to x at $x = 0$ is given by (see proof of Theorem 2.10)

$$\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} = -\frac{1}{2}\left(a\left(r(t)-\theta\right) - (a-b)Y(t)\right).$$

Solving the definition of c_B to $Y(t)$ returns

$$Y(t) = -\frac{\eta^2}{4b^2}c_B - \frac{\eta^2}{2b^2} - \rho\frac{\sigma\eta}{ab}.$$

With this we have

$$\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} = -\frac{1}{2}\left(a\left(r(t)-\theta\right) + (a-b)\left(\frac{\eta^2}{4b^2}c_B + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right)\right).$$

Since we choose $c_B > -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right) + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right)$, it holds

$$\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} > 0.$$

Next, we consider the difference between $y(t,x)$ and $\bar{y}$. $\frac{B_z(x)}{x}$ is positive and strictly monotonically decreasing in $z > 0$ for $x > 0$. Therefore, it holds

$$\frac{B_a(x)}{x} > \frac{B_b(x)}{x} > \frac{B_{a+b}(x)}{x} > 0. \tag{2.12}$$

With these inequalities, $c_B > 1$, and $1 - e^{-zx} > 0$ for $x, z > 0$, we can estimate the difference as follows:

$$\begin{aligned} y(t,x) - \bar{y} &= -\frac{\sigma^2}{4a^2}\frac{B_a(x)}{x}\left(c_A - \left(1 - e^{-ax}\right)\right) - \frac{\eta^2}{4b^2}\frac{B_b(x)}{x}\left(c_B - \left(1 - e^{-bx}\right)\right) \\ &\quad - \rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} \\ &> -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}c_A + \frac{\eta^2}{4b^2}c_B\right) - \rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x}. \end{aligned}$$

For $\rho \leqslant 0$ we have

$$-\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}c_A + \frac{\eta^2}{4b^2}c_B\right) - \rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} \geqslant -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}c_A + \frac{\eta^2}{4b^2}c_B\right).$$

With the choice of $c_A \leqslant -\frac{a^2\eta^2}{b^2\sigma^2}c_B$, we obtain $y(t,x) - \bar{y} > 0$ for all $x > 0$.

For $\rho > 0$ we have

$$-\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}c_A + \frac{\eta^2}{4b^2}c_B\right) - \rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} > -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}c_A + \frac{\eta^2}{4b^2}c_B + \rho\frac{\sigma\eta}{ab}\right).$$

Due to the choice of $c_A \leqslant -\frac{4a^2}{\sigma^2}\left(\frac{\eta^2}{4b^2}c_B + \rho\frac{\sigma\eta}{ab}\right)$, $y(t,x) - \bar{y} > 0$ for all $x > 0$.

Consequently, $y(t,x)$ is greater than $\bar{y}$ for the choice of c_A and c_B.

Since $r(t)$ is greater than or equal to $\bar{y}$, $\lim\limits_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x}$ is positive, $y(t,x)$ converges towards $\bar{y}$ on the long end, and $y(t,x)$ lies above $\bar{y}$, the yield curve has an odd number of local extreme values.

Due to $c_A \leqslant -14$ and $c_B > 1$, $y(t,x)$ is the sum of strictly concave and strictly convex functions. $y(t,x)$ has a local extreme value if for $\rho \leqslant 0$ it holds

$$\frac{\partial K_{c_A}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x} = -\frac{\partial K_{c_B}(x)}{\partial x}$$

and if for $\rho > 0$ it holds

$$\frac{\partial K_{c_A}(x)}{\partial x} = -\left(\frac{\partial K_{c_B}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x}\right).$$

$\frac{\partial K_{c_A}(x)}{\partial x}$ is strictly concave and $\frac{\partial K_{c_B}(x)}{\partial x}$ is strictly convex due to the choice of c_A and c_B. $\frac{\partial R_1(x)}{\partial x}$ is strictly concave for $\rho < 0$, it is constant zero for $\rho = 0$, and it is strictly convex for $\rho > 0$. Consequently, $\frac{\partial K_{c_A}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x}$ and $-\frac{\partial K_{c_B}(x)}{\partial x}$ have a maximum of two intersections for $\rho \leqslant 0$ according to Lemma 2.11. The same holds for $\frac{\partial K_{c_A}(x)}{\partial x}$ and $-\left(\frac{\partial K_{c_B}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x}\right)$ with $\rho > 0$. Therefore, the derivative

of $y(t,x)$ has a maximum of two zeros and thus $y(t,x)$ a maximum of two local extreme values.

Combined with the odd number of local extreme values that the yield curve must have, it follows that the yield curve is humped for the above choice of $r(t)$, c_A, and c_B.

For $a > b$ we choose c_A as

$$c_A \geqslant 1$$

and c_B for $\rho \leqslant 0$ as

$$c_B < min\left\{-14; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right)+\frac{\eta^2}{2b^2}+\rho\frac{\sigma\eta}{ab}\right); -\frac{b^2\sigma^2}{a^2\eta^2}c_A\right\}$$

and for $\rho > 0$ as

$$c_B < min\left\{-14; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right)+\frac{\eta^2}{2b^2}+\rho\frac{\sigma\eta}{ab}\right); -\frac{4b^2}{\eta^2}\left(\frac{\sigma^2}{4a^2}c_A+\rho\frac{\sigma\eta}{ab}\right)\right\}.$$

Analogously to $a < b$, it can be shown that $\lim\limits_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} > 0$ and $y(t,x) > \bar{y}$ for the choice of c_A and c_B. Due to the strict curvature of the derivative of $K_{c_A}(x)$, $K_{c_B}(x)$, and $R_1(x)$, $y(t,x)$ has a maximum of two local extreme values according to Lemma 2.11. Together with $r(t) \geqslant \bar{y}$, $\lim\limits_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} > 0$, $\lim\limits_{x\uparrow\infty} y(t,x) = \bar{y}$ as well as $y(t,x) > \bar{y}$, it follows that $y(t,x)$ is humped for the above choice of $r(t)$, c_A, and c_B.

(iv) We consider Transformation 1 of $y(t,x)$ and a short rate $r(t) \leqslant \bar{y}$. For $a < b$ we choose c_B as

$$c_B < min\left\{-14; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right)+\frac{\eta^2}{2b^2}+\rho\frac{\sigma\eta}{ab}\right)\right\}$$

and c_A for $\rho \leqslant 0$ as

$$c_A \geqslant max\left\{1; -\frac{4a^2}{\sigma^2}\left(\frac{\eta^2}{4b^2}(c_B-1)+\rho\frac{\sigma\eta}{ab}\right)+1\right\}$$

and for $\rho > 0$ as

$$c_A \geqslant max\left\{1; -\frac{a^2\eta^2}{b^2\sigma^2}\left(c_B-1\right)+1\right\}.$$

As in *(iii)*, the first derivative of $y(t,x)$ with respect to x at $x = 0$ is given by

$$\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} = -\frac{1}{2}\left(a\left(r(t)-\theta\right)+(a-b)\left(\frac{\eta^2}{4b^2}c_B+\frac{\eta^2}{2b^2}+\rho\frac{\sigma\eta}{ab}\right)\right).$$

Since $c_B < -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right)+\frac{\eta^2}{2b^2}+\rho\frac{\sigma\eta}{ab}\right)$, it holds

$$\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x} < 0.$$

Next, we consider the difference between $y(t,x)$ and $\bar{y}$. With the properties of $\frac{B_z(x)}{x}$ according to Inequality (2.12), $1-e^{-zx}<1$ for $x,z>0$, and $c_B<-14$, we obtain

$$\begin{aligned} y(t,x)-\bar{y} =& -\frac{\sigma^2}{4a^2}\frac{B_a(x)}{x}\left(c_A-\left(1-e^{-ax}\right)\right)-\frac{\eta^2}{4b^2}\frac{B_b(x)}{x}\left(c_B-\left(1-e^{-bx}\right)\right) \\ & -\rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} \\ <& -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}\left(c_A-1\right)+\frac{\eta^2}{4b^2}\left(c_B-1\right)\right)-\rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x}. \end{aligned}$$

For $\rho \leqslant 0$ we have

$$\begin{aligned} & -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}(c_A-1)+\frac{\eta^2}{4b^2}(c_B-1)\right)-\rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} \\ \leqslant & -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}(c_A-1)+\frac{\eta^2}{4b^2}(c_B-1)+\rho\frac{\sigma\eta}{ab}\right). \end{aligned}$$

Due to the choice of $c_A \geqslant -\frac{4a^2}{\sigma^2}\left(\frac{\eta^2}{4b^2}(c_B-1)+\rho\frac{\sigma\eta}{ab}\right)+1$, we obtain $y(t,x)-\bar{y}<0$ for all $x>0$.

For $\rho>0$ we have

$$\begin{aligned} & -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}(c_A-1)+\frac{\eta^2}{4b^2}(c_B-1)\right)-\rho\frac{\sigma\eta}{ab}\frac{B_{a+b}(x)}{x} \\ < & -\frac{B_a(x)}{x}\left(\frac{\sigma^2}{4a^2}(c_A-1)+\frac{\eta^2}{4b^2}(c_B-1)\right). \end{aligned}$$

With the choice of $c_A \geqslant -\frac{a^2\eta^2}{b^2\sigma^2}(c_B-1)+1$, we obtain $y(t,x)-\bar{y}<0$ for all $x>0$. Consequently, $y(t,x)$ is less than $\bar{y}$ for the choice of c_A and c_B.

Since $r(t)$ is less than or equal to $\bar{y}$, $\lim_{x\downarrow 0}\frac{\partial y(t,x)}{\partial x}$ is negative, $y(t,x)$ converges towards $\bar{y}$ on the long end, and $y(t,x)$ is less than $\bar{y}$, $y(t,x)$ has an odd number of local extreme values. It has a local extreme value if for $\rho \leqslant 0$ it holds

$$\frac{\partial K_{c_A}(x)}{\partial x} = -\left(\frac{\partial K_{c_B}(x)}{\partial x}+\frac{\partial R_1(x)}{\partial x}\right).$$

and if for $\rho>0$ it holds

$$\frac{\partial K_{c_A}(x)}{\partial x}+\frac{\partial R_1(x)}{\partial x} = -\frac{\partial K_{c_B}(x)}{\partial x}.$$

As $c_A \geqslant 1$ and $c_B < -14$, $\frac{\partial K_{c_A}(x)}{\partial x}$ is strictly convex and $\frac{\partial K_{c_B}(x)}{\partial x}$ is strictly concave. $\frac{\partial R_1(x)}{\partial x}$ is strictly concave for $\rho < 0$, it is constant zero for $\rho = 0$, and it is strictly convex for $\rho > 0$. Consequently, $\frac{\partial K_{c_A}(x)}{\partial x}$ and $-\left(\frac{\partial K_{c_B}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x}\right)$ have a maximum of two intersections for $\rho \leqslant 0$ according to Lemma 2.11. The same holds for $\frac{\partial K_{c_A}(x)}{\partial x} + \frac{\partial R_1(x)}{\partial x}$ and $-\frac{\partial K_{c_B}(x)}{\partial x}$ with $\rho > 0$. Therefore, the derivative of $y(t,x)$ has a maximum of two zeros and thus $y(t,x)$ a maximum of two local extreme values. Combined with the odd number of local extreme values that the yield curve must have, it follows that $y(t,x)$ is dipped for the above choice of $r(t)$, c_A, and c_B.

For $a > b$ we choose c_A as

$$c_A \leqslant -14$$

and c_B for $\rho \leqslant 0$ as

$$c_B > max\left\{1; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right) + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right); \right.$$
$$\left. -\frac{4b^2}{\eta^2}\left(\frac{\sigma^2}{4a^2}(c_A - 1) + \rho\frac{\sigma\eta}{ab}\right) + 1\right\}$$

and for $\rho > 0$

$$c_B > max\left\{1; -\frac{4b^2}{\eta^2}\left(\frac{a}{a-b}\left(r(t)-\theta\right) + \frac{\eta^2}{2b^2} + \rho\frac{\sigma\eta}{ab}\right); -\frac{b^2\sigma^2}{a^2\eta^2}(c_A - 1) + 1\right\}.$$

Analogously to $a < b$, it can be shown that $\lim_{x\downarrow 0} \frac{\partial y(t,x)}{\partial x} < 0$ and $y(t,x) < \bar{y}$. Due to the strict curvature of the derivative of $K_{c_A}(x)$, $K_{c_B}(x)$, and $R_1(x)$, $y(t,x)$ has a maximum of two local extreme values according to Lemma 2.11. Together with $r(t) \leqslant \bar{y}$, $\lim_{x\downarrow 0} \frac{\partial y(t,x)}{\partial x} < 0$, $\lim_{x\uparrow\infty} y(t,x) = \bar{y}$, and $y(t,x) < \bar{y}$, it follows that $y(t,x)$ is dipped for the above choice of $r(t)$, c_A, and c_B.

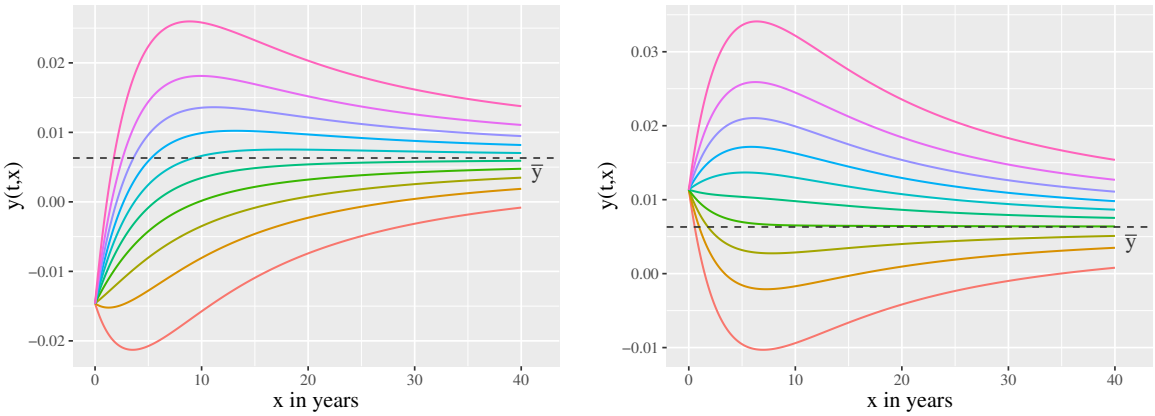
□

Hence, the theorem states that all three yield curve shapes of the one-factor Vasicek model are also produced by the two-factor Vasicek model. However, there is an additional yield curve shape, namely the dipped one for $a \neq b$. Note that we do not claim the list of the attainable yield curves in Theorem 2.13 to be definitive.

To analyze the behavior of the yield curve shapes at time t for $a \neq b$, we consider the $10\ \%$, the $50\ \%$, and the $90\ \%$ quantile of $r(t)$. Since $r(t)$ given r_0 is normally distributed, we can calculate those quantities explicitly. However, due to the two random sources $X(t)$ and $Y(t)$, we obtain different yield curve shapes for those quantities depending on the relation of $X(t)$ and $Y(t)$. To illustrate this, we have chosen to parameterize the set of curves by choosing ten quantile values of $Y(t)$ starting with the $5\ \%$ quantile and then increasing the quantile level in $10\ \%$ steps. The value of $\psi(t)$ is already fixed and the $X(t)$ values are then determined by $r(t)$, $\psi(t)$, and $Y(t)$.

(a) 10 % quantile of $r(t)$

(b) 50 % quantile of $r(t)$



(c) 90 % quantile of $r(t)$

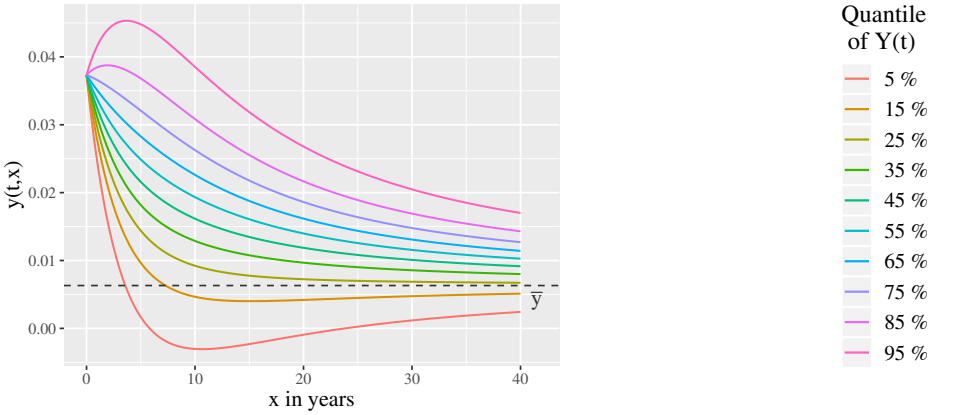


Figure 2.7: Yield curves generated by different quantiles of $r(t)$ and $Y(t)$ at time $t = 5$. Parameters: $r_0 = 0.001$, $a = 0.401$, $b = 0.178$, $\sigma = 0.0378$, $\eta = 0.0372$, $\theta = 0.01297$, $\rho = -0.996$

For the choice of parameters $r_0 = 0.001$, $a = 0.401$, $b = 0.178$, $\sigma = 0.0378$, $\eta = 0.0372$, $\theta = 0.01297$, and $\rho = -0.996$, the yield curves are illustrated in Figure 2.7 for $r(5)$ being equal to the 10 %, the 50 %, and the 90 % quantile. The first fact that highlights the difference of those figures to the one-factor Vasicek model is that there are more possible yield curves belonging to one value of the short rate. In Figure 2.7a the short rate lies below the long-term yield $\bar{y}$. Inverse yield curves do not occur in line with Theorem 2.13. However, dipped, normal, and humped yield curves exist. In Figure 2.7b and 2.7c the specific quantile of the short rate is larger than the long-term yield $\bar{y}$. In this case, dipped, inverse, and humped yield curves are generated. Normal ones cannot exist. Note also that the humped yield curves are above the inverse one which contrasts with the one-factor Vasicek model.

2.4 Measure Change and Yield Curve Shapes

To classify pension products according to their chance and risk characteristics, the use of Monte Carlo simulation methods to determine the distribution of their contract values at the end of the accumulation phase is the natural choice. This is mainly due to the complicated cost structures of the products and the way that profits are generated, assigned, and shared between the policy holders and the insurer. As life insurers typically invest in long-running interest rate products, the evolution of the yield curve over (calendar) time is a crucial ingredient for the simulation of pension products.

2.4.1 Pricing Measure, Physical Measure, and their Use in Simulation

The question about which measure to use when simulation methods are applied for chance and risk judgment is often not well understood in practical applications. However, it is clear that

- whenever there is the task to *price a financial product* in the future, a *risk-neutral measure* $\mathbb{Q}$ has to be used if the product depends only on the capital market,
- the evolution of the *stochastic input parameters* for pricing a financial product (such as the stock price for an option on that stock or the short rate determining the bond price) has to be simulated under a *physical measure* $\mathbb{P}$.

The reason for the first claim is arbitrage arguments for pricing a financial product. For the second claim, note that the dynamic evolution of the capital market moves under the physical measure, the so-called *real-world measure*.

As yield curves are derived from bond prices, given the choice of a pricing measure $\mathbb{Q}$ all possible yield curves in the Vasicek model are fully determined. For this, note also that the Vasicek model is time-homogeneous due to the constant coefficients. The actual shape of the yield curve at time t only depends on the value of $r(t)$ in the one-factor Vasicek model and on the values of $r(t)$, $X(t)$, and $Y(t)$ in the two-factor Vasicek model. However, the short rate evolves over time under the physical measure $\mathbb{P}$. Thus, while the possible yield curves are fully determined by $\mathbb{Q}$, the physical measure $\mathbb{P}$ determines the probability of their occurrence in the future.

Below we will concentrate on the one-factor Vasicek model as we have explicit forms of the distribution for the occurrence of the possible yield curve shapes. A similar analysis of the two-factor model is an aspect of future research.

2.4.2 Evolution of the Short Rate and Yield Curves under the Real-World Measure in the One-Factor Vasicek Model

In Section 2.3.1 we consider the distribution of future yield curve shapes under the risk-neutral measure $\mathbb{Q}$. However, as pointed out above, the probability of which

future yield curves actually occur depends on the distribution of the simulated short rate under a real-world $\mathbb{P}$.

A natural requirement on a real-world measure $\mathbb{P}$ is that the resulting short rate model stays in the Vasicek model class. To satisfy this requirement, we are constructing the dynamics of the short rate under $\mathbb{P}$ via a suitable change of measure by a Girsanov transformation. For this, we can change both the mean reversion level θ and the mean reversion speed a by adding suitable terms and obtain the stochastic differential equation of the short rate under $\mathbb{P}$ as

$$\begin{aligned} \mathrm{d}r(t) &= (a + \sigma d_a)(\theta + \sigma d_\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}\widetilde{W}(t) \\ &:= \widetilde{a}(\widetilde{\theta} - r(t))\mathrm{d}t + \sigma \mathrm{d}\widetilde{W}(t), \\ r(0) &= r_0. \end{aligned} \tag{2.13}$$

Here, $\widetilde{\theta}$ is a real valued constant, $\widetilde{a}$ is a positive constant, and $\widetilde{W}(t)$ is a one dimensional Brownian motion under $\mathbb{P}$ obtained via

$$\widetilde{W}(t) = W(t) + d_a \int_0^t r(s)\,\mathrm{d}s - (a d_\theta + \widetilde{\theta} d_a)t.$$

As required, Equation (2.13) defines a one-factor Vasicek short rate process under $\mathbb{P}$. It is still normally distributed with

$$r(t) \sim \mathcal{N}\left(r_0 e^{-\widetilde{a}t} + \widetilde{\theta}(1 - e^{-\widetilde{a}t}), \frac{\sigma^2}{2\widetilde{a}}(1 - e^{-2\widetilde{a}t})\right)$$

and an asymptotic behavior of

$$r(t) \sim \mathcal{N}\left(\widetilde{\theta}, \frac{\sigma^2}{2\widetilde{a}}\right).$$

Compared to the short rate under $\mathbb{Q}$, the additional drift d_θ shifts the mean of the short rate and d_a influences the variance of the short rate. More precisely, d_θ determines the asymptotic center of the distribution of the short rate, while d_a determines the intensity to be drawn back to this center. The zero-coupon bond price and the corresponding yield are obtained as in Equation (2.3) and (2.4). Only the input to those equations in form of the short rate $r(t)$ is obtained under the real-world measure $\mathbb{P}$. The bounds of the short rate for the different yield curve shapes to appear are the same as in Theorem 2.6.

Figure 2.8 illustrates the effect of the measure change of the short rate. First, we will explain the two parts of the figure separately. On the right hand side, different yield curves shapes depending on the level of the short rate $r(t)$ as well as the bounds b_{norm} and b_{inv} under the risk-neutral measure $\mathbb{Q}$ are depicted. This right side of the diagram is equivalent to Figure 2.5a. Here, one finds all possible yield curves since a yield curve always corresponds to bond prices, i.e. to the risk-neutral measure $\mathbb{Q}$. Also, the bounds are not affected by the change of measure. Since the actual shape

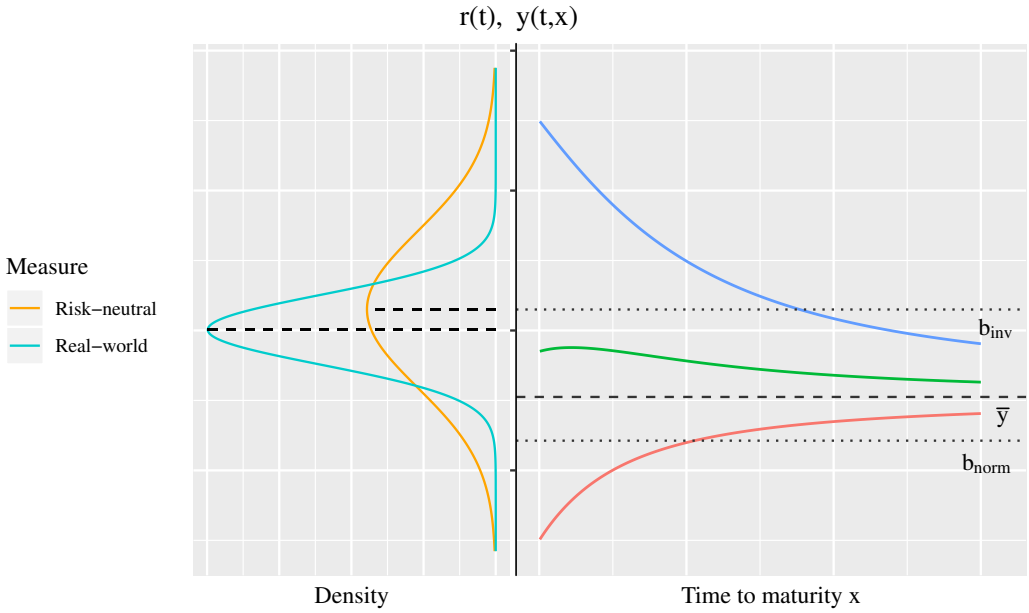


Figure 2.8: Effects on the frequency distribution of the different yield curve shapes displayed via the change of the density of the short rate caused by a change from the risk-neutral measure $\mathbb{Q}$ to the real-world measure $\mathbb{P}$

of the yield curve at a future time t only depends on the value $r(t)$, the physical measure $\mathbb{P}$ enters the left hand side of the figure. There, the asymptotic densities of the short rate under the risk-neutral measure $\mathbb{Q}$ as well as under a real-world measure $\mathbb{P}$ are shown. These densities are (asymptotically) responsible for the distribution of the short rate and thus also for the frequency of the occurrence of the different yield curve shapes. Consequently, while $\mathbb{Q}$ determines the yield curve shapes, $\mathbb{P}$ determines the frequency of their occurrence at future times. In Figure 2.8 the mean of the short rate under $\mathbb{P}$ is below the bound b_{inv}. As a result, we expect to observe inverse yield curves in the far future with a probability of less than $50\ \%$. This is not the case if we use the risk-neutral measure $\mathbb{Q}$ since then the asymptotic mean of the short rate equals b_{inv}.

As in the case of the risk-neutral measure (see Section 2.3.1), the probabilities for the three yield curve shapes depend on the distribution of the short rate. However, the distribution of the short rate under $\mathbb{P}$ is considered. The asymptotic behavior of the proportions of each single yield curve shape under $\mathbb{P}$ is given by

$$
\begin{aligned}
q_{norm} &\xrightarrow{t\to\infty} \Phi\left(-\left(\frac{3\sigma}{4a^2}+d_\theta\right)\sqrt{2\widetilde{a}}\right),\\
q_{hump} &\xrightarrow{t\to\infty} \Phi\left(d_\theta\sqrt{2\widetilde{a}}\right)-\Phi\left(-\left(\frac{3\sigma}{4a^2}+d_\theta\right)\sqrt{2\widetilde{a}}\right),\\
q_{inv} &\xrightarrow{t\to\infty} \Phi\left(d_\theta\sqrt{2\widetilde{a}}\right).
\end{aligned}
$$

Hence, an asymptotic proportion of inverse yield curves under $50\ \%$ can only be obtained for a negative d_θ. A positive d_θ increases the proportion of inverse and humped yield curves. The influence of d_a on the proportions of inverse and normal yield curves depends on d_θ. For $d_\theta > 0$ the proportion of inverse yield curves increases with d_a, while the proportion of normal yield curves decreases. For $d_\theta < 0$ inverse yield curves are less likely with higher d_a. The proportion of normal yield curves decreases with d_a for $-\frac{3\sigma}{4a^2} < d_\theta < 0$ and increases for $d_\theta < -\frac{3\sigma}{4a^2}$.

2.4.3 The Choice of the Real-World Measure $\mathbb{P}$

The choice of the real-world measure $\mathbb{P}$ is equivalent to the choice of the parameters $\widetilde{\theta}$ and $\widetilde{a}$ in Equation (2.13). As these parameters correspond to a subjective believe of future behavior of the interest market, their choice will always be up for debate. Of course, an additional positive drift parameter $\widetilde{\theta}$ would lead to a higher short rate. However, it would also lead to a higher proportion of future inverse yield curves. To obtain requirements on the parameters $\widetilde{\theta}$ and $\widetilde{a}$, one can on one hand use economic forecasts such as the annual ones given by the Organisation for Economic Co-operation and Development (OECD) or historical estimations on either the drift of the short rate or on the historical distribution of the yield curves. Below, we suggest a mixture of using forecasts for the mean of the short rate at some future time $\bar{t}$ and a condition on the probability to observe normal yield curves at a possible different future time $\bar{s}$.

The corresponding two possible requirements are:

(i) In $\bar{t}$ years a short rate of $\bar{r}$ is expected.

(ii) In $\bar{s}$ years a probability of at least $\bar{p} \cdot 100\ \%$ to observe normal yield curves is required.

These requirements are translated into the following equation and inequality

$$r_0 e^{-\widetilde{a}\bar{t}} + \widetilde{\theta}\left(1 - e^{-\widetilde{a}\bar{t}}\right) = \bar{r}, \tag{2.14}$$

$$r_0 e^{-\widetilde{a}\bar{s}} + \widetilde{\theta}\left(1 - e^{-\widetilde{a}\bar{s}}\right) + \frac{\sigma}{\sqrt{2\widetilde{a}}}\sqrt{1 - e^{-2\widetilde{a}\bar{s}}}Q_{\bar{p}} \leqslant b_{norm} \tag{2.15}$$

where $Q_{\bar{p}}$ is the $\bar{p}$ quantile of a standard normally distributed random variable. Of course, the inequality needs to be dealt with in a careful way: Either it is not possible to obtain $\bar{p} \cdot 100\ \%$ or more normal curves at time $\bar{s}$ (given that the requirement on the mean of the short rate at time $\bar{t}$ is already satisfied), or the inequality can even be solved as an equality, or it might only be possible to fulfill the relation as a strict inequality. We highlight that in our examples below.

We start to solve Equation (2.14) to obtain $\widetilde{\theta}$ as

$$\widetilde{\theta} = \frac{\bar{r} - r_0 e^{-\widetilde{a}\bar{t}}}{1 - e^{-\widetilde{a}\bar{t}}}. \tag{2.16}$$

Using this form of $\widetilde{\theta}$ in Equation (2.15) yields the inequality for $\widetilde{a}$ as

$$\theta - \frac{3\sigma^2}{4a^2} - r_0 e^{-\widetilde{a}\overline{s}} - \left(\overline{r} - r_0 e^{-\widetilde{a}\overline{t}}\right) \frac{1 - e^{-\widetilde{a}\overline{s}}}{1 - e^{-\widetilde{a}\overline{t}}} - \frac{\sigma}{\sqrt{2\widetilde{a}}} \sqrt{1 - e^{-2\widetilde{a}\overline{s}}} Q_{\overline{p}} \geqslant 0. \tag{2.17}$$

Depending on the parameters $\overline{r}$, $\overline{t}$, $\overline{p}$, and $\overline{s}$, Inequality (2.17) does not always have a solution. We first consider the extreme values for $\widetilde{a}$. Therefore, let the left side of Inequality (2.17) be denoted by $f(\widetilde{a})$. Its limits are

$$\lim_{\widetilde{a} \downarrow 0} f(\widetilde{a}) = \theta - \frac{3\sigma^2}{4a^2} - r_0 - (\overline{r} - r_0)\frac{\overline{s}}{\overline{t}} - \sigma\sqrt{\overline{s}} Q_{\overline{p}} =: f(0),$$

$$\lim_{\widetilde{a} \uparrow \infty} f(\widetilde{a}) = \theta - \frac{3\sigma^2}{4a^2} - \overline{r} = b_{norm} - \overline{r} =: f(\infty).$$

Obviously, Inequality (2.17) always has a solution $\widetilde{a}^*$ for $\overline{r} < b_{norm}$ as then for large values of $\widetilde{a}$ we have $f(\widetilde{a}) \geqslant 0$ due to $f(\infty) > 0$ and continuity of f. However, this might only be true as a strict inequality.

Next, let us look at small values of $\widetilde{a}$. $f(0)$ is non-negative if we have

$$\overline{r} \leqslant \left(\theta - \frac{3\sigma^2}{4a^2} - r_0 - \sigma\sqrt{\overline{s}} Q_{\overline{p}}\right) \frac{\overline{t}}{\overline{s}} + r_0 =: b_{f(0)}.$$

Whether $b_{f(0)}$ is larger than b_{norm} or vice versa depends on the relation of r_0 to $\overline{t}$, $\overline{s}$, and $\overline{p}$. The following relation holds

$$\text{(i) } \overline{t} \leqslant \overline{s}\text{:} \quad b_{f(0)} \begin{cases} \leqslant b_{norm}, & r_0 \leqslant b_{norm} + \sigma Q_{\overline{p}} \sqrt{\overline{s}} \dfrac{\overline{t}}{\overline{s} - \overline{t}} \\ > b_{norm}, & r_0 > b_{norm} + \sigma Q_{\overline{p}} \sqrt{\overline{s}} \dfrac{\overline{t}}{\overline{s} - \overline{t}} \end{cases}$$

$$\text{(ii) } \overline{t} > \overline{s}\text{:} \quad b_{f(0)} \begin{cases} \leqslant b_{norm}, & r_0 \geqslant b_{norm} - \sigma Q_{\overline{p}} \sqrt{\overline{s}} \dfrac{\overline{t}}{\overline{t} - \overline{s}} \\ > b_{norm}, & r_0 < b_{norm} - \sigma Q_{\overline{p}} \sqrt{\overline{s}} \dfrac{\overline{t}}{\overline{t} - \overline{s}} \end{cases}$$

Combining the achieved relations leads to

$$\begin{aligned} \overline{r} \leqslant \min\left\{b_{f(0)}; b_{norm}\right\}: &\quad f(0) \geqslant 0, f(\infty) \geqslant 0, \\ \min\left\{b_{f(0)}; b_{norm}\right\} < \overline{r} \leqslant \max\left\{b_{f(0)}; b_{norm}\right\}: &\quad f(0) < 0, f(\infty) \geqslant 0 \text{ or} \\ &\quad f(0) \geqslant 0, f(\infty) < 0, \\ \overline{r} > \max\left\{b_{f(0)}; b_{norm}\right\}: &\quad f(0) < 0, f(\infty) < 0. \end{aligned}$$

If signs of $f(0)$ and $f(\infty)$ differ, continuity of f implies the existence of a $\widetilde{a}$ solving Inequality (2.17) even as an equality. In the two cases of the same sign for $f(0)$ and $f(\infty)$, positive signs imply the existence of solutions of Inequality (2.17) as strict inequalities. For negative signs we can only say that there might be no solution at all but cannot give a general statement on existence/non-existence of a solution. These results are summarized in:

Proposition 2.14. *Let the risk-neutral short rate process be given by Equation* (2.2) *and the requirements determining the change of measure by Equation* (2.14) *and* (2.15)*. Set*

$$b_{f(0)} := \left(\theta - \frac{3\sigma^2}{4a^2} - r_0 - \sigma\sqrt{\bar{s}}Q_{\bar{p}}\right)\frac{\bar{t}}{\bar{s}} + r_0 \quad \text{and} \quad b_{norm} := \theta - \frac{3\sigma^2}{4a^2}$$

where $Q_{\bar{p}}$ is the $\bar{p}$ quantile of a standard normally distributed random variable. Subsequently, the following holds:

- *If $\bar{r} \leqslant \min\left\{b_{f(0)}; b_{norm}\right\}$, Inequality* (2.17) *is solvable. In addition, $f(\widetilde{a})$ could be positive for all positive $\widetilde{a}$. Inequality* (2.17) *is only satisfied as a strict inequality in this case.*
- *If $\min\left\{b_{f(0)}; b_{norm}\right\} < \bar{r} \leqslant \max\left\{b_{f(0)}; b_{norm}\right\}$, Inequality* (2.17) *is solvable and also possesses a solution if it is regarded as an equality.*
- *If $\bar{r} > \max\left\{b_{f(0)}; b_{norm}\right\}$, the solvability of Inequality* (2.17) *cannot be determined from the behavior of f at zero and infinity.*

Remark 2.15. Due to the more complicated distribution of the yield curve types in the two-factor model, there is no direct analogy to the proposition. One can only give the heuristic advice that adding a negative risk premium to the short rate will in principle avoid the occurrence of inverse yield curves. However, there is definitely room for future research.

We illustrate the findings by some numerical examples. Let us start with an observation: The behavior of $\mathbb{E}_{\mathbb{P}}(r(t))$ under the obtained real-world measure $\mathbb{P}$ is determined by r_0 and $\bar{r}$. For $\bar{r} > r_0$ the mean increases with time and decreases for $\bar{r} < r_0$. Hence, $\widetilde{\theta}$ is above r_0 for $\bar{r} > r_0$ and below for $\bar{r} < r_0$.

In our examples the parameters of the real-world measure are calculated as above for different values of $\bar{r}$. The following parameters of the risk-neutral measure and requirements are used:

$$r_0 = 0.016,\ a = 0.4,\ \sigma = 0.005,\ \theta = 0.02,\ \bar{t} = 10,\ \bar{s} = 20,\ \bar{p} = 0.7.$$

Here, we have

$$b_{norm} = 0.0199 < 0.0121 = b_{f(0)}.$$

We first get $\widetilde{\theta}$ from Equation (2.16). Then, $\widetilde{a}$ is obtained by treating Inequality (2.17) as an equality and solving it. Figure 2.9 shows $\mathbb{E}_{\mathbb{P}}(r(t))$ and the $70\ \%$ quantile of the short rate under the different real-world measures $\mathbb{P}$ over time. The requirements are marked. In Figure 2.9a $\bar{r}$ is chosen larger than $b_{f(0)}$ and r_0 but less than b_{norm}. Equation (2.17) is also solvable. The mean of the short rate increases. In contrast, the $\mathbb{E}_{\mathbb{P}}(r(t))$ decreases in Figure 2.9b since $\bar{r}$ is chosen to be less than r_0. Furthermore, $\bar{r}$ lies between $b_{f(0)}$ and b_{norm}. Therefore, a solution of Equation (2.17) is obtained. Figure 2.9c is based on $\bar{r}$ being less than $b_{f(0)}$. In this case $f(\widetilde{a})$ is positive for all $\widetilde{a}$. $\widetilde{a}$ is

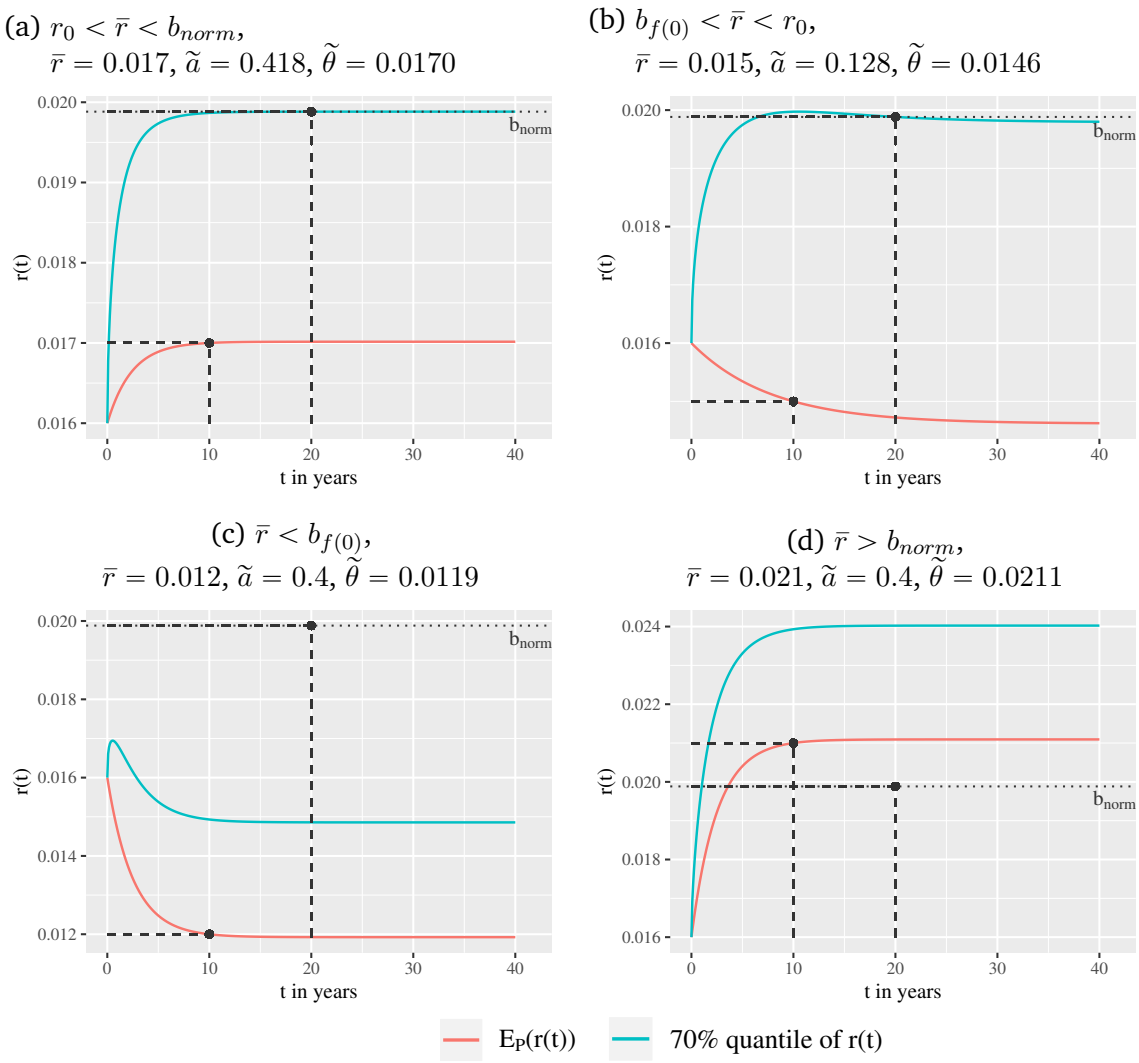


Figure 2.9: Mean and 70 % quantile of the short rate under a real-world measure $\mathbb{P}$ at different times t. Parameters: $r_0 = 0.016$, $\sigma = 0.005$, $\bar{t} = 10$, $\bar{p} = 0.7$, $\bar{s} = 20$

chosen equal to a of the risk-neutral measure as there is no need to change. The 70 % quantile of the short rate under $\mathbb{P}$ is notable under b_{norm}. Thus, we expect a fraction of future normal yield curves strictly above 70 % at time $\bar{s}$. In Figure 2.9d, $\bar{r}$ is chosen larger than b_{norm}. Here, $f(\tilde{a})$ is negative for all values of $\tilde{a}$. Thus, Equation (2.17) is not solvable. We chose $\tilde{a}$ as 0.4, the risk-neutral mean reversion speed. The desired 70 % quantile of the short rate as well as $\mathbb{E}_{\mathbb{P}}(r(\bar{t}))$ are well above b_{norm} for $\bar{s} = 20$. Therefore, the second requirement above can not be satisfied. This holds for all $\tilde{a}$.

2.5 Conclusion

In this chapter, we have dealt with two issues. Our first contribution highlights new findings for the Vasicek yield curves. In particular, we have shown significant advantages of using a two-factor Vasicek model instead of using a one-factor model such as

- the fact that the two-factor Vasicek model can overcome the complete dependence of the yield curve shape from the level of the short rate $r(t)$. This is particularly important when looking at the history of empirically observed yield curves in Europe where one has observed different yield curve shapes for the same value of the short rate at different time instants.
- the possibility to exhibit dipped yield curves, a fact which is particularly important in the current low interest rate setting.
- a better fit to market prices by introducing more parameters compared to a one-factor Vasicek model.

As a consequence, the actuary should add the two-factor Vasicek model and multi-factor models in general to his toolbox.

Our second main contribution concerns the question of the choice of a real-world measure $\mathbb{P}$ for simulating the evolution of the short rate over (calendar) time. While it is tempting to add a positive drift for the short rate and thus being able to show higher future returns, one has the disadvantage that a seemingly natural ordering between the returns of short-term and long-term fixed income investment gets lost. The price for the higher expected returns of – say – 10-year bonds is the fast increase of the probability to observe inverse yield curves in the future. It also contradicts the empirical observation that normal yield curves are observed much more frequently than inverse ones.

We have therefore equipped the actuary in Section 2.4 with a concept to handle this issue: A reasonably chosen measure transformation (i.e. a reasonably chosen drift and mean-reversion speed under $\mathbb{P}$) can preserve the natural ordering between short- and long-term investment in bonds.

3 Multi-Step Yield Curve Forecasting Using Machine Learning

The previous chapter focused on modeling the yield curve with the Vasicek interest rate model. Next, we consider the forecast of the yield curve. The evolution of the yield curve influences savings decisions of consumers, investment strategies of corporations, and policy decisions of governments. Furthermore, it is crucial for trading of financial products at the capital market and over-the-counter as well as bond portfolio and risk management. Thus a good forecast is important, particularly for the latter two.

Since computers have become more and more efficient over recent years techniques of machine learning are being used practically on an increasing scale. A lot of advancements have been made in this area and the theoretical understanding is improving also. In addition, more and more applications can be found in which existing machine learning algorithms have been successfully validated. In the field of yield curve forecasting, machine learning algorithms are not thoroughly studied in the literature as will be seen in the following section. Instead, the emphasis is on the extension and improvement of classical statistical methods. The studies using machine learning algorithms are partially not very detailed regarding the description and implementation of the hyperparameter optimization[1] and the validation of the models. Furthermore, the advantage of machine learning algorithms over classical statistical methods is not an object of thorough investigations.

In this chapter, we aim to compare linear models that have the advantage of low complexity and high interpretability with MLPs[2] (fully connected feed-forward neural networks) that are more complex and potentially more powerful in the context of yield curve forecasting. Our goal is to investigate whether it is beneficial to replace the commonly used techniques by MLPs that gained much popularity for tackling various regression tasks. Let us point out that we will not consider other popular machine learning techniques such as random forests since we want to focus on comparing very simple white box models with very complex black box algorithms. We will not be concerned with gray box models such as tree-based algorithms as well that lie somewhere in between. This may be the subject of further studies.

We evaluate these models based on the data of the European yield curve published by the ECB. For this, we establish an evaluation procedure and develop a new

[1] Hyperparameters are the parameters of the machine learning algorithms which are not learned during the training but have to be chosen prior (see also Definition 3.1 in Section 3.2.1). Hyperparameter optimization is the search for the best hyperparameter combinations.

[2] MLP: multilayer perceptron.

approach for hyperparameter optimization which allows us to search a larger hyperparameter space than standard methods. The latter can be transferred to other applications and machine learning algorithms. Since the performance strongly depends on the length of the forecast horizon, we consider several different ones resulting in different models that are appropriate.

The chapter is structured as follows. First, we give a short literature review of yield curve forecasting. In the following section, the underlying mathematical problem of time series forecasting is described. We introduce time series models, multi-step forecasting strategies, and selection procedures like cross validation and expanding windows. Then, we discuss the framework of our European yield curve forecasting. This includes the considered models, our approach of hyperparameter optimization which can also apply to other applications, the considered performance measures, and selection procedures. Subsequently, the data basis and results are described.

3.1 Literature Review

Diebold and Li (2006) are one of the first who pay attention to the practical problem of forecasting the yield curve, even though it plays an important role for bond portfolio, price and risk management. They interpolate the yield curve according to Nelson and Siegel (1987) and forecast the Nelson-Siegel parameters with a univariate first order autoregressive model. This is known as the dynamic Nelson-Siegel model. The short-term as well as long-term forecast quality of the model are evaluated out-of-sample. They compare the dynamic Nelson-Siegel model with different simple time series models. While the one-month-ahead forecasting of the yield curve using the Nelson-Siegel parameters produces no better results than those of simple time series models, the one-year-ahead results are significantly better.

Studies with different focus and extensions followed based on this groundwork. On the one hand, macroeconomic variables like the annual inflation rate are added to the models as additional exogenous variables (see Chen and Niu (2014) and De Pooter et al. (2010)). Thus, different states of the economy are considered. Other studies like Bernadell et al. (2005) take this aspect into account by using a regime-switching model. On the other hand, further models besides the one by Nelson and Siegel (1987) are used for estimating the yield curve. De Pooter (2007), De Rezende and Ferreira (2013), and Poklepović et al. (2014) replace Nelson and Siegel (1987) with the extension of Svensson (1994). De Pooter (2007) and De Rezende and Ferreira (2013) examine additional term structure models. Laurini and Hotta (2010) use a Bayesian estimation for forecasting the Nelson-Siegel and Nelson-Siegel-Svensson parameters. Christensen et al. (2011) introduce an arbitrage-free Nelson-Siegel approach, an extension of the dynamic Nelson-Siegel model to avoid arbitrage opportunities.

While a larger number of studies focus on forecasting the yield curve with classical statistic models, few investigations use machine learning techniques in this context. Täppinen (1998) compares neural networks with linear regression for forecasting the

change in yields with different times to maturity three months in advance. It shows that neural networks outperform the linear regression for yields with longer times to maturity. He uses one out-of-sample set. Zimmermann et al. (2002) also forecast the three and six month changes in the yields. They compare an error correction neural network to a time-delay recurrent neural network and a MLP. The error correction neural network outperforms the other neural networks. While Täppinen (1998) and Zimmermann et al. (2002) forecast the changes in the yields, Bose et al. (2006) and Poklepović et al. (2014) base the forecasting on the Nelson-Siegel parameters. Bose et al. (2006) apply a MLP and a generalized feed-forward network on the Nelson-Siegel parameters. Both architectures are fixed at one hidden layer with four neurons. They use one out-of-sample set evaluating and comparing both approaches. The generalized feed-forward network outperforms the MLP in their out-of-sample set. Poklepović et al. (2014) forecast the Nelson-Siegel parameters as well as the Nelson-Siegel-Svensson parameters using vector autoregression (VAR) and neural networks. The best result delivers the neural network based on the Nelson-Siegel parameters. In all four studies, the procedure of determining the network architecture and evaluating the models admit of improvement.

Sambasivan and Das (2017) use a Gaussian Process regression, a machine learning technique, in conjunction with a dynamic modeling strategy. Whenever new yield curve data is available the hyperparameters of the Gaussian Processes model are updated. This method is compared to a VAR of yields and the Nelson-Siegel parameters using a rolling window approach. Different short-, medium-, and long-term yields of the next day are forecast. The VARs produce better results for short-term yields, while the Gaussian Processes approach performs better in medium- and long-term yields. Therefore, they suggest a combination of both methods for forecasting the yield curve. Other forecasting horizons than one day are not considered in the paper.

Primarily, different methods of forecasting the yield curve are applied for the U.S. market. There are few investigations of other markets. Caldeira et al. (2016a) and Caldeira et al. (2016b) apply the different forecast methods on the Brazilian bond market. Caldeira et al. (2016a) investigate an approach combining different forecast models for different forecast horizons. In Caldeira et al. (2016b), different benchmark forecast models including the persistence model, VAR, dynamic Nelson-Siegel, and arbitrage-free dynamic Nelson-Siegel model are compared with regard to their forecasting performance. Bose et al. (2006) use different bonds of India as database. The German bond market is investigated in Zimmermann et al. (2002), while Poklepović et al. (2014) consider the Croatian financial market.

3.2 Setup and Objective

In this section, we formulate the underlying mathematical problem of time series forecasting in general and the approach as well as the idea of the considered time series models. While the one-time-step-ahead forecast is simple, there are different strategies used in the literature to tackle multi-step forecasts of time series. These

strategies are also introduced. Finally, we consider different selection procedures like cross validation and expanding windows to evaluate the models. This section is very general and presents the background of time series forecasting.

We consider N different variables that will be forecast. There is an observation of each variable at each time $t_k \in \mathbb{R}_0^+$, $k \in \{1, \dots, n\}$. The observations of the variables at time point t_k are denoted by $x_{t_k} \in \mathbb{R}^N$. The development of each variable in the course of time is depicted by n observations and are denoted by $x^{(i)} \in \mathbb{R}^n$, $i \in \{1, \dots, N\}$. $x_{t_k}^{(i)}$ refers to the observation of the i^{th} variable at time t_k.

The goal of a one-time-step-ahead forecast – which is the usual case – is to forecast the observation $x_{t_{k+1}}$ based on the current observation x_{t_k}. For this purpose, the functional relationship $f : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^N$ which maps the value of the time series at time t_k onto its value at time t_{k+1} has to be estimated and is represented by all observations x_{t_k}, $k \in \{1, \dots, n\}$, which are the true signal shifted by white noise. We assume that

$$x_{t_{k+1}} = \widehat{f}(x_{t_k}, t_k) + \epsilon_{t_k}$$

where $k \in \{1, \dots, n-1\}$, $\widehat{f} : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^N$ is the estimation of f, and $\epsilon_{t_k} \in \mathbb{R}^N$ is a vector of realizations of independent random variables with zero expectation and equal variance. We choose $\widehat{f}$ such that it minimizes

$$\left\| \begin{pmatrix} x_{t_2}^T \\ \vdots \\ x_{t_n}^T \end{pmatrix} - \begin{pmatrix} \widehat{f}(x_{t_1}, t_1)^T \\ \vdots \\ \widehat{f}(x_{t_{n-1}}, t_{n-1})^T \end{pmatrix} \right\|$$

with respect to a suitable norm $\|\cdot\|$ on $\mathbb{R}^{(n-1)\times N}$. The forecast value of $x_{t_{k+1}}$ is then

$$\widehat{x}_{t_{k+1}} = \widehat{f}(x_{t_k}, t_k).$$

The actual value of the time series can depend on more than the last value. The function f is then based not only on x_{t_k} but also $x_{t_{k-1}}$ and so on. Since we consider only the dependence of x_{t_k} on $x_{t_{k-1}}$, we restrict this section to this variant. However, it can be easily transferred to the case of a dependency to more previous values.

3.2.1 Forecasting Models

We aim at comparing linear models that have the advantage of low complexity and high interpretability with MLPs, fully connected feed-forward neural networks, that are potentially more capable in learning the underlying distribution of the data.

Persistence Model

The basic time series model is the persistence model. Here, the forecast for the next time step is simply the current value. Thus, the estimator $\widehat{f}$ is defined as the identity function

$$\widehat{f}\,(x_{t_k}, t_k) := x_{t_k}.$$

This trivial method is often called *naive forecast* as it does not take into account any underlying structure. For forecasts of time series without jumps or low volatility, the persistence model is often hard to beat if the forecast horizon is short. However, the persistence model quickly becomes inaccurate for long-term forecasts or time series that change significantly from one time step to the next.

Autoregression and Vector Autoregression

Besides the persistence model, we consider autoregressive and vector autoregressive models, classical time series forecast models. For literature, we refer to the corresponding chapters in Deistler and Scherrer (2018), Franke (2002), and Tsay (2010). Both autoregression (AR) and VAR are modeled by a linear function $\widehat{f}$.

In an autoregressive model, the value one-time-step-ahead of a variable only depends linearly on its own previous values and on a stochastic term. Since we consider a first order autoregressive model, denoted by AR(1), we have only a dependency on the current value of the variable. Further, we assume an additional linearity of the variable in time. AR(1) is defined by

$$\widehat{f}^{(i)}\left(x_{t_k}^{(i)}, t_k\right) := b^{(i)} + w^{(i)} \cdot x_{t_k}^{(i)} + w_t^{(i)} \cdot t_k \tag{3.1}$$

for $k \in \{1, \ldots, n\}$ and $i \in \{1, \ldots, N\}$ where $b^{(i)}$ is the bias and $w^{(i)}$ and $w_t^{(i)}$ are the coefficients of the linear model.

Since we want to forecast N variables that are described by an AR(1) process, each $f^{(i)}(x_{t_k}^{(i)}, t_k)$ is estimated separately by an autoregressive model $\widehat{f}^{(i)}(x_{t_k}^{(i)}, t_k)$. The biases $b^{(i)}$ and coefficients $w^{(i)}$ and $w_t^{(i)}$, $i \in \{1, \ldots, N\}$, of the collection of variables can be represented by a vector and a diagonal matrix, respectively. Hence, (3.1) can be reformulated in matrix notation as follows:

$$\widehat{f}\left(x_{t_k}, t_k\right) = b + W \cdot \begin{pmatrix} x_{t_k} \\ t_k \end{pmatrix} \tag{3.2}$$

where

$$b := \begin{pmatrix} b^{(1)} \\ \vdots \\ b^{(N)} \end{pmatrix} \quad \text{and} \quad W := \begin{pmatrix} w^{(1)} & \cdots & 0 & w_t^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & w^{(N)} & w_t^{(N)} \end{pmatrix}$$

denote the bias and weight matrix of the model, respectively.

In contrast to the AR(1) model, a vector autoregressive model includes further dependencies. In a first order VAR, denoted by VAR(1), $x_{t_{k+1}}^{(i)}$ is explained by its own current value $x_{t_k}^{(i)}$ and the current value of the other model variables $x_{t_k}^{(l)}$, $l \in \{1, \ldots, N\} \backslash \{i\}$. Analogously to (3.1) we define the VAR(1) as follows:

$$\widehat{f}^{(i)}\left(x_{t_k}^{(i)}, t_k\right) := b^{(i)} + \sum_{j=1}^{N} w_j^{(i)} \cdot x_{t_k}^{(j)} + w_t^{(i)} \cdot t_k$$

where $w_j^{(i)}$ is the coefficient of the variable $x^{(j)}$ in the estimation of $x^{(i)}$. The matrix notation of the VAR(1) model satisfies (3.2) with bias and weight matrix given by

$$b := \begin{pmatrix} b^{(1)} \\ \vdots \\ b^{(N)} \end{pmatrix} \quad \text{and} \quad W := \begin{pmatrix} w_1^{(1)} & \cdots & w_N^{(1)} & w_t^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ w_1^{(N)} & \cdots & w_N^{(N)} & w_t^{(N)} \end{pmatrix}.$$

Obviously, a VAR(1) is a generalization of an autoregressive model.

The biases and coefficients of autoregressive and vector autoregressive models are estimated by least squares. Here, the sum of the squared residuals occurring in the results of every single estimation is minimized. This has the effect that the biases and coefficients of these models can be obtained by a closed formula if the data matrix has full rank. Furthermore, these parameters and thus the models are interpretable which is a huge advantage in contrast to black box machine learning algorithms such as neural networks. A disadvantage is that only linear interrelations can be mapped by these models. If the variable can not be described by a linear process, ARs and VARs forecast evolutions poorly. However, these models are suitable as benchmarks to evaluate the model performance of more complex models due to their simplicity and interpretability.

Multilayer Perceptron

The above models are the benchmark for judging the suitability of MLPs. MLPs are a class of very powerful but not easily usable, fully connected feed-forward neural networks that constitute the foundational part of the deep learning framework. In our setting we define MLPs based on Wiese et al. (2020) as follows:

Let $\phi : \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, monotonic function that satisfies $\phi(0) = 0$. In addition, let $L, N_0, \ldots, N_{L+1} \in \mathbb{N}$, Θ an Euclidean vector space and let $a_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ be an affine mapping for any $l \in \{1, \ldots, L+1\}$. A function $\widehat{f}_\theta : \mathbb{R}^{N_0} \times \Theta \to \mathbb{R}^{N_{L+1}}$, defined by

$$\widehat{f}_\theta(x_{t_k}, t_k) := a_{L+1} \circ g_{\theta,L} \circ \cdots \circ g_{\theta,1}(x_{t_k}, t_k)$$

for $k \in \{1, \ldots, n\}$ where

$$g_{\theta,l} := \phi \circ a_l \quad \text{for all} \quad l \in \{1, \ldots, L\}$$

and ϕ being applied component-wise, is called a *multilayer perceptron with L hidden layers*. In this setting N_0 represents the *input dimension*, N_{L+1} the *output dimension*, and $N_1, \ldots, N_L$ the *hidden dimensions*. Furthermore, for any $l \in \{1, \ldots, L+1\}$ the function a_l is defined by

$$a_l(x) := b_l + W_l x$$

for every $x \in \mathbb{R}^{N_{l-1}}$ where, analogously to (3.2), $b_l \in \mathbb{R}^{N_l}$ is the bias and $W_l \in \mathbb{R}^{N_l \times N_{l-1}}$ the weight matrix of the model. With this representation, the MLP's parameters are defined by

$$\theta := (W_1, \ldots, W_{L+1}, b_1, \ldots, b_{L+1}) \in \Theta.$$

The function ϕ is referred to as *activation function*. Furthermore, note that MLPs have a similar underlying structure as the linear models considered above. Indeed, the main differences in MLPs are the non-linearity caused by the activation function ϕ as well as the increased complexity induced by the hidden layers. The parameters are estimated by minimizing a loss function. For the adjustment of the bias and weights, backpropagation is used which computes the gradient of the loss function with respect to each bias and weight.

Figure 3.1 illustrates one MLP with an input layer of 8 nodes, two hidden layers where the first one has 11 nodes and the second one 13, and an output layer of 7 nodes. The links of each node to every other node of the following layer can clearly be seen.

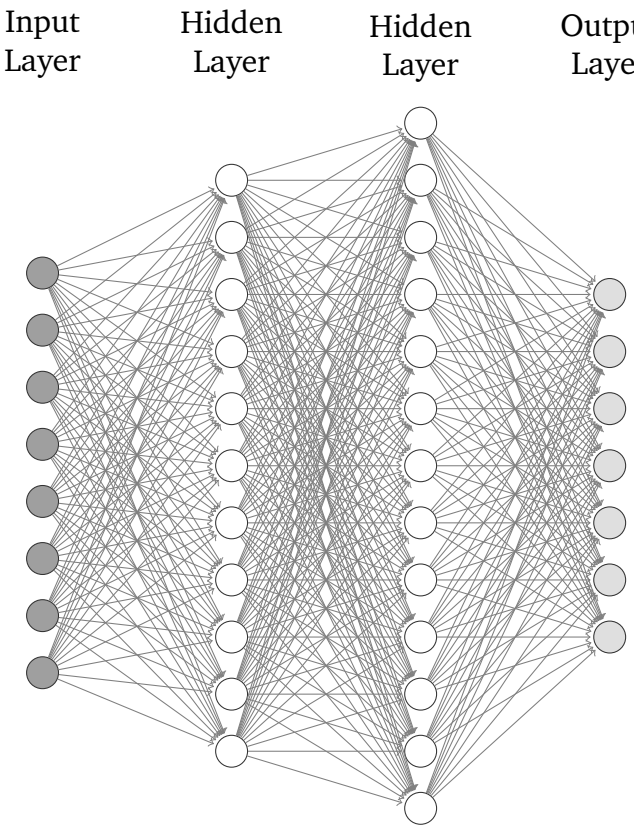


Figure 3.1: 2-layer MLP with input dimension $N_0 = 8$, output dimension $N_3 = 7$, and hidden dimensions $N_1 = 11$ and $N_2 = 13$

The usage of MLPs is justified since theoretically MLPs are able to approximate the underlying data structure arbitrarily well. The so-called *universal approximation theorem* of one-layer perceptrons states that a feed-forward neural network with a network architecture of a single hidden layer and a finite number of neurons can approximate any continuous functions on compact sets, see for instance Hornik (1991) or Buehler et al. (2019). Note that only one hidden layer is needed, but the number of neurons may be extremely large and potentially unfeasible to train the model. The

universal approximation theorem can be easily transferred to MLPs with an output dimension larger than one and more than one hidden layer which makes the training of the model easier. This theorem is one of the main reasons for such models being extremely popular nowadays.

Before we can calibrate the parameters of an MLP (the *training* process), we have to decide on some external choices such as the activation function or the number of hidden layers. Such parameters are examples of so-called hyperparameters that play an important role in training the model.

Definition 3.1 (Hyperparameters)**.** The parameters of the model that are chosen prior to the training process and are not learned during the training are called *hyperparameters*. The main hyperparameters of MLPs are

- Network topology: The number L of hidden layers and the numbers $N_1, \ldots, N_L$ of neurons per hidden layer,
- Activation function ϕ,
- Parameters of the optimization via backpropagation: e.g. solver, learning rate, stopping criterion.

Since those hyperparameters have a huge influence on the accuracy of the forecast obtained by MLPs, we aim to tune these hyperparameters to a quasi-optimal setting. This task is rather challenging and time-consuming. Therefore, in Section 3.3.2 we develop an approach how this task may be tackled in an automatized manner based on the standard hyperparameter optimization like Grid and Random Search.

3.2.2 Forecasting Strategies

The one-time-step-ahead forecast is clear and based on the last known values. However, many applications in practice require a forecast of a longer period than one time step resulting in a multi-step time series forecasting task. In this section, we introduce different strategies implementing a multi-step forecast which can be found in Chevillon (2007), Sorjamaa et al. (2007), and Taieb et al. (2012). These can be used for all models presented in Section 3.2.1.

A straightforward approach is to follow the idea of the step by step forecast but using the forecast values of $h-1$ previous time steps to forecast the h^{th} time step where h is the forecast horizon and $h > 0$. This leads to the recursive strategy defined as follows.

Definition 3.2 (Recursive strategy)**.** The *recursive strategy* involves using a one-time-step model multiple times where the forecast for the prior time step is used as an input for making a forecast on the following time step. The value $x_{t_{k+h}}$ is forecast recursively by

$$\widehat{x}_{t_{k+h}} = \widehat{f}_h(x_{t_k}, t_k) = \widehat{f}_1(\widehat{f}_{h-1}(x_{t_k}, t_k), t_{k+h-1}) = \underbrace{\widehat{f}_1 \circ \cdots \circ \widehat{f}_1}_{h \text{ times}}(x_{t_k}, t_k)$$

where $\widehat{f}_1 : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^N$ is the estimated one-time-step-ahead forecast model.

This strategy is also called *iterated multi-step* (see Chevillon (2007)). Although constituting a rather straightforward approach to the problem at hand, one quickly realizes the potential problems. Each forecast comes with an error and by using the forecast as input we amplify the error and accumulate all the errors along the way. An alternative approach avoiding the accumulation of errors is the multiple output strategy that directly forecasts all relevant time steps simultaneously, consequently using the same model for each time step.

Definition 3.3 (Multiple output strategy)**.** The *multiple output strategy* involves developing one model that is capable of forecasting the entire forecast sequence in a one-shot manner. Let $\mathscr{H}$ be the set of considered forecast horizons and set $H := \sharp(\mathscr{H})$. Furthermore, we define

$$x_{t_{k+\mathscr{H}}} = \begin{pmatrix} x_{t_{k+h_1}} \\ \vdots \\ x_{t_{k+h_H}} \end{pmatrix} \in \mathbb{R}^{N \cdot H}$$

for $k \in \{1, ..., \widetilde{n}\}$ with $\widetilde{n} = n - \max\{\mathscr{H}\}$. The relation between $x_{t_{k+\mathscr{H}}}$ and x_{t_k} is estimated by $\widehat{f} : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^{N \cdot H}$. Then, the forecast $\widehat{x}_{t_{k+\mathscr{H}}}$ of all forecast horizons $h \in \mathscr{H}$ is given by

$$\widehat{x}_{t_{k+\mathscr{H}}} = \widehat{f}(x_{t_k}, t_k)$$

where $\widehat{f}$ is an estimator for $f : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^{N \cdot H}$.

Here, the errors are not accumulated. Further, dependencies between different forecast horizons are considered. However, this strategy has the disadvantage that the same model is used for short- and long-term forecasts. It can be assumed that the long-term forecast has a different functional relation than the short-term.

A third method is the direct multi-step forecast strategy.

Definition 3.4 (Direct multi-step forecast strategy)**.** The *direct multi-step forecast strategy* involves developing a separate model for each forecast time step. The forecast value is computed by

$$\widehat{x}_{t_{k+h}} = \widehat{f}_h(x_{t_k}, t_k)$$

where $\widehat{f}_h : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^N$ an estimator for $f_h : \mathbb{R}^N \times \mathbb{R}_0^+ \to \mathbb{R}^N$ which maps the relation between $x_{t_{k+h}}$ and x_{t_k} for each forecast horizon $h \in \mathscr{H}$ and $k \in \{1, \ldots, n-h\}$.

Since separate models are used for different forecast horizons, dependencies between different time-step-ahead forecasts cannot be modeled. However, this method also prevents the accumulation of errors similar to the multiple output strategy. Moreover, this approach allows the usage of good short-term models and other models suitable for long-term forecasts availing the different strengths of various models. Therefore, we employ the direct multi-step forecast strategy and aim at finding appropriate models for the time steps under consideration.

For completeness let us mention that one can combine the direct multi-step and recursive strategies in order to benefit from the respective strengths of the two methods. This approach is referred to as *direct-recursive hybrid strategy*.

3.2.3 Selection Procedures

We want to find the best model for representing the structure of the data at hand. Therefore, the models have to be evaluated. For this, the data is split into a training and validation set. The models are estimated on the training set which means the parameters of the model, the bias and coefficients, are determined. Based on the resulting bias and coefficients, the performance measures are calculated on the validation set for each model. We refer to Section 3.3.3 for consideration of different performance measures and their calculation. The best model is the model with the best performance measure on the validation set. This can be either the smallest or the largest depending on the type. In the following, we assume that the smaller the performance measure, the better the model. Using only one training and validation set bears the risk of depending extensively on a chosen time period. Therefore, different approaches are developed to avoid this risk. Two approaches are introduced in the following: cross validation and expanding window approach.

Cross Validation

The cross validation approach is a classical procedure in machine learning that we adopt to the time series setting. It can be used if the time series is stationary[3]. During cross validation the data is partitioned into $K << n$ disjoint, approximately equally sized folds. We control this by dividing the time points t_k, $k \in \{1, \dots, n\}$, into K complementary sets $\mathcal{T}_l$, $l \in \{1, \dots, K\}$, so that

$$\bigcup_{l=1}^{K} \mathcal{T}_l = \{t_1, \dots, t_n\} \text{ and } (\mathcal{T}_i \cap \mathcal{T}_j = \varnothing) \wedge (|\sharp(\mathcal{T}_i) - \sharp(\mathcal{T}_j)| \leqslant 1) \ \forall\, i \neq j.$$

All observations x_{t_k} with $t_k \in \mathcal{T}_l$ build the l^{th} fold. Each fold will be used as the validation set, while the remaining $K-1$ folds constitute the training set. Note that the folds can consist of contiguous as well as non-contiguous data points. In the following, however, we consider the case of a temporally contiguous data block as a validation set.

Figure 3.2 illustrates a 4-fold cross validation ($K = 4$). In Figure 3.2a the validation set is contiguous, while in Figure 3.2b is non-contiguous. Each circle represents one data point x_{t_k}. If the circle is colored gray, the data point belongs to the training set and if it is black, to the validation set.

[3] A time series is weakly stationary if the mean of the time series is constant for all time points and the covariance between the values at any two time points depends only on the difference between the two time points and not on the location of the points along the time axis.

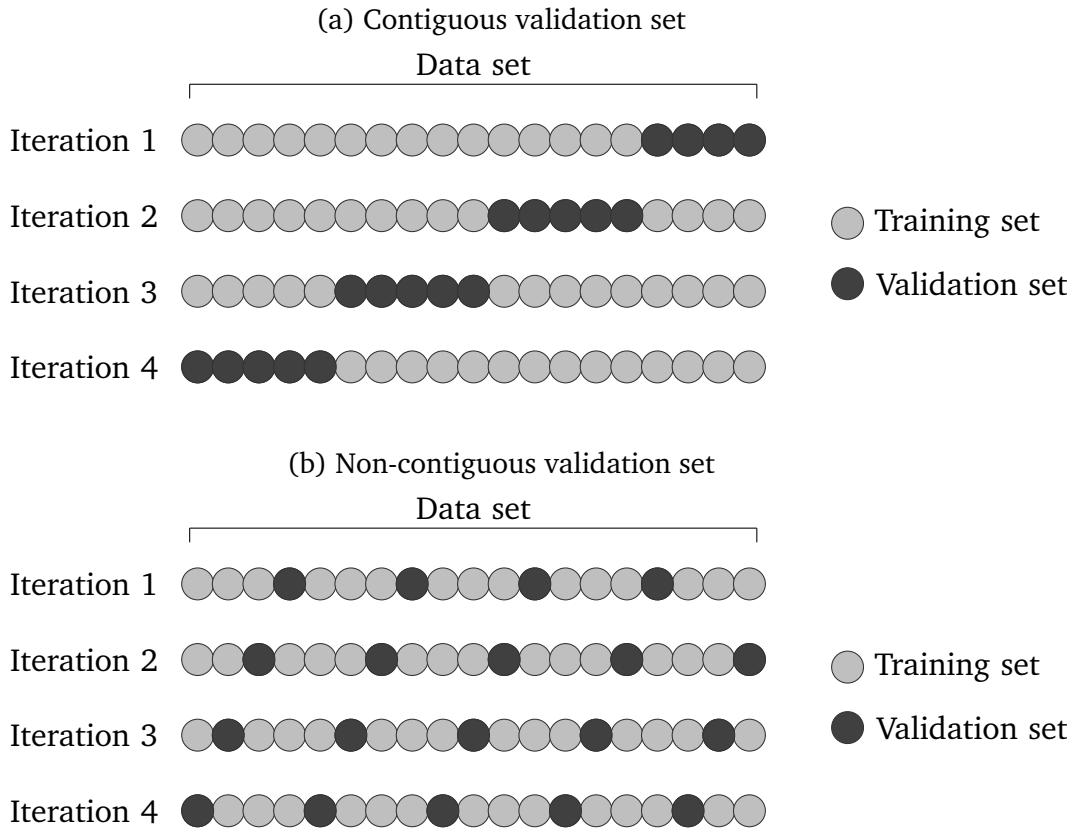


Figure 3.2: 4-fold cross validation

For each model $m \in \mathscr{M}$ where $\mathscr{M}$ is the set of models under consideration the splitting results in K scores μ_l^m, $l \in \{1, \ldots, K\}$, of the chosen performance measure μ, respectively computed on the l^{th} validation set. The validation score for model m is defined as

$$\overline{\mu}^m := \frac{1}{K} \sum_{l=1}^{K} \mu_l^m.$$

The model $m_\mu^{cv} := \underset{m \in \mathscr{M}}{\arg\min}\ \overline{\mu}^m$ with the lowest validation score is selected by this procedure as the best model with respect to the performance measure μ.

Expanding Windows

The second procedure for model selection is an expanding window approach. For this purpose, let λ, $\nu \in \mathbb{N}$ and set

$$K := \left\lfloor \frac{n - \lambda}{\nu} \right\rfloor.$$

Our approach is as follows: We start our iterative procedure with the training set

$$x_1^{tr} := \begin{pmatrix} x_{t_1} & \cdots & x_{t_\lambda} \end{pmatrix}^T \in \mathbb{R}^{\lambda \times N}$$

of the time series' first λ data points and the validation set

$$x_1^{val} := \begin{pmatrix} x_{t_{\lambda+1}} & \cdots & x_{t_{\lambda+\nu}} \end{pmatrix}^T \in \mathbb{R}^{\nu \times N}$$

consisting of the time series' subsequent ν data points. Thus, we train every model $m \in \mathscr{M}$ on x_1^{tr} and use it for forecast on x_1^{val}. For this training and test set, we obtain the score μ_1^m. The training set is then expanded by ν observations and we obtain

$$x_2^{tr} := \begin{pmatrix} x_{t_1} & \cdots & x_{t_{\lambda+\nu}} \end{pmatrix}^T \in \mathbb{R}^{(\lambda+\nu) \times N}$$

as the new training set. The new validation set on which we compute the score μ_2^m is given by

$$x_2^{val} := \begin{pmatrix} x_{t_{\lambda+\nu+1}} & \cdots & x_{t_{\lambda+2\nu}} \end{pmatrix}^T \in \mathbb{R}^{\nu \times N}.$$

We repeat this procedure further $K-2$ times and in each iteration $l \in \{1, \ldots, K\}$ we obtain an error score μ_l^m for model $m \in \mathscr{M}$. Figure 3.3 illustrates the training and validation sets for the expanding window approach exemplary. Each circle represents one data point. The color indicates whether the data point belongs to the training or validation set. Gray points are members of the training set, black ones correspond to the validation set.

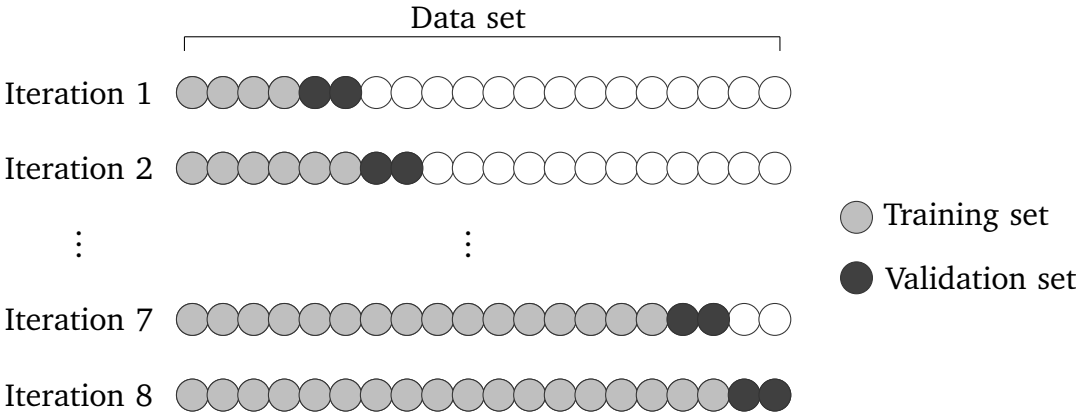


Figure 3.3: Expanding window approach with $\lambda = 4$ and $\nu = 2$

The validation score for model $m \in \mathscr{M}$ is defined as

$$\overline{\mu}^m := \frac{1}{K} \sum_{l=1}^{K} \mu_l^m.$$

The model $m_\mu^{ew} := \underset{m \in \mathscr{M}}{\arg\min}\ \overline{\mu}^m$ with the lowest validation score is selected by this procedure as the best model with respect to the chosen performance measure μ.

3.3 Approach

The purpose of this paper is to investigate whether deep learning has a benefit to forecasting the yield curve compared to classical models. More specifically, we aim at finding out which of the above-mentioned models is best suited for forecasting. For this, we use the European yield curve published by the ECB since it is a macroeconomic indicator of an economic area and is becoming increasingly important. Furthermore, the data are freely available. In this section, we describe our forecasting approach of the yield curve and the evaluation of the different models. This includes the specification of the considered models, the chosen performance measures, and the selection procedure. Additionally, we introduce our hyperparameter optimization approach. The description of the data and results of our analysis are summarized in Section 3.4.

The yield curve data consists of different observations $x_{t_k} \in \mathbb{R}^N$, $t_k \in \mathbb{R}_0^+$, where x_{t_k} contains the yields of N different times to maturity at the time point t_k. The considered times to maturity are the same for all t_k. Since the yield curve shows structural changes over time and is not stationary (see Section 3.4.1), we do not directly forecast the absolute yields of different times to maturity. Instead, we consider the increments of the absolute yields defined by

$$\triangle_h x_{t_k} = x_{t_k} - x_{t_{k-h}}$$

for the considered forecast horizons $h \in \mathscr{H}$. We use the same n time points t_k as input data for all forecast horizons h and estimate the relation between $\triangle_h x_{t_{k+h}}$ and $\triangle_h x_{t_k}$ for $k \in \{1, \dots, n\}$ by a function $\widehat{f}$. Since the level of the yield curve is not observable in the increments, we add the mean value of the considered yields $\overline{x}_{t_k}$ at time point t_k as an additional input variable supplementing the increments and the time point. The forecast increment is then

$$\widehat{\triangle}_h x_{t_{k+h}} = \widehat{f}_h(\triangle_h x_{t_k}, \overline{x}_{t_k}, t_k)$$

for any $k \in \{1, \dots, n\}$ and forecast horizon $h \in \mathscr{H}$ where $\widehat{f}_h : \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}_0^+ \to \mathbb{R}^N$. The function $\widehat{f}_h$ is estimated respectively by the different models mentioned above. The forecast for the yield curve $x_{t_{k+h}}$ is then given by

$$\widehat{x}_{t_{k+h}} = x_{t_k} + \widehat{\triangle}_h x_{t_{k+h}}.$$

Since it is not clear whether t_k as an additional feature has an impact on the estimation, we also consider a function independent of t_k for the relation between $\triangle_h x_{t_k}$ and $\triangle_h x_{t_{k+h}}$. Then, $\widehat{f}_h : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ is the estimation of $f_h : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ with

$$\widehat{\triangle}_h x_{t_{k+h}} = \widehat{f}_h(\triangle_h x_{t_k}, \overline{x}_{t_k}).$$

3.3.1 Models under Consideration

As a baseline, we consider a persistence model for yield curve forecasts, see Section 3.2.1. This is equivalent to forecasting an increment of zero. As linear models, we consider an AR(1) and VAR(1) model for the estimation of the increments for all forecast horizons. Thus, we assume that the current value only depends on its previous one. This is justified by the analysis of the autocorrelation and partial autocorrelation (see Section 3.4.1). The AR and the VAR with order 1 are the simplest models which also speaks for their usage since we want to compare models of low complexity with very complex one and see how strong the simplest models are. Furthermore, the order of the linear model coincides with the order chosen in the literature, see among others Diebold and Li (2006). For comparison, we consider MLPs with the same input data as the linear models. Consequently, the MLP has an input layer with dimension $N_0 = N+1$ and $N_0 = N+2$, respectively, and output dimension $N_{L+1} = N$ where N are the number of the forecast yields, see Section 3.4.1, and L is the number of hidden layers which is a result of the hyperparameter optimization.

3.3.2 Hyperparameter Optimization

In our application, the MLP is the only model we consider that has tunable hyperparameters. However, our hyperparameter optimization can be used for all machine learning algorithms with tunable hyperparameters and is introduced for such models in general.

Hyperparameter optimization is a crucial step in many advanced machine learning applications. One of the standard techniques in this regard is the Grid Search typically based on the cross validation method. It aims at finding the best performing hyperparameter combination from a given set $\mathscr{V}^*$, a set of hyperparameter combinations under consideration. For this, a cross validation is run for all hyperparameter combinations in $\mathscr{V}^*$ and the combination with the smallest performance measure is selected. Another method is Random Search. It differs from Grid Search in so far as that the performance measure is not calculated for every hyperparameter combination in $\mathscr{V}^*$ but for a random subset of $\mathscr{V}^*$ selected uniformly. The size of this random subset is prespecified, cf. Bergstra and Bengio (2012).

One obvious drawback of these methods, especially for Grid Search, is that the set $\mathscr{V}^*$ needs to be specified in advance. Often it is not known in which range the hyperparameters should be. Choosing $\mathscr{V}^*$ too small may result in not finding a suitable model, choosing the grid too large may not be feasible. Often the computing capacity and the available time determine the composition of $\mathscr{V}^*$. In order to specify $\mathscr{V}^*$ such that it allows for efficient optimization and yet yields an appropriate model for the task at hand, it is beneficial to choose $\mathscr{V}^*$ adjusted to the data.

In this section, we develop an automatic approach for hyperparameter optimization. Since our extension can apply to other applications, we introduce it in an abstract manner. It is based on the set of all hyperparameter combinations that may possibly be considered denoted by $\mathscr{V}$. It is not the set of all possible combinations

since this is in general unbounded. However, $\mathscr{V}$ is assumed to be very large and not feasible for Grid Search. Our approach aims at restricting $\mathscr{V}$ to a feasible subset $\mathscr{V}^*$ by selecting only those hyperparameters that are plausible good candidates for the best model.

Consequently, our approach for hyperparameter optimization consists of the following two steps:

(i) Preselection of hyperparameter combinations $\mathscr{V}^*$ from $\mathscr{V}$ based on multiple validation curves and a suitable selection procedure.

(ii) Finding the best hyperparameter combination in $\mathscr{V}^*$ via Grid Search or Random Search.

Step 1: Preselection of hyperparameters

In the first step, we find a suitable set of values for each hyperparameter $p \in \mathscr{P}$ where $\mathscr{P}$ is the set of the considered hyperparameters. For this, we choose an arbitrary hyperparameter p and the corresponding set $\mathscr{V}_p$ of possible values of the hyperparameter p, while the other hyperparameters remain fixed. $\mathscr{V}_p$ is a subset of $\mathscr{V}$. To obtain the hyperparameter sets, we first define $\mathscr{V}_p$ and $\mathscr{V}$ results by $\mathscr{V} := \times_{p\in\mathscr{P}} \mathscr{V}_p$. $\mathscr{V}_p$ is chosen very large, spanning a wide range of possible values. The grid $\mathscr{V}_p$ of possible hyperparameter values which constitutes the basis of the preselection technique is selected randomly. For this purpose, we specify lower and upper bounds for the respective hyperparameter and select uniformly at random a given number of values within this interval. Note that $\mathscr{V}_p$ may also be chosen in a different way.

Subsequently, a Grid Search with a performance measure μ is applied where only the one currently considered hyperparameter varies. Based on a considered selection procedure, it results in K training and corresponding validation sets. The validation score of the model m on the validation set, $l = 1, .., K$, is denoted by $\mu_l^{m,p,v}$ where $v \in \mathscr{V}_p$ denotes the value of the hyperparameter p. The averaged measure over the K validation scores as well as the empirical standard deviation are then defined by

$$\overline{\mu}^{m,p,v} := \frac{1}{K}\sum_{l=1}^{K} \mu_l^{m,p,v},$$

$$\sigma_{\mu^{m,p,v}} := \sqrt{\frac{1}{K-1}\sum_{l=1}^{K} (\mu_l^{m,p,v} - \overline{\mu}^{m,p,v})^2}.$$

Based on these values, we can compute the empirical mean and standard deviation on the whole grid $\mathscr{V}_p$ of the hyperparameter p as follows:

$$\overline{\mu}^{m,p} := \frac{1}{\sharp(\mathscr{V}_p)}\sum_{v\in\mathscr{V}_p} \overline{\mu}^{m,p,v},$$

$$\sigma_{\overline{\mu}^{m,p}} := \sqrt{\frac{1}{\sharp(\mathscr{V}_p)-1}\sum_{v\in\mathscr{V}_p} (\overline{\mu}^{m,p,v} - \overline{\mu}^{m,p})^2}.$$

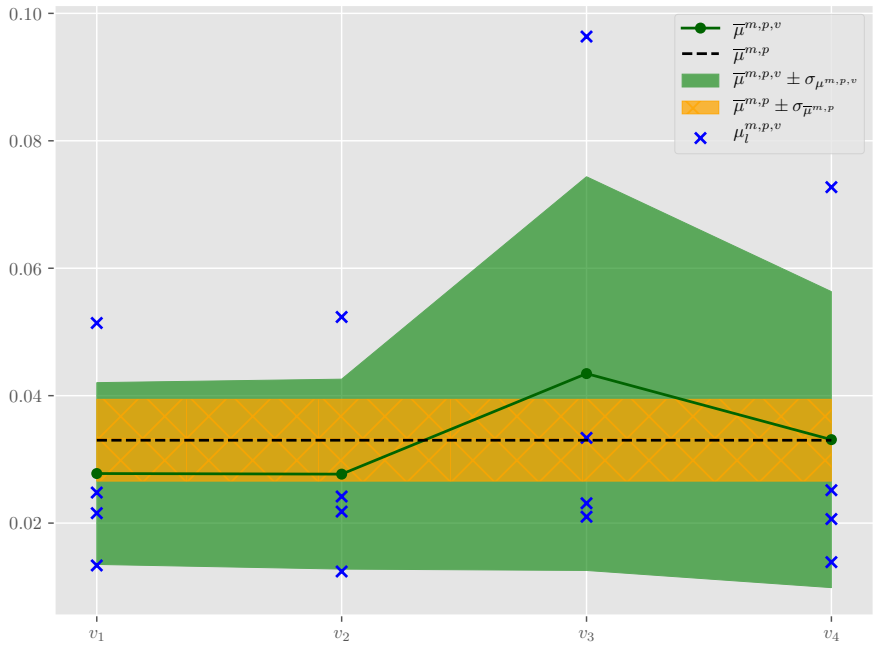


Figure 3.4: Validation curve, different means of the performance measure, and calculated quantities in the preselection step of the hyperparameter optimization, $K = 4$

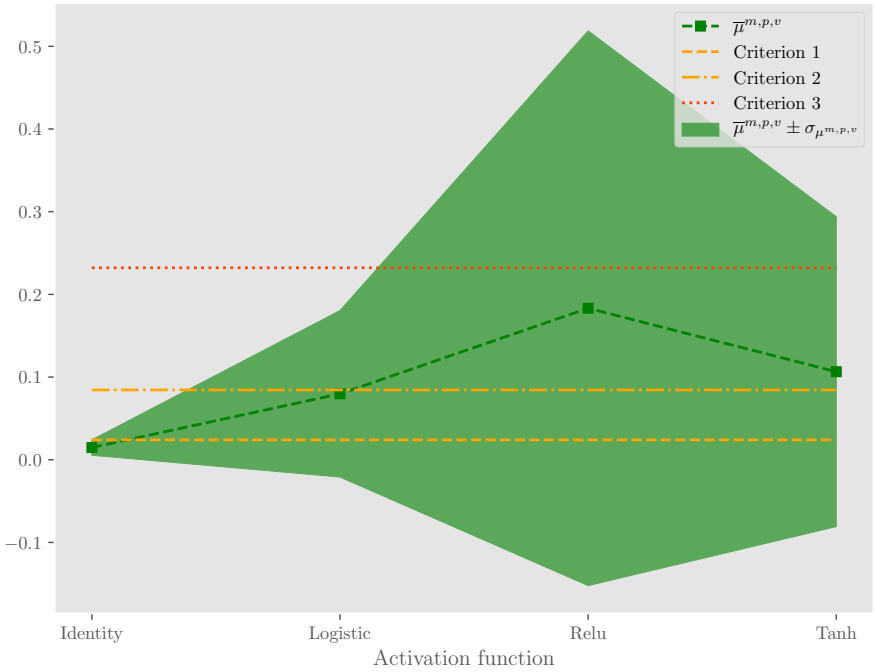


Figure 3.5: Preselection criteria of the hyperparameter optimization

The function mapping v onto $\overline{\mu}^{m,p,v}$ is referred to as *validation curve*. Figure 3.4 illustrates this curve, the various values under μ as well as the resulting statistical properties.

If $\overline{\mu}^{m,p,v}$ fluctuates only slightly among the various values of p, i.e. if $\sigma_{\overline{\mu}^{m,p}}$ is small, the hyperparameter p has no significant impact on the performance of the MLP and we consider the default value for p without loss of generality. If various values of the hyperparameter result in the same error, we choose the minimum value. The selection of possible values for the hyperparameter p is based on the following criteria which aim at ensuring that the forecast error is small and not very volatile:

- Criterion 1: Select all values $v \in \mathcal{V}_p$ satisfying

$$\overline{\mu}^{m,p,v} \leqslant \min_{j \in \mathcal{V}_p} \left(\overline{\mu}^{m,p,j} + \sigma_{\mu^{m,p,j}} \right) =: c_1.$$

- Criterion 2: Select all values $v \in \mathcal{V}_p$ satisfying

$$\overline{\mu}^{m,p,v} \leqslant \min_{j \in \mathcal{V}_p} \left(\overline{\mu}^{m,p,j} \right) + \sigma_{\overline{\mu}^{m,p}} =: c_2.$$

- Criterion 3: Select all values $v \in \mathcal{V}_p$ satisfying

$$\overline{\mu}^{m,p,v} + \sigma_{\mu^{m,p,v}} \leqslant \min_{j \in \mathcal{V}_p} \left(\overline{\mu}^{m,p,j} + \sigma_{\mu^{m,p,j}} \right) + std\left(\overline{\mu}^{m,p,v} + \sigma_{\mu^{m,p,v}} \right)$$

  where std is the standard deviation.

Criterion 1 is concerned with the forecast error for each hyperparameter value separately. In contrast to this, Criteria 2 and 3 are concerned with the relations among the performance measures of the other hyperparameter values. Criterion 2 takes into account the volatility of the validation curve. In Criterion 3 the standard deviation of the upper bound of the confidence interval of the validation curve is considered. Hence, Criterion 3 excludes those values having a low mean performance score but a high fluctuation of the scores across the K validation sets. That is to say, the performance measures vary strongly among the various folds and thus the forecast cannot be trusted.

In Figure 3.5 the three criteria with their boundaries are illustrated for the hyperparameter p being the activation function. Depending on the criterion different hyperparameter values are selected. Criterion 1 selects the identity function, while Criterion 2 additionally selects the logistic function. Criterion 3 chooses the identity and logistic function as Criterion 2. Since, as shown in Figure 3.5, the selected values of the hyperparameter p and thus the restriction of $\mathcal{V}_p$ to $\mathcal{V}_p^*$ depend on the selection criterion, we combine these criteria by taking the intersection of the individual selections. Note that the set $\mathcal{V}_p^*$ resulting from the intersection of Criteria 1 and 2 is non-empty since always one of these two criteria leads to a subset of the selection obtained by the other criterion. Indeed, we have

$$c_i \leqslant c_j \Longrightarrow \mathcal{V}_{p,i}^* \subseteq \mathcal{V}_{p,j}^*$$

for all $i, j \in \{1, 2\}$ and $i \neq j$. The parameter set selected by Criterion 3 can have an empty intersection with the parameter grids selected by Criterion 1 and Criterion 2. In this case, we take the union rather than the intersection since we cannot obtain a consistent set of suitable values for the hyperparameter p. Thus, all values that are relevant to any of the three criteria are chosen. Applying this approach for all hyperparameter $p \in \mathscr{P}$ results in a selected hyperparameter combination set $\mathscr{V}^* := \times_{p \in \mathscr{P}} \mathscr{V}_p^*$.

Step 2: Finding the best hyperparameters

Next, we apply Grid or Random Search based on the same selection procedure to the set $\mathscr{V}^*$ obtained in Step 1 in order to find the best hyperparameter combination of the considered machine learning model, i.e. leading to the smallest validation score $\overline{\mu}^m$ averaged over the K validation sets, cf. Section 3.2.3.

3.3.3 Performance Measures

After the description of the considered models in the previous sections, the different performance measures are listed in the following. We consider three performance measures evaluating the models: the mean squared error (MSE), the mean absolute error (MAE), and the maximum error (MAXE). Of all these measures, the MSE is the best known. It is defined by

$$\textit{MSE} := \frac{1}{N}\frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{n}\left(x_{t_k}^{(i)} - \widehat{x}_{t_k}^{(i)}\right)^2$$

where $\widehat{x}_{t_k}$ are the forecasts, x_{t_k} are the true values at time t_k, $k \in \{1, \ldots, n\}$, N the number of the considered yields, and n the number of observations for which a forecast is made. The MSE is not only used for evaluating the models but also as the loss function minimized during the training of the MLP.

The MAE is calculated by

$$\textit{MAE} := \frac{1}{N}\frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{n}|x_{t_k}^{(i)} - \widehat{x}_{t_k}^{(i)}|.$$

Mathematically, the MAE corresponds to the ℓ^1-norm and is more interpretable than the MSE since it indicates how far the forecast deviates from the true value on average. Last, we consider the MAXE calculated by

$$\textit{MAXE} := \max_{i=1,\ldots,N;k=1,..,n}\left(|x_{t_k}^{(i)} - \widehat{x}_{t_k}^{(i)}|\right).$$

Since these performance measures specify the error in the forecast, a smaller measure is better than a larger one.

3.3.4 Selection Procedure

In order to choose a suitable model and appropriate corresponding hyperparameters of the MLP, we combine cross validation with an expanding window approach. Our motivation is based on the following observations:

- Neural networks allow for a vast amount of hyperparameter combinations that have a significant impact on the quality of the forecast. The set of possible hyperparameter combinations cannot easily be reduced by a simple analysis of the data.

- Even after a preselection of hyperparameters, we have too many time periods for efficiently running a proper expanding window loop with reasonable time shifts for hyperparameter optimization.

- We want to choose the best model via expanding windows since this reflects a real-world application better by not looking into the future.

We optimize the hyperparameters of the MLP that belong to one of the categories in Definition 3.1 via our approach introduced in Section 3.3.2. Due to our motivation, we use a 4-fold cross validation with contiguous data points as a validation set in the approach. The loss function of the MLPs which is minimized is the MSE. After selecting the best hyperparameter combination, the classical time series models and the MLP are evaluated by an expanding window approach. We have a total of 60 validation sets. As performance measure the MSE, MAE, and MAXE of the validation sets are compared. Since the performance of the MLP depends on random states which specify e.g. the initial weights, 45 MLPs are run with the same optimal hyperparameters but different initial random states. The mean of the performance measures of all 45 MLPs is used as the performance of the MLP.

Since we do not want to find the best MLP but comparing the models and analyzing the benefit of MLPs towards classical models, cross validation and expanding windows can be combined. It is important that the evaluation of the models fits the practical application which is the case here. How we choose the hyperparameters of the MLP is independent of this. MLPs with better performance measures on the expanding window approach can be found if the hyperparameter optimization is based on the same selection procedure. However, it is more computationally intensive.

3.4 Forecasting the European Yield Curve

In this section, we describe the data of the European yield curve, the considered yields, time periods, and forecast horizons. The results of the evaluation are also presented in this section.

3.4.1 Yield Curve Data

For estimation and validation of the models we use the zero-coupon yield curve for triple-A-rated government bonds published by the ECB[4]. The zero coupon yields of times to maturity of 3, 6, 9 months, and from one year to 30 years in one year steps are provided. Data points are daily and compose a time span of 15 years, from 2004 September 6$^{\text{th}}$ until 2019 August 31$^{\text{th}}$. It is a sample of 3,834 daily observations. As calendar, the trading day calendar is used. According to this and the data, one year has 256 trading days on average, half a year 128 trading days, one quarter of a year 64 trading days, and one month 21 trading days. These are also the different forecast horizons we consider besides one trading day.

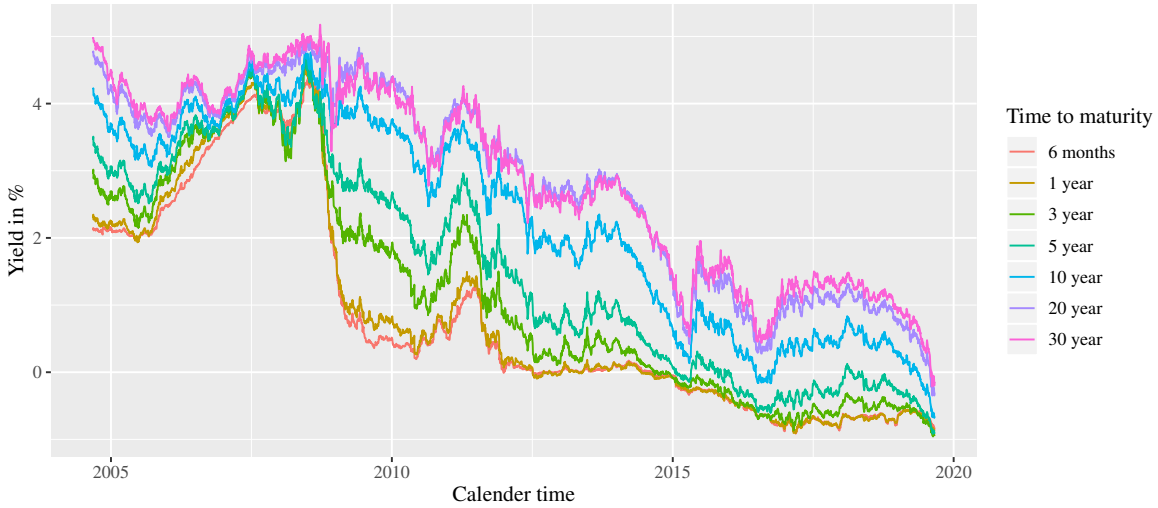


Figure 3.6: Evolution of considered zero-coupon yields from 2004 September 6$^{\text{th}}$ to 2019 August 31$^{\text{th}}$

Figure 3.6 shows the evolution of selected zero-coupon yields over calendar time. All yields decrease in the last quarter of 2004 and the first half of 2005. Then a significant increase of the short-term yields from about $2\,\%$ to $4\,\%$ can be seen before the financial crisis of 2007. Simultaneously, the long-term yields do not increase to the same extent, but all yields are close together. In the second half of 2008, the short-term yields rapidly decrease, while the other yields go down slower. Until 2017 there is a tendency of decline in the yields, while afterwards the level remains steady. This is followed by a decrease of the yields in 2019 once more.

Table 3.7 reports the descriptive statistics for selected yields based on the data. The time to maturity is measured in years. For each time series the mean, standard deviation, minimum, maximum, skewness, kurtosis, autocorrelation for a lag of 1, 256 as well as 512 trading days, and the augmented Dickey-Fuller (ADF) test statistic are shown. The following well-known facts regarding yield curve data[5] are also seen

[4] `https://sdw.ecb.europa.eu/browse.do?node=9691417`, accessed on 2020 July 14$^{\text{th}}$.
[5] See Diebold and Li (2006), p. 343.

here: The average mean yield curve is upward sloping, long-term yields are less volatile and more persistent than short-term yields, and the autocorrelations are very high even after one year (corresponds to the lag of 256). Due to the high lag-1 autocorrelation, we can assume a linear relation between the time series and its own with lag-1 delayed values. The ADF test suggests that each time series has a unit root and is consequently non-stationary.

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-256	Lag-512	ADF
0.25	0.8191	1.5317	-0.9300	4.3255	0.9203	-0.4952	0.9998	0.8154	0.5841	-0.9450
0.5	0.8453	1.5680	-0.9147	4.3570	0.9055	-0.5558	0.9999	0.8201	0.5957	-0.9169
0.75	0.8739	1.5875	-0.9126	4.4474	0.8768	-0.6046	0.9999	0.8313	0.6161	-0.8467
1	0.9047	1.5990	-0.9079	4.5396	0.8392	-0.6556	0.9999	0.8435	0.6393	-0.7662
2	1.0463	1.6156	-0.9337	4.7138	0.6537	-0.9012	0.9998	0.8819	0.7239	-0.6263
3	1.2117	1.6142	-0.9567	4.7385	0.4725	-1.1288	0.9998	0.9048	0.7823	-0.5547
4	1.3903	1.6052	-0.9484	4.7343	0.3187	-1.2921	0.9997	0.9184	0.8201	-0.4524
5	1.5704	1.5927	-0.9188	4.7304	0.1912	-1.3964	0.9997	0.9266	0.8442	-0.3571
6	1.7435	1.5788	-0.8755	4.7291	0.0857	-1.4565	0.9997	0.9313	0.8592	-0.2711
7	1.9040	1.5650	-0.8243	4.7358	-0.0019	-1.4856	0.9997	0.9337	0.8680	-0.1943
8	2.0494	1.5520	-0.7752	4.7484	-0.0745	-1.4939	0.9997	0.9345	0.8726	-0.1438
9	2.1786	1.5402	-0.7274	4.7624	-0.1346	-1.4892	0.9997	0.9343	0.8744	-0.0074
10	2.2920	1.5295	-0.6803	4.7763	-0.1837	-1.4769	0.9997	0.9334	0.8742	0.0366
15	2.6628	1.4877	-0.4807	4.8722	-0.3130	-1.3962	0.9997	0.9253	0.8627	0.0812
20	2.8213	1.4534	-0.3473	4.9849	-0.3279	-1.3546	0.9996	0.9188	0.8553	0.1051
25	2.8773	1.4230	-0.2659	5.0984	-0.2906	-1.3484	0.9995	0.9150	0.8560	-0.0294
30	2.8868	1.3984	-0.2105	5.1750	-0.2302	-1.3562	0.9994	0.9114	0.8587	-0.2253

Table 3.7: Descriptive statistics for specific yields of the ECB over the period 2004 September 4$^{\text{th}}$ to 2019 August 31$^{\text{th}}$ (3,834 daily observations)

For the yield curve forecasting, we consider different times to maturity and their relating yields. They are selected according to containing a mix of short-, medium-, and long-term yields. We use the yields with times to maturity of 6 months, 1, 3, 5, 7, 10, and 20 years. For all time points, these seven yields are considered. Remember that we do not directly forecast the absolute yields of different times to maturity but the increments of the yields. The descriptive statistic of the increments can be found in Appendix B. It contains the lag-1, lag-21, lag-256 autocorrelation, and the lag-2 partial autocorrelation. The lag-1 partial autocorrelation corresponds to the lag-1 autocorrelation and is therefore not listed separately. The difference between the lag-1 partial autocorrelation and the lag-2 one is very large for all increments except the lag-1 increments. Here, the lag-1 autocorrelation is already small. This observation indicates that its previous to last value does not have a large direct impact on the current value which justifies the assumption of using only the previous values as input variables. We use the mean value of the seven considered yields as an additional endogenous variable as an indicator of the yield level.

The data has a clear, structural change in the first three months of 2009, cf. Figure 3.6. To avoid including this time span, we consider the different increments of the yields from 2010 March 1$^{\text{st}}$ to 2019 August 31$^{\text{th}}$. Therefore, we use the incre-

ments of one year, as the maximum forecast horizon, after the structural change. The data set includes 2,430 daily observations.

3.4.2 Evaluation

In this section, we present the numerical results of the model comparison as explained in Section 3.3. The best hyperparameter combination of the MLPs is chosen via our approach introduced in Section 3.3.2 and a 4-fold cross validation over the period 2010 March 1$^{\text{st}}$ to 2019 August 31$^{\text{th}}$. Table 3.8 contains the respective best hyperparameter combination for each of the forecast horizons under consideration. These MLPs are used for the evaluation of the models via expanding windows. In Table 3.8a the time point t_k is not used as an additional feature, while it is used in Table 3.8b. The hyperparameters are described by the terms of the scikit-learn package. Here, the network architecture specifies the number of hidden layers and their neurons. The input layer and output layer are fixed. Lbfgs is an optimizer in the family of quasi-Newton methods, adam refers to a stochastic gradient-based optimizer proposed by Kingma and Ba (2014), and sgd to stochastic gradient descent.

(a) Without time point t_k as additional input variable

Hyperparameter	Forecast horizon (in trading days)				
	1	21	64	128	256
Network architecture	(55, 40, 60, 60)	(10, 65)	(55)	(25, 5, 80, 55)	(70, 85, 25)
Activation function	logistic	logistic	logistic	identity	tanh
Optimizer	lbfgs	adam	adam	sgd	sgd
Alpha	0.797	0.797	0.0855	0.0855	0.0015
Learning rate	constant	adaptive	constant	adaptive	adaptive
Initial learning rate	0.001	0.001	0.001	7.42e-05	0.000315
Momentum	0.9	0.729	0.9	0.767	0.951
Beta 1	0.9	0.684	0.92	0.9	0.9
Beta 2	0.999	0.999	0.262	0.999	0.999
Normalization	MinMaxScaler	MinMaxScaler	MinMaxScaler	MinMaxScaler	StandardScaler

(b) With time point t_k as additional input variable

Hyperparameter	Forecast horizon (in trading days)				
	1	21	64	128	256
Network architecture	(60, 45, 45, 80)	(10, 65)	(85)	(90, 20, 75, 40)	(35, 75, 70, 75)
Activation function	identity	logistic	logistic	tanh	relu
Optimizer	lbfgs	adam	adam	sgd	sgd
Alpha	0.797	0.791	0.797	0.797	0.00672
Learning rate	constant	adaptive	adaptive	adaptive	adaptive
Initial learning rate	0.001	0.000483	0.001	7.42e-05	0.000315
Momentum	0.9	0.767	0.729	0.767	0.886
Beta 1	0.9	0.61	0.075	0.563	0.72
Beta 2	0.999	0.999	0.999	0.343	0.999
Normalization	MinMaxScaler	MinMaxScaler	MinMaxScaler	MinMaxScaler	StandardScaler

Table 3.8: Hyperparameter values of the best MLP

The MLPs with these hyperparameters are used in the expanding window approach comparing the different models. The data from 2010 March 1st to 2013 August 31th constitute the first training set for training the models. The trained models are then evaluated on the next month from 2013 September 1st to 2013 September 30th. Subsequently, the training set is extended by the previous validation set and the next month is used as the new validation set. We proceed until 2018 August 31th. In total, we have 60 different validation sets and scores which are averaged to obtain one score. As performance measure, we consider the MSE, MAE, and MAXE of the validation sets. Remember that 45 MLPs are run with the same optimal hyperparameters but different initial random states. The mean of their performance measures is used as the performance of the MLP. The results of the models evaluation are summarized in Table 3.9.

The MSE is so low because the yields are less than one and the MSE becomes even smaller due to squaring. The other performance measures are larger by an order of magnitude. The MLP is the best model for all performance measures and forecast horizons except for the MAXE and a forecast horizon of one trading day as well as 128 trading days. The AR(1) model performs best in this case. The next best model after the MLPs is the persistence model for short-term up to medium-term forecast horizons under the performance measure of MSE. Under the MAE and MAXE there is one forecast horizon each of the short-term up to medium-term forecast horizons where an autoregressive model is the best classical statistical model instead of the persistence model. The AR(1) model would be chosen after the MLPs for a forecast horizon of 128 trading days and the VAR(1) model for a forecast horizon of 256 days. The differences between the performance measure of the best MLP and the best other models increase with the forecast horizon. The benefit of using a MLP is greatest for a forecast horizon of 256 trading days. For the MAE the difference of the best classical statistical model and the MLP is 0.06927 $\%$ which means that the MLP is 0.06927 $\%$ closer to the true value on average than the best classical statistical model. This difference is clearly noticeable for large amounts as well as in times of low interests.

Whether a forecast with time point t_k as an additional variable is better than without can not be said in general. However, the consideration of the time point t_k is better for the best model for medium-term forecast horizons and worse for short- and long-term forecast horizons.

In Table 3.10 the different runs of the MLP are compared to the best model of the classical statistic models for the different performance measures. For this, we state the minimum and maximum error of the different MLP runs as well as the maximum quantile $q \in \mathbb{N}$ of the MLP errors such that the error of the best classical statistic model is larger than the error in q $\%$ of the runs of the best MLP. A quantile of zero means that the error of the best classical statistic model is smaller than the minimum error of the MLPs. That is the case for the MAXE and the forecast horizon of one trading day and 128 trading days, see Table 3.10c. If the error of the best classical statistic model is larger than the maximum error of the MLPs the quantile is 100 $\%$. The MLP is the best model in all runs. This is given for the MSE and MAXE for two forecast

(a) Forecast horizon of 1 trading day

	MSE (10^{-8})	MAE (10^{-4})	MAXE (10^{-4})
Persistence	7.3187	1.8360	9.8059
AR(1)	7.3452	1.8421	**9.7803**
AR(1) with t_k	7.3472	1.8395	9.8234
VAR(1)	7.4556	1.8563	9.8211
VAR(1) with t_k	7.4532	1.8533	9.8594
MLP	**7.3182**	**1.8322**	9.8538
MLP with t_k	8.3390	1.9874	10.1513

(b) Forecast horizon of 21 trading days

	MSE (10^{-6})	MAE (10^{-4})	MAXE (10^{-4})
Persistence	1.4935	8.7691	29.5428
AR(1)	1.5534	8.9267	29.7112
AR(1) with t_k	1.5720	8.7395	29.6659
VAR(1)	1.7237	9.4537	30.9675
VAR(1) with t_k	1.7503	9.3621	30.8799
MLP	1.4808	8.4356	**29.2209**
MLP with t_k	**1.4717**	**8.3906**	29.2636

(c) Forecast horizon of 64 trading days

	MSE (10^{-6})	MAE (10^{-4})	MAXE (10^{-4})
Persistence	4.8529	15.9896	44.7575
AR(1)	5.0147	16.3902	44.8930
AR(1) with t_k	5.5063	17.1468	46.7420
VAR(1)	5.2913	17.4416	47.5278
VAR(1) with t_k	6.6261	19.7642	51.6741
MLP	4.7320	15.6084	44.6375
MLP with t_k	**4.5727**	**14.7706**	**43.6202**

(d) Forecast horizon of 128 trading days

	MSE (10^{-6})	MAE (10^{-4})	MAXE (10^{-4})
Persistence	10.5736	23.9950	59.8552
AR(1)	9.7324	23.7923	**56.8065**
AR(1) with t_k	12.8400	28.9881	62.6128
VAR(1)	13.5881	28.4764	62.2101
VAR(1) with t_k	11.5109	27.9145	59.3082
MLP	**9.4606**	**22.9513**	60.3836
MLP with t_k	9.5980	23.1332	61.2511

(e) Forecast horizon of 256 trading days

	MSE (10^{-6})	MAE (10^{-4})	MAXE (10^{-4})
Persistence	24.9299	36.4237	79.2252
AR(1)	15.3891	30.9599	72.9785
AR(1) with t_k	18.0064	34.3573	67.9758
VAR(1)	11.8774	27.4735	66.26655
VAR(1) with t_k	11.9210	26.9073	62.6852
MLP	**7.1175**	**19.9796**	**54.6273**
MLP with t_k	8.6477	22.0782	56.8995

Table 3.9: Performance measure of the different forecast models. For all models, the own last value and the mean of the yields are used as input. Additionally, the time point t_k is considered in the models.

(a) MSE

Forecast horizon (trading days)	Best classical statistical model	Best MLP	Best classical model (10^{-6})	Minimum MLP (10^{-6})	Maximum MLP (10^{-6})	Quantile (in %)
1	Persistence	MLP	0.073187	0.073129	0.074495	91
21	Persistence	MLP with t_k	1.4935	1.4454	1.4936	100
64	Persistence	MLP with t_k	4.8529	4.1820	5.0653	96
128	AR(1)	MLP	9.7324	8.8540	10.5607	80
256	VAR(1)	MLP	11.8774	6.3349	8.3300	100

(b) MAE

Forecast horizon (trading days)	Best classical statistical model	Best MLP	Best classical model (10^{-4})	Minimum MLP (10^{-4})	Maximum MLP (10^{-4})	Quantile (in %)
1	Persistence	MLP	1.8360	1.8314	1.8471	95
21	AR(1) with t_k	MLP with t_k	8.7395	8.3204	8.4804	100
64	Persistence	MLP with t_k	15.9896	13.6958	15.7232	100
128	AR(1)	MLP	23.7923	22.0694	24.3601	94
256	VAR(1) with t_k	MLP	26.9073	18.5571	22.6239	100

(c) MAXE

Forecast horizon (trading days)	Best classical statistical model	Best MLP	Best classical model (10^{-4})	Minimum MLP (10^{-4})	Maximum MLP (10^{-4})	Quantile (in %)
1	AR(1)	MLP	9.7803	9.8283	9.9216	0
21	Persistence	MLP	29.5428	28.8869	29.5264	100
64	Persistence	MLP with t_k	44.7575	42.4912	44.9933	96
128	AR(1)	MLP	56.8065	57.2432	64.2952	0
256	VAR(1) with t_k	MLP	62.6852	50.9110	59.3307	100

Table 3.10: Comparison of the best classical statistical model to the 45 runs of MLP with the best hyperparameter combination

horizons, cf. Table 3.10a, and for the MAE for three, cf. Table 3.10b. In all other cases, the best classical statistical model lies in the quantile over 90 % except for the MSE and forecast horizon of 128 trading days. Here the quantile is still large at 80 %. The better performance of the MLP compared to the classical model is clearly visible.

Summarizing, the MLP outperforms the classical statistical models for every considered forecast horizon under the MSE and MAE. Under the MAXE the MLP is the best in three out of five forecast horizons. The best model after the MLP is the persistence model for short- and medium-term forecast horizon (until 64 trading days). The benefit of using a MLP over the persistence model is very small for the forecast horizon of one trading day and 21 trading days. Therefore, the question arises whether the costs of hyperparameter optimization are justifiable in this case. Based on the results above, we suggest to use a persistence model for a forecast horizon of

one trading day since the yield curve does not fluctuate much within one trading day. For medium-term forecast horizons the benefit cost relation should be considered individually depending on the requirements of the model's accuracy and simplicity for a specific application. For long-term forecast horizons, however, we recommend to a MLP since it's forecast quality in that context is significantly superior to all other models under consideration.

3.5 Conclusion

The evolution of the yield curve is decisive for various areas and applications. For this reason, accurate forecasting is necessary. Therefore, we investigated the benefit of using deep learning algorithms that are potentially more capable in learning the underlying distribution compared to linear models that have the advantage of low complexity and high interpretability in the context of yield curve forecasting. As a data basis, we used the European yield curve data published by the ECB. For our examination, we compared the performance of persistence, autoregressive, and vector autoregressive models with MLPs. Since the performance strongly depends on the length of the forecast time interval, we considered short-, medium-, and long-term forecast horizons that result in different models that are appropriate. Evaluating the models an expanding window approach with 60 validation sets and the performance measures MSE, MAE, MAXE were used.

A crucial step for MLPs is the hyperparameter optimization. We extended standard procedures such as Grid and Random Search with a preselection of promising hyperparameter values. This is done automatically based on three criteria. Based on the preselection Grid or Random Search is run. Our method allows to search a larger hyperparameter space than usually possible with standard methods. It can be easily transferred to other applications and machine learning algorithms. We used our approach to optimize the hyperparameters of the MLPs on the data of the European yield curve.

The result of the hyperparameter optimization was the basis for the evaluation of the MLP. Since the performance of the MLP depends on random states which specify e.g. the initial random states, 45 MLPs were run with the same optimal hyperparameters but different initial random states. The performance measure of the MLP was then derived from the averaged performance of the 45 MLPs. Then, these were compared to the performance of the persistence and the simple linear models. For all considered forecast horizons the MLP outperformed the classical statistical models under the performance measure of MSE and MAE. The second-best model was the persistence model or the autoregressive model depending on the forecast horizon. Under the performance of MAXE, the MLP was the best model at three of five forecast horizons. For the remaining two forecast horizons the autoregressive model outperformed the MLP. However, we do not suggest to use MLPs for all forecast horizons. For short-term forecast horizons, the difference between the MLP and the persistence model is so small that the cost of training and optimizing the MLP should influence

the model choice considerably. For long-term forecast horizons, MLPs should be used in any case.

The models do not consider macroeconomic variables as well as different regimes of the yield curve as exogenous variables. These are enhancements for further studies since our focus was on the comparison of linear models with MLPs and we did not want to develop the best model for forecasting the yield curve including other exogenous variables which e.g. represent the economic situation. Besides the consideration of additional variables, the MLPs can be replaced by other machine learning algorithms like recurrent neural networks and Random Forest. This is left for further studies.

4 Chance-Risk Classification of a Portfolio consisting of State-Subsidized Pension Products and its Impact on Customer Consulting

Consulting in the finance industry mostly consists of two parts, at least during the initial consultation. The customer's risk profile is determined based on a questionnaire and is aggregated into a risk class. Afterwards, matching products should be offered. For this, investment products are classified into different asset classes. This procedure is intended to ensure that the customer is recommended products that are suited for him and match his risk preference. In Germany for example, this is regulated in § 64 (3) of the Wertpapierhandelsgesetz. The same procedure has also been established within the insurance industry for consulting on retirement provision products. This approach is simple and practical, but it does not include products already owned by customers. Due to diversification effects, however, the total view of the customer's portfolio may differ from an individual view of each product which may result in different investment recommendations.

For state-subsidized pension products, we extend the above described consulting approach to include the customer's portfolio. We choose this type of product since there is a unified classification by PIA[1] in Germany which is missing for other financial products. Therefore, we are restricting the scope of our study to this country only. Each state-subsidized pension product is assigned to one of five so-called CRCs[2]. Based on this classification of single products, we develop a method for determining the CRC of a portfolio. As a result, we are able to

- determine the maximum premium in a new pension product so that the customer's portfolio has no larger CRC than his risk preference.

- offer a wider range of products to the customer. This is particularly important in the insurance industry where each insurance company offers significantly fewer products than in the banking sector.

[1] PIA: Produktinformationsstelle Altersvorsorge gGmbH.
[2] CRC: chance-risk class.

- compensate for previously incorrect consulting via the portfolio.
- create new pension products with a specific CRC as upper bound.

Additionally, we introduce a procedure for determining the CRC of a state-subsidized pension product over the contract term since it can be affected by the current evolution of the capital market.

This chapter is divided into four sections. In Section 4.1, we describe the classification of state-subsidized pension products by PIA. This includes the description of the chance and risk parameters calculation of the classification in Section 4.1.1 as well as the class boundaries definition in Section 4.1.2. The following section investigates the properties of the chance and risk parameters as well as their mappings. On the basis of these, the existence of a diversification effect is shown and the CRC of a portfolio is determined. The results are summarized in Section 4.3. At this, both a portfolio consisting of two new pension products, Section 4.3.2, and one of an already purchased and a new pension product, Section 4.3.3, are considered. Furthermore, implications for customer consulting are given on the basis of our results. Finally, different approaches for determining the evolution of the pension product's CRC over the contract term are introduced in the last section.

4.1 Classification of State-Subsidized Pension Products

Since 2017 it has been a legal requirement that every state-subsidized pension product sold in Germany must be classified in one of five CRCs and that this classification is shown on the corresponding PIB[3]. This classification is assigned by PIA and based on a mathematical capital market simulation and the corresponding evolution of the contract value of the pension product. It is explained in the following section. For literature, we refer to Korn and Wagner (2018) and Section 3.2 of Korn and Wagner (2019).

4.1.1 The Classification Algorithm

The classification is based on the distribution of the simulated contract value at the end of the accumulation phase of an idealized customer who is defined in § 14 of the Altersvorsorge-Produktinformationsblattverordnung (AltvPIBV). For the idealized customer, contracts with different kinds of premium payments – regular payment or single payment – as well as with different accumulation phases T of 12, 20, 30, and 40 years are specified by law. For a regular premium payment, it is assumed that the customer pays 100 Euros (including state bonuses) to the insurer every month until the end of the accumulation phase T. For the same accumulation phase, the sum of all premium payments of the regular payment corresponds to the premium of the single payment which is paid at the beginning of the accumulation phase. Thus, the

[3]PIB: Produktinformationsblatt.

idealized customer pays $1,200 \cdot T$ Euros as single premium. We denote the premium of the idealized customer by P_T and it is defined by

$$P_T = \begin{cases} 100 \text{ Euros}, & \text{regular monthly premium in [0,T)} \\ 1200 \cdot T \text{ Euros}, & \text{single premium at the beginning of the accumulation phase.} \end{cases}$$

The idealized customer with a single premium payment is only considered if the pension product allows only a single premium payment and no regular premium according to § 14 (1) sentence 2 of the AltvPIBV. Since the contract value depends on the evolution of the capital market, 10,000 scenarios of the pension product's evolution are generated per simulation for each accumulation phase T. In the simulation, costs and product-specific properties are considered, whereas management decisions regarding strategy and investments are not simulated. We denote the simulated contract values at the end of a specific accumulation phase T by v^k, $k = 1, \ldots, 10,000$. Without loss of generality, we assume that v^k is non-negative in all scenarios. This corresponds to the assumption that the customer can at most loose his paid premiums over the accumulation phase. If the pension product has a contractual money-back guarantee, meaning that at least the sum of paid premiums including state bonuses has to be credited to the customers account at the end of the accumulation phase, v^k is the maximum of the simulated final contract values and $1,200 \cdot T$.

Based on the final contract values v^k, the chance and risk measure which present deterministic annual interest rates are calculated. For the chance measure, the mean of the simulated contract values is calculated by

$$\overline{V}^c = \mathbb{E}(V) \tag{4.1}$$

where $\mathbb{E}$ denotes the expected value with respect to the empirical distribution of the sample and V defines the final contract values of the pension product. The risk measure μ^r is based on the mean of the 2,000 lowest final contract values which is given by

$$\overline{V}^r = \mathbb{E}\left(V | V \leqslant Q_{0.2}\right) \tag{4.2}$$

where $Q_{0.2}$ is the $20\ \%$ quantile of V defined by $P\left(V \leqslant Q_p\right) = p$.

Afterwards, the largest constant interest rates μ^c and μ^r are determined based on the customer's monthly payments or single payment which lead to the above calculated means $\overline{V}^c$ and $\overline{V}^r$. In this calculation, a simple savings process without costs or randomness is assumed. Thus, the relation between the averaged final contract value $\overline{V}$ and its corresponding constant interest rate μ is

$$\overline{V} = \begin{cases} 100 \sum_{k=1}^{12 \cdot T} \left(1 + \frac{\mu}{12}\right)^k, & \text{regular premium} \\ 1200 \cdot T \left(1 + \frac{\mu}{12}\right)^{12T}, & \text{single premium.} \end{cases} \tag{4.3}$$

Then, μ^c is the largest value which solves Equation (4.3) with $\overline{V}^c$ as $\overline{V}$ and μ^c as μ and μ^r the largest value which solves the Equation for μ^r with $\overline{V}^r$ as $\overline{V}$ and μ^r as μ.

μ^c and μ^r span a space which is divided in different areas by the CRC boundaries determined using reference products, see Section 4.1.2. These areas define the CRCs and the products are classified according to their μ^c and μ^r combination.

PIA determines a CRC for the provider for each accumulation phase of the idealized customer depending on the simulated premium payment. According to § 5 (2) sentence 4 of the AltvPIBV, the agreed accumulation phase of the offered pension product, denoted by $\mathcal{T}$, determines the accumulation phase T of the classification and which of the four CRC is assigned to the product. It holds

$$\text{for} \left\{ \begin{array}{c} \mathcal{T} \in (0;12] \\ \mathcal{T} \in (12;20] \\ \mathcal{T} \in (20;30] \\ \mathcal{T} \in (30;\infty) \end{array} \right\} \text{ the CRC determined for } \left\{ \begin{array}{c} T = 12 \\ T = 20 \\ T = 30 \\ T = 40 \end{array} \right\}$$

has to be shown on the PIB.

4.1.2 Chance-Risk Class Boundaries

The classification of a pension product requires a clear allocation of the pair (μ^c, μ^r) to one of five CRCs. The CRC characterizes different earnings potential and risk whereby a higher CRC is accompanied by higher risk and higher earnings potential. They are defined in Produktinformationsstelle Altersvorsorge (2016) and in Schreiben des Bundesministeriums der Finanzen vom 14.03.2019 as follows:

The CRC specifies how the earnings potential and risks of this pension product are assessed compared to other state-subsidized pension products. For an idealized customer, the pension product is evaluated by the PIA for various capital market scenarios over a comparable accumulation phase obtained by § 5 (2) sentence 4 of the AltvPIBV. Thereby, a money-back guarantee is taken into account. The so called Riester pension products always contain one.[4] A money-back guarantee is given if at least the paid premiums and state bonuses are available for pension cover at the end of the accumulation phase. Premiums for an additional insurance as well as premiums which relate to already paid-out capital are deducted.[5]

[4] Original text: "Die Chancen-Risiko-Klasse (CRK) gibt an, wie die Ertragschancen und Risiken dieses Produkts gegenüber anderen steuerlich geförderten Altersvorsorgeprodukten einzuschätzen sind. Für einen Musterkunden hat die unabhängige Produktinformationsstelle Altersvorsorge dieses Produkt für verschiedene Kapitalmarktszenarien über eine vergleichbare Ansparphase von X Jahren untersucht und in die CRK Y eingeteilt. Dabei wurde berücksichtigt, ob dieses Produkt zu Beginn der Auszahlungsphase eine Beitragserhaltungszusage enthält. Riester-Produkte enthalten immer eine Beitragserhaltungszusage." (Schreiben des Bundesministeriums der Finanzen vom 14.03.2019, p. 9).

[5] Original text: "Mit einer Beitragserhaltungszusage sagt der Anbieter Folgendes zu: Zu Beginn der Auszahlungsphase stehen mindestens die eingezahlten Beiträge und ggf. gezahlten Altersvorsorgezulagen zur Altersabsicherung zur Verfügung. Beitragsanteile für eine Zusatzabsicherung

CRC 1 The pension product offers a secure investment through a fixed guaranteed (minimum) positive interest rate over the whole accumulation phase or an interest rate linked to a reference interest rate. The earnings potential is low. The irrevocably accumulated capital after deducting the costs continuously increases in the accumulation phase. The insurer issues a money-back guarantee.

CRC 2 The pension product offers a safety-oriented investment with limited earnings potential. The insurer issues a money-back guarantee.

CRC 3 The pension product offers a balanced investment with moderate earnings potential. If the insurer does not issue a money-back guarantee, there is a moderate risk of loss.

CRC 4 The pension product offers a return-oriented investment with higher earnings potential. If the insurer does not issue a money-back guarantee, there is a higher risk of loss.

CRC 5 The pension product offers an opportunity-oriented investment with high earnings potential. If the insurer does not issue a money-back guarantee, there is a high risk of loss.[6]

Note that CRC 1 and 2 also include additional qualitative criteria besides the earnings potential and risk. Products in CRC 1 and 2 need a money-back guarantee, while CRC 1 additionally has the requirement of a continuously increasing guaranteed capital after deducting the costs in the accumulation phase. The money-back guarantee must be contractually assured.

For classification of a pension product into one CRC, the classes have to be clearly defined by both chance and risk measure. To determine the boundaries of the CRCs,

und Beitragsanteile, die bereits ausbezahltes Kapital betreffen, werden dabei abgezogen." (Produktinformationsstelle Altersvorsorge (2016)).

[6] Original text: "CRK 1 Das Produkt bietet eine sichere Anlage durch eine bis zum Beginn der Auszahlungsphase festgelegte garantierte (Mindest-)Verzinsung oder an einen Referenzzins gekoppelte Verzinsung mit niedrigen Ertragschancen. Das unwiderruflich gebildete Kapital nach Abzug der Kosten steigt in der Ansparphase fortwährend an. Der Anbieter gibt eine Beitragserhaltungszusage.
CRK 2 Das Produkt bietet eine sicherheitsorientierte Anlage mit begrenzten Ertragschancen. Der Anbieter gibt eine Beitragserhaltungszusage.
CRK 3 Das Produkt bietet eine ausgewogene Anlage mit moderaten Ertragschancen. Gibt der Anbieter keine Beitragserhaltungszusage, so besteht ein moderates Verlustrisiko.
CRK 4 Das Produkt bietet eine renditeorientierte Anlage mit höheren Ertragschancen. Gibt der Anbieter keine Beitragserhaltungszusage, so besteht ein höheres Verlustrisiko.
CRK 5 Das Produkt bietet eine chancenorientierte Anlage mit hohen Ertragschancen. Gibt der Anbieter keine Beitragserhaltungszusage, so besteht ein hohes Verlustrisiko." (Schreiben des Bundesministeriums der Finanzen vom 14.03.2019, p. 10).

different reference portfolios without costs are simulated. No costs are considered so that the CRC does not depend on cost concepts. Furthermore, no product concepts are included in the reference portfolio to make the CRCs independent of them as well. The reference portfolios are chosen such that they represent the characteristic of the respective CRC. Especially, the money-back guarantee and the continuously increasing capital are considered in the definition of the reference portfolios.

Reference portfolio 1 Zero-coupon bonds maturing at the end of the accumulation phase are bought with every premium.

Reference portfolio 2 As many zero-coupon bonds that mature at the end of the accumulation phase as necessary to secure the premium are bought with every premium. From the remaining premium payment call options on the risky asset with at-the-money strike and a one-year term are bought. In the last year of the accumulation phase, the maturity of the call option equals the end of the accumulation phase. Any profits of the call options are credited to the customers account and also secured with zero-coupon bonds with maturity at the end of the accumulation phase.

Reference portfolio 3 A portfolio containing $50\ \%$ stocks is held. The remaining amount is invested in zero-coupon bonds with a constant maturity of 10 years.

Reference portfolio 4 A portfolio containing $75\ \%$ stocks is held. The remaining amount is invested in zero-coupon bonds with constant a maturity of 10 years.

Reference portfolio 5 A portfolio comprised only of stocks is held.

The reference portfolios are the same for single premium payments as well as for regular premium payments. However, only a one-off allocation of the premium in the investment components is made at the beginning of the accumulation phase.

For each reference portfolio, the chance and risk measure are calculated as explained in Section 4.1.1. Then, the boundaries between the CRCs are defined as a straight line with slope 1 in a μ^c-μ^r-diagram. With the exception of the boundary between CRC 4 and 5 which lies on the fifth reference portfolio the straight line goes through the center of the line segment of two neighboring reference portfolios. For the different combinations of accumulation phases T of 12, 20, 30, and 40 years and premium payment, regular or single, the CRC boundaries differ. Figure 4.1 illustrates one combination.

By the choice of slope 1 of the CRC boundaries, each CRC boundary can be specified uniquely as a point on the straight line. For this, we choose the μ^c-intercept which is the chance measure of the specific boundary at a risk measure of zero. The μ^c-intercept of the boundary between CRC j and $j+1$, where $j = 1, 2, 3, 4$, for the accumulation phase T is denoted by b_T^j in the following. One pension product with

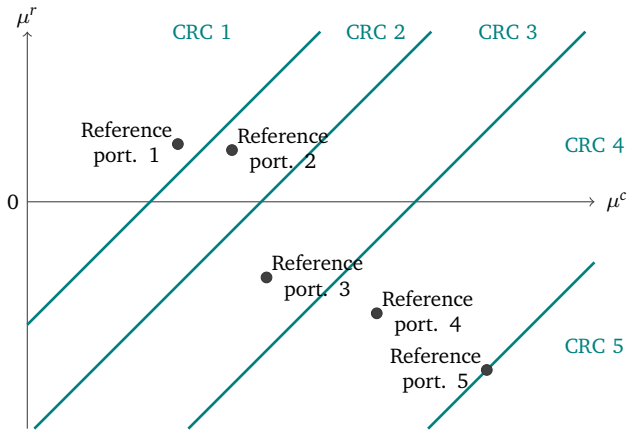


Figure 4.1: Illustration of reference portfolios and CRC boundaries

μ^c and μ^r can be then classified by the difference $\mu^c - \mu^r$. The interval between two successive b_T^j in which the difference lies determines the CRC.

4.2 Properties of the Chance and Risk Parameters

We examine the properties of the parameters and their mappings mentioned in Section 4.1.1. $\overline{V}^c$ and μ^c as well as $\overline{V}^r$ and μ^r quantify the return of a product, not the loss as is usual with classical risk measures. Therefore, μ^r is no risk measure in the classical sense. The name risk measure is misleading. Nevertheless, we will keep the term risk measure as this is the designation on the PIB.

The averaged final contract values $\overline{V}^c$ and $\overline{V}^r$ are non-negative due to $v^k \geqslant 0$. Therefore, μ^c and μ^r are larger than -12. Obviously, the mean of the final contract values is always larger than the mean of the 2,000 lowest final contract values: $\overline{V}^c > \overline{V}^r$. The same holds for the chance and risk measure: $\mu^c > \mu^r$. The chance and risk parameters only coincide if the final contract values v^k are equal in all 10,000 simulations. Furthermore, a larger risk parameter $\overline{V}^r$ or μ^r is better than a lower one because they do not quantify the loss of a pension product but the profit. Pension products with a money-back guarantee have a non-negative chance and risk measure.

We define the following mathematical framework for the analysis of the mappings. We denote the set of pension products with a simulated accumulation phase of T by $\mathcal{V}_T$. A pension product is defined by the mapping $V : \Omega \to \mathbb{R}_0^+$ where Ω is a fixed set of scenarios and $V(\omega)$ is the final contract value at the end of the accumulation phase T of the realization $\omega \in \Omega$. $\overline{V}^c$ is the result of the mapping $f^c : \mathcal{V}_T \to \mathbb{R}_0^+$ with $f^c(V) = \mathbb{E}(V)$ and $\overline{V}^r$ the result of $f^r : \mathcal{V}_T \to \mathbb{R}_0^+$ with $f^r(V) = \mathbb{E}(V|V \leqslant Q_{0,2})$. μ^c or μ^r is determined by the mapping $g_T \circ f^c : \mathcal{V}_T \to [-12; \infty)$ or $g_T \circ f^r : \mathcal{V}_T \to [-12; \infty)$ where $g_T : \mathbb{R}_0^+ \to [-12; \infty)$. Mathematically, the mapping f^c, f^r, $g_T \circ f^c$, and $g_T \circ f^r$ are risk measures analyzed in the following. The mappings are illustrated in Figure 4.2.

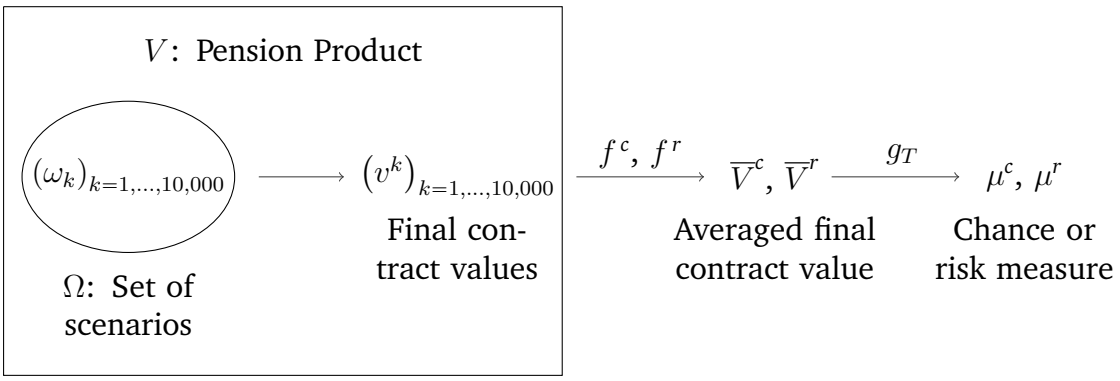


Figure 4.2: Illustration of mappings and notations

If there were no costs and the entire premium were invested directly in the corresponding asset, the pension product with a multiple of the premium would result in the same multiple of the final contract values. Linearity of the final contract values in the premium would be also given if all costs were proportional to the premium or the contract value. However, since fixed costs, e. g. custody costs or unit costs, exist, the assumption of linearity is only approximately fulfilled. The higher the premium, the lower the influence of fixed costs so that we are again closer to linearity. To simplify, we assume in the following that the costs are linear in the premium payment and the contract value. This results in the linearity of the final contract values in the premium.

For analyzing the mappings of the chance and risk parameters, we adapt some properties of classical risk measures to the profit view of the PIA and define them as follows.

Definition 4.1. Let $\mathcal{V}_T$ be a set of pension products with a simulated accumulation phase T. A mapping $\varrho : \mathcal{V}_T \to \mathbb{R}$ is

- **monotonic** if for $X, Y \in \mathcal{V}_T$ with $X \leqslant Y$ it holds

$$\varrho(X) \leqslant \varrho(Y),$$

- **cash invariant** if for $X \in \mathcal{V}_T$ and $m \in \mathbb{R}$ it holds

$$\varrho(X + m) = \varrho(X) + m,$$

- **concave** if for $X, Y \in \mathcal{V}_T$ and $0 \leqslant \lambda \leqslant 1$ it holds

$$\varrho(\lambda X + (1 - \lambda)Y) \geqslant \lambda\varrho(X) + (1 - \lambda)\varrho(Y),$$

- **positive homogeneous** if for $X \in \mathcal{V}_T$ and $\lambda \geqslant 0$ it holds

$$\varrho(\lambda X) = \lambda\varrho(X).$$

The meaning of monotonicity is clear: If the payoff of a pension product is larger than another one, the measure ϱ increases. The property of cash invariance, also called translation invariance or translation property, is given if by adding a constant amount m to X the risk will be reduced by the same amount. For diversification the property of concavity is important. Consider X and Y as different pension products which define the final contract values for a constant premium payment under the assumption of linearity. If the (monthly) available wealth for retirement is divided with a fixed proportion λ between both products, which means *diversifying*, a final capital of $\lambda X + (1 - \lambda)Y$ will be obtained. The risk should not increase by splitting wealth in different investments. This is guaranteed by concavity and depicts the idea of diversification. If the condition of concavity as equality is fulfilled, then the measure is linear.

Lemma 4.2. *Let $f^c : \mathcal{V}_T \to \mathbb{R}_0^+$ be defined as in Equation* (4.1) *and $f^r : \mathcal{V}_T \to \mathbb{R}_0^+$ as in Equation* (4.2)*. f^c has the following properties*

- *monotonicity,*
- *cash invariance,*
- *linearity (concavity as equality),*
- *positive homogeneity.*

f^r has the following properties

- *monotonicity,*
- *cash invariance,*
- *concavity,*
- *positive homogeneity.*

Proof. The properties of f^c follows from the linearity and monotonicity of the expected value.

f^r is a conditional Value at Risk which is defined by $\mathbb{E}(X|X \leqslant Q_p)$ where X is a random variable and Q_p the p quantile of X. The conditional Value at Risk is a coherent risk measure which is characterized by monotonicity, cash invariance, convexity, and positive homogeneity. For the definition of classical risk measures and their properties, we refer to Föllmer and Schied (2011). Transferring the classical conditional Value at Risk to the profit view instead of loss changes convexity to concavity and the properties monotonicity and cash invariance remain as defined above. □

μ^c and μ^r are determined by mappings $(g_T \circ f^c)(V)$ and $(g_T \circ f^r)(V)$. Here, $g_T(\overline{V})$ is the inverse function of Equation (4.3) depending on the kind of premium payment. Unfortunately, Equation (4.3) cannot be explicitly solved for μ for regular premium payment in contrast to single premium payment. Therefore, we analyze the inverse

function $g_T^{-1} : [-12; \infty) \to \mathbb{R}_0^+$ which corresponds to the right side of Equation (4.3) to obtain properties of $g_T(\overline{V})$. We start with the case of regular premium payment and define $g_T^{-1}(\mu)$ as

$$g_T^{-1}(\mu) = 100 \sum_{k=1}^{12 \cdot T} \left(1 + \frac{\mu}{12}\right)^k.$$

The first and second derivatives are

$$\frac{\partial g_T^{-1}(\mu)}{\partial \mu} = 100 \sum_{k=1}^{12 \cdot T} \frac{k}{12} \left(1 + \frac{\mu}{12}\right)^{k-1},$$

$$\frac{\partial^2 g_T^{-1}(\mu)}{\partial \mu^2} = 100 \sum_{k=1}^{12 \cdot T} \frac{k}{144}(k-1) \left(1 + \frac{\mu}{12}\right)^{k-2}.$$

For single premium payment we obtain

$$g_T^{-1}(\mu) = 1200 \cdot T \left(1 + \frac{\mu}{12}\right)^{12T},$$

$$\frac{\partial g_T^{-1}(\mu)}{\partial \mu} = 1200 \cdot T^2 \left(1 + \frac{\mu}{12}\right)^{12T-1},$$

$$\frac{\partial^2 g_T^{-1}(\mu)}{\partial \mu^2} = 1200 \cdot T^2 \left(T - \frac{1}{12}\right) \left(1 + \frac{\mu}{12}\right)^{12T-2}.$$

Clearly, $g_T^{-1}(\mu)$ is strictly monotonically increasing and convex for $\mu \geqslant -12$ for regular premium payment as well as single premium payment. Hence, $g_T(\overline{V})$ is strictly monotonically increasing and concave for $\overline{V} \geqslant 0$. Furthermore, $g_T^{-1}(\mu)$ is continuous. Due to the strict monotonicity of $g_T^{-1}(\mu)$ for $\mu \geqslant -12$, $g_T(\overline{V})$ is also continuous for $\overline{V} \geqslant 0$. Figure 4.3 illustrates the transformation function $g_T(\overline{V})$ for $\overline{V} \geqslant 0$ to sum of paid premiums and its properties for the different kinds of premium payment and accumulation phases T. In this figure the curves of the different accumulation phases intersect in one point since the x-axis shows the averaged final contract value divided by the sum of paid premiums. A value of one on the x-axis corresponds to an averaged final value of the paid premium sum which in turn corresponds to a constant interest rate of zero for all accumulation phases.

Due to the properties of g_T, the properties of the mappings $g_T \circ f^c$ and $g_T \circ f^r$ can be determined and are summarized in the following Lemma.

Lemma 4.3. *Let $g_T : \mathbb{R}_0^+ \to [-12; \infty)$ be defined as inverse function of Equation* (4.3)*, $f^c : \mathcal{V}_T \to \mathbb{R}_0^+$ as in Equation* (4.1)*, and $f^r : \mathcal{V}_T \to \mathbb{R}_0^+$ as in Equation* (4.2)*. $g_T \circ f^c$ and $g_T \circ f^r$ are monotonic and concave.*

Proof. We define f as f^c or f^r. Let $X, Y \in \mathcal{V}_T$ and $X \leqslant Y$. Due to Lemma 4.2, f is monotonic and we have

$$f(X) \leqslant f(Y).$$

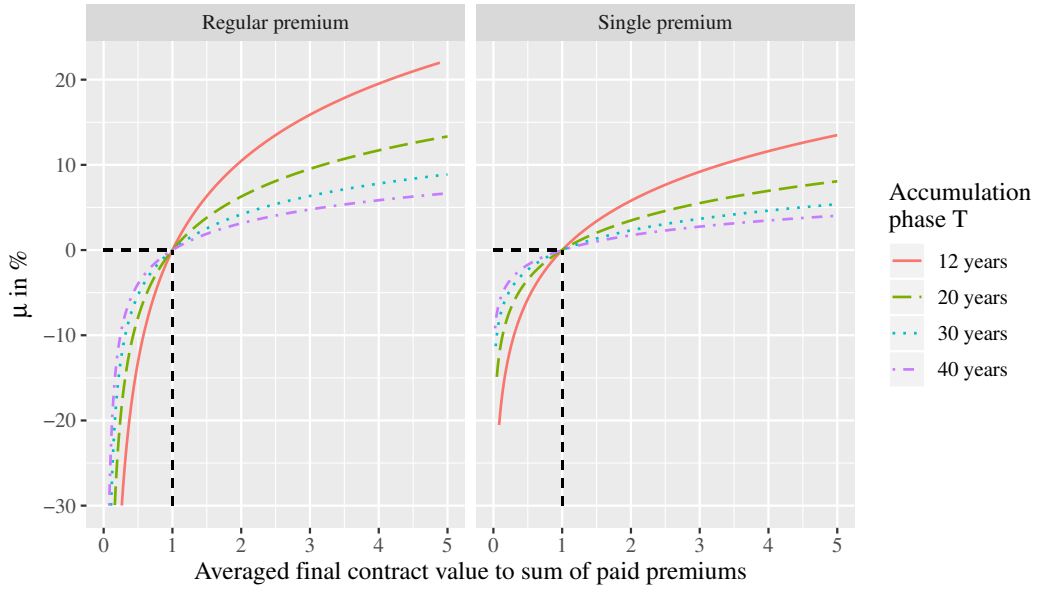


Figure 4.3: Illustration of function $g_T(\overline{V})$ for the different kinds of premium payment

As g_T increases strictly monotonically, we obtain

$$(g_T \circ f)(X) = g_T(f(X)) \leqslant g_T(f(Y)) = (g_T \circ f)(Y).$$

The chance and risk measure is also monotonic.

Let $X, Y \in \mathcal{V}_T$ and $0 \leqslant \lambda \leqslant 1$. Depending on the considered measure, f is linear or concave. In both cases, we have

$$f(\lambda X + (1-\lambda)Y) \geqslant \lambda f(X) + (1-\lambda) f(Y)$$

which is an equality in the case $f = f^c$. Due to the monotonicity and concavity of g_T, we have

$$\begin{aligned}(g_T \circ f)(\lambda X + (1 - \lambda Y)) &= g_T(f(\lambda X + (1-\lambda)Y)) \geqslant g_T(\lambda f(X) + (1-\lambda) f(Y)) \\ &\geqslant \lambda g_T(f(X)) + (1-\lambda) g_T(f(Y)) \\ &= \lambda (g_T \circ f)(X) + (1-\lambda)(g_T \circ f)(Y).\end{aligned}$$

Consequently, $g_T \circ f^c$ and $g_T \circ f^r$ are concave. □

Due to the non-linearity of g_T, $g_T \circ f^c$ and $g_T \circ f^r$ are neither cash invariant nor positive homogeneous.

4.3 Chance-Risk Class of a Portfolio and the Consequences for Customer Consulting

The determination of the CRC of a portfolio is of interest for insurance companies, especially for customer consulting of pension products. For consulting with respect to pension products, the European Institute for Quality Management for Financial Products and Methods (EI-QFM) in Kaiserslautern developed an algorithm to assign a customer also a CRC (for details see Korn and Andelfinger (2016)). Based on this, a pension product with a higher CRC than the customer's should not be sold to the customer, only pension products with the same or a lower CRC. Therefore, only the latter are suitable products. This seems to be common sense in the German insurance industry. Hence, the customer has only pension products with the same or lower risk than his own risk preferences. Furthermore, this standard is very easy and practicable. However, it does not take pension products or other assets which the customer already owns into account and thus might ignore possible diversification effects. This leads to the question of the existence of a diversification effect in the classification and how to determine the CRC of a portfolio. Moreover, it should be possible to determine the CRC of a portfolio (at least in the sense of an upper bound) during the consultation. For this, we examine the classification by PIA for diversification effects first. Based on these results, we take a closer look at determining the CRC of a portfolio. Here, we consider a portfolio consisting of two new pension products and then one of a pension product that has been running for t months and a new one. Since all of these questions arise in practice and are of great interest to the insurance industry, we will derive some practical implications based on our theoretical findings.

We consider two pension products with the same accumulation phase T and kind of premium payment. $V^i \in \mathcal{V}_T$, $i = 1, 2$, has a CRC of CRC^i. It is specified by the pairs $(\overline{V}^{c,i}, \overline{V}^{r,i})$ and $(\mu^{c,i}, \mu^{r,i})$ which results from the classification. Besides the CRC, μ^c and μ^r of both pension products are known and denoted by $\mu^{c,i}$ and $\mu^{r,i}$, $i = 1, 2$. The risk profile of the customer is described by CRC^{cust}. Furthermore, the CRC boundaries are known by b_T^j where $j = 1, 2, 3, 4$ and $T = 12, 20, 30, 40$.

We assume that $CRC^1 \leqslant CRC^2$. This is ensured if the chance and risk measure of the pension products are in the following relations:

(i) $\mu^{c,1} < \mu^{c,2}$ and $\mu^{r,1} > \mu^{r,2}$,

(ii) $\mu^{c,1} \leqslant \mu^{c,2}$ and $\mu^{r,1} \leqslant \mu^{r,2}$,

(iii) $\mu^{c,1} \geqslant \mu^{c,2}$ and $\mu^{r,1} \geqslant \mu^{r,2}$.

Figure 4.4 illustrates the different cases. In the following we focus on the first case since it corresponds to the best practice of consulting which is described in Weber (2009). According to this a customer should be consulted optimally, see Thesis 14 in Section 3 of Weber (2009). This means that products should only be recommended if they are not dominated by another product in the earnings potential as well as in the risk. This is only given in the first case. Therefore, we restrict our investigation

to this area. In the second and third case, one pension product is dominated by the other one. All money should be invested in the dominating pension product.

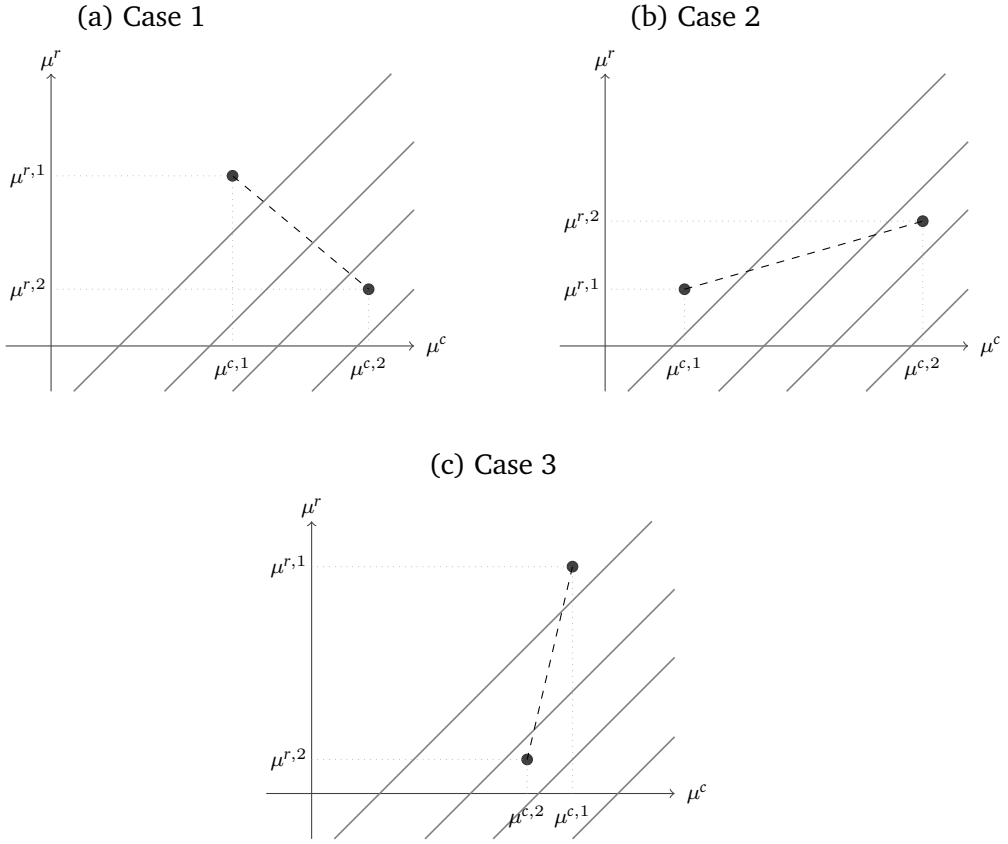


Figure 4.4: Illustration of the different relations of the chance and risk measure of each pension product such that $CRC^1 < CRC^2$

In the following a portfolio of two pension products is considered. This is determined by the distribution of the investment in each pension product. We denote the proportion invested in pension product V^1 by α with $0 \leqslant \alpha \leqslant 1$. The remaining proportion $(1-\alpha)$ is invested in pension product V^2. The investment can be a monthly as well as a single premium payment. We also consider a total investment of the premium payment of the idealized customer since for $\alpha = 1$ or $\alpha = 0$ the averaged final contract values $\overline{V}^{c,i}$ and $\overline{V}^{r,i}$ as well as the chance and risk measures $\mu^{c,i}$ and $\mu^{r,i}$, $i = 1, 2$, result. Under the assumption of linearity of the costs in the premium payment and the contract value, see Section 4.2, the allocation of the investment in two pension products is equivalent to the same allocation of the final contract values. Thus, the portfolio is defined by $\alpha V^1 + (1-\alpha)V^2$. The results can be transferred one-to-one to the same allocation of the premium payments.

4.3.1 Diversification Effect

In the last section, we examined the properties of the chance and risk parameters of PIA. We use these results to examine the classification for a diversification effect.

Due to the properties of PIA's chance and risk parameters which we showed in Section 4.2, the averaged final contract values and the chance and risk measure of a portfolio of two pension products can be estimated by each chance and risk parameters of the pension products. The results are summarized in the following Lemma.

Lemma 4.4. *Let $V^1, V^2 \in \mathcal{V}_T$ be two pension products with the same accumulation phase T. Let $\overline{V}^{c,i}$, $\overline{V}^{r,i}$ be defined as in Equation* (4.1) *and* (4.2) *and $\mu^{c,i}$, $\mu^{r,i}$ as in Equation* (4.3) *depending on the kind of premium payment for $i = 1, 2$. Let g_T be the inverse function of Equation* (4.3)*. Let the portfolio with fraction $0 \leqslant \alpha \leqslant 1$ be defined via $\alpha V^1 + (1-\alpha)V^2$. Let the interpolated averaged final contract values be defined as $\overline{V}^{j,int(\alpha)} := \alpha\overline{V}^{j,1} + (1-\alpha)\overline{V}^{j,2}$ and the interpolated measures as $\mu^{j,int(\alpha)} := \alpha\mu^{j,1} + (1-\alpha)\mu^{j,2}$, $j = c, r$.*

a) For the portfolio, we have

$$\begin{aligned} \overline{V}^{c,ptf(\alpha)} &= \overline{V}^{c,int(\alpha)}, \quad & \overline{V}^{r,ptf(\alpha)} &\geqslant \overline{V}^{r,int(\alpha)}, \\ \mu^{c,ptf(\alpha)} &\geqslant \mu^{c,int(\alpha)}, \quad & \mu^{r,ptf(\alpha)} &\geqslant \mu^{r,int(\alpha)}. \end{aligned}$$

b) Furthermore, we have

$$\begin{aligned} \overline{V}^{c,ptf(\alpha)} &\geqslant \overline{V}^{r,ptf(\alpha)}, \\ \mu^{c,ptf(\alpha)} &\geqslant \mu^{r,ptf(\alpha)}. \end{aligned}$$

c) For the chance measure of the portfolio, it holds

$$\mu^{c,ptf(\alpha)} \leqslant max\left(\mu^{c,1}; \mu^{c,2}\right).$$

Proof. Clearly, a) follows due to linearity of f^c and concavity of f^r, $g_T \circ f^c$, and $g_T \circ f^r$. b) results by the definition of $\overline{V}^c$ and $\overline{V}^r$ as well as μ^c and μ^r.
For c) we consider the calculation of $\mu^{c,ptf(\alpha)}$ by

$$\mu^{c,ptf(\alpha)} = g_T\Big(\underbrace{g_T^{-1}\left(\mu^{c,ptf(\alpha)}\right)}_{=\overline{V}^{c,ptf(\alpha)}}\Big) = g_T\Big(\alpha\underbrace{g_T^{-1}\left(\mu^{c,1}\right)}_{=\overline{V}^{c,1}} + (1-\alpha)\underbrace{g_T^{-1}\left(\mu^{c,2}\right)}_{=\overline{V}^{c,2}}\Big) = g_T\Big(\overline{V}^{c,int(\alpha)}\Big).$$

Taking the derivative of $\mu^{c,ptf(\alpha)}$ with respect to α results in

$$\frac{\partial \mu^{c,ptf(\alpha)}}{\partial \alpha} = \frac{\partial g_T\left(\overline{V}^{c,int(\alpha)}\right)}{\partial \overline{V}^{c,int(\alpha)}}\Big(g_T^{-1}\left(\mu^{c,1}\right) - g_T^{-1}\left(\mu^{c,2}\right)\Big).$$

$\mu^{c,ptf(\alpha)}$ increases in α if $\mu^{c,1} > \mu^{c,2}$ or rather $\overline{V}^{c,1} > \overline{V}^{c,2}$ and decreases if $\mu^{c,1} < \mu^{c,2}$ or rather $\overline{V}^{c,1} < \overline{V}^{c,2}$ since g_T and g_T^{-1} are strictly monotonically increasing. Consequently, it holds

$$\mu^{c,ptf(\alpha)} \leqslant max\left(\mu^{c,1}; \mu^{c,2}\right).$$

□

Unfortunately unlike the mean of the final contract values, the chance measure of a portfolio is not linear in the single chance measure of the pension products. Therefore, we use the linearity of the mean of the final contract values to restrict $\mu^{c,ptf(\alpha)}$, cf. proof of Lemma 4.4 c).

The linear combination of $\overline{V}^r$ and μ^r of two pension products represents the upper limit of the portfolio's risk. Furthermore, μ^c of the portfolio is bounded from below by the interpolation of the chance measure of each pension product. The risk is reduced, while the return is constant or increasing depending on the considered parameter. This is a classical diversification effect which exists due to the chance and risk definition of the classification by PIA.

Figure 4.5 illustrates Lemma 4.4. Figure 4.5a is based on the averaged final contract values $\overline{V}^c$ and $\overline{V}^r$, while Figure 4.5b shows the chance measure μ^c and risk measure μ^r. The dashed black line is the interpolation of the corresponding chance and risk parameters of two pension products for all values of $\alpha \in [0;1]$. The dashed green one represents one possible realization of $\overline{V}^{c,ptf(\alpha)}$ and $\overline{V}^{r,ptf(\alpha)}$ or $\mu^{c,ptf(\alpha)}$ and $\mu^{r,ptf(\alpha)}$ of the portfolio for all values of $\alpha \in [0;1]$. We consider one linear combination of the two pension products specified by one value of α. In Figure 4.5a the interpolated $\overline{V}^{c,int(\alpha)}$ and $\overline{V}^{r,int(\alpha)}$ combination is vertically shifted upwards to obtain the actual $\overline{V}^{c,ptf(\alpha)}$ and $\overline{V}^{r,ptf(\alpha)}$. The upward estimation of $\overline{V}^{r,ptf(\alpha)}$ by $\overline{V}^{c,int(\alpha)}$ on the y-axis results from b) in combination with a) of Lemma 4.4. Figure 4.5b illustrates the relation between the interpolation of the chance and risk measures of the pension products and of the portfolio. All possible actual chance and risk measures of the portfolio with specific proportion α lie on the green line between the two arrows. $\mu^{r,ptf(\alpha)}$ is estimated upward by $\mu^{c,2}$ on the y-axis which results from b) and c) of Lemma 4.4.

Besides the interpolation of only the averaged final contract values or only the chance and risk measures, the interpolation of a combination of these parameters can be considered as well. In Figure 4.6 we put the different possibilities of interpolation of the parameters and their resulting μ^c and μ^r in relation to each other. Each point represents one α. The red line is the actual chance and risk measure of the portfolio. It results from interpolation of the final contract values of each simulation run. The green one is the interpolation of the chance measure μ^c and risk measure μ^r of each pension product. Besides these two lines, the interpolation of $\overline{V}^c$ and μ^r, the turquoise line, is drawn as well as the interpolation of both $\overline{V}^c$ and $\overline{V}^r$, the purple one. The measure of the interpolated values of $\overline{V}^c$ is larger than the interpolated μ^c. This follows from the relation

$$\mu^{c,ptf(\alpha)} = g_T\Big(\alpha\overline{V}^{c,1} + (1-\alpha)\overline{V}^{c,2}\Big) \geqslant \alpha\mu^{c,1} + (1-\alpha)\mu^{c,2}.$$

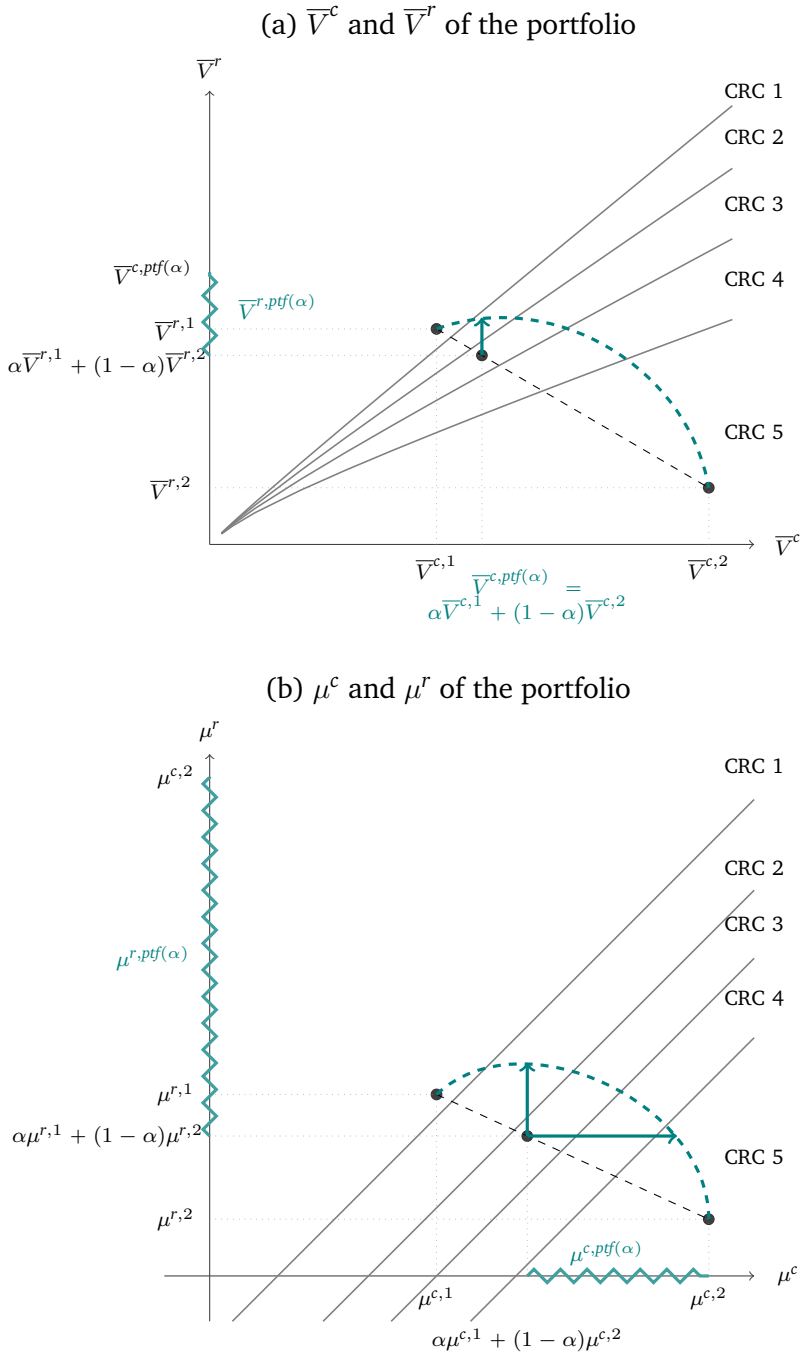


Figure 4.5: Estimation of the portfolio chance and risk parameters via their interpolation

Therefore, the turquoise line lies always right of the green line in a μ^c-μ^r diagram. Furthermore, the purple line illustrating the interpolation of the averaged final contract values $\overline{V}^c$ and $\overline{V}^r$ lies above the turquoise one. This always holds due to the

concavity of g_T

$$\mu^{r,ptf(\alpha)} \geqslant g_T\Big(\alpha \underbrace{g_T^{-1}\left(\mu^{r,1}\right)}_{=\overline{V}^{r,1}} + (1-\alpha)\underbrace{g_T^{-1}\left(\mu^{r,2}\right)}_{=\overline{V}^{r,2}}\Big) \geqslant \alpha\mu^{r,1} + (1-\alpha)\mu^{r,2}.$$

The actual portfolio is above the purple line. The position of the interpolated chance and risk measure, the green line, with respect to both the interpolated averaged final contract values, the purple line, and the actual chance and risk measure of the portfolio, the red one, is inconclusive. For $\mu^{c,1} < \mu^{c,2}$ and $\mu^{r,1} > \mu^{r,2}$, the focused case, the actual chance and risk measure of the portfolio as well as the measures of the interpolated final contract values lie above the interpolation of the measures due to the above estimations. If $\mu^{c,1} < \mu^{c,2}$ and $\mu^{r,1} < \mu^{r,2}$ or $\mu^{c,1} > \mu^{c,2}$ and $\mu^{r,1} > \mu^{r,2}$, the actual values (red) and the measure of the interpolated averaged final contract values (purple) can lie above as well as under the green line. These are the strict dominant cases where one pension product dominates the other (Case 2 and 3 mentioned in Section 4.3).

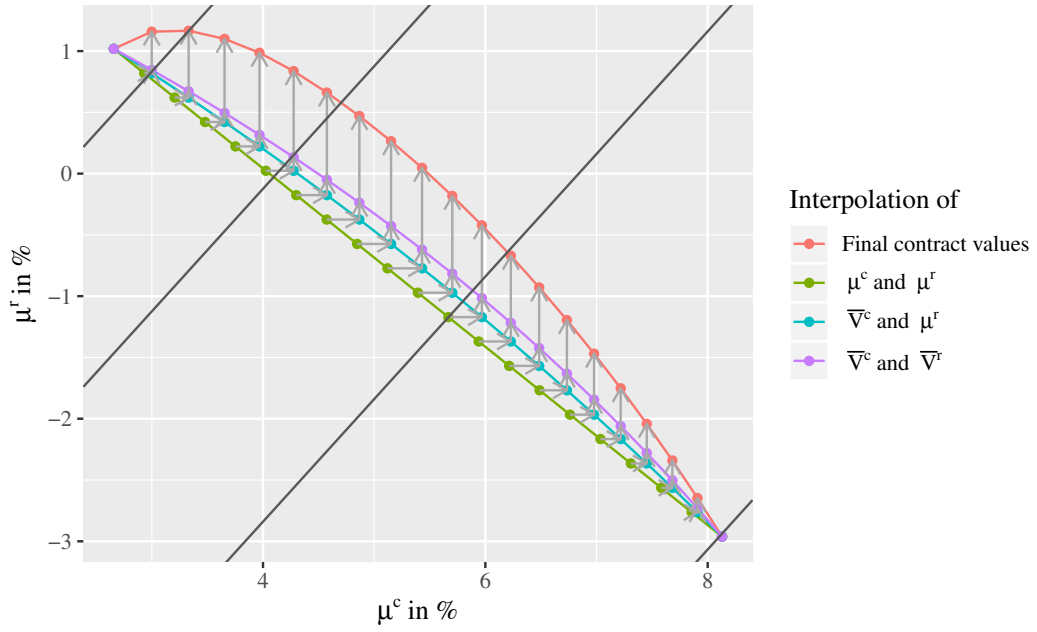


Figure 4.6: Different chance and risk parameter interpolation combinations

In Figure 4.6 pension products with a higher chance and risk measure than the chance and risk measure of pension product V^1 can be produced by a combination of both pension products, see the red line for this. Thus, pension product V^1 is not optimal since it is dominated by different combinations of pension product V^1 and V^2. Furthermore, all combinations of pension products left of the maximum of the red curve are not optimal. They are dominated by products right of the maximum and have a higher chance measure at same risk measure. The combinations

left of the maximum are inefficient, while the combinations right of the maximum are efficient. This reminds of Markowitz's portfolio theory and its efficient and inefficient frontier. With this knowledge the insurance company can create optimal pension products by combining an optimal with a non-optimal pension product. If two pension products are optimal, a combination thereof provides a decreasing line in a chance-risk-measure diagram.

4.3.2 Portfolio of Two New Pension Products

In the last section, we estimated the averaged final contract values as well as the chance and risk measure of a portfolio consisting of different pension products. These parameters determine the CRC. But what is the CRC of the portfolio and can it be restricted by the chance and risk measure of the pension products themselves? Clearly, if the distribution of the final contract values of the combining pension products are known, the CRC of the portfolio is determined by weighted summation of the final contract values of the pension products in every simulation run under the assumption of the (at least approximate) linearity of the costs. Based on these, the chance and risk measure are calculated according to Section 4.1.1. Hence, we obtain the CRC of the portfolio. However, the final contracts values are not available for the insurance consultant or the customer since they are not allowed to be published. The insurance consultant and the customer only know the CRC of the pension product printed on the PIB which is handed over in the customer consultation. Is this information enough to determine the CRC of a portfolio or is more information needed especially for the insurance consultant? This is one question we investigate in this section.

Lemma 4.4 enables us to limit the CRC of the portfolio resulting in the following theorem.

Theorem 4.5. *Let $V^1, V^2 \in \mathcal{V}_T$ be two pension products with $\overline{V}^{c,i}$, $\overline{V}^{r,i}$ defined as in Equation* (4.1) *and* (4.2) *for $i = 1, 2$. Let g_T be the inverse function of Equation* (4.3) *and $0 \leqslant \alpha \leqslant 1$. The CRC of the portfolio of two pension products defined as $\alpha V^1 + (1-\alpha)V^2$ is not greater than the CRC resulting from interpolation of the averaged final contract values $\overline{V}^{c,i}$ and $\overline{V}^{r,i}$ of each pension product with proportion α.*

Proof. Due to Lemma 4.4 it holds that $\overline{V}^{c,ptf(\alpha)}$ equals the interpolation of $\overline{V}^{c,1}$ and $\overline{V}^{c,2}$ with fraction α, while $\overline{V}^{r,ptf(\alpha)}$ is larger than the interpolation of $\overline{V}^{r,1}$ and $\overline{V}^{r,2}$. Therefore, the point $(\overline{V}^{c,ptf(\alpha)}, \overline{V}^{r,ptf(\alpha)})$ lies vertical above $(\overline{V}^{c,int(\alpha)}, \overline{V}^{r,int(\alpha)})$ in a $\overline{V}^c$-$\overline{V}^r$-diagram. The point $(\overline{V}^{c,int(\alpha)}, \overline{V}^{r,int(\alpha)})$ is located in the area above a line l which is the transformation by g_T^{-1} of the next larger CRC boundary. Due to the strict monotonicity of g_T^{-1}, l is also strict monotonic. Thus, the point $(\overline{V}^{c,ptf(\alpha)}, \overline{V}^{r,ptf(\alpha)})$ also lies above the line l because of its position vertically above $(\overline{V}^{c,int(\alpha)}, \overline{V}^{r,int(\alpha)})$. Since g_T is continuous, the transformation by g_T of both points lies in the area above the transformation of the line. Due to they way l is constructed, the transformation by g_T is exactly the CRC

boundary. Consequently, the CRC of the portfolio is not larger than the CRC resulting from the interpolation of $\overline{V}^{c,i}$ and $\overline{V}^{r,i}$. □

Corollary 4.6. *Theorem 4.5 cannot be transferred to the interpolation of the chance and risk measures of each product due to the additional horizontal shift of* $(\overline{V}^{c,ptf(\alpha)}, \overline{V}^{r,ptf(\alpha)})$ *to* $(\overline{V}^{c,int(\alpha)}, \overline{V}^{r,int(\alpha)})$.

We denote the CRC resulting from interpolation of the averaged final contract values by $\mathcal{CRC}$. $\mathcal{CRC}^{ptf(\alpha)}$ is then the CRC of the portfolio defined as $\alpha V^1 + (1-\alpha)V^2$ resulting from the interpolated averaged final contract values of the two pension products. Note that $\mathcal{CRC}^{ptf(\alpha)}$ corresponds to CRC^1 for $\alpha = 1$ and to CRC^2 for $\alpha = 0$.

Theorem 4.5 requires a function g_T which is continuous and invertible. Note that any strictly monotonic and continuous function $\mathbb{R}_0^+ \to D$, where $D \subseteq \mathbb{R}$, satisfies these requirements since the function is invertible due to the strict monotonicity. The inversion function is again strictly monotonic in the same direction and continuous. Consequently, Theorem 4.5 holds for other strictly monotonic and continuous functions than g_T as well.

Clearly, the qualitative criteria of CRC 1 and CRC 2 have to be taken into account. If a pension product does not satisfy them, the CRC of the portfolio cannot be less than CRC 2 or CRC 3. Due to Corollary 4.6, the statement of the theorem does not hold for the CRC resulting from interpolation of the chance and risk measure of each pension product. If only the chance and risk measure of the products are known, they have to be converted into the corresponding averaged final contract values. Instead of the interpolation of the mean of the 2,000 lowest final contract values, the interpolation of the risk measures of each pension product can be used for the estimation of $\mu^{r,ptf(\alpha)}$. The statement of Theorem 4.5 can be transferred to this interpolation. However, the resulting CRC is more inexact than the result which is obtained by interpolation of $\overline{V}^r$. This can be pointed out in Figure 4.6.

Recommendation for Customer Consulting. *The meaning of Theorem 4.5 for the practical application is clear. The CRC of the portfolio can be approximated via interpolation of the averaged final contract values. The use of the chance and risk measure or the CRC itself is not permitted since they can lead to an incorrect consultation. In this case the portfolio can have a greater CRC than the customer's CRC. Thus, the customer takes a higher risk than he prefers.*

Consequently, the knowledge of the CRCs alone is not sufficient. At any rate, the customer consultant has to be provided with the chance and risk measure. However, the knowledge of the averaged final contract values saves the transformation of the chance and risk measure to the averaged final contract values.

Figure 4.5 also shows the relation between the CRC of the interpolated chance and risk parameters and of the portfolio. When considering the averaged final contract values as in Figure 4.5a, the portfolio is in CRC 2 assuming the dashed green line is the actual realization of $\overline{V}^{c,ptf(\alpha)}$ and $\overline{V}^{r,ptf(\alpha)}$. However, lower CRCs are possible which can be obtained by a larger vertical shift upward of the linear combination. Obviously,

a CRC larger than 3, the CRC of the interpolated averaged final contract values of the pension products, cannot be generated. Figure 4.5b considers the pension products' interpolated chance and risk measures. All actually possible chance and risk measures of the portfolio lie on the green line between the two arrows. Obviously, the portfolio can have a lower as well as larger CRC than the CRC of the interpolated chance and risk measures which falls in CRC 3. The CRC of the actual portfolio cannot simply be characterized by the interpolation of the chance and risk measures compared to the interpolation of the averaged final contract values.

Next, we investigate the relationship between the proportion α and the CRC of the corresponding portfolio. We project the point $(\mu^{c,ptf(\alpha)}, \mu^{r,ptf(\alpha)})$ on the μ^c-axis parallel to the CRC boundaries by calculating $\mu^{c,ptf(\alpha)} - \mu^{r,ptf(\alpha)}$ to determine the CRC of a portfolio. Thus, we receive the corresponding $\widehat{\mu}^c = \mu^c - \mu^r$ at $\widehat{\mu}^r = 0$.

Theorem 4.7. *Let $V^1, V^2 \in \mathcal{V}_T$ be two pension products with $\mu^{c,i}$ and $\mu^{r,i}$ calculated by $f^c \circ g_T$ and $f^r \circ g_T$ where f^c is defined as in Equation* (4.1)*, f^r as in Equation* (4.2)*, and g_T as inverse function of Equation* (4.3) *for $i = 1, 2$. Let be $\mu^{c,1} < \mu^{c,2}$, $\mu^{r,1} > \mu^{r,2}$ in such a way that we have $CRC^1 \leqslant CRC^2$. Let CRC^{cust} be a given risk profile with $CRC^1 \leqslant CRC^{cust} \leqslant CRC^2$ and $b_T^{CRC^{cust}}$ the μ^c-intercept of the CRC boundary between CRC^{cust} and $CRC^{cust} + 1$ for the accumulation phase T. Let $u(\alpha)$ be defined as*

$$u(\alpha) := g_T\left(\alpha g_T^{-1}\left(\mu^{c,1}\right) + (1-\alpha) g_T^{-1}\left(\mu^{c,2}\right)\right) - g_T\left(\alpha g_T^{-1}\left(\mu^{r,1}\right) + (1-\alpha) g_T^{-1}\left(\mu^{r,2}\right)\right).$$

Let the premium distribution α^* be chosen such that α^* solves

$$u(\alpha^*) = \begin{cases} u(0), & CRC^{cust} = CRC^2 \\ b_T^{CRC^{cust}}, & \text{else.} \end{cases} \tag{4.4}$$

Then, the CRC of a portfolio of two pension products with proportion $\alpha > \alpha^*$ is not larger than CRC^{cust}.

Proof. We use the linearity of f^c and calculate $\mu^{c,ptf(\alpha)}$ by

$$\mu^{c,ptf(\alpha)} = g_T\left(\alpha g_T^{-1}\left(\mu^{c,1}\right) + (1-\alpha) g_T^{-1}\left(\mu^{c,2}\right)\right) = g_T\left(\overline{V}^{c,int(\alpha)}\right).$$

Using the relation

$$\overline{V}^{r,ptf(\alpha)} \geqslant \alpha g_T^{-1}\left(\mu^{r,1}\right) + (1-\alpha) g_T^{-1}\left(\mu^{r,2}\right) = \overline{V}^{r,int(\alpha)}$$

from Lemma 4.4 in combination with the strictly increasing monotonicity of g_T results in

$$\mu^{c,ptf(\alpha)} - \mu^{r,ptf(\alpha)} \leqslant g_T\left(\overline{V}^{c,int(\alpha)}\right) - g_T\left(\overline{V}^{r,int(\alpha)}\right) = u(\alpha).$$

Taking the derivative of $u(\alpha)$ with respect to α yields

$$\begin{aligned} \frac{\partial u(\alpha)}{\partial \alpha} =& \frac{\partial g_T\left(\overline{V}^{c,int(\alpha)}\right)}{\partial \overline{V}^{c,int(\alpha)}} \left(g_T^{-1}\left(\mu^{c,1}\right) - g_T^{-1}\left(\mu^{c,2}\right)\right) \\ & - \frac{\partial g_T\left(\overline{V}^{r,int(\alpha)}\right)}{\partial \overline{V}^{r,int(\alpha)}} \left(g_T^{-1}\left(\mu^{r,1}\right) - g_T^{-1}\left(\mu^{r,2}\right)\right). \end{aligned}$$

For $\mu^{c,1} < \mu^{c,2}$ the minuend is smaller than zero due to the strict monotonicity of g_T. The subtrahend is non-negative for $\mu^{r,1} > \mu^{r,2}$. Hence, $\frac{\partial u(\alpha)}{\partial \alpha} < 0$ in this case so that $u(\alpha)$ is strictly monotonically decreasing.

Furthermore, it holds

$$\begin{aligned} u(0) &= \mu^{c,2} - \mu^{r,2}, \\ u(1) &= \mu^{c,1} - \mu^{r,1} < b_T^{CRC^1} \end{aligned}$$

where b_T^j is the μ^c-intercept of the CRC boundary between CRC j and CRC $j+1$ and the accumulation phase T. For $CRC^1 < CRC^2$, it holds $u(0) > u(1)$ since $\mu^{c,1} < \mu^{c,2}$ and $\mu^{r,1} > \mu^{r,2}$. For $CRC^{cust} = CRC^2$, $\alpha^* = 0$ solves Equation (4.4). Due to the strict monotonicity of $u(\alpha)$, no further α which solves Equation (4.4) exists. It holds for $CRC^1 \leqslant CRC^{cust} < CRC^2$

$$u(0) \geqslant b_T^{CRC^2-1} \geqslant b_T^{CRC^1} > u(1).$$

Combined with the strict monotonicity of $u(\alpha)$ and the choice of CRC^{cust}, this results in the existence of exactly one α^* solving Equation (4.4).

Due to the strictly decreasing monotonicity of $u(\alpha)$, it holds $u(\alpha) < u(\alpha^*)$ for $\alpha > \alpha^*$. Consequently, for $\mu^{c,1} < \mu^{c,2}$, $\mu^{r,1} > \mu^{r,2}$ and $\alpha > \alpha^*$ we have the estimation

$$\mu^{c,ptf(\alpha)} - \mu^{r,ptf(\alpha)} \leqslant u(\alpha) < u(\alpha^*) = \begin{cases} u(0), & CRC^{cust} = CRC^2 \\ b_T^{CRC^{cust}}, & \text{else.} \end{cases}$$

The CRC of the portfolio of two pension products with proportion $\alpha > \alpha^*$ is not larger than CRC^{cust} in this case. □

Theorem 4.7 requires that g_T is strictly monotonically increasing. g_T can be replaced by any strictly monotonically increasing function $\mathbb{R}_0^+ \to D$, where $D \subseteq \mathbb{R}$, and Theorem 4.7 still holds.

Note that $\alpha^* = 0$ for $CRC^{cust} = CRC^2$. In this case the CRC of the portfolio of two pension products with proportion α is not larger than CRC^2 for all $0 \leqslant \alpha \leqslant 1$. This is in line with the expectation that it is not possible to end up with a CRC that is larger than the maximum CRC of both pension products when two pension products are combined. The qualitative criteria of CRC 1 and 2 are not considered in Theorem 4.7. This must be ensured by the combined pension products. The recommended policy is described in the next paragraph. Moreover, Theorem 4.7 can also be used to combine two pension products with different CRC to a new one with an CRC between.

Recommendation for Customer Consulting. *Theorem 4.7 provides an equation for determining the proportion of wealth to invest in the different pension products. This equation cannot be solved explicitly for α^*. Therefore, α^* has to be numerically determined.*

However, the qualitative criteria must be taken into account. For this, we give the following manageable rules:

- *The customer has a CRC of 1 ($CRC^{cust} = 1$): Only pension products with a money-back guarantee and continuously increasing capital should be combined.*
- *The customer has a CRC of 2 ($CRC^{cust} = 2$): Only pension products with a money-back guarantee should be combined. This is always guaranteed by Riester pension products.*
- *The customer has a larger CRC than 2 ($CRC^{cust} > 2$): All possible pension products can be combined.*

We illustrate Theorem 4.7 on an example with regular premium payment and accumulation phase T of 12 years. Pension product V^1 has CRC 1 with $\mu^{c,1} = 2.6577\ \%$ and $\mu^{r,1} = 1.0184\ \%$. The qualitative criteria of CRC 1 are given. Pension product V^2 is specified by $\mu^{c,2} = 8.1296\ \%$, $\mu^{r,2} = -2.9618\ \%$, and CRC 5. It has neither a money-back guarantee nor a continuously increasing capital. Obviously, the requirements of $\mu^{c,1} < \mu^{c,2}$ and $\mu^{r,1} > \mu^{r,2}$ are fulfilled. CRC 1 and 2 cannot be reached by the portfolio since the money-back guarantee is not contractually secured in total. Nonetheless, we also calculate α^* for a CRC^{cust} of 1 and 2 for illustration. Solving Equation (4.4) for α^* using different values for CRC^{cust} delivers the α^* of Table 4.7.

CRC^{cust}	1	2	3	4	5
α^* in %	94.86	75.18	46.56	0.2697	0

Table 4.7: α^* depending on CRC^{cust} for the example of two new pension products

For different α, we calculate $\mu^{c,ptf(\alpha)}$ and $\mu^{r,ptf(\alpha)}$ as well as the constant interest rate of both interpolated averaged final contract values as an estimation of the chance and risk measure of the portfolio. These parameters and the corresponding CRC are illustrated in Figure 4.8. Figure 4.8a shows the relation between the actual and the estimated CRC. The lower line represents the constant interest rates of the interpolation of the averaged final contract values, while the upper line shows the actual values of the portfolio. Every point of the lines belongs to one α. The same α on the interpolated and actual line are connected by a vertical arrow which also specifies the direction in which the interpolated value is moved obtaining the true value. Note that α goes from 1 to 0 for the direction from pension product V^1 to pension product V^2. The color clarifies in which interval α lies depending on α^* which depends on CRC^{cust}. We denote α^* which belongs to CRC^{cust} equal j, $j = 1, \ldots, 5$, by α_j^*. For the red points, it holds $\alpha > \alpha^*$. The olive points specify $\alpha^* \geqslant \alpha > \alpha_2^*$ and so on. The statement of Theorem 4.7 is clearly seen. For $\alpha > \alpha_j^*$, the actual CRC is not larger than the estimated CRC which is again not larger than CRC^{cust}. This is also clarified in Figure 4.8b which shows the actual and estimated CRC depending on α. The dashed vertical lines are α_j^*, $j = 1, \ldots, 5$. On the left side near α_j^*, a jump in the CRCs can be seen. The CRC of the portfolio according to weighted sum of the final contract values is smaller than the CRC resulting from the interpolation of the averaged final contract values. Before that, the CRCs are equal. Remember that CRC 1 and 2 are

not reachable due to the missing qualitative criteria of the portfolio. They have to be updated to CRC 3. Figure 4.8 shows the calculated values without consideration of the qualitative criteria.

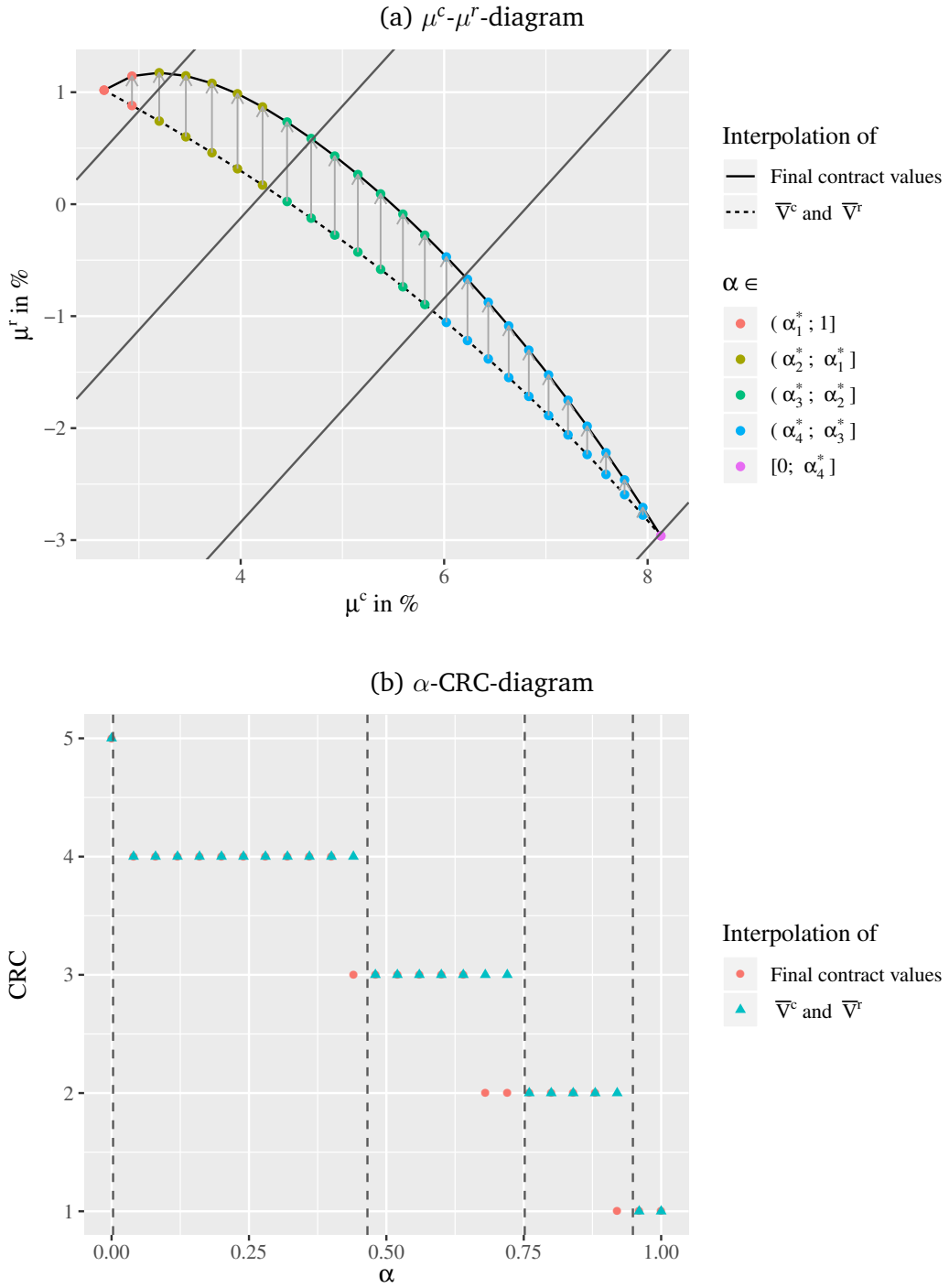


Figure 4.8: Relation between α, $\mathcal{CRC}$, and CRC of the portfolio with two new pension products

For the downward estimation of $\mu^{r,ptf(\alpha)}$ in Theorem 4.7, $\alpha\mu^{r,1} + (1-\alpha)\mu^{r,2}$ can also be used instead of $g_T\Big(\alpha g_T^{-1}\left(\mu^{r,1}\right) + (1-\alpha)g_T^{-1}\left(\mu^{r,2}\right)\Big)$. The so obtained α^* is larger than the one in Theorem 4.7. This is implied by the relation

$$\mu^{r,ptf(\alpha)} \geqslant g_T\Big(\alpha g_T^{-1}\left(\mu^{r,1}\right) + (1-\alpha)g_T^{-1}\left(\mu^{r,2}\right)\Big) \geqslant \alpha\mu^{r,1} + (1-\alpha)\mu^{r,2}$$

due to the concavity of g_T.

4.3.3 Portfolio of an Existing and New Pension Product

In this section, we consider the purchase of a pension product and the CRC determination of the portfolio if the customer already owns a pension product. The approach introduced by us is in accordance with the concept of PIA. The determination of the portfolio CRC is again based on the distribution of the premium payment.

The existing pension product was contracted with an agreed premium of $\mathcal{P}^{old}$ and accumulation phase of $\mathcal{T}^{old}$ t months ago where $t \in \{0, \ldots, 12 \cdot \mathcal{T}^{old}\}$. In addition, we have a new pension product with agreed accumulation phase of $\mathcal{T}^{new}$. The accumulation phases of both pension products end at the same time. This means that the remaining accumulation phase of the existing pension product corresponds to the accumulation phase of the new product: $\mathcal{T}^{new} = \mathcal{T}^{old} - {}^{t}/_{12}$. Furthermore, both pension products have the same kind of premium payment. The classification of the new pension product is based on the accumulation phase T^{new} and premium $P_{T^{new}}$. We assume that the averaged final contract values of the new pension product of the classification, denoted by $\overline{V}^{c,new}$ and $\overline{V}^{r,new}$, as well as the chance measure $\mu^{c,new}$ and risk measure $\mu^{r,new}$ are known besides its CRC. The customer will invest an amount of I where I is either the monthly or the single premium. First, we assume that the entire investment is used for the new pension product. Thus, it holds $\mathcal{P}^{new} = I$.

Determination of the Chance-Risk Class

By Theorem 4.5 the CRC of a portfolio can be estimated from above using the averaged final contract values. The averaged final contract values of the new pension product are known for the accumulation phase T^{new} and the premium $P_{T^{new}}$. However, the current averaged final contract values of the existing pension product have to be calculated. It is defined by the method of determining the evolution of the contract value over the accumulation phase of PIA based on the remaining contract term. Figure 4.9 illustrates the approach of calculating the current averaged final contract values.

The starting point is the current contract value v_t^{old} of the existing pension product which can be taken from the annual information the customer receives from the insurance company. The current contract value has to be scaled to the premium payment $P_{T^{new}}$ of the idealized customer to assign him an artificial but updated CRC. Under the assumption of the linearity of the final contract values in the premium

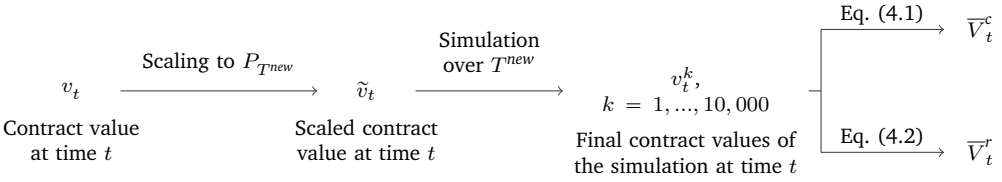


Figure 4.9: Calculation of the averaged final contract values at time t of a existing pension product.

payment, we obtain the scaled contract value by

$$\widetilde{v}_t^{old} = v_t^{old} \cdot \frac{P_{T^{new}}}{\mathcal{P}^{old}}$$

where $\widetilde{v}_t^{old}$ is the contract value at time t scaled to the premium of the idealized customer. The scaling to $P_{T^{new}}$ results from the choice of the function g_T below (see Equation (4.5) and (4.6)). Both sides of the equations are calculated based on the same premium payment. Furthermore, scaling ensures a consistent approach to the classification by PIA. If the contract value is made up of different assets (for example a three-pot hybrid product), all the assets components have to be scaled as above.

Based on $\widetilde{v}_t^{old}$ as starting capital, the final contract values are simulated over the same accumulation phase T^{new} and the same capital market development as the new pension product. The latter includes the knowledge of the current market model parameters and the stochastic scenarios of the PIA model both of which we assume as known. In the simulation 100 Euros are used as regular monthly premium and no further premium payments for single premium payment. Furthermore, the simulation is performed for the contract parameters at contract start with the valid contract costs at their time of withdrawal. The sum of the premiums in the simulation is defined as $100 \cdot (12 \cdot T^{new} + t)$ for regular premium and $1,200 \cdot T^{new}$ for single premium payment according to the assumed premium payments. The single premium payment sum results from the assumption that contract values are scaled to $P_{T^{new}}$. Consequently, costs tied on the sum of the premiums, e.g. the acquisition costs of a classical life insurance pension product, have to be adjusted such that they are calculated on the basis of the premium sum in the current simulation. This ensures that all values are consistently considered. The acquisition costs are usually charged over the first five years according to the zillmerization. These are still to be taken into account in the simulation at $t \leqslant 60$ with an adjusted basis and at the time of withdrawal. As a result, we obtain the simulated final contract values, denoted by v_t^k, $k = 1, \ldots, 10,000$. Based on these, the averaged final contract values $\overline{V}_t^c$ and $\overline{V}_t^r$ are calculated by Equation (4.1) and (4.2).

The CRC of the portfolio can be estimated by the interpolation of the averaged final contract values. The proportion α_{old} of the final contract value of the existing pension product is the same as the one of the total premium payment which is invested in the existing pension product due to the assumption of linearity of the final contract

values in the premium payment. Thus, α_{old} results from

$$\alpha_{old} = \frac{\mathcal{P}^{old}}{\mathcal{P}^{old} + \mathcal{P}^{new}},$$

while

$$1 - \alpha_{old} = \frac{\mathcal{P}^{new}}{\mathcal{P}^{old} + \mathcal{P}^{new}}$$

is the proportion of the final contract values of the new pension product. Then, the interpolated averaged final contract values of the two pension products are obtained by

$$\begin{aligned}\overline{V}^{c,int(\alpha_{old})} &= \alpha_{old}\overline{V}_t^{c,old} + (1-\alpha_{old})\,\overline{V}^{c,new},\\ \overline{V}^{r,int(\alpha_{old})} &= \alpha_{old}\overline{V}_t^{r,old} + (1-\alpha_{old})\,\overline{V}^{r,new}.\end{aligned}$$

Next, the chance and risk measure of the interpolated averaged final contract values have to be determined for the upper bound of the CRC of the portfolio. Calculating them according to Equation (4.3) with T^{new} as T and the corresponding $\overline{V}^{int(\alpha)}$ as $\overline{V}$ leads, generally speaking, to an increased chance and risk measure of the portfolio. The reason for this is the non-consideration of the already paid up premiums of the existing contract. They are assessed as return. Therefore, the function for calculation the chance and risk measure has to be modified. The chance and risk measure are defined as the annual constant interest rate over the accumulation phase T by PIA. At the beginning of the contract, T can be considered either as the entire or the remaining accumulation phase. Consequently, the chance and risk measure can be seen as an annual return over the entire contract term as well as over the remaining one. In the first approach, the chance and risk measure leading to the corresponding interpolated averaged final contract values over the entire accumulation phases are calculated by

$$\overline{V} = \begin{cases} \alpha \cdot 100 \sum_{j=1}^{t} \left(1+\frac{\mu}{12}\right)^{12T+j} + 100 \sum_{j=1}^{12T} \left(1+\frac{\mu}{12}\right)^{j}, & \text{regular premium} \\ 1200 \cdot T \left(\alpha \cdot \left(1+\frac{\mu}{12}\right)^{12T+t} + (1-\alpha) \cdot \left(1+\frac{\mu}{12}\right)^{12T}\right), & \text{single premium.} \end{cases} \tag{4.5}$$

In the second approach, the chance and risk measure as annual constant yield over the remaining contract term solve the equation

$$\overline{V} = \begin{cases} \alpha \cdot \widetilde{v}_t \cdot \left(1+\frac{\mu}{12}\right)^{12T} + 100 \sum_{j=1}^{12T} \left(1+\frac{\mu}{12}\right)^{j}, & \text{regular premium} \\ (\alpha \cdot \widetilde{v}_t + (1-\alpha) \cdot 1200 \cdot T) \cdot \left(1+\frac{\mu}{12}\right)^{12T}, & \text{single premium.} \end{cases} \tag{4.6}$$

The chance and risk measure of the interpolated averaged final contract values are then obtained by solving Equation (4.5) or (4.6) with α_{old} as α, T^{new} as T, the appropriate interpolated averaged final contract value $\overline{V}^{int(\alpha_{old})}$ as $\overline{V}$, and the corresponding μ.

We define the function $g_{T,\alpha}(\overline{V})$ as the inverse function of Equation (4.5) or (4.6) solved for μ. Unfortunately, Equation (4.5) or (4.6) cannot be explicitly solved. Since $g_{T,\alpha}^{-1}(\mu)$ consists of summands which are strictly monotonically increasing and convex for $\mu \geqslant -12$ (cf. Section 4.2), the function $g_{T,\alpha}(\overline{V})$ is again strictly monotonically increasing and concave for $\overline{V} \geqslant 0$. Additionally, $g_{T,\alpha}(\overline{V})$ is continuous in this domain since $g_{T,\alpha}^{-1}(\mu)$ is strictly monotonically increasing and continuous for $\mu \geqslant -12$.

Both approaches differ in the consideration of the performance of the existing pension product in the calculation of the chance and risk measure. In Equation (4.5), the portfolio is classified by both the past performance of the existing pension product as well as the future performance of the portfolio, while the past performance of the existing pension product does not matter in Equation (4.6). Here, only the future performance is decisive for classification. For a balanced performance of both pension products over their entire contract term, the chance and risk measure are similar in both approaches. Normally, the performance of a pension product varies over its contract term. There are phases of high and low returns. Assuming the existing pension product has a tremendous performance on the first t months which leads to a very high contract value at time t and a low performance afterwards. In Equation (4.6) the performance of the last t months dominates the future performance which leads to a low measure, although the overall performance of the existing pension product and thus of the portfolio is tremendous. Conversely, despite a poor performance in the first t months, the measure of Equation (4.6) can be large if the future performance is very good. In both cases, Equation (4.5) leads to a more balanced measure that better reflects the overall performance of each pension product and the portfolio.

Depending on the type of pension product, part or all of the contract value at time t is irrevocably allocated to the customer which means that at least this amount is available at the end of the accumulation phase. From this perspective only the future return of the portfolio is of interest to the customer. Therefore, Equation (4.6) should be chosen for the classification of the portfolio. However, as we discussed before, this approach can misrepresent the existing pension product and its performance. Furthermore, the customer cannot fully dispose of his current contract value that has been earned up to now. If he dissolves his pension product in order to do so, he has to pay a cancellation fee. Consequently, the return generated up to that time point cannot be completely realized. Since pension products are contracted to finance the pension phase, they are usually held until the beginning of the retirement. Therefore, the return over the entire accumulation phase is important to the customer. Due to these arguments, we recommend the calculation of the chance and risk measure according to Equation (4.5) compared to Equation (4.6). Nevertheless, a final specification of the legislator about the valid approach is required.

Based on μ^c and μ^r, the CRC of the portfolio $\alpha_{old} V^{old} + (1 - \alpha_{old})\, V^{new}$ resulting from the interpolation of the averaged final contract values, denoted by $\mathcal{CRC}^{ptf(\alpha_{old})}$, is determined by using the current CRC boundaries of the accumulation phase T^{new}. The qualitative criteria of CRC 1 and 2 are to be taken into account. According to Theorem 4.5, the CRC of the portfolio is not greater than $\mathcal{CRC}^{ptf(\alpha_{old})}$. Therefore,

we obtain an upper bound for the CRC of the portfolio with the above described approach. For $\alpha_{old} = 0$ and $\alpha_{old} = 1$, $\mathcal{CRC}^{ptf(\alpha_{old})}$ of the portfolio is not an estimation from above since the entire capital is invested in a single pension product and the risky averaged final contract value of the pension product is known. The calculation of the estimated CRC for $\alpha_{old} = 0$ results in CRC^{new}. It also corresponds to the classification of PIA.

Recommendation for Customer Consulting. *As in Section 4.3.2, the average final contract values should be used to determine an upper bound of the CRC of the portfolio via interpolation. For the new pension product, its averaged final contract values of the classification can be used if they are known. The final contract values of the existing pension product have to be simulated. For this, the current contract value of the existing pension product has to be known. The chance and risk measure of the portfolio approximated by the interpolation of the averaged final contract values are then calculated by a modified function which takes the already paid premiums into account.*

As an example we consider a classical life insurance pension product which was contracted four years ago ($t = 48$) with an agreed monthly premium of 100 Euros and contract term of 32 years. The pension product has a guaranteed interest rate of $1.25\ \%$ and a current contract value of $3,931.50$ Euros ($v_{48}^{old} = 3,931.50$ Euros). Additionally, a new fund saving plan with a monthly premium of 200 Euros is purchased which invests the premium after costs in a pure equity fund with a volatility of $25\ \%$. The agreed contract term is 28 years and corresponds to the remaining accumulation phase of the existing pension product. The simulated accumulation phase of the classification is 30 years ($T^{new} = 30$). The existing pension product has a money-back guarantee and a continuously increasing capital, while the fund saving plan has not. However, both products have the same contract costs: acquisition costs of $2.5\ \%$ of the premium sum extracted constantly over the first five years and monthly administrative costs of $7\ \%$ of the premium. The fund of the fund saving plan additionally has ongoing charges of $0.3\ \%$.

First, the current averaged final contract values of the classical life insurance pension product have to be calculated. The scaled current contract value is equal to the current contract value since the agreed premium corresponds to P_{30}: $\tilde{v}_{48}^{old} = v_{48}^{old} = 3,931.50$ Euros. Based on $\tilde{v}_{48}^{old}$, we simulate the classical life insurance pension products over 30 years with the current market model parameters of PIA, the current parameters of the reserve fund, a guaranteed interest rate of $1.25\ \%$, and the current valid contract costs. There are still acquisition costs to be taken into account for one year. Basis of these costs are the premium sum which is $40,800$ Euros in the simulation. This is also the minimum final contract value due to the money-back guarantee of the pension product. As current parameters of the reserve fund, we use a proportion of equities of $20\ \%$, a volatility of the equities proportion of $12\ \%$, a duration of the bond proportion of 12 years, costs of the reserve fund of $0.1\ \%$, a current declared total return of $3\ \%$ as well as the return of the equities of the last year of $5\ \%$ and the second to last year of $10\ \%$. Based on the simulated final contract values, we obtain $\overline{V}_{48}^{c,old} = 58,432.80$ Euros and $\overline{V}_{48}^{r,old} = 52,231.78$ Euros.

The current averaged final contract values of the fund saving plan are known due to classification and it holds $\overline{V}^{c,new} = 126,958.29$ Euros as well as $\overline{V}^{r,new} = 36,000$ Euros. The proportion α_{old} of the total premium payments which is invested in the classical life insurance pension product is obtained by

$$\alpha_{old} = \frac{100}{100 + 200} = \frac{1}{3}.$$

Then, the interpolated averaged final contract values with the proportion α_{old} in the existing pension product and the remaining proportion in the fund saving plan are $\overline{V}^{c,int(1/3)} = 104,116.46$ Euros and $\overline{V}^{r,int(1/3)} = 41,410.59$ Euros. With Equation (4.5) for calculating chance and risk measure, we receive $\mu^c = 5.6468\ \%$, $\mu^r = 0.6028\ \%$ and with Equation (4.6) $\mu^c = 5.7768\ \%$, $\mu^r = 0.6588\ \%$. In both cases, the CRC resulting from the interpolation of the averaged final contract values, $\mathcal{CRC}^{ptf(1/3)}$, is 3 according to the current CRC boundaries of 30 years. Consequently, the CRC of the portfolio consisting of the existing and new pension product is not larger than 3. Due to the missing qualitative criteria for CRC 1 and 2, the CRC cannot be smaller than 3.

The Choice of the Proportion α

Assuming the above calculated CRC of the portfolio as an upper bound does not fit the risk profile of the customer meaning $CRC^{cust} < \mathcal{CRC}^{ptf(\alpha)}$. The investment of the entire amount I in the new pension product leads to an unsuitable portfolio's CRC for the customer. Therefore, we investigate in this section how the premium in the new pension product or the proportion α should be chosen such that the customer does not take more risk than he prefers. For this, two different cases are considered:

(i) The premium of the existing pension product can be increased.

(ii) The premium of the existing pension product cannot be increased.

(i) The existing pension product can be raised. We consider the increase in accordance with the contract conditions as well as the purchase of the contract on the same conditions. In case of the latter, the contract still has to be sold. Both variants can cause different costs which are taken into account in the calculation. We treat both kinds of increases as a separate product.

Our considered portfolio in this case consists of three pension products: the existing one, the increase of the existing one, and the new pension product. The entire investment I is used split in both new pension products. The proportion β, $0 \leqslant \beta \leqslant 1$, is used for increasing the existing pension product. The remaining proportion $(1 - \beta)$ is invested in the entirely new pension product. Relative to the total premium payment we obtain the following proportion:

- α_{old} which is invested in the existing pension product and constant

$$\alpha_{old} = \frac{\mathcal{P}^{old}}{\mathcal{P}^{old} + I}$$

- α_{incr} which is taken for increasing the existing pension product

$$\alpha_{incr} = \frac{\beta I}{\mathcal{P}^{old} + I}$$

- α_{new} which is invested in the new pension product

$$\alpha_{new} = \frac{(1-\beta)\, I}{\mathcal{P}^{old} + I} = 1 - \alpha_{old} - \alpha_{incr}.$$

α_{incr} and α_{new} can only take values between 0 and $\frac{I}{\mathcal{P}^{old}+I}$ depending on β. Both add up to $\frac{I}{\mathcal{P}^{old}+I}$.

For the calculation of the CRC of the portfolio with these three products, we need the averaged final contract values. Besides the current averaged final contract values of the existing pension product $\overline{V}_t^{old}$ (see Section 4.3.3) and the one of the entirely new pension product $\overline{V}^{new}$ – these values are known – we need the averaged final contract values of the increase of the existing, denoted by $\overline{V}_0^{c,incr}$ and $\overline{V}_0^{r,incr}$. Here, the different costs are to be taken into account. We specify the averaged final contract values of the increasing by a subscript of 0 to emphasize that it does not have to correspond to the values of the classification but may have to be simulated itself. If the same pension product is purchased again in order to increase the existing pension product, the averaged final contract values of the classification are used as the averaged final contract values of the increasing: $\overline{V}_0^{c,incr} = \overline{V}^{c,old}$ and $\overline{V}_0^{r,incr} = \overline{V}^{r,old}$. We assume that the averaged final contract values of the classification are known. If the increase of the existing pension product according to the contract conditions has the same costs as the new purchase of the same pension product and they are also taken at the same time points in relation to the contract term, the averaged final contract values of the classification are also employed. In addition to the costs, the other contract parameters also have to be the same. If the pension product is not sold anymore, the averaged final contract values have to be simulated as described in Section 4.1.1. The same applies if the costs or other contract parameters of increasing the premium according to the contract conditions differ from the new purchase which is normally the case. The corresponding costs as well as the contract parameters of the increase of the existing pension product are used in the simulation.

Under the assumption of the linearity of the final contract values in the premium, we obtain the following interpolation of the averaged final contract values:

$$\begin{aligned}
\overline{V}^{c,int(\alpha_{old},\alpha_{incr})} &= \alpha_{old}\overline{V}_t^{c,old} + \alpha_{incr}\overline{V}_0^{c,incr} + \alpha_{new}\overline{V}^{c,new}, \\
\overline{V}^{r,int(\alpha_{old},\alpha_{incr})} &= \alpha_{old}\overline{V}_t^{r,old} + \alpha_{incr}\overline{V}_0^{r,incr} + \alpha_{new}\overline{V}^{r,new}.
\end{aligned}$$

We extend Theorem 4.7 to the case of a portfolio of three pension products with one fixed proportion of one pension product.

Theorem 4.8. *Let $g : \mathbb{R}_0^+ \to D$, where $D \subseteq \mathbb{R}$, be a strictly monotonically increasing function calculating the chance and risk measure from the averaged final contract values. Let $V^1 \in \mathcal{V}_T$ be an already t months existing pension product with $\overline{V}_t^{c,1}$ and $\overline{V}_t^{r,1}$ calculated as described in Section 4.3.3 over a simulation phase of T. Let α_1 be the proportion which is invested in V^1 and be fixed. Let $V^2, V^3 \in \mathcal{V}_T$ be two pension products with $\overline{V}^{c,i}, \overline{V}^{r,i}$ defined as in Equation (4.1) and (4.2) for $i = 2, 3$. Let α_i be the proportion which is invested in pension product V^i with $0 \leqslant \alpha_i \leqslant (1 - \alpha_1)$ and $\alpha_2 + \alpha_3 = 1 - \alpha_1$. Let $\overline{V}^{c,2} < \overline{V}^{c,3}$ and $\overline{V}^{r,2} > \overline{V}^{r,3}$ such that $\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)} \leqslant \mathcal{CRC}^{ptf(\alpha_1,0)}$ where $\mathcal{CRC}^{ptf(\alpha_1,\alpha_2)}$ is the estimated $\mathcal{CRC}$ of the portfolio $\alpha_1 V^1 + \alpha_2 V^2 + (1-\alpha_1-\alpha_2)V^3$. Let CRC^{cust} be a given risk profile with $\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)} \leqslant CRC^{cust} \leqslant \mathcal{CRC}^{ptf(\alpha_1,0)}$ and $b_T^{CRC^{cust}}$ the μ^c-intercept of the current boundary between CRC^{cust} and $CRC^{cust}+1$ for the accumulation phase T. Let $u(\alpha_1, \alpha_2)$ be defined as*

$$u(\alpha_1,\alpha_2) := g\Big(\alpha_1\overline{V}_t^{c,1} + \alpha_2\overline{V}^{c,2} + (1-\alpha_1-\alpha_2)\overline{V}^{c,3}\Big) - g\Big(\alpha_1\overline{V}_t^{r,1} + \alpha_2\overline{V}^{r,2} + (1-\alpha_1-\alpha_2)\overline{V}^{r,3}\Big)$$

Let α_2^* be chosen such that α_2^* solves

$$u(\alpha_1,\alpha_2^*) = \begin{cases} u(\alpha_1,0), & CRC^{cust} = \mathcal{CRC}^{\alpha_1,0} \\ \\ b_T^{CRC^{cust}}, & \textit{else.} \end{cases} \tag{4.7}$$

Then, the CRC of the portfolio of $\alpha_1 V^1 + \alpha_2 V^2 + (1 - \alpha_1 - \alpha_2)V^3$ with $\alpha_2 > \alpha_2^*$ is not larger than CRC^{cust}.

Proof. Analogous to proof of Theorem 4.7, we use the linearity of f^c for calculation of $\mu^{c,ptf(\alpha_1,\alpha_2)}$ as

$$\mu^{c,ptf(\alpha_1,\alpha_2)} = g\Big(\alpha_1\overline{V}_t^{c,1} + \alpha_2\overline{V}^{c,2} + (1-\alpha_1-\alpha_2)\overline{V}^{c,3}\Big) = g\Big(\overline{V}^{c,int(\alpha_1,\alpha_2)}\Big),$$

the estimation of

$$\mu^{r,ptf(\alpha_1,\alpha_2)} \geqslant g\Big(\alpha_1\overline{V}_t^{r,1} + \alpha_2\overline{V}^{r,2} + (1-\alpha_1-\alpha_2)\overline{V}^{r,3}\Big) = g\Big(\overline{V}^{r,int(\alpha_1,\alpha_2)}\Big),$$

and the strictly increasing monotonicity of g to obtain the estimation

$$\mu^{c,ptf(\alpha_1,\alpha_2)} - \mu^{r,ptf(\alpha_1,\alpha_2)} \leqslant g\Big(\overline{V}^{c,int(\alpha_1,\alpha_2)}\Big) - g\Big(\overline{V}^{r,int(\alpha_1,\alpha_2)}\Big) = u(\alpha_1,\alpha_2).$$

Since α_1 is fixed, u depends only on α_2. Taking the derivative of u with respect to α_2 yields

$$\frac{\partial u(\alpha_1,\alpha_2)}{\partial \alpha_2} = \frac{\partial g\left(\overline{V}^{c,int(\alpha_1,\alpha_2)}\right)}{\partial \overline{V}^{c,int(\alpha_1,\alpha_2)}}\Big(\overline{V}^{c,2} - \overline{V}^{c,3}\Big) - \frac{\partial g\left(\overline{V}^{r,int(\alpha_1,\alpha_2)}\right)}{\partial \overline{V}^{r,int(\alpha_1,\alpha_2)}}\Big(\overline{V}^{r,2} - \overline{V}^{r,3}\Big).$$

For $\overline{V}^{c,2} < \overline{V}^{c,3}$ the minuend is smaller than zero due to the strict monotonicity of g. The subtrahend is non-negative for $\overline{V}^{r,2} > \overline{V}^{r,3}$. Hence, $\frac{\partial u(\alpha_1,\alpha_2)}{\partial \alpha_2} < 0$ in this case so that $u(\alpha_1, \alpha_2)$ is strictly monotonically decreasing.

Furthermore, it holds

$$u(\alpha_1, 0) = g\Big(\alpha_1 \overline{V}_t^{c,1} + (1-\alpha_1)\overline{V}^{c,3}\Big) - g\Big(\alpha_1 \overline{V}_t^{r,1} + (1-\alpha_1)\overline{V}^{r,3}\Big),$$
$$u(\alpha_1, 1-\alpha_1) = g\Big(\alpha_1 \overline{V}_t^{c,1} + (1-\alpha_1)\overline{V}^{c,2}\Big) - g\Big(\alpha_1 \overline{V}_t^{r,1} + (1-\alpha_1)\overline{V}^{r,2}\Big) < b_T^{\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)}}$$

where b_T^j is the μ^c-intercept of the CRC boundary between j and $j+1$ for the accumulation phase T. Due to $\overline{V}^{c,2} < \overline{V}^{c,3}$ and $\overline{V}^{r,2} > \overline{V}^{r,3}$, we have $u(\alpha_1, 0) > u(\alpha_1, 1-\alpha_1)$.

For $CRC^{cust} = \mathcal{CRC}^{ptf(\alpha_1,0)}$, $\alpha_2^* = 0$ solves Equation (4.7). Due to the strict monotonicity of $u(\alpha_1, \alpha_2)$ no further α_2 solving Equation (4.7) exists. It holds

$$u(\alpha_1, 0) \geqslant b_T^{\mathcal{CRC}^{ptf(\alpha_1,0)}-1} \geqslant b_T^{\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)}} > u(\alpha_1, 1-\alpha_1)$$

for $\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)} \leqslant CRC^{cust} < \mathcal{CRC}^{ptf(\alpha_1,0)}$. Combined with the strict monotonicity of $u(\alpha_1, \alpha_2)$ and the choice of CRC^{cust} this results in the existence of exactly one α_2^* solving Equation (4.7).

Due to the strictly decreasing monotonicity of $u(\alpha_1, \alpha_2)$, it holds $u(\alpha_1, \alpha_2) < u(\alpha_1, \alpha_2^*)$ for $\alpha_2 > \alpha_2^*$. Consequently, for $\overline{V}^{c,2} < \overline{V}^{c,3}$, $\overline{V}^{r,2} > \overline{V}^{r,3}$ and $\alpha_2 > \alpha_2^*$ we have the estimation

$$\mu^{c,ptf(\alpha_1,\alpha_2)} - \mu^{r,ptf(\alpha_1,\alpha_2)} \leqslant u(\alpha_1, \alpha_2) < u(\alpha_1, \alpha_2^*) = \begin{cases} u(\alpha_1, 0), & CRC^{cust} = \mathcal{CRC}^{\alpha_1,0} \\ b_T^{CRC^{cust}}, & \textit{else.} \end{cases}$$

The CRC of the portfolio $\alpha_1 V^1 + \alpha_2 V^2 + (1-\alpha_1-\alpha_2)V^3$ with $\alpha_2 > \alpha_2^*$ is not larger than CRC^{cust} in this case. □

Note that the portfolio's CRC estimated from above can be attained via Theorem 4.8. If $\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)} = \mathcal{CRC}^{ptf(\alpha_1,0)}$, Theorem 4.8 provides $\alpha_2^* = 0$. No other upper estimated CRC than $\mathcal{CRC}^{ptf(\alpha_1,1-\alpha_1)}$ or $\mathcal{CRC}^{ptf(\alpha_1,0)}$ can be generated via the choice of α_2 in this case.

The new pension product is chosen so that the requirements of Theorem 4.8 are satisfied. As function g for the chance and risk measure, the inverse function of Equation (4.5) or (4.6) of Section 4.3.3 is employed with T^{new} as T and α_{old} as α. Applying Theorem 4.8 delivers α_{incr}^*. All $\alpha_{incr} > \alpha_{incr}^*$ ensure that the CRC of the portfolio is not larger than the desired CRC^{cust}. The corresponding proportion β^* of I can be calculated by

$$\beta^* = \alpha_{incr}^* \frac{\left(\mathcal{P}^{old} + I\right)}{I}.$$

The maximum premium in the new pension product, $\mathcal{P}^{*,new}$, ensuring that the CRC of the portfolio is not larger than CRC^{cust} is calculated by $(1-\beta^*)I$.

Recommendation for Customer Consulting. *If the existing pension product can be increased, a maximum premium to be invested in the increase can be calculated such that the CRC of the portfolio is not larger than a given one. For this, the averaged final contract values of the increase have to be simulated and calculated in the PIA model in addition to the present model of the existing pension product if necessary.*

However, the qualitative criteria must be taken into account. For this, the consultant has to ensure that the existing and new pension product fulfill the qualitative criteria for CRC 1 if the customer has a CRC of 1 ($CRC^{cust} = 1$) and for CRC 2 if the customer has a CRC of 2 ($CRC^{cust} = 2$).

We take the example of the previous section even if CRC 1 and 2 cannot be reached due to the missing qualitative criteria. Nevertheless, we also calculate these cases for illustrative purposes. The existing pension product can be increased incurring the same costs as previously and the acquisition costs being charged over the next 60 months. The simulation of the increase over 30 years results in the averaged final contract values: $\overline{V}_0^{c,incr} = 48,379.26$ Euros and $\overline{V}_0^{r,incr} = 43,655.14$ Euros. The simulation is based on the current market model parameters of PIA, the current parameters of the reserve fund as above, a guaranteed interest rate of 1.25%, and the valid contract costs.

We consider the case where the entire investment of 200 Euros is used for the fund saving plan and the case where it is used for the increase. The CRC of the portfolio of the existing pension product and the fund saving plan is not larger than 3 as calculated in Section 4.3.3: $\mathcal{CRC}^{ptf(1/3,0)} = 3$. The same calculation for the portfolio of the existing pension product and the entire investment of 200 Euros in the increase delivers $\mathcal{CRC}^{ptf(1/3,2/3)} = 1$ for both functions of $g_{T,1/3}(\overline{V})$ and the current CRC boundaries of 30 years. Consequently, $\mathcal{CRC}^{ptf(1/3,2/3)} < \mathcal{CRC}^{ptf(1/3,0)}$ and Theorem 4.8 can be applied. We obtain α^*_{incr} in $\%$ and α^*_{new} in $\%$ listed in Table 4.10 for different CRC^{cust} with $\mathcal{CRC}^{ptf(1/3,2/3)} \leqslant CRC^{cust} \leqslant \mathcal{CRC}^{ptf(1/3,0)}$ besides the constant α_{old}.

	$g_{T,1/3}(\overline{V})$	CRC^{cust} 1	2	3
α_{old} in $\%$		33.33	33.33	33.33
α^*_{incr} in $\%$	Eq. (4.5)	55.40	33.40	0
	Eq. (4.6)	55.66	34.09	0
α^*_{new} in $\%$	Eq. (4.5)	11.26	33.26	66.66
	Eq. (4.6)	11.01	32.58	66.66

Table 4.10: α_{old}, α^*_{incr}, and α^*_{new} depending on CRC^{cust} for the example of an existing and new pension product with the possibility of increasing the existing one

For different $\alpha_{incr} \in [0; 2/3]$ we calculate the actual $\mu^{c,ptf(1/3,\alpha_{incr})}$ and $\mu^{r,ptf(1/3,\alpha_{incr})}$ of the portfolio as well as μ^c and μ^r of the interpolated averaged final contract values $\overline{V}^{c,int(1/3,\alpha_{incr})}$ and $\overline{V}^{r,int(1/3,\alpha_{incr})}$ as estimation of the chance and risk measure of the portfolio. These parameters and the corresponding CRC are illustrated in Figure 4.11 for

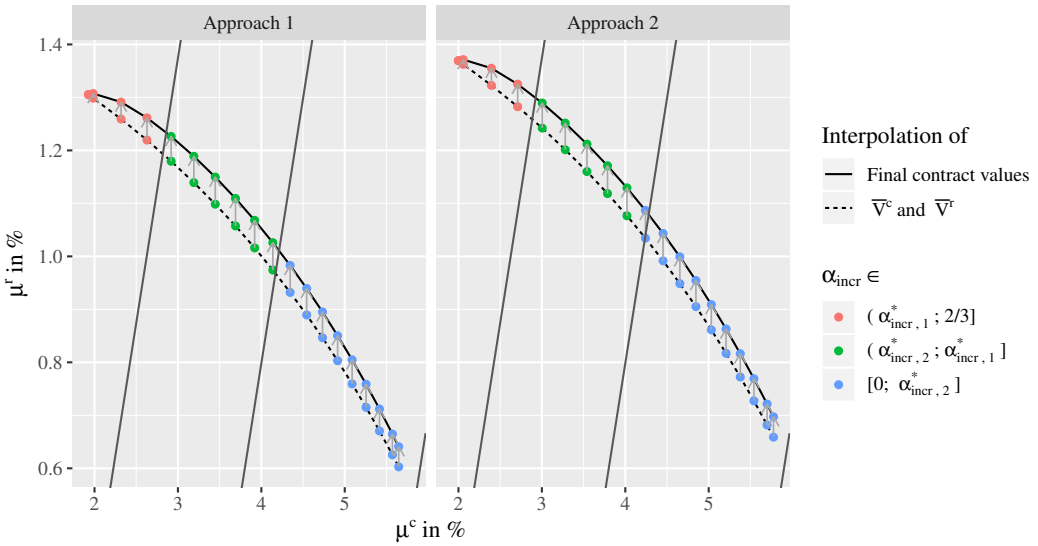


Figure 4.11: Relation between α_{incr}, $\mathcal{CRC}$, and CRC of the portfolio with an existing and new pension product

both functions of $g_{T,1/3}(\overline{V})$. Just like in Figure 4.8a the lower line represents the constant interest rates of the interpolation of the averaged final contract values, while the upper line shows the actual values of the portfolio. Every point of the lines belong to one α_{incr}. The same α_{incr} on the interpolated and actual line are connected by a vertical arrow which also specifies the direction in which the interpolated value is moved to obtain the true values. Note that α_{incr} goes from $2/3$ to 0 from the point at the top left to the point at the bottom right. For $\alpha_{incr} = 2/3$ it looks like the actual chance and risk measure and the one of the interpolation are the same, but they are not albeit with only a very small difference. As for $\alpha_{incr} = 0$ a part of the total premium payment is always invested in the existing pension product, another product with another order of the final contract values. Therefore, the interpolation of the averaged final contract values is an estimation from above of the portfolio. The color clarifies the interval depending on α^*_{incr} which in turn depends on CRC^{cust} wherein α_{incr} lies. We denote α^*_{incr} which belongs to CRC^{cust} equal j, $j = 1, 2, 3$, by $\alpha^*_{incr,j}$. The statement of Theorem 4.8 is clearly seen. For $\alpha_{incr} > \alpha^*_{incr,j}$ the actual CRC is not larger than the CRC resulting from the interpolation of the averaged final contract values which is again not larger than CRC j.

(ii) In the second case the existing contract cannot be increased: Whether the contract conditions of the existing pension product exclude an increase or the contract is not sold anymore. Consequently, the CRC of the portfolio can only be controlled with the investment in the new pension product. This means that not the entire available amount I will be invested in any case. Here, the premium of the new pension product $\mathcal{P}^{new}$ and thus α have to be adjusted until the desired CRC of the customer results

from the calculation in Section 4.3.3. For CRC 1 or 2 as CRC^{cust}, the existing and new pension product have to fulfill the qualitative criteria of CRC 1 or 2.

Recommendation for Customer Consulting. *The CRC of the portfolio has to be estimated according to Section 4.3.3 for every possible premium of the new pension product. The premium of the existing pension product remains the same. The qualitative criteria for CRC 1 and 2 have to be taken into account.*

We consider the same example as before with the knowledge that CRC 1 and 2 cannot actually be reached. Here, the classical life insurance pension product cannot be increased. For different premiums in the fund saving plan with $\mathcal{P}^{new} \in [0; 200]$, we calculate the estimated CRC of the portfolio of the existing and new pension product as described in Section 4.3.3 and obtain the premium $\mathcal{P}^{*,new}$ in Euros for the different CRC^{cust} which are listed in Table 4.12. If the customer has a risk profile of 2, a fund saving plan with a maximum premium of 59.11 Euros or 56.26 Euros should be contracted such that the portfolio does not have a larger CRC than 2. Furthermore, 100 Euros are invested in the classical life insurance pension product.

	$g_{T,\alpha}(\overline{V})$	CRC^{cust}		
		1	2	3
$\mathcal{P}^{*,new}$ in Euros	Eq. (4.5)	15.81	59.11	200
	Eq. (4.6)	14.99	56.26	200

Table 4.12: $\mathcal{P}^{*,new}$ depending on CRC^{cust} for the example of an existing and new pension product without the possibility of increasing the existing one

4.4 Chance-Risk Class of a Pension Product over the Contract Term

The classification by PIA only determines the CRC at the time of purchase. However, the CRC of the pension product will not remain the same over the contract term. Factors such as the past performance of the pension product, the amortization of the acquisition costs, or the current financial market situation might influence the CRC. Unfortunately, there are no legal specifications on the determination of the current CRC of an existing contract already running a certain period of time. The legislator only specifies the framework for the CRC at purchase. We introduce different approaches of determining the CRC over the contract term in this section. Thereby, we apply the legal situation of the classification to this question first. Afterwards, further conceivable procedures are presented. These are in accordance with the concept of the classification by PIA. Additionally, the approaches in this section are consistent with the CRC determination of a portfolio consisting of an existing and new pension product.

We consider an existing pension product contracted at the age of x with agreed accumulation phase $\mathcal{T}$ and premium payment $\mathcal{P}$. Let v_t denote the contract value after t months just before the next premium payment where $t = 0, \ldots, 12 \cdot \mathcal{T}$. Clearly, the current contract value after zero months is zero: $v_0 = 0$. The remaining accumulation phase after t months, denoted by $\mathcal{T}_t$, results from

$$\mathcal{T}_t = \mathcal{T} - \frac{t}{12}.$$

The values of the classification by PIA are calculated at the contract start $t = 0$. We use the above notation of these values without a subscript of 0 in order to emphasize that these values are coming from the classification by PIA. The calculated values for determining the CRC at later time points $t > 0$ are identified by a subscript.

The CRC of the pension product at the beginning of the contract is issued by PIA and printed on the PIB. It is determined for a simulated accumulation phase of

$$T = \begin{cases} 12, & \mathcal{T} \in [0; 12] \\ 20, & \mathcal{T} \in (12; 20] \\ 30, & \mathcal{T} \in (20; 30] \\ 40, & \mathcal{T} \in (30; \infty] \end{cases}$$

according to § 5 (2) sentence 4 of the AltvPIBV. Furthermore, the simulation is based on the assumption of a premium payment of the idealized customer (see Section 4.1.1).

Figure 4.13 shows the chronological sequence of the pension product. In addition, it illustrates the CRC at different times t and the corresponding simulated accumulation phases for this calculation explained below. The simulated accumulation phases T and T_t extend beyond the age of 67 years since the end of both phases do not necessarily end with reaching this age. For example, a contract with $\mathcal{T}$ of 13 years is simulated over an accumulation of 20 years for classification.

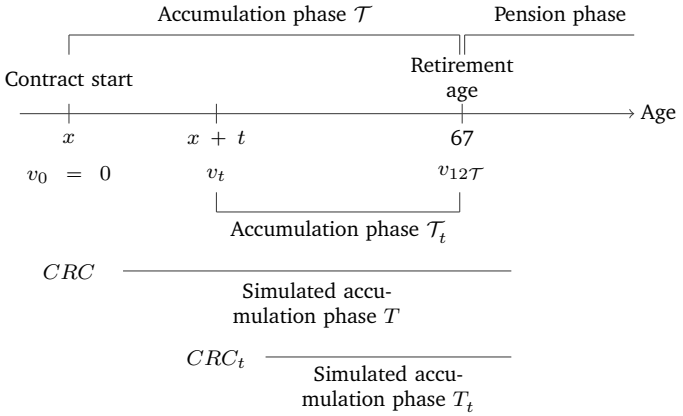


Figure 4.13: Chronological sequence of a pension product

We assume that the pension product already runs t months and we want to determine the current CRC of the pension product denoted by CRC_t. According to the legislator four different accumulation phases are considered in the classification of the pension product. For each of them the CRC is calculated under the terms described above. The pension product receives then the CRC of one of the four accumulation phases corresponding to its own accumulation phase. Transferring this procedure to our issue implies that the running pension product obtains one of the four CRC issued by PIA depending on the remaining accumulation phase $\mathcal{T}_t$. CRC_t results from

$$CRC_t = \begin{cases} CRC \text{ of } T = 12, & \mathcal{T}_t \in [0; 12] \\ CRC \text{ of } T = 20, & \mathcal{T}_t \in (12; 20] \\ CRC \text{ of } T = 30, & \mathcal{T}_t \in (20; 30] \\ CRC \text{ of } T = 40, & \mathcal{T}_t \in (30; \infty] \end{cases}$$

where CRC is the current CRC of the pension product issued by PIA at time t. The CRC can jump if the classification interval changes from one time point to the next, e.g. from $\mathcal{T}_t = 12\frac{1}{12}$ to $\mathcal{T}_t = 12$. However, the current CRC can also change, while two sequential remaining accumulation phases lie in one classification interval since the pension products are checked every year and the capital market parameters are changed.

An advantage of this approach is that no further simulations have to be performed. However, costs included in the classification may differ in each accumulation phase considered by PIA. Therefore, the costs may change when the underlying accumulation phase of the CRC changes over the contract term. Consequently, the costs of the contract could not be correctly taken into account which is problematic. Furthermore, the already paid premiums are not included in the classification. In addition, the CRC has to exist for each of the four accumulation phases which is not necessarily the case. For example, if the minimum offered contract term of the pension product is larger than 12 years or the pension product is no longer sold, then the corresponding CRC of the pension product is not calculated by PIA. These are problems where it is not clear how to deal with them.

Therefore, we develop further approaches which have the disadvantage that additional simulations have to be run, but the correct costs and paid premiums are considered and the existence of a CRC for every remaining accumulation phase is ensured. In order to determine the CRC of an existing pension product, the current averaged final contract values $\overline{V}_t$ of the pension product have to be calculated first which is not straightforward. Based on these values, the running pension product is assigned a CRC via the chance and risk measure.

The calculation of the averaged final contract values is similar to the one in Section 4.3.3 except for the choice of the scaling and of the simulation phase. The basis of the determination of the averaged final contract values and thus of CRC_t is the current contract value v_t of the pension product. This value has to be scaled to the

premium payment of the idealized customer by

$$\widetilde{v}_t = v_t \cdot \frac{P_T}{\mathcal{P}}$$

where $\widetilde{v}_t$ is the contract value scaled to the premium of the idealized customer at time t. Scaling of the contract value ensures a consistent approach to the classification by PIA. As in Section 4.3.3, we assume that the contract value is linear in the premium payment. If the contract value is made up of different assets (for example a three-pot hybrid product), all assets components have to be scaled as above.

Next, the final contract values of the pension product have to be simulated over the remaining accumulation phase T_t in the PIA model under consideration of the contract value $\widetilde{v}_t$ as starting capital and the premium payment P_T. No further premium payments are considered in the simulation in the single premium payment case. The distribution of the contract value in different asset components has to be taken into account. For the simulation of the market model, the capital market developments with the current model parameters of the PIA model, which we assume as known, and asset costs are used, while the contract is simulated on the contract parameters which hold at the contract start. Thus, the current market situation as well as the specific contract conditions are represented. The premium sum in the simulation is defined as $100 \cdot \lfloor 12 \cdot T_t + t \rfloor$ for regular premium and $1,200 \cdot T$ for single premium payment according to the assumed premium payments. Consequently, costs with the premium sum as basis, e.g. the acquisition costs of a classical life insurance pension product, have to be adjusted so that they are calculated on the basis of the premium sum of the current simulation.

The simulated accumulation phase T_t cannot be straightforwardly determined. All of the following approaches are conceivable. For a better understanding, we refer to Table 4.14 which compares the different methods and shows the problems involved.

(i) T_t results by deducting the past contract term t from the accumulation phase T of the classification:

$$T_t = T - \frac{t}{12}.$$

The simulation phase decreases each month by one month over the entire contract term. However, the simulation phase T_t can significantly differ from the actual remaining accumulation phase $\mathcal{T}_t$. This is especially problematic at the end of the accumulation phase of the contract where $T_{12 \cdot \mathcal{T}} = T - \mathcal{T}$. $T_{12 \cdot \mathcal{T}}$ is non-negative for $\mathcal{T} < 40$ and negative for $\mathcal{T} > 40$, while $\mathcal{T}_{12 \cdot \mathcal{T}}$ is equal to zero. No simulation can be performed for a negative simulation phase T_t which is already the case for $t > 12 \cdot T$ and $\mathcal{T} > 40$. The difference between the simulated and the actual remaining accumulation phase remains the same over the contract term and can be up to 9 years for $\mathcal{T} < 40$ and larger for $\mathcal{T} > 40$. It is only zero for $\mathcal{T} = T$ and thus for $\mathcal{T} = 12, 20, 30, 40$.

(ii) T_t results by deduction of the past contract term t from the actual accumulation phase $\mathcal{T}$ of the contract:

$$T_t = \mathcal{T} - \frac{t}{12}.$$

T_t is the actual remaining accumulation phase of the contract in contrast to the first approach. However, there is a gap in the simulated accumulation phase of the classification of PIA, T, and the simulation phase one month later, T_1. It can be up to 9 years for $\mathcal{T} < 40$ and larger for $\mathcal{T} > 40$. It is only zero in the case of $\mathcal{T} = T$ and thus for $\mathcal{T} = 12, 20, 30, 40$.

(iii) The past contract term is adjusted to the simulated accumulation by the proportion of T to $\mathcal{T}$ and deducted from T. T_t results from

$$T_t = \frac{T}{\mathcal{T}} \left(\mathcal{T} - \frac{t}{12} \right).$$

This approach has the advantage that the simulated accumulation phase T_t uniformly decreases from T to zero over the whole accumulation phase $\mathcal{T}$ of the contract. We have no gap between the simulated accumulation phase T and T_1 and it holds that $T_{12 \cdot \mathcal{T}}$ is zero. Furthermore, the approach can deal with $\mathcal{T} < 40$ and $\mathcal{T} > 40$. However, the difference between two successive simulated accumulation phases T_t and T_{t+1} are not equal to one month. This is only the case if $\mathcal{T} = T$ and thus if $\mathcal{T} = 12, 20, 30, 40$. Therefore most of the time, the simulated accumulation phase is not a full month.

	Passed time t in months									
	0	12	24	36	48	$\cdots$	120	132	144	156
$\mathcal{T}_t$	13	12	11	10	9	$\cdots$	3	2	1	0
T_t as (i)	20	19	18	17	16	$\cdots$	10	9	8	7
T_t as (ii)	20	12	11	10	9	$\cdots$	3	2	1	0
T_t as (iii)	20	18.46	16.92	15.38	13.85	$\cdots$	4.62	3.08	1.54	0

Table 4.14: Comparison of the actual remaining accumulation phase$\mathcal{T}_t$ in years to the simulation phase T_t in years based on the different approaches

In order to illustrate the different approaches, we consider a contract with an accumulation phase $\mathcal{T}$ of 13 years. Consequently, the accumulation phase T of the classification is 20 years. Table 4.14 shows the simulated accumulation phase T_t in years for different passed times t in months and the different approaches. The above-mentioned problems of the approaches are highlighted.

Due to the uniform distribution of the accumulation phase over the simulated accumulation phase at the time of purchase, we recommend the third approach. Furthermore, this approach does not lead to any inconsistencies in the CRC determination over the contract term.

After determining T_t, the simulation can be performed obtaining 10,000 final contract values denoted by v_t^k, $k = 1, \ldots, 10,000$. Based on these, the averaged final contract values $\overline{V}_t^c$ and $\overline{V}_t^r$ are calculated by Equation (4.1) and (4.2). In the classification by PIA, the CRC is determined by the chance and risk measure which represent the annual constant interest rate over the accumulation phase T leading to the averaged simulated final contract values. As in Section 4.3.3, T can be considered as the entire accumulation phase as well as the remaining accumulation phase. Therefore, the same possibilities for calculating the chance and risk measure and thus the CRC as above consistent with the PIA classification exist.

(i) The chance and risk measure at time t is the annual constant interest rate over the entire accumulation phase T leading to the current averaged final contract values resulting from the simulation at time t. The chance and risk measure μ_t at time t solve the following equation which corresponds to Equation (4.3) with the appropriate $\overline{V}_t$ as $\overline{V}$ and μ_t as μ:

$$\overline{V}_t = \begin{cases} 100 \sum_{k=1}^{12 \cdot T} \left(1 + \frac{\mu_t}{12}\right)^k, & \text{regular premium} \\ 1200 \cdot T \left(1 + \frac{\mu_t}{12}\right)^{12T}, & \text{single premium.} \end{cases}$$

Based on μ_t^c and μ_t^r, the pension product is assigned to a CRC. Thereby, the CRC of the pension product has to be determined by the current CRC boundaries of the corresponding accumulation phase T since the pension product is simulated under the current capital market model parameters. The usage of the CRC boundaries which were valid during PIA classification while simultaneously employing of the current model parameters would lead to an inconsistent result. In addition, the qualitative criteria of CRC 1 and 2 have to be taken into account for determining the CRC.

(ii) The chance and risk measure are defined as the annual constant interest rate over the remaining accumulation phase leading to the current averaged final contract values resulting from the current simulation. For this, Equation (4.3) has to be modified. The chance and risk measure then solve the equation

$$\overline{V}_t = \begin{cases} \widetilde{v}_t \cdot \left(1 + \frac{\mu_t}{12}\right)^{12T_t} + 100 \sum_{j=1}^{12T_t} \left(1 + \frac{\mu_t}{12}\right)^j, & \text{regular premium} \\ \widetilde{v}_t \cdot \left(1 + \frac{\mu_t}{12}\right)^{12T_t}, & \text{single premium} \end{cases}$$

with the appropriate averaged final contract value and measure at time t. Based on μ_t^c and μ_t^r, the pension product is assigned to a CRC. Since CRC boundaries do not exist for each accumulation phase T_t, we adapt the procedure of PIA using the existing boundaries of the next larger accumulation phase. For $T_t > 40$ no such boundaries exist. The CRC boundaries of an accumulation phase of 40 years are used instead.

Both approaches differ in the consideration of the past in the calculation of the chance and risk measure. In the first approach, the pension product is classified by its performance of the past as well as the future. In the second approach, the performance of the past does not matter in the calculation of the measures. Only the future performance is decisive for classification. Thus, using the second approach might result into a low chance-measure, although the pension product had a tremendous performance on the first t months and vice versa. The same statements as in Section 4.3.3 apply here and we recommend the first calculation of the chance and risk measure as return over the entire contract term for the same reasons. Nevertheless, a final specification of the legislator about the valid approach is required.

The CRC will be decreasing over the contract term in the above introduced approaches. This is caused by the declining simulated remaining accumulation phase. Thus, the variance of the prices of the underlying assets in the simulation also decreases over the contract term. This in turn reduces the variance of the simulated final contract values. Thereby, the averaged final contract values $\overline{V}_t^c$ and $\overline{V}_t^r$ converge towards each other. This also holds for μ_t^c and μ_t^r. At the end of the accumulation phase the averaged final contract values are equal and correspond to the contract value $v_{12\cdot\mathcal{T}}$. Thus, the chance and risk measure have the same value. Therefore, the pension product has a CRC between 1 and 3 depending on the qualitative criteria at the end of the accumulation phase of the contract. In a μ^c-μ^r-diagram, the contract moves towards the straight line through the point (0,0) with slope of 1 over the contract term.

We take a monthly view as a basis of the CRC determination of an existing pension product. Equally, we can use an annual view and calculate the CRC after one year, two years, and so on. This corresponds to $t = 0, 12, 24, \ldots, 12 \cdot \mathcal{T}$. The customer receives information of his contract value once a year from the insurance company. This can be used for the determination of the contract values v_t after t months with $t = 0, 12, 24, \ldots, 12 \cdot \mathcal{T}$.

It cannot be definitively answered which approach should be taken. All presented approaches are conceivable and consistent with the PIA classification. The approach following the legislator is easy to use but does not necessarily consider the correct costs and paid premium payments. This is eliminated in the other introduced approaches albeit with a drawback: The requirement of a simulation. As long as all approaches lead to the same CRC no miscounselings can happen regardless of the used approach. The discrepancy in the CRC of the legislator approach compared to the succeeding increases with the difference between the remaining accumulation phase and the used accumulation phase according to the legislator's approach. In order to favor one approach, a specification of the legislator is required.

Recommendation for Customer Consulting. *The CRC issued by PIA can be used for determining the CRC of an existing pension product. In order to avoid the abovementioned drawbacks of this approach, simulations of the future contract values are required. For this, a simulation application of the PIA model has to be provided to*

the consultant by the insurance company so that he can determine the CRC during the consultation.

4.5 Conclusion

The classification of state-subsidized pension products by PIA is an important instrument for consumer protection as it enables the customer to assess such a product in terms of its chance potential and risk. Together with the procedure developed by EI-QFM to determine the customer's chance-risk profile, the customer consultant has two useful instruments providing optimal and customer individual consulting with regard to suitable pension products.

However, in order to provide suitable consulting to a customer who already has pension products, a CRC determination of the resulting portfolio of existing and new product is necessary. We developed an appropriate procedure for this. It is based on an examination of the properties of the various chance and risk parameters and their mappings. Only the mapping to the mean of the final contract values, i.e. the contract values at the end of the accumulation phase, is linear in the combination of two products, while all others are concave. We assume that – at least approximately justified – the costs of the products are linear in the premium payment and the contract value. Based on the properties, we showed that a diversification effect exists in the classification according to PIA. It is also possible, for example, to include products with a higher CRC than the customer's in the portfolio by combining products and still maintain the desired overall CRC while increasing the earnings potential at the same time.

Another important result of our analysis is that the averaged final contract values should be used as upper bound of the CRC of the portfolio. These must be interpolated according to their proportions and the corresponding chance and risk measure must be calculated. In the case of a portfolio consisting of two new pension products, the proportions of each product can be determined so that the CRC of the portfolio is not greater than a given CRC. We also applied our diversification results to the case of a portfolio consisting of an existing and a new pension product where it is possible to increase the premium of the existing pension product. If this is not the case, the upper bound of the portfolio's CRC must be determined for each proportion in the new pension product until it has the desired CRC.

This results in the necessity for the legislator to create corresponding regulations on how the consultant can use the final contract values without explicitly knowing them since under the current regulation neither the explicit values of the chance and risk measure nor the simulated final contract values of the products used by PIA may be communicated externally by the provider. Only in this way our *recommendations for customer consulting* derived from our theoretical results can actually be implemented by the consultant.

Our results are not only useful for consulting but can also be interesting to the insurer in terms of cost-effective and customer-fit product development. For example,

determining or restricting the CRC of a portfolio enables the provider to construct (approximately) optimal products for the respective CRC from his existing products by appropriate portfolio construction without development of completely new product concepts. Moreover, in principle, our analysis can also include other assets of the customer as state-subsidized pension products. However, these must be classified according to the PIA model.

Finally, we suggested different approaches of the CRC determination over the contract term since only the CRC at the beginning of the accumulation phase is obtained by the PIA classification. The current legal situation is applied to the CRC determination over the contract term. Here, however, the pension product and the development of its CRC is not considered correctly. Therefore, we developed an approach that eliminates these disadvantages but requires further simulations.

A Analysis of $B_z(x)/x$

Lemma A.1. *Let be $z > 0$ and $\frac{B_z(x)}{x}$ be defined as*

$$\frac{B_z(x)}{x} := \frac{1 - e^{-zx}}{zx}.$$

For $x > 0$ it holds

- *$\frac{B_z(x)}{x}$ is positive,*
- *the first derivative of $\frac{B_z(x)}{x}$ is negative,*
- *the second derivative of $\frac{B_z(x)}{x}$ is positive,*
- *the third derivative of $\frac{B_z(x)}{x}$ is negative.*

Proof. Since e^{-zx} is lower than 1 for $z, x > 0$, the numerator of $\frac{B_z(x)}{x}$ is positive. Consequently, $\frac{B_z(x)}{x}$ is positive for $z, x > 0$.

The first derivative of $\frac{B_z(x)}{x}$ is

$$\frac{\partial B_z(x)/x}{\partial x} = \frac{(1 + zx)e^{-zx} - 1}{zx^2}.$$

$(1 + zx)e^{-zx}$ is lower than 1 for $z, x > 0$ since $1 + zx < e^{zx}$ for $z, x \neq 0$. Consequently, the first derivative of $\frac{B_z(x)}{x}$ is negative for $z, x > 0$.

The second derivative of $\frac{B_z(x)}{x}$ is

$$\frac{\partial^2 B_z(x)/x}{\partial x^2} = \frac{2\left(1 - (1 + zx)e^{-zx}\right) - z^2x^2e^{-zx}}{zx^3}.$$

We consider the numerator and denominator separately. The denominator is always positive because of $z, x > 0$. We denote the numerator by $f(x)$. For x going to 0 we have

$$\lim_{x \downarrow 0} f(x) = 0.$$

Furthermore, the first derivative of $f(x)$ is

$$\frac{\partial f(x)}{\partial x} = z^3x^2e^{-zx}$$

and is positive for $z > 0$. Therefore, the numerator increases from zero on and is always positive. Due to $f(x) > 0$ and $zx^3 > 0$, the second derivative of $\frac{B_z(x)}{x}$ is positive for $z, x > 0$.

The third derivative of $\frac{B_z(x)}{x}$ is

$$\frac{\partial^3 B_z(x)/x}{\partial x^3} = \frac{6\left((1+zx)\,e^{-zx}-1\right)+3z^2x^2e^{-zx}+z^3x^3e^{-zx}}{zx^4}.$$

We consider the numerator and denominator separately. The denominator is always positive because of $z, x > 0$. We denote the numerator by $g(x)$. For x going to 0 we have

$$\lim_{x \downarrow 0} g(x) = 0.$$

Furthermore, the first derivative of $g(x)$ is

$$\frac{\partial g(x)}{\partial x} = -z^4x^3e^{-zx}$$

and is negative for $z, x > 0$. Therefore, the numerator decreases from zero on and is always negative. Due to $g(x) < 0$ and $zx^4 > 0$, the second derivative of $\frac{B_z(x)}{x}$ is negative for $z, x > 0$. □

B Descriptive Statistics of the Yield Increments with Different Time-Lags

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-21	Lag-256	Lag-2 partial autocorr.
0.5	-0.0005	0.0167	-0.1545	0.1707	0.0890	16.4943	0.0108	0.0035	0.0228	-0.0005
1	-0.0006	0.0189	-0.2637	0.1799	-1.0828	25.0857	0.0861	0.0289	0.0195	0.0220
3	-0.0010	0.0298	-0.3295	0.1894	-0.7038	12.0425	0.0533	0.0462	0.0261	0.0205
5	-0.0013	0.0336	-0.2258	0.1773	-0.0493	4.0109	0.0405	0.0299	0.0230	0.0237
7	-0.0015	0.0356	-0.1772	0.1770	0.1585	2.3174	0.0414	0.0208	0.0204	0.0184
10	-0.0017	0.0376	-0.1930	0.1775	0.2229	2.0838	0.0458	0.0147	0.0191	0.0048
20	-0.0018	0.0419	-0.2410	0.2403	0.1043	3.1072	0.0619	0.0028	0.0257	-0.0169

Table B.1: Increments of 1 trading day

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-21	Lag-256	Lag-2 partial autocorr.
0.5	-0.0101	0.0932	-0.5859	0.3183	-0.9473	6.2709	0.9680	0.1447	0.1257	-0.0323
1	-0.0128	0.1024	-0.6374	0.4495	-0.8694	7.6962	0.9666	0.1106	0.1162	-0.1017
3	-0.0222	0.1389	-0.7565	0.6057	-0.3475	4.1167	0.9558	0.0345	0.0546	-0.0864
5	-0.0283	0.1563	-0.7468	0.5421	-0.0509	1.5862	0.9550	0.0162	0.0508	-0.0808
7	-0.0323	0.1668	-0.6907	0.5577	0.1852	0.8623	0.9554	0.0192	0.0593	-0.0830
10	-0.0360	0.1764	-0.6732	0.5930	0.3012	0.7540	0.9553	0.0297	0.0527	-0.0875
20	-0.0387	0.1928	-0.7717	0.8101	0.1795	1.4881	0.9531	0.0056	0.0358	-0.1090

Table B.2: Increments of 21 trading days

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-21	Lag-256	Lag-2 partial autocorr.
0.5	-0.0295	0.1717	-0.9794	0.6237	-1.1834	7.1673	0.9907	0.7486	-0.1831	0.0387
1	-0.0382	0.1859	-0.9677	0.7995	-0.7924	6.8138	0.9899	0.7480	-0.2363	-0.0620
3	-0.0668	0.2500	-1.1645	1.0069	-0.2235	3.6195	0.9862	0.7252	-0.1415	-0.0707
5	-0.0844	0.2778	-1.0957	0.8702	0.1009	1.3621	0.9858	0.7099	-0.0448	-0.0681
7	-0.0955	0.2946	-1.0513	0.8555	0.3534	0.4725	0.9858	0.7000	-0.0045	-0.0713
10	-0.1052	0.3128	-1.0268	0.8256	0.4994	0.2287	0.9859	0.6969	-0.0082	-0.0719
20	-0.1117	0.3410	-0.9746	1.1690	0.3951	0.7889	0.9850	0.6935	-0.0357	-0.0765

Table B.3: Increments of 64 trading days

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-21	Lag-256	Lag-2 partial autocorr.
0.5	-0.0571	0.2584	-1.2996	0.8041	-0.9825	4.9503	0.9957	0.8449	-0.2715	-0.0149
1	-0.0737	0.2833	-1.3480	0.8512	-0.5677	3.7765	0.9953	0.8497	-0.3430	-0.0920
3	-0.1310	0.3729	-1.3205	1.2814	0.2045	2.1300	0.9933	0.8427	-0.2561	-0.0550
5	-0.1642	0.4074	-1.4424	1.3571	0.3275	1.1683	0.9931	0.8324	-0.1581	-0.0470
7	-0.1844	0.4253	-1.4217	1.2826	0.3246	0.2222	0.9930	0.8289	-0.1268	-0.0446
10	-0.2016	0.4456	-1.3094	1.1368	0.2561	-0.4732	0.9930	0.8310	-0.1274	-0.0409
20	-0.2117	0.4829	-1.4947	1.0533	0.0509	-0.2739	0.9927	0.8331	-0.1501	-0.0523

Table B.4: Increments of 128 trading days

Time to maturity (in years)	Mean (10^{-2})	Std. Dev. (10^{-2})	Min (10^{-2})	Max (10^{-2})	Skew	Kurt	Autocorrelation Lag-1	Lag-21	Lag-256	Lag-2 partial autocorr.
0.5	-0.1267	0.3469	-1.3116	1.0237	-0.2295	2.2305	0.9976	0.9300	-0.4886	0.0053
1	-0.1558	0.3753	-1.4208	1.1403	-0.3440	2.7763	0.9973	0.9268	-0.5365	-0.0654
3	-0.2646	0.4745	-1.7613	1.2126	-0.4406	0.9804	0.9958	0.9104	-0.5332	-0.0362
5	-0.3237	0.5075	-1.6993	0.9654	-0.3092	-0.3919	0.9955	0.9029	-0.4620	-0.0203
7	-0.3585	0.5223	-1.6392	0.6775	-0.2144	-0.9988	0.9953	0.8987	-0.4209	-0.0168
10	-0.3879	0.5415	-1.7044	0.8099	-0.2495	-0.8526	0.9951	0.8961	-0.3956	-0.0231
20	-0.4039	0.5782	-2.1340	0.9790	-0.3829	-0.0473	0.9947	0.8916	-0.3880	-0.0523

Table B.5: Increments of 256 trading days

Bibliography

S.-K. Acar, R. Korn, K. Natcheva-Acar, and J. Wenzel. A two-factor HJM interest rate model for use in asset liability management. In G. Mitra and K. Schwaiger, editors, *Asset and Liability Management Handbook*, 62–76, Palgrave Macmillan, UK, 2011.

J. Bergstra and Y. Bengio. Random search for hyper-parameter opimization. *Journal of Machine Learning Research*, 13:281–305, 2012.

C. Bernadell, J. Coche, and K. Nyholm. Yield curve prediction for the strategic investor. *ECB Working Paper Series*, 472, 2005.

S. K. Bose, J. Sethuraman, and S. Raipet. Forecasting the term structure of interest rates using neural networks. In J. Kamruzzaman, R. Begg, and R. Sarker, editors, *Artificial Neural Networks in Finance and Manufacturing*, pages 124–138. IGI Global, 2006. ISBN 9781591406709. doi: 10.4018/978-1-59140-670-9.ch007.

A. Brace, D. Gatarek, and M. Musiels. The market model of interest rate dynamics. *Mathematical Finance*, 7:127–155, 1997.

D. Brigo and F. Mecurio. *Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit*. Springer, 2nd edition, 2007. ISBN 978-3-540-22149-4.

H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

J. F. Caldeira, G. V. Moura, and A. A. Santos. Predicting the yield curve using forecast combinations. *Computational Statistics & Data Analysis*, 100:79–98, 2016a. doi: 10.1016/j.csda.2014.05.008.

J. F. Caldeira, G. V. Moura, A. A. Santos, and F. Tourrucôo. Forecasting the yield curve with the arbitrage-free dynamic Nelson-Siegel model: Brazilian evidence. *EconomiA*, 17(2):221–237, 2016b. doi: 10.1016/j.econ.2016.06.003.

Y. Chen and L. Niu. Adaptive dynamic Nelson-Siegel term structure model with applications. *Journal of Econometrics*, 180(1):98–115, 2014.

G. Chevillon. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785, 2007. doi: 10.1111/j.1467-6419.2007.00518.x.

J. H. Christensen, F. X. Diebold, and G. D. Rudebusch. The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, 164(1):4–20, 2011. doi: 10.1016/j.jeconom.2011.02.011.

J. E. Cox, J. C. ans Ingersoll and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

M. De Pooter. Examining the Nelson-Siegel Class of term structure models. *Tinbergen Institute Discussion Paper*, 043/4, 2007.

M. De Pooter, F. Ravazzolo, and D. van Dijk. Term structure forecasting using macro factors and forecast combination. *International Finance Discussion Papers*, 993, 2010.

R. B. De Rezende and M. S. Ferreira. Modeling and forecasting the yield curve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting*, 32(2):111–123, 2013. doi: 10.1002/for.1256.

M. Deistler and W. Scherrer. *Modelle der Zeitreihenanalyse*. Mathematik Kompakt. Birkhäuser, Cham, 1st edition, 2018. ISBN 978-3-319-68663-9.

S. Desmettre and R. Korn. *Moderne Finanzmathematik – Theorie und praktische Anwendungen: Band II: Erweiterungen des Black-Scholes-Modells, Zins, Kreditrisiko und Statistik*. Springer Spektrum, 2018. ISBN 978-3-658-20999-5.

F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006. doi: 10.1016/j.jeconom.2005.03.005.

F. Diez and R. Korn. Yield curve shapes of Vasicek interest rate models, measure transformations and an application for the simulation of pension products. *European Actuarial Journal*, 10:91–120, 2020. doi: 10.1007/s13385-019-00214-0.

F. Diez, R. Horsky, and R. Korn. Chancen-Risiko-Klassifizierung eines Portfolios aus staatlich geförderten Altersvorsorgeprodukten und Empfehlungen für die Kundenberatung. 2020. Manuscript submitted for publication.

E. F. Fama and R. R. Bliss. The information in long-maturity forward rates. *American Economic Review*, 77(4):680–692, 1987.

H. Föllmer and A. Schied. *Stochastic finance: An introduction in discrete time*. De Gruyter graduate. De Gruyter, Berlin and New York, NY, 3rd rev. and extended edition, 2011. ISBN 978-3-11-021804-6.

J. Franke. *Einführung in die Statistik der Finanzmärkte*. Springer, Berlin, 2002. ISBN 3-540-41722-2.

D. Heath, R. Jarrow, and A. Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60:77–105, 1992.

T. Ho and S. Lee. Term structure and pricing interest rate contingent claims. *Journal of Finance*, 41:1011–1029, 1986.

K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

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

M. Keller-Ressel. Correction to: 'Yield curve shapes and the asymptotic short rate distribution in affine one-factor models'. *Finance and Stochastics*, 22(2):503–510, 2018.

M. Keller-Ressel. Total positivity and the classification of term structure shapes in the two-factor vasicek model, 2019.

M. Keller-Ressel and T. Steiner. Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. *Finance and Stochastics*, 12(2):149–172, 2008.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. 2014. URL `https://arxiv.org/abs/1412.6980`. Online accessed 23-09-2020.

R. Korn and V. P. Andelfinger. Der Kunde - Chance und Risiko im Beratungsgespräch. *Zeitschrift für Versicherungswesen*, 17:538–540, 2016.

R. Korn and A. Wagner. Chance-risk classification of pension products: Scientific concepts and challenges. *Innovations in Insurance, Risk- and Asset Management*, pages 381–398, 2018.

R. Korn and A. Wagner. *Praxishandbuch Lebensversicherungsmathematik: Simulation und Klassifikation von Produkten*. VVW GmbH, Karlsruhe, 1st edition, 2019. ISBN 978-3-96329-074-9.

M. P. Laurini and L. K. Hotta. Bayesian extensions to diebold-li term structure model. *International Review of Financial Analysis*, 19(5):342–350, 2010. doi: 10.1016/j.irfa.2010.08.010.

K. Miltersen, K. Sandmann, and D. Sondermann. Closed-form solutions for term structure derivatives with log-normal interest rates. *Journal of Finance*, 52:409–430, 1997.

C. R. Nelson and A. F. Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489, 1987.

T. Poklepović, Z. Aljinović, and B. Marasović. The ability of forecasting the term structure of interest rates based on Nelson-Siegel and Svensson model. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 8(3):718–724, 2014.

Produktinformationsstelle Altersvorsorge. Beschreibungstexte für die Chance-Risiko-Klassen im Produktinformationsblatt. 2016. URL `https://produktinformationsstelle.de/wp-content/uploads/2018/08/CRK-Texte-PIB_082016.pdf`. Online accessed 20-07-2020.

R. Sambasivan and S. Das. *A Statistical Machine Learning Approach to Yield Curve Forecasting*. IEEE, Piscataway, NJ, 2017.

Schreiben des Bundesministeriums der Finanzen vom 14.03.2019. Produktinformationsblatt gemäß §7 Altersvorsorgeverträge-Zertifizierungsgesetz (AltZertG); Amtlich vorgeschriebenes Muster gemäß §13 Altersvorsorge-Produktinformationsblattverordnung (AltvPIBV). *IV C PIA - S 2220 - a/16/10003:004, BStBl I 2019*, page 240, 2019.

A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse. Methodology for long-term prediction of time series. *Neurocomputing*, 70(16-18):2861–2869, 2007. doi: 10.1016/j.neucom.2006.06.015.

L. E. O. Svensson. Estimating and interpreting forward interest rates: Schweden 1992 - 1994. *NBER Working Paper*, 1994.

S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8):7067–7083, 2012. doi: 10.1016/j.eswa.2012.01.039.

J. Täppinen. Interest rate forecasting with neural networks. *VATT-Discussion papers*, 170, 1998.

R. S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley and Chichester : John Wiley [distributor], Hoboken, N.J., 3rd edition, 2010. ISBN 978-0-470-41435-4.

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

M. Weber. *Kodex zur Anlageberatung: Gute Sitten und optimale Entscheidungen*. Forschung für die Praxis. Behavioral Finance Group, Lehrstuhl für ABWL Finanzwirtschaft, Insbesondere Bankbetriebslehre, Univ. Mannheim, 2009.

M. Wiese, R. Knobloch, R. Korn, and P. Kretschmer. Quant gans: deep generation of financial time series. *Quantitative Finance*, 2015(1):1–22, 2020. doi: 10.1080/14697688.2020.1730426.

H. G. Zimmermann, C. Tietz, and R. Grothmann. Yield curve forecasting by error correction neural networks and partial learning. *ESANN'2002 proceedings - European Symposium on Artificial Neural Networks*, pages 407–412, 2002.

Scientific career

01/2018 – 10/2020	PhD Student supervised by Prof. Dr. Ralf Korn Fraunhofer Institute for Industrial Mathematics ITWM and Technische Universität Kaiserslautern, Kaiserslautern, Germany
01/2016 – 12/2017	Research Associate at the Department of Financial Mathematics Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany
10/2011 – 04/2014	Master of Science (M.Sc.) Industrial Mathematics Julius-Maximilians Universität, Würzburg, Germany
10/2008 – 07/2011	Bachelor of Science (B.Sc.) Industrial Mathematics Julius-Maximilians Universität, Würzburg, Germany
09/1999 – 06/2008	Secondary School Frobenius-Gymnasium, Hammelburg, Germany

Wissenschaftlicher Werdegang

01/2018 – 10/2020	Doktorandin von Prof. Dr. Ralf Korn Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM und Technische Universität Kaiserslautern, Kaiserslautern, Deutschland
01/2016 – 12/2017	Wissenschaftliche Mitarbeiterin der Abteilung Finanzmathematik, Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM, Kaiserslautern, Deutschland
10/2011 – 04/2014	Master of Science (M.Sc.) Wirtschaftsmathematik Julius-Maximilians Universität, Würzburg, Deutschland
10/2008 – 07/2011	Bachelor of Science (B.Sc.) Wirtschaftsmathematik Julius-Maximilians Universität, Würzburg, Deutschland
09/1999 – 06/2008	Gymnasium Frobenius-Gymnasium, Hammelburg, Deutschland