State price density models for the term structure of interest rates

Applications to insurance and expansions to the stock market and macroeconomic variables

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik des Karlsruher Instituts für Technologie (KIT) genehmigte

DISSERTATION

von

Dipl.-Math. oec. Udo Robin Pfeiffer

Tag der mündlichen Prüfung: 21. Juli 2010

Referent: Prof. Dr. Nicole Bäuerle
Korreferent: JProf. Dr. Luitgard Veraart

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Stochastik des Karlsruher Instituts für Technologie als Beitrag zum Projekt “Alternative Investments”, gefördert durch das Bundesministerium für Bildung und Forschung unter Förderkennzeichen *03BAPAD2*.
Die Chance, an diesem Projekt mitzuarbeiten, verdanke ich Frau Prof. Dr. Nicole Bäuerle. Sie hat weiterhin durch Diskussionen und Vorschläge maßgeblich zum Entstehen dieser Arbeit beigetragen, mir aber auch stets weitestgehende Freiheiten gelassen in der Weiterentwicklung des Themas. Dafür möchte ich mich bedanken.

Bei Frau JProf. Dr. Luitgard Veraart möchte ich mich bedanken für die freundliche Übernahme des Korreferats.

Herrn Dr. Bierbaum und den anderen Mitarbeitern der Allianz Lebensversicherungs-AG möchte ich für Ihre Diskussionsbereitschaft und ihre Anregungen ebenso danken wie für ihre Bereitschaft, mir während einiger Wochen in Stuttgart die Anwenderseite der Lebensversicherungsmathematik näher zu bringen.

Bedanken möchte ich mich auch bei allen Mitarbeitern des Instituts für Stochastik für die angenehme Atmosphäre und die gute Zusammenarbeit, insbesondere bei meinen Büro-”Mitbewohnern” Florian Ketterer, Birte Muhsal, Sebastian Urban und nicht zuletzt Jonathan Ott, mit dem ich nun schon fast drei Jahre ein Büro teile.

Schließlich möchte ich mich bei meiner Familie bedanken, die mir mein - überlanges - Studium ermöglicht hat und mich auch während meiner Promotion stets unterstützt hat.

Contents

List of Figures

Chapter 1

Introduction

A basic task in mathematical finance lies in comparison of cash flows occurring at different points in times. In many cases, the current term structure might be sufficient for such a comparison. However, reliable current term structure data is typically limited to less than ten years of maturity due to liquidity constraints. Furthermore, the current term structure can not reflect dependencies between discounting functions and future payoffs subject to interest rate risk. A stochastic term structure model is therefore required, which allows to derive the joint distribution of stochastic discounting functions with arbitrary time to maturity and future payoffs.

So far, most term structure models presented in the literature were developed for banking applications. Regulatory requirements considering mathematical finance models also emerged earlier in the banking sector than for insurance companies. Two main frameworks to term structure modeling dominate in banking, namely so called shortrate models, in particular the affine model class, and so called market models, in particular the well known Libor market model. An overview of current term structure models and their applications in banking may be found in the books of Filipovic [Fil09], Brigo and Mercurio [BM01] or Musiela and Rutkowski [MR05].

For insurance companies, discounting future payoffs dependent on term structure dynamics play a central role. What is special in insurance applications is that, first, the duration of life or pension insurance contracts typically exceed the available maturities of the currently observable term structure. On the other side, the insurer requires a term structure model to discount his contractual liabilities, which typically requires Monte Carlo simulation due to complexity and dependence on other factors of risk such as biometrical risk or cancellation. Finally, life and pension insurers at least in continental Europe mainly invest in fixed income securities. Future payoffs of insurance contracts therefore depend on intermediate returns achieved in fixed income markets and intermediate portfolio allocation, which implements path-dependence for most insurance applications. This path-dependence in turn requires that simulated yield curves are among those historically observed. To conclude we have to consider the joint dynamics of stochastic discounting and payoff functions. This shows the importance of long-term interest rate models for insurance applications, but also points to general problems in implementation.

In this work, we repeat estimation and implementation of the Cairns [Cai04a] model using the Extended Kalman filter already known in the literature. Additionally, we present the cosh model proposed by Rogers [Rog97] in its first estimation and implementation, again using the Extended Kalman filter. By its theoretical properties and also by the measures of historical fit which we derived, the Cairns model is superior. On the other side, we find that the cosh model, albeit not guaranteeing positive interest rates, is vastly superior computationally to the Cairns model. As the cosh model and the Cairns model show several basic similarities in implementation and stochastic dynamics, we find that the cosh model can be used as a fast approximation of the otherwise superior Cairns model.
In a second step, we show how to expand the pure bond market cosh model to a full investment model covering equity, government bond and inflation-indexed bond markets. In general, the techniques provided for expansion of the cosh model would work with the Cairns model as well, yet are typically computationally unfeasible. The cosh model allows for efficient implementation of such expanded models, which are another contribution of this work. As we show how to include macroeconomic variables as well as monetary policy rules, the cosh model may also be used as a macro-finance model in monetary policy applications as well as to examine the impact of macroeconomic variables on insurance products.
Another contribution of this work is the explicit insurance focus we take in our examinations. Both the Cairns and cosh models were evaluated with respect to their applicability in insurance, the model extensions presented reflect actual demand of insurance companies. A further example would be the recent discussion of regulatory specification of the asymptotic long rate, because of which we discuss the ability of the two models considered to estimate the asymptotic long rate and to implement sensitivity analysis of the asymptotic long rate.

The thesis is organized as follows. We present in section 2.1 a selection of criteria on term structure models and discuss their importance both for insurance and banking applications. Considering examples for insurance applications, we generally think of long-term life or pension insurance contracts, whereas considering banking applications, we think of plain vanilla interest rate derivatives such as caps, floors or swaptions. The main difference hence lies in contractual time to maturity, risk factors included and availability of market prices for comparison.

In section 2.2.1, we introduce the Rogers framework for definition of term structure models based upon the state price density. The state price density approach provides an alternative pricing approach to the better known risk-neutral approach, which requires discounting by the integrated shortrate under the risk-neutral measure. We find that the state price density approach is computationally superior to the risk-neutral pricing approach once path dependent payoffs with infrequent and irregular payment dates are considered. The Rogers framework defines the state price density by the specification of a state vector process and the choice of a function f with rather general properties. We discuss how the choices of f and the dynamics of X may be restricted by the criteria in 2.1.

In sections 2.2.5 and 2.2.6, we present the Cairns [Cai04a] and cosh [Rog97] models, respectively. Both are state price density models based upon multi-dimensional Ornstein-Uhlenbeck processes. Cairns proposed his model explicitly for long-term interest rate modeling, as it provides sustained periods of both high and low interest rates. Furthermore, the Cairns model guarantees positive interest rates. However, it is computationally slow. We found that the cosh model does not guarantee positive interest rates, yet it provides sustained periods of both high and low interest rates as the Cairns model, and in fact with similar underlying dynamics driving the model. The cosh model however offers improved computational efficiency.

In section 2.2.7, both models are proved to be free of arbitrage. Both models allow to choose freely the market price of risk, the drift correction term from the risk-neutral measure to the physical measure.

In section 2.3, we shortly discuss standard approaches to estimate term structure models. Both models can be interpreted as factor models. A measurement equation links the factor process to the observable term structure. This state space model provides the defining framework of the Kalman filter or, as the measurement equation is nonlinear in both models, its extended form.

In section 2.3.3 we estimation data for term structure models. Riskless interest rates from government bonds provide the underlying dynamics, macroeconomic data may be used to improve long-term dynamical properties, interest rate derivative data may improve volatility fit.

We present estimation results for both models. Sections 2.3.4 and 2.3.5 provide the implementation of the Extended Kalman filter and its estimation results for the two-dimensional Cairns and cosh models, respectively. Section 2.3.7 provides results for the respective three-factor models. Simulation exercises for the two-factor cases demonstrate the ability of the Extended Kalman filter to provide the "true" parameters. Historical fit is examined by calculating mean absolute errors as well as cross-correlation and autocorrelation of the time series of residuals.

Generally, we found that the underlying state vector components coincide with the principal components of the term structure. This result was used in section 2.3.8 to specify the long-term mean of the Ornstein-Uhlenbeck state process, which the Kalman filter underestimated in the sense that the term structure implied by the long-term mean of the state

vector was too low in comparison to those historically observed.
Another important aspect of both models discussed in section 2.3.9 is that one model parameter equals the asymptotic long rate, the limit of the yield curve for infinite maturity. The asymptotic long rate in both models is thus constant – as in many other currently used term structure models.
Section 2.4 concludes with a comparison of the Cairns and cosh model.

Pure term structure models are insufficient to simulate investment success of an insurance company's investment portfolio. A major task therefore is to extend the financial market consistently. The first, and most important, extension in section 3.1 considers equity, since it is generally assumed that on the long run stocks provide higher returns than bonds and mixed bond and stock portfolios provide lower volatilities thanks to diversification effects. We first consider in general stock pricing within the state price approach, particularly considering dividend payments. We then derive and implement two approaches to include stock price data. Both are estimated on bond and stock market data using the Extended Kalman filter. Both approaches, dividend discount and Black-Scholes based, guarantee no-arbitrage if used with the cosh model. For both models, estimation results are provided and historical fit of both the bond and stock market model is examined.
First, in section 3.1.3, we use a dividend discount approach which interprets the stock price at time t as the value of all future dividends discounted at time t. Such an approach easily fits into the state price density framework, which is used to discount the dividends. We find that the dividend discount model is unfeasible to be implemented in the Cairns framework due to computational limitations. For the cosh model, on the other side, it provides an arbitrage-free, implementable stock pricing framework.
The second stock market approach, presented in section 3.1.4, is based on the Black-Scholes stock market model. Here, stock price dynamics are defined under the risk-neutral measure using the shortrate as provided by the bond market model. Using the market price of risk, stock price dynamics under the physical measure can be derived.
In section 3.2, we expand the cosh model to macroeconomic variables. It is well known that macroeconomic variables improve forecasting ability of term structure models. Particularly long-term dynamics of the term structure should improve with the inclusion of macroeconomic data. We present a rate-based and an index-based approach to the inclusion. Furthermore, we present how monetary policy rules should be included into the model framework to ensure that these rules hold on average for simulated yields as well. We also provide ideas how macroeconomic variables might be used in insurance applications besides improved dynamics of the term structure model.
Section 4 concludes.

Chapter 2

The basic bond market model

2.1 Criteria of term structure models

Term structure models are a major tool in the finance industry. First, term structure models are the basic pricing tools for fixed income markets. Second, yet more important, term structure models are required to implement stochastic discounting functions. In many pricing applications, interest rates are assumed as constant, an example being the Black-Scholes framework. As soon as financial instruments are considered with long maturities for which no interest rate is observed, or with path-dependent payoffs conditional on terms structures observed during maturity, stochastic interest rates have to be considered.

Life insurance companies are particularly dependent on reliable term structure models. Insurance companies invest large parts of their reserves in the fixed income markets, particularly in government bonds or other investment grade fixed income securities. For these assets, the term structure of domestic government bonds is a benchmark describing market dynamics. On the other side, life insurance products may provide cash flows so far into the future that the currently observed term structure can not be used for discounting. In these cases, term structure models are used to provide the discounting functions required.

The second major group of financial actors are banks. In case of banks, term structure models are predominantly required to price interest rate derivatives. Furthermore, as discussed earlier, exotic derivatives with path dependent intermediate cash flows might require term structure models as well. Banks do typically not face the very long times to maturity life insurance companies are forced to handle. This implies that the currently observed term structure is sufficient to discount the cash flows encountered in banking applications.

All in all, we can conclude that requirements on interest rates vary substantially with the implied application. In particular, life insurance applications have special requirements considering the stochastics of the model and the maturities involved. During the last decades, various term structure models were developed and presented in the literature. One goal of this work is to derive criteria which allow to evaluate the applicability of a term structure model for different tasks, in particular life insurance applications. In the following, we will present a number of characteristics for term structure models and their respective importance for insurance and banking applications.

2.1.1 No-Arbitrage

An arbitrage strategy exploits price differentials between markets or assets which allow to gain a riskless profit. Although such price differentials might exist in reality, their exploitation by arbitrageurs closes the price differential quickly, see for example [Hul00]. In highly liquid markets such as government bond markets or swap markets, we can reasonably assume that no arbitrage holds in reality as well. In pricing models, the no-arbitrage condition becomes a basic consistency assumption which links all fixed income submarkets. Requiring no-arbitrage therefore means that within the pricing model, no systematic inconsistencies between fixed income securities exist. Namely, the no arbitrage condition guarantees that if the model is fitted to a certain set of market data at a certain point in time, all prices of contingent claims derived using the model are consistent with observed prices and therefore, in a sense "fair".

In banking applications, the primary goal is pricing of contingent claims and their hedging. In case of pricing, the no-arbitrage condition is required to guarantee consistency of the derived prices with observed prices. Typically, pricing is based on fitting the model to a collection of observed prices of certain liquid contingent claims and then it is assumed that model-derived arbitrage-free prices of assets of the same type are consistent and hence "fair". This approach is called *calibration* and is discussed in [Reb02] or [RSM04] for interest rate derivatives. In case of hedging, the basic assumption of using some assets to hedge against price changes of other assets is valid only if the prices move consistently, as guaranteed by no-arbitrage.

Note that term structure models typically have problems in fitting several types of fixed income assets at once, see again [Reb02] or [RSM04]. The most important example would be differences between swaption and cap markets, see [LSCS01]. Even if a term structure model is arbitrage-free, this does not imply that cross-asset hedging is possible within the model. As a consequence, to price contingent claims the model should be fitted to observed prices of related assets, and to hedge assets of one type by assets of another type the model must be fitted to observed prices of both types. Typically, we encounter no problems if we consider only an underlying and one type of derivative. Problems might be substantial if we consider cross-asset hedging for two types of derivatives. This is of particular interest for insurance applications and we will discuss this further in section 2.3.3.

In pricing of contingent claims in the insurance world, the same arguments hold as in pricing and hedging of contingent claims of the banking sector. Although insurance products are typically too complex and too illiquid to exploit possible arbitrage opportunities, consistency between insurance contingent claims and other financial data is still required.

Of further importance is the aspect of investment, which is a major determinant of the insurance contract. If the model allows for arbitrage, an investment policy may exploit these arbitrage possibilities. Given a specified allocation rule, the simulated investment success may therefore be due to arbitrage rather than realistic investment policies.

2.1.2 Boundedness

Boundedness of interest rates considers two aspects. For once, we have a lower bound at zero exists. Second, we take as granted that interest rates cannot be arbitrarily high. Considering the lower bound, there is a simple economical explanation. As the lender of money has to abstain from consumption now until being repaid, even in case of no credit risk the lender will demand a compensation for this abstain. This reflects the basic assumption of time preferences that consumption of a certain value today is preferred over consumption of the same value at some later date. This general idea of borrowing and lending implies that interest rates are compensations for time transfer of consumption for which the lender requires a compensation in the form of interest.
Such a simplistic argumentation however not necessarily holds with interest rates as observed in financial markets. An interest rate swap, for example, exchanges the cash flow from time varying interest rates on a nominal amount against a stream of fixed interest rate payments on the same nominal amount, neither direct lending nor borrowing occur. The fixed rate is called the swap rate, LIBOR is typically used as the floating rate. As LIBOR is the interest rate for inter-bank lending, by the above described framework LIBOR is positive. A fair swap contract therefore requires a positive fixed swap rate as well.
Bonds are traded assets which can be interpreted as securitized accounts receivable. As traded assets, their current price is subject to supply and demand. The implicit interest rate can be calculated from the current value and the (expected) cash flows. Therefore, implicit interest rates are subject to supply and demand of the underlying bond. Now if demand is high enough so that the current price of the bond is higher than the sum of its future payments, the interest rate implied by this specific bond is negative. Such a situation occurred for US treasury bills on December 9, 2008, in the aftermath of the collapse of Lehman brothers. In this case investors were willing to pay a premium to hold highly liquid treasury bills. As this was a short episode only and happened due to a special year's end effect[1] we can reasonably assume positivity for bond yields.

Negative interest rates are of major concern in applications with guaranteed returns, as in many insurance contracts. If interest rates within a term structure model are guaranteed to be positive, losses due to failure of achieving the guaranteed return are bounded, whereas with negative interest rates the gap between guaranteed and achieved returns is unbounded. Furthermore, discounting with negative interest rates increases the impact of

[1] All information from Bloomberg,
http://www.bloomberg.com/apps/news?pid=20601087&sid=aOGXsWKEI6F4

negative interest rate scenarios on current prices.
In most applications, conditional probabilities of negative interest rates are low given that current term structures are sufficiently away from zero. This justifies pricing of short- or intermediate term contingent claims using term structure models which allow for negative interest rates, see for example Brigo and Mercurio [BM01] on the Hull-White model [HW90]. If, however, current interest rates are low, the conditional probabilities of negative interest rates might be high and therefore should be considered in pricing. In these cases, term structure models which guarantee positivity are recommended. Furthermore, if we simulate interest rates over long time periods, the conditional probability of negative interest rates occurring once during the simulated paths may be substantial even if current interest rates seem sufficiently high. As this is typically the case in insurance applications, we can conclude that positivity of yields is of particular concern for insurance companies. For an overview to positive interest rate models, see [Cai04b].

Considering a possible upper bound, note that competition among lenders implies that compensation for the abstain in consumption will not get arbitrarily high. To derive an upper bound, another economic argument comes into play, based predominantly on monetary policy with respect to inflation. Rational investors should always demand interest rates above (expected) inflation rates to preserve the real value of the money lend out. Considering developed countries, we can assume that the respective central bank will keep inflation in check so that competition among lenders guarantees that interest rates will not explode.
Whereas this does not allow for a fixed upper bound of interest rates, we can derive that interest rates are "bounded in probability", which means that the probability $P(Y \geq M)$ decreases with M. Given the experience of the stagflation era and the subsequent monetary experiment, inflation rates in western countries may well reach double digit values and interest rates may likewise reach more than 20%[2], albeit with small probability.

To conclude, we see that negative interest rates are of minor concern if the term structure model is used for short-term and intermediate-term pricing and the conditional probabilities of negative interest rates are small. In case current interest rates are low or long maturities are to be simulated, we recommend interest rate models which guarantee the zero lower bound by definition.
Considering high interest rates, we can not impose a fixed upper bound for all interest rates. We therefore recommend an upper bound by probability, whereby the probability of interest rates beyond the record ones observed during the monetary experiment should be essentially zero. In particular, this implies that the question of possible extremal scenarios

[2]The maximal 3-month rate observed at month end for the US was 16% on july 1981, the maximal 10-year rate was 15.18% on august 1981.

for risk management is closely related to the question of boundedness of the yield curve within the model. Note also that viability of an implicit upper bound crucially depends on prior beliefs considering inflation and the central bank's ability to keep inflation in check.

2.1.3 Positivity

Positivity not only considers the fixed zero lower bound of interest rates, but may also be of major importance considering the behavior of very low interest rates. Due to the recent Japanese zero interest rate policy as well as low short-term interest rates worldwide in the aftermath of the 2007 financial crisis, the behavior of interest rates in the vicinity of the zero lower bound became more important. In fact, due to generally lower inflation rates in recent years and subsequently low interest rates, hitting the zero lower bound may be a recurring problem of the future[3]. Periods of prolonged low interest rates over the whole term structure – the so called *Japan scenario*[4] – are of major concern to insurance companies, since during a Japan scenario bond market returns are below the guaranteed returns of many contracts.

Historically, low interest rates were a matter of the short end of the yield curve, only. Short-term rates were low, yet at the same time the term structure was very steep. For example, in the US, the 3-month rate from November 2001 to October 2004 was below 2%, whereas the 10-year rate during that period was at least 221 basis points higher. A similar example for Germany would be 2003 to 2004 with an average 1-year rate of 2.25% and an average slope of 169 basis points. The conduct of monetary policy typically implies that the central bank sets its target rate, yet long-term rates react to a smaller extent. Interest rate cuts therefore increase the slope, whereas if the central bank increases the target rate, the slope decreases. Any term structure model estimated from historical data therefore implies that decreasing short-term rates tend to coincide with an increasing slope and increasing short-term rates tend to coincide with a decreasing slope. The Japan scenario is the main exception to this normal functioning of monetary policy. Under normal conditions, the central bank determines short-term interest rates according to the overall macroeconomic situation. If the policy instrument reaches the zero lower bound, yet the macroeconomic situation would require further interest rate cuts, the central bank has to rely on alternative instruments. Bernanke, Reinhart and Sack [BRS04] discuss two alternative monetary policies to be employed:

1. The central bank pledges to keep the policy rate close to zero for a sustained period of time. Particularly, the central bank makes clear to the market what macroeconomic situation may lead to the end of the zero interest rate policy.

[3] Note that the possibility of hitting the zero lower bound more frequently in the future was the basis for an IMF paper advocating an increased inflation target of 4%. See [BMD10].

[4] In Japan, the 10 year yield was below 2% from february 1999 at least to 2007

2. The central bank uses *quantitative easing*, thus the central bank buys long-term government bonds or MBS to inject liquidity in the economy which decreases long-term interest rates.

Both alternative approaches to monetary policy imply a Japan scenario. Committing itself to a policy rate close to zero over the time period $[0, T]$ implies forward rates close to zero and a term structure close to zero for maturities up to T by an arbitrage argument, hence the slope decreases. Buying zerobonds increases bond prices and therefore decreases the implied long-term interest rates, hence the central bank directly decreases the slope.
The zero lower bound is not only a challenge for monetary policy, but for interest rate models as well. A Japan scenario is caused by alternative instruments of monetary policy in case short-term interest rates reach the zero lower bound, yet conventional monetary policy would require the policy instrument to be cut further.

A steep slope of the term structure does not necessarily imply problems for insurance companies as increasing duration of the bond portfolio should be sufficient to achieve the guaranteed return. Furthermore, the slope can be used to forecast interest rate movements, see [Fam84], whereby a steep yield curve implies that short-end yields will rise. During a Japan scenario, increasing the duration of the bond portfolio may not be sufficient to exceed guaranteed returns. Furthermore, the flat yield curve of a Japan scenario implies that short-end yields will remain low. This implies that the insurance company might fail to achieve the guaranteed return in bond markets for a prolonged period of time.
To make matters worse, we saw that a Japan scenario is a result of alternative monetary policy instruments if the zero lower bound is reached yet further interest rate cuts were needed. Such a monetary situation can only occur due to a steep recession. Therefore, domestic equity, real estate and commodity investments might fail to produce sufficient returns as well.

We can conclude that Japan scenarios are effectively worst case scenarios for insurance companies which should be considered in risk management and pricing. We saw that Japan scenarios are a result of alternative monetary policy instruments in case the zero lower bound is reached, which highlights the special role of an explicit bound at zero for interest rates. Therefore, interest rates are required to be positive to implement the singularity in monetary policy which constitutes the Japan scenario.

2.1.4 Time steps and Jumps

Following Black and Scholes [BS73], Brownian motion and hence continuous time models became a standard in mathematical applications in finance and therefore in term structure modeling as well. Only a few discrete-time models exist, for example Black and Karasinski [BK91] – although for these models typically continuous-time specifications were derived.

In practice, we will encounter discrete time observations only so that for all practical purposes estimation and simulation are based on discrete time steps. The time steps required can be derived from the application intended. Overnight-hedging of interest rate derivatives implies daily or sub-daily observations, in general trading applications most likely require high-frequency data. Considering insurance applications, on the other side, monthly or quarterly time steps should be sufficient.

The choice of the time steps is particularly important with respect to jumps within the data. An important example for jumps in term structure data would be the arrival of new data within the market and immediate reaction of prices, see for example in [BEG97]. It is therefore frequently assumed that jump processes are required to model high-frequency term structure data. Considering monthly or quarterly data, stochastic volatility might still be a problem. First, we can assume that jumps in monthly data are rare and typically coincide with historical events such as the monetary experiment, German unification or the 2007 financial crisis. Second, we can assume that the jump distribution are asymmetric, as most jumps in term structure data reflect quick reactions of the central bank to an economic downturn, whereas fighting inflation by increasing rates follows a more gradual approach. Third, long-term rates are known to be highly persistent, which implies that jumps in term structure data are predominantly associated with short-term rates. We expect that estimation of rare, asymmetric jumps which only occur on the short end of the yield curve is extremely difficult. We therefore recommend diffusion models for insurance applications with low frequency requirements, whereas for high-frequency data we generally recommend using jump-diffusion models.

2.1.5 Computational efficiency

The question of computational efficiency of a term structure model is essentially a question of either analytically tractable formulae or numerical algorithms to derive prices of term structure contingent claims. Most term structure models in usage right now can be described as *factor models*, that is the dynamics of the whole fixed income market are described by a factor process X, typically Markovian. The prices of contingent claims at time t are therefore functions of the factor process X_t and parameters describing the payoff function of the claim. These functions are not only required for pricing, but also for estimation and calibration of the model to current or historical data. If these functions are analytical, derivation of prices is deterministic and computationally fast. In some cases, prices can be derived numerically, which may require substantial computational resources. Finally, pricing function could be approximated by Monte Carlo methods: the payoffs of the contingent claim are simulated and discounted, using simulated yields. The empirical mean of these simulations then approximates the true price. Monte Carlo methods are indeterministic and can approximate the price only for sufficiently high numbers of trials. A trade-off exists between the quality of the approximation and computational efficiency.

We can conclude that term structure models should avoid Monte Carlo methods in pricing plain-vanilla products as far as possible. Furthermore, as term structure data is likely an input for any estimation or calibration approach, we would require analytical bond pricing formulas. Other restrictions considering computational efficiency then stem from applications, obviously term structure models used for pricing of certain interest rate derivatives should allow for these derivatives to be priced as efficiently as possible.

In insurance applications, the payoff is often path-dependent, for example to reflect portfolio allocation rules. Path-dependent contingent claims typically require higher numbers of trials than contingent claims dependent on a single payoff at time T. It is often difficult or even impossible to derive closed formula or numerical approaches for these contingent claims, thus term structure models in insurance applications should allow for efficient Monte Carlo simulation of path-dependent contingent claims. State price density models are well suited for Monte Carlo approaches, whereas affine models are well suited to derive closed formula or numerical approximations to derivative prices, see for example [Fil09] or [BM01] for an overview.

2.1.6 Number of factors

In 1991, Litterman and Scheinkman [LS91] used principal component analysis to prove that the dynamics of the term structure are determined by three principal factors:

- *Level*: the overall niveau of the term structure
- *Slope*: the steepness of the term structure
- *Curvature*: the bend of the term structure

These three factors describe 94% of the dynamics of the term structure, whereby the level is the most important factor followed by the slope. We repeated Litterman and Scheinkman's derivation for figure 2.1. The left column shows the time series of the three principal components level, slope and curvature, the right column provides empirical proxies of these principal components[5]. As these three factors are by construction uncorrelated, term structure models based on a one-dimensional state vector are not able to describe the dynamics of the whole term structure realistically. We recommend at least a two-factor term structure model to catch the majority of cross-sectional dynamics, hence the model covers level and slope. If curvature is likely to have an important impact in the application, which might be the case if duration of bond portfolios prominently features within the claim to be priced, a three-factor model should be used, otherwise, an additional curvature factor might be dispensable.

[5]As the empirical proxy of the level we propose the 10-year rate. As empirical proxy of the slope we propose the 10-year rate minus the 3-month rate. The empirical proxy of curvature is taken as the 10-year rate plus the 3-month rate minus two times the 2-year rate.

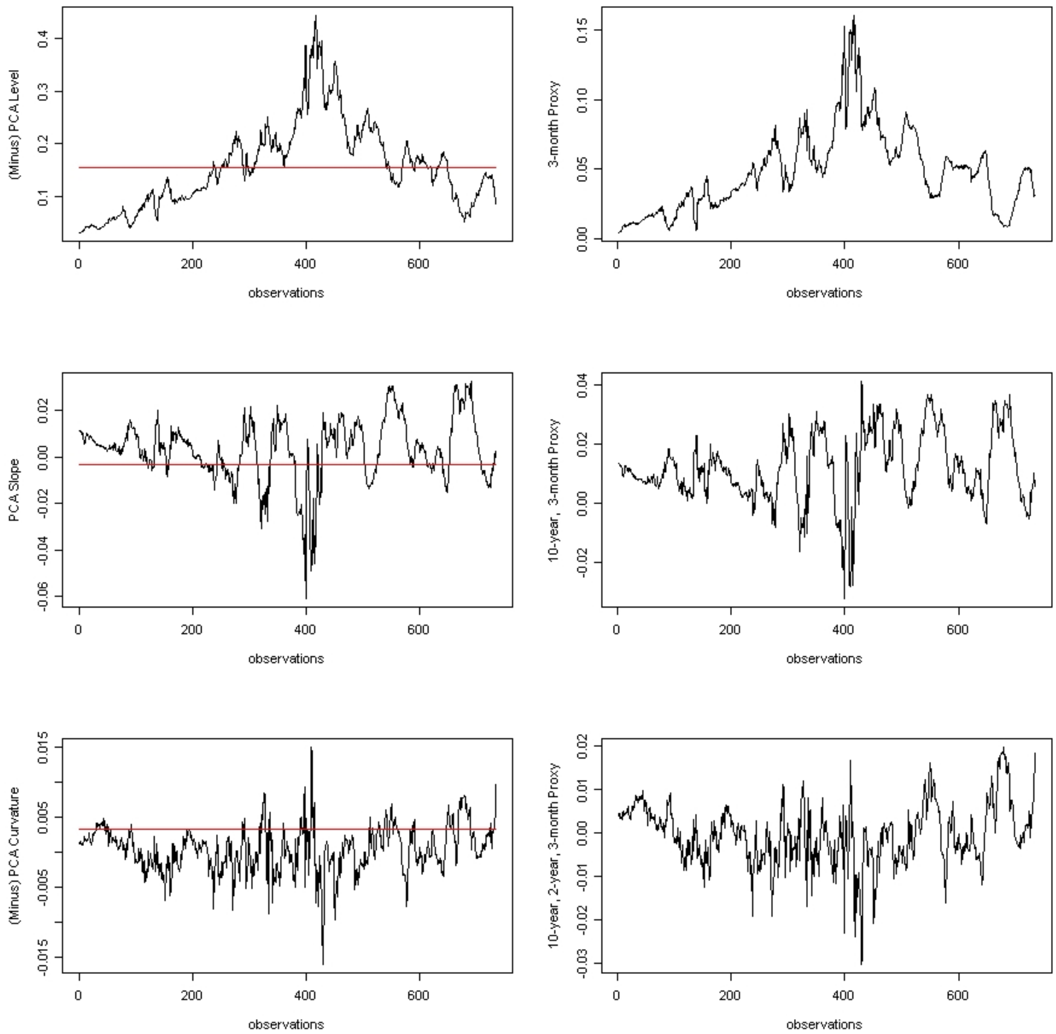

Figure 2.1: The first three factors derived by principal component analysis for US data from 1947 to 2008 (left column) with their respective empirical proxy (right column) and the mean factor (red line), from the top to the bottom level, slope and curvature. As the empirical proxy of the level we propose the 10-year rate. As empirical proxy of the slope we propose the 10-year rate minus the 3-month rate. The empirical proxy of curvature is taken as the 10-year rate plus the 3-month rate minus two times the 2-year rate.

Litterman and Scheinkman derived their three principal components from term structure data. The impact of these three factors on interest rate derivatives may differ from their impact on the term structure. Collin-Dufresne and Goldstein [CDG02] find that swap rates have only limited explanatory power for the returns of at-the-money straddle-portfolios, for which volatility in the swap rates is an important factor. This implies that the three principal factors of Litterman and Scheinkman which determine most of the dynamics of swap rates do not describe sufficiently the price dynamics of these portfolios. Collin-Dufresne and Goldstein call this *unspanned stochastic volatility*. A possible explanation would be that the unspanned factor is a principal component with minor effect on the term structure, yet deep impact for certain derivatives. We can conclude that in certain pricing applications, notably in interest rate derivatives dependent on stochastic volatility, factors unspanned by level, slope and curvature are needed. This, in turn, implies that models to price interest rate derivatives might well deviate from the above recommendation.

Note that a frequent question in term structure models is the ability to fit the different shapes of the term structure historically observed. In particular, inverted, normal, flat and hump-shaped curves were observed. Typically, even one factor models are able to produce normal, flat and inverted yield curves. However, they are obviously unable to vary level and slope at the same time, which implies that at a given point in time, any given level implies a unique slope and vice versa. In particular, one factor models typically can not produce high and normal or low and inverted yield curves. A good example for this would be the Hull-White Model [HW90], see also [Fil09] or [BM01] for a general overview, in which every shortrate $r(t)$ implies a fixed shape of the term structure, reflecting the fact that only one stochastic factor is available. In particular, such a model can neither produce very high normal yield curves nor very low inverted yield curves. Consequently, the ability to produce a wide variety of yield curves is closely related to the number of stochastic factors. To produce normal and inverted yield curves at varying height, a level and a slope factor are required. The varying Hump-shapes observed depend on curvature and maybe even on an additional factor describing the very short end.

In insurance applications, correct bond market dynamics and correct discounting functions are of primary concern. For the necessity of multi-factor models in actuarial applications see also Fischer, May and Walther [FMW03] and [FMW04]. The term structure enters pricing of insurance products in two ways: as discounting function of all payoffs and to describe bond market investments. Now considering the discounting function, if the level factor is matched realistically, this should be sufficient to discount future cash flows. Considering the investment return, however, the model must be able to realistically simulate returns on bond portfolios with varying duration, hence again at least level and slope are required. The more important the bond market is for investment success, the more realistically should cross-sectional behavior be implemented, which sooner or later includes curvature as well, and hence the more factors should be used.

In banking applications, our primary focus lies on fixed income derivatives. For these, as discussed above, the question of unspanned factors is of major importance. Nevertheless, fitting the underlying should already require a level and possibly a slope factor. We can conclude that multi-factor models are important for banking applications as well, yet the numbers of factors and their respective interpretation depends on the derivative to be priced. In particular, we expect that slope and curvature are of lesser importance to banking applications, whereas stochastic volatility should be a prominent driving factor for many derivatives.

2.1.7 Mean Reversion

As we can see in figure 2.1, the principal components of Litterman and Scheinkman showed mean reverting behavior, again under the physical measure. Mean reversion is a major ingredient in term structure models as it implicitly guarantees upper and lower boundedness by probability as well as correct long- and medium term dynamics. If we take, for example, 10-year yields as proxy for the level factor and the difference between 10-year and 1-year rates as proxy for the slope, then in case of a high slope, the slope mean reverts faster than the level factor, thus the slope decreases predominantly due to short-end movement, as found in reality.

It can be regarded a stylized fact that in multi-factor models the state process components coincide with empirical approximations to level, slope and curvature. This points to the question which observed yield best approximates the level factor. If the level factor coincides with a short-end yield, highly persistent long-end yields are functions of highly volatile level and slope factors. As long-end yields are persistent, in many term structure models long-end yields are a result of high-volatility factors downscaled in volatility. In such models, long-end volatility is frequently scaled down excessively, which implies that real data shows excess volatility in comparison to model-implied long-term yields, see Gürkaynak, Sack and Swanson [GSS03].

A second problem arises due to mean reversion of the level factor in case of long-term simulations. Historically, US long-term rates increased from around 3% following world war II up to almost 20% during the monetary experiment 1979 to 1982. Subsequently, a falling trend can be observed in long-term yields which may continue even today. Consequently, interest rates reverted both from very low and very high observations, yet mean reversion is very slow. As short-term interest rates varied around these long-term rates historically, short-term rates show strong mean reversion to long-end yields, yet only very weak mean reversion over long data sets. Only in short- or medium-term applications, persistency of long-end yields implies that short-end yields show considerable mean reversion. As a consequence, we recommend less volatile long-end yields as an empirical proxy for the level factor and we recommend term structure models which implement the principal components of Litterman and Scheinkman by a low mean reversion (long-term) level factor and high mean-reversion slope and curvature factors. This might, however, imply considerable problems in estimating the long-term mean of the level factor, something we will indeed encounter in section 2.3.8.

2.1.8 The macroeconomic role of interest rates

It is well known that the term structure is the major link between macroeconomy and finance. The question is whether macroeconomic information should be used in estimation and whether macroeconomic information could be used to determine cross-sectional and dynamical behavior of the term structure.
Several stylized facts emerged in the literature:

- The slope is related to monetary policy, as the central bank sets its target rate and intervenes in the overnight interbank market to enforce the target rate on the very short end of the riskless interbank term structure. By no-arbitrage the short end of the government bond implied yield curve and the target rate are highly correlated.
- The slope is related to the business cycle, as the monetary authority sets its policy rate according to inflation expectations and economic growth, as for example explained in the Taylor rule, see 3.2.
- Long-term yields are determined by inflation expectations, based on the previously discussed behavior of investors demanding a real compensation for their abstain in consumption. Rational investors therefore demand interest rates higher than expected inflation for the respective time to maturity.

A first benefit of incorporating macroeconomic data stems from the improved forecasting ability, hence overall improved dynamical properties of the model to be examined. This is particularly true for factor models driven by a Markovian state process X. In these cases, macroeconomic information provides additional information about the current and future conduct of monetary policy and hence about current and future term structures. To give a simple example, assume two points in time t_1, t_2 at which we observe approximately the same yield curves $Y_{t_1} \approx Y_{t_2}$. In a standard factor model, this should imply $X_{t_1} \approx X_{t_2}$. Yet now we assume that for the inflation rate $I(t_1) \geq \bar{I} \geq I(t_2)$ and for output growth we have $O(t_1) \leq \bar{O} \leq O(t_2)$. Now since at time t_1 inflation is higher than its long-term average $\bar{I}$ and output is lower than its long-term average $\bar{O}$, the central bank will cut interest rates, thus the slope will increase. At time t_2, the opposite will occur. We see that current inflation rate and output growth reasonably forecast term structure dynamics, assuming certain policy rules by the central bank. This is due to the fact that macroeconomic indicators contain information about past and future term structures so that the information contained in the current state is increased whereby at the same time

the Markovian framework is kept.
Second, if several assets are to be priced, macroeconomic indicators contain information about inter-market dependencies. To give an example, a recession typically implies falling stock prices, yet also decreasing interest rates so that bond prices rise. As long-term modeling and multi-asset frameworks are typical for insurance applications, macroeconomic indicators should be particularly helpful in these cases.
Additionally, insurance applications typically have to consider model parameters considering customer behavior, particularly cancellation. Obviously, cancellation should be negatively correlated with overall economic growth. Third, many pension and life insurance contracts offer guaranteed returns. Inclusion of macroeconomic variables might offer new product specifications for insurance companies, for example guaranteeing real returns rather than nominal returns or indexation of invalidity insurance payments on inflation.

2.1.9 Volatility

Litterman and Scheinkman [LS91] found three factors dominating term structure dynamics. Notably, volatility was not among them. Christiansen and Lund [CL02] show that term structure volatility can explain curvature changes, which in turn points toward a more general link between stochastic volatility and curvature. Given the results of Collin-Dufresne and Goldstein [CDG02] however, there seems to exists an additional, unspanned factor describing volatility which therefore is not linked to curvature. It is well known that derivatives depend on the volatility of the underlying, see for example [BM01] or [Sad09]. We therefore recommend stochastic volatility for term structure models used in pricing of standard term structure derivatives. In insurance applications, however, the main focus lies on the term structure and its dynamics, so that by definition the unspanned factor of fixed income volatility should be of minor interest.

Besides an additional factor driving volatility, note that stochastic volatility might also be implemented endogenously. If a factor model is used, Ornstein-Uhlenbeck processes imply a constant spot volatility of the factors. Cox-Ingersoll-Ross processes as factors on the other side imply stochastic spot volatilities. Now if level, slope and curvature are described by CIR processes, this should imply some sort of stochastic volatility for model-implied yields as well. We therefore recommend at least the possibility to implement a given factor model with CIR processes to test on the impact of stochastic volatility.
In either case, an important aspect for practitioners is the possibility to consider volatility shocks. To implement these in factor models, note that both OU and CIR processes include constant parameters scaling spot volatilities which can be used to implement volatility shocks.

In insurance applications, as already discussed previously, using monthly or quarterly time steps decreases the impact of volatility to the ability of the model to produce overall variability of yields and shapes of the yield curve comparable to what is observed historically. Specifically, the question of stochastic volatility becomes a question of the range of level, slope and curvature. The overall dynamics of the principal components are more important than their respective stochastic volatility.

2.1.10 Summary

To summarize, we find only a handful of properties which are likewise required in both insurance and banking applications. These are no-arbitrage and multiple, mean reverting factors driving the model. On the other side, we found plenty of differences between banking and insurance applications. In particular, these two differ in their required times to maturity. As insurance applications have longer time horizons, they imply different requirements considering long-term variability, cross-sectional behavior, positivity and extremal scenarios during lifetime. Furthermore, path-dependence is typical for insurance contingent claims, which first implies that in most cases closed-form solutions are not available and second has special requirements in simulation-based approaches. Banking applications, on the other side, are subject to short-term and high-frequency aspects, which implies that jumps and more generally stochastic volatility has to be considered. Finally, we find that cross-asset requirements show significant differences: insurance applications typically require additional financial markets to be included consistently to reflect investment policy of the insurance company. Banking applications are typically restricted to the fixed income market, but contain various fixed income derivatives, which may differ substantially. Therefore, insurance applications have to cover several financial markets, but within these markets typically only basic assets such as bonds and stock. Banking applications on the other side have to consider multiple assets within the same market, in particular (multiple) interest rate derivatives. Of special interest in these cases are unspanned factors driving only certain derivative markets but not or only to a minor extent the underlying. As insurance applications are predominantly interested in these underlyings, such unspanned factors are of minor concern. Table 2.1.10 will provide a general overview.

Properties	Insurance	Banking
No-Arbitrage	important	important
Boundedness	important	important
Positivity	important	Depends on current situation
Japan-Scenario	important	Depends on current situation
Long times to maturity	important	unimportant
Mean Reversion	important	important
Path dependence	important	depends on application
Closed bond pricing formula	crucial	important
Closed derivative pricing formula	less important	crucial
Number of factors	important	important
Unspanned factors	unimportant	crucial
Stochastic volatility	less important	important
Jumps	unimportant	important
High-frequency data	unimportant	important
Calibration	important	crucial
Macroeconomy	important	unimportant

Table 2.1: Comparison of various properties and their importance for

2.2 Two term structure models

2.2.1 The general framework of Rogers

Typically, definition of a term structure model is based on the specification of an underlying stochastic driver X and a mapping g which links the state X_t at time t to the respective term structure Y_t. In standard shortrate models, this is done in two steps: first, the state process X is mapped into a one-dimensional process $r : \mathcal{X} \to \mathbb{R}$, which is interpreted as the shortrate under the risk-neutral measure $\mathcal{Q}$. Interest rates of higher maturities are then calculated by the standard formula of risk-neutral pricing

$$P(t,T) = E^{\mathcal{Q}} \left[e^{-\int_t^T r(X_s)ds} \middle| \mathcal{F}_t \right], \tag{2.1}$$

where the expectation is conditional on a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ with $\mathcal{F}_t := \sigma(X_s,\ s \leq t)$ and taken under the risk-neutral measure $\mathcal{Q}$. We face the dual problem that for once $r(X_t)$ has to fulfill the empirical properties of the shortrate, for example mean reversion and boundedness, but at the same time we require (2.1) to provide a closed solution. Then, in a second step, these closed solutions to bond prices must fulfill certain empirical properties as well. It is difficult to specify a shortrate which solves both initial problems, not to speak about catching all empirical properties of the bond market.

Rogers [Rog97] presents an alternative framework for the construction of term structure models. This framework is based on the choice of a positive function $f : \mathcal{X} \to (0, \infty)$, which, together with a parameter α and a stochastic state process $X = (X_t)$, a continuous-time Markov process, is sufficient to define a complete term structure model. Rogers does not impose any restrictions on the state process. However, as we will see in 2.2.3, the empirical dynamics of the term structure imply certain properties that we have to acknowledge in our choice. Unlike typical shortrate models, the Rogers framework is not based on risk-neutral pricing. Instead, the so-called *state price density* is defined by

$$\varsigma_t = e^{-\alpha t} f(X_t), \tag{2.2}$$

and the dynamics of the state vector (X_t) are given under the so called *reference measure* $\tilde{\mathcal{P}}$. With the state price density defined, the current price $C(t)$ of a contingent claim at time t which pays $C(T)$ at time $T > t$ is given by

$$C(t) \quad = \quad \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T C(T)|X_t\right]}{\varsigma_t} = e^{-\alpha(T-t)} \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)C(T)|X_t\right]}{f(X_t)}. \tag{2.3}$$

where the expectations are conditional on the state X_t and evaluated under the reference measure, which therefore is the measure used in pricing based on state price densities. In section 2.2.7, we will see that the reference measure $\tilde{\mathcal{P}}$ must be equivalent to the risk-neutral measure $\mathcal{Q}$ to guarantee no-arbitrage of the bond market as well as the physical measure $\mathcal{P}$ to allow for estimation and forecasting. A major assumption frequently used later is that expectations $E^{\tilde{\mathcal{P}}}[\varsigma_T|X_t]$ under the reference measure exist for all $0 \leq t \leq T$. Interpreting zerobonds as contingent claims with payoff $C(T) := 1$ at the time of maturity T yields the following theorem.

Theorem 2.2.1. *Within the Rogers framework, the price of a zerobond at time t which pays 1 at maturity T is given by*

$$P(t,T) \quad = \quad \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right]}{\varsigma_t} = e^{-\alpha(T-t)} \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}$$

Proof. A zerobond is a derivative with payoff $C(T) = 1$ at time T. The definition of the state price density in (2.3) provides the formula. □

We see now that existence of expectations $E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right]$ is a necessary requirement for closed form bond prices.
In the literature, the Rogers framework is often called the *potential approach*. If we assume that $E^{\tilde{\mathcal{P}}}\left[\varsigma_t|X_0\right] = P(0,t) \to 0$ for $t \to \infty$, a natural assumption in bond pricing, then (ς_t) is a *potential*, which coined the name.

Corollary 2.2.2. *Within the Rogers framework, interest rates $y(t,T)$ at time t with time to maturity $T-t$ are given by*

$$y(t,T) \quad := \quad \alpha - \frac{1}{T-t} \log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right). \tag{2.4}$$

Proof. By definition of zerobond yields

$$y(t,T) = -\frac{\log(P(t,T))}{T-t}.$$

□

We can also derive general formulae for instantaneous forward rates and the shortrate.

Theorem 2.2.3. *Within the Rogers framework, instantaneous forward rates are given by*

$$f(t,T) := \alpha - \frac{\frac{\partial}{\partial t} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]} \tag{2.5}$$

and the shortrate is given by

$$r_t = \frac{(\alpha - G)f}{f}(X_t)$$

where G is the infinitesimal generator[6] *of the state process X.*

Note that the definition of shortrates and instantaneous forward rates therefore requires the function f to be twice continuously differentiable to guarantee that the infinitesimal generator is defined.

Proof. Obviously, the state vector process X_t does not depend on the payment date T, hence by definition,

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\log(P(t,T)) \\ &= -\frac{\partial}{\partial T}\left(-\alpha(T-t) + \log\left(E^{\tilde{\mathcal{P}}}[f(X_T)|X_t]\right) - \log(f(X_t))\right) \\ &= \alpha - \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}[f(X_T)|X_t]}{E^{\tilde{\mathcal{P}}}[f(X_T)|X_t]}. \end{aligned}$$

The shortrate can be derived as

$$\begin{aligned} r_t = f(t,t) &= \alpha - \frac{E^{\tilde{\mathcal{P}}}\left[Gf(X_t)|X_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_t)|X_t\right]} \\ &= \alpha - \frac{Gf(X_t)}{f(X_t)} \\ &= \frac{(\alpha - G)f}{f}(X_t). \end{aligned}$$

□

Corollary 2.2.4. *If f is twice continuously differentiable and has compact support, instantaneous forward rates of the Rogers model are given by*

$$f(t,T) = \alpha - \frac{E^{\tilde{\mathcal{P}}}\left[Gf(X_T)|X_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}.$$

[6]See appendix A.

Proof. With Dynkin's formula[7], we get

$$\alpha - \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}[f(X_T)|X_t]}{E^{\tilde{\mathcal{P}}}[f(X_T)|X_t]} \quad = \quad \alpha - \frac{E^{\tilde{\mathcal{P}}}\left[Gf(X_T)|X_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}. \tag{2.6}$$

□

We see that the Rogers framework allows for definition of nominal interest rates and forward rates of arbitrary maturities once the function f and the state vector process X is specified. The only restrictions so far are that f must be positive and twice continuously differentiable on the state space $\mathcal{X}$ of X and α must be positive as well. Nevertheless, to specify a model within the general framework, any additional information which might restrict the choice of f may be helpful. Section 2.2.3 will try to present such restrictions. The next section will discuss advantages of state price density models over the standard risk-neutral pricing.

2.2.2 Risk-neutral pricing and the state price density

In this subsection, we will motivate the usage of state price density models rather than the shortrate models predominantly used in banking. The standard approach in finance to price a random future payoff Y at time T under stochastic interest rates is based on the expected value under the risk neutral measure of the discounted value of Y. Therefore, the price P_t at time t of a contingent claim paying Y at time T is given by

$$P_t \quad = \quad E^{\mathcal{Q}}\left[e^{-\int_t^T r_s ds} Y|\mathcal{F}_t\right]. \tag{2.7}$$

where Y is $\mathcal{F}_T$-measurable. Obviously, with a stochastic shortrate process (r_t) it is often difficult to derive a closed form solution to this expression, particularly if the payoff Y depends on the shortrate process, as is for example the case in life insurance. In many cases, the so called T-forward measure $\mathcal{Q}^T$ may be applied. The expectation with respect to the T-forward measure of a bounded $\mathcal{F}_T$-measurable random payoff Y is given by

$$E^{\mathcal{Q}^T}[Y|\mathcal{F}_T] \quad = \quad \frac{E^{\tilde{\mathcal{P}}}[Y\varsigma_T|\mathcal{F}_t]}{E^{\tilde{\mathcal{P}}}[\varsigma_T|\mathcal{F}_t]}.$$

For the payoff Y at time T and the price under the reference measure we therefore have

$$\begin{aligned} P_t &= \frac{E^{\tilde{\mathcal{P}}}\left[Y\varsigma_T|\,\mathcal{F}_t\right]}{\varsigma_t} \\ &= \frac{E^{\mathcal{Q}^T}[Y|\mathcal{F}_T]E^{\tilde{\mathcal{P}}}[\varsigma_T|\mathcal{F}_t]}{\varsigma_t} \\ &= P(t,T)E^{\mathcal{Q}^T}[Y|\mathcal{F}_T] \end{aligned}$$

In many applications, the T-forward measure significantly simplifies pricing. Particularly, if simulation-based approaches are required and the payoff depends on the interest rates

[7] See A.

and hence the shortrate path over $[t, T]$, as is typically the case with insurance applications, the T-forward measure simplifies simulation considerably as path dependence in simulation may vanish. The same is implicitly the case with the state price density. Using the state price density, the price at time t of our random payoff Y is given by

$$P_t = \frac{E^{\tilde{\mathcal{P}}}[Y\varsigma_T|\mathcal{F}_t]}{\varsigma_t}.$$

The main advantage over risk-neutral pricing obviously is that we do not have to consider path-dependence over $[t, T]$ if the payoff Y depends on the term structure at time T. If path-dependent payoffs occur, as is typically the case in insurance applications due to portfolio allocation decisions or premia payed, the state price density allows for simple implementation as well. To discount a cash flow payed at various payment dates $T_1, \ldots, T_n$, we simply require intermediate state price densities which can be constructed from the states $X_{T_1}, \ldots, X_{T_n}$. Note that we are not restricted to equidistant payment dates, nor do we have to approximate the discounting functions. In contrast to that, simulation-based risk-neutral pricing always has to rely on approximated shortrate paths, and we are generally forced to simulate the shortrate at additional points in time besides $T_1, \ldots, T_n$. To give an example: for an insurance contract which provides the first payoff in 30 years not a single intermediate state between today and the first payoff date in 30 years is required in the state price density approach. For the risk-neutral approach, however, monthly time steps require simulation of 359 intermediate states, yet provide only an approximation to the true discounting function.
Comparing the state price density approach to the T-forward measure, note that $P(t, T)$ must be taken from market data. Typically, however, for times to maturity $T-t$ larger than 10 years market data does not exist or is rather unreliable due to liquidity constraints. For long time horizons, as we encounter in life and pension insurance, the state price density approach is therefore superior to the pricing approach using the T-forward measure as well.

2.2.3 Restricting the choices in the general framework

The general framework of Rogers is extremely flexible considering the choices of the state vector X or the function f. Nevertheless, in section 2.1 we found several criteria term structure models must fulfill. These criteria might in turn restrict the choices of f and X in Rogers' generic approach. In this section, we try to identify such restrictions. Of particular importance are the following points from section 2.1:

1. Mean reversion and
2. Boundedness of interest rates
3. Availability of analytical bond pricing formulae
4. No-arbitrage

5. Positivity of interest rates

6. The behavior of the long end of the term structure.

Note that the empirical properties we use to derive restrictions of the Rogers model a priori only apply to the dynamics under the physical measure, not necessarily under the reference measure. However, as we will see in section 2.2.7, the reference measure must be equivalent to the risk-neutral measure to guarantee no-arbitrage. By choosing a market price of risk we then construct the physical measure as a further equivalent measure to both the reference and the risk-neutral measure. As all three measures are equivalent, they have the same null sets. If we can describe certain empirical properties of term structures by null sets under the physical measure, the same properties hold under the other equivalent measures as well.

The state process X

Rogers does not restrict the state process X. However, empirical research as well as implementation practice with various interest rate models allow several assumptions about the state process. First, according to Litterman and Scheinkman [LS91], principal component analysis showed that the first three components, level, slope and curvature, explain about 97% of the dynamics of the term structure. As a change of measures does not change the dimensionality of the underlying process, Litterman and Scheinkman's work implies that a multifactor model is required to catch the dynamics of the whole term structure.

A major requirement for interest rate dynamics under the physical measure was mean reversion. The main idea behind mean reversion is that we expect any historically observed term structure to reemerge with positive probability, hence the probability of a certain term structure to emerge only once is a null set. Therefore, mean reversion of interest rates must hold under all equivalent measures.

Analytical pricing formulae

A major criterion on term structure models from a practitioners point of view is the availability of closed form solutions to zero bond and interest rate derivative prices, as stated in 2.1. Given the bond pricing formula

$$P(t,T) \quad = \quad e^{-\alpha(T-t)} \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}$$

the critical question is whether the conditional expectation $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]$ can be calculated analytically. Given a state process X and its conditional distribution of $X_T|X_t$, it is often straightforward to decide whether closed-form bond prices are available or not[8].

[8]Note that the conditional distribution of $X_T|X_t$ is available for most choices of state vector processes, such as Brownian motion, finite state space Markov chains in discrete time, Ornstein-Uhlenbeck and Cox-Ingersoll-Ross processes.

We can demonstrate this using the examples presented by Rogers [Rog97]. If we require X to be an Ornstein-Uhlenbeck process, we know the conditional distribution of $X_T|X_t$ to be normal. The first example of Rogers defines $f(x) := \exp(a^T x)$. As $a^T X_T|X_t$ follows a normal distribution, the closed form solution of bond prices is based on the first moment of a lognormal distribution. Example 2 sets $f(x) := \exp(\frac{1}{2}(x-c)^T Q(x-c))$ and the bond pricing function uses results of [LW99] about expected values of functions of normally distributed random variables. Example 3 defines $f(x) := \gamma + \frac{1}{2}(x-c)^T Q(x-c)$ and the bond pricing formula is based on the second moment of the normal distribution. The fourth example, which we will examine below, defines

$$f(x) := \cosh(\gamma^T x + c) = \frac{\exp(\gamma^T x + c) + \exp(-\gamma^T x - c)}{2},$$

which also uses the first moment of a lognormal distribution.

Whether additional closed formulae of contingent claim prices are required is a question of the purpose the model is developed for. Due to asymmetry in payoff functions, deriving closed form solutions to derivative prices requires more elaborate examinations and may not provide a simple criterion on a given choice of f and X.

No-arbitrage

In section 2.1, we discussed why no-arbitrage as a basic consistency criterion is important in term structure modeling. In his derivation of the state price density framework, Rogers assumes existence of a risk-neutral measure and hence no-arbitrage. In his generic approach, however, Rogers specifies the state price density given a function f and the dynamics of the state vector under the reference measure $\tilde{\mathcal{P}}$. Existence of the risk-neutral measure is thus not guaranteed and therefore no-arbitrage has to be proved for any specified model. In case the state price density is a supermartingale, Rutkowski [Rut97] provides the following theorem.

Theorem 2.2.5. *If the state price density (ς_t) is a strictly positive supermartingale, the bond market defined by ς_t is arbitrage free.*

Proof. See Rutkowski [Rut97], proposition 1, page 154. □

If the state-price density is no supermartingale, no-arbitrage has to be proved. We will demonstrate the standard approach in section 2.2.7. The idea is to take bond price dynamics under the reference measure, and then to construct a drift correction term which provides an equivalent measure under which discounted asset prices are martingales. The resulting measure is therefore a risk-neutral measure. This may be a difficult and time consuming task. Note furthermore that the derivation of bond price dynamics under the reference measure might be impossible if closed form solutions to bond prices are not available.

Positivity

Besides analytical bond pricing formulas, another important requirement of interest rate models is positivity of the resulting yields. Empirically, negative interest rates should not occur, as derived in 2.1. Because the probability of negative interest rates under the physical measure is zero, interest rates are positive under the other equivalent measures as well. The following theorem links positivity of interest rates to the supermartingale property of the state price density.

Theorem 2.2.6. *Suppose (ς_t) is a positive state price density process. In the Rogers framework, interest rates for arbitrary maturities are always non-negative if and only if the state-price density is a positive supermartingale under the reference measure $\tilde{\mathcal{P}}$.*

Proof. (ς_t) is a positive supermartingale, if and only if for all $t, T \in \mathbb{R}^+$, $t \leq T$

$$\begin{array}{rccc} & E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right] & \leq & \varsigma_t \\ \overset{\varsigma_t>0}{\Leftrightarrow} & \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right]}{\varsigma_t} & \leq & 1 \\ \Leftrightarrow & P(t,T) & \leq & 1 \\ \Leftrightarrow & -\frac{1}{T-t}\log\left(P(t,T)\right) & \geq & 0 \\ \Leftrightarrow & y(t,T) & \geq & 0 \end{array}$$

hence interest rates are non-negative. □

A simple corollary can be derived based on Rutkowski's theorem 2.2.5 regarding no-arbitrage.

Corollary 2.2.7. *Suppose (ς_t) is a positive state price density process. If interest rates of a Rogers framework are always positive, the model is free of arbitrage.*

Proof. By theorem 2.2.6, positivity of all interest rates implies the state price density to be a positive supermartingale. By theorem 2.2.5, the model is free of arbitrage if the state price density is a positive supermartingale. □

A rather simple approach to guarantee that ς_t is a positive supermartingale would be to require $(f(X_t))$ to be a martingale or a positive supermartingale. In this case, the state price density is a supermartingale as

$$E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right] = e^{-\alpha T}E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right] \leq e^{-\alpha t}f(X_t) = \varsigma_t$$

holds in either case. However, if $(f(X_t))$ is a martingale, then

$$y(t,T) = \alpha - \frac{1}{T-t}\log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right) = \alpha - \frac{\log(1)}{T-t} = \alpha,$$

and hence all interest rates are equal to α. We can conclude the following lemma.

Lemma 2.2.8. *In a Rogers model, $(f(X_t))$ must not be a martingale.*

In the next section, we will find a simple condition under which $(f(X_t))$ might be a supermartingale.

The long end of the term structure

Given the yield formula of the Rogers model in 2.2.2, we know that the parameter α is influential for the asymptotic long rate of the yield curve $\lim_{T\to\infty} y(t,T)$. In particular, section 2.3.9 will show that the asymptotic long rate for all practical purposes is constant and equal to α. By now, we can derive the following general lemma.

Lemma 2.2.9. *In a Rogers model with $(f(X_t))$ being a supermartingale, $\alpha = 0$ and α is a lower bound for the asymptotic long rate.*

Proof. Choosing f so that $(f(X_t))$ is a supermartingale implies that for all $0 \leq t \leq T$

$$\begin{aligned}
f(X_t) &\geq E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right] \\
1 &\geq \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)} \\
0 &\geq \frac{1}{T-t}\log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right) \\
0 &\leq -\frac{1}{T-t}\log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right) \\
\alpha &\leq \alpha - \frac{1}{T-t}\log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right) = y(t,T).
\end{aligned}$$

As we require that interest rates may get arbitrarily close to 0, this effectively implies $\alpha = 0$. In section 2.3.9, we will find that α, dependent on the current state X_t, either equals the asymptotic long rate, is an upper bound of it or a lower bound. From economic reasons, we require that yield curves are continuous functions of the time to maturity. Therefore, α is either a lower bound of the asymptotic long rate as well or α equals the asymptotic long rate. Because $\alpha = 0$ if $(f(X_t))$ is a supermartingale, α cannot be equal to the asymptotic long rate as this would imply investors to require no compensation for lending money on the very long term, a contradiction to economic rationality. Therefore, α is a lower bound for the asymptotic long rate. □

A similar lemma based on economic reasoning can be found for convex functions f.

Lemma 2.2.10. *If the function $f : \mathcal{X} \to \mathbb{R}^+$ is convex, and the state vector process (X) is mean reverting to μ so that $E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] \to \mu$ for $T \to \infty$, then α is the upper bound of the asymptotic long rate.*
If the function f is concave, α is the lower bound of the asymptotic long rate.

Proof. Let $f : \mathcal{X} \to \mathbb{R}^+$ be a convex function. By Jensen's inequality,

$$E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right] \quad \geq \quad f\left(E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]\right) \to f(\mu)$$

since f is continuous by definition. Thus

$$\begin{aligned}\lim_{T\to\infty} y(t,T) \quad &= \quad \alpha - \lim_{T\to\infty} \frac{1}{T-t} \log\left(\frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]}{f(X_t)}\right) \\ &\leq \quad \alpha - \lim_{T\to\infty} \frac{1}{T-t} \log\left(\frac{f\left(E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]\right)}{f(X_t)}\right) \to \alpha\end{aligned}$$

The proof for concave f is analogously. □

These lemmata may imply a conflict for many term structure models if Rogers generic approach is used and $g(X_t) := \alpha - Gf(X_t)$ is not guaranteed to be positive. In this case, higher α decreases the probability of negative short rates. If f is convex, this also increases the upper bound of the asymptotic long rate. If f is concave, this increases the lower bound of the asymptotic long rate. In section 2.3.9, we will see that according to theorem 2.3.3 α implies a bound for the asymptotic long rate $\lim_{T\to\infty} y(t,T)$ depending on the limiting behavior of $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ for arbitrary $x \in \mathcal{X}$.

2.2.4 The Rogers framework and the Flesaker-Hughston framework

Within the general framework of Rogers, we will compare two special cases: the cosh model proposed by Rogers himself and the Cairns model [Cai04a], which is a special case of the framework of Flesaker and Hughston [FH96], but which can be defined in terms of the Rogers framework as well.

Flesaker and Hughston [FH96] proposed a general framework to define term structure model which guarantees positivity of all yields. They start with defining the bond price by

$$P(t,T) := \frac{\int_T^\infty M(t,s)\phi(s)ds}{\int_t^\infty M(t,s)\phi(s)ds} \tag{2.8}$$

where ϕ is a deterministic function. $M(t,s)$ for $0 \leq t \leq s < \infty$ is a family of strictly positive diffusion processes over the index s which are martingales with respect to t under the reference measure $\tilde{\mathcal{P}}$. Rutkowski [Rut97] defined $A_t := \int_t^\infty \phi(s)M(t,s)ds$, which leads to the bond pricing formula

$$\begin{aligned}P(t,T) \quad &= \quad \frac{\int_T^\infty M(t,s)\phi(s)ds}{\int_t^\infty M(t,s)\phi(s)ds}, \\ &\overset{Fubini}{=} \quad \frac{E^{\tilde{\mathcal{P}}}\left[\left.\int_T^\infty \phi(s)M(T,s)ds\right|\mathcal{F}_t\right]}{\int_t^\infty \phi(s)M(t,s)ds} \\ &= \quad \frac{E^{\tilde{\mathcal{P}}}\left[A_T|X_t\right]}{A_t}\end{aligned}$$

demonstrating the relation between Rogers state price density approach and the bond price model of Flesaker and Hughston. In particular, A_t is a strictly positive state price density. By definition of the bond price in (2.8),

$$1 \geq P(t,T) = \frac{E^{\tilde{\mathcal{P}}}\left[A_T|X_t\right]}{A_t}$$

and hence (A_t) is also a supermartingale. Rutkowski proves absence of arbitrage given that the state-price density A_t is a strictly positive supermartingale, see 2.2.5. Therefore, no-arbitrage in a Flesaker-Hughston model implicitly holds.

Whereas in Rogers generic approach we had to choose f and X so that ς_t is a supermartingale, the Flesaker-Hughston framework requires definition of a martingale $M(t,s) := M(X_t,t,s)$. In both cases the state vector process must be multidimensional and mean reverting by section 2.1. The main difference between the Rogers framework and the Flesaker-Hughston framework lies in closed bond price formulae. Such closed formulae in the Flesaker-Hughston framework require that the integrals are analytically solvable. In the Rogers framework, closed formulae require that $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t\right]$ is available in closed form. This criterion is generally easier to handle than the solvability of the integral in Flesaker-Hughston.

2.2.5 The Cairns model

The Cairns model can be introduced in two ways: Using the framework of Flesaker and Hughston, as Cairns did originally, or using the Rogers framework, which we will do later. Flesaker and Hughston define their model as a bondpricing model which guarantees positivity of all interest rates. Namely, Flesaker and Hughston define zero-coupon bond prices by

$$P(t,T) = \frac{\int_T^\infty M(t,s)\phi(s)ds}{\int_t^\infty M(t,s)\phi(s)ds}$$

where $(M(t,s))_{0\leq t\leq s<\infty}$ is a family of strictly positive diffusion processes under a reference measure $\tilde{\mathcal{P}}$. $(M(t,s))_{t\geq 0}$ is also a martingale for all $s \geq t$ and $\phi(\cdot)$ is a deterministic function. Cairns chooses the family of martingales M by

$$\begin{aligned} M(0,T) &= 1 \;\forall T \\ dM(t,T) &= M(t,T)\sigma(t,T)' dW^{\tilde{\mathcal{P}}}(t) \end{aligned} \tag{2.9}$$

where[9]

$$dW^{\tilde{\mathcal{P}}}(t) = C dZ^{\tilde{\mathcal{P}}}(t)$$

[9] We will denote correlated Brownian motion by W_t and independent Brownian motion by Z_t, respectively.

and $Z_1^{\tilde{\mathcal{P}}}(t), \ldots, Z_d^{\tilde{\mathcal{P}}}(t)$ are d independent Brownian motions under the reference measure $\tilde{\mathcal{P}}$ and the matrix C is chosen such that $CC' = ((\rho_{ij}))_{i,j=1}^d$ is an instantaneous correlation matrix with $d\left\langle W_i^{\tilde{\mathcal{P}}}(t), W_j^{\tilde{\mathcal{P}}}(t)\right\rangle = \rho_{ij}$. Now using the Ito-Doeblin formula on (2.9) implies

$$d\log\left(M(t,T)\right) \;=\; \sigma(t,T)^T dW^{\tilde{\mathcal{P}}}(t) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{n}\sigma_i(t,T)\sigma_j(t,T)\rho_{ij}dt$$

and by defining $\sigma_i(t,T) := \sigma_i \exp\left[-\kappa_i(T-t)\right]$

$$\begin{aligned}
&\log M(t,T) \\
&= \sum_{i=1}^{n}\sigma_i \int_0^t e^{-\kappa_i(T-s)} dW_i^{\tilde{\mathcal{P}}}(s) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\sigma_i\sigma_j \int_0^t e^{-(\kappa_i+\kappa_j)(T-s)}ds \\
&= \sum_{i=1}^{d}\sigma_i e^{-\kappa_i(T-t)}\hat{X}_i(t) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j}e^{-(\kappa_i+\kappa_j)(T-t)}\left(1-e^{-(\kappa_i+\kappa_j)t}\right)
\end{aligned}$$

with

$$d\hat{X}_i(t) = -\kappa_i \hat{X}_i(t)dt + dW_i^{\tilde{\mathcal{P}}}(t)$$

and $\hat{X}_i(0) = 0$. $\hat{X}_1, \ldots, \hat{X}_d$ are therefore correlated Ornstein-Uhlenbeck processes with mean reversion factor κ_i. Cairns defines the deterministic function $\phi(\cdot)$ using some parameters $\eta, \alpha, \hat{x}_1, \ldots, \hat{x}_d$ as

$$\phi(s) = \eta \exp\left[-\alpha s + \sum_{i=1}^{d}\sigma_i\hat{x}_i e^{-\kappa_i s} - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j}e^{-(\kappa_i+\kappa_j)s}\right].$$

This specification is chosen to simplify the combined integrand

$$\begin{aligned}
M(t,s)\phi(s) \;=\;& \eta\exp\Bigg[-\alpha s + \sum_{i=1}^{d}\sigma_i\hat{x}_i e^{-\kappa_i s} - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j}e^{-(\kappa_i+\kappa_j)s} \\
& + \sum_{i=1}^{d}\sigma_i e^{-\kappa_i(T-t)}\hat{X}_i(t) - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j}e^{-(\kappa_i+\kappa_j)(T-t)}\left(1-e^{-(\kappa_i+\kappa_j)t}\right)\Bigg] \\
=\;& \eta\exp\left[-\alpha s + \sum_{i=1}^{d}\sigma_i e^{-\kappa_i(s-t)}X_i(t) - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j}e^{-(\kappa_i+\kappa_j)(s-t)}\right]
\end{aligned}$$

with $X_i(t) = \hat{x}_i\exp(-\kappa_i t) + \hat{X}_i(t)$. $X_1, \ldots, X_d$ are Ornstein-Uhlenbeck processes under $\tilde{\mathcal{P}}$ with $X_i(0) = \hat{x}_i$ and $dX_i(t) = -\kappa_i X_i(t)dt + dW_i^{\tilde{\mathcal{P}}}(t)$ for $i = 1, \ldots, d$. Alternatively, one could follow the Flesaker-Hughston approach and define $\phi(s) := \frac{\partial}{\partial s}P(0,s)$, which guarantees that the model fits perfectly the current term structure. This approach resembles the calibration of the Hull-White model [HW90] to the current term structure. Both approaches share the problem that reliable data of the yield curve is only available up to maturities of 10 years, which implies an upper integration bound of the calibrated Cairns

model of only 10 years which is insufficient.
Integrating over $M(t,s)\phi(s)$, we can define a simplifying function $H : [0,\infty)\times\mathcal{X} \to (0,\infty)$,

$$\begin{aligned}
&\int_T^\infty M(t,s)\phi(s)ds \\
= \quad & \eta \int_T^\infty \exp\left[-\alpha s + \sum_{i=1}^d \sigma_i e^{-\kappa_i(s-t)} X_i(t) - \frac{1}{2}\sum_{i,j=1}^n \frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j} e^{-(\kappa_i+\kappa_j)(s-t)}\right] ds
\end{aligned}$$

with substitution $u := s - t$ this equals

$$\begin{aligned}
& \eta e^{-\alpha t} \int_{T-t}^\infty \exp\left[-\alpha u + \sum_{i=1}^d \sigma_i e^{-\kappa_i u} X_i(t) - \frac{1}{2}\sum_{i,j=1}^d \frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j} e^{-(\kappa_i+\kappa_j)u}\right] du \\
:= \quad & \eta e^{-\alpha t} \int_{T-t}^\infty H(s,X(t))ds \qquad (2.10)
\end{aligned}$$

and hence

$$\int_t^\infty M(t,s)\phi(s)ds \quad = \quad \eta e^{-\alpha t}\int_0^\infty H(s,X(t))ds,$$

whereby

$$H(u,x) \quad = \quad \exp\left[-\alpha u + \sum_{i=1}^d \sigma_i x_i e^{-\kappa_i u} - \frac{1}{2}\sum_{i,j=1}^d \frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j} e^{-(\kappa_i+\kappa_j)u}\right].$$

This provides the following theorems describing bond prices and nominal yields.

Theorem 2.2.11. *Within the Cairns model, the price of a zerobond at time t which pays 1 at maturity T is given by*

$$P(t,T) = \frac{\int_{T-t}^\infty H(s,X(t))ds}{\int_0^\infty H(s,X(t))ds}.$$

Corollary 2.2.12. *Within the Cairns model, zerobond rates $y(t,T)$ at time t with time to maturity $T-t$ are given by*

$$y(t,T) \quad := \quad -\frac{1}{T-t}\left(\log\left(\int_{T-t}^\infty H(s,X(t))ds\right) - \log\left(\int_0^\infty H(s,X(t))ds\right)\right).$$

Using these theorems, we can derive instantaneous forward rates and shortrates using standard formulae.

Theorem 2.2.13. *Within the Cairns model, instantaneous forward rates are given by*

$$f(t,T) = \frac{H(T-t,X(t))}{\int_{T-t}^\infty H(s,X(t))ds},$$

the shortrate is is given by

$$r(t) \quad = \quad \frac{H(0,X(t))}{\int_0^\infty H(s,X(t))ds}$$

Proof. The instantaneous forward rates in Cairns can be derived by

$$f(t,T) = -\frac{\partial}{\partial T}\log(P(t,T)) = \frac{H(T-t,X(t))}{\int_{T-t}^{\infty} H(s,X(t))ds}.$$

The shortrate is then given as

$$r(t) = f(t,t) = \frac{H(0,X(t))}{\int_0^{\infty} H(s,X(t))ds}$$

□

If we want to introduce the Cairns model within the Rogers framework, the starting point is the specification of the state-price density.

Theorem 2.2.14. *Within the Cairns model, the state price density process* (ς_t) *is given by*

$$\varsigma_t := \eta e^{-\alpha t}\int_0^{\infty} H(s,X(t))ds$$

for all $t \geq 0$. (ς_t) *is a positive supermartingale.*

Note that the value of η is actually irrelevant for pricing and hence omitted.

Proof. By (2.10), for $0 \leq t \leq T$

$$\eta e^{-\alpha t}\int_{T-t}^{\infty} H(s,X(t))ds = \int_T^{\infty} M(t,s)\phi(s)ds$$

and hence

$$\eta e^{-\alpha T}\int_0^{\infty} H(s,X(T))ds = \int_T^{\infty} M(T,s)\phi(s)ds$$

and

$$\eta e^{-\alpha t}\int_0^{\infty} H(s,X(t))ds = \int_t^{\infty} M(t,s)\phi(s)ds,$$

hence for $T \geq t$

$$\begin{aligned}
& \eta E^{\tilde{\mathcal{P}}}\left[e^{-\alpha T}\int_0^{\infty} H(s,X(T))ds \middle| \mathcal{F}_t\right] \\
= \; & E^{\tilde{\mathcal{P}}}\left[\int_T^{\infty} M(T,s)\phi(s)ds \middle| \mathcal{F}_t\right] \\
\overset{Fubini}{=} \; & \int_T^{\infty} E^{\tilde{\mathcal{P}}}\left[M(T,s)\middle| \mathcal{F}_t\right]\phi(s)ds \\
= \; & \int_T^{\infty} M(t,s)\phi(s)ds.
\end{aligned}$$

Therefore

$$\begin{aligned}
& \frac{E^{\tilde{\mathcal{P}}}\left[e^{-\alpha T}\int_0^{\infty} H(s,X(T))ds \middle| \mathcal{F}_t\right]}{e^{-\alpha t}\int_0^{\infty} H(s,X(t))ds} \\
= \; & \frac{\int_T^{\infty} M(t,s)\phi(s)ds}{\int_t^{\infty} M(t,s)\phi(s)ds} \\
= \; & P(t,T),
\end{aligned}$$

by definition of the Cairns model. This is the defining equation for the state price density. Since exp and H are positive functions, $e^{-\alpha T} \int_0^\infty H(s, X(T))ds$ is positive as well. Finally,

$$\begin{aligned} \eta E^{\tilde{\mathcal{P}}} \left[e^{-\alpha T} \int_0^\infty H(s, X(T))ds \middle| \mathcal{F}_t \right] &= \int_T^\infty M(t,s)\phi(s)ds \\ &\leq \int_t^\infty M(t,s)\phi(s)ds \\ &= \eta e^{-\alpha t} \int_0^\infty H(s, X(t))ds, \end{aligned}$$

whereby we used that $M(t,s)\phi(s)$ is positive for all $T \geq t$ and all s. □

2.2.6 The cosh model

As an alternative to the Cairns model, we present the fourth example of Rogers, the so called *cosh* model. The cosh model is introduced by Rogers as an example for his generic approach. The term structure model is specified by $f(x) := \cosh(x)$ and the state vector being an Ornstein-Uhlenbeck process. This defines the state price density of the cosh model by

$$\varsigma_t := e^{-\alpha t} \cosh(\gamma^T X_t + c). \tag{2.11}$$

The standard example for the Rogers framework is the exponential affine model $f(X_t) := \exp\left(\gamma^T X_t + c\right)$ with X being an Ornstein-Uhlenbeck process. This is Rogers' first example for the generic approach, it was later examined further by Leippold and Wu [LW99]. Although the cosh models seems by far more complicated, it is in fact simply a combination of affine exponential models, since $\cosh(x) = \frac{1}{2}\left[\exp(x) + \exp(-x)\right]$. In both models the state price density is not a supermartingale, reflecting the difficulties to define a model with this property. Consequently, both models allow for negative interest rates. Nevertheless, the subset of states which imply negative interest rates is obviously a halfspace in case of the exponential affine model, whereas the set of states which imply negative interest rates in the cosh model is only a subset of a halfspace. This rather geometrical argument indicates that the probability of negative interest rates should be smaller with the cosh model than with the exponential affine model. This is the main reason we preferred the cosh model to the exponential affine model.
The cosh model as well as the exponential affine model offer simple closed-form solutions of bond prices, based on moments of lognormal distributions. These closed bond prices guarantee computational efficiency for both models.

We choose the state process X to follow Ornstein-Uhlenbeck dynamics under the reference measure $\tilde{\mathcal{P}}$. Specifically, we assume

$$dX_t = \kappa(\tilde{\mu} - X_t)dt + C dZ_t^{\tilde{\mathcal{P}}} \tag{2.12}$$

where $Z_t^{\tilde{\mathcal{P}}} = (Z_t^{\tilde{\mathcal{P}},1}, \ldots, Z_t^{\tilde{\mathcal{P}}n})$ is an n-dimensional Brownian motion under the reference measure $\tilde{\mathcal{P}}$ with $Z_t^{\tilde{\mathcal{P}},i}$ and $Z_t^{\tilde{\mathcal{P}},j}$ mutually uncorrelated for $i \neq j$. The state vector components are correlated by the instantaneous correlation matrix

$$CC^T = \rho = ((\rho_{ij})),$$

hence

$$d\langle X_i(t), X_j(t)\rangle = \rho_{ij} dt.$$

This dependence structure is chosen following the previous definition of the (correlated) state vector in Cairns. The matrix κ is a $n \times n$ diagonal matrix, again as in Cairns. The long-term mean $\tilde{\mu}$ of the state process under the reference measure $\tilde{\mathcal{P}}$ is an n-dimensional vector, which in the Cairns model was implicitly taken to be zero. For the individual state vector component $X^{(i)}$ we have

$$dX_t^{(i)} = \kappa_i(\tilde{\mu}_i - X_t^{(i)})dt + \sum_{j=1}^{d} C_{ij} dZ_t^{\tilde{\mathcal{P}},j}. \tag{2.13}$$

Using the Ito-Doeblin formula, we can easily derive a solution to this stochastic differential equation.

Theorem 2.2.15. *A process with dynamics (2.13) has solution*[10]

$$X_t^{(i)} = X_0^{(i)} e^{-\kappa_i t} + \tilde{\mu}_i(1 - e^{-\kappa_i t}) + \sum_{j=1}^{d} \int_0^t e^{\kappa_i(s-t)} C_{ij} dZ_s^{\tilde{\mathcal{P}},j}. \tag{2.14}$$

Proof. We take $f(t, X_t^{(i)}) := X_t^{(i)} e^{\kappa_i t}$. Then by the Ito-Doeblin-formula

$$\begin{aligned} df(t, X_t) &= \left[\kappa_i X_t^{(i)} e^{\kappa_i t} + \kappa_i(\tilde{\mu}_i - X_t^{(i)}) e^{\kappa_i t}\right] dt + e^{\kappa_i t} \sum_{j=1}^{d} C_{ij} dZ_t^{\tilde{\mathcal{P}},j} \\ &= \kappa_i \tilde{\mu}_i e^{\kappa_i t} dt + e^{\kappa_i} \sum_{j=1}^{d} C_{ij} dZ_t^{\tilde{\mathcal{P}},j}. \end{aligned}$$

[10]We write for readability

$$e^{-\kappa(T-t)} := diag(e^{-\kappa_{ii}(T-t)})$$

and

$$1 - e^{-\kappa(T-t)} := diag\left(1 - e^{-\kappa_{ii}(T-t)}\right).$$

Note that for κ being a diagonal matrix we encounter here this notation coincides with the definition of an exponential of a matrix $exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$. In the following sections, diagonality of κ and $e^{-\kappa(T-t)}$ is frequently used.

Next we integrate this stochastic differential equation from 0 to t and get

$$\begin{aligned}
f(t,X_t)-f(0,X_0) &= \int_0^t \kappa_i\tilde{\mu}_i e^{\kappa_i s}ds+\sum_{j=1}^{d}\int_0^t e^{\kappa_i s}C_{ij}dZ_s^{\tilde{\mathcal{P}},j}.\\
X_t^{(i)}e^{\kappa_i t}-X_0^{(i)} &= \tilde{\mu}_i(e^{\kappa_i t}-1)+\sum_{j=1}^{d}\int_0^t e^{\kappa_i s}C_{ij}dZ_s^{\tilde{\mathcal{P}},j}.\\
X_t^{(i)} &= X_0^{(i)}e^{-\kappa_i t}+\tilde{\mu}_i(1-e^{-\kappa_i t})+\sum_{j=1}^{d}\int_0^t e^{\kappa_i(s-t)}C_{ij}dZ_s^{\tilde{\mathcal{P}},j}.
\end{aligned}$$

□

Using the solution of the stochastic differential equation, the following theorem provides the distribution of X_T conditional on X_t, which can directly be used to derive the transition equation for the Kalman filter.

Theorem 2.2.16. *The conditional distribution of X_T given X_t whereby (X_t) is an Ornstein-Uhlenbeck process with dynamics as described in (2.12) and (2.13) is normal with conditional mean*

$$E\left[X_T|\,X_t\right] = e^{-\kappa(T-t)}X_t+(1-e^{-\kappa(T-t)})\tilde{\mu} \tag{2.15}$$

and conditional covariance matrix

$$Cov^{\tilde{\mathcal{P}}}\left[X_T^{(i)},X_T^{(j)}\middle|\,X_t\right]=\frac{\rho_{ij}}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)(T-t)}\right). \tag{2.16}$$

Proof. As the Ornstein-Uhlenbeck processes are driven by a d-dimensional Brownian motion, we have that the distribution of X_T given X_t is (multivariate) normal. For $i=1,\dots,d$ we have, using (2.14),

$$\begin{aligned}
&E^{\tilde{\mathcal{P}}}\left[X_T^{(i)}\middle|\,X_t\right]\\
&= E^{\tilde{\mathcal{P}}}\left[X_t^{(i)}e^{-\kappa_i(T-t)}+\tilde{\mu}_i(1-e^{-\kappa_i(T-t)})+\sum_{j=1}^{d}\int_t^T e^{\kappa_i(s-T)}C_{ij}dZ_s^{\tilde{\mathcal{P}},j}\middle|\,X_t\right]\\
&= X_t^{(i)}e^{-\kappa_i(T-t)}+\tilde{\mu}_i(1-e^{-\kappa_i(T-t)})+\sum_{j=1}^{d}E^{\tilde{\mathcal{P}}}\left[\int_t^T e^{\kappa_i(s-T)}C_{ij}dZ_s^{\tilde{\mathcal{P}},j}\middle|\,X_t\right]\\
&= X_t^{(i)}e^{-\kappa_i(T-t)}+\tilde{\mu}_i(1-e^{-\kappa_i(T-t)})
\end{aligned}$$

and hence

$$E\left[X_T|\,X_t\right] = e^{-\kappa(T-t)}X_t+\left(1-e^{-\kappa(T-t)}\right)\tilde{\mu}.$$

Let $k, l \in \{1, \ldots, d\}$. Then

$$\begin{aligned}
& Cov^{\tilde{\mathcal{P}}}\left[X_T^{(k)}, X_T^{(l)} | X_t\right] \\
= \; & E^{\tilde{\mathcal{P}}}\left[\left(X_T^{(k)} - E^{\tilde{\mathcal{P}}}\left[X_T^{(k)}|X_t\right]\right)\left(X_T^{(l)} - E^{\tilde{\mathcal{P}}}\left[X_T^{(l)}|X_t\right]\right)\Big| X_t\right] \\
= \; & E^{\tilde{\mathcal{P}}}\left[\left(\sum_{i=1}^d \int_t^T e^{\kappa_k(s-T)} C_{ki} dZ_s^{\tilde{\mathcal{P}}\,(i)}\right)\left(\sum_{j=1}^d \int_t^T e^{\kappa_l(s-T)} C_{lj} dZ_s^{\tilde{\mathcal{P}}\,(j)}\right)\Bigg| X_t\right] \\
= \; & \sum_{i=1}^d \sum_{j=1}^d E^{\tilde{\mathcal{P}}}\left[\left(\int_t^T e^{\kappa_k(s-T)} C_{ki} dZ_s^{\tilde{\mathcal{P}}\,(i)}\right)\left(\int_t^T e^{\kappa_l(s-T)} C_{lj} dZ_s^{\tilde{\mathcal{P}}\,(j)}\right)\Bigg| X_t\right] \\
= \; & e^{-(\kappa_k+\kappa_l)T} E^{\tilde{\mathcal{P}}}\left[\left(\int_t^T e^{\kappa_k s} dW_s^{\tilde{\mathcal{P}}\,(k)}\right)\left(\int_t^T e^{\kappa_l s} dW_s^{\tilde{\mathcal{P}}\,(l)}\right)\Bigg| X_t\right] \\
= \; & e^{-(\kappa_k+\kappa_l)T} E^{\tilde{\mathcal{P}}}\left[\int_t^T e^{(\kappa_k+\kappa_l)s} \rho_{lk} ds | X_t\right] \\
= \; & e^{-(\kappa_k+\kappa_l)T}\left[\frac{\rho_{lk}}{\kappa_k+\kappa_l} e^{(\kappa_k+\kappa_l)T} - \frac{\rho_{lk}}{\kappa_k+\kappa_l} e^{(\kappa_k+\kappa_l)t}\right] \\
= \; & \frac{\rho_{lk}}{\kappa_k+\kappa_l}\left(1 - e^{-(\kappa_k+\kappa_l)(T-t)}\right)
\end{aligned}$$

□

For notational simplicity, we set

$$Cov^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] \;:=\; \Sigma(t,T) = \left(\frac{\rho_{lk}}{\kappa_k+\kappa_l}\left(1 - e^{-(\kappa_k+\kappa_l)(T-t)}\right)\right)_{l,k=1,\ldots,d}.$$

We can now derive bond prices and yields.

Theorem 2.2.17. *For the cosh model with state price density process* (ς_t) *with*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c)$$

for all $t \geq 0$ *and* (X_t) *an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13), the price of a zerobond at time* t *with maturity* T *is*

$$\begin{aligned}
P(t,T) &= \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T|X_t\right]}{\varsigma_t} \\
&= \frac{\cosh\left(\sum_{i=1}^d\left(\gamma_i e^{-\kappa_i(T-t)} X_t^{(i)} + (1-e^{-\kappa_i})\mu_i\right) + c\right)}{\cosh(\gamma^T X_t + c)} \\
&\quad \exp\left(-\alpha(T-t) + \frac{1}{2}\sum_{i,j=1}^d \frac{\gamma_i \rho_{ij}\gamma_j}{\kappa_i+\kappa_j}\left(1 - e^{-(\kappa_i+\kappa_j)(T-t)}\right)\right).
\end{aligned}$$

We will frequently use vector notation, hence

$$P(t,T) \;=\; e^{-\alpha(T-t)}\frac{\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c\right)}{\cosh(\gamma^T X_t + c)} e^{\frac{1}{2}\gamma^T\Sigma(t,T)\gamma}.$$

Proof. Since $X_T|X_t$ follows a multivariate normal distribution, $\gamma^T X_T + c$ conditional on X_t follows a normal distribution as well with mean

$$E^{\tilde{\mathcal{P}}}\left[\gamma^T X_T + c\middle|\, X_t\right] \;=\; \gamma^T\left(e^{-\kappa(T-t)}X_t + \left(1-e^{-\kappa(T-t)}\right)\tilde{\mu}\right) + c$$

and variance

$$Var^{\tilde{\mathcal{P}}}\left[\gamma^T X_T + c|X_t\right] \;=\; \gamma^T\Sigma(t,T)\gamma.$$

Now $\exp(\gamma^T X_T + c)$ is, conditionally on X_t, lognormally distributed. Therefore

$$\begin{aligned}
& E^{\tilde{\mathcal{P}}}\left[\exp(\gamma^T X_T + c)\middle|\, X_t\right] \\
= \;& \exp\left(E^{\tilde{\mathcal{P}}}\left[\gamma^T X_T + c\middle|\, X_t\right] + \frac{Cov\left[\gamma^T X_T + c|X_t\right]}{2}\right) \\
= \;& \exp\left(\gamma^T\left(e^{-\kappa(T-t)}X_t + \left(1-e^{-\kappa(T-t)}\right)\tilde{\mu}\right) + c + \frac{\gamma^T\Sigma(t,T)\gamma}{2}\right).
\end{aligned}$$

We arrive at

$$\begin{aligned}
& E^{\tilde{\mathcal{P}}}\left[\cosh(\gamma^T X_T + c)\middle|\, X_t\right] \\
= \;& \frac{1}{2}E^{\tilde{\mathcal{P}}}\left[\exp(\gamma^T X_T + c) + \exp(-\gamma^T X_T - c)\middle|\, X_t\right] \\
= \;& \frac{1}{2}\left(\exp\left(\gamma^T\left(e^{-\kappa(T-t)}X_t + \left(1-e^{-\kappa(T-t)}\right)\tilde{\mu}\right) + c\right)\right. \\
& \left. + \exp\left(-\gamma^T\left(e^{-\kappa(T-t)}X_t + \left(1-e^{-\kappa(T-t)}\right)\tilde{\mu}\right) - c\right)\right)\exp\left(\frac{\gamma^T\Sigma(t,T)\gamma}{2}\right) \\
= \;& \cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|\, X_t\right] + c\right)\exp\left(\frac{\gamma^T\Sigma(t,T)\gamma}{2}\right)
\end{aligned}$$

which yields the result required. □

Now the bond pricing formula allows to derive yields of higher maturities.

Corollary 2.2.18. *For the cosh model with state price density process (ς_t) with*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c)$$

for all $t \geq 0$ and (X_t) an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13), nominal zerobond rates $y(t,T)$ at time t with time to maturity $T-t$ are given by

$$y(t,T) \;=\; \alpha - \frac{\log\cosh\left(\gamma^T E\left[X_T|\, X_t\right] + c\right)}{T-t} - \frac{\log\cosh\left(\gamma^T x + c\right)}{T-t} - \frac{\gamma^T\Sigma(t,T)\gamma}{2(T-t)}.$$

Proof.

$$\begin{aligned}
y(t,T) \;&=\; -\frac{\ln P(t,T)}{T-t} \\
&=\; \alpha - \frac{\log\cosh\left(\gamma^T E\left[X_T|\, X_t\right] + c\right)}{T-t} - \frac{\log\cosh\left(\gamma^T x + c\right)}{T-t} - \frac{\gamma^T\Sigma(t,T)\gamma}{2(T-t)}.
\end{aligned}$$

□

Shortrates and instantaneous forward rates can be derived from Rogers general formulae.

Theorem 2.2.19. *For the cosh model with state price density process* (ς_t) *and*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c)$$

for all $t \geq 0$ *and* (X_t) *an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13), the shortrate is given by*

$$r_t = \alpha - \gamma^T\kappa(\tilde{\mu} - X_t)\tanh(\gamma^T X_t + c) - \frac{1}{2}\gamma^T\rho\gamma$$

and instantaneous forward rates $f(t,T)$ *are given by*

$$f(t,T) = \alpha - \tanh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)\gamma^T\kappa e^{-\kappa(T-t)}(\tilde{\mu} - X_t) - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma$$

Proof. The shortrate within the Rogers framework is given by

$$r_t = \frac{(\alpha - G)f(X_t)}{f(X_t)}.$$

With $f(X_t) = \cosh(\gamma^T X_t + c)$ and the state vector dynamics given by (2.12),

$$\begin{aligned}
&(\alpha - G)f(X_t)\\
=\ & \alpha f(X_t) - \sum_{i=1}^{n}\kappa_i(\tilde{\mu}_i - X_t)\frac{\partial f}{\partial x_i}(X_t) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(CC^T)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)\\
=\ & \alpha\cosh(\gamma^T X_t + c) - \sum_{i=1}^{n}\kappa_i(\tilde{\mu}_i - X_t)\sinh(\gamma^t X_t + c)\gamma_i\\
& -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(CC^T)_{ij}\cosh(\gamma^T X_t + c)\gamma_i\gamma_j\\
=\ & \alpha\cosh(\gamma^T X_t + c) - \gamma^T\kappa(\tilde{\mu} - X_t)\sinh(\gamma^t X_t + c) - \frac{1}{2}\gamma^T\rho\gamma\cosh(\gamma^T X_t + c).
\end{aligned}$$

Hence

$$\begin{aligned}
r_t &= \frac{\alpha\cosh(\gamma^T X_t + c) - \gamma^T\kappa(\tilde{\mu} - X_t)\sinh(\gamma^t X_t + c) - \frac{1}{2}\gamma^T\rho\gamma\cosh(\gamma^T X_t + c)}{\cosh(\gamma^T X_t + c)}\\
&= \alpha - \gamma^T\kappa(\tilde{\mu} - X_t)\tanh(\gamma^t X_t + c) - \frac{1}{2}\gamma^T\rho\gamma.
\end{aligned}$$

Instantaneous forward rates are given by

$$\begin{aligned}
f(t,T) &= -\frac{\partial}{\partial T}\log\left(P(t,T)\right)\\
&= -\frac{\partial}{\partial T}\log\left(e^{-\alpha(T-t)}\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)}{\cosh(\gamma^T X_t + c)}e^{\frac{1}{2}\gamma^T\Sigma(t,T)\gamma}\right)\\
&= -\frac{\partial}{\partial T}\left(-\alpha(T-t) + \log\left(\cosh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)\right) + \frac{1}{2}\gamma^T\Sigma(t,T)\gamma\right)\\
&= \alpha - \frac{\sinh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)}{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)}\frac{\partial}{\partial T}\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial T}\frac{\gamma_i\rho_{ij}\gamma_j}{\kappa_i + \kappa_j}e^{-(\kappa_i+\kappa_j)(T-t)}\\
&= \alpha - \tanh(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + c)\gamma^T\kappa e^{-\kappa(T-t)}(\tilde{\mu} - X_t) - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma
\end{aligned}$$

which yields the result. ☐

2.2.7 Change of measures

Equivalent martingale measures

In section 2.2.6 we introduced the cosh and the Cairns model. Both models were introduced using a so called reference measure which is used for the pricing formula based on the state price density. The reference measure is a theoretical measure introduced to facilitate pricing formulae based on the state price density. Analogously, the risk-neutral measure is a theoretical measure introduced to facilitate pricing based on the martingale property of discounted asset prices under this measure. Historical data, however, is available under the physical or historical measure. The state price density approach requires the physical and reference measure to be equivalent in the same way as the risk-neutral approach requires the risk-neutral measure and the physical measure to be equivalent: to allow for estimation. Furthermore, the reference measure and the risk-neutral measure must be equivalent as a prerequisite for the no-arbitrage condition.
Given an asset price process $(S_t)_{t\geq 0}$, the existence of an equivalent measure under which the discounted asset price process $\left(e^{-\int_0^t r_s ds} S_t\right)$ is a martingale implies no-arbitrage. This equivalent martingale measure is the risk-neutral measure, see for example [KS91] or [MR05]. The mentioned market consists of the risky asset S, the bank account paying the shortrate r_t and any derivative on the underlying S. To guarantee no-arbitrage in state price density models defined under the reference measure, we have to prove existence of an equivalent risk-neutral measure.
A main difference between no arbitrage in bond markets and no-arbitrage in a Black-Scholes market lies in the number of risky assets. In a Black-Scholes market, a single stock S_t is typically the only risky asset. In a bond market with stochastic term structure dynamics, infinitely many risky assets $(P(t,T))_{T\geq t}$ exist. To prove no-arbitrage of the whole bond market a single equivalent measure is required under which all discounted zerobond prices regardless of time to maturity are martingales. We will demonstrate the standard approach to derive no-arbitrage for bond market models for the cosh model.

In standard shortrate models, the dynamics of the shortrate are typically defined under the risk-neutral measure, which therefore implicitly exists and hence no-arbitrage holds. Rogers generic approach defines term structure models under a reference measure, existence of an equivalent risk-neutral measure is not guaranteed. The standard approach to prove no-arbitrage is based on the construction of a risk-neutral measure. We know that discounted asset price processes $\left(e^{-\int_0^t r_s ds} S_t\right)$ are martingales under the risk-neutral measure. Hence given dynamics of a risky asset under the risk-neutral measure

$$dS_t = \mu(t,S_t)dt + \sigma(t,S_t)dW_t^{\mathbb{Q}},$$

we can apply the Ito-Doeblin formula[11] for the discounted asset price process $\left(e^{-\int_0^t r_s ds} S_t\right)$

$$de^{-\int_0^t r_s ds} S_t = \left[-r_t e^{-\int_0^t r_s ds} S_t + \mu(t, S_t) e^{-\int_0^t r_s ds} + 0\right] dt + \sigma(t, S_t) e^{-\int_0^t r_s ds} S_t dW_t^{\mathcal{Q}}$$

which simplifies to

$$\frac{de^{-\int_0^t r_s ds} S_t}{e^{-\int_0^t r_s ds} S_t} = \left[-r_t + \frac{\mu(t, S_t)}{S_t}\right] dt + \sigma(t, S_t) dW_t^{\mathcal{Q}}.$$

By assumption, the discounted asset price under the risk neutral measure must be a martingale, hence $\mu(t, S_t) = r_t S_t$. Therefore, in a stochastic term structure model, the drift of any zerobond under the risk-neutral measure must equal the stochastic model-implied shortrate. This allows to construct a measure by specifying a drift correction term which, see also [KS91] or [MR05],

1. ensures that under the newly constructed measure the drift of the zerobond dynamics equals the shortrate and
2. ensures that the constructed measure is equivalent to the initial measure.

If these conditions hold, the constructed measure is a risk-neutral measure and hence the underlying market is free of arbitrage. As the underlying market so far consists merely of the bank account and the single bond $P(t, T)$ whose dynamics were used to derive the risk-neutral measure, the bond market as a whole is free of arbitrage if and only if the derived drift correction term is independent of the time to maturity $T - t$. In this case, the above described algorithm constructs the same risk-neutral measure for all zerobonds and hence the whole bond market consisting of the bank account, zerobonds of arbitrary times to maturity $(P(t, T))_{T \geq t}$ and their derivatives is arbitrage-free.

Since historical data was observed under the physical measure, one has to specify the dynamics under the physical measure as well, both for shortrate models defined under the risk-neutral measure and state price density models defined under the reference measure. In shortrate models defined under the risk-neutral measure, the market price of risk defines the physical measure. We will see that in state price density models the specification of the physical dynamics is also equivalent to the choice of a market price of risk.
Best practice in term structure models is to specify the market price of risk in such a way that dynamics under the original measure and under the physical measure imply similar state factor dynamics and distributions. To give an example, Dai and Singleton [DS00] choose[12] a market price of risk in the affine framework which guarantees that the state factor follows the same mean reverting dynamics under both the risk-neutral and the physical measure, yet with distinct constant long-term means. We will see that Cairns

[11] See the appendix A

[12] As affine term structure models are defined under the risk-neutral measure, one is indeed free to choose a market price of risk as long as it implies that risk-neutral measure and physical measure are equivalent.

followed best practice in his model as well.
In subsequent papers, however, Duffie [Duf00] and Duarte [Dua04] found that alternative, more complicated market prices of risk may improve historical fit, forecasting and risk premia. We can therefore expect that different choices of a market price of risk can improve certain model aspects of the cosh or Cairns model as well. The price we have to pay for this improvement are more complicated conditional distributions of the state vector under the physical measure.

In the following, we will derive the dynamics under the risk-neutral and physical measures and thus the no-arbitrage condition for both the Cairns and the cosh model and the respective market prices of risk.

The risk-neutral measure within the Cairns model

Cairns followed the standard approach to prove no-arbitrage. Specifically, he derived the dynamics of the bond price under the reference measure, calculated the required drift-correction term and proved the Novikov condition. The drift correction term from the risk-neutral to the reference measure is given by Cairns as

$$dZ_j^{\mathcal{Q}} = dZ_j^{\tilde{\mathcal{P}}} - V_j(t,t)dt. \tag{2.17}$$

with

$$V_i(t,t) = \frac{\int_0^\infty \sum_{j=1}^d \sigma_j e^{-\kappa_j u} C_{ji} H(u,X_t)du}{\int_0^\infty H(u,X_t)du}. \tag{2.18}$$

We summarize by the following theorem.

Theorem 2.2.20. *The Cairns model is free of arbitrage.*

Proof. By theorem 2.2.14, the Cairns model can be interpreted as a state price density model whereby the state price density is a strictly positive supermartingale. By theorem 2.2.5, no-arbitrage holds. □

The physical measure within the Cairns model

Cairns again follows the standard approach and specifies the market price of risk $\Lambda(t, X_t)$ in such a way that the state vector X remains an Ornstein-Uhlenbeck process under the physical measure.
We are essentially free to choose a market price of risk once no-arbitrage holds by existence of the risk-neutral measure. Note, however, that by construction the market price of risk is a drift correction term from a single risk-neutral measure for all bonds to a single physical measure, therefore it must be independent of a time to maturity $T-t$. If the market price of risk fulfills the Novikov condition, risk-neutral and physical measure are equivalent.

Theorem 2.2.21. *In the Cairns model, the market price of risk*

$$\Lambda(t, X_t)^{\mathcal{Q},\mathcal{P}} := V(t,t) - C^{-1}\kappa\mu$$

defines a physical measure equivalent to both the risk-neutral and the reference measure.

Proof. As in the Cairns model, we prove the Novikov condition to derive equivalence of the physical and the risk-neutral measure

$$E^{\tilde{\mathcal{P}}}\left[\exp\left(\frac{1}{2}\int_0^t \sum_{i=1}^d (V_i(s,s) - (C^{-1}\kappa\mu)_i)\right)\right] \quad < \quad \infty.$$

We know by definition that $\sigma_i(t,T)$ are bounded for all $T > t$ and $i = 1, \dots, d$.

Since $H(u,x) > 0$ for all $u > 0$ and $-\infty < x < \infty$, we get

$$\begin{aligned} |V_j(t,t)| &= \left|\sum_{i=1}^d \sigma_i C_{ij} \frac{\int_0^\infty H(u,X_t)e^{-\kappa_i u}du}{\int_0^\infty H(u,X_t)du}\right| \\ &\leq \left|\sum_{i=1}^d \sigma_i C_{ij}\right| \frac{\int_0^\infty H(u,X_t)e^{-\kappa_i u}du}{\int_0^\infty H(u,X_t)du}. \end{aligned}$$

Now since $e^{-\kappa_i u} \leq 1$ for $u \geq 0$ we have

$$\int_0^\infty H(u,X_t)e^{-\kappa_i u}du \quad \leq \quad \int_0^\infty H(u,X_t)du$$

and hence

$$|V_j(t,t)| \quad \leq \quad \left|\sum_{i=1}^d \sigma_i C_{ij}\right| \frac{\int_0^\infty H(u,X_t)e^{-\kappa_i u}du}{\int_0^\infty H(u,X_t)du} \leq \sum_{i=1}^d |\sigma_i C_{ij}|\,,$$

which guarantees that $V_j(t,t)$ is bounded for all $j = 1, \dots, d$ and $t \geq 0$. Therefore the integrand of the Novikov condition $\sum_{i=1}^d (V_i(s,s) - (C^{-1}\kappa\mu)_i)$ is bounded since θ is constant. As the expected value of a bounded random variable is itself bounded, the Novikov condition is fulfilled. □

In shortrate models, state vector dynamics are defined under the risk-neutral measure, and best practice implies that the market price of risk ensures similar dynamics under the physical measure as well. In the state price density approach, pricing is implemented using the reference measure. Consequently, best practice would imply to choose the market price of risk so that dynamics under the physical measure resemble dynamics under the reference measure. The following theorem will show that this is equivalent to specifying directly a drift correction term from the reference measure $\tilde{\mathcal{P}}$ to the physical measure $\mathcal{P}$.

Corollary 2.2.22. *In the Cairns model with market price of risk*

$$\Lambda^{\mathcal{Q},\mathcal{P}}(t, X_t) \quad = \quad V(t,t) - C^{-1}\kappa\mu,$$

the dynamics of the state vector under the physical measure are given by

$$dX_t \quad = \quad \kappa(\mu - X_t)\,dt + C dZ_t^{\mathcal{P}}$$

Proof. We have

$$\begin{aligned} dZ^{\mathcal{P}} &= dZ^{\mathcal{Q}} + \Lambda(t, X_t)^{\mathcal{Q},\mathcal{P}} dt \\ &= dZ^{\tilde{\mathcal{P}}} - V(t,t)dt + \Lambda(t, X_t)^{\mathcal{Q},\mathcal{P}} dt \end{aligned} \tag{2.19}$$

with (2.17). By (2.2.21), $\Lambda(t, X_t)^{\mathcal{Q},\mathcal{P}} := V(t,t) + \theta$. Now considering the resulting state vector dynamics of component $i \in \{1, \dots, d\}$ we get

$$\begin{aligned} dX_t^{(i)} &= -\kappa_i X_t^{(i)} dt + \sum_{j=1}^{d} C_{ij} dZ^{\tilde{\mathcal{P}},j} \\ &= -\kappa_i X_t^{(i)} dt + \sum_{j=1}^{d} C_{ij} \left(dZ^{\mathcal{P},j} - \theta dt \right) \\ &= -\sum_{j=1}^{d} C_{ij}\theta_j dt - \kappa_i X_t^{(i)} dt + \sum_{j=1}^{d} dZ_t^{\mathcal{P},j} \\ &= [-(C\theta)_i - \kappa_i X_t^{(i)}]dt + \sum_{j=1}^{d} dZ_t^{\mathcal{P},j}. \end{aligned}$$

which implies $\theta = -C^{-1}\kappa\mu$. The drift correction term between the reference and the physical measure is given by θ following (2.19). As θ is constant, the Novikov condition holds and therefore the reference measure and the physical measure are equivalent. □

The risk-neutral measure within the cosh model

To derive the dynamics under the risk-neutral measure, we follow the standard approach. First, we derive the dynamics of the bond price $P(t,T)$ defined by the cosh model under the reference measure using the Ito-Doeblin formula, then we examine the drift correction term which guarantees that the drift under the risk-neutral measure equals the shortrate. We begin with the dynamics of the bond price under the reference measure.

$$\begin{aligned} dP(t,T) &= \left[\frac{\partial}{\partial t} P(t,T) + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} P(t,T)\mu_i(t, X_t) + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} P(t,T)\rho_{ij} \right] dt \\ &\quad + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} P(t,T) \sum_{j=1}^{d} C_{ij} dZ_j^{\tilde{\mathcal{P}}}(t). \end{aligned}$$

The derivative in t is

$$\begin{aligned}
\frac{\partial}{\partial t}P(t,T) &= \frac{\partial}{\partial t}e^{-\alpha(T-t)}\frac{E^{\tilde{\mathcal{P}}}\left[\cosh(\gamma^T X_T + c)|X_t\right]}{\cosh(\gamma^T X_t + c)} \\
&= \frac{\partial}{\partial t}e^{-\alpha(T-t)}\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)}{\cosh(\gamma^T X_t + c)}e^{\frac{\gamma^T\Sigma(t,T)\gamma}{2}} \\
&= \alpha e^{-\alpha(T-t)}\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)}{\cosh(\gamma^T X_t + c)}e^{\frac{\gamma^T\Sigma(t,T)\gamma}{2}} \\
&\quad + e^{-\alpha(T-t)}\frac{\sinh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)\frac{\partial}{\partial t}\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]}{\cosh(\gamma^T X_t + c)}e^{\frac{\gamma^T\Sigma(t,T)\gamma}{2}} \\
&\quad e^{-\alpha(T-t)}\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)}{\cosh(\gamma^T X_t + c)}e^{\frac{\gamma^T\Sigma(t,T)\gamma}{2}}\frac{1}{2}\frac{\partial}{\partial t}\gamma^T\Sigma(t,T)\gamma \\
&= P(t,T)\left[\alpha + \tanh\left(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c\right)\gamma^T\kappa e^{-\kappa(T-t)}(X_t - \tilde{\mu})\right. \\
&\quad \left. -\frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma\right]
\end{aligned}$$

For the derivative in x_i, note that the (conditional) covariance $\Sigma(t,T)$ does not depend on the current state $X_t = x$, see theorem 2.2.16. Therefore

$$\begin{aligned}
\frac{\partial}{\partial x_i}P(t,T) &= e^{-\alpha(T-t)+\frac{\gamma^T\Sigma(t,T)\gamma}{2}}\left(\frac{\frac{\partial}{\partial x_i}\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)\cosh(\gamma^T X_t + c)}{\cosh^2(\gamma^T X_t + c)}\right. \\
&\quad \left. -\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)\frac{\partial}{\partial x_i}\cosh(\gamma^T X_t + c)}{\cosh^2(\gamma^T X_t + c)}\right) \\
&= e^{-\alpha(T-t)+\frac{\gamma^T\Sigma(t,T)\gamma}{2}}\left(\frac{\sinh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)e^{-\kappa_i(T-t)}\gamma_i}{\cosh(\gamma^T X_t + c)}\right. \\
&\quad \left. -\frac{\cosh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)\sinh(\gamma^T X_t + c)\gamma_i}{\cosh^2(\gamma^T X_t + c)}\right) \\
&= P(t,T)\left(\tanh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)e^{-\kappa_i(T-t)}\gamma_i - \tanh(\gamma^T X_t + c)\gamma_i\right) \\
&= \gamma_i P(t,T)\left[\tanh(\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right] + c)e^{-\kappa_i(T-t)} - \tanh(\gamma^T X_t + c)\right]
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{d}\mu_i(t,X_t)\frac{\partial}{\partial x_i}P(t,T) \\
=\ &\sum_{i=1}^{d}\kappa_i\left(\tilde{\mu}_i - X_t^{(i)}\right)\gamma_i P(t,T)\left[\tanh\left(\gamma^T E[X_T|X_t] + c\right)e^{-\kappa_i(T-t)} - \tanh\left(\gamma^T X_t + c\right)\right] \\
=\ &P(t,T)\left[\tanh\left(\gamma^T E[X_T|X_t] + c\right)\gamma^T\kappa e^{-\kappa(T-t)}(\tilde{\mu} - X_t)\right. \\
&\left. -\tanh\left(\gamma^T X_t + c\right)\gamma^T\kappa(\tilde{\mu} - X_t)\right].
\end{aligned}$$

Finally, we get

$$\begin{aligned}
&\frac{\partial^2}{\partial x_i \partial x_j} P(t,T) \\
= \;& \gamma_i \gamma_j P(t,T) \Big[-\tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \Big[e^{-\kappa_i (T-t)} + e^{-\kappa_j (T-t)} \Big] \\
& +2\tanh^2\left(\gamma^T X_t + c\right) + e^{-(\kappa_i + \kappa_j)(T-t)} - 1 \Big]
\end{aligned}$$

and thus

$$\begin{aligned}
& \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \frac{\partial^2}{\partial x_i \partial x_j} P(t,T) \\
= \;& \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \gamma_i \gamma_j P(t,T) \left[-\tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \right. \\
& \left. \cdot \left(e^{-\kappa_i (T-t)} + e^{-\kappa_j (T-t)} \right) + 2\tanh^2\left(\gamma^T X_t + c\right) + e^{-(\kappa_i + \kappa_j)(T-t)} - 1 \right] \\
= \;& P(t,T) \Big[-\gamma^T e^{-\kappa (T-t)} \rho \gamma \tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \\
& +\gamma^T \rho \gamma \tanh^2\left(\gamma^T X_t + c\right) - \frac{1}{2} \gamma^T \rho \gamma + \frac{1}{2} \gamma^T e^{-\kappa (T-t)} \rho e^{-\kappa (T-t)} \gamma \Big].
\end{aligned}$$

So the dynamics of the bond price under the reference measure are given by

$$\begin{aligned}
dP(t,T) \;=\; & \left[\frac{\partial}{\partial t} P(t,T) + \sum_{i=1}^{d} \tilde{\mu}_i \frac{\partial}{\partial x_i} P(t,T) + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} P(t,T) \rho_{ij} \right] dt \\
& + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} P(t,T) \sum_{j=1}^{d} C_{ij} dZ_j^{\tilde{\mathcal{P}}}(t) \\
= \;& P(t,T) \Big[\alpha - \tanh\left(\gamma^T E\left[X_T|X_t\right] + c\right) \gamma^T \kappa e^{-\kappa (T-t)} (\tilde{\mu} - X_t) \\
& - \frac{1}{2} \gamma^T e^{-\kappa (T-t)} \rho e^{-\kappa (T-t)} \gamma + \tanh\left(\gamma^T E[X_T|X_t] + c\right) \gamma^T \kappa e^{-\kappa (T-t)} (\tilde{\mu} - X_t) \\
& - \tanh\left(\gamma^T X_t + c\right) \gamma^T \kappa (\tilde{\mu} - X_t) \\
& - \gamma^T e^{-\kappa (T-t)} \rho \gamma \tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \\
& + \gamma^T \rho \gamma \tanh^2\left(\gamma^T X_t + c\right) - \frac{1}{2} \gamma^T \rho \gamma + \frac{1}{2} \gamma^T e^{-\kappa (T-t)} \rho e^{-\kappa (T-t)} \gamma \Big] dt \\
& + P(t,T) \Big[\tanh\left(\gamma^T E[X_T|X_t] + c\right) \gamma^T e^{-\kappa (T-t)} C \\
& - \tanh\left(\gamma^T X_t + c\right) \gamma^T C \Big] dZ_t^{\tilde{\mathcal{P}}}(t) \\
= \;& P(t,T) \Big[r_t - \gamma^T e^{-\kappa (T-t)} \rho \gamma \tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \\
& + \gamma^T \rho \gamma \tanh^2\left(\gamma^T X_t + c\right) \Big] dt + P(t,T) \Big[\tanh\left(\gamma^T E[X_T|X_t] + c\right) \gamma^T e^{-\kappa (T-t)} C \\
& - \tanh\left(\gamma^T X_t + c\right) \gamma^T C \Big] dZ_t^{\tilde{\mathcal{P}}}(t)
\end{aligned}$$

whereby we used the formula of the shortrate from theorem 2.2.19. Now we need a drift correction term $\Lambda(X_t, t)$ with

$$dZ_t^{\tilde{\mathcal{P}}} = dZ_t^{\mathcal{Q}} + \Lambda(X_t, t) dt,$$

such that the drift of the bond price under the constructed measure equals the shortrate r_t. Assuming a drift correction term $\Lambda(X_t, t)$, the bond price dynamics become

$$\begin{aligned} \frac{dP(t,T)}{P(t,T)} &= \Big[r_t - \gamma^T e^{-\kappa(T-t)}\rho\gamma \tanh\left(\gamma^T X_t + c\right) \tanh\left(\gamma^T E[X_T|X_t] + c\right) \\ &\quad + \gamma^T \rho\gamma \tanh^2\left(\gamma^T X_t + c\right)\Big]\, dt \\ &\quad + \Big[\tanh\left(\gamma^T E[X_T|X_t] + c\right)\gamma^T e^{-\kappa(T-t)} C \\ &\quad - \tanh\left(\gamma^T X_t + c\right)\gamma^T C\Big] \left(dZ_t^{\mathcal{Q}} + \Lambda(X_t, t)dt\right). \end{aligned}$$

First, we define $\Lambda(X_t, t) := C^T \Lambda'(X_t, t)$ for simplicity, then

$$\begin{aligned} & P(t,T)\Big[\tanh\left(\gamma^T E[X_T|X_t] + c\right)\gamma^T e^{-\kappa(T-t)} C - \tanh\left(\gamma^T X_t + c\right)\gamma^T C\Big] C^T \Lambda'(X_t, t)dt \\ = \; & P(t,T)\Big[\tanh\left(\gamma^T E[X_T|X_t] + c\right)\gamma^T e^{-\kappa(T-t)} \rho \Lambda'(X_t, t) \\ & - \tanh\left(\gamma^T X_t + c\right)\gamma^T \rho \Lambda'(X_t, t)_t\Big]\, dt. \end{aligned}$$

The combined drift term now must equal the shortrate, hence

$$\begin{aligned} r_t &= r_t - \gamma^T e^{-\kappa(T-t)}\rho\gamma \tanh\left(\gamma^T X_t + c\right)\tanh\left(\gamma^T E[X_T|X_t] + c\right) + \gamma^T\rho\gamma\tanh^2\left(\gamma^T X_t + c\right) \\ &\quad + \tanh\left(\gamma^T E[X_T|X_t] + c\right)\gamma^T e^{-\kappa(T-t)}\rho\Lambda'(X_t, t) - \tanh\left(\gamma^T X_t + c\right)\gamma^T\rho\Lambda'(X_t, t), \end{aligned}$$

for which

$$C^T \Lambda'(X_t) = C^T \gamma \tanh\left(\gamma X_t + c\right)$$

is the obvious choice. As the last step, we examine the Novikov condition to prove equivalence of the reference measure and the constructed measure. Thus we require

$$E^{\tilde{\mathcal{P}}}\left[\exp\left(\frac{1}{2}\int_0^t \sum_{i=1}^d (C\gamma)_i^2 \tanh^2(\gamma^T X_s + c)ds\right)\right] < \infty.$$

As tanh is bounded so is the integrand and therefore the integral. As the expected value of a bounded random variable is again bounded the Novikov condition holds for our drift correction term. According to the Cameron-Martin-Girsanov theorem the measures $\tilde{\mathcal{P}}$ and the constructed measure $\mathcal{Q}$ are equivalent. As therefore $\mathcal{Q}$ is an equivalent measure under which discounted bond prices are martingales, $\mathcal{Q}$ is a risk-neutral measure.

Note that the drift correction term $\Lambda(X_t)$ does not depend on the current time t nor on the time to maturity $T-t$ of the bond $P(t,T)$ used to derive the term. Therefore, the same drift correction term $\Lambda(X_t)$ applies for all zerobond $(P(t,T))_{T\geq t}$, which is required to derive no-arbitrage of the whole bond market. We summarize this in the following theorem.

Theorem 2.2.23. *For the cosh model with state price density process* (ς_t) *with*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c)$$

and (X_t) an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13) under the reference measure, there exists an equivalent martingale measure $\mathcal{Q}$ with

$$dZ_t^{\tilde{\mathcal{P}}} = dZ_t^{\mathcal{Q}} + C^T \gamma \tanh\left(\gamma^T X_t + c\right) dt,$$

and

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + C dZ_t^{\mathcal{Q}}.$$

The measure $\mathcal{Q}$ is the usual risk-neutral measure.

By the fundamental theorem of asset pricing, the existence of a risk-neutral measure implies no-arbitrage of the market considered. So far, this market consists of the banking account, the single risky asset $P(t,T)$ and its derivative securities. As the drift correction term $\Lambda(X_t)$ does not depend on the time to maturity $T-t$, we can construct the same risk-neutral measure for arbitrary bonds $(P(t,T))_{T\geq t}$, hence the following corollary holds.

Corollary 2.2.24. *For the cosh model with state price density process (ς_t) with*

$$\varsigma_t = e^{-\alpha t} \cosh(\gamma^T X_t + c)$$

and (X_t) an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13), the market consisting of bank account, zerobonds $P(t,T)$ of arbitrary maturities $T \geq t$ and their derivative securities is free of arbitrage.

The physical measure within the cosh model

As the cosh model is defined using the reference measure, best practice implies that we choose the dynamics of the physical measure similar to those of the reference measure and not the risk-neutral measure. This yields a drift correction term $\Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(t, X_t)$ from the reference to the physical measure, hence

$$dZ^{\mathcal{P}} = dZ^{\tilde{\mathcal{P}}} + \Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) dt$$

and

$$\begin{aligned}
dX_t^{(i)} &= \kappa_i \left(\tilde{\mu}_i - X_t^{(i)}\right) dt + \sum_{j=1}^d C_{ij} dZ_t^{\tilde{\mathcal{P}},j} \\
dX_t^{(i)} &= \kappa_i \left(\tilde{\mu}_i - X_t^{(i)}\right) dt + \sum_{j=1}^d C_{ij} \left(dZ_t^{\mathcal{P},j} - \Lambda_j^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) dt\right) \\
dX_t^{(i)} &= \left(-\sum_{j=1}^d C_{ij} \Lambda_j^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) + \kappa_i \tilde{\mu}_i - \kappa_i X_t^{(i)}\right) dt + \sum_{j=1}^d C_{ij} dZ_t^{\mathcal{P},j}.
\end{aligned}$$

If we want the state factor under the physical measure to be an Ornstein-Uhlenbeck process with long-term mean μ, this implies

$$\begin{aligned}\kappa_i\mu_i &\stackrel{!}{=} -\sum_{j=1}^{d} C_{ij}\Lambda_j^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) + \kappa_i\tilde{\mu}_i \\ \kappa_i(\tilde{\mu}_i - \mu_i) &= \sum_{j=1}^{d} C_{ij}\Lambda_j^{\tilde{\mathcal{P}},\mathcal{P}}(X_t).\end{aligned}$$

Therefore $\Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) := C^{-1}\kappa(\tilde{\mu} - \mu)$, which results in

$$dX_t = \kappa\,(\mu - X_t)\,dt + dZ_t^{\mathcal{P}}. \tag{2.20}$$

As the drift correction term is constant, the Novikov condition is fulfilled and both measures are equivalent. As with the Cairns model, we will see that this specification of the physical measure also defines our market price of risk.

Theorem 2.2.25. *In the cosh model with state price density process* (ς_t) *with*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c)$$

and (X_t) *an Ornstein-Uhlenbeck process with dynamics (2.12) under the reference measure and (2.20) under the physical measure, the market price of risk is given by*

$$\Lambda^{\mathcal{Q},\mathcal{P}}(X_t) = C\gamma\tanh(\gamma^T X_t + c) - C^{-1}\kappa(\mu - \tilde{\mu}).$$

Proof. The dynamics under the reference measure (2.12) are given by definition, the dynamics under the physical measure (2.20) are given by choice. The drift correction term between these measures is given by

$$dZ^{\mathcal{P}} = dZ^{\tilde{\mathcal{P}}} + \Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(X_t)dt.$$

With the drift correction term from the reference measure to the risk-neutral measure

$$dZ^{\mathcal{Q}} = dZ^{\tilde{\mathcal{P}}} + \Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)dt$$

we get

$$dZ^{\mathcal{P}} = dZ^{\mathcal{Q}} + \left[\Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(X_t) - \Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)\right]dt,$$

which defines the market price of risk

$$\begin{aligned}\Lambda^{\mathcal{Q},\mathcal{P}}(X_t) &= \left[\Lambda^{\tilde{\mathcal{P}},\mathcal{P}}(t, X_t) - \Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)\right] \\ &= C^{-1}\kappa(\mu - \tilde{\mu}) - C\gamma\tanh(\gamma^T X_t + c).\end{aligned}$$

□

2.3 Estimation

2.3.1 Overview

In the literature, several approaches to estimating term structure models exist. Duffee and Stanton [DS04] provide an overview of the most influential ones: Maximum Likelihood estimation, the efficient method of moments (EMM) and the Kalman filter.

Maximum Likelihood estimation maximizes a so called (Log-)Likelihood function, which is a conditional probability function. Many term structure models follow a so called *state space formulation*, where a vector of observations $y_t \in \mathbb{R}^n$ is determined as the mapping under some function g of a certain state $x_t \in \mathbb{R}^d$, thus $g(x_t) = y_t$, whereby the state follows some stochastic process. The state x_t is typically unobservable. Now a frequently used approach to maximum likelihood estimation is to derive the state vector path $\{x_t : 0 \leq t \leq T\}$ by inversion, thus $x_t := g^{-1}(y_t)$. With the inverted state vector path given, the Likelihood function can be expressed as the solution to a partial differential equation of the mean and volatility parameters of the diffusion process X. Then model parameters can be estimated using Maximum Likelihood. The solution to this partial differential equation however is not necessarily available in closed form so that numerical algorithms or approximations are required. Furthermore, the initially used inversion approach does typically not provide a state vector path x_t if $d < n$, which is typically the case. Many authors therefore assume rather arbitrarily that a subset of d yields are observed without error, whereas all other observations are subject to measurement error. Other problems of classical ML estimation stem from its finite-sample properties.

The second important estimation technique in term structure modeling is the efficient method of moments developed by Gallant and Tauchen in [GT96], which is essentially a generalized method of moments as known in the econometrics literature. In this framework, simulations produced with the dynamic model are used to derive indirect inferences about the conditional (log-)density function of the observations. The following introduction is taken from [ACS99]. Let f be some auxiliary function which approximates the log density of y_t conditional on all previous information Y_{T-1} and an auxiliary parameter vector θ to be estimated. The auxiliary function provides a (Pseudo-)Loglikelihood function, which one has to maximize with respect to the parameter vector θ as a first step, hence $\hat{\theta}_T$ satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}f(y_t|Y_{t-1},\theta)\bigg|_{\theta=\hat{\theta}_T} = 0. \tag{2.21}$$

Even if the auxiliary model is misspecified, standard QML theory implies under suitable regularity that $\hat{\theta}_T \to \theta_0$. A simulated series $\hat{y}_n(\rho)$, $n = 1, \ldots, N$, is generated from the structural model for a given parameter set ρ and used op evaluate the sample moments at the QML estimate of the auxiliary model $\hat{\theta}_T$

$$m_T(\rho,\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}f(\hat{y}(\rho)|\hat{Y}_{t-1}(\rho),\theta)\bigg|_{\theta=\theta_0}.$$

As $N\to\infty$, $m_N(\rho,\hat{\theta}_T)\to m(\rho,\hat{\theta}_T)$ almost surely. For the simulated sample large enough, the Monte Carlo error becomes negligible and can be ignored. For computational reasons, the GMM criterion in the moment vector is minimized to obtain the EMM estimator of ρ

$$\hat{\rho}_T := \arg\min_{\rho}\left[m_T(\rho,\hat{\theta}_T)^T\hat{I}_T^{-1}m_T(\rho,\hat{\theta}_T)\right] \tag{2.22}$$

where $\hat{I}_T^{-1}$ denotes a consistent estimator of the asymptotic covariance matrix. Dependent on the choice of the auxiliary function, EMM reaches asymptotically the efficiency of ML estimation. However, finite-sample properties are poor as term structure data consists of highly persistent and highly correlated time series. Duffee and Stanton [DS04] show that these problems cannot be solved by reducing the number of moments to be matched, nor by choosing different moments. The problem lies in the weighting matrix $\hat{I}_T^{-1}$ of the moments in case of highly persistent data.

The Kalman filter allows to derive Maximum Likelihood estimates by filtering in case that first the state vector dynamics are Gaussian and second there exists a linear link between the state vector and the observation. The Kalman filter estimates the unobservable state vector of a state space model conditional on some parameter vector θ. The measurements depend on the state vector by an affine function and all measurements are assumed to be observed with an error. As the Kalman filter allows to derive the Loglikelihood function based on conditional one-step densities, maximizing the Loglikelihood value with respect to the parameter vector θ provides us both with the parameter set and the filtered state vector which fit the specified state vector dynamics and the measurement equation best. If the state vector dynamics or the measurement equation cannot be expressed in an affine form dependent on Gaussian innovations, first order Taylor expansion could be used to approximate the nonlinear case by a Gaussian, linear model which yields a Quasi-Loglikelihood function for estimation. This effectively allows to include a variety of measurements which depend on the state vector, which, together with the assumption of measurement with error allows for panel data to be used. As the Kalman filter produces an estimate of the historical state process path, historical yields or security prices that were not observed or not included in measurement in the first place can easily be derived. These implied measurements can also be used to examine the historical fit of the model.

For the choice of estimation methods, Duffee and Stanton provide three conclusions:

1. ML estimation yields highly biased estimates for most term structure models considered by these authors, especially if flexibility in the market price of risk is allowed.

2. the Efficient Method of Moments is an “unacceptable alternative” to ML estimation in finite samples. In fact, it seems that EMM requires substantially larger samples than ML estimation to reach its asymptotic behavior.

3. the Kalman filter is a reasonable alternative to ML estimation even in case of non-Gaussian settings. However, if feasible, the ML estimator is still superior.

In general, Duffee and Stanton recommend using Monte Carlo simulations to test the ability of the estimation technique to derive the parameter sets required, especially considering small-sample behavior.

2.3.2 The Kalman filter

The basis of the Kalman filter is the so called *state space formulation*. This formulation is based on the two parts of the Kalman filter: the *transition equation* which describes the dynamical evolution of the state process and the *measurement equation*, which relates the state variable at time t with the observations at the same time. The Kalman filter is therefore a natural choice for models in which an unobservable stochastic state process X describes the dynamics of observable measurements Y. The measurements must depend on the state process X by some function $g : \mathcal{X} \to \mathcal{Y}$. All observations are assumed to be measured with error, whereby the standard deviation of the error can be estimated as well. For a general introduction of the Kalman filter see [Har91], here, we follow [Kel01].

In the original Kalman filter, measurement and transition equations are linear in the state process at time t. Let θ denote the vector of all model parameters, then the measurement equation in its most general form is given by

$$Y_t = a_t(\theta) + B_t(\theta)X_t + \epsilon_t(\theta)$$

whereby $Y_t \in \mathcal{Y}$ and $dim(\mathcal{Y}) = n$, $\epsilon_t(\theta), a_t(\theta) \in \mathbb{R}^n$, the state vector $X_t \in \mathcal{X}$ with $dim(\mathcal{X}) := d \leq \mathcal{Y}$ and thus $B_t(\theta) \in \mathbb{R}^{n\times d}$. For the measurement error $\epsilon_t(\theta)$, we assume a (multivariate) normal distribution with

$$E\left[\epsilon_t(\theta)\right] = 0$$

and

$$E\left[\epsilon_s(\theta)\epsilon_t(\theta)^T\right] = H_t(\theta)$$

for $t = s$ and

$$E\left[\epsilon_s(\theta)\epsilon_t(\theta)^T\right] = 0$$

otherwise, whereby the covariance matrix $H_t(\theta) \in \mathbb{R}^{n\times n}$ has to be estimated under the assumption that the vectors of error terms for different observations in time $t_1, t_2, \ldots$, say a sequence of term structures, are uncorrelated in t. This reflects the basic assumption that time-dependence of two measurements y_t and y_s is fully described by the transition equation

$$X_t = c_t(\theta) + \Phi_t(\theta)X_{t-1} + \eta_t(\theta).$$

In this case, $\eta_t(\theta), c_t(\theta) \in \mathbb{R}^d$ and $\Phi_t(\theta) \in \mathbb{R}^{d\times d}$. The error term $\eta_t(\theta)$ is again assumed to be multivariate normal with

$$E\left[\eta_t(\theta)\right] = 0$$

and

$$E\left[\eta_s(\theta)\eta_t(\theta)^T\right] := Q_t(\theta)$$

for $s = t$ and

$$E\left[\eta_s(\theta)\eta_t(\theta)^T\right] := 0$$

otherwise. The matrix $Q_t(\theta)$ must be estimated as well as the matrix $H_t(\theta)$.

In most cases, the Euler-Maruyama scheme could be used to derive the discretization of the state vector dynamics, whereby the distribution of the state X_t conditional on X_{t-1} is normal. In the models we consider here, the underlying state process is an Ornstein-Uhlenbeck process, for which the conditional distribution is known. This can be used to derive the transition equation directly without approximation as in Euler-Maruyama.
Many term structure models however do not provide measurements which are linear in the state vector. For these cases, the *Extended Kalman filter* has to be used. In its most general form, both the transition equation and the measurement equation are non-linear in the state X_{t-1}. In the Cairns and cosh models, we only have to consider a non-linear measurement equation

$$Y_t = g_t(X_t, \epsilon_t(\theta), \theta).$$

The Extended Kalman filter approximates the non-linear function $g_t(X_t, \epsilon_t(\theta), \theta)$ around the conditional mean of the stochastic inputs

$$\begin{aligned} E[(X_t, \epsilon_t)|\mathcal{F}_{t-1}] &= (E[X_t|\mathcal{F}_{t-1}], E[\epsilon_t|\mathcal{F}_{t-1}]) \\ &= (X_{t|t-1}, 0), \end{aligned} \tag{2.23}$$

whereby $X_{t|t-1} := E[X_t|X_{t-1}]$ and $\mathcal{F}_s := \{Y_s, Y_{s-1}, \dots, Y_1\}$. This requires that error terms are uncorrelated over time as well as uncorrelated with the state vector, which reflects our assumption that the state space formulation covers all systematic movements. The first-order Taylor series expansion around $(X_{t|t-1}, 0)$ then yields

$$Y_t \approx g_t(X_{t|t-1}, 0, \theta) + B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t-1}\epsilon_t(\theta)$$

with

$$
\begin{aligned}
B_{t|t-1} &= \left.\frac{\partial g_t(x,\epsilon,\theta)}{\partial x}\right|_{(x,\epsilon)=(X_{t|t-1},0)} \\
R_{t|t-1} &= \left.\frac{\partial g_t(x,\epsilon,\theta)}{\partial \epsilon}\right|_{(x,\epsilon)=(X_{t|t-1},0)}.
\end{aligned}
$$

Note that both $B_{t|t-1}$ and $R_{t|t-1}$ do not depend on the current state X_t, but only on $X_{t|t-1}$, the optimal forecast of the current state given the previous state $X_{t-1|t-1}$. The approximated measurement equation is linear in X_t and we can employ the filtering technique of the Kalman filter. The (Extended) Kalman filter now works as a linear filtering technique to derive the (unobservable) state vector by two steps:

1. **Prediction Step:** First, we form an optimal prediction of the next measurement y_{t+1}, given all current information $\mathcal{F}_t$, whereby for all practical reasons $\mathcal{F}_t := \sigma(X_s, s \leq t)$. The optimal prediction is the conditional expectation

   $$X_{t|t-1} = E[X_t|\mathcal{F}_{t-1}] = c_t(\theta) + \Phi_t(\theta)X_{t-1|t-1}$$

   whereby we used the transition equation. For each t we denote by $X_{t|t}$ the best estimate of the state at time t based on both the observation y_t of the current time t and the best estimation of the current state X_t conditional on the previous state $X_{t|t-1}$. The second prediction equation is the conditional covariance matrix of the prediction error $X_t - X_{t|t-1}$, given by

   $$
   \begin{aligned}
   \Sigma_{t|t-1} &:= E\left[(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T \middle| \mathcal{F}_{t-1}\right] \\
   &= E\Big[(c_t(\theta) + \Phi_t(\theta)X_{t-1} + \eta_t(\theta) - c_t(\theta) - \Phi_t(\theta)X_{t-1|t-1}) \\
   &\quad (c_t(\theta)^T + X_{t-1}^T\Phi_t(\theta)^T + \eta_t(\theta)^T - c_t(\theta)^T - X_{t-1|t-1}^T\Phi_t(\theta)^T)\Big| \mathcal{F}_{t-1}\Big] \\
   &= \Phi_t(\theta)E[(X_{t-1} - X_{t-1|t-1})(X_{t-1} - X_{t-1|t-1})^T|\mathcal{F}_{t-1}]\Phi_t(\theta)^T \\
   &\quad + E[\eta_t(\theta)\eta_t(\theta)^T|\mathcal{F}_{t-1}] \\
   &= \Phi_t(\theta)\Sigma_{t-1|t-1}\Phi_t(\theta)^T + Q_t(\theta).
   \end{aligned}
   $$

   Where we defined $\Sigma_{t|t} := E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T|\mathcal{F}_t]$, that is the optimal estimate of the error covariance matrix at time $t-1$ using all information available at time $t-1$. The prediction step therefore yields a-priori estimates of the state vector and the covariance matrix of the state vector.

2. **Updating Step:** The a-priori estimate is then "updated", hence the a-priori estimate $X_{t|t-1}$ based on the available information at time $t-1$ is combined with the new measurement y_t at time t to the a-posteriori estimate.

The update step is based on the prediction error $v_t = y_t - E[y_t|\mathcal{F}_{t-1}] = y_t - y_{t|t-1}$ and its covariance matrix. Namely

$$\begin{aligned}
v_t &:= y_t - y_{t|t-1} \\
&\approx y_t - E[g_t(X_{t|t-1}, 0, \theta) + B_{t|t-1}(\theta)(X_t - X_{t|t-1}) + R_{t|t-1}(\theta)\epsilon_t(\theta)|\mathcal{F}_{t-1}] \\
&= y_t - g_t(X_{t|t-1}, 0, \theta) - B_{t|t-1}(\theta)E[X_t - X_{t|t-1}|\mathcal{F}_{t-1}] - R_{t|t-1}(\theta)E[\epsilon_t(\theta)|\mathcal{F}_{t-1}] \\
&= y_t - g_t(X_{t|t-1}, 0, \theta).
\end{aligned}$$

The covariance matrix of the prediction error is given by

$$\begin{aligned}
F_{t|t-1} &:= Cov[v_t|\mathcal{F}_{t-1}] \\
&= E[(y_t - g_t(X_{t|t-1}, 0, \theta))(y_t - g_t(X_{t|t-1}, 0, \theta))^T|\mathcal{F}_{t-1}] \\
&\approx E\left[(g_t(X_{t|t-1}, 0, \theta) + B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t}\epsilon_t(\theta) - g_t(X_{t|t-1}, 0, \theta))\right. \\
&\quad \left.(g_t(X_{t|t-1}, 0, \theta) + B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t}\epsilon_t(\theta) - g_t(X_{t|t-1}, 0, \theta))^T|\mathcal{F}_{t-1}\right] \\
&= E\left[(B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t}\epsilon_t(\theta))(B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t}\epsilon_t(\theta))^T|\mathcal{F}_{t-1}\right].
\end{aligned}$$

Since the state vector and the measurement error are uncorrelated and the expected measurement error is zero,

$$\begin{aligned}
& E[(B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t-1}\epsilon_t(\theta))(B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t-1}\epsilon_t(\theta))^T|\mathcal{F}_{t-1}] \\
=\ & E[B_{t|t-1}(X_t - X_{t|t-1})(B_{t|t-1}(X_t - X_{t|t-1}))^T|\mathcal{F}_{t-1}] \\
& +E[R_{t|t-1}\epsilon_t(\theta)(R_{t|t-1}\epsilon_t(\theta))^T|\mathcal{F}_{t-1}]. \\
=\ & B_{t|t-1}E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T|\mathcal{F}_{t-1}]B_{t|t-1}^T + R_{t|t-1}E[\epsilon_t(\theta)\epsilon_t(\theta)^T|\mathcal{F}_{t-1}]R_{t|t-1}^T. \\
=\ & B_{t|t-1}\Sigma_{t|t-1}B_{t|t-1}^T + R_{t|t-1}H_t(\theta)R_{t|t-1}^T.
\end{aligned}$$

Now we want to update the prediction $X_{t|t-1}$ due to the information y_t at time t. With

$$\begin{aligned}
& E[(y_t - y_{t|t-1})(X_t - X_{t|t-1})^T|\mathcal{F}_{t-1}] \\
=\ & E[(B_{t|t-1}(X_t - X_{t|t-1}) + R_{t|t-1}\epsilon_t)(X_t - X_{t|t-1})^T|\mathcal{F}_{t-1}] \\
=\ & B_{t|t-1}E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T|\mathcal{F}_{t-1}] \\
=\ & B_{t|t-1}\Sigma_{t|t-1}
\end{aligned}$$

and analogously

$$\begin{aligned}
& E[(X_t - X_{t|t-1})(y_t - y_{t|t-1})^T|\mathcal{F}_{t-1}] \\
=\ & \Sigma_{t|t-1}B_{t|t-1}^T
\end{aligned}$$

we get that the vector $\begin{pmatrix} X_t \\ y_t \end{pmatrix}$ conditional on the information $\mathcal{F}_{t-1}$ is distributed according to

$$\begin{pmatrix} X_t \\ y_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} X_{t|t-1} \\ y_{t|t-1} \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1}B_{t|t-1}^T \\ B_{t|t-1}\Sigma_{t|t-1} & F_{t|t-1} \end{pmatrix}\right).$$

Using the following lemma, we can derive the updating step.

Lemma 2.3.1. *Let Z_1 and Z_2 be vectors of random variables with joined normal distribution*

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{pmatrix}\right).$$

Then the distribution of Z_1 conditional on Z_2 is $\mathcal{N}(m, \Sigma)$ where

$$\begin{aligned} m &= \mu_1 + \Omega_{12}\Omega_{22}^{-1}(Z_2 - \mu_2) \\ \Sigma &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}^T. \end{aligned}$$

Proof. See Kellerhals, [Kel01], Lemma 4.2.1, page 19. □

This implies that

$$E[Z_1|Z_2] = m,$$

hence the optimal forecast of Z_1 conditional on Z_2 is given by m as derived above. Furthermore, the covariance matrix of the state vector conditional on Z_2 is given by

$$E[(Z_1 - m)(Z_1 - m)^T] = \Sigma.$$

In our case, $Z_1 = X_t$ and $Z_2 = y_t$. With the above derived joint distribution of $(X_t^T, y_t^T)^T$ this implies the updates

$$\begin{aligned} X_{t|t} := E[X_t|y_t] &= X_{t|t-1} + \Sigma_{t|t-1}B_{t|t-1}^T F_{t|t-1}^{-1}(y_t - y_{t|t-1}) \\ &= X_{t|t-1} + \Sigma_{t|t-1}B_{t|t-1}^T F_{t|t-1}^{-1} v_t \\ &= X_{t|t-1} + K_t v_t \end{aligned}$$

where K_t is called the *Kalman gain matrix*. By the lemma, $X_{t|t}$ is the optimal forecast of X_t given the new observation y_t. Analogously, the covariance matrix of the state vector conditional on information y_t is given by

$$\begin{aligned} \Sigma_{t|t} &:= E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T|y_t] \\ &= \Sigma_{t|t-1} - \Sigma_{t|t-1}B_{t|t-1}^T F_{t|t-1} B_{t|t-1}\Sigma_{t|t-1}. \\ &= \Sigma_{t|t-1} - K_t B_{t|t-1}\Sigma_{t|t-1} \end{aligned}$$

To summarize, we can describe the Kalman filtering algorithm given starting points $X_{0|0}$ and $\Sigma_{0|0}$ by the following definition:

1. **Prediction Step:** estimate X_t based upon information $\mathcal{F}_{t-1}$, particularly $X_{t-1|t-1}$

$$\begin{aligned} X_{t|t-1} &= c_t(\theta) + \Phi_t(\theta)X_{t-1|t-1} \\ \Sigma_{t|t-1} &= \Phi_t(\theta)\Sigma_{t-1|t-1}\Phi_t(\theta)^T + Q_t(\theta) \end{aligned}$$

2. **Updating step:** correct the previous estimates $X_{t|t-1}$ and $\Sigma_{t|t-1}$ using the Kalman gain matrix K_t. First, derive the measurement error

$$v_t = y_t - g_t(X_{t|t-1}, 0, \psi)$$

then derive the Jacobi matrix $B_{t|t-1}$ of the measurement equation for the error covariance matrix

$$F_{t|t-1} = B_{t|t-1}\Sigma_{t|t-1}B_{t|t-1}^T + R_{t|t-1}H_t(\theta)R_{t|t-1}^T.$$

The Kalman gain matrix $K_t = \Sigma_{t|t-1}B_{t|t-1}^T F_{t|t-1}^{-1}$ is then required for the updating steps

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + K_t v_t \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - K_t F_{t|t-1} K_t^T. \end{aligned}$$

The choice of the starting points $X_{0|0}$ and $\Sigma_{0|0}$ has a significant impact on the performance of the filter. By definition, the filter approaches the "true" state process over time. The choice of starting values now can speed up or slow down the approximation of the true state vector. An often recommended choice for the starting points is $X_{0|0} := E[X]$ and $\Sigma_{0|0} := E[XX^T]$, if available. The basic assumption here is that the state vector shows some form of mean reversion, varies around its long-term mean $E[X]$ and therefore $E[X]$ is a sensible choice of starting point.

In term structure modeling, we have however additional information available. As our state vectors are assumed to be mean reverting, we can improve our choices of starting values by explicitly taking into account the deviations of the state vector from its long-term mean as reflected in the deviation of the term structure observed y_t from its long-term mean $E[Y]$. If the measurement function $y_t = g(X_t)$ is injective, one can invert it by

$$\min_{x \in \mathcal{X}} |y_t - g(x)|$$

whereby $|\cdot|$ is a reasonable norm. Although we cannot assume the function g to be injective in general, because we assume a d dimensional factor driving the dynamics of n observations with $n > d$ and because state vector components are frequently found to coincide with the three principal components of the term structure, we conclude that the assumption of an injective measurement function g is indeed valid. In case the time steps $(t, t+1)$ are small, note that $g(X_{t+1|t})$ is a reasonable forecast of y_{t+1} given the current observation y_t and the estimate of the current state X_t. We can use this measurement forecast to improve the starting point by minimizing the deviation of the current state-implied measurement $g(X_t)$ and the forecasted measurement $g(X_{t+1|t})$ from both respective empirical observations

$$\min_{x \in \mathcal{X}} (w_1 |y_t - g(x)| + w_2 |y_{t+1} - g(E[X_{t+1}|X_t = x])|)$$

whereby $w_i \geq 0$ are reasonable weights, typically $w_1 \geq w_2$. The mapping g not being injective implies that there exist two states x_1 and x_2 with $y_t = g(x_1)$ and $x_1 = g(x_2)$. Then by mean reversion $E[X_{t+1}|X_t = x_1] \neq E[X_{t+1}|X_t = x_2]$ and hence for the weighted sum

$$\begin{aligned} & w_1|y_t - g(x_1)| + w_2|y_{t+1} - g(E[X_{t+1}|X_t = x_1])| \\ \neq \quad & w_1|y_t - g(x_2)| + w_2|y_{t+1} - g(E[X_{t+1}|X_t = x_2])|. \end{aligned}$$

We used both approaches in our estimations. If the parameter set θ is close to the true parameter sets, or at least close to a local maximum of the Loglikelihood function which describes well term structure dynamics, both approaches were essentially equivalent. If, however, the parameter set θ does not describe well the term structure dynamics, the second approach is vastly superior in fitting reasonable starting points of the Kalman filter. Therefore, estimation approaches should start using the second approach to calibrate $X_{0|0}$. If estimation allows for an iterative approach, sooner or later the second calibration approach can be replaced by the first approach, thereby increasing speed of the estimation algorithm. If estimated parameter sets θ are already available, as is the case for example in derivative pricing, the first calibration approach is sufficient for all purposes.
For the initial matrix $\Sigma_{0|0}$ we used $\Sigma_{0|0} = Cov[X_{t+\Delta}|X_t]$, which in case of X being an Ornstein-Uhlenbeck process does not depend on the initial state $X_{0|0}$ but only on θ. This should guarantee a high stability of the estimate of the covariance of the state vector and in turn the proposed choice of a starting vector should be reasonable. Note however that $\Sigma_{0|0} := Id_d$ was only slightly worse if the parameter set θ did not describe term structure dynamics well. In initial estimation steps the measurement error ν defining the matrix $Q_t(\theta)$ is typically huge and therefore dominates $\Sigma_{t|t} = \Phi_t(\theta)\Sigma_{t-1|t-1}\Phi_t(\theta)^T + Q_t(\theta)$. All in all the impact of the initial value $\Sigma_{0|0}$ was considerably smaller than the impact of our choice of $X_{[0|0}$.

Parameter Estimation

The Kalman filter can be used to derive a (Quasi-)Maximum Likelihood estimation approach. The likelihood function dependent on the parameter vector θ of the state space model is given by the joint density of the observations $(y_T, y_{T-1}, \ldots, y_1)$

$$\begin{aligned} l(y;\theta) &= p(y_T, y_{T_1}, \ldots, y_1) \\ &= p(y_T|y_{T-1}, \ldots, y_1;\theta) \cdot p(y_{T-1}|y_{T-2}, \ldots, y_1;\theta) \cdot \ldots \cdot p(y_1|\mathcal{F}_0;\theta). \end{aligned}$$

Whereby $\mathcal{F}_0$ contains all prior information, for example from an implementation point of view the starting values $X_{0|0}$ and $\Sigma_{0|0}$. If the state process is Markovian, as in all our applications,

$$\begin{aligned} & p(y_T|y_{T-1}, \ldots, y_1;\theta) \cdot p(y_{T-1}|y_{T-2}, \ldots, y_1;\theta) \cdot \ldots \cdot p(y_1|\mathcal{F}_0;\theta) \\ = \quad & p(y_T|y_{T-1};\theta) \cdot p(y_{T-1}|y_{T-2};\theta) \cdot \ldots \cdot p(y_1|\mathcal{F}_0;\theta). \end{aligned}$$

Now by definition of the measurement equation, the distribution of y_t given y_{t-1} is Gaussian. Specifically, y_t conditional on $\mathcal{F}_{t-1}$ is Gaussian with mean $E[y_t|\mathcal{F}_{t-1}]$ and covariance matrix

$$
\begin{aligned}
Cov[y_t|\mathcal{F}_{t-1}] &= E[(y_t - E[y_t|\mathcal{F}_{t-1}])(y_t - E[y_t|\mathcal{F}_{t-1}])^T|\mathcal{F}_t] \\
&= E[v_t v_t^T|\mathcal{F}_{t-1}] \\
&= F_{t|t-1}.
\end{aligned}
$$

The conditional density $p(y_t|y_{t-1};\theta)$ is thus given as

$$
\begin{aligned}
p(y_t|y_{t-1};\theta) &= \frac{1}{(2\pi)^{\frac{n}{2}}|F_{t|t-1}|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_t - E[y_t|\mathcal{F}_{t-1}])^T F_{t|t-1}^{-1}(y_t - E[y_t|\mathcal{F}_{t-1}])} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|F_{t|t-1}|^{\frac{1}{2}}} e^{-\frac{1}{2}v_t^T F_{t|t-1}^{-1} v_t},
\end{aligned}
$$

a function of the prediction error v_t and its covariance matrix. The parameter n is the measurement dimension. The resulting Likelihood function is only an approximation, hence a Quasi-ML function, which can be expressed in terms of prediction errors v_t and their covariance $F_{t|t-1}$. If we consider the Quasi-Loglikelihood function, we get

$$
\begin{aligned}
\ln(l(y;\theta)) &= \sum_{t=1}^{T} \ln\left(p(y_t|y_{t-1};\theta)\right) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(n\ln(2\pi) + \ln|F_{t|t-1}| + v_t^T F_{t|t-1}^{-1} v_t\right).
\end{aligned}
$$

As we calculate v_t and $F_{t|t-1}$ in each step for the filtering algorithm, each filtering step provides us with an iterative update of the Loglikelihood function due to the current prediction error v_t and its covariance matrix $F_{t|t-1}$. The Loglikelihood function is therefore a side result of the filtering algorithm.
In order to estimate the parameter vector θ, an optimization algorithm is required. Note that as single function evaluation $\ln(l(y;\theta))$ requires the whole Kalman filter to be applied, estimation by Kalman filtering might be computationally slow.

Estimation was conducted using MATLAB, which provides two optimization algorithms: *fminunc* and *fminsearch*. The *fminunc* algorithm attempts to find a minimum of a scalar function of several variables, starting at a specified starting point. It does not guarantee to find the global minimum. The function computes a finite-difference approximation to the Hessian matrix of the scalar function to be optimized. The BFGS Quasi-Newton method with a cubic line search procedure is used. The function to be minimized must be continuous.
Likewise, *fminsearch* also attempts to find a minimum of a scalar function of several variables, starting at a specified starting point. To do so, *fminsearch* uses a simplex search

method, a method which does not use gradients as in *fminunc*. The *fminsearch* algorithm is generally less efficient than *fminunc* for problems of order greater than two. However, when the problem is highly discontinuous, *fminsearch* might be more robust as it can often handle discontinuity, particularly if it does not occur near the solution. As we can not guarantee continuity of our models for all model parameters, this is an important aspect. Namely, for expanded models as the stock model in chapter 3.1, the *fminsearch* algorithm proved to be superior. Again, *fminsearch* gives only local solutions.
In practice, optimization by the *fminunc* algorithm was faster, yet less stable then optimization with the *fminsearch* algorithm. On the other side, *fminsearch* was able to further increase Loglikelihood values given *fminunc* results. Therefore, the *fminunc* algorithm was applied in a first step, the *fminsearch* algorithm in a second step.

Another practical problem is finding reasonable starting values θ_0 for the optimization algorithms. Most parameters do not have an economical interpretation we can use to derive a-priori specifications.
Obviously, the correlation parameters $-1 \leq \rho_{ij} \leq 1$. By definition, the parameter α is positive and we will later see that it equals the asymptotic long rate within the model, so we can impose an upper bound as well. The mean reversion factors κ_i are positive. To guarantee sufficient variation in the state vector components we can assume an upper bound as well. In both models, however, the scaling factors σ and γ remain. Note that if

$$dX_t^{(i)} = \kappa_i(\mu_i - X_t^{(i)})dt + dW_t^{(i)}$$

then by the Ito-Doeblin lemma for $Y_t^{(i)} := \gamma_i X_t^{(i)}$

$$\begin{aligned} dY_t^{(i)} &= [\gamma_i \kappa_i(\mu_i - X_t^{(i)})]dt + \gamma_i dW_t^{(i)} \\ &= \kappa_i(\gamma_i \mu_i - Y_t^{(i)})dt + \gamma_i dW_t^{(i)} \end{aligned}$$

So γ and σ scale volatility of the state vector, which we standardized with $\sigma := Id_d$. As γ and σ can therefore considered as volatility parameters, we can assume these parameters to be bounded as well.
As the bounds we can impose on the model parameters are typically the only prior information available, a simple approach would be to choose starting parameters θ_0 arbitrarily by a uniform distribution on the bounded intervals for each parameter. Generally, imposing sharp bounds increases the chance of reasonable starting points. Nevertheless, sharp bounds might exclude viable parameter choices, which implies again a trade-off between efficiency and estimation quality. In general, if the bounds were chosen too restrictive, the optimization approach implied parameter sets in which the initial bounds were reached, so that in further estimation steps the bounds were widened. Particularly considering the long-term mean μ sharper bounds were crucial to find reasonable initial values, whereas

the estimated parameters μ then frequently laid outside the initial bounds[13].

Both Matlab optimization algorithms use unconstrained data. To implement the limits within the unconstrained fminsearch and fminunc algorithms, we used a simple model to map the constrained parameters θ_i^c to unconstrained parameters θ_i^u, namely transformation works according to

$$\begin{aligned}\theta_i^u &= log((\theta_i^c - b_i^l)/(b_i^u - \theta_i^c)) \\ \theta_i^c &= (b_i^u - b_i^l)\frac{\exp(\theta_i^u)}{1+\exp(\theta_i^u)} + b_i^l.\end{aligned}$$

Starting points can now be derived by choosing every model parameter θ_i^c according to a uniform distribution $\mathcal{U}\left(l_i^l, l_i^u\right)$. In a second step, we run the Kalman filter for each starting set θ_0^c. Then we can sort the starting vectors θ_0^c according to their respective Loglikelihood values. Only parameter sets with sufficiently high Loglikelihood values are then used for further optimization. As initial parameter sets θ_0^c were chosen by a uniform distribution, a sufficiently large set of initial parameter sets θ_0 covers the whole parameter space and choosing those starting values with higher Loglikelihood values should be equivalent to restricting the sets of starting points to a reasonable subset of the parameter space. Furthermore, as optimization algorithms typically provide local extrema only, launching optimization from various starting points is a reasonable approach to check on whether resulting extrema are local or not.

Due to this local maximum problem and the different properties of *fminunc* and *fminsearch* algorithms, it has proved to be effective to follow an iterative approach. After each optimization step, results have to be examined. If the improvement of the Loglikelihood function is small, several possible explanations exist:

1. a local maximum is reached
2. the limits of the *fminunc* algorithm is reached
3. the bounds of the parameter space are reached for at least one parameter θ_i.

In the third case, another optimization step should be started using widened bounds. In the second step, we should continue with *fminunc*, which is typically able to increase Loglikelihood values even more. In the first case, we have to compare the current Loglikelihood value to other Loglikelihood values derived in the recent optimization step. If the respective Loglikelihood value is significantly smaller than other Loglikelihood values, the optimization algorithm is stuck in a local maximum, yet we know that (local) maxima with higher Loglikelihood values exist, so that the current parameter set can be sorted out. This implies that a large set of initial values is required, as many initial values have low

[13] For a discussion of the difficulties in estimating the parameter μ, see also 2.3.8.

Loglikelihood values and therefore do not even enter the first optimization step, and after each step additional parameter sets may be sorted out. Such an approach saves computation time on the long run as it is possible to avoid being stuck in early local maxima of the Loglikelihood function. On the other side, this points to the general problem that we can only derive local maxima, not global maxima. Generally, we believe that if the starting points of the optimization algorithm are sufficiently scattered across the parameter space, the global maximum should be accessible.
Another problem is that due to using a Quasi-ML estimator only, we can not reasonably choose between two distinct local maxima if their respective Loglikelihood values are close as the Loglikelihood value derived is only an approximation to the true Loglikelihood value. Further examinations are required, some of which we will discuss later.

To summarize, the estimation algorithm used can be described as such:

1. Choose N starting sets θ_0^j, $j = 1, \ldots, N$ uniformly distributed within the implicit bounds for each parameter value $b_i^l \leq \theta_i \leq b_i^u$.

2. Calculate $\ln(l(y; \theta_0^j))$ for all starting values $\theta_0^j \in \Theta_0 := \{\theta_0^j : j = 1, \ldots, N\}$. Specify a subset of starting values $\Theta_1 := \{\theta_0^j : j \in J_1\}$, $J_1 \subset \{1, \ldots, N\}$ with high Loglikelihood values for further examination.

3. Start optimization function *fminunc* with a limited number of steps to get θ_{i+1}^j for all $j \in J_i$. Specify a subset of starting values $\Theta_{i+1} := \{\theta_{i+1}^j : j \in J_{i+1}\}$, $J_{i+1} \subset J_i$ for which the Loglikelihood value is high $\ln(l(y; \theta_{i+1}^j) \approx \max_{k \in J_i} \ln(y; \theta_{i+1}^k)$ or has improved sufficiently $\ln(l(y; \theta_{i+1}^j) > \ln(y; \theta_i^j)$ to exclude local maxima. Repeat.

4. Repeat the previous step with the *fminsearch* algorithm.

We found that, as a rule of thumbs, for the cosh model around 75% of the initial parameter sets resulted in reasonable estimates, whereas in the Cairns model it was only around 50%. To estimate extended models as the stock market or exchange rate expansions, which have significantly more parameters, the fraction of reasonable starting values θ_0 was considerably smaller, yet the above described algorithm proved particularly effective in excluding inferior parameter sets as well as increasing computational efficiency.

2.3.3 Data

In specifying the data set to be used in estimating a term structure model we first encounter some basic questions. First, we have to decide on the time horizon of the underlying data. Second we have to decide on the type of market data to be chosen. We will discuss these questions with respect to both banking and insurance applications.

Time horizon

Most term structure models currently in use were developed to consistently price interest rate derivatives. The primary focus therefore lies in consistently fitting the current market situation. The model is calibrated to current data on some observable assets, then model-implied "fair" prices of other assets can be derived. Relative to the cross-sectional fit, the time series behavior of the theoretical price is of minor interest. Some term structure models allow to extract most or even all model parameters from a set of market prices observed at a single point in time. An important example would be the Hull-White model [HW90], which is an extension of Vasicek [Vas77] where the model parameters are specified as deterministic functions in time, which can be calibrated to current market data, for example the current term structure, the current term structure of (implied) volatility and cap data. Whereas this allows to fit parts of fixed income markets exactly, the time series properties as defined by the time-dependent Vasicek parameters depend on the current cross-sectional information only. This approach therefore can not fit any time series properties of interest rates whatsoever.

Now, if we want to price contingent claims with long maturities, path dependent payoffs or if we use simulations in pricing, time series properties of the stochastic factors driving the prices become important as well. The most important aspect of time series properties of interest rates is mean reversion. If we assume interest rates to be mean reverting, the speed of mean reversion and particularly the long-term mean are time series aspects which typically can not be derived from the current term structure alone. The longer the maturity of the asset to be priced, the more important mean reversion and other time series properties become. For example, it is well known that the slope of the term structure is mean reverting and closely related to the business cycle as well as monetary policy. If the slope is an important input for a long-term asset, we should require the model parameters to be estimated from data covering a full business cycle to cover the full variability of the slope and furthermore sufficient examples of the implementation of monetary policy. As this implies sufficient inflation-growth samples, data covering multiple business cycles might be necessary.

Life and pension insurance contracts incorporate path-dependent portfolio allocation decisions as well as path dependent distribution of returns. Pricing typically requires simulation and many insurance contracts have very long times to maturity. These aspects of insurance contracts require realistic time series properties of the model. If we consider, for example, a life insurance contract started by a 35 year old in 1980 to end at the age of 65, the lifetime of the contract covers several recessions, interest rates varied by more than 1000 basis points and stock market indices multiplied. Consequently, for the pricing of insurance products we recommend models which are able to cover such variations and to fit time series properties of interest rates and possibly additional assets, which requires estimation on historical data. As a rule of thumb, the data set used for estimation should span time horizons equivalent to the time horizons required for pricing. In case of insurance applications, this may imply several decades of market data. In general, we can conclude that pricing of contingent claims with very long time horizons and path-dependence require a reasonable fit of time series properties of the underlying variables.

Another aspect to be reckoned with therefore lies in the availability and quality of the data itself, particularly early historical data. For many interest rate derivatives, trading started in the 1980s. Considering government bond data, auctioned maturities may have changed over time[14]. Quality is yet another aspect. In case of interest rate swaps, increasing liquidity, collateralization and other credit enhancements have significantly changed the swap market since its beginning, see for example [Ape03]. US Long-term bond data of the fifties heavily relies upon callable bonds, see [MK93]. Changes in taxation regimes might have substantially altered the after-tax returns of fixed income assets which resulted in changes in portfolio allocation of private investors, see [GO97]. To summarize, one should choose market data which is sufficiently liquid over the whole time period of the dataset and which either did not undergo significant regime changes or whose regime changes might be covered by the model. A typical example in term structure modeling using US data is to either exclude the monetary experiment 1979 to 1982 from estimation data or to choose estimation data which starts significantly earlier to cover a sufficient initial subsample of "normal" term structure behavior.

Market data

We showed that long-term historical market data should be used if time series properties are of importance, which is the case for most insurance applications. The second question considers the type of market data to use. Available historical data for estimation can be partitioned into three groups:

1. interest rates,
2. derivative data and
3. macroeconomic data.

To price interest rate derivatives in the risk-neutral approach, but also in the state price density approach, dynamics of riskless interest rates are required. In general, however, riskless interest rates are not observable. As proxies either the term structure as implied by domestic government bonds or the term structure as given by swap rates is used. In the following, we will discuss these two proxies to the term structure of riskless interest rates.

[14] To give an example, treasury departments have an incentive to end issuance of government bonds with very long times to maturity in case current yields are high, whereas on the other side there is an incentive to increase duration of debt outstanding in case current yields are low.

Government bonds Bonds are traded debt securities in which the borrower owes the owner of the securitized debt the payment of a specified notional amount at maturity as well as coupon payments at specified intermediate dates. The "fair" current value of the bond as observed in the market can then be used to derive the implied term structure. A zerobond is a bond for which all coupons equal zero. In this special case, a simple relation between the price at time t of a bond P which matures at time T, $P(t,T)$, and the spot interest rate, $y(t,T)$, exists by

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

This implies that the annualized return up to maturity of a zerobond P equals the respective spot rate $y(t,T)$. As coupon bonds can be interpreted as portfolios of zerobonds, interpolation algorithms can be used to derive the term structure of interest rates from prices of traded coupon bonds, although this typically implies measurement errors. Published government yield curves therefore necessarily are only approximations to the true yield curve.

Note that the "fair" bond prices depend on liquidity of the bond issue, credit risk associated with the issuer, tax regulations and other factors. For developed nations, bonds denominated in domestic currency and issued by the domestic national government are generally considered free of default risk, although this may be reconsidered due to the aftermath of the 2007 financial crisis, which saw soaring state deficits and a general fear of rating changes even for some of the largest developed economies in the world. Domestic government bonds are typically the most liquid financial assets available in domestic currency, thus liquidity premia are small[15]. For all practical reasons, one can assume that government bond implied spot rates are essentially domestic risk-free interest rates. The usage of government bond implied yield curves as a benchmark is so well established that, according to [(Ch02], Singapore and Hong Kong began issuing government debt without financing needs for the economical benefits of introducing a government bond yield curve benchmark.

Note however that even before the financial crisis positive yield spreads between member states of Eurozone existed. For many Eurozone members the German government-bond implied yield curve was below the domestic government bond implied curve. For some countries, particularly Portugal, Italy and Greece, this should be attributed predominantly to default premiums, whereas for others such as France, Austria or the Netherlands liquidity should be the predominant driver of these yield spreads. Consequently, we can assume the German government bond-implied curve to provide the riskless term structure for the whole of Eurozone. This shows that using government-bond implied yields as a proxy for the riskless term structure is not beyond doubt.

[15]For the impact of liquidity premia, see for example Longstaff [Lon02]

An important aspect of bonds is that new bonds are only issued at specified dates according to the auction cycle. The most recently issued (so called “on-the-run”) treasuries of a specified maturity are typically “on special” in the repo market[16], which means that their respective repo rates are lower than the repo rates on other treasuries due to supply constraints of the treasury securities “on-special”, see Fisher [Fis02] or [Duf96]. Consequently, two treasury securities which are identical besides specialness have a different price and therefore imply different yields, hence treasury data should be adjusted for repo specials prior to usage. Furthermore, considering the approximation of the riskless yield curve, it remains an open question whether treasuries “on special” imply true riskless interest rates or just reflect supply constraints which allow some market participants arbitrage possibilities. Unfortunately, as data considering specialness is not readily available, such adjustments are rarely made[17].

In banking applications, the term structure is predominantly required to discount cash flows of certain assets. In insurance applications we require term structures for two purposes: discounting and bond portfolio modeling. Given their benchmark character for fixed income markets, government bond implied yields can reasonably be used for discounting. On the other side, domestic government bonds are a major part of investment portfolios of most insurance companies, particularly those of continental Europe. Therefore, term structure models used in insurance applications should be estimated and calibrated with government bond implied yield curves.

Swap rates Plain vanilla interest rate swaps are over-the-counter agreements to exchange a cash flow of constant interest payments against a cash flow of floating interest payments, based on a fixed notional amount which is not exchanged. LIBOR is usually used to index the floating payment, whereas the fixed payment is based on the *swap rate*, which is quoted for varying maturities of the respective swap and hence forms a term structure, see also [Sad09] or [RSM04]. The question arises whether swap rates are a reasonable approximation of the riskless interest rate required in pricing.

Government bond implied interest rates and swap rates are typically highly correlated, yet not equal. The difference between the swap rates and government bond implied rates is the swap spread. These spreads vary stochastically, particular in times of economic

[16] A ”repo” involves one investor selling treasuries today and agreeing to a buy back of these treasuries at a specific price on a specific future date. A repo can therefore be interpreted as a collateralized loan, whereby the seller of the treasuries provides these treasuries as collateral for a loan whose interest rate is specified by the price difference between the sell and buy-back price of the collateral treasuries. This interest rate is called the repo rate. Those treasuries not on special can be interpreted as ”general collateral” or interchangeable for repo loans, hence their implied interest rate is independent of the respective general collateral used for the repo. This rate is called the general collateral rate. For a general introduction to repos, see [RSM04].

[17] We use the datasets of the Federal Reserve and the Bundesbank, respectively, for which such adjustments were not made either.

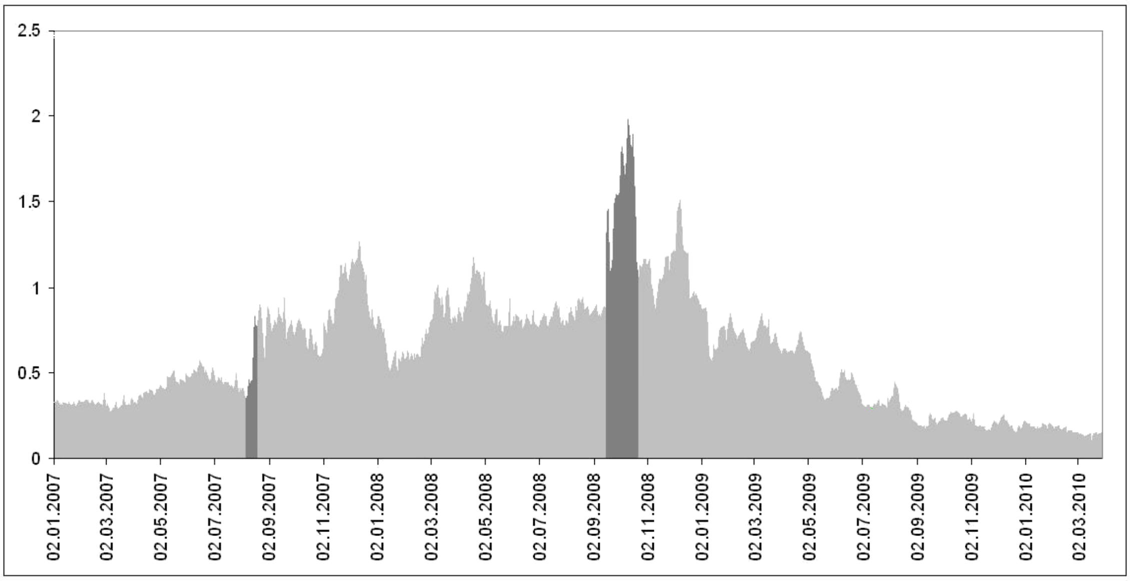

Figure 2.2: Daily swap spreads during the 2007 to 2010 financial crisis. Dark gray are first the "black swan in the money market", see [TW08] and the jumps associated with the collapse of Lehman brothers and the rescue of AIG and the reverse jump due to the recapitalization of the US banking system by 250 billion dollars.

crisis, see figure 2.2. Sun, Sundaresan and Wang [SSW93] show that there exists a default premium in the swap spread, albeit smaller than the default premium in the bond market. The collapse of the inter-banking market following the Lehman Brothers bankruptcy resulted in a sudden change in swap rates in the same way as LIBOR rates grew. As a consequence, credit risk in swap contracts may have been underestimated during previous years. Liu, Longstaff and Mandell [LLM02] support liquidity risk as a primary determinant of the swap spread. As the swap spread is typically positive and may be explained by positive risk premia, treasury yields are a superior approximation of riskless interest rates. This is of particular importance through financial crisis, as can again be seen by figure 2.2. The higher the swap spread, the higher the deviation of swap yields from riskless interest rates.

In times of financial crisis, swap rates should not be used as an approximation for riskless interest rates. In normal times, identified by low and persistent swap spreads, such an approximation might be valid. Nevertheless, modeling investment results in fixed income markets should use the benchmark of government bonds. It might be of interest whether swap spreads comove with certain investment grade fixed income securities such as asset-backed securities like Pfandbriefe, MBS or corporate bonds. If this is the case, the swap spread might be used as an approximation to returns until maturity of these risky assets. Note, however, that Sun, Sundaresan and Wang [SSW93] showed that the default premium in the swap market is smaller than in bond markets. Considering the long time horizons of insurance contracts and therefore the higher probability of a financial crisis during lifetime of the contract, we recommend using government bond-implied yields for insurance applications. Note also that swap spreads are typically positive, thus discounting future payoffs with government bonds is more conservative than discounting with swap rates.

Another important factor which prohibits usage of swap yields for insurance applications is availability of data. Whereas treasury securities are traded since before world war II, swap trading began only in the eighties with OTC swap contracts. In particular, available swap yields do not include the periods of the oil price shocks of the early 1970s or the monetary experiment in the early 1980s.

On the other side, swaps are by construction predominantly used for interest rate risk hedging, and thus became very popular and highly liquid instruments. As swap rates for constant maturities are effectively quoted continuously, swap markets provide true constant maturity yield data, whereas in government bond markets the available maturities depend on the auction cycle[18].

The usage for hedging already shows the importance of swap rates in the banking sector. For banking applications, a major criteria for the choice of the appropriate term structure is the underlying of the contingent claim to be priced. For government bond futures, government bond term structures should be used to capture the dynamics of the underlying. Caps, Floors and Swaptions work on inter-bank offered rates and swap rates, hence these underlyings should be used. Note, however, that this does not necessarily exclude the usage of government bond implied term structures. The question whether interest rates which contain sizable credit spreads should be used for discounting remains. In case of sizable swap spreads, we recommend using both government bond-implied yields and swap yields, the first for discounting, the second as underlying. Therefore the importance of government bond implied yield curves for banking applications rises with the swap spread.

Derivative data Collin-Dufresne and Goldstein [CDG02] find that swap rates have only limited explanatory power for the returns of at-the-money straddle-portfolios, that is portfolios of at-the-money caps and floors highly dependent on swap rate volatility. The authors call this finding *unspanned stochastic volatility*. Now as the three principal components of Litterman and Scheinkman determine term structure dynamics of swap rates, this finding implies that straddle portfolios are subject to factors unspanned by level, slope and curvature. In a related paper, Heidari and Wu [HW01] find that the three factors of Litterman and Scheinkman are sufficient to match bond market movements, pricing derivatives however requires three additional factors[19]. For a general overview to calibration and inconsistent interest rate derivative markets see for example [RSM04] and [Reb02]. Two explanations for these findings come into mind:

[18] See Dai and Singleton [DS00]

[19] Note that contrary to that Fan, Gupta and Ritchken [FGR03] find that the effect of unspanned factors on swaptions is minor.

- The unspanned factors principal components, yet not among level, slope and curvature,
- The unspanned factors are indeed unspanned by the term structure.

In either case, it is obvious that interest rate derivatives depend on the driving factors in an utterly different way than the term structure[20]. Inclusion of derivative data into estimation would therefore give a better insight in dynamics of factors besides the three factors of Litterman an Scheinkman. If unspanned factors exist, the dynamics of these unspanned factors can only be derived using derivatives of the same type. If the derivatives depend crucially on spanned factors beside level, slope and curvature, then again using derivative data in estimation is recommended as the factor to be derived is clearly dominated by level, slope and curvature, and possibly additional principal components with higher impact then the factor in question. Particularly, if unspanned factors exist for certain derivatives, even a perfect fit of the initial term structure is not sufficient to price these derivatives. On the other side, if we are interested in term structure data only, a 3-factor term structure model estimated with swap or bond yields is sufficient.
As a consequence, we can fully recommend the practitioners' approach of fitting the term structure model to derivatives of the same type as those to be priced. To give an example, to price an option on government bonds, the model should be estimated on term structure data as well as options on government bonds with varying maturities both of the options themselves and of the underlying government bonds. Term structure data then covers dynamics of the underlying, in particular level, slope and curvature, whereas option data covers dynamics unspanned by level, slope and curvature. Such an approach guarantees that potential unspanned factors as well as higher principal components enter into estimation. This approach is especially recommended for pricing in the banking sector.

In insurance applications, our main interest lies in the correct specification of term structure dynamics for discounting payoffs and simulation of investment returns, thus dynamics of the term structure are sufficient and therefore level, slope and curvature. Furthermore, observable market prices of "similar" insurance derivatives typically do not exist. Floors provide minimum rates, in a sense similar to guaranteed returns within insurance applications. However, the underlying for traded floors is LIBOR, whereas for insurance applications a single underlying does not exist but rather insurance returns depend on bond and stock returns at least. If the LIBOR-treasury spread is small, the differences are of minor interest. Recent experience however implies that Japan scenarios, hence worst case scenarios for insurance companies for which hedging would be crucial, coincide with increasing LIBOR-treasury spreads, hence in the moment the floor is required for protection the LIBOR rate might be significantly higher than the treasury rate and maybe also

[20]Note that this has important consequences on the question whether the bond market is complete or not.

higher than the return achieved by the insurance company so that only an imperfect hedge results. We can conclude that insurance instruments differ from most interest rate derivatives in terms of underlying, maturity, path-dependence and extremal situations so that typically no interest rate derivatives of a type sufficiently close to the insurance contract are available.

For the usage of interest rate derivatives for estimation more practical points regarding implementation exist as well. Such practical considerations exist mainly due to availability of data on one side and the ability of the respective model to derive prices on the other side.
Whereas term structure data is available for decades, interest rate derivatives are rather new financial innovations. Time series of interest rate derivative prices typically start in the 80s or even later. For financial innovations, regime changes are frequently found in the early years, see [Ape03] for the swap markets. Second, liquidity may have been low during the early years. Furthermore, many interest rate derivatives were initially traded over-the-counter. Extraction of quotes in OTC markets may be difficult.
Considering implementation, note that the model used not necessarily allows for closed formulae of interest rate derivatives. Therefore, Monte Carlo approaches have to be used for pricing. These approaches are not deterministic, the resulting prices approximate the true prices only for high numbers of trials. There exists a trade-off between computational speed in estimation on one side and deterministic input data on the other side which makes estimation based on Monte Carlo methods difficult.
Finally, model restrictions might determine the usage of derivative data. Of particular interest in this case is the volatility smile, which interest rate derivatives show as well, see [JLZ07]. Analogously to the Black-Scholes framework, certain term structure models might be unable to fit a collection of "‘smiling"’ interest rate derivatives at the same time.

Macroeconomic data The dependence of interest rates on macroeconomic data is well known. Whereas long-term interest rates predominantly reflect inflation expectations, short-term interest rates and particularly the slope reflect monetary policy, enacted by the central bank according to the current outlook on inflation and economic activity. It is therefore natural to assume that macroeconomic variables contain information about the term structure. Indeed, according to Ang and Piazzesi [AP03], macroeconomic variables explain up to 85% of the dynamics of short- and intermediate-term yields, but explain only around 40% of the dynamics of the long end of the yield curve. These authors also derive that Litterman and Scheinkman's level factor remains almost intact if macroeconomic variables are incorporated, but macroeconomic variables, particularly inflation, explain a significant part of the variation of the slope. Moreover, they find macroeconomic variables in a term structure model to improve forecasts. The problem arises which macroeconomic

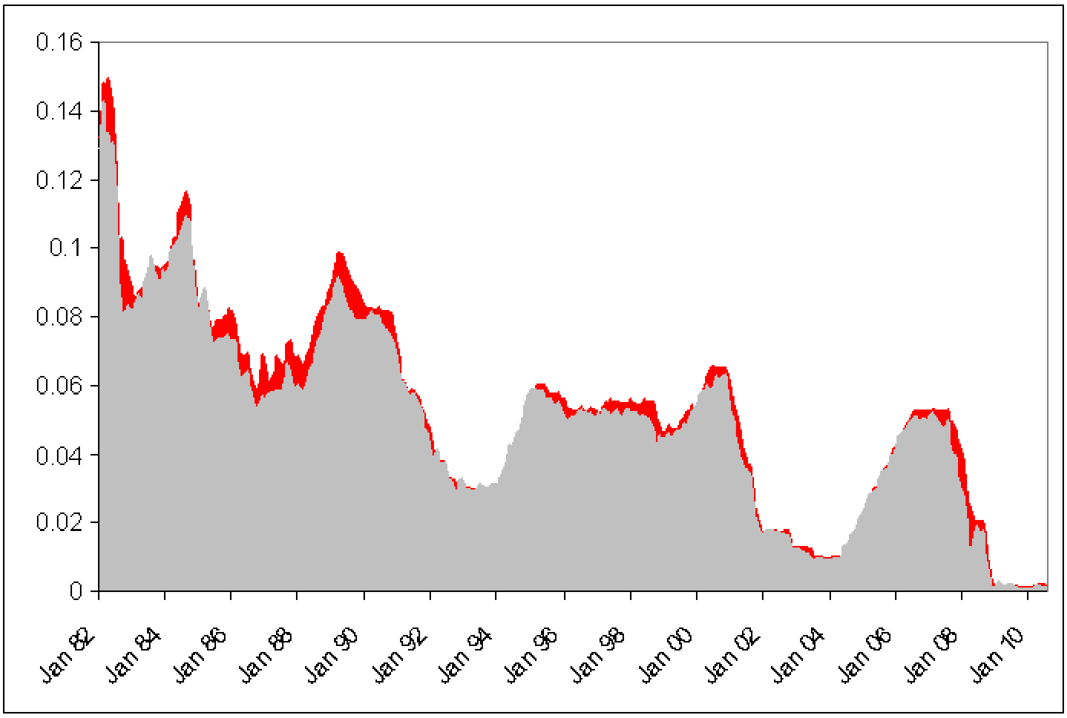

Figure 2.3: Time series of the 3-month Treasury bill rate (light gray) and the (positive) spread of the federal funds rate in US data from 1984 to 2008.

data could add significant information.

Monetary policy sets the current target rate according to lagged, coincident and leading macroeconomic variables. In turn, the monetary transition mechanism determines the yield curve according to the macroeconomic information. Finally, the changes in the yield curve affect economic activity and inflation through investment decisions of the private actors in the market. Therefore, macroeconomic variables together with the yield curve describe the current state in the macroeconomic continuum, whereby each variable holds information about current, past and future states. We therefore can include macroeconomic data to overcome the restrictions of a Markovian context in term structure data. Although macroeconomic variables can easily be described by Markov processes, they contain information about past yield curves, which determined current inflation and economic activity, and future yield curves as implied by the reaction of the monetary authority to the current and expected macroeconomic situation. This essentially explains the stylized fact in term structure modeling that inclusion of macroeconomic variables improves forecasting ability of term structure models.

Considering the macroeconomic indicators to be included, the domestic inflation rate comes into mind. Note, however, that there is typically not a unique published inflation rate. In case of the US, for example, different inflation rates are published with emphasis on urban versus rural communities and inclusion of energy, food and tobacco costs. Another stylized fact is that long-term interest rates are determined by inflation expectations,

so rather than including current inflation rates one could include inflation expectations directly taken from survey data or extracted from prices of inflation-indexed government bonds.
Even more choices exist for measures of macroeconomic activity. A standard choice would be to use aggregate economic activity, hence GDP or GNP growth, either in nominal or real terms respectively. Again, market forecasts could be included directly to mirror the fact that the central bank incorporates expectations in the same way as lagged values. On the other side, note that purchasing manager indices as well as stock market indices are important leading variables closely related to economic activity.
In the literature, the third macroeconomic factor included is typically the central bank's policy instrument, for example the federal funds rate in the US. From an arbitrage argument, however, it is clear that an overnight government bond yield should be extremely close to the overnight federal funds rate. It is close and not equal since the federal funds rate as an inter-bank reference rate is subject to counterparty risk, seasonal and regulatory effects, notably year-end effects, due to sudden liquidity changes in banks. Figure 2.3 shows the 3-month treasury bill rate and the federal funds rate over time. Obviously, including federal funds rate data will not include genuinely new information, but only a highly volatile additional proxy of the shortrate subject to externalities such as inter-bank market liquidity, particularly at month and year's end.

If we include more than one macroeconomic time series, we have to explicitly consider dependencies of these time series. An important example would be the famous Taylor rule [Tay93], which links current inflation, economic growth and short-term interest rates with their respective long-term means. As the Taylor rule describes very well historical behavior of many central banks, we can assume that the joined development of inflation, economic growth and short-term interest rates persists into the future, which must be considered in simulations. By assuming that the deviation of the joint indicators from the Taylor rule is observed with mean zero, we can easily implement the rule into Kalman filter and EMM estimation. This does not, however, imply that simulated data follows the rule as well.

Whereas using derivative data in estimation is expected to improve the model fit, particularly considering so called "unspanned" factors like stochastic volatility, implementation of estimation approaches which employ derivative data is often difficult. Augmenting the estimation data set with macroeconomic data, however, is rather simple. Most term structure models are factor models, that is an underlying state process X drives the dynamics of the term structure through time. The factor process components typically coincide with Litterman and Scheinkman's principal components. Many macroeconomic indicators can be expressed as mean reverting rate processes. We can define a component process to coincide with the observable macroeconomic rate[21]. For a discussion of implementation

[21] For example, we can define the inflation rate i_t to be measured by component k of the state vector

possibilities see also 3.2. To summarize, we can expect the inclusion of macroeconomic data to be significantly easier than the inclusion of derivative data. The question is which macroeconomic indicators to choose. Following the literature, we would recommend using current annualized inflation and GDP growth rates. An interesting approach to the problem of choice can be found again in Dai and Singleton [DS00], who derive the first principal component of a group of inflation measures and a group of measures for economic activity, respectively, as their macroeconomic variables to be included.

Considering the impact of macroeconomic variables on banking applications, the main focus lies on derivatives which have the respective macroeconomic variable as an underlying, for example inflation caps or derivatives on output. A term structure model which explicitly takes into consideration the dependencies between the macroeconomic variable and the term structure allows to derive at least Monte Carlo simulated prices of derivatives on the macroeconomic variable with stochastic interest rates[22]. Another application stems from improved forecasting ability, as essentially such a term structure model can forecast arbitrage-free term structure changes contrary to classical statistical forecasting approaches which can not guarantee no-arbitrage. Besides these special cases, however, the benefit of including macroeconomic variables into term structure models for banking applications is limited.

Considering insurance applications, however, macroeconomic variables would be important in long-term simulations, as they help to derive realistic long-term dynamics of the term structure model, particularly with respect to business cycles. Furthermore, long-term interest rates should be more realistic due to inflation expectations. Finally, we can expect that any additional asset included into the framework, particularly the stock market, should depend on the very same macroeconomic variables as the term structure. As a consequence, inclusion of macroeconomic variables into both a term structure and a stock market model implies a more realistic handling of interdependencies of these financial markets.

The dataset

The dataset used consists of end-of-month US term structures implied by treasury securities from january 1984 to january 2008 with maturities of 3 and 6 months and 1, 2, 3, 5, 7, 10, 20 and 30 years. The dataset is obtained from the Federal Reserve download portal.

---

process X_t within a state space approach by

$$i_t = e_k^T X_t + \epsilon_t^{(k)},$$

where e_k is the k-th unit vector. This also defines the required measurement equation for a Kalman filter.

[22] Note that in such models we have to derive the no-arbitrage condition for each of these derivatives respectively.

Although for many maturities substantially longer time series are available, rates for treasury bills with maturities of less than a year are only available from january 1982 on. We excluded the years 1983 and 1984 due to the monetary experiment[23], since the starting point of the Kalman Filter approach is crucial for estimation and yields as observed during the monetary experiment are not valid as starting points. The data ends in january 2008 to avoid problems due to the excessive rate cutting and non-traditional instruments of monetary policy in the aftermath of the subprime crisis. During the 23 years of observations the US treasury department ceased to issue and restarted issuance of both the 20 and 30 year treasuries, thus only for about half of the data sample all rates are available. However, for all observation points at least one rate with a maturity over 10 years is available.

Unlike the US, Germany does not issue government bonds with maturities of less than one year – at least the Bundesbank does not provide them. We estimated model parameters using German end-of-month term structure data with maturities of 1 to 10 years from the Bundesbank statistical download portal. The dataset starts september 1972 up to July 2008, again due to availability questions considering the different maturities and to guarantee a reasonable starting point for the Kalman filter before the monetary experiment and the equivalent period in Germany.

2.3.4 Estimation of the Cairns model

Implementation of the Kalman filter

To implement the Extended Kalman filter in case of the Cairns model, we provide the respective equations following 2.3. Implementation of an Extended Kalman filter for the Cairns model may also be found in [Lut07]. First, we have to specify the starting points $X_{0|0}$ and $\Sigma_{0|0}$. To do so, we follow the previously described approach and fit $X_{0|0}$ to the first two yield curves Y_0 and Y_1 in initial estimation iterations and to Y_0 for later iteration steps. For the initial covariance matrix $\Sigma_{0|0}$, we took the conditional covariance matrix of the state vector X over a time period of length 1.
For the Cairns model, the state vector dynamics follow an Ornstein-Uhlenbeck process

$$dX_t \quad = \quad \kappa(\mu - X_t)dt + CdZ_t^{\tilde{\mathcal{P}}}.$$

Following theorem 2.2.16, the vector X_{t+1} conditional on X_t is distributed according to

$$\begin{aligned} X_{t+1}|X_t \quad &\sim \quad E[X_{t+1}|\mathcal{F}_t] + \eta_t \\ &= \begin{pmatrix} (1-e^{-\kappa_1 1})\mu_1 \\ \vdots \\ (1-e^{-\kappa_d 1})\mu_d \end{pmatrix} + \begin{pmatrix} e^{-\kappa_1 1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-\kappa_d 1} \end{pmatrix} \begin{pmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(d)} \end{pmatrix} + \eta_t(\theta) \end{aligned}$$

[23]In january 1982, the lowest rate was the 3-month treasury bill rate of 12.92%.

whereby η_t is normally distributed with mean zero and covariance matrix $Q_t(\theta)$ given by

$$\begin{aligned} Q_t(\theta) &= Cov\left[X_{t+1}|X_t\right] \\ &= \left(\left(\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\rho_{ij}}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)}\right)\right)\right)_{i,j=1,\ldots,d}. \end{aligned} \quad (2.24)$$

and θ denotes the vector of model parameters to be estimated. Since $X_{t|t-1} = E[X_t|X_{t-1}]$ and $\eta_t(\theta)$ is normally distributed, this defines the transition equation of the state vector by

$$\begin{aligned} X_{t|t-1} &= c_t(\theta) + \Phi_t(\theta)X_{t-1|t-1} \\ &= (1-e^{-\kappa})\mu + e^{-\kappa}X_{t-1|t-1}. \end{aligned}$$

For transition of the conditional state covariance matrix Σ, we have

$$\begin{aligned} \Sigma_{t|t-1} &= \Phi_t(\theta)\Sigma_{t-1|t-1}\Phi_t(\theta)^T + Q_t(\theta) \\ &= e^{-\kappa}\Sigma_{t-1|t-1}(e^{-\kappa})^T + Q_t(\theta). \end{aligned}$$

This completes the transition step. Next, we derive the updating step of the filter, which requires definition of the measurement equation, measurement errors and their covariance matrix $F_{t|t-1}$ and finally the Kalman gain matrix. In the Cairns model, a bond price with time to maturity τ_i, $i = 1,\ldots,n$, is given by

$$P(t, t+\tau_i) = \frac{\int_{\tau_i}^{\infty} H(u, X(t))du}{\int_0^{\infty} H(u, X(t))du}$$

hence the spot rate for the same maturity is given by[24]

$$\begin{aligned} y(t, t+\tau_i) &= -\frac{1}{\tau_i}\log\left(P(t, t+\tau_i)\right) \\ &= \underbrace{\frac{1}{\tau_i}\log\left(\frac{\int_{\tau_i}^{\infty} H(u, X(t))du}{\int_0^{\infty} H(u, X(t))du}\right)}_{g_i(X_t;\theta)} \end{aligned}$$

for $i = 1,\ldots,n$. Assuming all observations are subject to a measurement error ϵ_t, this defines the measurement equation by

$$\begin{pmatrix} y^M(t, t+\tau_1) \\ \vdots \\ y^M(t, t+\tau_n) \end{pmatrix} = \underbrace{\begin{pmatrix} g_1(X_t;\theta) \\ \vdots \\ g_n(X_t;\theta) \end{pmatrix}}_{:=g(X_t;\theta)} + \epsilon_t(\theta)$$

[24] In order to emphasize the role of the time to maturity $\tau_i := T_i - t$ for each yield, we write the time of maturity by $T_i = t + \tau_i$.

with $y^M(t, t+\tau_i)$, $i = 1, \ldots, n$ being the interest rates as observed in the market and ϵ_t being a multivariate normal error term with covariance matrix

$$Cov(\epsilon_t) = H_t(\theta) = diag(\nu, \ldots, \nu) \in \mathbb{R}^{n \times n}.$$

This is a simplifying choice to derive a single model parameter for measurement errors. In reality, model misspecifications will likely lead to both cross- and autocorrelated errors, hence possibly $Cov(\epsilon_i(t), \epsilon_j(t)) \neq 0$ and $Cov(\epsilon_i(t), \epsilon_i(t+h)) \neq 0$. However, correlated measurement errors ν_{ij} imply up to $\frac{n(n-1)}{2} - 1$ additional model parameters. These additional parameters do not contribute to explain model dynamics, but are merely an analytical instrument for the errors.

Next we require the first order Taylor expansion of the measurement equation at $X_{t|t-1}$,

$$g(X_t; \theta) \approx g(X_{t|t-1}; \theta) + B_{t|t-1}(X_t - X_{t|t-1})$$

for $i = 1, \ldots, n$ with

$$B_{t|t-1} = \left. \begin{pmatrix} \frac{\partial}{\partial x_1} g_1(x;\theta) & \ldots & \frac{\partial}{\partial x_d} g_1(x;\theta) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} g_n(x;\theta) & \ldots & \frac{\partial}{\partial x_d} g_n(x;\theta) \end{pmatrix} \right|_{x = X_{t|t-1}} \in \mathbb{R}^{n \times d}.$$

The derivatives are given by

$$\begin{aligned} \frac{\partial}{\partial x_j} g_i(x;\theta) &= \frac{1}{\tau_i} \frac{\partial}{\partial x_j} \log \left(\frac{\int_{\tau_i}^{\infty} H(u,x)du}{\int_0^{\infty} H(u,x)du} \right) \\ &= \frac{1}{\tau_i} \frac{\int_0^{\infty} H(u,x)du}{\int_{\tau_i}^{\infty} H(u,x)du} \left(\frac{\partial}{\partial x_j} \frac{\int_{\tau_i}^{\infty} H(u,x)du}{\int_0^{\infty} H(u,x)du} \right) \\ &= \frac{1}{\tau_i} \frac{\int_0^{\infty} H(u,x)du}{\int_{\tau_i}^{\infty} H(u,x)du} \left(\frac{\int_{\tau_i}^{\infty} \frac{\partial}{\partial x_j} H(u,x)du \int_0^{\infty} H(u,x)du}{\left(\int_0^{\infty} H(u,X(t))du\right)^2} \right. \\ &\quad \left. - \frac{\int_{\tau_i}^{\infty} H(u,x)du \int_0^{\infty} \frac{\partial}{\partial x_j} H(u,x)du}{\left(\int_0^{\infty} H(u,x)du\right)^2} \right) \\ &= \frac{1}{\tau_i} \left(\frac{\int_{\tau_i}^{\infty} \frac{\partial}{\partial x_j} H(u,x)du}{\int_{\tau_i}^{\infty} H(u,x)du} - \frac{\int_0^{\infty} \frac{\partial}{\partial x_j} H(u,x)du}{\int_0^{\infty} H(u,x)du} \right). \end{aligned}$$

With

$$\frac{\partial}{\partial x_j} H(u,x) = \sigma_j e^{-\kappa_j u} H(u,x)$$

we get

$$\begin{aligned} &\frac{\partial}{\partial x_j} g_i(x;\theta) \\ &= \frac{1}{\tau_i} \left(\frac{\int_0^{\infty} \sigma_j e^{-\kappa_j u} H(u,x)du}{\int_0^{\infty} H(u,x)du} - \frac{\int_{\tau_i}^{\infty} \sigma_j e^{-\kappa_j u} H(u,x)du}{\int_{\tau_i}^{\infty} H(u,x)du} \right). \end{aligned}$$

The covariance of the prediction error $F_{t|t-1}$ is given by

$$\begin{aligned} F_{t|t-1} &= Cov\left[v_t|\mathcal{F}_{t-1}\right] \\ &= B_{t|t-1}\Sigma_{t|t-1}B_{t|t-1}^T + H_t, \end{aligned}$$

for which all components are known by now. Therefore the Kalman gain matrix

$$K_t = B_{t|t-1}\Sigma_{t|t-1}F_{t|t-1}^{-1}$$

is fully specified and completes the updating steps according to the Extended Kalman filter algorithm given in page 61.

Two preliminary observations can be made from this derivation of the Extended Kalman filter for the Cairns model:

1. Each Kalman filter step requires evaluation of the integrals
$$\int_{\tau_j}^{\infty} H(u,x)du$$
and
$$\int_{\tau_j}^{\infty} e^{-\kappa_i u}H(u,x)du$$
for $j = 0, 1, \ldots, n$ with $\tau_0 = 0$ and $i = 1, \ldots, d$ and $x \in \mathcal{X}$. As these integrals can not be solved in closed form, numerical integration has to be used. This is computationally demanding.

2. For practicability, a finite upper integration bound has to be assumed for numerical integration. This is an additional approximation within the filter, although by definition $H(u,x) \to 0$ and $e^{-\kappa_i u}H(u,x) \to 0$ for $u \to \infty$, respectively. The impact of a finite integration bound should therefore be minimal if it is chosen large enough.

3. We can not guarantee that the prediction error covariance $F_{t|t-1}$ is invertible. In practice, MATLAB provides by $pinv(A)$ a pseudo-inverse matrix B to the matrix A, for which $ABA = A$, $BAB = B$ holds and both AB and BA are symmetric. This implies yet another approximation in the filter. This final approximation is, however, a typical problem of the Kalman filter. Since the matrix $F_{t|t-1}$ depends on the current state X_t, we can not guarantee a priori that $F_{t|t-1}$ is invertible for all $t \geq 0$. In fact, we found this to be the main reason the Kalman filter algorithm stopped from an error.

Testing the Kalman filter

To test the ability of the Kalman filter to estimate the model parameters, we specify exogenously a set of model parameters. Using these parameters, we simulate a sample of

κ_1	κ_2	α	γ_1	γ_2	ρ_{12}	μ_1	μ_2	ν	LogL
0.6	0.06	0.04	0.6	0.4	-0.5	0.0	0.0	0.0	
0.5998	0.0600	0.0400	0.5717	0.3993	-0.5178	-0.66	6.30	0.00003	26168
0.5993	0.0601	0.0400	0.5679	0.4010	-0.4013	0.47	9.34	0.00002	26302
0.5993	0.0601	0.0400	0.5694	0.4007	-0.4220	0.54	3.78	0.00002	26304
0.6007	0.0600	0.0400	0.5774	0.3991	-0.4868	0.41	-4.37	0.00002	27180
0.0600	0.6006	0.0400	0.3993	0.5783	-0.4912	-6.39	-0.22	0.00001	28210

Table 2.2: Estimates of the Kalman filter for data simulated by the Cairns model using the parameters of the first line.

289 yield curves, the same number of observations as in the empirical data used later. The Kalman filter can thus be used to estimate the model parameters from the simulated dataset, for which the true model parameters are known. If the Kalman filter is correctly specified, the estimated model parameters should coincide with the "true" model parameters. We propose to use the exemplary parameters used by Cairns himself in his paper, $\kappa = (0.6, 0.06)'$, $\alpha = 0.04$, $\sigma = (0.6, 0.4)'$, $\rho = -0.5$ and $\mu = (0, 0)$.

We followed the iterative approach discussed earlier, starting with the *fminunc* algorithm and continuing with *fminsearch*. However, we limited the amount of *fminsearch* steps for computational efficiency, whereas with real data iteration was stopped only if Loglikelihood values did not improve any further. We derived 20 estimates of the Cairns model. Many estimates had to be excluded during the iterative estimation approach because the implicit parameter bounds were reached or the algorithm reached local minima. Furthermore, in many cases the filter stopped since $F_{t|t-1}$ was singular.

According to table 2.2, the Kalman filter estimates most parameters remarkably well, including correlation. The last estimate shows that the state vector components are exchangeable, which means that the filter detects state vector dynamics, yet the order of state vector components is not specified uniquely. Note, however, that one could determine the order by imposing the limits of the parameters accordingly in estimation, particularly the limits of the mean reversion parameters κ_i.
The filter has significant problems in estimating the long-term mean μ_i of the low mean-reversion factor. We will discuss this further in section 2.3.8. Considering correlation, the parameter ρ is estimated very well, although it varied considerable during the iteration steps and typically only the last *fminsearch* steps produced a reasonable estimate of ρ. All in all, these results justify the usage of the Kalman filter for the Cairns model.

Results

As our first estimation step for the Cairns model, several starting values were chosen randomly from a uniform distribution between specified upper and lower bounds of the model

α	σ_1	σ_2	κ_1	κ_2	ρ_{12}	μ_1	μ_2	ν	LogL
0.0053	0.695	0.690	0.575	0.023	-0.031	-0.677	4.752	0.0014	16609
0.0052	0.693	0.695	0.580	0.023	0.013	-0.595	5.204	0.0014	16609
0.0179	0.330	0.561	0.623	0.020	-0.171	-1.723	-2.36	0.0015	16562

Table 2.3: QML parameter estimates of the Cairns model using the Extended Kalman filter.

parameters. Of these starting values, those who provided the highest Loglikelihood values were taken for further iterative estimation steps using both $fminunc$ and $fminsearch$ until the Loglikelihood did not change anymore.

All in all, 3 viable parameter sets were estimated given in table 2.3. We see that the mean reversion factors κ_i are rather close for all estimates. There exists a high-mean reversion factor and a low-mean reversion factor, as proposed by Cairns. The correlation parameter ρ and the weighting parameters γ_i are relatively stable as well, with only the last estimate deviating a bit more. The long-term means μ_i of the state vector components vary substantially, particularly for the low mean reversion state vector component, as we expected from our previous test. As ν is the estimated measurement error, lower values of ν indicate superior historical fit. Therefore, the first two estimates seem to be slightly superior.

As a second step, we recommend analyzing historical fit of the model. To do that, we analyze the residuals defined as the difference between the model implied yields

$$y(\hat{X}_t; t, t + \tau_i) := g_i(\hat{X}_t; \theta)$$

and the observed yields $y(t, T)$. Table 2.4 shows mean absolute pricing errors in basis points for the parameter estimates introduced above, defined by

$$MAE(\tau) := \left| y(t, t + \tau) - y(\hat{X}_t; t, t + \tau) \right|.$$

Mean absolute errors are a suitable criterion to derive overall time series fit of the term structure model in an economically interpretable way if provided in basis points. The MAEs vary around 10 basis points, which implies that the model fits historical data rather well. As one can clearly see in table 2.4, the mean absolute error of the 7-year yield is the global minimum and at a maturity of 1 or 0.5 years respectively there is a local minimum. The short and long ends show the highest pricing errors. This is a pattern which we would also expect from curvature mismatch, that is the model covers level and slope, but fails to describe curvature dynamics.

Note that as a stylized fact in multi-factor term structure modeling, the estimated historical factors coincide with some of the principal components of Litterman and Scheinkman. Consequently, we compared the filtered state processes to empirical proxies of the principal

Mean Absolute Pricing Errors in Basis Points									
0.25	0.5	1	2	3	5	7	10	20	30
15.12	7.25	7.14	12.34	11.45	8.15	6.43	8.98	10.13	11.34
15.12	7.26	7.18	12.36	11.49	8.19	6.50	9.01	10.20	11.32
15.28	6.73	7.63	14.02	12.19	8.24	6.54	9.41	11.32	12.37

Table 2.4: Mean absolute pricing errors (MAEs) in basis points for the US term structure 1984 to 2008 for three parameter sets from QML-estimation using the Extended Kalman filter on the Cairns model.

components. The results are given in figure 2.4. We find a clear correlation between the low-mean reversion factor and the level measured as the long-term yield. In section 2.1 we recommended such a specification to solve the excess volatility problem in long-end yields. Since the high mean reversion factor is related to the slope, although less clear. The factor deviates from the slope if the level is low, we assume the zero lower bound of the Cairns model to be responsible. In a three-factor model, we can expect that the additional factor covers curvature.

Besides measurement error and historical fit by MAEs we also recommend examination of cross-correlations of the residuals $g_i(\hat{X}_t;\theta) - y(t, t+\tau_i)$ for each $i = 1, \ldots, n$. We assumed mutually independent measurement errors $\epsilon_t^{(i)}$, $i = 1, \ldots, n$. Consequently, if residuals merely reflect measurement errors, cross-correlations of residuals should be close to zero. On the other side, if cross-correlations are significantly away from zero, this points to systematic deviations of residuals which might be explained by a more parsimonious model. If the model does not catch curvature dynamics correctly, which we assume to be the case for a two-factor model, then cross-correlation should show a certain pattern of positive and negative correlations of the time series of residuals. If the model fails to explain curvature dynamics, then model-implied short and long rates will tend to deviate in the same direction from the true observed yields. On the other side, model-implied medium rates will tend to deviate from the true observed yields in the opposite direction. We would expect a correlation matrix with a pattern given in figure 2.5. Table 2.6 shows the calculated cross-correlations of the residuals of the first estimate, the cross-correlation matrix of errors[25] indeed shows the predicted pattern.

Another important criterion in model analysis is residual autocorrelation. High autocorrelation may imply a systematic, transitionary factor in the residuals. On the other side, any misspecification of the state $X_{t-1|t-1}$ may only be corrected over time by the updating step. If $X_{t-1|t-1}$ is misspecified, $X_{t|t}$ tends to be misspecified as well, albeit to a

[25] Note that 20- and 30-year rate data is censored and in particular not sufficient to estimate correlation between the 20- and 30-year rates.

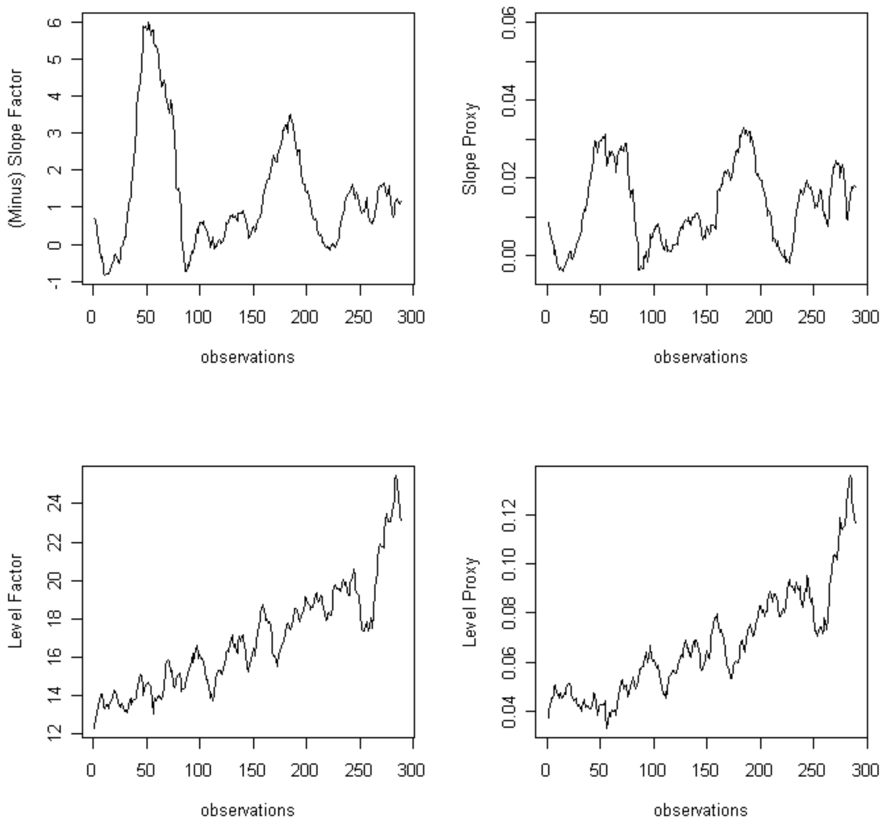

Figure 2.4: Filtered state vector components (left) and empirical proxies of the first two principal components of term structure dynamics.

+	-	+
-	+	-
+	-	+

Table 2.5: Expected pattern in residual cross-correlation matrices in case of curvature mismatch.

maturities	0.25	0.5	1	2	3	5	7	10	20	30
0.25	1.00	0.68	-0.60	-0.90	-0.89	-0.63	-0.32	0.13	0.52	0.30
0.5	0.68	1.00	-0.09	-0.77	-0.87	-0.81	-0.61	0.00	0.42	0.36
1	-0.60	-0.09	1.00	0.43	0.30	-0.06	-0.37	-0.26	-0.20	-0.04
2	-0.90	-0.77	0.43	1.00	0.94	0.63	0.32	-0.32	-0.60	-0.31
3	-0.89	-0.87	0.30	0.94	1.00	0.79	0.52	-0.15	-0.57	-0.39
5	-0.63	-0.81	-0.06	0.63	0.79	1.00	0.73	0.18	-0.47	-0.45
7	-0.32	-0.61	-0.37	0.32	0.52	0.73	1.00	0.42	-0.05	-0.44
10	0.13	0.00	-0.26	-0.32	-0.15	0.18	0.42	1.00	0.18	0.14
20	0.52	0.42	-0.20	-0.60	-0.57	-0.47	-0.05	0.18	1.00	-
30	0.30	0.36	-0.04	-0.31	-0.39	-0.45	-0.44	0.14	-	1.00

Table 2.6: Cross-Correlation matrix of the residuals of the first estimate of the two-factor Cairns model with US data from 1984 to 2008.

Maturities	0.25	0.5	1	2	3	5	7	10
1 month	0.85	0.80	0.86	0.88	0.85	0.78	0.74	0.73
2 months	0.65	0.57	0.74	0.72	0.63	0.48	0.47	0.48
3 months	0.51	0.38	0.66	0.56	0.44	0.26	0.36	0.42
4 months	0.42	0.22	0.59	0.43	0.28	0.08	0.30	0.37
5 months	0.34	0.08	0.55	0.34	0.15	-0.06	0.23	0.35

Table 2.7: Residual Autocorrelations for the first estimate of the Cairns model on US data from 1984 to 2008.

lesser degree. This in turn implies residual autocorrelation. Particularly due to misspecifications in the calibrated initial state $X_{0|0}$, we can expect historical autocorrelation to be overestimated. Generally, low autocorrelations will indicate that the model does not systemically deviate from the true measurements. For our monthly observations, table 2.7 shows the autocorrelations of the residuals to be substantial[26].

2.3.5 Estimation of the cosh model

Implementation of the Kalman filter

In our model definition, we assumed the state process to follow the same Ornstein-Uhlenbeck dynamics under the physical measure as in the Cairns model. For the initial values $X_{0|0}$ and $\Sigma_{0|0}$ we followed the Cairns model and calibrated $X_{0|0}$ to the first two measurements, whereas for the state covariance matrix $\Sigma_{0|0}$ we chose the covariance matrix $Cov[X_{t+1}|\mathcal{F}_t]$. The prediction step of the Cairns model can also be applied for the cosh model since both

[26]Again, due to censored data, we do not examine the autocorrelations of the residual errors for 20 and 30 years of maturity.

frameworks assume the same state vector dynamics. Therefore we have

$$X_{t|t-1} = e^{-\kappa} X_{t-1|t-1} + (1 - e^{-\kappa})\mu,$$

and

$$\Sigma_{t|t-1} = e^{-\kappa} \Sigma_{t-1|t-1} (e^{-\kappa})^T + Q_t(\theta).$$

The covariance matrix $Q_t(\theta)$ is given in (2.24). The measurement equation is defined analogously to the Cairns model by

$$\begin{pmatrix} y^M(t, t+\tau_1) \\ \vdots \\ y^M(t, t+\tau_n) \end{pmatrix} = \underbrace{\begin{pmatrix} g_1(X_t; \theta) \\ \vdots \\ g_n(X_t; \theta) \end{pmatrix}}_{:=g(X_t;\theta)} + \epsilon_t(\theta)$$

with $y^M(t, t+\tau_i)$, $i = 1, \dots, n$ the interest rates as observed in the market and

$$g_i(X_t; \theta) = \alpha - \frac{\cosh\left(\gamma^T E[X_T | X_t]\right)}{T - t}$$

and $\epsilon_t(\theta) \in \mathbb{R}^{n \times n}$ being a multivariate normal error term with covariance matrix

$$Cov\,(\epsilon_t) := H_t(\theta) = diag\,(\nu, \dots, \nu) \in \mathbb{R}^{n \times n}.$$

For the updating step, we require the Kalman gain matrix

$$K_t = \Sigma_{t|t-1} B_{t|t-1}^T F_{t|t-1}^{-1}$$

where $B_{t|t-1}$ is the Jacobi matrix of the non-linear yield function of the cosh model given by

$$B_{t|t-1} = \left.\begin{pmatrix} \frac{\partial}{\partial x_1} g_1(x, \theta) & \dots & \frac{\partial}{\partial x_d} g_1(x; \theta) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} g_n(x; \theta) & \dots & \frac{\partial}{\partial x_d} g_n(x; \theta) \end{pmatrix}\right|_{x = X_{t|t-1}} \in \mathbb{R}^{n \times d}.$$

The derivatives are given by

$$\begin{aligned}
&\frac{\partial}{\partial x_j} g_i(x;\theta) \\
= &-\frac{\partial}{\partial x_j}\log\left(e^{-\alpha(T_i-t)}\frac{\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)}{\cosh\left(\gamma^T x+c\right)}\right) \\
= &\frac{\cosh\left(\gamma^T x+c\right)}{\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)}\left(\frac{\frac{\partial}{\partial x_j}\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)\cosh\left(\gamma^T x+c\right)}{\cosh^2\left(\gamma^T x+c\right)}\right. \\
&\left.-\frac{\frac{\partial}{\partial x_j}\cosh\left(\gamma^T x+c\right)\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)}{\cosh^2\left(\gamma^T x+c\right)}\right) \\
= &\left(\frac{\sinh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)\gamma_j e^{-\kappa_j(T_j-t)}}{\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)}\right. \\
&\left.-\frac{\sinh\left(\gamma^T X_t+c\right)\gamma_j\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)}{\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)\cosh\left(\gamma^T x+c\right)}\right) \\
= &\left(\gamma_j\tanh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_{T_i}|X_t=x]+c\right)e^{-\kappa_j(T_i-t)}-\gamma_j\tanh\left(\gamma^T x+c\right)\right) \qquad (2.25)
\end{aligned}$$

where $i=1,\ldots,n$ and $j=1,\ldots,d$. The prediction error v_t can be calculated from the new information y_t and the measurement implied by the prediction $X_{t|t-1}$, hence

$$\begin{aligned}
v_t &= y_t - g_t(X_{t|t-1},\theta) \\
&= \begin{pmatrix} y(t,T_1)-\frac{1}{T_1-t}\log\left(e^{-\alpha(T_1-t)}\frac{\cosh\left(\gamma^T E[X_{T_1}|X_t=X_{t|t-1}]+c\right)}{\cosh\left(\gamma^T X_{t|t-1}+c\right)}\right) \\ \vdots \\ y(t,T_n)-\frac{1}{T_n-t}\log\left(e^{-\alpha(T_n-t)}\frac{\cosh\left(\gamma^T E[X_{T_n}|X_t=X_{t|t-1}]+c\right)}{\cosh\left(\gamma^T X_{t|t-1}+c\right)}\right) \end{pmatrix}.
\end{aligned}$$

The error covariance matrix $F_{t|t-1}$ is given by

$$F_{t|t-1} = B_{t|t-1}\Sigma_{t|t-1}B_{t|t-1}^T + H_t$$

whereby we assume as in the Cairns model $H_t(\theta) = diag(\nu,\ldots,\nu) \in \mathbb{R}^{n\times n}$, that is measurement errors are mutually uncorrelated. This yields the Kalman gain matrix K_t and hence the updating steps

$$X_{t|t} = X_{t|t-1} + K_t v_t$$

and

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t B_{t|t-1}\Sigma_{t|t-1}.$$

Unlike the Cairns model, the cosh model allows for derivation of all formulae without numerical integration. We can thus expect the cosh model to be significantly faster than the Cairns model computationally. Considering the inverse of the matrix $F_{t|t-1}$, we face the same problem as in the Cairns model in that we can not guarantee $F_{t|t-1}$ to be invertible for all t and all $X_{t|t-1}$. Indeed, singular $F_{t|t-1}$ was again the main reason for the Kalman filter to fail.

Testing the Kalman filter

We follow the same approach as with the Cairns model to test the ability of the Kalman filter to estimate the required parameters. We choose a set of parameters, then simulate times series of yields. As mentioned previously, we recommend using simulated datasets of the same size as the real datasets to be used later. Finally, the Kalman filter is used to estimate the model parameters from these simulated datasets. If the Kalman filter is correctly specified, the estimated model parameters should coincide with the "true", previously specified model parameters.

We derived 20 estimates of the cosh model. The fraction of estimates which had to be excluded because the implicit parameter bounds were reached was considerably higher than in case of the Cairns model. Likewise, we had to exclude more estimates as local minima. As in the Cairns model, we limited the number of iteration steps in estimation, hence the estimates given above are not final. We can conclude that the Cairns model is less dependent on the initial parameter set for maximization of the Loglikelihood value and therefore less estimates have to be excluded for the Cairns model than for the cosh model. Nonetheless, due to computational efficiency the cosh model is still faster than the Cairns model.

To summarize, the Kalman filter provides in most cases very stable estimates of the true model parameters. Notable exceptions are the parameters c, μ and $\tilde{\mu}$. The reason for the considerable instability of these parameters is shown in (2.26). As with the Cairns model, we find that the high- and low-mean reversion factors are exchangeable in the sense that the order of the factors is not fixed. Furthermore, the sign of the factors is not fixed, either. In particular, we find that for the estimated values γ_i^e we have $\{|\gamma_1^e|, |\gamma_2^e|\} = \{|\gamma_1|, |\gamma_2|\}$ and also $|\rho_{12}^e| = |\rho_{12}|$. To summarize, the Kalman filter is able to estimate the model parameters properly.

Results

We start estimation in the same way as in the Cairns model. The first model presented assumes $\tilde{\mu} = 0$, hence the state vector under the reference measure follows the same dynamics as proposed by Cairns. The remaining parameters were chosen from a uniform distribution between upper and lower bounds for each model parameter. Of these initial parameter sets, those with the highest Loglikelihood values are used in further estimation, as described in the algorithm on page 61. As seen in the previous section, the cosh model

α	γ_1	γ_2	c	μ_1	μ_2	κ_1	κ_2	ρ	ν	LogL
0.006	0.022	-0.473	-30.87	1.97	-0.76	0.488	0.021	0.54	0.00163	16424
0.004	0.479	-0.022	-2.70	-0.85	-1.99	0.02	0.488	0.54	0.00163	16424
0.006	0.022	0.473	-19.68	1.97	0.75	0.488	0.021	-0.54	0.00163	16424
0.006	0.022	-0.474	0.03	1.99	0.72	0.488	0.021	0.54	0.00163	16424
0.067	0.024	0.315	-0.45	-0.01	0.86	0.485	0.026	-0.39	0.00160	16435

Table 2.8: QML parameter estimates of the cosh model using the Extended Kalman filter on US data from 1984 to 2008.

tends to end in local minima, break due to singular $F_{t|t-1}$ and reach the parameter bounds more often than the Cairns model. We therefore required a higher number of estimates than for the Cairns model, a subsample of these results is presented in table 2.8.

We find two subgroups of estimates which differ prominently in the parameter α. For each subgroup, we find one state vector component to show high mean reversion and the other state vector component to show low mean reversion as in the Cairns model. Estimates of the mean reversion parameters are highly stable. The correlation coefficient ρ is highly stable as well with its sign depending on γ_1 and γ_2, respectively. These scaling factors γ showed considerable stability within the subgroups.

The parameters μ and c showed varied significantly even within the subgroups. Note that this is likely due to instability of the long-term means μ which transforms into instability of c. Assuming $c = \gamma^T b$ for some $b \in \mathbb{R}^n$ we have for $Y_t := X_t + b$ by the Ito-Doeblin formula

$$\begin{aligned} dY_t^{(i)} &= \left[0 + \kappa_i(\mu_i - X_t^{(i)}) + 0\right] dt + \sum_{j=1}^{d} C_{ij} dZ_t^{(j)} \\ &= \kappa_i \left((\mu_i - b_i) - Y_t^{(i)}\right) dt + \sum_{j=1}^{d} C_{ij} dZ_t^{(j)}. \end{aligned} \quad (2.26)$$

α	γ_1	γ_2	c_i	μ_1	μ_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	κ_1	κ_2	ρ	ν	LogL
0.060	0.300	-0.020	0.3	-9.00	-12.00	-2.00	-10.00	0.025	0.500	0.500	0.001	
0.059	-0.301	-0.021	4.24	19.43	0.13	14.69	1.36	0.025	0.503	-0.604	0.0001	20486
0.060	0.300	-0.020	0.02	-7.04	-9.88	-1.09	-8.78	0.025	0.502	0.515	0.0002	20490
0.060	-0.022	0.300	-3.95	-0.55	12.98	0.30	13.37	0.500	0.025	0.450	0.0001	20503
0.060	0.300	-0.022	0.16	-0.92	-5.51	-1.18	-6.35	0.025	0.500	0.451	0.0001	20503
0.060	0.218	-0.015	-4.55	-5.72	17.89	-10.04	18.63	0.034	0.502	0.539	0.0001	20507

Table 2.9: Estimates of the Kalman filter for data simulated by the cosh model using the “true” parameters in the first line.

0.25	0.5	1	2	3	5	7	10	20	30
16.7	6.8	8.9	16.3	13.8	8.1	6.6	10.4	12.0	13.9
16.7	6.8	8.9	16.3	13.8	8.1	6.6	10.4	12.0	13.8
16.7	6.8	8.9	16.3	13.8	8.1	6.6	10.4	12.0	13.9
16.7	6.8	8.9	16.3	13.8	8.1	6.6	10.4	12.0	13.9
16.4	6.3	9.0	16.0	13.5	7.7	6.4	10.0	12.8	14.2

Table 2.10: MAEs of the cosh model using the Extended Kalman filter on US data from 1984 to 2008.

Obviously, as the linear equation system $\gamma^T b = c$ does not necessarily provide a unique solution, infinitely many "true" estimates of the model parameters with varying μ and c are possible. The model presented so far is hence overparameterized. A first idea would set $c = 0$. However, assuming $\gamma^T b = 0$ may still allow for infinitely many b, hence instability of μ prevails. This is a direct consequence of X_t entering the bond pricing formula through affine transformation. We found that flexible c improved stability of the Kalman filter, particularly in later expansions of the model. We therefore recommend to keep the overparameterized model including c.

Examining MAEs in table 2.10, the high- and low-α estimates differ in their ability to fit the long and short ends of the term structure, respectively. High-α estimates imply a better historical fit of the short end, yet decreased fit of the long end. In section 2.3.9, we will see that this is likely a result of censored data on the very long end of the yield curve.

Considering the interpretation of the state vector components, as can be seen in figure 2.5 for the first parameter set, we again find one state vector component to coincide with a long-term rate and the other to capture slope dynamics. As the state vector describing the level is highly correlated to the 10-year yield and mean reversion is weak, we must expect the same problems in estimating μ^L as encountered in Cairns. As we assume that the state vector components follow Ornstein-Uhlenbeck dynamics and our dataset covers 25 years, we would assume that all state vector components crossed their respective long-term mean at least once from 1984 to 2008. We therefore calculate for each state process component the distance between the empirical mean of the filtered path and the estimated mean μ_i. To account for differences in the scaling factor γ_i, we standardize this difference by the range of the filtered state process. We measure this range as the distance between the upper-0.1-quantile $Q^{0.9}((\hat{X}_t^L)_{t=1,\dots,T})$ and the lower-0.1-quantile $Q^{0.1}((\hat{X}_t^L)_{t=1,\dots,T})$, whereby $\hat{X}^L$ denotes the level component of the filtered state process. A ratio higher than 1 implies that the distance between the estimated long-term mean and the empirical long-term mean is higher than the range of the filtered path. In particular, this usually implies that $\mu^L \notin conv\left\{\hat{X}_t^L, t = 1, \dots, T\right\}$. $\mu^L \notin conv\left\{\hat{X}_t^L, t = 1, \dots, T\right\}$ therefore is a clear

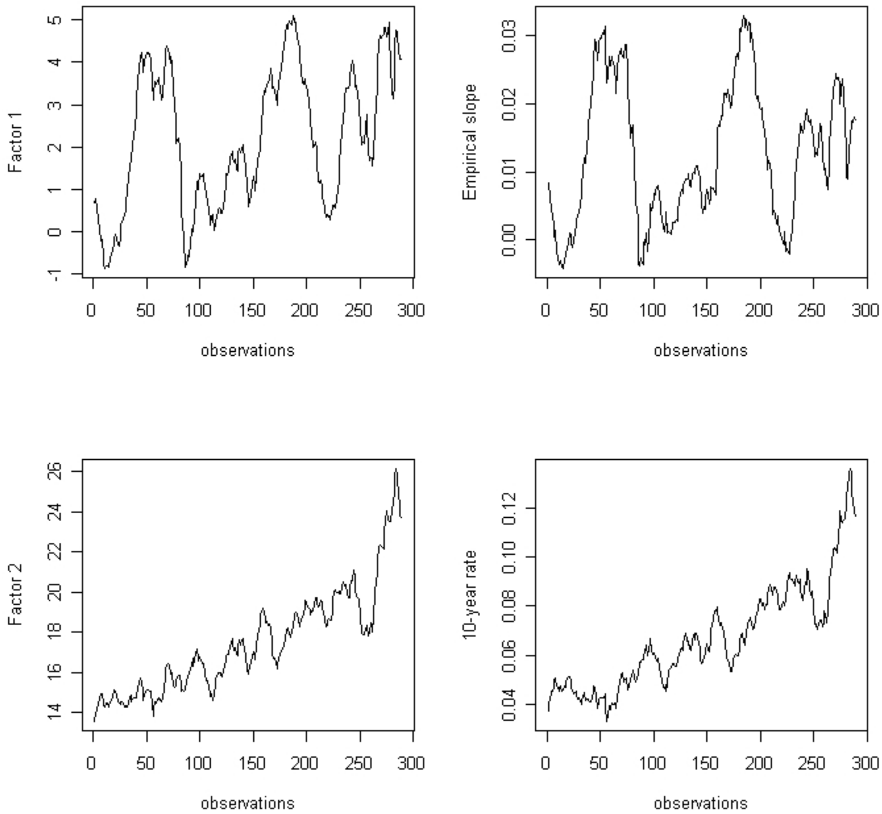

Figure 2.5: Filtered state vector components (left) and empirical proxies of the first two principal components of term structure dynamics.

Slope	Level
3.08	0.06
3.08	0.06
2.90	0.05
3.08	0.06
1.02	0.45

Table 2.11: The ratios of the (absolute) difference between the estimated and empirical long-term mean μ and the overall range of the respective filtered state vector component, for both the high- and low-mean reversion components.

indication that the Kalman filter failed to estimate μ^L properly. As a measure of this failure we calculate

$$\frac{\left|\frac{1}{n}\sum_{t=1}^{n}\hat{X}_t^L - \mu_i\right|}{Q^{0.9}((\hat{X}_t^L)_{t=1,\dots,T}) - Q^{0.1}((\hat{X}_t^L)_{t=1,\dots,n})}. \tag{2.27}$$

In table 2.11, we find that estimates of μ^L are generally poor. The last estimate of μ^L seems to be superior to the previous estimates, although the ratio of μ^S is worse. However, if we examine the filtered state paths, we find again that the estimated value μ^L still underestimates the long-term mean.

As in the Cairns model, we examined auto- and cross-correlation of the time series of residuals. As we implemented a 2-factor model, the cross-correlation matrix of the 2-factor cosh model is expected to show the same pattern of positive and negative correlations implied by curvature mismatch we presented in figure 2.5. Indeed in tables 2.12 and 2.13 we find the predicted pattern in the cross-correlation matrix of errors[27] for both high- and low-α estimates. Interestingly, cross-correlation matrices are in both cases pretty much the same for maturities lower than 7 years, yet differ in higher maturities. This hints to a connection between α and long-end curvature misfit we will examine further in section 2.3.9.

Considering autocorrelation, we expect the same pattern we found in the Cairns model. We already found proof for the failure of the cosh model to cover curvature dynamics in cross-correlation matrices. For our monthly observations, tables 2.14 and 2.15 show the autocorrelations of the residuals to be substantial. Differences between the high- and low-α estimates can only be found in the higher maturities, indicating again a special role of α for the long end of the yield curve. The general pattern of autocorrelations is the same as in the Cairns model, indicating that both models fail to cover the dynamics of a transitionary

[27]Note that due to censored data, correlation between the 20- and 30-year rates are not reliable and therefore omitted.

maturities	0.25	0.5	1	2	3	5	7	10	20	30
0.25	1.00	0.61	-0.73	-0.93	-0.91	-0.58	-0.07	0.45	0.62	0.20
0.5	0.61	1.00	-0.12	-0.67	-0.76	-0.72	-0.43	0.19	0.36	0.23
1	-0.73	-0.12	1.00	0.65	0.56	0.10	-0.35	-0.44	-0.41	-0.11
2	-0.93	-0.67	0.65	1.00	0.96	0.56	0.03	-0.59	-0.68	-0.17
3	-0.91	-0.76	0.56	0.96	1.00	0.72	0.23	-0.45	-0.66	-0.28
5	-0.58	-0.72	0.10	0.56	0.72	1.00	0.66	0.08	-0.43	-0.39
7	-0.07	-0.43	-0.35	0.03	0.23	0.66	1.00	0.52	0.17	-0.45
10	0.45	0.19	-0.44	-0.59	-0.45	0.08	0.52	1.00	0.47	0.03
20	0.62	0.36	-0.41	-0.68	-0.66	-0.43	0.17	0.47	1.00	-
3ß	0.20	0.23	-0.11	-0.17	-0.28	-0.39	-0.45	0.03	-	1.00

Table 2.12: Cross-Correlation matrix of the residuals of the first estimate of the cosh model of the low-α subgroup.

maturities	0.25	0.5	1	2	3	5	7	10	20	30
0.25	1.00	0.61	-0.74	-0.93	-0.91	-0.58	-0.14	0.36	0.49	0.26
0.5	0.61	1.00	-0.14	-0.69	-0.78	-0.76	-0.53	0.02	0.13	0.44
1	-0.74	-0.14	1.00	0.63	0.53	0.06	-0.34	-0.47	-0.46	0.02
2	-0.93	-0.69	0.63	1.00	0.95	0.56	0.11	-0.49	-0.49	-0.30
3	-0.91	-0.78	0.53	0.95	1.00	0.73	0.32	-0.31	-0.48	-0.40
5	-0.58	-0.76	0.06	0.56	0.73	1.00	0.72	0.21	-0.30	-0.47
7	-0.14	-0.53	-0.34	0.11	0.32	0.72	1.00	0.59	0.19	-0.52
10	0.36	0.02	-0.47	-0.49	-0.31	0.21	0.59	1.00	0.38	0.02
20	0.49	0.13	-0.46	-0.49	-0.48	-0.30	0.19	0.38	1.00	-
30	0.26	0.44	0.02	-0.30	-0.40	-0.47	-0.52	0.02	-	1.00

Table 2.13: Cross-Correlation matrix of the residuals of the first estimate of the cosh model of the high-α subgroup.

Maturities	0.25	0.5	1	2	3	5	7	10
1-month	0.87	0.80	0.91	0.92	0.88	0.77	0.08	0.79
2-month	0.72	0.60	0.81	0.80	0.72	0.48	0.59	0.58
3-month	0.62	0.43	0.74	0.70	0.60	0.29	0.51	0.49
4-month	0.55	0.29	0.67	0.61	0.49	0.14	0.46	0.43
5-month	0.50	0.18	0.63	0.54	0.40	0.02	0.40	0.38

Table 2.14: Residual autocorrelations for the first estimate of the high-α subgroup of the cosh model.

Maturities	0.25	0.5	1	2	3	5	7	10
1-month	0.87	0.82	0.91	0.92	0.88	0.79	0.77	0.83
2-month	0.72	0.62	0.82	0.80	0.73	0.53	0.53	0.66
3-month	0.62	0.46	0.75	0.70	0.61	0.35	0.44	0.59
4-month	0.55	0.32	0.68	0.61	0.50	0.21	0.39	0.54
5-month	0.50	0.22	0.64	0.55	0.42	0.11	0.33	0.50

Table 2.15: Residual autocorrelations for the first estimate of the low-α subgroup of the cosh model.

factor in residuals, quite likely curvature. Note however that error autocorrelations in the Cairns model were generally smaller than error autocorrelations in the cosh model.

We can conclude that the Kalman filter provides reasonable estimates. However, the Kalman filter fails to produce reasonable estimates of the long-term mean of the level factor under the physical measure μ. Overall, historical fit measured in MAEs is 1 to 1.5 basis points worse than in the Cairns model, yet estimation and implementation is vastly more efficient. Implied state vector dynamics show that the cosh model resembles the dynamics of the Cairns model. Aside from the zero lower bound implemented in the Cairns approach, the cosh model seems to be a viable alternative, particularly if computational efficiency is required.

Cairns essentially assumed that the state vector under the reference measure is mean reverting to a zero long-term mean. Due to computational efficiency of the cosh model, we can generalize this assumption by allowing for $\tilde{\mu} \neq 0$, which increases the number of model parameters by d. As this introduces additional model parameters to be estimated, a minimum requirement is an improved historical fit.

The parameter estimates in table 2.18 show the well-known pattern of high- and low-α estimates. Mean reversion parameters κ_i and scaling parameters γ_i are highly stable within each subgroup, as are estimates of the correlation parameter ρ. Parameters c and μ are highly unstable, due to (2.26). Obviously, the same problem should hold for $\tilde{\mu}$ and c under

0.25y	0.5y	1y	2y	3y	5y	7y	10y	20y	30y
17.3	6.5	8.0	15.8	13.3	7.5	5.6	9.7	11.6	13.8
17.3	6.5	8.0	15.8	13.4	7.5	5.6	9.7	11.7	13.7
16.7	6.3	8.2	15.8	13.1	7.1	5.4	9.4	12.1	14.3
16.7	6.3	8.2	15.8	13.1	7.1	5.4	9.4	12.1	14.3

Table 2.16: MAEs in basis points for the US term structure 1984 to 2008 estimated for the cosh model with separated long-term means μ and $\tilde{\mu}$ under the respective measures.

High	Low
0.05	2.91
0.06	3.08
0.00	0.89
0.00	0.90

Table 2.17: Ratios of deviation of the empirical mean of the filtered state process component from the estimated long-term mean of the same process for the US term structure 1984 to 2008 estimated for the cosh model with separated long-term means μ and $\tilde{\mu}$ under the respective measures.

the reference measure as well.

Table 2.16 shows MAEs of the generalized model. Historical fit slightly improved by an average of 0.4 basis points in comparison to the restricted case $\tilde{\mu} = 0$. Clearly, such a small improvement in historical fit does not justify the introduction of d additional model parameters.

According to table 2.17, deviation-to-range ratios for estimated long-term means under the physical measure are comparable to the previous implementations or slightly better. In particular, the ratios for the high-α estimates seem promising. However, a ratio smaller than 1 does not necessarily imply that the long-term mean μ^L is indeed reached. Indeed this is the case only for the first high-α estimate, but even in this case the long-term mean is close to the edge of $conv(\hat{X}_t = 1, \ldots, T)$. Overall, the slight improvement in these ratios does not justify the introduction of d additional model parameters either.
Although $\tilde{\mu}$ is a long-term mean of the state vector as well, we do not examine the deviation-to-range ratios of $\tilde{\mu}$. To derive a comparable ratio for $\tilde{\mu}$ we would require filtered state vectors under the measure $\tilde{P}$, which is unfeasible as term structure dynamics under the reference measure are not observed.

Finally, we examine auto- and cross-correlation of the time series of residuals of the cosh model with generalized $\tilde{\mu}$. We find the predicted pattern in the cross-correlation matrix of errors for both high- and low-α estimates. Changes in cross-correlation to the model with $\tilde{\mu} = 0$ were generally less than 0.05 with both high- and low-α subgroups. In the same way, we encounter only minor improvements in residual autocorrelations of maximal 0.05 in comparison to the model with $\tilde{\mu} = 0$. The tables are thus omitted.

To summarize, introducing $\tilde{\mu} \neq 0$ slightly improved MAEs. We can expect $\tilde{\mu}$ to be rather stable since it is a pure cross-sectional parameter. Nevertheless, cross-correlation and autocorrelation resemble those of the simpler case $\tilde{\mu} = 0$. Obviously, a third stochastic factor would be superior to flexible $\tilde{\mu}$. Due to computational efficiency of the cosh model, we were able to derive the estimates rather quick, yet overall improvement of model fit and model dynamics does not justify the introduction of d additional model parameters.

2.3.6 Result Summary

We found that the Kalman filter was able to estimate both the Cairns and the cosh model rather well. In particular mean reversion parameters κ, scaling parameters γ and σ and correlation parameters ρ were estimated properly. In the cosh model, (2.26) implies a problem due to the affine transformation of the state vector, which makes unique identification of c impossible. Nevertheless, we found that the parameter c stabilizes the Kalman filter and hence we recommend using c.

In both models, one state vector component coincided with the slope, whereas the other component coincided with the level factor, measured as the 10-year rate[28]. Both models failed in estimation of the long-term mean μ^L of the level factor. A possible explanation might be low mean reversion of the state vector. We require additional examinations considering the factor μ^L, presented in section 2.3.8.

For the cosh model we found two distinct subgroups of model estimates according to the parameter α. Comparing MAEs, cross-correlation matrices and autocorrelation, we found that the parameter α has a specific long-end effect, which we already showed in 2.2.1. Section 2.3.9 will examine further this role of α and the question how more stable estimates can be derived.

Both models described historical term structures rather well, with average MAEs of 7 basis points for the Cairns model and 10 basis points for the cosh model. On the other side we found significant autocorrelation in the time series of residuals of all maturities. One major contribution to this autocorrelation could be the lack of a curvature factor. Cross-correlation matrices and MAEs hinted to a systematic failure to catch curvature dynamics. We expect that MAEs of both models could be reduced further by inclusion of a third state vector component, which will likely coincide with curvature.

The cosh model is more efficient computationally than the Cairns model. As both models showed many similarities, in particular considering the state vector behavior, we can conclude that the cosh model is a viable approximation to the Cairns model if computational speed is crucial and the zero lower bound is of minor importance.

[28] We will see in section 2.3.9, that the level factor in both models coincides with the observed yield with the highest maturity.

α	γ_1	γ_2	c	μ_1	μ_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	κ_1	κ_2	ρ	ν	LogL
0.004	0.022	-0.479	12.67	44.40	-18.18	46.42	-18.05	0.489	0.020	0.542	0.0016	16424
0.009	0.022	0.464	-55.26	12.80	-49.02	10.82	-49.52	0.484	0.021	-0.537	0.0016	16424
0.006	0.473	0.022	0.15	-17.83	33.97	-18.58	32.00	0.021	0.488	-0.543	0.0016	16424
0.067	0.024	0.316	2.05	54.88	-10.69	56.99	-12.28	0.475	0.027	-0.438	0.0016	16448
0.067	0.315	0.024	1.92	-5.99	-0.14	-7.62	2.01	0.027	0.476	-0.433	0.0016	16448

Table 2.18: QML Parameter estimates of the cosh model with $\tilde{\mu} \neq 0$ using the Extended Kalman filter on US data from 1984 to 2008.

2.3.7 Three factor models

In the previous section, we demonstrated how to estimate both the Cairns and the cosh model and presented estimates of two-factor models. Generally, as discussed in 2.1, insurance applications require at least two stochastic factors to govern the level of the yield curve and its slope. This should be sufficient to provide stochastic discount factors as well as realistic cross-sectional behavior. However, in both models we saw that error autocorrelations were substantial, indicating that the two-factor model misses a systematical factor driving the term structure. According to Litterman and Scheinkman, the third principal component driving the yield curve is curvature. MAEs as well as error cross-correlation matrices of both models indeed showed a pattern indicating curvature mismatch. In this section, we will present estimation results for the respective three-factor models. In particular, we are interested in improving curvature fit. Of general interest however is the question whether the Kalman filter is able to identify a third distinct state vector component at all, particularly in case of the cosh model where we already found problems in identifying the parameters c and μ.

Cairns three-factor model

We estimated the Cairns three-factor model using the US dataset beginning in 1984. We employed both the original Kalman filter augmented by a third state factor which also estimated μ^L and the alternative approach discussed in 2.3.8 to derive μ^L exogenously, given in the third row of table 2.22. The main change between a two-factor model and a three-factor models is the increase in the number of correlation parameters.

In table 2.22, we get a low-mean reversion factor with $\kappa \approx 0.02$, which is the level factor, a medium-mean reversion factor with $\kappa_i \approx 0.7$ which is highly correlated to curvature and a high-mean reversion factor with $\kappa_i \approx 1$ which is also highly correlated to curvature. The sum of the later two state factors, though, is closely related to the slope. This indicates that for higher-dimensional models, the interpretational simplicity considering the state vector components vanishes, although the state vector as a whole clearly contains information about the three main principal components of the yield curve. It might be possible to restrict model parameters for higher-dimensional models to preserve the coincidence of each state vector component with a a single principal component, for example by limiting mean reversion and correlation accordingly. Scaling factors σ_i are highly stable as well. Correlation estimates, however, differ significantly with the exception of the high correlation between the two curvature factors. It may well be that high correlation between the two curvature factors leads to spurious correlation between the remaining factors. Another possible explanation might be a more general problem of estimating correlation matrices for higher-dimensional Ornstein-Uhlenbeck processes.

As expected, the original Kalman filter underestimates μ^L systematically in the sense that

0.25	0.5	1	2	3	5	7	10	20	30
4.73	4.83	6.66	3.57	2.96	4.28	5.68	6.43	6.16	7.59
4.73	4.83	6.66	3.57	2.96	4.28	5.68	6.43	6.16	7.59
4.91	4.71	6.56	3.72	2.82	4.34	5.39	6.94	6.21	7.43

Table 2.19: Mean Absolute Errors for the Three-factor Cairns model.

maturities	0.25	0.5	1	2	3	5	7	10
1-month	0.70	0.76	0.87	0.76	0.68	0.81	0.80	0.74
2-months	0.46	0.61	0.75	0.56	0.47	0.64	0.63	0.53
3-months	0.35	0.49	0.66	0.44	0.40	0.53	0.55	0.48
4-months	0.29	0.43	0.58	0.36	0.31	0.41	0.51	0.43
5-months	0.31	0.40	0.54	0.34	0.23	0.34	0.46	0.40

Table 2.20: Error autocorrelations for the first estimate of the three-factor Cairns model.

$g_i(\mu, \theta)$ implies a Japan scenario. Contrary to that, the alternative approach provides an estimate of μ^L which shows extremely low deviation-to-range ratios and guarantees a reasonable implied curve $g(\mu)$ and $\mu^L \in conv\left\{\hat{X}_{t=1,\ldots,nT}\right\}$.

Examining MAEs in table 2.19, we clearly see a substantial improvement in historical fit as MAEs were reduced to an average of 5.3 basis points. This is remarkable given that the yield curve data used for estimation is the result of an interpolation approach to observable coupon bond prices and hence holds a fitting error of a few basis points by itself. Improvement in historical fit will likely result in overfitting to yield data, which is itself merely an approximation to real bond data.

Table 2.20 presents error autocorrelations for the three-factor model. We find again some evidence for a further systematic factor, the model does not catch all systematic variation in the term structure. Autocorrelations have, however, substantially reduced for lags of 1 and 2 months. For higher lags autocorrelations seem to converge around 0.35. One reason might be that the Kalman filter tends to produce autocorrelated errors, as discussed previously. If this is the case, remaining autocorrelation of 0.35 is due to the estimation procedure and can not be reduced any further. Another reason would be the fourth principal component of Litterman and Scheinkman with extremely high mean reversion implying autocorrelation for lags of 4 or 5 months.

The cosh three-factor model

The three-factor cosh model was implemented with the recommended alternative approach to estimate μ^L from section 2.3.8. This was done in order to reduce the number of model parameters to be estimated by the Kalman filter. On the other side, we assumed the general approach with flexible $\tilde{\mu}$. Results are given in table 2.23.

0.25	0.5	1	2	3	5	7	10	20	30
4.84	4.29	6.67	4.24	2.62	4.54	5.55	7.16	5.01	6.37
4.91	4.28	6.81	3.88	2.52	4.44	5.53	7.38	5.31	7.76
4.82	4.32	6.58	4.37	2.60	4.57	5.33	7.80	5.24	8.00
5.00	4.31	6.49	4.78	2.73	4.63	5.07	8.91	5.22	8.41

Table 2.21: Mean absolute errors for the three-factor cosh model.

We see that a low-mean reversion parameter exists with $\kappa \approx 0.02$ and $\gamma \approx 0.4$, a medium-mean reversion parameters with $\kappa \approx 0.53$ and $\gamma \approx 0.04$ and a high-mean reversion parameter with $\kappa \approx 1.2$ and $\gamma \approx 0.02$. Therefore mean reversion of the state factor of the cosh model resembles mean reversion as found for the Cairns model in table 2.22. The asymptotic long rate α again shows the two subgroups already encountered in the two-factor case. As can be found in section 2.3.9, this should vanish in case 20- and 30-year rates are omitted. Furthermore, mean reversion of the level factor should increase in this case. Note also that κ and γ show no dependence on the respective estimate of α. As we will see later, the third factor is indeed a curvature factor.

In the two-factor model, we already found a possible link between α and curvature in the differences between cross-correlation and autocorrelation of high- and low-α estimates. With a third factor describing curvature dynamics, this pattern vanishes, indicating that the link between α and curvature is a two-factor problem only. As on the other side high- and low-α subgroups persist, we can conclude that curvature mismatch can not explain these subgroups. As we will see in section 2.3.9, censored data is the most likely reason for the two subgroups encountered.
Correlation estimates are very stable. Depending on the sign of γ_i we have that the medium and high mean-reversion components are highly correlated, the high and low mean-reversion components are effectively independent, and for the medium and low mean-reversion components the absolute value of correlation is 0.5.
As in the two-factor case, parameters μ, $\tilde{\mu}$ and c vary substantially. This was expected as the fundamental problem of (2.26) remains, which shows that X_t enters through an affine transformation which and therefore not allow to identify c and μ uniquely in the cosh model.

Examining MAEs in table 2.21 shows that historical fit improved considerably, again to an average of 5.3 basis points. We also see that the higher α, the better the historical fit of the high end of the yield curve. As discussed previously, it is questionable whether a further improvement of historical fit is worth a fourth state vector component.

κ_1	κ_2	κ_3	α	σ_1	σ_2	σ_3	ρ_{12}	ρ_{13}	ρ_{23}	μ_1	μ_2	μ_3	ν	LogL
0.021	0.674	1.054	0.001	0.745	1.149	1.208	0.460	-0.547	-0.980	9.41	0.48	-0.61	0.00089	17667
0.021	0.694	0.982	0.001	0.746	1.456	1.453	-0.070	-0.007	-0.987	8.56	-0.01	-0.26	0.00089	17647
0.020	0.704	0.993	0.001	0.736	1.426	1.473	0.369	-0.452	-0.985	20.04	0.37	-0.45	0.00089	17651

Table 2.22: Estimation results of the three-factor Cairns model.

α	γ_1	γ_2	γ_3	c	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$	κ_1	κ_2	κ_3	ρ_{12}	ρ_{13}	ρ_{23}	ν	LogL
0.0502	-0.455	0.023	-0.045	1.25	5.95	20.01	6.62	0.025	1.179	0.527	0.447	0.046	0.849	0.00091	17642
0.0130	0.041	0.019	0.516	-2.88	25.07	-0.23	10.58	0.533	1.281	0.020	-0.842	0.005	-0.405	0.00091	17649
0.0013	0.541	-0.042	-0.019	0.10	8.42	1.14	-5.20	0.019	0.524	1.282	-0.003	0.392	-0.850	0.00091	17653
0.0018	-0.538	-0.040	-0.019	-1.04	-9.39	-15.93	-6.44	0.019	0.524	1.286	0.124	-0.490	-0.847	0.00091	17655

Table 2.23: Estimation results of the three-factor cosh model.

maturities	0.25	0.5	1	2	3	5	7	10
1-month	0.75	0.73	0.87	0.80	0.67	0.82	0.80	0.77
2-months	0.54	0.59	0.74	0.62	0.45	0.65	0.61	0.56
3-months	0.44	0.45	0.64	0.53	0.41	0.55	0.53	0.51
4-months	0.41	0.36	0.55	0.42	0.25	0.44	0.49	0.46
5-months	0.41	0.32	0.51	0.37	0.17	0.36	0.42	0.43

Table 2.24: Error autocorrelations of the three-factor cosh model with high-α.

Considering cross-correlation, the curvature pattern vanished. There seems, however, to exist a new pattern we already encountered with the three-factor Cairns model. We believe this to be connected to the fourth principal component of Litterman and Scheinkman. Considering autocorrelation, the results were very stable and mirror our previous results for the Cairns model. As the results were very stable independent of α we only show autocorrelations for high α in table 2.24. Autocorrelation is still significant, yet diminished considerably for shorter horizons in comparison to the two-factor approach. On the other side, note that the decrease of autocorrelation is weaker. As discussed previously, the Kalman filter tending to autocorrelated residuals by definition or failure to catch an additional state factor with very high mean reversion may explain the remaining autocorrelation for higher lags.

Examining the filtered state processes, we find that one vector component is highly correlated to the level and shows clearly the pattern already known from the two-factor case. One state vector component is highly correlated to the slope, the third vector is highly correlated to a curvature proxy. Note, however, that the slope and the curvature components are highly correlated among themselves, although the slope and curvature proxies based on empirical data are effectively independent over the dataset used. This may indicate that it becomes increasingly complicated to differ between the state vector components in higher-dimensional cosh models. In fact, a short examination of the four factor model provided four highly correlated components which *all* were highly correlated to the slope, yet principal component analysis of the filtered state processes showed that they still contained level, slope and curvature information. This is similar to the finding of two highly correlated curvature factors in the three-factor Cairns model which nevertheless contained both slope and curvature information. We expect that using models with $d \geq 3$ may still provide state vectors which contain the first d principal components of Litterman and Scheinkman, yet the simple interpretation for the state vector components vanishes. Furthermore, we must examine closely whether the estimated state vector provides reasonable simulations in case of several highly correlated Ornstein-Uhlenbeck processes describing term structure dynamics.

Summary

We estimated both three-factor Cairns and three-factor cosh models. In both cases, we found that the additional factor describes stochastic curvature as a high mean reversion process. This decreased MAEs of both models to an average of clearly less than 6 basis points. The patterns in MAEs and cross-correlations which hinted to a curvature mismatch vanished as well. In both cases, the parameter estimates were very stable, particularly those describing model dynamics, indicating that a four-factor model could be estimated as well. Interestingly, both models provided state vector components with high, medium and low mean reversion. The low mean reversion state vector component, which described the level, had $\kappa \approx 0.02$ in both models, the intermediate mean reversion component had $\kappa \approx 0.5$ in the Cairns and 0.7 in the cosh model. The high mean-reversion component had $\kappa \approx 1$ in the Cairns and 1.2 in the cosh model.
Error autocorrelations decreased, yet still indicated that the models fail to include all systematic factors driving the term structure. Whereas all this implies that historical fit could be increased by introducing a fourth state vector component, note that the dataset used for estimation is the result of an interpolation algorithm based on coupon bond prices to derive a continuous yield curve. As this interpolation algorithm commands a measurement error of a few basis points by itself, it is questionable whether historical fit to erroneous data should be improved further. A four-factor model should be fitted to bond data directly rather than interpolated yields.

Considering estimation speed, we found again the cosh model to be vastly superior computationally. Whereas estimation of a three-factor Cairns model is a question of several days if not weeks, estimates of the cosh model can be derived in a couple of hours. Similarity of state vector behavior again shows that the cosh model is closely related to the Cairns model and may be a viable proxy to the Cairns model if the zero lower bound is of minor interest.

2.3.8 The parameter μ

In section 2.3.5, we found that both in the Cairns and the cosh model we had problems to identify the long-term mean μ^L of the level factor X_t^L. These problems in estimating stem from different sources - the role of μ in the Kalman filter estimation approach, the nature of the factor processes, and underlying data.

If mean reversion is low, the impact of the long-term mean on the one-step-ahead distributions $X_{t+1}|X_t$ used in the transition equation of the Kalman Filter

$$X_{t|t-1}^{(i)} = e^{-\kappa_i} X_{t-1|t-1}^{(i)} + (1 - e^{-\kappa_i})\mu_i \tag{2.28}$$

for $i = 1, \ldots, n$ is small, since $1 - \exp(-\kappa_i \Delta) \approx 0$ for small κ_i. Indeed, the long-end factor μ^L in all models considered was characterized by very low mean reversion, and rightly so

given the dynamics of the empirical proxy in figure 2.5.
One reason for low mean reversion of the long end factor is trend behavior. The 10-year rate shows an increasing trend from the fifties up to the monetary experiment 1979-1982 and a decreasing trend ever since, see figure 2.6. Throughout the whole dataset used for estimation, the long end yield therefore shows a falling trend. In an Ornstein-Uhlenbeck process, one way to fit a trend in a low mean reversion factor is to underestimate the long-term mean. The fact that estimates of μ^L generally imply a Japan scenario in case of the Cairns model and very low or negative interest rates in case of the cosh model supports the assumption that the Kalman filter underestimates μ^L in order to fit the falling trend. Higher frequency of the data does not change the low mean reversion of historical long-end data. We can only hope to improve estimation of low mean reversion processes by increasing the length of the time series. The McCulloch, Kwon dataset starts in December 1946, thus increasing the available data by 38 years. Note however that the early years of this dataset relied heavily on callable bonds and that in the meantime many fundamental changes in the Treasury markets took place[29]. Nevertheless, even if we used the full post-war dataset despite quality considerations, the above mentioned trend behavior of the long end of the term structure implies that the mean reversion process X^L was close to its constant long-term mean μ^L only during two short time periods.

Given these problems, it becomes clear that the Kalman filter has problems in estimating μ^L properly. Whereas mean reversion of the 10-year rate is still a reasonable assumption to guarantee both variability and boundedness, the mean reversion we find in historical data is too low for estimation of the long-term mean μ^L. Nevertheless, based upon figure 2.6 we could easily specify the long-term mean of the 10-year rate exogenously at around 5%. Several questions arise:

1. What impact does an exogenous specification of μ^L have on the remaining factors?
2. How can we reasonably specify the long-term mean μ^{10} of the 10-year rate?
3. How can we derive a long-term mean μ^L from a specified long-term mean μ^{10} of the 10-year rate?

The following sections try to solve these problems.

[29]Events to consider for example were the purchase program of the Fed to guarantee a maximal interest rate on 10 year treasury bonds thus fixing the long end of the term structure until the 1951 Treasury-Federal Reserve Accord, "operation twist" where treasury department and Federal Reserve tried to actively change the slope of the term structure, changes in transaction costs and taxation with implied impact on demand, ineligibility for commercial bank purchase (an important factor prior to the 1951 Treasury-Federal Reserve Accord), the ability to be surrendered at par in payment of estate taxes, the end of the gold standard, the oil crises and the infamous monetary experiment.

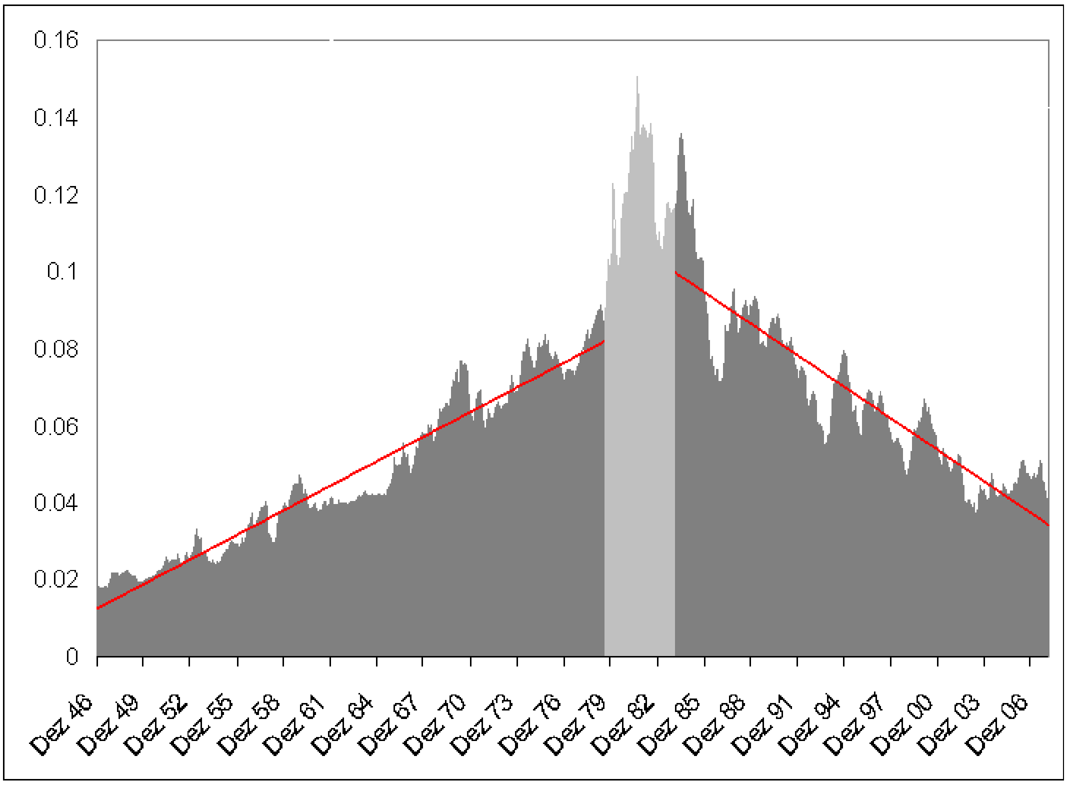

Figure 2.6: The historical US 10-year yield with linear trends estimated for the subsamples before and after the monetary experiment, here set as July 1979 to December 1983. The later, light-gray subsample is used for estimation.

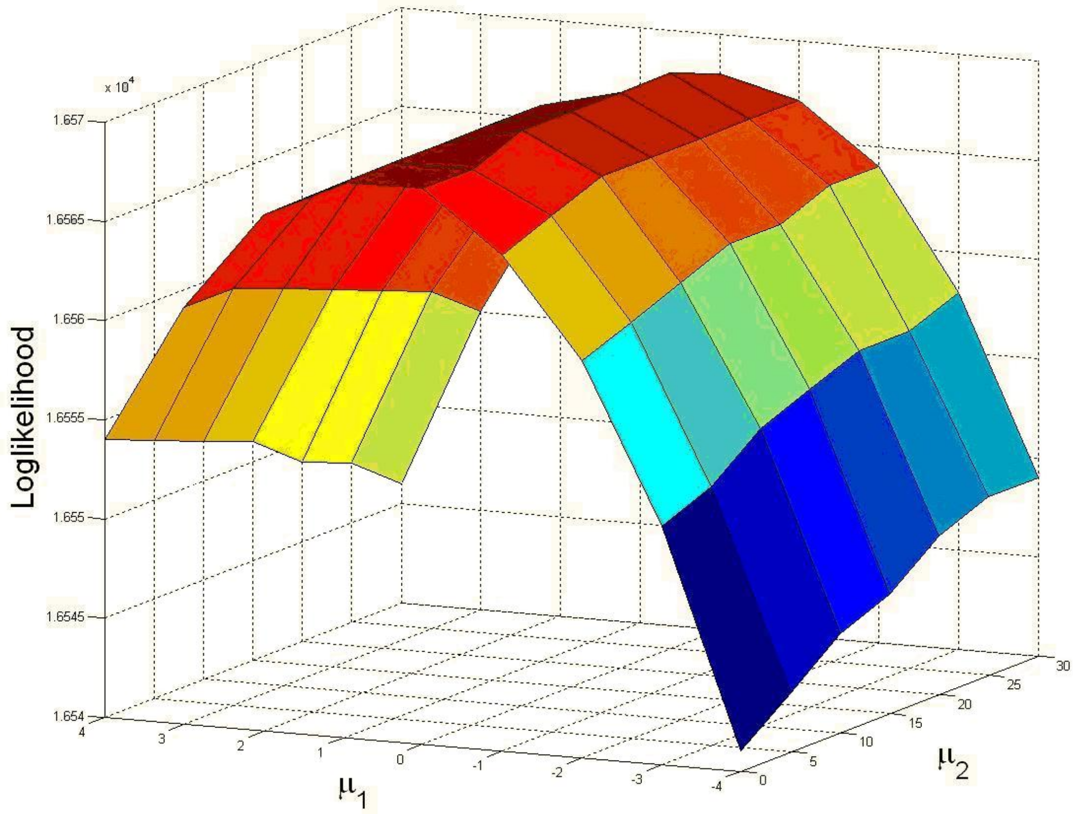

Figure 2.7: Loglikelihood values according to the Kalman Filter for the first Parameter estimate of the Cairns model with varying values for μ.

Loglikelihood sensitivities

A first analysis must consider the sensitivity of the Loglikelihood values on changes in μ^L. If sensitivities are low, we can specify μ^L exogenously with only minor changes in Loglikelihood values, hence reestimation of the other model parameters is not necessarily required. These sensitivities can also be used as a measure for the ability of the Kalman filter to estimate μ. We thus calculated Loglikelihood values for a set of possible μ estimates for both examples of the Cairns model and the cosh model. Figure 2.7 shows the Loglikelihoods dependent on μ for the Cairns model, figures 2.8 and 2.9 show Loglikelihood sensitivities for the first low-α and the first high-α estimate of the cosh model, respectively.

For the Cairns model, sensitivities of the Loglikelihood function on changes in μ^L are very low, resulting in the problems encountered to identify the true value of μ^L. In case of the cosh model, for the low-α estimate we clearly see that the Loglikelihood value is sensitive with respect to μ^S, yet hardly reacts to changes in μ^L, repeating our difficulties with the Cairns model. In case of the high α estimates, we find that Loglikelihood values are non-continuous and sensitivities considering both μ^S and μ^L are low. Consequently, we reject the high-α estimates altogether. In case of the low α-estimates we need to reconsider the estimate of μ^L in the same way as in the Cairns model.

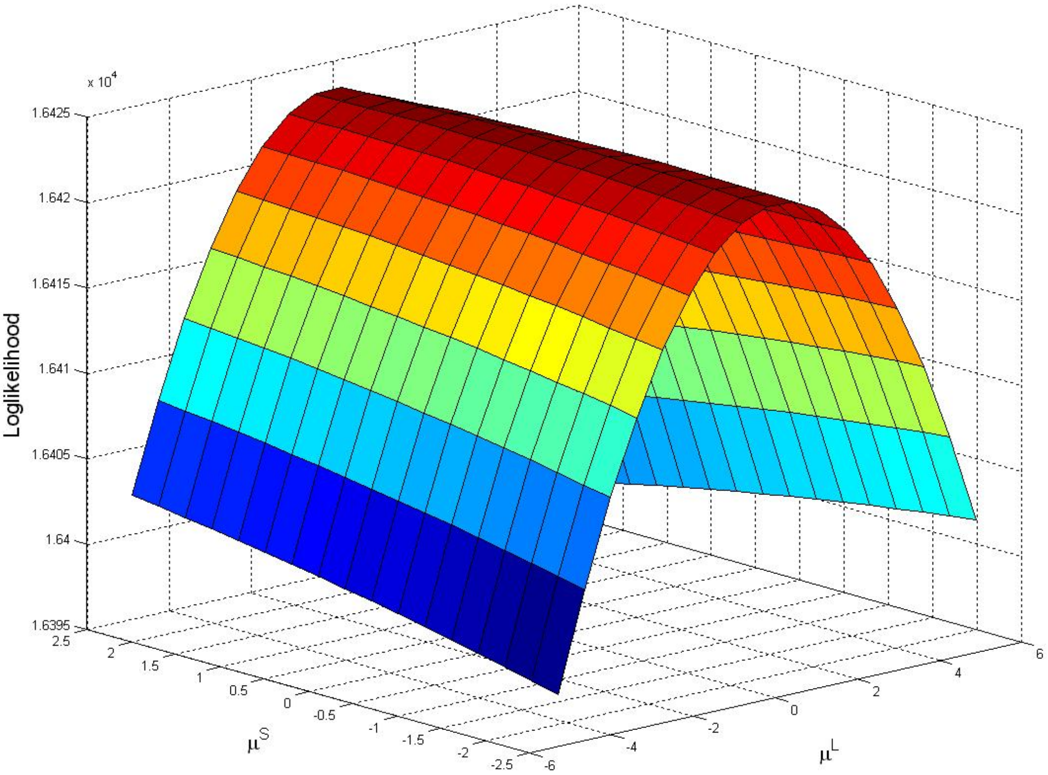

Figure 2.8: Loglikelihood values according to the Kalman Filter for the first Parameter estimate of the low-α subgroup of the cosh model with varying values for μ, assuming flexible $\tilde{\mu}$.

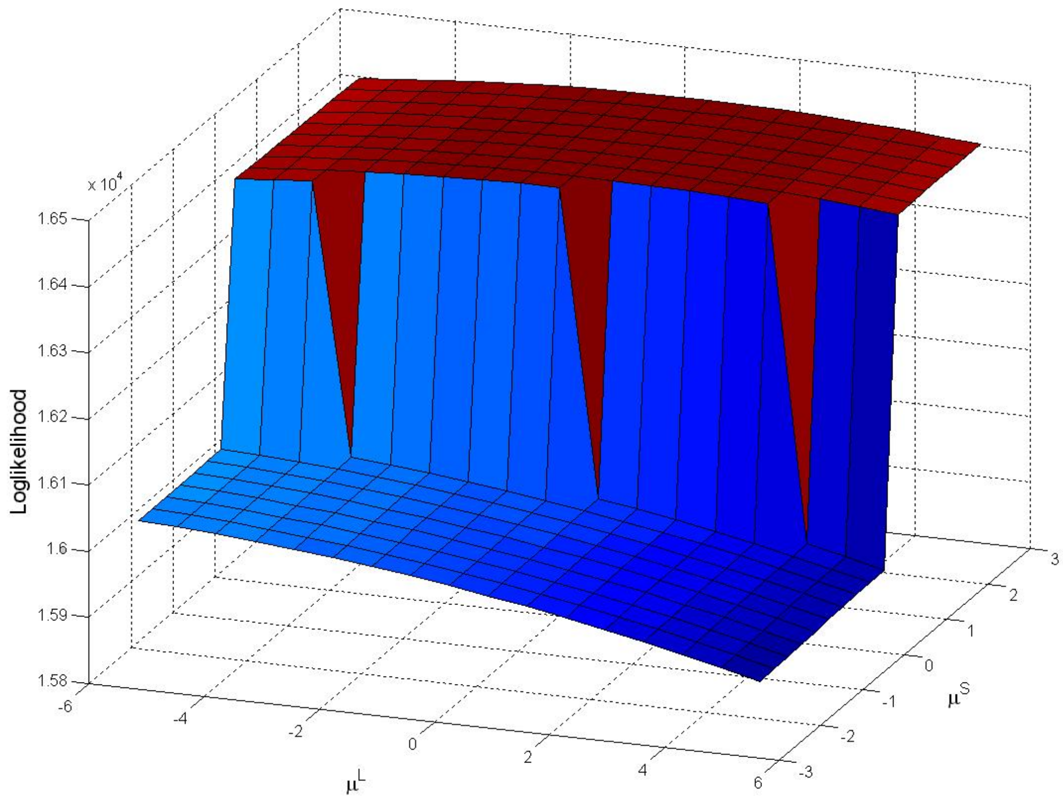

Figure 2.9: Loglikelihood values according to the Kalman Filter for the first Parameter estimate of the high-α subgroup of the cosh model with varying values for μ, assuming flexible $\tilde{\mu}$.

α	γ_1	γ_2	c	κ_1	κ_2	ρ	ν	LogL
0.0054	0.0226	-0.4758	36.95	0.48	0.0205	0.4967	0.0016	16409
0.0023	-0.4847	-0.0225	-10.08	0.02	0.4843	-0.4909	0.0016	16409
0.0615	-0.0237	0.2943	0.00	0.47	0.0275	0.3943	0.0016	16413
0.0612	-0.2964	0.0245	0.00	0.03	0.4658	0.4057	0.0016	16414

Table 2.25: Estimates of the cosh model assuming $\mu = \tilde{\mu} = 0$.

Slope	Level
0.38	2.97
0.63	2.98
0.55	0.99
0.35	1.00

Table 2.26: Ratios of the distances between empirical and estimated mean of the state process components and the empirical range of the state process components in case $\mu = \tilde{\mu} = 0$.

Restricting μ

In both the Cairns and the cosh model, we are free to choose the dynamics of the physical measure. Following the Cairns model, we specified the market price of risk so that the state factor processes in both models follow Ornstein-Uhlenbeck dynamics with long-term mean μ. Thanks to the computational efficiency of the cosh model, we are able to examine how restricting μ can avoid the instability of the estimates encountered. To measure improvements of the restricted models we employ again the ratio criterion introduced in 2.27.

First, we assume $\tilde{\mu} = \mu = 0$. This is the simplest implementation, which requires only 8 model parameters to be estimated for the two-factor case. Implicitly, this model identifies the reference measure $\tilde{\mathcal{P}}$ and the physical measure $\mathcal{P}$. Estimation results are given in table 2.25. We find again the low-mean reversion factor to coincide with long-end yields whereas the high mean reversion factor coincides with the slope. In table 2.26 we examine the ratio of deviations between empirical and estimated long-term mean μ to the range of the filtered state process component paths. Again, the ratio of the level factor is generally close to or above 1, indicating that the estimated long-term mean μ^L might never be reached by the filtered state process. In all cases $\mu = 0$ implies a very low term structure $g(\mu)$. Consequently, the restriction led to misspecification of the long-term mean in the same way as encountered in the unrestricted case.

In a second approach, we again identify the measures $\tilde{\mathcal{P}}$ and $\mathcal{P}$, yet assume flexible $\mu = \tilde{\mu}$. The parameter estimates are given in table 2.27. Again, the estimates are separated into

α	γ_1	γ_2	c	μ_1	μ_2	κ_1	κ_2	ρ	ν	LogL
0.006	0.023	0.475	2.15	-47.80	-4.24	0.485	0.021	-0.494	0.0016	16409
0.006	-0.023	0.475	-7.05	29.71	14.59	0.485	0.021	0.495	0.0016	16409
0.061	-0.024	0.297	-1.46	-25.28	2.88	0.467	0.028	0.406	0.0016	16414
0.061	-0.024	-0.299	-2.20	39.31	-10.55	0.471	0.028	-0.411	0.0016	16414

Table 2.27: Estimates of the cosh model assuming $\mu = \tilde{\mu}$.

Slope	Level
0.50	2.96
0.50	2.97
0.45	1.01
0.45	1.02

Table 2.28: Ratios of the distances between empirical and estimated mean of the high- and low mean-reversion state process components and the empirical range of the state process components in case $\mu = \tilde{\mu}$.

two groups according to the parameter α. The ratios in table 2.28 imply that the estimated long-term mean $\mu^L = \tilde{\mu}^L$ might never be reached by the filtered state process. A closer examination again showed that for all estimates, the long-term mean μ^L systematically underestimated the level in the sense that the long-term mean term structure $g(\mu; \theta)$ implies a very low term structure.

Finally, we separate the physical and the reference measure, allow for flexible $\tilde{\mu}$ yet restrict $\mu = 0$. This is a straightforward restriction of the general framework discussed before, for which we found instabilities in estimation of μ. Results are given in table 2.29, deviation-to-range ratios are given in table 2.30. In all cases the assumed value $\mu = 0$ does not lie within the range of the path processes, and $\mu = 0$ implied very low interest rates.

We can conclude that restricting $\mu = 0$ does not solve our problem. In general, the parameters μ, whether they are restricted or not, reflect very low term structures. This is most likely a result of the falling trend in long-term interest rates. As model restrictions do not improve the estimates, alternative specifications of the long-term mean μ are required.

α	γ_1	γ_2	c	$\tilde{\mu}_1$	$\tilde{\mu}_2$	κ_1	κ_2	ρ	ν	LogL
0.006	0.022	0.473	55.00	1.97	0.77	0.488	0.020	-0.543	0.0016	16424
0.006	-0.473	-0.022	42.58	-0.73	-1.97	0.021	0.488	-0.543	0.0016	16424
0.065	0.305	-0.024	0.08	-1.46	-2.15	0.027	0.480	0.427	0.0016	16446
0.066	0.314	0.024	-0.03	1.56	-2.06	0.026	0.489	-0.422	0.0016	16448

Table 2.29: Estimates of the cosh model assuming $\mu = 0$ but flexible $\tilde{\mu}$.

Slope	Level
3.08	0.06
3.08	0.06
0.86	0.01
0.89	0.01

Table 2.30: Ratios of the distances between empirical and estimated mean of the state process components to the empirical range of the state process components in case $\mu = 0$ yet flexible $\tilde{\mu}$.

We will discuss these alternative approaches in the next section.

Using historical data directly

We found that the Kalman filter can not estimate the long-term mean μ^L of the level factor properly. On the other side, we know that the filtered path of the level factor is highly correlated to the 10-year rate. Particularly, there seems to exist an affine link between the state vector X_t^L and the 10-year rate Y_t^{10}. Consequently, we can perform a linear regression

$$X_t^L = a + bY_t^{10} + \epsilon_t \tag{2.29}$$

which we expect to have high explanatory power. Now instead of estimating μ^L through the Kalman filter, we can use this prior information for both the cosh and the Cairns model to specify μ^L. As Loglikelihood sensitivities of μ^L are extremely small, we can estimate the full model using the Kalman filter and then redefine μ^L to arrive at a model equivalent in Loglikelihood values yet superior in fitting the long-term mean of the level process. Our primary condition of a superior fit of μ^L is $\mu^L \in conv\left\{\hat{X}_t^L, t = 1, \ldots, T\right\}$. To achieve this, we have several choices:

1. Taking the empirical mean of a mean-reverting process as approximation of its long-term mean, we get

$$\mu^L \approx \frac{1}{T}\sum_{t=1}^{T}\hat{X}_t^L,$$

thus the long-term mean of the state vector component X^L can be calculated approximately estimating the regression coefficients in (2.29) and calculating the empirical mean 10-year rate, hence

$$\mu^L \approx a + b\frac{1}{T}\sum_{t=1}^{T}Y_t^{10}.$$

2. Instead of using an approximation, we estimate the long-term mean μ^{10} of the observed 10 year yields, which we assume to follow an Ornstein-Uhlenbeck process. We

get

$$\mu^L = a + b\mu^{10}$$

by the Ito-Doeblin formula applied on $f(Y^{10}) := a + bY^{10}$. Such an estimation approach is based on a single directly observable time series. We therefore expect the estimate of the long-term mean μ^{10} to be more stable in comparison to the indirect Kalman filter-based estimation.

3. The previous approaches "inverted" first in the sense that the Kalman filter derived the implied state vector process $\hat{X}$ from term structure data observed. In a second step, we then took the average. Another idea would be to change this in the sense that we first calculate the average yield curve, and then "invert" in the sense that we calculate the state $\hat{X}$ which best fits the average historical yield curve. Namely, we can use g^{-1} to calculate the (approximate) state for any given yield curve, hence we define $\mu := g^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} Y_{t_i}\right)$. As μ^S is estimated properly, another approach would be to minimize $||g(\mu^S, \cdot) - \frac{1}{N}\sum_{i=1}^{N} Y_{t_i}||$, thus we estimate μ^S by the Kalman filter but calibrate μ^L to the long-term mean curve and hence the long-term mean level.

For the first approach, $\mu^L \in conv\left\{\hat{X}_t^L, t = 1, \ldots, T\right\}$ is guaranteed since $\frac{1}{T}\sum_{t=1}^{T} \hat{X}_t^L$ is a convex combination of historical states $\hat{X}_t^L$. The remaining alternative specifications of μ^L at least provide a significantly higher probability that the long-term mean of the level factor lies in the range of the filtered level factor. Considering the second approach, we assume that direct observability of Y^{10} should facilitate estimation of μ^{10} within the range of observations. Considering the last approach, fitting μ to the mean curve should be equivalent to fitting μ^L to the mean observed level and μ^S to the mean observed slope, which should correspond to a state within the range of filtered states $\hat{X}$.

Based on our original estimates of the Kalman filter, we tested the above described approaches. In all cases, the linear regression in (2.29) implied R^2 of at least 97.4%, thus comovement between the level factor and the 10-year yield is considerable[30]. This justifies the usage of the regression in the subsequent approaches.

- The first approach generally implied alternative estimates of μ^L which provided ratios of the level factor around 0.05 for the low-α estimates and 0.005 for the high-α estimates, hence a considerable improvement is found. Loglikelihood values decreased by 5 for all low-α estimates and 2 for all high-α estimates.

[30] Note that in section 2.3.9, we found that the level factor coincides with the highest observable yield. This would be the 30-year rate. As this is not available for the full sample, we used the 10-year rate, which is fully available. Note, however, that the small deviation between the level factor and the 10-year rate could be attributed to differences between the 10-year rate and the 30-year rate. Therefore, if we estimated the model using only maturities up to ten years, we can expect R^2 to be even higher.

- The second approach provides ratios of 0.05 for the slope and 0.4 for the level factor in case of the low-α estimates, for the high-α estimates the ratios for the slope decrease to 0.002. Loglikelihoods decreased by 4 in the low-α estimates and by 1 in the high-α estimates.

- the third approach examined estimated only μ^L by an inversion approach, that is

$$\mu^L = \arg\min_{x \in \mathbb{R}^+} \left| g(x, \mu^S) - \frac{1}{T} \sum_{t=1}^{T} Y_t \right|.$$

  Ratios for the level factor estimates were 0.015 for the high-α estimates and 0.013 for the low-α estimates. Loglikelihoods decreased by 4 for the low and by 1 for the high-α estimates.

- Finally, we estimated both μ^L and μ^S by inversion, resulting in ratios of μ^L of 0.055 and 0.015 for the low- and high-α estimates, respectively. For μ^S, the ratios were 0.02 for the low- and 0.002 for the high-α estimates. Loglikelihoods decreased by 5 and 1, respectively.

In figure 2.10, we provide the Kalman filter estimates of μ^L, the red line, and the range of alternative estimates, the shaded area. In figure 2.11, we provide the same for estimates of μ^S. We see that Kalman filter estimates and alternative estimates of μ^L differ considerably. On the other side, the estimates of μ^S coincide for all approaches including the Kalman filter. Deviation-to-range ratios were superior for all alternative approaches in comparison to the Kalman filter estimates. The second approach provides the highest ratios of the alternative approaches. It also still underestimates μ^L in the sense that the yield curve $g(\mu)$ implied by the Kalman filter estimate implies very low interest rates, albeit it underestimates μ^L to a lesser extent than the original Kalman filter. The estimates of μ by the other alternative approaches imply term structures well within the range of observable yield curves and in fact typically rather close to the empirical mean yield curve.

All alternative approaches imply a decrease in Loglikelihood values, which is, however, negligible. Neither from Loglikelihood values nor from an economic or implementational viewpoint can we reject the first, third or fourth alternative approach. We can conclude that these approaches are equivalent in identifying more reasonable estimates of μ^L given a parameter estimate. Note also that the exogenous specification of μ^S by the fourth approach was equivalent to estimates using the Kalman filter.

A major difference remaining within the first, third and fourth approach lies in the prior information they require. For the first and third approach, prior estimation using the original Kalman filter is still required and μ is only corrected in a second step. The first approach requires the filtered state space $\hat{X}$ of a Kalman filter estimate for the regression, the third approach requires that we can identify level factor, which according to estimation results is not clear a priori. Contrary to that, all prior information we have in the fourth

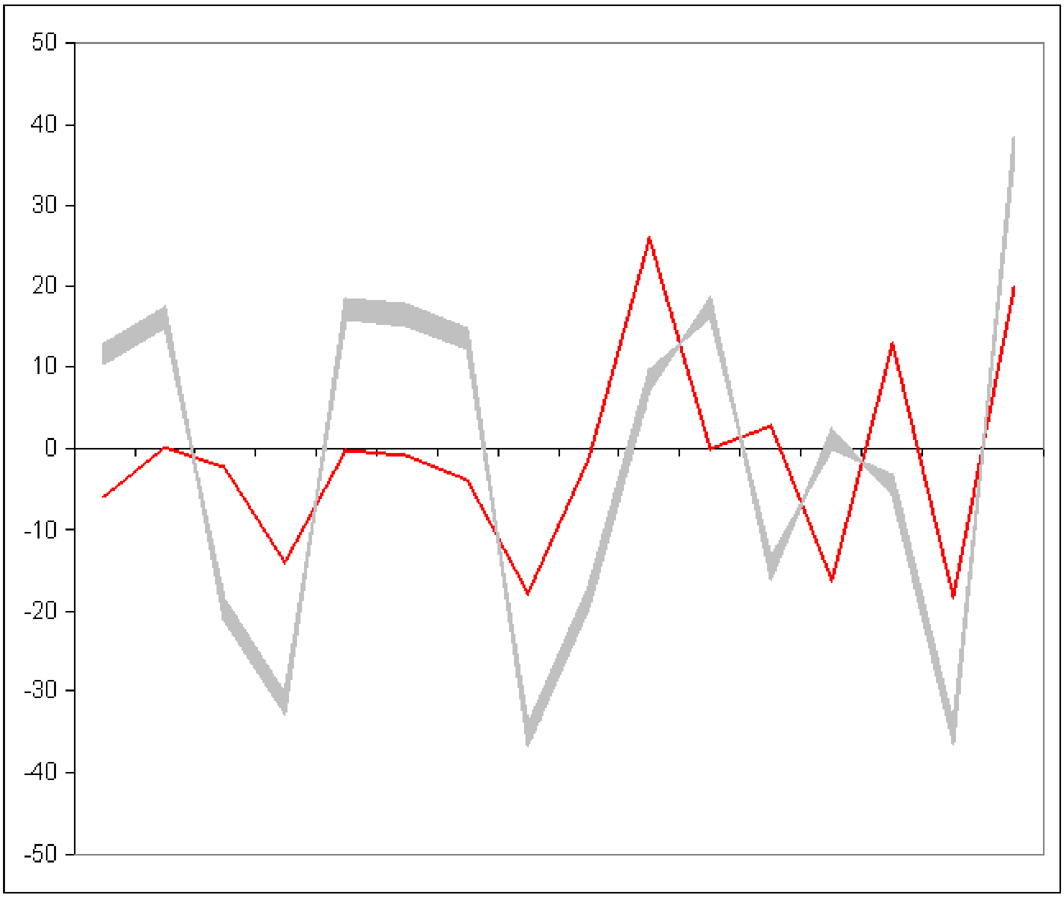

Figure 2.10: Kalman filter estimates of μ^L (red line) and range of alternative estimates of μ^L (shaded area) for several estimates of the cosh model.

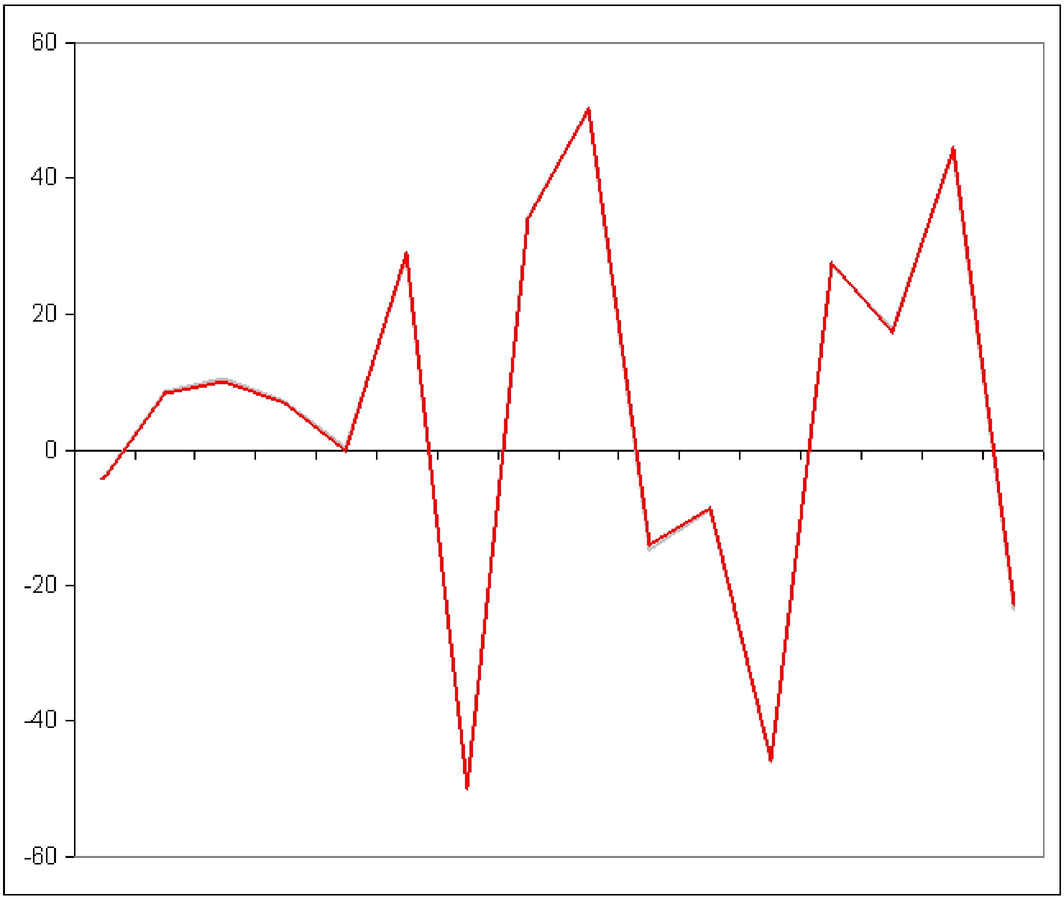

Figure 2.11: Kalman filter estimates of μ^S (red line) and range of alternative estimates of μ^S (shaded area, fully covered) for several estimates of the cosh model.

approach is the model definition. We do not assume that one particular state vector component coincides with a certain measurement. We therefore can employ the fourth approach directly in estimation: the Kalman filter starts with a parameter set θ which does not include μ. As the remaining model parameters are sufficient to define the measurement function g in both the Cairns and the cosh models, we can define $\mu := g^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Y_t\right)$ as an initial step in the Kalman filter as our calibration of $X_{0|0}$, and in fact based on the very same calibration algorithm. The Kalman filter then starts with a parameter set which implies a reasonable μ at the beginning of the filtering algorithm. As the fourth approach provides estimates of μ^S as well which were equivalent to Kalman filter estimates, this reduces the number of model parameters by d and at the same time avoids the trend-fitting problem of the original Kalman filter. We therefore strongly recommend to use the fourth approach for estimation.

Summary

In section 2.3, we found that both the Cairns and the cosh model had significant problems in identifying the long-term mean of the level factor. In particular, estimated long-term means μ^L implied Japan scenarios for the Cairns model and very low or even negative yields in the cosh model. In a sense, the Kalman filter therefore underestimated μ^L in both models. Examining this further we found that low mean reversion of the level factor together with a falling trend in historically observed long-end yields are the most likely explanation for this underestimation. Alternative approaches in specifying more realistic μ^L were required.

We provided several approaches to specify μ^L exogenously. In general, these approaches are based on the idea that within both models, the term structure is a function of a mean reverting process, which are closely related to the principal components of the yield curve. Therefore, the long-term mean of the underlying state process and the long-term mean of the observed yields should be related. As the long-end factor coincides with the 10-year rate, we were able to develop estimation approaches of μ^L which make use of this close relation. In particular, we were able to derive a regression equation which could be used to derive μ^L from the long-term mean μ^{10} of the 10-year yield.

Alternatively, we used the yield formula to invert the long-term mean term structure. This last approach proved to be superior as it required no prior information on state vector behavior and hence can be used already in the filtering algorithm itself. This expansion of the Kalman filter allows to estimate all parameters jointly, reduces the parameters in the optimization algorithm by d, and at the same time guarantees reasonable estimates of μ^L. It is therefore strongly recommended.

2.3.9 The parameter α

In section 2.3, we saw that in both models, yet particularly in the cosh model, high- and low-α estimates were derived. Among all model parameters of the Rogers framework, the parameter α has a special role, as α exists independent from the choice of the function f and the state vector dynamics X. Given the definition of the state price density

$$\varsigma_t := e^{-\alpha t} f(X_t),$$

it is clear that the parameter α shapes discounting functions for long time horizons. Namely insurance applications should therefore depend on realistic estimates of the parameter α. The question arises why the differences encountered in the estimation of α exist and whether we can improve the stability of our estimates of α.
Cairns showed that the parameter α in his model equals the asymptotic long (instantaneous) forward rate $\alpha = \lim_{T\to\infty} f(t,T)$. We will find that the parameter α is closely related to the asymptotic long rate $\lim_{T\to\infty} y(t,T)$ in all Rogers frameworks.
Note that in any term structure model,

$$\lim_{T\to\infty} P(t,T) = 0$$

holds. If we consider the general bond pricing formula of the Rogers framework, we get

$$0 = \lim_{T\to\infty} P(t,T) = \lim_{T\to\infty} e^{-\alpha(T-t)} \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}{f(X_t)}.$$

Such a limiting behavior can only be observed if either

1. The limit of the expected value is finite

$$c(x,t) := \lim_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < \infty$$

for all $t \geq 0$ and all $x \in \mathcal{X}$. Or

2. the limit does not exist, hence $c(t,x) = \infty$, but the term $e^{\alpha T}$ goes faster to infinity, thus for T high enough $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < e^{\alpha T}$ and $\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < \alpha e^{\alpha T}$. Finally,

3. the limit does not exist due to

$$\liminf_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < \limsup_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right].$$

We start with the third possibility. Considering the instantaneous forward rate

$$f(t,T) \quad = \quad \alpha - \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]},$$

we see that $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ must be differentiable and hence continuous in T for the instantaneous forward rate to exist. This implies that $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ is a continuous function varying between the lim inf and the lim sup of $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ for T large enough. Therefore $\frac{\partial}{\partial T}E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ repeatedly changes its sign, so the instantaneous forward rate repeatedly varies between $f(t,T) > \alpha$ and $f(t,T) < \alpha$, something we should exclude from economic reasons[31]. Now if $\liminf_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] = 0$ or $\limsup_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] = \infty$, we can easily define a sequence of times T_n for which the instantaneous forward rate explodes, which we must exclude from economic reasons, see also section 2.1. Nevertheless, we will already exclude $\frac{\partial}{\partial T}E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ changing its sign repeatedly, the reason being the variation in instantaneous forward rates.
We now consider the asymptotic long rate within a Rogers framework,

$$\begin{aligned}
\lim_{T\to\infty} y(t,T) &= -\lim_{T\to\infty} \frac{\log\left(P(t,T)\right)}{T-t} \\
&= \lim_{T\to\infty} \frac{\log\left(e^{-\alpha(T-t)}\frac{E^{\tilde{\mathcal{P}}}[f(X_T)|X_t=x]}{f(X_t)}\right)}{T-t} \\
&= \alpha - \lim_{T\to\infty} \frac{\log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]\right)}{T-t} + \lim_{T\to\infty} \frac{f(X_t)}{T-t} \\
&= \alpha - \lim_{T\to\infty} \frac{\log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]\right)}{T-t}
\end{aligned}$$

For

$$0 < \liminf_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] \le \limsup_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < \infty$$

we see that the asymptotic long rate equals α. This naturally holds for $0 < c(t,x) < \infty$ as well[32]. In case $c(x,t) = 0$, the general yield formula implies that α is a lower bound for the asymptotic long rate conditional on $X_t = x$, as for T large enough $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] < 1$ and hence the logarithm is negative. If $c(x,t) = \infty$, the assumed finiteness of $y(t,T)$ and the long-term limit of $P(t,T)$ imply that the expected value goes to infinity, yet slower than $\exp(\alpha(T-t))$. We find that in this case the parameter α is an upper bound for long-term nominal interest rates conditional on $X_t = x$.
Given this, we can partition the state space $\mathcal{X}$ according to the role of the parameter α conditional on $X_t = x$ as

$$\begin{aligned}
M_1 &:= \{x \in \mathcal{X} : \lim_{T\to\infty} y(x;t,T) < \alpha\}, \\
M_2 &:= \{x \in \mathcal{X} : \lim_{T\to\infty} y(x;t,T) = \alpha\}, \\
M_3 &:= \{x \in \mathcal{X} : \lim_{T\to\infty} y(x;t,T) > \alpha\}.
\end{aligned}$$

[31] We may also exclude this due to the Dybvig-Ingersoll-Ross theorem 2.3.2.

[32] In the following, we will only consider the cases $c(t,x) \in \mathbb{R}_0^+ \cup \{\infty\}$ as, from an economic viewpoint, variation between $f(t,T) > \alpha$ and $f(t,T) < \alpha$ for $T \to \infty$ should be excluded as well. Furthermore, if the Rogers framework is implemented with a sufficiently smooth twice differentiable function f and mean reverting state vector X we can reasonably expect to only find models with $c(t,x) \in \mathbb{R}_0^+ \cup \{\infty\}$.

Mean reversion of the state vector now implies that if $P(X_t \in M_i, X_{t+h} \in M_j) > 0$ then also $P(X_t \in M_j, X_{t+h} \in M_i) > 0$ for $h > 0$. The following theorem of Dybvig, Ingersoll and Ross [DIR96] helps to generalize the role of α to all $x \in \mathcal{X}$.

Theorem 2.3.2 (Dybvig, Ingersoll, Ross (1996))**.** *The asymptotic long rate* $\lim_{T\to\infty} y(t,T)$ *and the asymptotic long forward rate* $\lim_{T\to\infty} f(t,\tau,T)$ *can never fall in* t *in the absence of arbitrage.*

Proof. McCulloch (2000) [McC00] shows that there is a crucial error in the proof of Dybvig, Ingersoll and Ross for their theorem, but that the error can be corrected and the conclusion remains valid, however, the long-end limit of the yield curve is indeterminate. According to Hubalek, Klein, Teichmann (2002) [HKT02] the strategy in the proof of McCulloch (2000) is anticipative, so not admissible for a no-arbitrage argument. These authors provide a proof of the Dybvig-Ingersoll-Ross theorem without any additional assumptions. Schulze (2008) [Sch08] also provides an alternative proof of the Dybvig-Ingersoll-Ross theorem without anticipation by providing an explicit arbitrage strategy. □

Given the partition of $\mathcal{X}$ according to the role of α as introduced above, the fact that the asymptotic long rate is non-decreasing implies $P(X_t \in M_i, X_{t+h} \in M_j) = 0$ for $i < j$ and $h > 0$. If the state vector is mean reverting, two out of the three sets M_1, M_2 and M_3 must be null sets under the reference measure $\tilde{P}$ and by equivalence of the measures also under the physical measure $\mathcal{P}$ and the risk-neutral measure $\mathcal{Q}$. Consequently, the parameter α uniformly bounds asymptotic long rates $\lim_{T\to\infty} y(x;t,T)$ for all $x \in \mathcal{X}$. This is expressed in the following theorem.

Theorem 2.3.3. *Given a Rogers model with mean reverting state vector and function* f*, with*

$$\liminf_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] = \limsup_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$$

for all $x \in \mathcal{X}$ *and* $t \geq 0$ *and*

$$c(t,x) := \lim_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$$

for $c(t,x) \in \mathbb{R}_0^+ \cup \{\infty\}$*, we have that*

1. *if* $c(t,x) = 0$ *for one* $x \in \mathcal{X}$*, then* $\tilde{P}\left(x \in \mathcal{X} :\ c(t,x) = 0\right) = 1$ *and* $\alpha < \lim_{T\to\infty} y(t,T)$ $\tilde{P}$*-almost surely,*

2. *if* $0 < c(t,x) < \infty$ *for one* $x \in \mathcal{X}$*, then* $\tilde{P}\left(x \in \mathcal{X} :\ 0 < c(t,x) < \infty\right) = 1$ *and* $\alpha = \lim_{T\to\infty} y(t,T)$ $\tilde{P}$*-almost surely,*

3. *if* $c(t,x) = \infty$ *for one* $x \in \mathcal{X}$*, then* $\tilde{P}\left(x \in \mathcal{X} :\ c(t,x) = \infty\right) = 1$ *and* $\alpha > \lim_{T\to\infty} y(t,T)$ $\tilde{P}$*-almost surely.*

The theorem allows to derive the role of α directly from the choice of the function f and a single calculation of $c(t,x)$. As we required closed-form bond price formulae, the expected value $E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$ must be available in closed form and so derivation of $c(t,x)$ should be simple.
We will use the theorem to derive the role of α in the cosh model. For the Cairns model, we follow an alternative approach based on Cairns finding that α is the asymptotic long forward rate.

Theorem 2.3.4. *Within a Rogers model specified by the dynamics of the state process X under the reference measure $\tilde{\mathcal{P}}$ and a function $f : \mathcal{X} \to \mathbb{R}^+$ and under the conditions*

$$\liminf_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] = \limsup_{T\to\infty} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right]$$

and

$$\lim_{T\to\infty} \frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|X_t = x\right] = 0$$

for all $x \in \mathcal{X}$ the asymptotic long forward rate equals the asymptotic long rate

$$\lim_{T\to\infty} y(t,T) = \lim_{T\to\infty} f(t,T).$$

Proof. We have to distinct again the three cases $c(t,x) = 0$, $0 < c(t,x) < \infty$ and $c(t,x) = \infty$. First, we assume $c(t,x) = 0$. Consider the general formula for the asymptotic long rate

$$\begin{aligned}
\lim_{T\to\infty} y(t,T) &= \lim_{T\to\infty} \left[\alpha - \frac{\log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]\right)}{T-t} + \frac{\log(f(X_t))}{T-t}\right] \\
&= \alpha - \lim_{T\to\infty} \frac{\log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]\right)}{T-t}.
\end{aligned}$$

As $c(t,x) = 0$ we have $\lim_{T\to\infty} \log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]\right) = -\infty$, furthermore $\lim_{T\to\infty} T - t = \infty$, so we can use the rule of L'Hospital. We get

$$\begin{aligned}
\alpha - \lim_{T\to\infty} \frac{\log\left(E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]\right)}{T-t} &= \lim_{T\to\infty} \left(\alpha - \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}\right) \\
&= \lim_{T\to\infty} f(t,T).
\end{aligned}$$

Second, we assume a constant limit $0 < c(t,x) < \infty$. By the second condition,

$$\lim_{T\to\infty} \frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right] = 0$$

which implies

$$\lim_{T\to\infty} \left(\alpha - \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}\right) = \alpha.$$

Note that without the second condition, the asymptotic long forward rate still converges to α, yet it may regularly switch its sign, contrary to what the Dybvig-Ingersoll-Ross theorem states.

Third, we assume $c(t,x) = \infty$. In this case, we can again use L'Hospitals rule and get

$$\begin{aligned}\lim_{T\to\infty} y(t,T) &= \alpha - \lim_{T\to\infty} \frac{E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}{T-t} \\ &= \alpha - \lim_{T\to\infty} \frac{\frac{\partial}{\partial T} E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]}{E^{\tilde{\mathcal{P}}}\left[f(X_T)|\mathcal{F}_t\right]},\end{aligned}$$

which concludes the proof. □

Note that theorem 2.3.4 not necessarily holds for general term structure models. Given

$$y(t,T) = -\frac{\log\left(P(t,T)\right)}{T-t}$$

we have

$$\begin{aligned}f(t,T) &= -\frac{\partial}{\partial T}\log\left(P(t,T)\right) \\ &= \frac{\partial}{\partial T}\left[y(t,T)(T-t)\right] \\ &= (T-t)\frac{\partial}{\partial T}y(t,T) + y(t,T).\end{aligned}$$

Now for $T \to \infty$,

$$\lim_{T\to\infty} f(t,T) = \lim_{T\to\infty}(T-t)\frac{\partial}{\partial T}y(t,T) + \lim_{T\to\infty} y(t,T)$$

where the second term equals the asymptotic long rate, whereas in the first term the first half $(T-t)$ converges to infinity and the second half $\frac{\partial}{\partial T}y(t,T)$ converges to zero, so we can not derive a general result.

We can, however, use theorem 2.3.4 to derive the asymptotic long rate in the Cairns model in the following corollary.

Corollary 2.3.5. *The asymptotic long rate and the asymptotic long forward rate of the Cairns model equal the parameter α.*

Proof. Cairns already proved that the parameter α equals the asymptotic long forward rate. By Theorem 2.3.4

$$\lim_{T\to\infty} y(t,T) = \lim_{T\to\infty} f(t,T) = \alpha.$$

□

To derive the role of α in the cosh model, we can use theorem 2.3.3.

Theorem 2.3.6. *The asymptotic long rate and the asymptotic long forward rate of the cosh model equal the parameter α.*

Proof. Since the limit

$$\begin{aligned}\lim_{T\to\infty} E\left[\cosh(\gamma X_T + c)|X_t = x\right] &= \lim_{T\to\infty} \cosh(\gamma E\left[X_T|X_t = x\right] + c)\exp\left(\frac{\gamma^T\Sigma\gamma}{2}\right)\\ &= \cosh(\gamma^T\mu + c)\exp\left(\frac{\gamma^T\Sigma\gamma}{2}\right)\end{aligned}$$

is finite and positive for all $x \in \mathcal{X}$, α is the constant asymptotic long rate in the cosh model by theorem 2.3.3. By theorem 2.3.4, the asymptotic long instantaneous rate equals α. □

As f is a positive function and X is mean reverting we can assume that for all practical purposes choices of f and X result in $0 < c(t, x) < \infty$, so that the asymptotic long rate is constant and equals the parameter α. The question arises what impact the unobservable asymptotic long rate has on observable yields, which in turn shows whether α can be estimated properly.

Figure 2.12 shows the difference between a yield curve implied by a high-α estimate and a low-α estimate of the cosh model, each of these two extrapolated to 50 years of maturity. In figure 2.13, we examine the same for the Cairns model. We see that the parameter α shapes long-term interest rates and thus long-term state price densities, as expected. The parameter α is therefore crucial to discount payoffs with maturities beyond 20 years and hence to the pricing of life and pension insurance contracts. On the other side, the impact of α on observable yields is rather small, which may explain the instability of our estimates. We will have to examine further whether the parameter α can be estimated properly and, if not, how we can determine α alternatively.

The asymptotic long rate within other term structure models

Before examining our ability to estimate the asymptotic long rate, we compare our results to other standard models which are often used in insurance applications, namely the Hull-White model, the Black-Karasinski model and the affine model framework, see for example the books of Filipovic [Fil09] or Brigo and Mercurio [BM01] for an overview of all these models. Our primary concern is to show that most term structure models suffer from problems in determining the asymptotic long rate.

For the Hull-White model [HW90], (simulated) yields at all maturities depend on the initial term structure of nominal interest rates and forward rates. We expect the asymptotic long rate to depend on the initial term structure as well. The Hull-White model is based on shortrate dynamics

$$dr_t = (\vartheta_t - ar_t)dt + \sigma dW_t \tag{2.30}$$

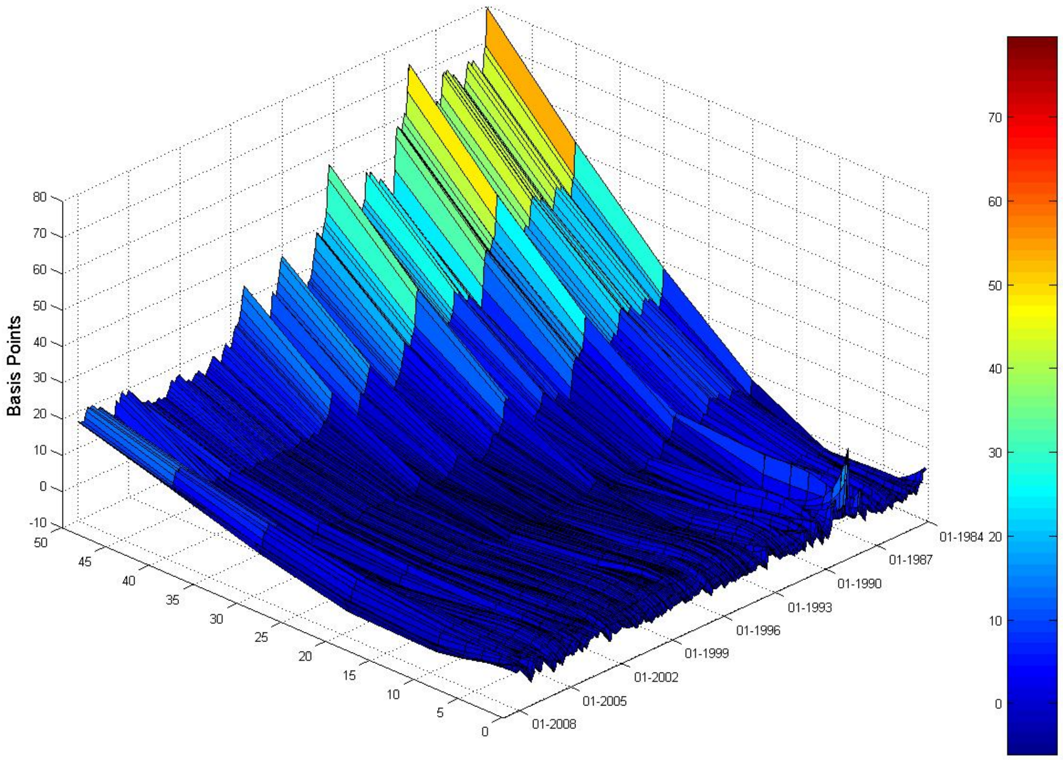

Figure 2.12: Difference in Basis points for the model-implied yield curves of a high-α estimate minus a low-α estimate within the Cosh model.

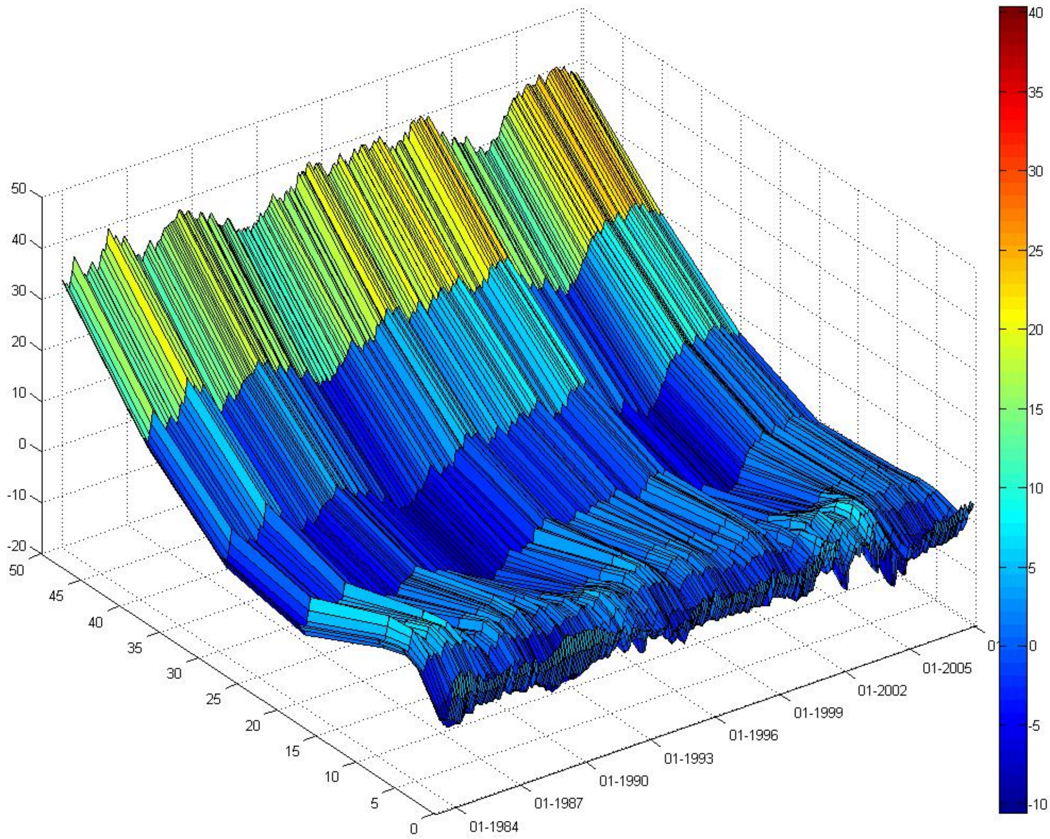

Figure 2.13: Difference in Basis points for the model-implied yield curves of a higher-α estimate minus a lower-α estimate within the Cairns model.

The asymptotic long rate is hence given by

$$\begin{aligned}
\lim_{T\to\infty} y(t,T) &= \lim_{T\to\infty} -\frac{1}{T-t}\log\left(P(t,T)\right) \\
&= \lim_{T\to\infty} -\frac{1}{T-t}\log\left(A(t,T)e^{-B(t,T)r_t}\right) \\
&= -\lim_{T\to\infty}\frac{1}{T-t}\left(\log\left(\frac{P^M(0,T)}{P^M(0,t)}\right) + B(t,T)f^M(0,t)\right. \\
&\quad \left. -\frac{\sigma^2}{4a}(1-e^{-2at})B(t,T)^2 - B(t,T)r_t\right)
\end{aligned}$$

whereby $P^M(0,t)$ denotes the observed market price at time $t=0$ of a zerobond which matures at time t and $B(t,T) = \frac{1}{a}\left(1-e^{-a(T-t)}\right)$. Note that by simple no-arbitrage relations[33]

$$\log\left(\frac{P^M(0,T)}{P^M(0,t)}\right) = -f^M(0,t,T)(T-t).$$

We get

$$\begin{aligned}
\lim_{T\to\infty} y(t,T) &= -\lim_{T\to\infty}\frac{1}{T-t}\Big(-f^M(0,t,T)(T-t) + B(t,T)f^M(0,t) \\
&\quad -\frac{\sigma^2}{4a}(1-e^{-2at})B(t,T)^2 - B(t,T)r_t\Big) \\
&= -\lim_{T\to\infty}\left(-f^M(0,t,T) + \frac{B(t,T)f^M(0,t)}{T-t}\right. \\
&\quad \left. -\frac{\sigma^2}{4a(T-t)}(1-e^{-2at})B(t,T)^2 - \frac{B(t,T)r_t}{T-t}\right) \\
&= \lim_{T\to\infty} f^M(0,t,T)
\end{aligned}$$

since $B(t,T)$ is bounded for all $T \geq 0$. This implies that the asymptotic long rate in the Hull-White model is deterministic, yet unobservable.

This points to a major problem in using the Hull-White model for long-term pricing, as market instantaneous forward rates $f^M(0,t)$ are only available for t up to the highest observable bond maturity in the market. For the US, this is not more than 30 years. If simulations are required for higher times to maturity, forward rates must be extrapolated from market data. This typically implies the specification of the asymptotic long rate as well, an approach we will discuss later. Generally speaking, implementing a constant asymptotic long forward rate to extrapolate the observed instantaneous forward rate curve

[33] Investing until T at the current interest rate $y(0,T)$ is equivalent to investing until $t < T$ at rate $y(0,T)$ and then securing the instantaneous forward rate $f(0,t,T)$ for the time $[t,T]$. This implies

$$\begin{aligned}
\exp\left(y(0,T)T\right) &= \exp\left(y(0,t)t\right)\exp\left(f(0,t,T)(T-t)\right) \\
\frac{\exp\left(y(0,T)T\right)}{\exp\left(y(0,t)t\right)} &= \exp\left(f(0,t,T)(T-t)\right) \\
\frac{P(0,t)}{P(0,T)} &= \exp\left(f(0,t,T)(T-t)\right).
\end{aligned}$$

consistently for varying applications is recommended.

The Black-Karasinski model [BK91] does not provide closed form bond prices, so that we can not calculate the limit of the yield curve directly. Nevertheless, Yao [Yao98] provides a criterion which helps to identify model parameters which ensure a finite upper bound for the asymptotic long rate. Using Jensen's inequality, we get that

$$\begin{aligned} 0 &\leq \lim_{T\to\infty} y(t,T) = -\lim_{T\to\infty} \frac{1}{T-t} \log\left(P(t,T)\right) \\ &= -\lim_{T\to\infty} \frac{1}{T-t} \log\left(E^{\mathcal{Q}}\left[e^{-\int_t^T r_s ds}|\mathcal{F}_t\right]\right) \\ &\leq -\lim_{T\to\infty} \frac{1}{T-t} \log\exp\left(-E^{\mathcal{Q}}\left[\int_t^T r_s ds|\mathcal{F}_t\right]\right) \\ &= \lim_{T\to\infty} \frac{1}{T-t} \int_t^T E^{\mathcal{Q}}\left[r_s|\mathcal{F}_t\right] ds, \end{aligned} \tag{2.31}$$

thus boundedness of the expected shortrate implies boundedness of the asymptotic long rate. Based on the stochastic differential equation defining the log-shortrate dynamics in the Black, Derman and Toy (1990) [BDT90] model, Yao finds a closed formula for the expected shortrate which may be used to derive those model parameters which ensure boundedness of the shortrate and hence the asymptotic long rate. As interest rate explosion is a frequently found problem of Black-Karasinski models, it is not clear whether there exists a practical approach to derive model estimates which ensure bounded interest rates including the asymptotic long rate.

Considering the affine model, we refer again to Yao [Yao98], who already proved for the framework of Duffie and Kan [DK96] that the asymptotic long rate is constant. For the one-factor Vasicek [Vas77] model with shortrate dynamics

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t^{\mathcal{Q}}$$

and the Cox-Ingersoll-Ross [CIR85] model with shortrate dynamics

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t^{\mathcal{Q}} \tag{2.32}$$

Yao calculates the asymptotic long rates $l(t) = \lim_{T\to\infty} y(t,T)$ as

$$\begin{aligned} l^{Vasicek}(t) &= \theta - \frac{\sigma^2}{2\kappa^2} \\ l^{CIR}(t) &= \frac{2\kappa\theta}{\kappa + \sqrt{\sigma^2 + \kappa^2}} \end{aligned}$$

whereby for the Vasicek model $\kappa > 0$ is required and for the CIR model $\sigma \neq 0$.
We will furthermore present the asymptotic long rate for the Chen-Scott framework [CS92], which is essentially a multi-factor affine model whose state vector components all follow CIR dynamics as given in (2.32). Because affine term structure models guarantee positive

interest rates only if all state vector components follow CIR dynamics, these models were recommended by Fischer, May and Walther [FMW04] for insurance applications. The bond pricing formula for the Chen-Scott framework is given by

$$P(t,T) = \prod_{i=1}^{n} A_i(T-t)e^{-B_i(T-t)X_t^{(i)}}$$

where

$$A_i(T-t) := \left[\frac{2h_i e^{(\kappa_i+\lambda_i+h_i)\frac{T-t}{2}}}{2h_i + (\kappa_i+\lambda_i+h_i)(e^{(T-t)h_i}-1)}\right]^{2\kappa_i\mu_i/\sigma_i^2}$$

and

$$B_i(T-t) = \frac{2(e^{(T-t)h_i}-1)}{2h_i + (\kappa_i+\lambda_i+h_i)\left(e^{(T-t)h_i}-1\right)}$$

with $h_i = \sqrt{(\kappa_i+\lambda_i)^2 + 2\sigma_i^2}$. The parameters λ_i define the components of the market price of risk $\Lambda(X_t) \in \mathbb{R}^d$ by $\Lambda_i(X_t) = \lambda_i\sqrt{X_t^{(i)}}$. The asymptotic long rate is given by

$$\lim_{T\to\infty} y(t,T) \quad = \quad -\lim_{T\to\infty}\frac{1}{T-t}\sum_{i=1}^{n}\left(\log(A_i(T-t)) - B_i(T-t)X_t^{(i)}\right),$$

we have to calculate the limits of $\log(A_i(T-t))$ and $B_i(T-t)$ for $T \to \infty$. Thus

$$\begin{aligned}
&\lim_{T\to\infty}\frac{\log(A_i(T-t))}{T-t} \\
=\ & \lim_{T\to\infty}\frac{2\frac{\kappa_i\mu_i}{\sigma_i^2}\left(\log(2h_i) + (\kappa_i+\lambda_i+h_i)\frac{T-t}{2} - \log\left(2h_i + (\kappa_i+\lambda_i+h_i)\left(e^{h_i(T-t)}-1\right)\right)\right)}{T-t} \\
=\ & 2\frac{\kappa_i\mu_i}{\sigma_i^2}\frac{\kappa_i+\lambda_i+h_i}{2} + \lim_{T\to\infty}\frac{-\log\left(2h_i + (\kappa_i+\lambda_i+h_i)\left(e^{h_i(T-t)}-1\right)\right)}{T-t} \\
\overset{L'H}{=}\ & 2\frac{\kappa_i\mu_i}{\sigma_i^2}\frac{\kappa_i+\lambda_i+h_i}{2} - \lim_{T\to\infty}\frac{h_i(\kappa_i+\lambda_i+h_i)\left(e^{h_i(T-t)}-1\right)}{2h_i + (\kappa_i+\lambda_i+h_i)\left(e^{h_i(T-t)}-1\right)} \\
=\ & \frac{(\kappa_i+\lambda_i+h_i)\kappa_i\mu_i}{\sigma_i^2} - h_i
\end{aligned}$$

and

$$\begin{aligned}
\lim_{T\to\infty}\frac{B_i(T-t)}{T-t} &= \lim_{T\to\infty}\frac{2(e^{h_i(T-t)}-1)}{(T-t)\left(2h_i + (\kappa_i+\lambda_i+h_i)(e^{h_i(T-t)}-1)\right)} \\
&= \lim_{T\to\infty}\underbrace{\frac{1}{T-t}}_{\to 0}\underbrace{\frac{2(1-e^{-h_i(T-t)})}{\left(2h_ie^{-h_i(T-t)} + (\kappa_i+\lambda_i+h_i)(1-e^{-h_i(T-t)})\right)}}_{\to\frac{2}{\kappa_i+\lambda_i+h_i}} = 0.
\end{aligned}$$

Thus, the asymptotic long rate is given by

$$\begin{aligned}
\lim_{T\to\infty} y(t,T) &= \lim_{T\to\infty}\sum_{i=1}^{n}\left(\log(A_i(T-t)) - B_i(T-t)X_t^{(i)}\right) \\
&= \sum_{i=1}^{n}\left(\frac{(\kappa_i+\lambda_i+h_i)\kappa_i\mu_i}{\sigma_i^2} - h_i\right).
\end{aligned}$$

To summarize, a constant asymptotic long rate is frequently found within term structure models, notably the affine model framework. In case of the Black-Karasinski and Black, Derman and Toy models, the asymptotic long rate can not be calculated due to lack of closed form bond prices. However, Yao [Yao98] shows that if we require the short-rate to be bounded, the asymptotic long rate is bounded as well. It is not clear, however, whether we are able to derive parameter estimates which ensure boundedness for all interest rates in the Black-Karasinski framework.
A constant asymptotic long rate within any term structure model is a function of the model parameters to be estimated. Consequently, the implicit assumption of an asymptotic long rate introduces mutual dependencies between the model parameters. In particular, note that within shortrate models, all parameters have a direct interpretation as to how they influence the short rate and in estimation, short-end influences typically dominate. This implies that the asymptotic long rate in most models is a function of short-end model parameters.

Estimation problems

The question arises whether a constant asymptotic long rate can be estimated properly from available data. In the following we will discuss stylized facts regarding the observable long end of the term structure and the implied difficulties in estimating the asymptotic long rate, be it constant or stochastic.

- The first problem lies in availability of long-term interest rate data. For the US, the highest maturity observable for government bonds is 30 years. In Germany, the highest maturity is only 15 years. The question arises whether 15 years of maturity or even 30 years provide sufficient data with respect to estimation of the asymptotic long rate.

- A second problem lies in censored long-term data. In the US, 20- and 30-year treasury bonds were not auctioned continuously. Only for a subsample of the dataset used for estimation a full term structure was available, weakening the database for estimation of the asymptotic long rate even more.
  Note that the U.S. Treasury likely managed its debt duration according to expected interest rates. If long-term yields are considered low, the treasury increases duration of the debt outstanding, thus continues or reintroduces 20- and 30-year bonds. If long-term yields are considered high, the treasury likely decreases duration and cancels or diminishes the issue of long-end bonds. The overall goal is to preserve favorable interest rates. The available dataset of long-term yields may therefore be biased toward low long-end yields due to debt management.

- It is a stylized fact that interest rate volatility decreases with maturity. As a consequence, volatility of asymptotic long forward rates decreases with maturity as well. Nevertheless, Gürkaynak, Sack and Swanson [GSS03] show that observable forward rates at long horizons do react significantly to macroeconomic and monetary policy surprises, especially regarding expected inflation. They find that the observable long end of the term structure still shows significant volatility, which increases our doubts regarding the usability of observable long-term yields to estimate the asymptotic long rate.

- Brown and Schaefer [SB00] and Christiansen [Chr01] found[34] that observable long-term forward as well as discount bond rates[35] almost always are downward sloping based on daily treasury STRIPS data from 1985 to 1994. Figure 2.14 plots the 30-year rate against the 20- and 30-year slope, showing that indeed the slope was negative most of the time. Examining treasury yield data from 1947 to 2008, we find that the slope between the 10-year and 20-year rate was negative in 170 out of 580 available month-end observation pairs, but the slope between the 20- and 30-year rates was negative in 163 out of 208 observation pairs, providing anecdotal evidence for a twist in the slope between 20 and 30 years of maturity. The full data set, on the other side, consists of 734 observation dates. Again, censored data might introduce a bias in observations.
  The negative slope implies that observable long-end yields form an upper bound for the asymptotic long rate whenever the negative slope was observed. Given again figure 2.14, this implies that the asymptotic long rate is below 5%. In particular, this implies that observable long-end yields provide very little information about the height of the constant asymptotic long rate.

- In some countries, regulators require insurance companies and pension funds to follow a strict approach to match assets and liabilities, which requires these investors to hold a certain amount of very long-term bonds according to their long-term liabilities. Typically, this imposes a regulatory requirement to buy and hold domestic long-term government bonds. Although such a regime might have some benefits in risk management, the primary beneficiary would be the government since excess demand

[34] Brown and Schaefer [SB00] relied in their analysis on an affine term structure model. Christiansen, on the other side, came to the same results as the previous authors by using time series models in her analysis, thus her results are based solely on empirical data and therefore are free of model risk.

[35] To see this

$$
\begin{aligned}
f(t,T,T+2) &< f(t,T,T+1) \\
(T+2-t)y(t,T+2) - (T-t)y(t,T) &< (T+1-t)y(t,T+1) - (T-t)y(t,T) \\
y(t,T+2) + (T+1-t)(y(t,T+2) - y(t,T+1)) &< 0
\end{aligned}
$$

Now given positive interest rates, this implies a negative slope in nominal yields as well.

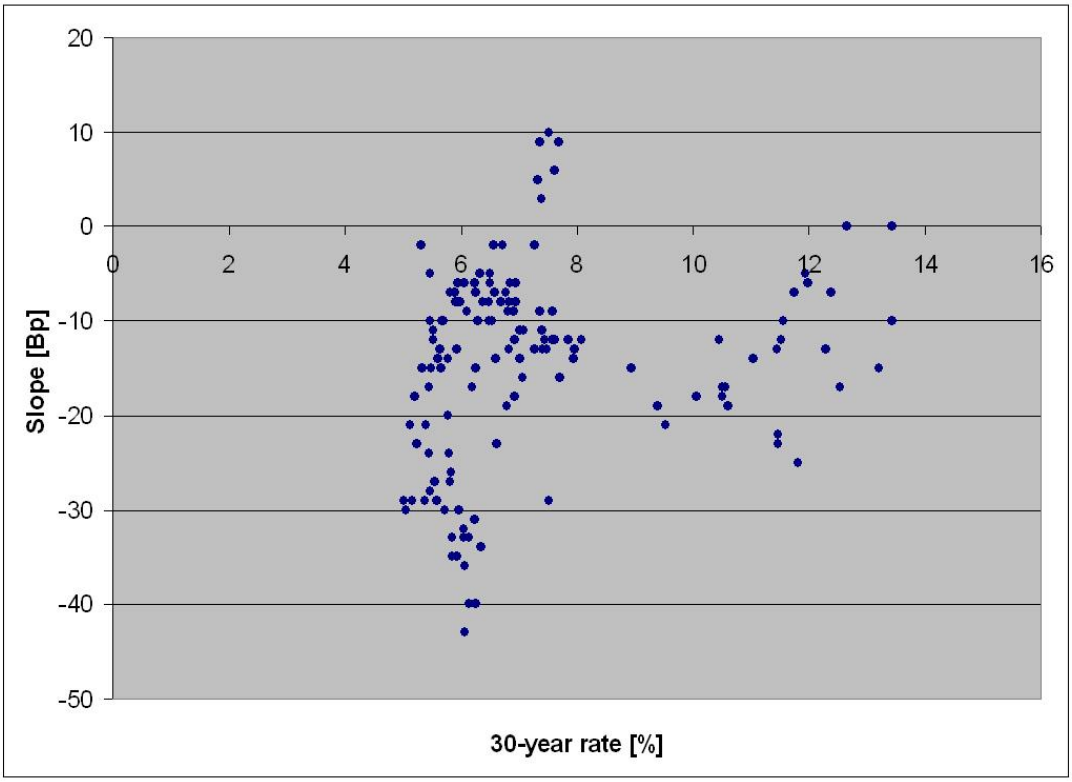

Figure 2.14: Plot of the Slope between the 20 and 30 year rates in basis points - if observed - against the 30-year rate

decreases long-term bond yields. If regulatory requirements determine long-term bond prices rather than demand and supply by rational investors, long-term rates do not reflect equilibrium riskless interest rates.

- Market liquidity for government bonds with maturities above 10 years is generally lower than for shorter maturity government bonds[36]. This implies a liquidity premium on treasury bonds with maturities beyond 10 years. If there exists a positive liquidity premium in long-term yields, observable yields overestimate riskless interest rates and the twist in the term structure is likely underestimated.

We see that the data on the very long end of the yield curve implies some problems in estimation. First, the observed maturities might not be sufficient to cover the very long end of the term structure. Second, long-end data might be censored and we expect that censored data implies a significant bias toward lower long-term yields in observable data. Third, available long-end data shows considerable variation whereas the asymptotic long rate is either constant or moves only rarely. Finally, the twist in the term structure and liquidity premia imply that long-term yields may systematically deviate from long-term riskless interest rates.

[36]One reason might be again that institutional investors are forced to buy and hold long-term bonds by asset-liability regulatory rules.

To summarize, we have to expect significant difficulties in estimating the asymptotic long rate. We have to examine the estimates of α closely and should examine sensitivities of term structure models to changes in the asymptotic long rate. At the end, we might have to specify the asymptotic long rate exogenously if estimation does not work properly.

Estimation by the cosh and Cairns models

In the previous examination of various term structure models, we found that most models imply a constant asymptotic long rate, which in turn is a function of the model parameters. This introduces an unwelcome restriction on model parameters, in particular dependencies between the model parameters which we typically do not account for in estimation. In the Rogers models we considered so far, only the parameter α describes the asymptotic long rate, whereas the other model parameters are unaffected. This makes the Cairns and cosh models particularly interesting for examinations of the asymptotic long rate. For once, the models will estimate the asymptotic long rate without restricting the remaining model parameters, as would, for example, be the case in affine models. Second, the models can be used to derive sensitivities of term structures or derivative prices to the asymptotic long rate, which in affine models requires non-identifiable changes in several model parameters[37]. Finally, if we are not able to estimate the asymptotic long rate properly, the models can be used to compare exogenous specifications of the asymptotic long rate without these specifications determining short-end dynamics, as would, again, be the case in affine models.

A first analysis of our ability to estimate the parameter α is based on the sensitivity of the Loglikelihood values and historical errors on said parameter. Note again that such an analysis would be infeasible in term structure models in which the asymptotic long rate depends on multiple model parameters. Figures 2.15 and 2.16 show a clear distinction between the high- and low-α estimates for both models. Sensitivities to changes of the asymptotic long rate is greater for the Cairns model than for the cosh model. Given unobservability of the asymptotic long rate, we can conclude that stability of the Loglikelihood value with respect to α is sufficient within both models.

In both models, we find the extent of the curvature pattern within cross-sectional errors to depend on α. Now obviously α governs the long-end of the yield curve. It might be that α takes over the role of the stochastic curvature factor on the very long end of the yield

[37]Specifically, if the asymptotic long rate is a function $f(\theta)$ of model parameters $\theta \in \Theta$, then each asymptotic long rate α implies a subset $M(\alpha)$ of the parameter space Θ such that $f(\theta) = \alpha$ for all $\theta \in M(\alpha)$. Considering sensitivities of changes from α_1 to α_2 then implies two parameter sets $M(\alpha_1) \subset \Theta$ and $M(\alpha_2) \subset \Theta$ such that infinitely many pairs of parameter sets (θ_1, θ_2) lead to the desired change in the asymptotic long rate $f(\theta_1) - f(\theta_2) = \alpha_1 - \alpha_2$.

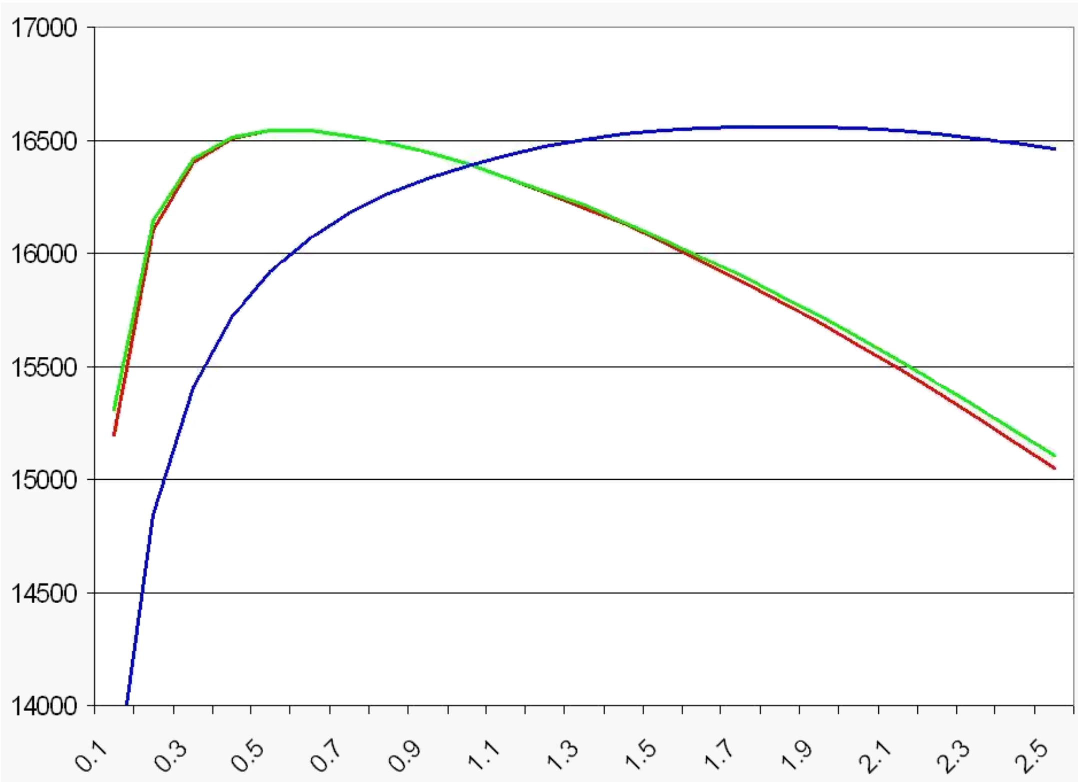

Figure 2.15: Loglikelihood values for varying α in percentage points for the three estimates available of the Cairns model.

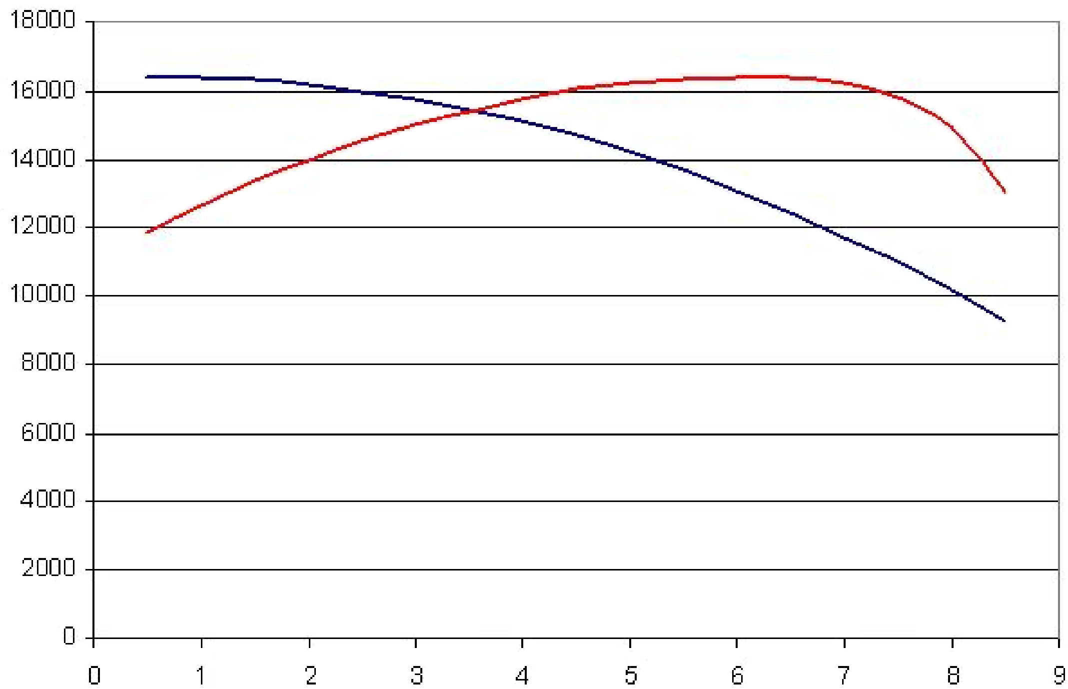

Figure 2.16: Loglikelihood values for varying α in percentage points for the three estimates available of the cosh model.

curve. As a third state vector component increases curvature fit in both models, we would expect to get more reasonable estimates of α in a three factor model. As seen in section 2.3.7, however, a three factor cosh model still provides the high- and low-α subgroups in estimation. Considering curvature, the typical "curvature pattern" in error correlations vanishes as well as the curvature pattern in MAEs and mean errors. Curvature can not explain why the the asymptotic long rate is estimated either high or low.

Another cause of this variability of α might be censored data. To exclude the impact of censored data, we extended the dataset to provide us with a full time series of 20- and 30-year interest rates. Extension uses the mean slope between the 20 and 30-year rates, as far as observed. Missing 20 year rates were defined by the 30 year rate minus the (negative) mean slope, missing 30 year rates were defined as the 20 year rate plus the (negative) mean slope. The idea now is that if censored data is to blame for the variability of our estimates of α, then estimation using the extended data set should result in stable estimates of α. Re-estimation of the model using the extended dataset indeed resulted in uniformly low estimates of $\alpha \approx 0$. We can conclude that the high-α estimates are likely a result of fitting curvature in censored data. Second, note that model-implied yields failed to produce the twist in the slope, particularly if long-end yields are low. Third, whereas the state vector components still coincide with level and slope proxies as used earlier, in this case the 30 year-rate was the best proxy for the level factor.
To examine this further, we re-estimated the cosh model twice, once excluding 30-year rates from the extended dataset, once excluding both 20- and 30-year rates. Note that both datasets do not contain empirical evidence for the twist in the term structure. In the first case, we get $\alpha \approx 5\%$, in the second case $\alpha \approx 4.5\%$. We also found that in both cases the high end of the observable term structure provided the level proxy. Interestingly, mean reversion increased substantially for the shorter datasets.
We repeated these exercises for the Cairns model as well, with equivalent results: estimated α increases to around 4.5% if the 30-year rates are omitted, indicating that low α is a result of the twist in observable data. The level factor always coincides with the long-end of the observable yield curve. Finally, shortening the dataset increases mean reversion. We can conclude several stylized facts for both models:

- The long end of the observable term structure is the best level proxy. This implies that censored long-end data should be avoided.

- The higher the maturity of the interest rate the level proxy coincides with, the lower mean reversion of the level factor.

- The estimate of the asymptotic long rate is highly dependent on the highest maturity observable. Inclusion of the twist in the estimation dataset implies $\alpha \approx 0$.

- Even $\alpha \approx 0$ is not sufficient to reproduce the twist observed in real data in model-implied yield curves.
- Without the twist in the estimation dataset, $\alpha \approx 4\%$ in both models, which is still rather low given historically observed long-term yields.

We found significant estimation problems considering estimation of the asymptotic long rate with our proposed dataset. The question arises how to increase the dataset. Rogers models allow for joint international bond market models. This effectively increases the dataset by multiple term structures and exchange rates:

- Real rates can be extracted from inflation-indexed bonds. If we use the Rogers model on inflation-indexed bonds and achieve a stable estimate of α^{real}, the nominal asymptotic long rate is given by

$$\alpha^{nominal} = \bar{i} + \alpha^{real}.$$

  The asymptotic inflation rate $\bar{i}$ can be determined "‘forward looking"’ as the inflation target of the central bank, or $\bar{i}$ can be specified "‘backward looking"’ as a long-term mean of past inflation rates. Basically, however, the problem of estimating α for nominal rates is exchanged for the problem of estimating α in real rates.

- Differences in local asymptotic long rates for two bond markets imply that long-term forward exchange rates explode[38]. The conclusion is that there exists a unique,

[38]The exchange rate Y^{ij} between countries i and j is given by the fraction of the respective state price densities

$$Y^{ij} := \frac{\varsigma_t^j}{\varsigma_t^i}. \tag{2.33}$$

Using covered interest arbitrage, we can derive that in an international bond market model based on the Rogers framework, the forward exchange rate is either exploding for one currency and going to zero for the other, or a unique international asymptotic long rate exists.
Without loss of generality we assume $\alpha^i > \alpha^j$. Then there exists a time of maturity T^* such that $y^i(t,T) > y^j(t,T)$ for all $T > T^*$. According to covered interest arbitrage, investing an amount of currency i over a time to maturity τ in the bond market of country i is equivalent to changing the amount of currency i into currency j at the current exchange rate Y_t^{ij}, investing it in the bond market of country j at $y^j(t,T)$ and taking a forward exchange rate contract which guarantees an $F^{ji}(t,T)$ to change back the amount of j plus compounded interest at time T. In formula

$$\begin{aligned} e^{y^i(t,T)(T-t)} &= Y_t^{ij} e^{y^j(t,T)(T-t)} / F^{ji}(t,T) \\ \exp\Big((y^i(t,T) - y^j(t,T))(T-t)\Big) &= \exp\Big((y^i(t,T) - y^j(t,T))(T-t)\Big) = \frac{Y_t^{ij}}{F^{ji}(t,T)} \\ F^{ji}(t,T) &= Y_t^{ij} \exp\Big(-(y^i(t,T) - y^j(t,T))(T-t)\Big). \end{aligned}$$

As there exists a T^* for which this inequality holds, all forward exchange rate curves are exponentially increasing. Obviously, this should not be the case assuming a long-term equilibrium in global trade. If $\alpha^i = \alpha^j$, for T large enough $y^i(t,T) \approx y^j(t,T)$ and hence the forward exchange rate is approximately the current exchange rate Y_t^{ij}.

world-wide asymptotic long rate. This allows to estimate multiple international term structures jointly with a single parameter α, thereby increasing the dataset substantially.

To conclude, we found both models to have significant problems in estimating the asymptotic long rate. Using multiple term structures adds additional data, yet even augmented datasets seem to be insufficient to estimate α.
Estimation using the original dataset produced estimates of the asymptotic long rate which differed by 600 basis points. Estimation using the extended dataset produced stable estimates, indicating that the previous results were due to censored data. However, whether or not 30 year rates and hence a small negative slope of -14.5 basis points were included in the extended dataset made the asymptotic long rate vary by 500 basis points. We have to acknowledge that empirical data seems to be insufficient to estimate a constant asymptotic long rate properly. Note that the same problem implicitly holds for other term structure models with a constant asymptotic long rate as well. As discussed previously, for most models a constant asymptotic long rate implies a restricting equation of the model parameters. Difficulties in specifying the asymptotic long rate imply that this restricting equation is not well defined. In the next section, we will discuss alternative approaches. Furthermore, our examinations showed dependencies of the level factor on the highest observable maturity. We recommend estimation both the Cairns and the cosh model with yield data up to 10 years of maturity only. This guarantees higher mean reversion of the level factor and avoids any impact of censored data on this factor.

Fit α to exogenous data

As with the long-term mean μ of the level component of the state process, we are not able to estimate α thoroughly from empirical data. Again as with μ, the parameter α has an economic interpretation as the asymptotic long rate which might be used to specify it exogenously. Note that in most models, such an exogenous specification is equivalent to introducing a restricting equation on model parameters. Consequently, changes in the asymptotic long rate for example in sensitivity analysis requires reestimation of the whole model. The asymptotic long rate being a distinct model parameter in the Cairns and cosh model provides a unique opportunity to examine the asymptotic long rate and its impact on prices of long-term assets.

Figure 2.17 shows that for fixed α, estimating the remaining parameters with historical data results in parameter sets which differ only slightly in Loglikelihood values and MAEs. Minimal and maximal Loglikelihood values differ by only 50, and minimal and maximal MAEs differ by less than 0.3 basis points. Due to approximations in the Kalman filter, such differences are not necessarily significant. For these results, we used the standard dataset including censored data, which is again reflected in both Loglikelihood values and MAEs.

Given our results in the previous section, this pattern should vanish using the extended dataset. In particular, we can assume that exogenous specifications of α and subsequent estimation of the remaining model parameters should provide good Loglikelihood values and MAEs. We can conclude that any exogenous specification of α provides a model with high historical fit measured in MAEs and Loglikelihood values.

A simple rule of thumb equates the asymptotic long rate with an average long rate. Within the underlying dataset, the average 10 year rate was 6.68%, reflecting both a good MAE and Loglikelihood value in our estimates with fixed α above. Such a high specification, however, implies that the model is unable to reproduce the twist for most historically observed levels of the term structure. Furthermore, the asymptotic long rate in this case is crucially dependent on the underlying dataset over which the average of the long-end yield is taken.
Using the expectations hypothesis (EH), we can derive an asymptotic long rate based on the same idea yet using short-end data. Namely, if we equate the asymptotic long rate with a long-term yield, we get

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} y(t_i, t_i+\tau) &\overset{EH}{=} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\tau}\sum_{k=0}^{\tau-1} E_t\left[r_{t_i+k}\right] + \phi(t_i, t_i+\tau)\right)\\ &= \frac{1}{\tau}\sum_{k=0}^{\tau-1}\left(\frac{1}{n}\sum_{i=1}^{n} E_t\left[r_{t_i+k}\right]\right) + \frac{1}{n}\sum_{i=1}^{n}\phi(t_i, t_i+\tau)\end{aligned}$$

where $\phi(t_i, \tau)$ denotes the term premium[39] at time t_i. For $n \to \infty$, this formula approximates the asymptotic long rate as an average over average conditional shortrate forecasts $E_t[r_{t_i+k}]$ for varying forecasting horizons. Of particular interest here is that shortrate forecasts depend crucially on the slope, whereas averaging over a single long-term interest rate depends predominantly on the level. The EH-based approach might therefore make better use of the information contained in historical yields. There remains, however, a significant problem in specifying the term premium, see for example [KO07].

As a second approach, we propose to use macroeconomic data to specify the asymptotic long rate. Obviously, a rational investor demands interest rates which at least compensate losses in real value of the notional. Assuming a successful monetary policy of the central bank with regards to inflation, we can therefore expect the inflation target $\bar{i}$ of the central bank to be a lower bound of the asymptotic long rate. It is a stylized fact that real interest rates are very persistent. In fact, as we will see in 3.2, the equilibrium real rate is often set as 2% or around 2%. This implies an asymptotic long rate of around 4%. An asymptotic long rate of 4.2% is currently discussed in regulatory boards as a possible regulatory specification[40].

[39] The yield premium in the notation of Kim and Orphanides [KO07].

[40] As we saw previously for the Rogers frameworks, the asymptotic long rate has a major impact on dis-

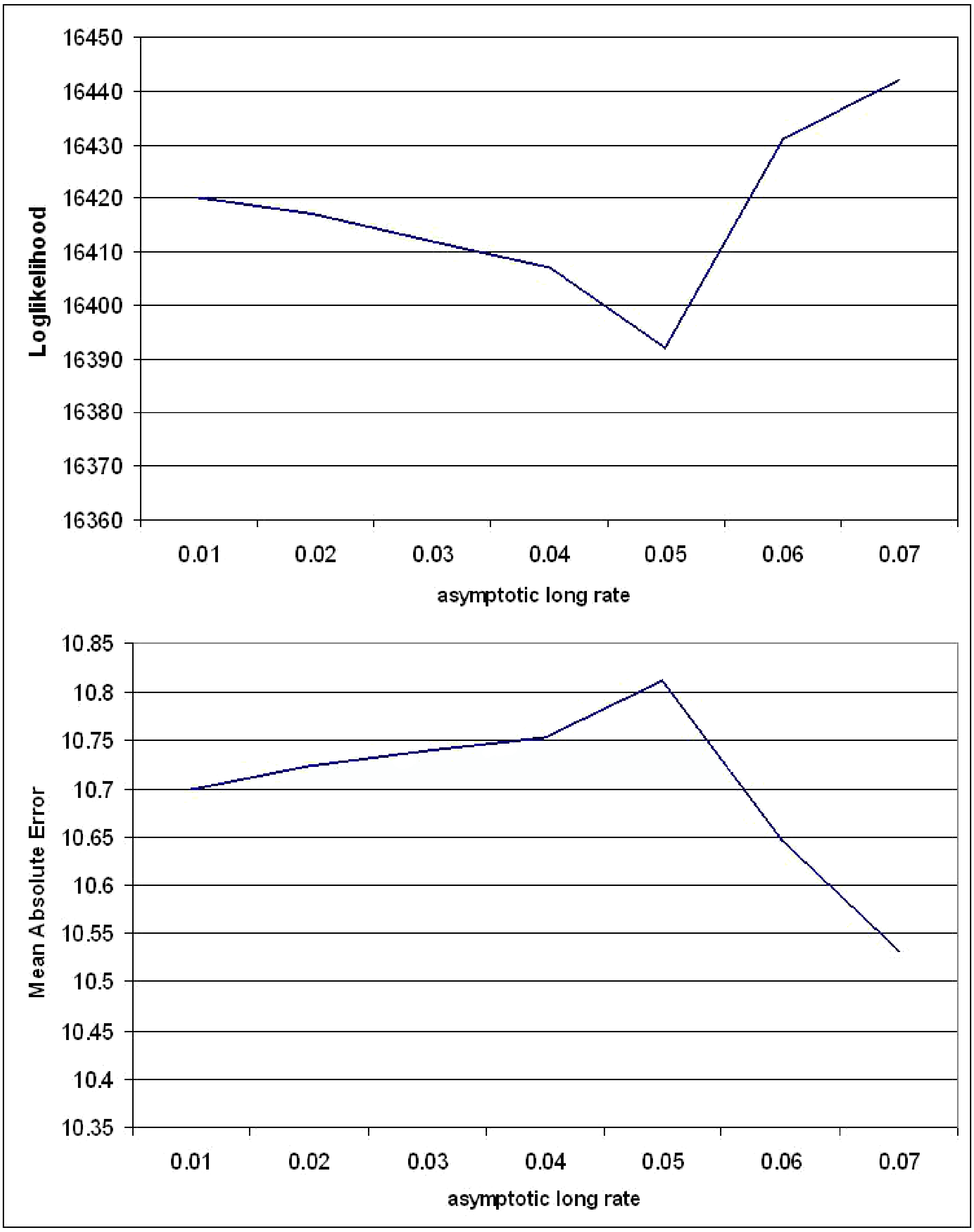

Figure 2.17: Implied Mean absolute errors and Loglikelihood values for reestimates of the cosh model with exogenously specified asymptotic long rate α.

Using the Taylor rule [Tay93] and the Expectations Hypothesis[41], we are able to derive a sound economic foundation for this approach. Particularly,

$$\begin{aligned} y(t,t+n)-\phi_n(t) &\overset{EH}{=} \frac{1}{n}\sum_{i=0}^{n-1} E_t[r_{t+i}] \\ &\overset{Taylor}{=} \frac{1}{n}\sum_{i=0}^{n-1} E_t[i_t + r_t^R + a_i(i_t-\bar{i}) + a_u(u_t)] \end{aligned}$$

where i_t denotes the inflation rate process with long-term mean $\bar{i}$, r_t^R is the real short rate and u_t denotes the deviation of output from its equilibrium trend growth. For n large enough, particularly $n > 30$ years, this approximates the asymptotic long rate. For $n \to \infty$, assuming successful monetary policy, the average deviation of inflation and output from their respective equilibrium values can be taken as zero, thus

$$\begin{aligned} 0 &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} E_t\left[i_{t+i}-\bar{i}\right] \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} E_t[i_{t+i}] - \bar{i} \end{aligned} \tag{2.34}$$

counting long-term cash flows, for example in pension insurance. Higher α reduces the liabilities insurance companies face, hence fomr a regulatory point of view an upper bound for α might have to be imposed for internal models of insurance companies.

[41] The Expectations Hypothesis states that long-term interest rates $y(t, t+n\Delta)$ for Δ being a specified time period equal average expected shortrates over periods of length Δ plus a constant risk premium ϕ_n which only depends on the tenor n.

$$y(t,t+n\Delta) = \frac{1}{n}\sum_{i=0}^{n-1} E_t[y(t+i\Delta, t+(i+1)\Delta)] + \phi_n$$

The Expectations Hypothesis therefore assumes that the shape of the term structure depends on market participants' expectations of future interest rates. Simple algebra shows that the Expectations Hypothesis implies that the current slope of the term structure forecasts future interest yields in a simple regression approach. However, repeating this regression with empirical data does not provide regression coefficients as implied by the Expectations Hypothesis. The fact that the Expectations Hypothesis does not hold with empirical data became a stylized fact in term structure models, hence term structure models are require to fail the Expectations Hypothesis in the same way as empirical data. According to Fama [Fam84] and Hardouvelis [Har88], the failure of the slope to forecast future interest rates is due to the omission from the regression of a time-varying risk premium, hence

$$y(t,t+n\Delta) = \frac{1}{n}\sum_{i=0}^{n-1} E_t[y(t+i\Delta, t+(i+1)\Delta)] + \phi_n(t).$$

The Taylor rule implies that the central bank sets current shortrates according to current inflation and output variables, hence

$$r_t = i_t + r_t^R + a_i(i_t - \bar{i}) + a_u u_t$$

where i_t is the current inflation rate, r_t^R is the real shortrate, $\bar{i}$ is the inflation target of the central bank and u_t describes deviation of output growth from its equilibrium trend.

and

$$0 \quad = \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} E_t[u_{t+i}].$$

This implies

$$\lim_{n\to\infty} (y(t,t+n) - \phi_n(t)) \quad = \quad \lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} E_t[i_{t+i}] + \frac{1}{n} \sum_{i=0}^{n-1} E_t[r^R_{t+i}] \right)$$

for n large enough. By (2.34), $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} E_t[i_{t+i}] = \bar{i}$. Historical real rates were more persistent than historical nominal rates. Given an inflation target of the ECB of $\bar{i} = 2\%$, an average historical real rate of about 2% and a positive term premium this approach implies $\alpha \geq 4\%$, which coincides with the suggestion of 4.2% in regulatory boards.

To conclude, we find equilibrium macro-based approaches to be most promising, as they provide a sound economic basis for specifications of the asymptotic long rate. Using the Expectations Hypothesis with time-varying risk premia and the Taylor rule, and assuming successful monetary policy by the central bank, we were able to derive an economically sound explanation for the regulatory suggestion of $\alpha = 4.2\%$. Using average long-term rates provides a simpler approach, which however crucially depends on the underlying data.

Note that all specifications of the asymptotic long rate proposed here were rather high. In particular, neither of these approaches will reproduce the twist of the term structure in the Rogers models we proposed.

Summary

Since the asymptotic long rate determines the discounting function for long time horizons, it is of major importance for long-term investors and in particular insurance companies. This may also be seen in the fact that a regulatory approach to specify a constant asymptotic long rate for insurance companies is considered.

We found that in the cosh and Cairns models, the parameter α equals the asymptotic long rate, whereas in many other term structure models the asymptotic long rate is a function of the model parameters and therefore constant in time as well. This implies a restricting equation in estimation of most term structure models if the asymptotic long rate is exogenously specified. Furthermore, this close relation between model parameters and long-term behavior of term structure models is frequently ignored in estimation. As such restricting equations do not exist for the cosh and Cairns models, these models are of particular interest for estimation of the asymptotic long rate and examination of sensitivities.

Testing the ability of the Kalman filter to estimate α, we found significant problems due to censored data and various empirical properties of the long-end of the yield curve, for example excess volatility, liquidity premia and the twist in the term structure. In general, we

have to acknowledge that observable long-term yields do not provide sufficient data to determine the asymptotic long rate, let alone riskless long-term interest rates. Consequently, the asymptotic long rate should be specified exogenously. As both the cosh and Cairns model allow for exogenous specification of the asymptotic long rate α without restricting the remaining model parameters, these models are of particular interest in analyzing such exogenous choices of the asymptotic long rate empirically.
We presented two approaches, one based on taking average long-term rates as a proxy for the asymptotic long rate, and a second one which derives the asymptotic long rate from the inflation target and average real rates. Using the second approach, we present a sound economic derivation of the recently discussed regulatory suggestion of $\alpha = 4.2\%$ based on the Expectations Hypothesis with time-varying term premia and the Taylor rule describing monetary policy.

2.4 Model Comparison

In the previous sections, we presented two realizations of the Rogers framework, the cosh and the Cairns model. Both models are state price density models. We found that the state price density approach is superior to standard risk-neutral pricing of long-term cash flows with infrequent and irregular payments are to be priced by simulation. In particular, the state price density approach does not require to approximate the shortrate by pathwise simulation, all that is required are the states of the underlying state vector at the respective payment dates. In both models dynamics are provided by a d-dimensional Ornstein-Uhlenbeck process. Implementation and estimation of the cosh model is a new contribution to the literature. We found two main differences between these models:

1. The choice of f in the Cairns model guarantees the state price density to be a supermartingale. This, in turn, guarantees no-arbitrage of the bond market model and positivity of all interest rates. However, it is difficult to specify a function f such that $e^{-\alpha t} f(X_t)$ is a supermartingale for a mean reverting state vector (X_t). The Rogers framework also allows for definition of models where the state price density is not a supermartingale, an example of which is the cosh model. Whereas we can prove that no-arbitrage holds for the cosh model, interest rates are not guaranteed to be positive.

2. The choice of f in the Cairns model requires numerical integration to derive bond prices. This makes the Cairns model computationally slow. On the other side, the choice of f in the cosh model guarantees analytic solutions to bond prices, making the cosh model much more efficient computationally.

The question of positivity in interest rates is of major interest for risk management in insurance companies. The worst case scenario for insurance companies is a Japan scenario, a persistent low and flat term structure, as we already discussed in section 2.1. We showed that Japan scenarios are a result of alternative monetary policy instruments in case the zero lower bound of the policy rate is reached. Similar dynamics may only apply in term structure models guaranteeing positivity. If negative interest rates are possible, the monetary transition mechanism of monetary policy implies that falling interest rates coincide with an increasing slope, hence very low long-end rates coincide with very steep yield curves and therefore a high probability of negative short-term rates. Higher mean reversion as implied if the model is estimated from the shortened dataset omitting censored 20- and 30-year rates should reduce the probability of negative interest rates in the cosh model if the long-term mean of the level process is sufficiently high.

Besides these differences, both models still share several important properties. First, we considered historical fit of both term structure models and found that both fit historical term structure data remarkably well, yet the Cairns model is superior considering historical fit. We also found that the state vector contains information about the principal components of the term structure in both models. In two-factor models, one state vector component coincides with the long-end of the yield curve, in particular the interest rate of the highest maturity in the estimation dataset. A second state vector component coincides with the slope. In a three-factor model, the level component remains, whereas the other two components describe slope and curvature. Furthermore, in both models the asymptotic long rate $\lim_{T\to\infty} y(t,T)$ equals the model parameter α. Together with our findings considering the level factor this implies similar dynamics of the long-end of the term structure in both models. In further examinations omitted here, we found that both models provide highly correlated term premia. This, in turn, implies that the LPY term structure criteria of Dai and Singleton [DS01] should hold in a similar way for both models. Finally, we also found that forecasting power of both term structure models is more or less equivalent. Forecasting errors of both models are highly correlated, an out-of-sample forecast conditional on the state at January 2008 provided very similar forecasts s well.

To summarize, we see that the Cairns and cosh models behave remarkably similar aside from computational speed and negative interest rates. The Cairns model is superior from its theoretical properties as well as considering historical fit. We also found improved slightly forecasting ability of the Cairns model. Nevertheless, the Cairns model is computationally slow. Since the cosh model shares several basic properties with the Cairns model, it can be used as a fast approximation of the Cairns model due to its computational simplicity and speed.

A sample for an application of the cosh model as an approximation of the Cairns model would be the analysis of the asymptotic long rate, as discussed in section 2.3.9. Both models allow for examination of various exogenous specifications of α using historical data or sensitivity analysis of long-term asset prices with respect to α. In this case, the Cairns model can not be recommended due to computational inefficiency, yet the results derived using the cosh model may be used to calibrate the Cairns model as well.

In the next sections, we will present another major advantage of the cosh model. Due to simplicity of the state price density in the cosh model, it allows to expand the underlying financial market to equity as well as macroeconomic data rather easily. It is important that the occurrence of Japan scenarios or low interest rates in general is closely linked to macroeconomic data and therefore realistic implementation of macroeconomic variables should help to govern the probability of negative interest rates in the cosh model. We will find that the methods proposed for expansion of the model apply to the Cairns model as well, yet typically are unfeasible computationally. On the other side, the cosh model allows for efficient implementation of these expanded models. Additional asset classes such as equity are required for insurance applications, and macroeconomic variables improve long-term interest rate dynamics and cross-asset correlations. The possibility of simple expansion implies that for many tasks in insurance applications, the cosh model should actually be recommended over the Cairns model.

Chapter 3

Additional Asset classes

In the framework presented so far we considered a financial market which only consisted of default-free bonds, interpreted as domestic government bonds. To model long-term insurance contracts over their lifetime, we must be able to simulate the investment policy of an insurance company. Now, obviously, insurance companies are not restricted in their investment choices to domestic government bonds, even though these bonds indeed form a large part of insurance companies' portfolios. To diversify their holdings, insurance companies will buy additional assets besides government bonds. For an example of such assets, Wilkie [Wil84], [Wil86] presents the consol yield, the stock price, the dividend yield and inflation as

> what seems to [...] be the minimum model that might be used to describe the total investments of a life office or pension fund.

In a related paper, Wilkie [Wil95] expanded his investment model by an earnings index, short-term interest rates, property rentals and prices and yields on index-linked stock. In particular, income or earnings indices may be used to describe changes in cancellation and underwriting, short-term interest rates are included to provide a full term structure and real estate variables provide yet another asset class.
The stock market as a first expansion to the government bond market is an obvious choice. Campbell and Ammer [CA93] show that correlation between stock and bond returns are low. They find that only real interest rate changes influence both stock and bond returns, but these are very persistent. Low correlation is important for diversification according to modern portfolio theory. Dividends are of major interest to insurance companies, as dividends provide a steady stream of cash flows which increase liquidity and may be used to match intermediate liabilities.
The consol rate is included as a measure of long-end interest rates. In Wilkie's actuarial model the consol yield is therefore used to describe the bond market with a specific long-end focus. Since insurance companies hold large bond portfolios of varying duration, a single interest rate is not sufficient to cover the insurance company's exposure to the bond

market. As we recommended previously, level and slope are a minimum requirement of term structure models.
Wilkie proposed his actuarial model in 1984 and 1986, respectively. The then recent experience of stagflation in the late 70s and record inflation during the monetary experiment 1979 to 1982 made inflation a major factor driving bond and stock markets. Since then central banks around the world kept inflation rates in check so that the importance of inflation for investment decisions of insurance companies and policy holders decreased. Due to the recent financial crisis, the surge in state deficits worldwide and specifically excess liquidity provided by central banks for the banking system may indicate rising inflation rates in the future, hence inflation may become a serious concern for insurance companies again.

A more theoretical reason to include additional assets in our model stems from the underlying assumption of a state price density. The assumed states depend on all investment choices possible within the economy. Therefore, state price densities naturally depend on *all* investment opportunities and hence pricing kernels should not be estimated using nominal bond data alone. An example for estimation of a pricing kernel from various financial assets may be found in Chernov [Che03].

In the following, we will present how to expand both the cosh and the Cairns term structure models to a joint actuarial model which describes the default-free bond market, stocks and their dividends.

3.1 Consistent stock market models

In this section, we try to establish a combined state factor model for term structure as well as stock price dynamics. We will first discuss two basic approaches to model stock market dynamics - the return-based and the price-based approach - as well as their respective properties. In a second step, we will examine restrictions on a joint bond and stock market model due to usage of the state price density approach. Finally, we will present several ways to define stock price dynamics under the cosh and Cairns models, respectively.

3.1.1 Stock prices or stock returns?

A major difference between bond and stock markets are the key figures used in daily practice to describe the current market situation. Bond markets are generally described by interest rates, whereas stock markets are described by current stock prices and historical returns. One major differences therefore lies in bond markets providing data about deterministic *future returns* and stock markets providing data about *historical returns*. In bond market modeling, we followed the standard approach and used interest rates for estimation and simulation. Given a series of historical returns and an initial stock price, the current price can be calculated. Given on the other side historical prices, historical returns may be calculated with similar ease. We will now discuss empirical properties of the respective approaches and then derive their relative advantages.

- **Explosion:** Unlike bond prices or interest rates, stock prices can "explode" in the sense that they may follow an exponential growth trend.
- **Positivity**: Stock prices are generally assumed to be positive. An exception might be default of the underlying firm, which implies the stock price being zero thereafter.
- **Dividend jump**: In case of discrete dividend payments, the stock price decreases at the payment date, since at time T^- the stock price S_{T^-} contains a claim on the dividend to be payed in T.
- **Leverage**: The absolute value of stock price changes depends on the underlying stock price. The higher the stock price, the higher on average absolute daily or monthly stock price changes.
- **Positive long-term equity risk premium**: It is a standard assumption that the equity risk premium, the excess expected return of stocks over bonds, is positive over the long term. This implies that there exists a positive drift term in stock prices. Obviously, this assumption should be employed on stock funds and not on individual stocks which may well be subject to default.
- **Default:** A single firm may default after some time, which implies that the stock price reaches zero and remains zero. If a fund of stocks with regular asset reallocation is considered, default should not occur and hence the value of the fund is always positive.

On the other side, stock returns share several properties of returns on bonds *not* held until maturity, for example:

- **Heavy Tails:** the distribution of stock returns is heavy-tailed[1] in the sense that stock returns are not normally distributed since they put more probability weight on extremal events.
- **Mean reversion:** Stock returns may be described as mean reverting, albeit with very high mean reversion factor and high volatility.
- **Negativity:** Stock returns can be negative.
- **Positive long-term equity risk premium:** A positive equity risk premium implies that stock returns are on average positive and higher than bond market returns, for example interest rates.

[1]The tails of heavy-tailed distributions are not exponentially bounded.

- **Volatility:** As mentioned previously, stock returns are highly volatile and indeed show stochastic volatility. Events of high volatility often coincide with significantly negative stock returns.

The main differences between these two approaches lies in applications. Whereas most interest rate derivatives depend on yields and hence bond returns, stock derivatives depend on stock prices, not returns. A price-based model which provides the distribution of $S_T|S_t$ is therefore vastly superior to a return-based approach for evaluation of stock derivatives.

A major obstacle for price-based models are dividend payments. In case dividend payments are assumed, the so called *total return* of a stock consists of the *price return* and the *dividend return* of the stock. We will now see that once dividend payments are considered, the return-based approach might be superior.
Implementation of total return models typically take a basic *reinvestment assumption* of dividend payments. It is typically assumed that the dividend is reinvested into the stock. Given a discrete-time dividend payment process (D_T) at time T, this implies that $\frac{D_T}{S_T}$ stocks are bought[2] at time T. Thus, assuming at time 0 one stock is held and dividends are payed in τ_i for $i = 1, \ldots$, the value process $(W_t)_{t\geq 0}$, which describes the complete wealth of an investor, is given by

$$
\begin{aligned}
W_0 &= S_0 \\
W_{\tau_1} &= S_{\tau_1} + D_{\tau_1} = \left(1 + \frac{D_{\tau_1}}{S_{\tau_1}}\right) S_{\tau_1} \\
W_{\tau_1+\Delta} &= \left(1 + \frac{D_{\tau_1}}{S_{\tau_1}}\right) S_{\tau_1+\Delta} \\
W_{\tau_2} &= \left(1 + \frac{D_{\tau_1}}{S_{\tau_1}}\right) (S_{\tau_2} + D_{\tau_2}) = \left(1 + \frac{D_{\tau_1}}{S_{\tau_1}}\right) \left(1 + \frac{D_{\tau_2}}{S_{\tau_2}}\right) S_{\tau_2} \\
&\vdots \quad .
\end{aligned}
$$

where $0 < \tau_1 < \tau_i + \Delta < \tau_2$. Assuming dividend payment dates $\tau_1, \ldots, \tau_n \in [0, T]$, and $W_0 = S_0$ we can generalize this to

$$
W_T = S_T \prod_{i=1}^{n} \left(1 + \frac{D_{\tau_i}}{S_{\tau_i}}\right). \tag{3.1}
$$

[2]We hereby assume that no taxes are payed on dividend income. Note that there exist total return indices with reinvestment assumptions which consider taxes of private investors. As institutional investors such like insurance companies are often tax-exempt on assets under management, this is a reasonable assumption.

The total return over $[0,T]$ with dividend payment dates $\tau_1,\dots,\tau_n \in [0,T]$, is therefore, again assuming $W_0 = S_0$, defined by

$$\begin{aligned}\log\left(\frac{W_T}{W_0}\right) &= \log\left(\frac{S_T\prod_{i=1}^n\left(1+\frac{D_{\tau_i}}{S_{\tau_i}}\right)}{S_t}\right)\\ &= \log\left(\frac{S_T}{S_t}\right)+\sum_{i=1}^n\log\left(1+\frac{D_{\tau_i}}{S_{\tau_i}}\right)\\ &= \log\left(\frac{S_T}{S_t}\right)+\sum_{i=1}^n\log\left(\frac{S_{\tau_i}+D_{\tau_i}}{S_{\tau_i}}\right)\end{aligned}$$

for $T \geq 0$, which can be interpreted as the sum of the dividend returns over $[0,T]$ and the price return over $[0,T]$. Without reinvestment, the total return over $[0,T]$ with dividend payments $D_{\tau_1},\dots,D_{\tau_n}$ and $\tau_1,\dots,\tau_n \in [0,T]$ is given by

$$\log\left(\frac{W_T}{W_0}\right) = \log\left(\frac{S_T+\sum_{i=1}^n D_{\tau_i}}{S_0}\right).$$

We see that for short time spans T, total returns without considering reinvestment assumptions are a possible alternative. Over long horizons, however, total returns without reinvestment assumption lead to underestimation of investment success. Total return with reinvestment assumption requires path dependent simulation with intermediate points $\{\tau_i : 0 \leq \tau_i \leq T\}$. The main advantage of the price-based approach therefore vanishes in case total return and the wealth process are of interest, such as in simulation of the portfolio success of a life insurance company.

Note that a simple solution to the dividend problem may be in transforming the data and reinterpreting the model. So called total return indices describe development of an investment into a stock market index whereby it is assumed that all dividends payed are instantly reinvested into the index[3]. Now if we consider a stock price model estimated on a total return index, then dynamics and in particular trend behavior estimated are those of total returns and hence implicitly incorporate dividend payments.

3.1.2 General considerations

Both the Cairns and the cosh model were specified in terms of a state price density model. The question arises how the stock price relates to the state price density framework. To implement a joint bond and stock market model dependent on a state process (X_t), we require first, following Rogers generic approach, a function $f : \mathcal{X} \to \mathbb{R}^+$ such that $(e^{-\alpha t} f(X_t))$ is a positive supermartingale[4]. The function f together with the dynamics of (X) under

[3] Note that there exist with- and without tax total return indices.

[4] Note, however, that the cosh model does not provide a positive supermartingale. We will discuss in this section the fundamental case of the potential approach of Rogers, which assumes the supermartingale property of the state price density.

the reference measure define the bond market model. Second, we require a stock price process (S_t) with $g : \mathcal{X} \to \mathbb{R}^+$ such that $S_t = g(X_t)$ for all $t \geq 0$. The following theorems provide further criteria on $(g(X_t))$ within the state price density framework conditional on dividend payments.

Theorem 3.1.1. *Given a state price density model with positive state price density process $(\varsigma_t)_{t\geq 0}$ under the reference measure $\tilde{\mathcal{P}}$ and a positive stock price process $(S_t)_{t\geq 0}$, then $(\varsigma_t S_t)$ is a positive supermartingale in an arbitrage-free market.*

Proof. By assumption, both the stock (S_t) and the state price density (ς_t) are positive processes. For any fixed future time T, we can define a contingent claim $\Pi_t(S_T)$ which pays the stock price S_T at time T. Using the general pricing formula under the reference measure, we get for the price of the contingent claim at time t

$$\Pi_t(S_T) \quad = \quad \frac{E^{\tilde{\mathcal{P}}}\left[S_T \varsigma_T | \mathcal{F}_t\right]}{\varsigma_t}$$

where the expectation is conditional on the filtration $\mathcal{F}_t = \sigma\{X_s, 0 \leq s \leq t\}$, the natural filtration of the state vector process (X_t). Now assume that at time t $\Pi_t(S_T) > S_t$. Then shorting the derivative and buying the stock provides a positive cash flow in t since $\Pi_t(S_T) - S_t > 0$. At time T, we can sell the stock so that both positions cancel out, since $\Pi_T(S_T) = S_T$. This offers an arbitrage strategy, hence

$$\Pi_t(S_T) \leq S_t \tag{3.2}$$

holds. As we are long in the stock, dividend payments do not change our conclusion as $S_t \leq S_t + D_\tau$ with non-negative dividend payments for all $\tau \in [t, T]$. □

Now if we additionally require the stock to pay no dividends, we can prove the following theorem.

Theorem 3.1.2. *Given a state price density model with positive state price density process $(\varsigma_t)_{t\geq 0}$ under the reference measure $\tilde{\mathcal{P}}$, and a stock with price process $(S_t)_{t\geq 0}$ which pays no dividend, then $(\varsigma_t S_t)$ is a martingale under the reference measure $\tilde{\mathcal{P}}$ in an arbitrage-free market.*

Proof. We assume $\Pi_t(S_T) < S_t$ and consider the following strategy:

- at time t, we short a stock S_t and buy the derivative $\Pi_t(S_T)$. By assumption, $S_t - \Pi_t(S_T) > 0$.
- at time T, the derivative pays S_T and we end our short position in the stock, hence $S_T - S_T = 0$.

Obviously, this is an arbitrage strategy, hence we can conclude that $\Pi_t(S_T) \geq S_t$ and hence $S_t\varsigma_t \leq E^{\tilde{\mathcal{P}}}\left[S_T\varsigma_T|\mathcal{F}_t\right]$. By theorem 3.1.1, $(\varsigma_t S_t)$ is a supermartingale whether dividends are payed or not, which implies that $(\varsigma_t S_t)$ is a martingale if the stock pays no dividends. □

The price of a stock which pays dividends can not be a martingale due to the dividend jump. Before the dividend payment date T, the fair stock price contains a discounted dividend similarly to the dirty price of a coupon bond containing "accrued interest" before coupon payment. Economically speaking, the firm value prior to dividend payment includes the dividend to be payed. In T, the stock price jumps due to the dividend payed out, hence discounted dividend value for the dividend jumps to zero or, economically, the firm value decreases by the dividend payed. Obviously, the dividend jump does not contradict the supermartingale property of $(S_t\varsigma_t)$.

Theorems 3.1.2 and 3.1.1 show that the stock price function g should guarantee that $(S_t\varsigma_t) = (e^{-\alpha t}f(X_t)g(X_t))$ is a positive supermartingale under the reference measure. If we additionally assume that no dividends are payed, $(S_t\varsigma_t) = (e^{-\alpha t}f(X_t)g(X_t))$ must be a martingale. Considering the problems we encountered in choosing a positive supermartingale $(e^{-\alpha t}f(X_t))$, we can expect the additional task of finding a simple function g such that $(e^{-\alpha t}f(X_t)g(X_t))$ is a positive supermartingale or a positive martingale to be even more challenging.
The question arises whether we can derive a viable stock market model without $(S_t\varsigma_t)$ being a supermartingale similar to $(e^{-\alpha t}\cosh(\gamma^T X_t + c))$ defining a viable bond market model without being a supermartingale. There are at least two further restrictions on the stock market model which are required to hold. These are no-arbitrage and positivity of the stock price. Now since the state price density ς_t defines the arbitrage-free bond market, the market price of risk is already specified. This allows to derive a partial differential equation which any definition of the stock price has to fulfill.
Let $S_t := g(X_t, t)$ for all $t \geq 0$. Then, by the Ito-Doeblin formula, the dynamics of the stock under the reference measure are given by

$$\begin{aligned} dg(X_t,t) &= \left[\frac{\partial}{\partial t}g(X_t,t) + \sum_{i=1}^{d}\kappa_i(\tilde{\mu}_i - X_t^{(i)})\frac{\partial}{\partial x_i}g(X_t,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}g(X_t,t)\right]dt \\ &\quad + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}g(X_t,t)\sum_{j=1}^{d}C_{ij}dZ_j^{\tilde{\mathcal{P}}}(t). \end{aligned}$$

Assuming a drift correction term $\Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)$ derived from the bond market with

$$dZ_i^{\mathcal{Q}}(t) = dZ_i^{\tilde{\mathcal{P}}}(t) + \Lambda_i^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)dt$$

implies

$$\begin{aligned} dg(X_t,t) &= \left[\frac{\partial}{\partial t}g(X_t,t) + \sum_{i=1}^{d}\kappa_i(\tilde{\mu}_i - X_t^{(i)})\frac{\partial}{\partial x_i}g(X_t,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}g(X_t,t)\right]dt \\ &\quad + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}g(X_t,t)\sum_{j=1}^{d}C_{ij}\left(dZ_j^{\mathcal{Q}}(t) - \Lambda_j^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)dt\right). \\ &= \left[\frac{\partial}{\partial t}g(X_t,t) + \sum_{i=1}^{d}\kappa_i(\tilde{\mu}_i - X_t^{(i)})\frac{\partial}{\partial x_i}g(X_t,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}g(X_t,t)\right. \\ &\quad \left. - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}g(X_t,t)\sum_{j=1}^{d}C_{ij}\Lambda_j^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)\right]dt + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}g(X_t,t)dZ_i^{\mathcal{Q}}(t). \end{aligned}$$

Now assuming $\Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t,t) = C^T\tilde{\Lambda}^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t,t)$ we get

$$\sum_{j=1}^{d}C_{ij}\Lambda_j^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t) = C\Lambda^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t) = CC^T\tilde{\Lambda}^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t) = \rho\tilde{\Lambda}^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t)$$

and hence under the risk-neutral measure,

$$\begin{aligned} r_t &\overset{!}{=} \frac{\partial}{\partial t}g(X_t,t) + \sum_{i=1}^{d}\kappa_i(\tilde{\mu}_i - X_t^{(i)})\frac{\partial}{\partial x_i}g(X_t,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}g(X_t,t) \\ &\quad - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}g(X_t,t)\tilde{\Lambda}_i^{\tilde{\mathcal{P}},\mathcal{Q}}(X_t). \end{aligned}$$

Which provides a partial differential equation the stock price function g has to fulfill given an arbitrage-free bond market model with shortrate r_t driven by a state vector X_t. $g(0, X_0) = s_0$ becomes a boundary condition, positivity of $g(X_t)$ is a further condition to be fulfilled. Given a certain bond price model within Rogers' framework, this partial differential equation provides a consistent stock price.

Nevertheless, simpler approaches to define a joint bond and stock market model may be derived. First, note that the previous assumption of stock prices defining martingales or supermartingales, the two theorems above yield the following corollaries.

Corollary 3.1.3. *Given a state price density model with state price density process* (ς_t) *under the reference measure* $\tilde{\mathcal{P}}$*, and a stock with positive price process* $(S_t)_{t\geq 0}$ *which pays no dividends, then, if no arbitrage is possible,*

$$S_t = \frac{E^{\tilde{\mathcal{P}}}\left[S_T\varsigma_T|\mathcal{F}_t\right]}{\varsigma_t}.$$

Corollary 3.1.4. *Given a state price density model with state price density process* (ς_t) *under the reference measure* $\tilde{\mathcal{P}}$*, and a stock with positive price process* $(S_t)_{t\geq 0}$ *which pays stochastic dividends at discrete times, then, if no arbitrage is possible,*

$$S_t \leq \frac{E^{\tilde{\mathcal{P}}}\left[S_T\varsigma_T|\mathcal{F}_t\right]}{\varsigma_t}.$$

Proof. Both corollaries are easily proved using theorems 3.1.2 and 3.1.1. □

Note that by construction, the derivative $\Pi_t(S_T)$ resembles a forward contract. The difference lies in the payment date. For the derivative, we have to pay its price $\Pi_t(S_T)$ at time t to earn S_T at time T, whereas we pay the current forward price F_t at time T to receive S_T in the forward contract. The following theorems relate the forward price to the price of the derivative $\Pi_t(S_T)$.

Theorem 3.1.5. *By no-arbitrage and with* $\Pi_t(S_T)$ *defined as in the proof of theorem 3.1.1,*

$$\Pi_t(S_T) = P(t,T)F_t,$$

where F_t *is the forward price at time* t *to buy a single unit of the stock* S *at time* T.

Proof. First, we assume an investor shorting the derivative $\Pi_t(S_T)$ and concluding a forward contract in t on the stock S with forward price F_t. Then we get the following payoffs

Time	t	T
Derivative	$+\Pi_t(S_T)$	$-S_T$
Forward Contract	0	$S_T - F_t$
Payoff	$+\Pi_t(S_T)$	$-F_t$

The price of the derivative in t must therefore be equal to the discounted forward price F_t which is $\mathcal{F}_t$ measurable, hence

$$\Pi_t(S_T) \quad = \quad \frac{E^{\tilde{\mathcal{P}}}\left[F_t\varsigma_T|\mathcal{F}_t\right]}{\varsigma_t} = F_t\frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T|\mathcal{F}_t\right]}{\varsigma_t} = F_t P(t,T).$$

□

We did not consider dividend payments explicitly in the proof as neither the forward contract nor the derivative pays dividends. Note, however, that both the fair derivative price and the fair forward price incorporate market expectations about future dividend payments. For a stock which pays no dividends,

$$F_t = \frac{S_t}{P(t,T)} \tag{3.3}$$

holds, see for example [Shr04] or [MR05]. This leads back to corollary 3.1.3 by

$$\Pi_t(S_T) \quad \overset{3.1.5}{=} \quad P(t,T)F_t \overset{(3.3)}{=} S_t.$$

Now assuming on the other side dividends are payed at discrete time, then the forward price is given by

$$F_t = \frac{S_t - I(t,T)}{P(t,T)} \tag{3.4}$$

where $I(t,T)$ is the discounted value of all dividend payments in $[t,T]$, see for example [Hul00]. Consequently,

$$\Pi_t(S_T) \overset{3.1.5}{=} P(t,T)F_t \overset{(3.4)}{=} S_t - I(t,T) \tag{3.5}$$

and hence

$$\Pi_t(S_T) \leq S_t$$

which leads back to corollary 3.1.4 as $I(t,T) \geq 0$. Writing (3.5) in terms of the state price density, we get

$$\begin{aligned} S_t &= \Pi_t(S_T) + I(t,T) \\ &= \frac{E^{\tilde{\mathcal{P}}[S_T\varsigma_T|\mathcal{F}_t]}}{\varsigma_t} + \sum_{i=1}^{\infty} \frac{E^{\tilde{\mathcal{P}}[D_{\tau_i}\varsigma_{\tau_i}|\mathcal{F}_t]}}{\varsigma_t} \end{aligned} \tag{3.6}$$

where $(\tau_i)_{i=1,\dots,n}$ is the series of dividend payment dates in $[t,T]$. Now for $T \to \infty$ discounted dividends $I(t,T)$ must be monotonically increasing, since dividend payments D_{τ_i} are assumed to be non-negative. If we reasonably assume $\lim_{T\to\infty} \Pi_t(S_T) = 0$, (3.6) motivates the dividend discount model discussed in the following section.

3.1.3 Dividend discount models

The *dividend discount model*[5] defines the stock price as the discounted sum of all future dividend payments. Equivalently, the current stock price equals the net present value of the future cash flows it promises. Note that the dividend discount model does not assume reinvestment of the dividends. As the stock price merely reflects the right to receive future dividends, the model does not specify what happens with dividends already payed. In the notation of the previous section, $S_t^{DDM} = I(t,\infty)$ for all $t \geq 0$.

Continuous dividend yield

In many models, assuming a continuous dividend yield allows for simple calculations, see for example [Shr04]. If we assume that a stock is an asset which pays a continuous dividend (δ_s), then following Di Graziano and Rogers [GR06], the stock price at time t in the state price framework is given by

$$S_t = \frac{E^{\tilde{\mathcal{P}}}\left[\int_t^\tau \varsigma_s \delta_s ds | \mathcal{F}_t\right]}{\varsigma_t}$$

where Di Graziano and Rogers assumed the (stochastic) upper integration bound τ to be the random time of default of the stock. To provide closed formula, this requires first a solution to the stochastic integral over the discounted dividend process $(\varsigma_s\delta_s)$ and

[5]Name and idea are taken from Gordon [Gor59].

second a closed formula for its expected value. Furthermore, this model does not allow to easily include historical dividend yields directly. In most cases, the model assuming discrete dividend payments described in the next section is easier to be implemented with historical data and easier to be interpreted.

Discrete dividend payments

Using the standard pricing formula of the reference measure for the discounted discrete dividend payment process $(D_{\tau_i})_{i\in\mathcal{I}}$ we get

$$S_t := \sum_{\tau_i \geq t, i\in\mathcal{I}} \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i} D_{\tau_i} | \mathcal{F}_t\right]}{\varsigma_t} \tag{3.7}$$

where $\tau_1, \tau_2, \ldots, i \in \mathcal{I}$, are the known dividend payment dates of the stock. With the stock price defined in this way we can easily check that $(S_t\varsigma_t)$ is a supermartingale under the reference measure.

Theorem 3.1.6. *With the stock price process $(S_t)_{t\geq 0}$ defined as in (3.7), $(\varsigma_t S_t)$ is a supermartingale under the reference measure $\tilde{\mathcal{P}}$.*

Proof. Let $T > t$ and $\{\tau_i : i \in \mathcal{I}\}$ be the set of all dividend payment dates, then

$$\begin{aligned} E^{\tilde{\mathcal{P}}}\left[S_T\varsigma_T|\mathcal{F}_t\right] &= E^{\tilde{\mathcal{P}}}\left[\sum_{\tau_i\geq T, i\in\mathcal{I}} \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i}D_{\tau_i}|\mathcal{F}_T\right]}{\varsigma_T}\varsigma_T \middle| \mathcal{F}_t\right] \\ &\overset{(*)}{=} \sum_{T_i\geq T, i\in\mathcal{I}} E^{\tilde{\mathcal{P}}}\left[E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i}D_{\tau_i}|\mathcal{F}_T\right]\middle| \mathcal{F}_t\right] \\ &= \frac{\sum_{T_i\geq T, i\in\mathcal{I}} E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i}D_{\tau_i}|\mathcal{F}_t\right]}{\varsigma_t}\varsigma_t \\ &\leq \frac{\sum_{T_i\in\mathcal{I}} E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i}D_{\tau_i}|\mathcal{F}_t\right]}{\varsigma_t}\varsigma_t \\ &= S_t\varsigma_t \end{aligned}$$

where $(*)$ uses the dominated convergence theorem, whereby the sequence of positive dividend payments is dominated by the stock price almost everywhere. □

This is in line with theorem 3.1.1, which requires the stock price to be a supermartingale in a state price density model. Theorem 3.1.2 implies that if no dividends are payed, $(\varsigma_t S_t)$ must be a martingale. Although so far the model did not require any reinvestment assumption, the dividends payed in $[t, T]$ make $(\varsigma_t S_t)$ a supermartingale.

Theorem 3.1.7. *With the stock price S_t defined as in (3.7) and $(W_t)_{t\geq 0}$ the wealth process with reinvestment of dividends into the money market account, $(\varsigma_t W_t)$ is a martingale under the reference measure $\tilde{\mathcal{P}}$.*

Proof. The wealth at time T consists of the stock hold at T plus dividend payments in $\tau_1, \ldots, \tau_n$, whereby we assume that $\{\tau_i : i \in \mathcal{I}\} \cap [t,T] = \{\tau_1, \ldots, \tau_n\}$. Furthermore, we assume all dividends payed to be reinvested in the money market account. At time T, the dividend payed at τ_i then values $\frac{\varsigma_{\tau_i}}{\varsigma_T} D(X_{\tau_i})$ for $i = 1, \ldots, n$. Hence,

$$
\begin{aligned}
E^{\tilde{\mathcal{P}}}\left[\varsigma_T W_T | \mathcal{F}_t\right] &= E^{\tilde{\mathcal{P}}}\left[\varsigma_T \left(S_T + \sum_{i=1}^{n} \frac{\varsigma_{\tau_i}}{\varsigma_T} D_{\tau_i} \right) \middle| \mathcal{F}_t \right] \\
&= E^{\tilde{\mathcal{P}}}\left[\varsigma_T \left(\frac{\sum_{\tau_i > T} E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i} D_{\tau_i} | \mathcal{F}_T\right]}{\varsigma_T} + \sum_{i=1}^{n} \frac{\varsigma_{\tau_i}}{\varsigma_T} D_{\tau_i} \right) \middle| \mathcal{F}_t \right] \\
&= E^{\tilde{\mathcal{P}}}\left[\sum_{\tau_i > T} E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i} D_{\tau_i} | \mathcal{F}_T\right] + \sum_{i=1}^{n} \varsigma_{\tau_i} D_{\tau_i} \middle| \mathcal{F}_t \right] \\
&= \frac{E^{\tilde{\mathcal{P}}}\left[\sum_{i \in \mathcal{I}} \varsigma_{\tau_i} D_{\tau_i} \middle| \mathcal{F}_t \right]}{\varsigma_t} \varsigma_t \\
&\overset{(*)}{=} S_t \varsigma_t = W_t \varsigma_t .
\end{aligned}
$$

where $(*)$ again uses the dominated convergence theorem. □

Note that the discrete dividend discount model provides *price returns* only. The pricing formula is based on the discounted value of the future cash flow received by the stockholder. To derive total returns based on the dividend discount model, dividends payed have to be reinvested according to the respective reinvestment assumption.

Implementation

For implementation, we have to specify the dividend payment process (D_t). We can rely on several stylized facts considering dividend payments as well as the criteria on the stock price of section 3.1.1.

- The dividend payments must be non-negative.
- Considering a single stock, the probability of dividend payments to become zero must be positive. Considering a well diversified stock portfolio, the probability of the whole portfolio paying no dividend or defaulting can be assumed as zero.
- The dividend payments must be a function of the underlying state vector $D_t := D(X_t)$ for all $t \geq 0$. This generalizes the Markovian state vector approach used in the bond market model for the joint bond and stock market model.
- We require closed form solutions of the expected values in (3.7). Note also that calculating stock prices based on the dividend discount model will be computationally more costly, as for zerobonds and hence interest rates only a single expected value must be calculated, whereas for the stock a series of expected values for each dividend payment date has to be calculated.

- We expect nominal dividend payments to increase in time. A simple economic explanation for this property of dividend payments is inflation. Dividend payments can be interpreted as a share in the issuing company's income. Assuming that nominal income rises with inflation, so do dividends.

- At the dividend payment date τ, the stock price should "jump", since at time τ^-, the stock price encompasses the future dividend payment D_τ, whereas this is not the case at time τ.

If we require that a single state process drives both the term structure and dividend payments, X must be mean reverting. Assuming a constant drift in dividends exists,

$$D_t = D(t, X_t) := \exp(\bar{\mu}t + \gamma^D X_t),$$

provides a reasonable model. In this framework, dividend payments follow a deterministic drift, whereas the state vector describes deviations from the drift over time. The exponential function guarantees positive dividend payments, the (positive) trend $\bar{\mu}$ guarantees growing dividend payments in nominal terms and the affine transformation $(\gamma^D)^T X_t$ implements dependency on the state vector process. The parameter $\bar{\mu}$ also governs the equity risk premium. If $\bar{\mu}$ is sufficiently high, stock investments are expected to produce higher returns over the long term than bond investments.

The Cairns model

It is sufficient to evaluate one expected value of formula (3.7)

$$\begin{aligned}
& E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right] \\
= \;& E^{\tilde{\mathcal{P}}}\left[e^{\bar{\mu}T + \sum_{i=1}^d \gamma_i^D X_T^{(i)}} \int_T^\infty \phi e^{-\alpha s + \sum_{i=1}^d \sigma_i e^{-\kappa_i(s-T)} X_T^{(i)} - \frac{1}{2}\sum_{i,j=1}^d \frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)}} ds \,\middle|\, X_t \right] \\
= \;& \phi E^{\tilde{\mathcal{P}}}\left[\int_T^\infty e^{\bar{\mu}T + \sum_{i=1}^d \gamma_i^D X_T^{(i)} - \alpha s + \sum_{i=1}^d \sigma_i e^{-\kappa_i(s-T)} X_T^{(i)} - \frac{1}{2}\sum_{i,j=1}^d \frac{\rho_{ij}\sigma_i\sigma_j}{\kappa_i+\kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)}} ds \,\middle|\, X_t \right]
\end{aligned}$$

As in the original Cairns model, we can interchange expected value and integration, getting

$$
\begin{aligned}
& \phi \int_T^\infty E^{\tilde{\mathcal{P}}} \left[\exp \left(\bar{\mu} T + \sum_{i=1}^d \gamma_i^D X_T^{(i)} - \alpha s + \sum_{i=1}^d \sigma_i e^{-\kappa_i(s-T)} X_T^{(i)} \right.\right. \\
& \left.\left. -\frac{1}{2} \sum_{i,j=1}^d \frac{\sigma_i \rho_{ij} \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)} \right) \middle| X_t \right] ds \\
= & \phi \int_T^\infty \exp \Big(-\alpha s + \bar{\mu} T + (\gamma^D + \sigma e^{-\kappa(s-T)})^T E^{\tilde{\mathcal{P}}} \left[X_T | X_t \right] \\
& + \left(\gamma^D + \sigma e^{-\kappa(s-T)} \right)^T \Sigma(t,T) \left(\gamma^D + \sigma e^{-\kappa(s-T)} \right) - \frac{1}{2} \sum_{i,j=1}^d \frac{\rho_{ij} \sigma_i \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)} \Big) ds \\
= & \phi \int_T^\infty \exp \Big(-\alpha s + \bar{\mu} T + \left(\gamma^D + \sigma e^{-\kappa(s-T)} \right)^T \left(e^{-\kappa(T-t)} X_t + (1 - e^{-\kappa(T-t)}) \tilde{\mu} \right) \\
& + \frac{1}{2} \sum_{i,j=1}^d \frac{\left(\gamma_i^D + \sigma_i e^{-\kappa_i(s-T)} \right) \rho_{ij} \left(\gamma_j^D + \sigma_j e^{-\kappa_j(s-T)} \right)}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(T-t)} \\
& - \frac{1}{2} \sum_{i,j=1}^d \frac{\rho_{ij} \sigma_i \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)} \Big) ds \\
= & \phi \exp(\bar{\mu} T) \int_T^\infty \exp \Big(-\alpha s + \sum_{i=1}^d \left(\gamma_i^D e^{-\kappa_i(T-t)} + \sigma_i e^{-\kappa_i(s-t)} \right) X_t^{(i)} \\
& + \frac{1}{2} \sum_{i,j=1}^d \frac{\left(\gamma_i^D + \sigma_i e^{-\kappa_i(s-T)} \right) \rho_{ij} \left(\gamma_j^D + \sigma_j e^{-\kappa_j(s-T)} \right)}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(T-t)} \\
& - \frac{1}{2} \sum_{i,j=1}^d \frac{\rho_{ij} \sigma_i \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)} \Big) ds \\
= & \phi \exp \left(\bar{\mu} T + \sum_{i=1}^d \gamma_i^D e^{-\kappa_i(T-t)} X_t^{(i)} + \sum_{i,j=1}^d \frac{\gamma_i^D \rho_{ij} \gamma_j^D}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(T-t)} \right) \int_T^\infty \exp \Big(-\alpha s \\
& + \sum_{i=1}^d \sigma_i e^{-\kappa_i(s-t)} X_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^d \frac{\sigma_i \rho_{ij} \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-t)} - \frac{1}{2} \sum_{i,j=1}^d \frac{\rho_{ij} \sigma_i \sigma_j}{\kappa_i + \kappa_j} e^{-(\kappa_i+\kappa_j)(s-T)} \\
& + \frac{1}{2} \sum_{i,j=1}^d \frac{\sigma_i \rho_{ij} \gamma_j^D}{\kappa_i + \kappa_j} e^{-\kappa_i(s-t)} e^{-\kappa_j(T-t)} + \frac{1}{2} \sum_{i,j=1}^d \frac{\gamma_i^D \rho_{ij} \sigma_j}{\kappa_i + \kappa_j} e^{-\kappa_i(T-t)} e^{-\kappa_j(s-t)} \Big) ds
\end{aligned}
$$

where we used that in the Cairns model $\tilde{\mu} = 0$. The stock price can then be derived by numerical integration similar to the evaluation of bond prices. A single evaluation of the stock price requires numerical integration for each dividend payment date T_i because of the last two sums in the integrand. Therefore, computational effort to derive the current price of a single dividend paying security is equivalent to deriving a single interest rate. As a large sum of dividend paying securities is required, this leaves the dividend discount model inapplicable for the Cairns model from computational reasons.

The cosh model

For the cosh model, our choice of D_T allows for closed form solutions, as the product of the two lognormally distributed variables D_T and ς_T is again lognormally distributed.

$$\begin{aligned}
& E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right] \\
= \;& E^{\tilde{\mathcal{P}}}\left[e^{-\alpha T}\cosh(c+\gamma^T X_T)\exp(\bar{\mu}T + (\gamma^D)^T X_T)|X_t\right] \\
= \;& \frac{1}{2}e^{-\alpha T}\Big[\exp(c+\bar{\mu}T)E^{\tilde{\mathcal{P}}}\left[\exp((\gamma^D+\gamma)^T X_T)|X_t\right] \\
& +\exp(-c+\bar{\mu}T)E^{\tilde{\mathcal{P}}}\left[\exp((\gamma^D-\gamma)^T X_T)|X_t\right]\Big] \\
= \;& \frac{1}{2}\exp(-\alpha T+\bar{\mu}T)\left[\exp\left(c+(\gamma^D+\gamma)^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]+\frac{1}{2}(\gamma^D+\gamma)^T\Sigma(t,T)(\gamma^D+\gamma)\right)\right. \\
& \left.+\exp\left(-c+(\gamma^D-\gamma)^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]+\frac{1}{2}(\gamma^D-\gamma)^T\Sigma(t,T)(\gamma^D-\gamma)\right)\right] \\
= \;& \exp\left(-\alpha T+\bar{\mu}T+(\gamma^D)^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]+\frac{1}{2}\left(\gamma^T\Sigma(t,T)\gamma+(\gamma^D)^T\Sigma(t,T)\gamma^D\right)\right) \\
& \cosh\left(c+\gamma^T E^{\tilde{\mathcal{P}}}\left[X_T|X_t\right]+\gamma^T\Sigma(t,T)\gamma^D\right) \qquad (3.8)
\end{aligned}$$

As the sum of discounted dividend payments in the cosh model is again a deterministic function of the current state factor X_t, calculation of current stock prices may be computationally easy. For implementation, we have to cut off the infinite sums of discounted dividend payments. Due to $\bar{\mu} > 0$, discounted values of long-term dividends may still provide significant value. A likely outcome is that the model overestimates dividend payments in the near future to make up for omitted dividend payments in the far future. If this is he case, joint estimation using stock price and dividend yield data will require a sufficient number of dividend payment dates considered, which likely requires considerable computational effort.

Change of measures

In the dividend discount model with discrete dividend payments at time T_i, the stock can be interpreted as a portfolio of infinitely many derivatives which, conditional on the future state X_T, pay $D_T(T, X_T)$ at time T. Thus, if the price $\Pi^{D_T}(t)$ at time t of such a "single dividend security" can be priced arbitrage-free within the cosh model, the stock can be priced arbitrage-free as well. We follow the standard approach used in proving no-arbitrage for the pure bond market model: first, we derive the dynamics of the dividend security under the reference measure, then we derive the required market price of risk to arrive at the risk-neutral measure, which has to fulfill the Novikov condition. Due to computational infeasibility of the Cairns model, we only consider the cosh model.
We first have to derive the dynamics of the single dividend security under the reference

measure $\tilde{\mathcal{P}}$ by

$$
\begin{aligned}
d\Pi^{D_T}(t) &= \left[\frac{\partial}{\partial t}\Pi^{D_T}(t) + \sum_{i=1}^{d}\kappa_j\left(\tilde{\mu}_i - X_t^{(i)}\right)\frac{\partial}{\partial x_i}\Pi^{D_T}(t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\Pi^{D_T}(t)\right]dt \\
&\quad + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Pi^{D_T}(t)\sum_{j=1}^{d}C_{ij}dZ_j^{\tilde{\mathcal{P}}}(t).
\end{aligned}
$$

We begin with the derivative in t.

$$
\begin{aligned}
\frac{\partial}{\partial t}\Pi^{D_T}(t) &= \frac{\partial}{\partial t}\frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right]}{\varsigma_t} \\
&= \frac{\frac{\partial}{\partial t}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right]}{\varsigma_t} - \frac{\frac{\partial}{\partial t}\varsigma_t}{\varsigma_t^2}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right].
\end{aligned}
$$

Now as a first step

$$
\begin{aligned}
&\frac{\partial}{\partial t}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T | X_t\right] \\
= &\frac{\partial}{\partial t}\exp((\bar{\mu}-\alpha)T)\exp\left((\gamma^D)^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + \frac{1}{2}\gamma^T\Sigma(t,T)\gamma + \frac{1}{2}(\gamma^D)^T\Sigma(t,T)\gamma^D\right) \\
&\cosh\left(\gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + \gamma^T\Sigma(t,T)\gamma^D + c\right) \\
= &\exp((\bar{\mu}-\alpha)T)\left[\frac{\partial f(t,T,X_t)}{\partial t}\exp(f(t,T,X_t))\cosh(g(t,T,X_t))\right. \\
&\left. + \frac{\partial g(t,T,X_t)}{\partial t}\exp(f(t,T,X_t))\sinh(g(t,T,X_t))\right] \\
= &E^{\tilde{\mathcal{P}}}\left[\varsigma_{T_i} D_{T_i} | X_t\right]\left(\frac{\partial f(t,T,X_t)}{\partial t} + \frac{\partial g(t,T,X_t)}{\partial t}\tanh(g(t,T,X_t))\right)
\end{aligned}
$$

and

$$
\begin{aligned}
f(t,T,X_t) &:= (\gamma^D)^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + \frac{1}{2}\gamma^T\Sigma(t,T)\gamma + \frac{1}{2}(\gamma^D)^T\Sigma(t,T)\gamma^D \\
g(t,T,X_t) &:= \gamma^T E^{\tilde{\mathcal{P}}}[X_T|X_t] + \gamma^T\Sigma(t,T)\gamma^D + c
\end{aligned}
$$

with

$$
\begin{aligned}
\frac{\partial}{\partial t}f(t,T,X_t) &:= (\gamma^D)^T\kappa e^{-\kappa(T-t)}(X_t - \tilde{\mu}) - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\
&\quad - \frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D \\
\frac{\partial}{\partial t}g(t,T,X_t) &:= \gamma^T\kappa e^{-\kappa(T-t)}(X_t - \tilde{\mu}) - \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D.
\end{aligned}
$$

Second, note that

$$
\frac{\partial}{\partial t}\varsigma_t = \frac{\partial}{\partial t}e^{-\alpha t}\cosh\left(\gamma^T X_t + c\right) = -\alpha\varsigma_t,
$$

hence

$$\begin{aligned}
\frac{\partial}{\partial t}\Pi^{D_T}(t) &= \frac{\frac{\partial}{\partial t}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right]}{\varsigma_t} - \frac{\frac{\partial}{\partial t}\varsigma_t}{\varsigma_t^2}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right] \\
&= \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right]\left(\frac{\partial f(t,T,X_t)}{\partial t} + \frac{\partial g(t,T,X_t)}{\partial t}\tanh(g(t,T,X_t))\right)}{\varsigma_t} + \alpha\frac{E^{\tilde{\mathcal{P}}}[\varsigma_T D_T|X_t]}{\varsigma_t} \\
&= \Pi^{D_T}(t)\left[\alpha + \frac{\partial f(t,T,X_t)}{\partial t} + \frac{\partial g(t,T,X_t)}{\partial t}\tanh(g(t,T,X_t))\right] \\
&= \Pi^{D_T}(t)\left[\alpha + (\gamma^D)^T\kappa e^{-\kappa(T-t)}(X_t-\tilde{\mu}) - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma\right. \\
&\quad -\frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D + \Big(\gamma^T\kappa e^{-\kappa(T-t)}(X_t-\tilde{\mu}) \\
&\quad \left.-\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D\Big)\tanh(g(t,T,X_t))\right].
\end{aligned}$$

Next we derive the derivative with respect to x_i by

$$\begin{aligned}
\frac{\partial}{\partial x_i}\Pi^{D_T}(t) &= \frac{\partial}{\partial x_i}\frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T D(X_T)|X_t\right]}{\varsigma_t} \\
&= \frac{\frac{\partial}{\partial x_i}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_{T_j}|X_t\right]}{\varsigma_t} - \frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_{T_j}|X_t\right]}{\varsigma_t}\frac{\frac{\partial}{\partial x_i}\varsigma_t}{\varsigma_t} \\
&= \frac{\frac{\partial}{\partial x_i}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right]}{\varsigma_t} - \Pi^{D_T}(t)\frac{\frac{\partial}{\partial x_i}\varsigma_t}{\varsigma_t}.
\end{aligned}$$

Again, we calculate first

$$\begin{aligned}
&\frac{\partial}{\partial x_i}E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right] \\
&= \exp((\bar{\mu}-\alpha)T)\frac{\partial}{\partial x_i}\exp(f(t,T,X_t))\cosh(g(t,T,X_t)) \\
&= \exp((\bar{\mu}-\alpha)T)\left[\left(\frac{\partial}{\partial x_i}f(t,T,X_t)\right)\exp(f(t,T,X_t))\cosh(g(t,T,X_t))\right. \\
&\quad \left.+\exp(f(t,T,X_t))\sinh(g(t,T,X_t))\frac{\partial}{\partial x_i}g(t,T,X_t)\right] \\
&= E^{\tilde{\mathcal{P}}}\left[\varsigma_T D_T|X_t\right]\left(\frac{\partial}{\partial x_i}f(t,T,X_t) + \tanh(g(t,T,X_t))\frac{\partial}{\partial x_i}g(t,T,X_t)\right)
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial}{\partial x_i}f(t,T,X_t) &= \gamma_i^D e^{-\kappa_i(T-t)} \\
\frac{\partial}{\partial x_i}g(t,T,X_t) &= \gamma_i e^{-\kappa_i(T-t)}.
\end{aligned}$$

Furthermore

$$\begin{aligned}
\frac{\frac{\partial}{\partial x_i}\varsigma_t}{\varsigma_t} &= \frac{e^{-\alpha t}\frac{\partial}{\partial x_i}\cosh(\gamma^T X_t + c)}{e^{-\alpha t}\cosh(\gamma^T X_t + c)} \\
&= \varsigma_t\gamma_i\tanh(\gamma^T X_t + c)
\end{aligned}$$

thus

$$\begin{aligned}
\frac{\partial}{\partial x_i}\Pi^{D_T}(t) &= \frac{\frac{\partial}{\partial x_i}E_t\left[\varsigma_{T_k} D_{T_k}\right]}{\varsigma_t} - \Pi^{D_T}(t)\frac{\frac{\partial}{\partial x_i}\varsigma_t}{\varsigma_t} \\
&= \Pi^{D_T}(t)\Big(\frac{\partial}{\partial x_i}f(t,T,X_t) + \tanh(g(t,T,X_t))\frac{\partial}{\partial x_i}g(t,T,X_t) \\
&\quad - \gamma_i \tanh(\gamma^T X_t + c)\Big) \\
&= \Pi^{D_T}(t)\Big(\gamma_i^D e^{-\kappa_i(T-t)} + \tanh(g(t,T,X_t))\gamma_i e^{-\kappa_i(T-t)} \\
&\quad - \gamma_i \tanh(\gamma^T X_t + c)\Big). \qquad (3.9)
\end{aligned}$$

For the Ito-Doeblin formula, we need

$$\begin{aligned}
\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Pi^{D_T}(t)\mu^i(t,X_t) &= \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Pi^{D_T}(t)\kappa_i(\tilde{\mu}_i - X_t^{(i)}) \\
&= \Pi^{D_T}(t)\Big((\gamma^D)^T e^{-\kappa(T-t)}\kappa(\tilde{\mu} - X_t) \\
&\quad + \tanh(g(t,T_k,X_t))\gamma^T e^{-\kappa(T-t)}\kappa(\tilde{\mu} - X_t) \\
&\quad - \tanh(f(t,T,X_t))\gamma^T \kappa(\tilde{\mu} - X_t)\Big).
\end{aligned}$$

This makes, using the shortrate formula,

$$\begin{aligned}
&\frac{\partial}{\partial t}\Pi^{D_T}(t) + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Pi^{D_T}(t)\mu^i(t,X_t) \\
= {}& \Pi^{D_T}(t)\Big[\alpha + (\gamma^D)^T\kappa e^{-\kappa(T-t)}(X_t - \tilde{\mu}) - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\
& - \frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D + \gamma^T\kappa e^{-\kappa(T-t)}(X_t - \tilde{\mu})\tanh(g(t,T,X_t)) \\
& - \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D \tanh(g(t,T,X_t)) \\
& + (\gamma^D)^T\kappa e^{-\kappa(T-t)}(\tilde{\mu} - X_t) + \gamma^T\kappa e^{-\kappa(T-t)}(\tilde{\mu} - X_t)\tanh(g(t,T_k,X_t)) \\
& - \gamma^T\kappa(\tilde{\mu} - X_t)\tanh(f(t,T,X_t))\Big] \\
= {}& \Pi^{D_T}(t)\Big[\alpha - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\
& - \frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D - \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D \tanh(g(t,T,X_t)) \\
& - \gamma^T\kappa(\tilde{\mu} - X_t)\tanh(f(t,T,X_t))\Big] \\
= {}& \Pi^{D_T}(t)\Big[r_t + \frac{1}{2}\gamma^T\rho\gamma - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\
& - \frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D - \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D \tanh(g(t,T,X_t))\Big].
\end{aligned}$$

Next we derive the second derivative

$$\begin{aligned}
\frac{\partial^2}{\partial x_i \partial x_j}\Pi^{D_T}(t) &= \frac{\partial}{\partial x_i}\Pi^{D_T}(t)\left(\gamma_j^D e^{-\kappa_j(T-t)} + \tanh(g(t,T,X_t))\gamma_j e^{-\kappa_j(T-t)}\right. \\
&\quad \left. - \gamma_j \tanh(\gamma^T X_t + c)\right) \\
&= \Pi^{D_T}(t)\left(\gamma_i^D e^{-\kappa_i(T-t)} + \tanh(g(t,T,X_t))\gamma_i e^{-\kappa_i(T-t)} - \gamma_i \tanh(\gamma^T X_t + c)\right) \\
&\quad \left(\gamma_j^D e^{-\kappa_j(T-t)} + \tanh(g(t,T,X_t))\gamma_j e^{-\kappa_j(T-t)} - \gamma_j \tanh(\gamma^T X_t + c)\right) \\
&\quad + \Pi^{D_T}(t)\left(\gamma_i\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)}(1 - \tanh^2(g(t,T,X_t)))\right. \\
&\quad \left. - \gamma_i\gamma_j(1 - \tanh^2(\gamma^T X_t + c))\right) \\
&= \Pi^{D_T}(t)\left(\gamma_i^D\gamma_j^D e^{-(\kappa_i+\kappa_j)(T-t)} + \gamma_i^D\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)}\tanh(g(t,T,X_t))\right. \\
&\quad - \gamma_i^D\gamma_j e^{-\kappa_i(T-t)}\tanh(\gamma^T X_t + c) + \gamma_i\gamma_j^D e^{-(\kappa_i+\kappa_j)(T-t)}\tanh(g(t,T,X_t)) \\
&\quad + \gamma_i\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)}\tanh^2(g(t,T,X_t)) \\
&\quad - \gamma_i\gamma_j e^{-\kappa_i(T-t)}\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) \\
&\quad - \gamma_i\gamma_j^D e^{-\kappa_j(T-t)}\tanh(\gamma^T X_t + c) \\
&\quad - \gamma_i\gamma_j e^{-\kappa_j(T-t)}\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) \\
&\quad + \gamma_i\gamma_j\tanh^2(\gamma^T X_t + c) + \gamma_i\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)} \\
&\quad \left. - \gamma_i\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)}\tanh^2(g(t,T,X_t)) - \gamma_i\gamma_j + \gamma_i\gamma_j\tanh^2(\gamma^T X_t + c)\right) \\
&= \Pi^{D_T}(t)\left(\gamma_i^D\gamma_j^D e^{-(\kappa_i+\kappa_j)(T-t)} + \gamma_i^D\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)}\tanh(g(t,T,X_t))\right. \\
&\quad - \gamma_i^D\gamma_j e^{-\kappa_i(T-t)}\tanh(\gamma^T X_t + c) + \gamma_i\gamma_j^D e^{-(\kappa_i+\kappa_j)(T-t)}\tanh(g(t,T,X_t)) \\
&\quad - \gamma_i\gamma_j^D e^{-\kappa_j(T-t)}\tanh(\gamma^T X_t + c) \\
&\quad - \gamma_i\gamma_j\left(e^{-\kappa_i(T-t)} + e^{-\kappa_j(T-t)}\right)\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) \\
&\quad + \gamma_i\gamma_j e^{-(\kappa_i+\kappa_j)(T-t)} \\
&\quad \left. - \gamma_i\gamma_j + 2\gamma_i\gamma_j\tanh^2(\gamma^T X_t + c)\right).
\end{aligned}$$

Which using the Ito-Doeblin formula is required in the form

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2 S_t}{\partial x_i \partial x_j} \\
= \; &\frac{1}{2}\Pi^{D_T}(t)\left((\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D + (\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma\tanh(g(t,T,X_t))\right. \\
&- (\gamma^D)^T e^{-\kappa(T-t)}\rho\gamma\tanh(\gamma^T X_t + c) + \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D\tanh(g(t,T,X_t)) \\
&- \gamma^T\rho e^{-\kappa(T-t)}\gamma^D\tanh(\gamma^T X_t + c) \\
&- 2\gamma^T\rho e^{-\kappa(T-t)}\gamma\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) + \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\
&\left. - \gamma^T\rho\gamma + 2\gamma^T\rho\gamma\tanh^2(\gamma^T X_t + c)\right),
\end{aligned}$$

whereby we used that due to symmetry of ρ and the diagonal matrix $e^{-\kappa(T-t)}$ we have

$$(e^{-\kappa(T-t)}\rho)^T = \rho^T(e^{-\kappa(T-t)})^T = \rho e^{-\kappa(T-t)}$$

and

$$\begin{aligned}\gamma^T e^{-\kappa(T-t)}\rho\gamma &= (\gamma^T e^{-\kappa(T-t)}\rho\gamma)^T = (e^{-\kappa(T-t)}\rho\gamma)^T\gamma \\ &= \gamma^T(e^{-\kappa(T-t)}\rho)^T\gamma = \gamma^T\rho e^{-\kappa(T-t)}\gamma,\end{aligned}$$

hence

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\frac{\partial^2\Pi^{D_T}(t)}{\partial x_i\partial x_j} \\ =& \frac{1}{2}\Pi^{D_T}(t)\Big((\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D \\ &+2\gamma^D e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^T\tanh(g(t,T,X_t)) \\ &-2(\gamma^D)^T e^{-\kappa(T-t)}\rho\gamma\tanh(\gamma^T X_t + c) \\ &-2\gamma^T\rho e^{-\kappa(T-t)}\gamma\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) + \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\ &-\gamma^T\rho\gamma + 2\gamma^T\rho\gamma\tanh^2(\gamma^T X_t + c)\Big).\end{aligned}$$

Now for the drift term of the Ito-Doeblin formula we get

$$\begin{aligned}&\Pi^{D_T}(t)\Big[r_t + \frac{1}{2}\gamma^T\rho\gamma - \frac{1}{2}\gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\ &-\frac{1}{2}(\gamma^D)^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D - \gamma^T e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D\tanh(g(t,T,X_t)) \\ &+\frac{1}{2}\gamma^D e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma^D + \gamma^D e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma\tanh(g(t,T,X_t)) \\ &-\gamma^D e^{-\kappa(T-t)}\rho\gamma\tanh(\gamma^T X_t + c) \\ &-\gamma\rho e^{-\kappa(T-t)}\gamma\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) + \frac{1}{2}\gamma e^{-\kappa(T-t)}\rho e^{-\kappa(T-t)}\gamma \\ &-\frac{1}{2}\gamma^T\rho\gamma + \gamma\rho\gamma\tanh^2(\gamma^T X_t + c)\Big] \\ =& \Pi^{D_T}(t)\Big[r_t - \gamma^D e^{-\kappa(T-t)}\rho\gamma\tanh(\gamma^T X_t + c) \\ &-\gamma^T\rho e^{-\kappa(T-t)}\gamma\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) \\ &+\gamma\rho\gamma\tanh^2(\gamma^T X_t + c)\Big]. \qquad (3.10)\end{aligned}$$

To derive the drift correction term $\Lambda_\Pi(t, X_t)$, we follow the same approach as in the bond market. Again, for simplicity, we define $\Lambda_\Pi(t,T,X_t) := C^T\Lambda'_\Pi(t,T,X_t)$ thus we require

$$dZ_t^{\tilde{\mathcal{P}}} = dZ_t^{\mathcal{Q}} + \Lambda_\Pi(t,T,X_t)dt.$$

and hence with $CC^T = \rho$

$$CdZ_t^{\mathcal{Q}} = CdZ_t^{\tilde{\mathcal{P}}} - \rho\Lambda'_\Pi(t,T,X_t)dt.$$

Which leads us to

$$\begin{aligned}
& \Pi^{D_T}(t)\Big((\gamma^D)^T e^{-\kappa(T-t)}\rho\gamma\tanh(\gamma^T X_t + c) \\
& +\gamma^T\rho e^{-\kappa(T-t)}\gamma\tanh(\gamma^T X_t + c)\tanh(g(t,T,X_t)) \\
& -\gamma\rho\gamma\tanh^2(\gamma^T X_t + c)\Big) \\
= & \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Pi^{D_T}(t)\sum_{j=1}^{d}\rho_{ij}\Lambda_{\Pi}^{(i)}(t,T,X_t) \\
= & \Pi^{D_T}(t)\Big((\gamma^D)^T e^{-\kappa(T-t)}\rho\Lambda'_{\Pi}(t,T,X_t) + \tanh(g(t,T,X_t))\gamma^T e^{-\kappa(T-t)}\rho\Lambda'_{\Pi}(t,T,X_t) \\
& -\tanh(\gamma^T X_t + c)\gamma^T\rho\Lambda'_{\Pi}(t,T,X_t)\Big).
\end{aligned}$$

This implies

$$\Lambda^{\Pi}(t,T,X_t) = \Lambda(X_t) = C^T\gamma\tanh(\gamma^T X_t + c),$$

which in turn yields the following theorem.

Theorem 3.1.8. *For the cosh model with state price density process (ς_t) with*

$$\varsigma_t = e^{-\alpha t}\cosh(\gamma^T X_t + c),$$

for all $t \geq 0$ and (X_t) an Ornstein-Uhlenbeck process with dynamics (2.12) and (2.13) and the stock price process (S_t) defined as the infinite sum of discounted dividends with payoff at time T

$$D(T, X_T) = \exp\left(\bar{\mu}T + (\gamma^D)^T X_T\right), \tag{3.11}$$

the joined bond and stock market is arbitrage-free.

Proof. We constructed a measure, under which the price dynamics of a security which pays a stochastic dividend $D(T, X_T)$ at time T have a drift which equals the shortrate. The resulting drift correction term is the same as in case of the bond market. Therefore, first, the Novikov condition holds, making the measure an equivalent measure. Second, as the drift equals the shortrate, it is a risk-neutral measure. As it is the same risk-neutral measure we derived for the bond market, the market consisting of all dividend paying securities $\Pi^D(T)$, $T \geq 0$, bonds and the bank account is arbitrage-free. As a consequence, the stock as a portfolio of dividend paying securities is priced arbitrage-free as well. □

The market price of risk is already given by the bond market, see theorem 2.2.25.

Estimation

We want to estimate the joint bond and stock market model using the Extended Kalman filter. We implement the second dividend approach $D_\tau = \exp\left(\bar{\mu}\tau + \gamma^T X_\tau\right)$ whereby (X_t) is a multi-dimensional Ornstein-Uhlenbeck process. The transition equations of the Cairns and cosh models hold in the expanded model as well. What is left is to specify the measurement equation. For the cosh model, the stock price is given as a portfolio of securities which pay a dividend at time τ_i, $i \in \mathcal{I}$, hence

$$\begin{aligned} S_t &= \frac{\sum_{\tau_i > t, i \in \mathcal{I}} E^{\tilde{\mathcal{P}}}\left[\varsigma_{\tau_i} D_{\tau_i} | X_t\right]}{\varsigma_t} \\ &= \sum_{\tau_i > t, i \in \mathcal{I}} \frac{1}{2\varsigma_t} \exp(-\alpha\tau_i + \bar{\mu}\tau_i) \Big[\exp\Big(c + (\gamma^D + \gamma)^T E^{\tilde{\mathcal{P}}}\left[X_{\tau_i} | X_t\right] \\ &\quad + \frac{1}{2}(\gamma^D + \gamma)^T \Sigma(t, \tau_i)(\gamma^D + \gamma)\Big) \\ &\quad + \exp\Big(-c + (\gamma^D - \gamma)^T E^{\tilde{\mathcal{P}}}\left[X_{\tau_i} | X_t\right] + \frac{1}{2}(\gamma^D - \gamma)^T \Sigma(t, \tau_i)(\gamma^D - \gamma)\Big)\Big] \end{aligned}$$

this can be used directly to derive the measurement equation.

$$\begin{pmatrix} y^M(t, t+\tau_1) \\ \vdots \\ y^M(t, t+\tau_n) \\ S_t^M \end{pmatrix} = \begin{pmatrix} g_1(X_t; \theta) \\ \vdots \\ g_n(X_t; \theta) \\ g_{n+1}(X_t; \theta) \end{pmatrix} + \epsilon_t(\theta)$$

whereby $\epsilon_t(\theta) \in \mathbb{R}^{n+1}$ is a multivariate normal error term with $Cov\left(\epsilon_t(\theta)\right) := H_t(\theta) \in \mathbb{R}^{(n+1)\times(n+1)}$ and $y^M(t, t+\tau_i)$, $i = 1, \ldots, n$ and S_t^M are market observations. Unlike the pure term structure model, however, we recommend to distinguish between the measurement errors of interest rates on one side and the measurement errors of the stock market on the other side. This implies

$$H_t(\theta) \quad := \quad diag\left(\nu, \ldots, \nu, \nu^S\right).$$

Next, we require the matrix $B_{t|t-1}$ to derive the Kalman gain matrix

$$K_t \quad = \quad \Sigma_{t|t-1} B_{t|t-1}^T F_{t|t-1}^{-1}.$$

The derivatives of $g_1(X_t; \theta), \ldots, g_n(X_t; \theta)$ may be taken from equation (2.25) in section 2.3.5. With (3.9),

$$\begin{aligned} &\frac{\partial}{\partial x_i} S_t \\ = \; &\frac{\partial}{\partial x_i} \frac{\sum_{\tau_i > t, i \in \mathcal{I}} \Pi^{D_{\tau_i}}(t)}{\varsigma_t} \\ = \; &\sum_{i=1}^{\infty} \Pi^{D_{\tau_i}}(t) \left(\gamma_i^D e^{-\kappa_i(\tau_i - t)} + \tanh(g(t, \tau_i, X_t))\gamma_i e^{-\kappa_i(\tau_i - t)} - \gamma_i \tanh(\gamma^T X_t + c)\right) \end{aligned}$$

This yields the matrix $B_{t|t-1}$. With $F_{t|t-1} = B_{t|t-1}\Sigma_{t|t-1}B_{t|t-1}^T + H_t$, the Kalman gain matrix can be calculated and hence the Kalman filter is completely specified. We face, however, several implementational problems.

- We have to cut off the infinite sum of dividend paying securities. As the stock price is a sum of discounted dividend payments, even for lower interest rates an additional $n+1$-th dividend payment increases the stock price, yet only marginally in comparison to the first n discounted dividends, hence cutting off additional dividend payments should be valid.
  Note that this may change if we additionally use actual dividend data. In this case, historically observed dividend payments have to be considered both in height and frequency. As dividend payments are rather low, yet increasing in nominal terms, a high number of dividend payment dates will have to be included.
  In general, as mentioned before, we expect that cutting off dividend payments of later dates imply that expected dividends payed early will be overestimated to make up for the later dividends.

- Whereas calibrating $X_{0|0}$ to fit the initial yield curve was fairly easy and computationally fast, calibration to S_0 is computationally slow and often inaccurate. Another frequent finding using the Kalman filter was that the derivative of the stock price dominated the derivative of the yields, therefore correction was due to stock market misfit only. We therefore propose to introduce a scaling factor d to the dividends

$$D_T = \exp\left(d + \bar{\mu}T + (\gamma^D)^T X_T\right).$$

  This implies that $\exp(d)$ is a multiplicative factor to the stock price, but also to the derivatives of the stock price with respect to x_i. Therefore, d might allow to downscale the impact of the stock price derivatives in the updating step. We either can specify $\exp(d) = S_0$ directly, or we estimate d to incorporate the impact of $X_{0|0}$ fitted to term structure data. Furthermore, d may scale down stock price data. Indeed, stability of the Kalman filter increased once d was estimated alongside the remaining model parameters.

- For simplicity in estimation and due to data availability, we only used price data. One could additionally use dividend data, including dividend forecasts. We can expect that such an approach would determine our cut-off level of dividends as well as improve our estimate of $\bar{\mu}$.

For estimation, we use $S\&P500$ price index data. This reflects well the assumption that the insurance company invests in a well diversified stock fund rather than a single stock. Furthermore, $S\&P500$ data shows a growing trend over the dataset from 1984 to 2009, with deviations from the trend which grow in absolute terms as the index increases, reflecting leverage. As we assumed a constant long-term trend $\bar{\mu}$, $S\&P500$ data should fit well into our approach and furthermore might allow to derive empirical proxies for $\bar{\mu}$. Finally, note that anglo-saxon stock markets typically imply semiannual or even quarterly dividend payments, which reduces the dividend jump in the data.

Maturity	0.25	0.5	1	2	3	5	7	10
MAEs DDM	15.5	7.1	11.1	13.8	11.4	6.7	9.8	16.2
MAEs two-factor pure bond cosh	16.7	6.8	8.9	16.3	13.8	8.1	6.6	10.4

Table 3.1: Mean Absolute Errors in basis points of the term structure in the cosh model augmented by a dividend discount stock market model, and MAEs of the two-factor pure-bond cosh model for comparison.

Results

The model was implemented using only stock price data, omitting historical dividend payment date. Therefore, we had to make assumptions regarding the dividend payment dates. In a first implementation, we assumed that all dividends are payed in December. Under this assumption, model-implied stock prices showed vastly overestimated dividend jumps in December. Assuming more frequent dividend payments naturally reduced the dividend jumps. Nevertheless, without using dividend data directly and specifying a cut-off level describing the maximal maturities of discounted dividends included, the easiest way to implement a dividend discount model was assuming monthly dividend payments. Note that for stock market indices or stock funds, the assumption of monthly dividend payments might actually be realistic.

Due to implementational difficulties described above and computational limits, we were only able to derive a single reasonable estimate of the dividend discount model. Whereas we found it fairly easy to derive a model which fits well the term structure and provides model-implied stock prices highly correlated to the true stock price, it is fairly difficult to derive a model which fits stock prices in absolute terms. The main problem seems to be initial fit. In most cases, the initial stock price was overestimated and as a consequence, the dividend drift $\bar{\mu}$ was rather low.

The model is given by $\alpha = 0.051$, $\gamma = (-0.47, -0.18, 0.01)^T$ and $\gamma^D = (-1.28, 1.94, 0.69)^T$, $\mu = (1.9, -12.5, -10.8)^T$, $\tilde{\mu} = (-4.9, 5.7, -12.9)^T$, $\bar{\mu} = -0.0023$, $\kappa = (0.05, 0.062, 0.7)^T$, $\rho_{12} = 0.34$, $\rho_{13} = -0.04$, $\rho_{23} = 0.1$ and the scaling factor $c = -14.3$. Measurement errors for the term structure were $\nu = 0.0014$ and for the stock price data $\nu^S = 0.0082$, which already indicates a decent historical fit. Note that generally the estimates provided very low measurement errors for the term structure data, yet substantially higher measurement errors for stock data ranging up to 0.05.
Table 3.1 provides mean absolute errors in basis points for the term structure. On average, historical term structure fit of the three-factor joint bond and stock market model is half a basis point worse than historical fit of the two-factor pure bond market model.

Considering the stock market, the mean absolute error of the $S\&P500$ index is at merely 20 ticks. Figure 3.1 shows both the model implied and the observed stock prices. We see that the calibration algorithm provided a starting point $X_{0|0}$ which significantly

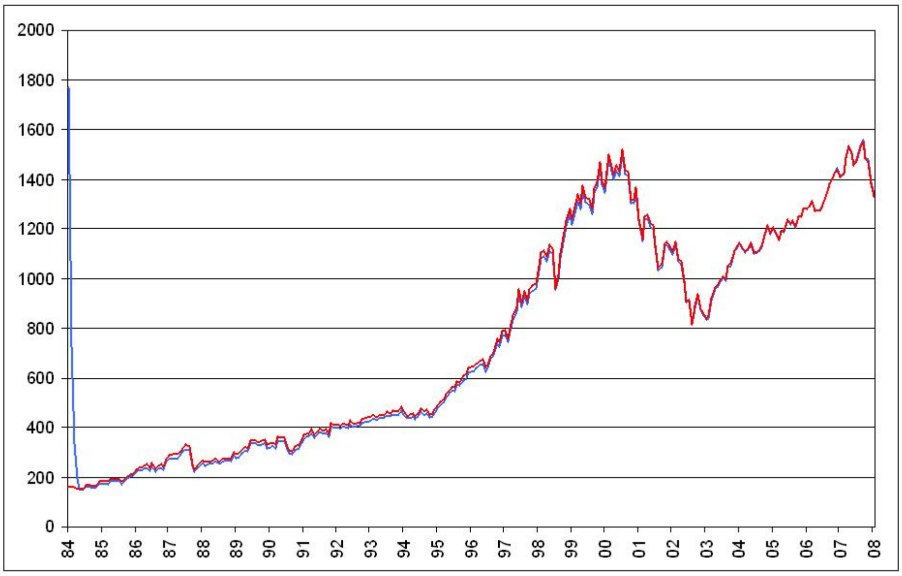

Figure 3.1: Filtered state vector components (left) and empirical proxies of the first two principal components of term structure dynamics.

overestimated the stock price. However, the Kalman filter corrected this initial overestimation over the first 4 steps. Excluding these, the mean absolute error reduces to 12 ticks. This is an extremely close historical fit over 28 years of stock price data. Note in particular that the model fits well the 1987 stock market crash, the 1998 turmoil, the bust of the dot-com-bubble and at least the starting of the recent financial crisis. As the Kalman filter calculates a term added to the Loglikelihood function at each time step, excluding the contribution of the first four steps might improve estimation.

We examine the filtered underlying state vector in figure 3.2. We find clear correlation between one of the state factor components and the slope and a second state factor component and the stock price. The third state factor is correlated to the 10-year rate, yet it does not cover the trend behavior so far encountered in the level factor. One possibility is that trend behavior is covered by the stock price factor. Both the dynamics of the level and the stock price factor provide very small mean reversion, as could be seen in our estimates of κ_2 and κ_3.

Note that the long-term trend in dividend payments $\bar{\mu}$ is slightly negative. This does not necessarily imply that stock prices show a slightly negative trend as well. Furthermore, estimates of $\bar{\mu}$ were highly unstable in general. As the long-end yield shows a falling trend throughout the dataset, causing difficulties in deriving stable estimates of μ^L as seen in section 2.3.8, even without a positive trend in dividend payments falling long-end yields imply decreased discounting of later dividends and therefore a rising trend in stock prices due to the dividend discount model.

Note also that estimates of α were very unstable, even though we excluded long-end term structure data and therefore expected stable estimates of $\alpha \approx 4.5\%$. Given how α enters

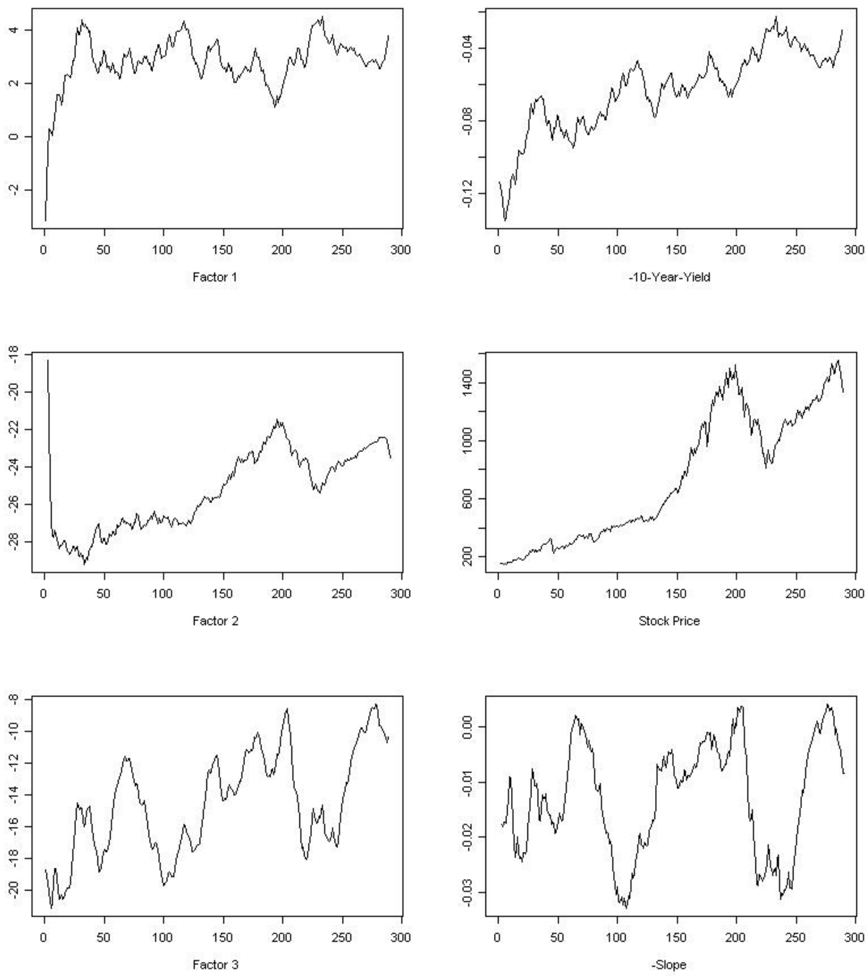

Figure 3.2: Filtered state vector components (left) and empirical proxies of the first two principal components of term structure dynamics.

the stock price model in (3.8) it is reasonable to assume that the problems of unstable α and $\bar{\mu}$ are closely related. We expect that exogenous specification of both α and $\bar{\mu}$ should provide more stable results. Note that the difference of α and $\bar{\mu}$ should be closely related to a long-term equity risk premium.

To summarize, we found that estimation of the dividend discount approach still requires further work, particularly considering the inclusion of historical dividend data and the implementation of historical dividend payment dates including the examination of dividend jumps. Note also that historically observed dividends and dividend forecasts are crucial in specifying a cut-off level for the dividend payment dates considered. In particular, we recommend using historical dividend data for estimation, and we recommend exogenous specification of both the asymptotic long rate α and the long-term dividend growth trend $\bar{\mu}$. Furthermore, parallelized computation is recommendable. Nevertheless, these preliminary results already show that the dividend discount model can provide joint bond and stock

market models which fit historical data extremely well.
Note also that the dividend discount model provides a stock price formula which only depends on the current state X_t. This simplifies Monte Carlo-based pricing of stock options under interest rate risk substantially. In this case, the current price of a call on the stock S_t is given by

$$\frac{E^{\tilde{\mathcal{P}}}\left[\varsigma_T (S_T - K)^+ | X_t\right]}{\varsigma_t}$$

It may easily be simulated as only the conditional distribution of $X_T|X_t$ is required. As stock options with very long times to maturity therefore depend jointly on the stock model and the state price density, it might be interesting to include these options into estimation of the joint model.

3.1.4 The Black-Scholes stock market model

In a basic Black-Scholes model, stock price dynamics under the risk-neutral measure are given by

$$dS_t = rS_t dt + \sigma S_t dW_t^{\mathcal{Q}}$$

where the shortrate r is assumed as constant, see for example [MR05] or [Shr04]. Now for the joint bond and stock market model, we have to use the stochastic shortrate process (r_t) provided by the arbitrage-free bond market model, hence under the the riskless measure

$$dS_t = r_t S_t dt + \sigma S_t dW_t^{\mathcal{Q}},$$

see for example [Shr04]. Using the market price of risk as defined by the bond market model, we can derive stock market dynamics under the reference or the physical measure. As the stock price is defined under the risk-neutral measure, no-arbitrage holds for the stock market. Closed form solutions to the stock price can be derived as solutions to the stochastic differential equation of the stock price above. Alternatively, using the Euler-Maruyama scheme [KP99], an iterative approximation may be used.
An example for this approach may be found in Albrecht [Alb07], where a one-factor Vasicek [Vas77] model is used for the shortrate and a Black-Scholes model for the stock dynamics. The Vasicek model provides the shortrate, which is used as the drift in a Black-Scholes model of the stock price. A two-dimensional correlated Brownian motion provides the stochastic driver of the model, whereby the shortrate depends on one component and the stock depends on both, although this correlation assumption could be generalized. Contingent claims dependent on both bond and stock market instruments can then be priced using the standard formula under the risk-neutral measure.
We will demonstrate this approach using the cosh model, for which the shortrate is given by

$$r_t \quad = \quad \alpha - \gamma^T \kappa(\tilde{\mu} - X_t) \tanh(\gamma^t X_t + c) - \frac{1}{2}\gamma^T \rho \gamma.$$

Therefore, the Black-Scholes model implies a stock price

$$\begin{aligned} dS_t &= r_t S_t dt + S_t \sigma^T C dZ_t^{\mathcal{Q}} \\ &= \left(\alpha - \gamma^T \kappa(\tilde{\mu} - X_t)\tanh(\gamma^t X_t + c) - \frac{1}{2}\gamma^T \rho \gamma\right) S_t dt + S_t \sigma^T C dZ_t^{\mathcal{Q}} \end{aligned} \tag{3.12}$$

where we used that $dW_t^{\mathcal{Q}} = CdZ_t^{\mathcal{Q}}$ with $CC^T = \rho$ as defined previously. Solving this equation for S_t provides us with a closed-form solution for the stock price. To do so, we use the following theorem.

Theorem 3.1.9. *Let $Z = \left(Z^{(1)}, \ldots, Z^{(d)}\right)$ be a d-dimensional Brownian motion. Furthermore, let $s_0 \in \mathbb{R}$ and μ, ν, λ_j, σ_j, $j = 1, \ldots, .d$ be progressive measurable real-valued processes with*

$$\int_0^t (|\mu_s| + |\nu_s|) ds \quad < \quad \infty \tag{3.13}$$

$\mathcal{P}$-almost surely for all $t \geq 0$ and

$$\int_0^t (\lambda_j^2(s) + v_j^2(s)) ds < \infty \tag{3.14}$$

$\mathcal{P}$-almost surely for all $t \geq 0$ and $j = 1, \ldots, d$. Then the stochastic differential equation

$$dS_t \quad = \quad (\mu_t S_t + \nu_t)\, dt + \sum_{j=1}^{d} (\lambda_j(t) X_t + v_j(t))\, dW_j(t) \tag{3.15}$$

with $S_0 = s_0$ has the (almost surely) unique solution

$$S_t \quad = \quad K_t \left(s_0 + \int_0^t \frac{1}{Z_s} \left(\nu_s - \sum_{j=1}^{d} \lambda_j(s) v_j(s) \right) ds + \sum_{j=1}^{d} \int_0^t \frac{v_j(s)}{K_s} dW_j(s) \right)$$

with

$$K_t \quad = \quad \exp\left(\int_0^t \left(\mu_s - \frac{1}{2} ||\lambda_s||^2 \right) ds + \sum_{j=1}^{d} \int_0^t \lambda_j(s) dW_j(s) \right)$$

where $(W_1, \ldots, W_d)$ is a Brownian motion.

Proof. See Korn [KK01], page 63. □

Theorem 3.1.10. *The price of a stock which pays no dividend under the Black-Scholes framework with cosh shortrate dynamics is given by*

$$S_t \quad = \quad s_0 \exp\left(\int_0^t r_s ds - \frac{1}{2} \sigma^T \rho \sigma t + \sigma^T C Z^{\mathcal{Q}}(t) \right)$$

for all $t \geq 0$ whereby $Z^{\mathcal{Q}} = \left(Z_1^{\mathcal{Q}}, \ldots, Z_d^{\mathcal{Q}}\right)$ is a Brownian motion under the risk-neutral measure.

Proof. By (3.12), we have $\mu_t := r_t$, $v_t = 0$, $\nu_t = 0$ and

$$\begin{aligned} S_t \sigma^T C dZ_t^Q &= S_t \sum_{i=1}^d \sigma_i \sum_{j=1}^d C_{ij} dZ_j^Q(t) \\ &= \sum_{j=1}^d S_t \left(\sum_{i=1}^d \sigma_i C_{ij} \right) dZ_j^Q(t), \end{aligned}$$

hence

$$\lambda_j(t) := \sum_{i=1}^d \sigma_i C_{ij}.$$

With $\lambda_t := \sigma^T C$

$$||\lambda(t)|| := \sqrt{\sigma^T C (\sigma^T C)^T} = \sqrt{\sigma^T \rho \sigma}.$$

Hence by theorem 3.1.9

$$S_t = s_0 K_t$$

with

$$\begin{aligned} K_t &= \exp\left(\int_0^t \left(r_s - \frac{1}{2}\sigma^T \rho \sigma \right) ds + \sum_{j=1}^d \int_0^t \sum_{i=1}^d \sigma_i C_{ij} dZ_j^Q(t) \right) \\ &= \exp\left(\int_0^t r_s ds - \frac{1}{2}\sigma^T \rho \sigma t + \sigma^T C Z^Q(t) \right) \end{aligned}$$

which provides the desired result. □

We employ the drift correction terms derived using the bond market model to implement the change of measure to the reference or the physical measure. Hence, in general, we have

$$\begin{aligned} dS_t &= r_t S_t dt + \sigma^T S_t C \left(dZ_t^P + \Lambda^P(X_t) \right) \\ &= S_t \left(r_t + \sigma^T C \Lambda^P(X_t) \right) dt + \sigma^T S_t C dZ_t^P \end{aligned}$$

for some equivalent measure $P \in \{\tilde{\mathcal{P}}, \mathcal{P}\}$. Following the proof of theorem 3.1.10, we have

$$\mu_t = r_t + \sigma^T C \Lambda^P(X_t)$$

whereas ν_t, v_t and $\lambda(t)$ are given as in theorem 3.1.10. This provides the stock price under an equivalent measure as follows.

Theorem 3.1.11. *The price process of a stock which pays no dividend under the Black-Scholes framework with cosh shortrate dynamics under a measure P equivalent to the risk-neutral measure with a drift correction term $\Lambda(X_t)$ between the risk-neutral measure and the measure P is given by*

$$S_t = s_0 \exp\left(\int_0^t \left(r_s + \sigma^T C \Lambda^P(X_t) \right) ds - \frac{1}{2}\sigma^T \rho \sigma t + \sigma^T C Z^P(t) \right).$$

for all $t \geq 0$ whereby $Z^P = (Z_1^P, \ldots, Z_d^P)$ a Brownian motion under the measure P.

Corollary 3.1.12. *The price process of a stock which pays no dividend under the Black-Scholes framework with cosh shortrate dynamics under the physical measure $\mathcal{P}$ is given by*

$$\begin{aligned} S_t &= s_0 \exp\left(\int_0^t \left(r_s + \sigma^T \rho\gamma \tanh(\gamma^T X_s + c) \right) ds - \sigma^T \kappa(\mu - \tilde{\mu})t \right. \\ &\quad \left. -\frac{1}{2}\sigma^T \rho\sigma t + \sigma^T C Z(t) \right) \end{aligned}$$

for all $t \geq 0$ where $Z = (Z_1, \ldots, Z_d)$ is a Brownian motion under $\mathcal{P}$ with uncorrelated components.

Proof. In the cosh model, the market price of risk is given by

$$\Lambda^{\mathcal{Q},\mathcal{P}}(X_t) = C\gamma \tanh(\gamma^T X_t + c) - C^{-1}\kappa(\mu - \tilde{\mu}),$$

see theorem 2.2.25. Since the Novikov condition holds for $\Lambda^{\mathcal{Q},\mathcal{P}}(X_t)$, we can apply theorem 3.1.9. Hence, using theorem 3.1.11, we get

$$\begin{aligned} S_t &= s_0 \exp\left(\int_0^t \left(r_s + \sigma^T C \Lambda^{\mathcal{Q},\mathcal{P}}(X_s) - \frac{1}{2}\sigma^T \rho\sigma \right) ds + \sigma^T C Z(t) \right) \\ &= s_0 \exp\left(\int_0^t \left(r_s + \sigma^T C \left(C\gamma \tanh(\gamma^T X_s + c) - C^{-1}\kappa(\mu - \tilde{\mu}) \right) \right) ds - \frac{1}{2}\sigma^T \rho\sigma ds \right. \\ &\quad \left. + \sigma^T C Z(t) \right) \\ &= s_0 \exp\left(\int_0^t \left(r_s + \sigma^T \rho\gamma \tanh(\gamma^T X_s + c) \right) ds \right. \\ &\quad \left. -\sigma^T \kappa(\mu - \tilde{\mu})t - \frac{1}{2}\sigma^T \rho\sigma t + \sigma^T C Z(t) \right). \end{aligned}$$

□

Dividend payments

To implement dividend payments, we can follow Björk [Bjö98] and assume a continuous dividend yield process

$$dD_t = S_t \delta(S_t) dt. \tag{3.16}$$

Then the dynamics of the stock under the risk-neutral measure become

$$dS_t = (r_t - \delta(S_t))S_t dt + S_t \sigma^T C dZ_t^{\mathcal{Q}}.$$

Using theorem 3.1.9, we can again derive a closed form solution to the stock price, this time dependent on $\delta(\cdot)$.

Theorem 3.1.13. *Let (S_t) be the price process of a stock which pays a continuous dividend yield as given in (3.16) with $(\delta(S_t))$ bounded, non-negative and progressively measurable.*

Under a Black-Scholes framework with stochastic shortrate taken from the cosh model the stock price is given by

$$S_t = s_0 \exp\left(\int_0^t (r_s - \delta_s)ds - \frac{1}{2}\sigma^T \rho \sigma t + \sum_{i=1}^{d}\sum_{j=12}^{d} \sigma_i C_{ij} Z_j^{Q}(t)\right)$$

for $t \geq 0$.

Proof. In this case,

$$\mu_t := r_t - \delta_t = \alpha - \gamma^T \kappa(\tilde{\mu} - X_t) \tanh\left(\gamma^T X_t + c\right) - \frac{1}{2}\gamma^T \rho \gamma - \delta_t.$$

μ_t is progressive measurable. Furthermore, $\nu_t = 0$. For r_t, $\exp\left(\int_0^t r_s ds\right) < \infty <$ is easily fulfilled from the same reasons as in the previous case without dividends. As we assumed dividend returns to be bounded, (3.13) holds. On the other side, $v_t = 0$ and λ_t is defined as previously, in particular (3.14) holds. Now if we apply again theorem 3.1.9, then

$$S_t = s_0 K_t$$

with

$$\begin{aligned} K_t &= \exp\left(\int_0^t (r_s - \delta_s)ds - \frac{1}{2}\sigma^T \rho \sigma t + \sum_{i=1}^{d}\sum_{j=1}^{d} \int_0^t \sigma_i C_{ij} dZ_j^{Q}(t)\right) \\ &= \exp\left(\int_0^t (r_s - \delta_s)ds - \frac{1}{2}\sigma^T \rho \sigma t + \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i C_{ij} Z_j^{Q}(t)\right). \end{aligned}$$

□

This may be used to specify δ_s in such a way that the integral $\int_0^t r_s - \delta_s ds$ has a closed form solution. Note, however, that δ_s must still be mean reverting and positive.

Implementation

The main implementational problem here is the identification of a state vector process under the physical measure for the state space model required in the Extended Kalman filter. First, we have an Ornstein-Uhlenbeck process

$$dX_t^{OU} = \kappa\left(\mu - X_t^{OU}\right)dt + CdZ_t$$

driving the bond market where $Z = (Z_1, \ldots, Z_d)$ is a Brownian motion under the physical measure with uncorrelated components. Second, the stock price formula depends on the previous Ornstein-Uhlenbeck process through both the shortrate and the market price of risk. Nevertheless, the stock price process also depends on a Brownian motion

$$dX_t^{BM} = dZ_t.$$

As both processes (X^{OU}) and $\left(X^{BM}\right)$ are driven by $Z = (Z_1, \ldots, Z_d)$, they are clearly correlated. Whereas this joint dependence on (Z) is mathematically simple, we will see that this requires special attention in implementing a Kalman filter. Neither $\left(X^{BM}\right)$ nor $\left(X^{OU}\right)$ can be defined as the state vector since we cannot derive one process from the other for $t > 0$. If we consider the respective discretized transition equations of both $\left(X^{OU}\right)$ and $\left(X^{BM}\right)$, we get

$$X_t^{OU} \;=\; e^{-\kappa} X_{t-1}^{OU} + \eta_t \tag{3.17}$$

and

$$X_t^{BM} \;=\; X_{t-1}^{BM} + \eta_t \tag{3.18}$$

whereby η_t is multivariate normal with zero mean and covariance matrix $Q(\theta)$ given by

$$\begin{aligned} Q(\theta) &= Cov\left[X_{t+1}|X_{t+\Delta}\right] = Cov[X_{t+\Delta}] \\ &= \left(\left(\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\rho_{ij}}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)\Delta}\right)\right)\right)_{i,j=1,\ldots,d} \end{aligned} \tag{3.19}$$

where θ denotes again the vector of model parameters. The idea now is to use the white noise process (η_t) as the new state process. Then, using the Kalman filter as defined in section 2.3, the transition equation of the new state vector is given by

$$X_t \;=\; \eta_t(\theta).$$

This implies

$$X_{t|t-\Delta} \;=\; 0$$

and

$$\Sigma_{t|t-\Delta} \;=\; Q(\theta).$$

For the measurement equations of the stock price and interest rates, however, we still require both $\left(X^{OU}\right)$ and $\left(X^{BM}\right)$. We can use the discretizations (3.17) and (3.18) to derive the current state of X_t^{OU} and X_t^{BM}, respectively, as a function of $X_{t-\Delta}^{OU}$ and $X_{t-\Delta}^{BM}$ and the current state X_t. This implies

$$E[X_{t+n\Delta}^{OU}|X_t^{OU}] \;=\; e^{-\kappa n\Delta} X_t^{OU} = e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t.$$

Hence, the measurement equations for observed interest rates $y^M(t, t+\tau_i)$ become

$$\begin{aligned}
y^M(t,t+\tau_i) &= g_i(X_t;\theta) + \epsilon_t^{(i)}(\theta) \\
&= \alpha - \frac{\log\cosh\left(\gamma^T E\left[X_{t+n\Delta}^{OU}|X_t^{OU}\right] + c\right)}{n\Delta} + \frac{\log\cosh\left(\gamma^T X_{t-\Delta}^{OU} + \gamma^T X_t + c\right)}{n\Delta} \\
&\quad + \frac{\gamma^T \Sigma(t,t+n\Delta)}{2n\Delta} + \epsilon_t^i(\theta) \\
&= \alpha - \frac{\log\cosh\left(\gamma^T e^{-\kappa n\Delta} X_t^{OU} + c\right)}{n\Delta} + \frac{\log\cosh\left(\gamma^T X_{t-\Delta}^{OU} + \gamma^T X_t + c\right)}{n\Delta} \\
&\quad + \frac{\gamma^T \Sigma(t,t+n\Delta)}{2n\Delta} + \epsilon_t^i(\theta) \\
&= \alpha - \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa n\Delta} e^{-\kappa\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right)}{n\Delta} \\
&\quad + \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right)}{n\Delta} + \frac{\gamma^T \Sigma(t,t+n\Delta)}{2n\Delta} + \epsilon_t^i(\theta) \\
&= \alpha - \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right)}{n\Delta} \\
&\quad + \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right)}{n\Delta} + \frac{\gamma^T \Sigma(t,t+n\Delta)}{2n\Delta} + \epsilon_t^i(\theta).
\end{aligned}$$

As $e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU}$ can be regarded a constant, we get for $i = 1, \dots, d$

$$\begin{aligned}
\frac{\partial}{\partial x_j} g_i(X_t;\theta) &= \frac{\partial}{\partial x_j}\left(\alpha - \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right)}{n\Delta}\right. \\
&\quad \left. + \frac{\log\cosh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right)}{n\Delta} + \frac{\gamma^T \Sigma(t,t+n\Delta)}{2n\Delta}\right) \\
&= \frac{1}{n\Delta}\left(-\frac{\sinh\left(\gamma^T \left(e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right)}{\cosh\left(\gamma^T \left(e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right)} \gamma_j e^{-\kappa_j n\Delta}\right. \\
&\quad \left. + \frac{\sinh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right)}{\cosh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right)} \gamma_j\right) \\
&= \frac{1}{n\Delta}\left(-\tanh\left(\gamma^T \left(e^{-\kappa(n+1)\Delta} X_{t-\Delta}^{OU} + e^{-\kappa n\Delta} X_t\right) + c\right) \gamma_j e^{-\kappa_j n\Delta}\right. \\
&\quad \left. + \tanh\left(\gamma^T \left(e^{-\kappa\Delta} X_{t-\Delta}^{OU} + X_t\right) + c\right) \gamma_j\right).
\end{aligned}$$

For the second measurement equation describing the stock market, we recommend return data. Given the definition of the stock price under the physical measure in corollary 3.1.12, we require the integral

$$\int_0^t r_s + \sigma^T C \Lambda(X_s) ds,$$

which has to be approximated. On the other side, stock returns are given by

$$\begin{aligned}
\log\left(\frac{S_t}{S_{t-\Delta}}\right) &= \int_{t-\Delta}^t \left(r_s + \sigma^T \rho\gamma \tanh(\gamma^T X_s + c)\right) ds - \sigma^T \kappa(\mu - \tilde{\mu})\Delta \\
&\quad - \frac{1}{2}\sigma^T \rho\sigma\Delta + \sigma^T C\left(Z(t) - Z(t-\Delta)\right).
\end{aligned}$$

In this case, we only have to approximate the integral

$$\int_{t-\Delta}^{t} \left(r_s + \sigma^T \rho\gamma \tanh(\gamma^T X_s + c)\right) ds,$$

which is preferable. As approximations we propose

$$\int_{t-\Delta}^{t} r_s ds \quad = \quad y(t-\Delta, t)\Delta$$

which is a reasonable assumption for Δ short enough. Furthermore

$$\int_{t-\Delta}^{t} \sigma^T \rho\gamma \tanh\left(\gamma X_s + c\right) ds \quad = \quad \frac{1}{2}\left(\tanh\left(\gamma X_{t-\Delta} + c\right) + \tanh\left(\gamma X_t + c\right)\right)\sigma^T \rho\gamma.$$

Therefore, discretizing the stock return we get the $(n+1)$-th measurement equation, the stock returns $R^M(t-\Delta, t)$ observed in the stock market, by

$$\begin{aligned}
& R^M(t-\Delta,t) \\
= \;& g_{n+1}(X_t;\theta) + \epsilon_t^{n+1}(\theta) \\
= \;& \frac{1}{\Delta}\log\left(\frac{S_t}{S_{t-\Delta}}\right) + \epsilon_t^{n+1}(\theta) \\
= \;& \frac{1}{\Delta}\Bigg(y(t-\Delta,t)\Delta + \frac{1}{2}\left(\tanh\left(\gamma^T X_{t-\Delta}^{OU} + c\right) + \tanh\left(\gamma^T X_t^{OU} + c\right)\right)\sigma^T\rho\gamma\Delta + \sigma^T\kappa(\mu-\tilde{\mu})\Delta \\
& -\frac{1}{2}\sigma^T\rho\sigma\Delta + \sigma^T\left(X_t^{BM} - X_{t-\Delta}^{BM}\right)\Bigg) + \epsilon_t^{n+1}(\theta) \\
= \;& \frac{1}{\Delta}\Bigg(y(t-\Delta,t)\Delta + \frac{1}{2}\left(\tanh\left(\gamma^T X_{t-\Delta}^{OU} + c\right) + \tanh\left(\gamma^T X_t^{OU} + c\right)\right)\sigma^T\rho\gamma\Delta + \sigma^T\kappa(\mu-\tilde{\mu})\Delta \\
& -\frac{1}{2}\sigma^T\rho\sigma\Delta + \sigma^T\left(X_{t-\Delta}^{BM} + X_t - X_{t-\Delta}^{BM}\right)\Bigg) + \epsilon_t^{n+1}(\theta) \\
= \;& \frac{1}{\Delta}\Bigg(y(t-\Delta,t)\Delta + \frac{1}{2}\left(\tanh\left(\gamma^T X_{t-\Delta}^{OU} + c\right) + \tanh\left(\gamma^T\left(e^{-\kappa\Delta}X_{t-\Delta}^{OU} + X_t\right) + c\right)\right)\sigma^T\rho\gamma\Delta \\
& +\sigma^T\kappa(\mu-\tilde{\mu})\Delta - \frac{1}{2}\sigma^T\rho\sigma\Delta + \sigma^T X_t\Bigg) + \epsilon_t^{n+1}(\theta).
\end{aligned}$$

With this, we can calculate the derivatives by

$$\begin{aligned}
& \frac{\partial}{\partial x_j}\left(g_{n+1}(X_t;\theta) + \epsilon_t^{(d+1)}(\theta)\right) \\
= \;& \frac{\partial}{\partial x_j}\frac{1}{\Delta}\Bigg(y(t-\Delta,t)\Delta + \frac{1}{2}\left(\tanh\left(\gamma^T X_{t-\Delta}^{OU} + c\right) + \tanh\left(\gamma^T\left(e^{-\kappa}X_{t-\Delta}^{OU} + X_t\right) + c\right)\right)\sigma^T\rho\gamma\Delta \\
& +\sigma^T\kappa(\mu-\tilde{\mu})\Delta - \frac{1}{2}\sigma^T\rho\sigma\Delta + \sigma^T X_t\Bigg) \\
= \;& \frac{1}{\Delta}\left(\frac{1}{2}\frac{\partial}{\partial x_j}\tanh\left(\gamma^T\left(e^{-\kappa}X_{t-\Delta}^{OU} + X_t\right) + c\right)\sigma^T\rho\gamma\Delta + \sigma_j\right) \\
= \;& \frac{1}{\Delta}\left(\frac{1}{2}\left(1 - \tanh^2\left(\gamma^T\left(e^{-\kappa}X_{t-\Delta}^{OU} + X_t\right) + c\right)\right)\sigma^T\rho\gamma\Delta\gamma_j + \sigma_j\right).
\end{aligned}$$

Now the measurement error v_t can be derived as

$$F_{t|t-1} \quad = \quad B_{t|t-\Delta}\Sigma_{t|t-\Delta}B_{t|t-\Delta}^T + H_t$$

whereby $H_t = diag\left(\nu, \ldots, \nu, \nu^S\right) \in \mathbb{R}^{(n+1)\times(n+1)}$. The Kalman gain matrix is defined by

$$K_t = \left(B_{t|t-\Delta}\Sigma_{t|t-\Delta}\right)^T F_{t|t-\Delta}^{-1}$$

and finishes the updating steps

$$X_{t|t} = X_{t|t-\Delta} + K_t v_t = K_t v_t$$

and

$$\Sigma_{t|t} = \Sigma_{t|t-\Delta} - K_t F_{t|t-\Delta} K_t^T = Q(\theta) - K_t F_{t|t-\Delta} K_t^T.$$

Note that these updates merely provide $X_{t|t}$ and $\Sigma_{t|t}$, which will, however, not contribute to the conditional expectations $X_{t|t-\Delta}$ and $\Sigma_{t|t-\Delta}$ by definition of the state vector as white noise. On the other side, we require additional updating steps of the Ornstein-Uhlenbeck process $\left(X^{OU}\right)$ and the Brownian motion $\left(X^{BM}\right)$

$$X_{t|t}^{OU} = e^{-\kappa\Delta} X_{t|t-\Delta}^{OU} + X_{t|t}$$

and

$$X_{t|t}^{BM} = X_{t|t-\Delta}^{BM} + X_{t|t}.$$

Whereas this implements an Extended Kalman filter, note that we must expect that the model provides a poorer fit than the previous implementations of the Extended Kalman filter. For once, we had to approximate the integral in the stock return. Furthermore, the Extended Kalman filter as implemented does not provide a useful transition equation. Every estimate of the state vector X_t effectively is based solemnly on the updating step. Only through $\left(X^{OU}\right)$ and $\left(X^{BM}\right)$ do we have a correction.

Results

We implement a three-factor model. One main difference between the Black-Scholes based approach and the dividend discount model is computational speed. Whereas the dividend discount model is computationally slow, the Black-Scholes based approach is very fast in estimation and simulation alike, since calculation of the current stock return is computationally equivalent to computation of an interest rate with given maturity. We were therefore able to derive several estimation results.

First note that stock returns differ substantially from interest rates in autocorrelation and variance. Whereas interest rates are highly autocorrelated, autocorrelation of monthly stock returns is less than 0.01 for the dataset used. Standard deviation of interest rate time series used for estimation varies between 0.021 and 0.024, whereas standard deviation of monthly stock returns is at 0.5. To account for these differences, we implemented two approaches, based on restrictions of the parameter vector γ. The framework presented by

Assuming $\gamma_i \neq 0$, $i = 1, \ldots, 3$							
5.3	4.4	5.8	5.3	4.0	3.5	4.5	5.9
5.4	4.4	5.8	5.4	4.1	3.6	4.5	5.9
5.4	4.4	5.8	5.4	4.0	3.5	4.5	5.9
Assuming $\gamma_3 = 0$							
12.7	5.2	8.8	12.5	8.8	6.1	5.4	9.9
12.7	5.2	8.7	12.5	8.8	6.0	5.4	9.9
12.7	5.2	8.8	12.5	8.8	6.0	5.4	9.8

Table 3.2: Mean Absolute Errors of the term structure by joint bond market and Black-Scholes-based stock market model based on the cosh model.

Albrecht [Alb07] implies a pure stock market factor, which in our model would be equivalent to $\gamma_3 = 0$. We will consider two model frameworks, one without restrictions on γ, which implies that the state vector drives both stock and bond markets, and one with $\gamma_3 = 0$, which implies two state vector components driving bond and stock markets, and one state vector component driving only the stock market. Note that implementation of these restrictions is very easy, and pricing formulae still hold in contrast to Albrecht, which had significant problems in implementing correlated stochastic factors. In this section, all state vector components are assumed to be correlated and correlations are estimated due to historical data.

Table 3.3 provides the parameter estimates. Note also that the parameter estimates are fairly stable. We see that Loglikelihood values of the restricted model are significantly lower than in the unrestricted model.

Table 3.2 provides MAEs of implied yield curves. We see that restricting γ implies a poorer term structure fit. The reason, confirmed later by examining the filtered vectors (X^{BM}) and (X^{OU}), is that all three vector components contain term structure data for general γ, whereas $\gamma_3 = 0$ guarantees that the third state vector component drives stock returns only and therefore improves stock return fit.

Considering historical fit of the stock price, we have MAEs in basis points of more than 1700 for general γ and 18 basis points for $\gamma_3 = 0$. The reason is that for $\gamma_3 = 0$, the Extended Kalman filter fits Z_3 to the observed stock price, whereas with general γ a trade-off exists between fitting the stock price and the term structure. Given the lower Loglikelihood values of the restricted approach, we can expect that the distribution of Z_3 according to the filtering in the restricted case deviates from the theoretical model-implied distribution of Z_3. By definition of the model, stock returns are normally distributed, whereas it is well known that this is not the case in reality. Therefore the good historical fit of the model assuming $\gamma_3 = 0$ does not reflect the basic problems this approach takes

α	γ_1	γ_2	γ_3	κ_1	κ_2	κ_3	ρ_{12}	ρ_{13}	ρ_{23}	c	σ_1	σ_2	σ_3	ν	ν^S	LogL
Unrestricted γ																
0.061	0.049	0.317	0.022	0.550	0.031	1.421	-0.001	-0.889	-0.329	0	0.001	0.091	0.102	0.00066	0.03000	14326
0.060	-0.049	-0.317	0.022	0.552	0.032	1.452	-0.001	0.894	0.320	0	-0.007	-0.093	0.110	0.00066	0.02900	14327
0.060	-0.048	-0.318	0.022	0.560	0.031	1.410	0.006	0.892	0.329	0	-0.007	-0.092	0.107	0.00066	0.02999	14328
Restricted γ																
0.051	0.268	0.051	0	0.071	0.398	0.296	-0.849	0.501	-0.511	0.645	0.026	-0.052	-0.175	0.00126	0.00003	13574
0.051	0.268	0.051	0	0.071	0.398	0.194	-0.849	0.503	-0.512	0.644	0.026	-0.052	-0.174	0.00126	0.00005	13574
0.051	0.051	0.268	0	0.398	0.071	3.565	-0.849	-0.485	0.438	-0.644	0.068	0.000	0.193	0.00126	0.00001	13574

Table 3.3: Parameters estimates of the Black-Scholes based stock market extension to the cosh model with joint dynamics $\gamma \neq 0$ and a pure stock market factor $\gamma_3 = 0$. μ is derived using the calibration approach discussed in 2.3.8.

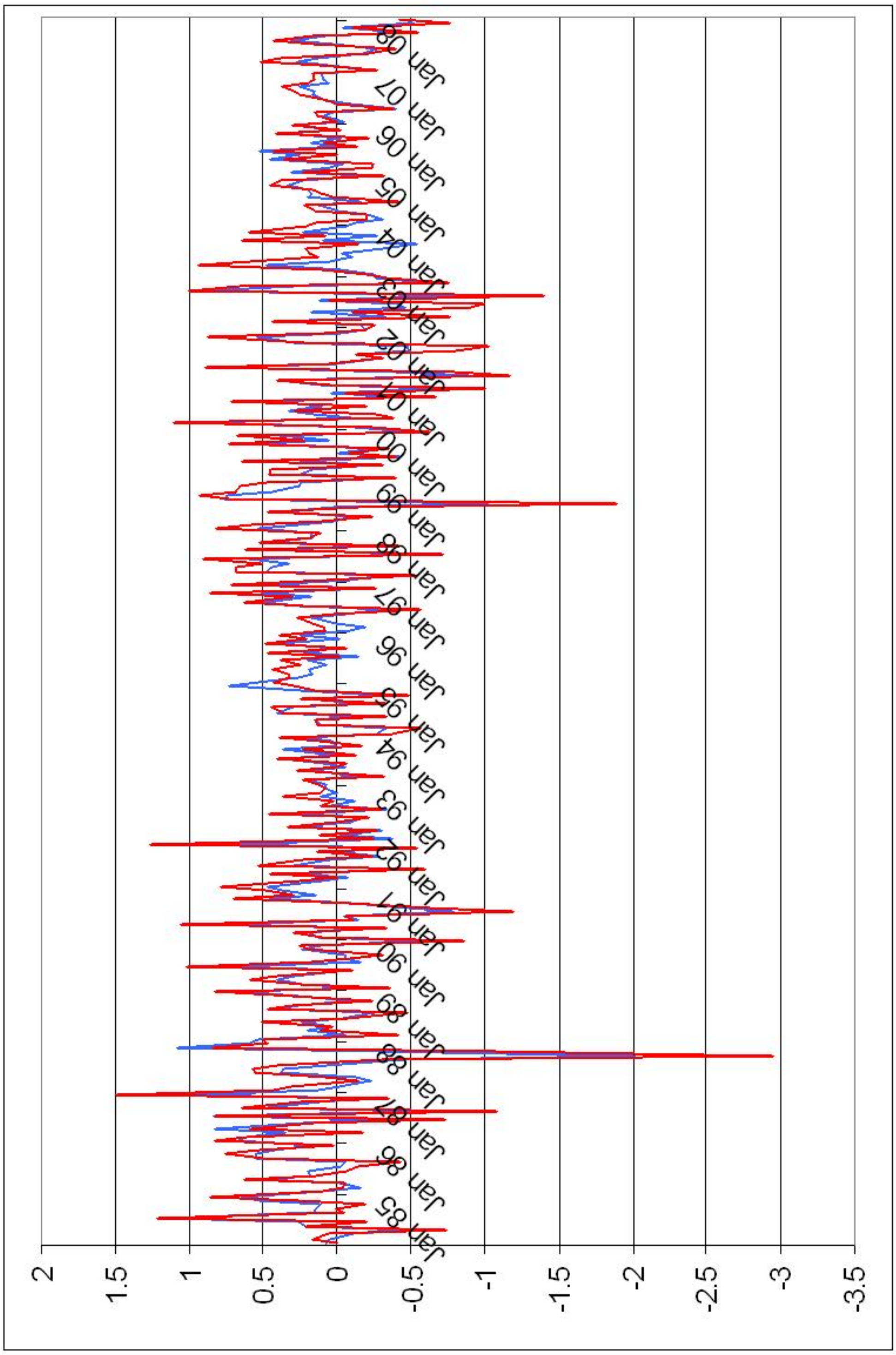

Figure 3.3: Historical monthly stock returns (red) and model-implied monthly stock returns (blue) of the Black-Scholes-based stock market extension of the cosh model. Returns are depicted annualized.

from the Black-Scholes model. We expect that non-normality of stock returns is responsible for the lower Loglikelihood values of the restricted model. A straightforward improvement of the joint model would allow for stochastic volatility of the state vectors, thereby introducing stochastic volatility in the stock market as well as the bond market.

Figure 3.3 shows historical and model-implied stock returns for general γ. We see that, although model-implied returns deviate substantially, they are nevertheless highly correlated.
We next examine the filtered factors X^{BM} and X^{OU}. For $\gamma_d = 0$, we find a level and a slope factor, as usually. The remaining factor drives the stock returns. For general γ, however, the third factor and curvature are correlated by -0.78. This implies that the third factor describes curvature rather than stock returns. It also explains the significantly better bond market fit found in table 3.2 for the model assuming general γ. It seems that term structure dynamics dominate the stock market observations. It is thus surprising that model-implied stock returns nevertheless show such a correlation with observed stock market returns.

3.1.5 Summary

In this section we presented two approaches to expand state price density models of the bond market to joint bond and stock market models. In general, stock market models may be implemented using return-based or price-based approaches. Both approaches have their merits: banking applications typically consider stock derivatives, which are based on stock prices rather than returns. Therefore, price-based approaches are superior for banking applications. Once dividend payments are introduced, however, the situation changes. To realistically implement discrete dividend payments, we require path-dependent approaches and consider reinvestment of dividends payed. In insurance applications, reinvestment of dividends is an important task since, over the long run, dividend returns make up a sizeable part of overall stock returns and furthermore intermediate dividend payments provide free cash flows without the need to liquidate assets under management.

We then considered in general implementation of stock models within the state price density approach. We found that simple restrictions imply that assuming dividend payments, the product of the state price density and the stock price ($\varsigma_t S_t$) must be a supermartingale, whereas without dividend payments this product must be a martingale. The wealth process including reinvested dividends is a martingale. Considering Rogers generic approach to define term structure models, these findings imply that a joint arbitrage-free bond and stock market model requires definition of two functions f and g which guarantee that the state price density $(e^{-\alpha t} f(X_t))_{t \geq 0}$ is a supermartingale and $(e^{-\alpha t} f(X_t) g(X_t))_{t \geq 0}$ is either a martingale or a supermartingale dependent on whether dividend payments occur or not. Given the problems in finding a function f such that $e^{-\alpha t} f(X_t)$ is a supermartingale, it became obvious that Rogers generic approach does not allow for a simple definition of a joint bond and stock market model.

We defined two alternative approaches. Both are based on simple ideas independent from the bond market model and its definition. First, we use the so-called dividend discount model, which assumes that the stock price at time t equals the discounted value of all future dividends the stockholder will receive. Using the state price density to discount the dividends links the bond market model and the stock market model. All that is yet required is a model which describes dividend payments as a function of the underlying state vector. Based upon straightforward economic considerations, we defined dividends as functions of the state vector in such a way that the current price of any dividend payed can be expressed in closed form for the cosh model. For the Cairns model, the same approach is mathematically possible, yet computationally infeasible. For the cosh model, we derive no-arbitrage of the joint bond and stock market. Furthermore, we present an Extended Kalman filter which allows to jointly estimate stock and bond market dynamics. Whereas implementation requires further work, preliminary results show that the joint model is able to jointly fit historical bond and stock market data remarkably well. Furthermore, as the stock price at time T in this model is a function of the current state X_T only, Monte Carlo-based pricing of stock options under stochastic interest rates is remarkably easy as well.

The second approach presented is based on the Black-Scholes stock market model under the risk-neutral measure. Using the drift correction terms as derived for the bond market model, we can derive a formula for the stock price under the physical measure. Again, we present how to estimate the joint model using the Extended Kalman filter. The resulting model is computationally superior to the dividend discount model, yet it does not fit historical data to the same extent as the dividend discount model.
The basic ideas of the two approaches presented may be applied to both the cosh and the Cairns model. Nevertheless, the dividend discount approach is computationally unfeasible for the Cairns model. The Black-Scholes approach, on the other side, proved to be fairly easy in implementation and estimation and therefore should allow for expansion of the Cairns model as well.
Additional work is yet required to improve estimation speed and stability in both approaches. To test the joint model, long-term stock options may provide a reasonable test of both stochastic discounting and stock price dynamics. It is important to note, however, that both models do not incorporate stochastic volatility, which is likely required to price stock market derivatives.

3.2 Macroeconomic variables

It is well known that the term structure is the major link between macroeconomy and finance. Some stylized facts were already mentioned in 2.1, namely the connection of the slope with monetary policy and the business cycle on one hand and the connection of long-term yields with inflation expectations on the other. We also saw how inclusion of macroeconomic data improves forecasting. In this section, we will describe possibilities to define joint macro-finance models and how to implement these approaches in case of the Cairns and cosh models.

3.2.1 Literature overview

As an introduction, we present a literature overview first covering the mutual dependencies between the term structure and macroeconomic information, and second considering the approaches how macroeconomic information is implemented within term structure models. In particular, three questions are of major interest:

- How does macroeconomic information affect the term structure and vice versa?
- Should dependencies between macroeconomic variables and the term structure be implemented uni-directional or bi-directional?
- which macroeconomic variables are used?

Evans and Marshall [EM01] examine macroeconomic shocks on the nominal yield curve. They present empirical evidence that macroeconomic variables explain most of the variation of interest rates with maturities ranging from 1 month through 5 years. This implies a clear uni-directional link from macroeconomic information to term structure dynamics.
Another important part of the literature describes monetary policy as an uni-directional link from macroeconomic variables to the term structure. Taylor [Tay93] presents a simple rule which describes how the central bank sets its policy rate based on inflation and output and their respective long-term trends. Now whereas such a policy rule implements an uni-directional link from macroeconomic variables to the term structure, note that the underlying goal of monetary policy is to control inflation and promote economic growth. If the current term structure reflects monetary policy, we should be able to forecast future macroeconomic variables based on the current term structure or, more precisely, based on the stand of current monetary policy as reflected in the current term structure. Bernanke and Blinder [BB92] indeed prove that the slope reflects the conduct of monetary policy and therefore predicts the true goal of the monetary authority: limiting inflation and promoting economic growth on the very long run. Estrella and Mishkin [EM97] also show that monetary policy is an important determinant of the slope, but not exclusively. They also find a significant predictive power of the term structure for both real activity and

inflation. Estrella and Hardouvelis [EH91] showed that term structure-based forecasts even outperform survey forecasts of macroeconomic variables, which proves the existence of an important link from the term structure to macroeconomic variables.
We can conclude that monetary policy implies a bi-directional approach to implement macroeconomic variables within a term structure model: monetary policy sets the current term structure in response to the current macroeconomic outlook with the explicit goal to influence future dynamics of macroeconomic variables. Macroeconomic variables and the term structure should therefore be implemented as a joint model of mutually dependent factors. Monetary policy rules can be used to implement this in a joint macro-finance model.

Dewachter and Lyrio [DL03] present such a bi-directional approach. They define an essentially affine term structure model[6] which includes macroeconomic factors. They implement output gap, inflation, stochastic central tendency of inflation and the instantaneous real interest rate as (partly) observable factors driving the term structure. In particular, this implies that at least some state factors are observable in the Kalman filter approach used. Dewachter and Lyrio find that the level of the term structure is related to inflation expectations, whereas the slope captures the business cycle and curvature is related to monetary policy. Diebold, Rudebusch and Aruoba [DRA04] also find that level and inflation on one side and slope and real activity on the other side are highly correlated. The curvature factor, however, is unrelated to any of the main macroeconomic variables. Rudebusch and Wu [WR04] first consider a yields-only model and find that the latent term structure factors are closely related to macroeconomic and monetary policy factors, namely in the same way as Dewachter and Lyrio [DL03]. Then they provide an affine term structure model as a joint macro-finance model with two latent factors and two state factors linked to output and inflation, respectively. This assumes again some state factors to be observable.
Dewachter and Lyrio's approach seems to the the simplest approach to include macroeconomic variables into factor models. In particular, it provides a bi-directional framework with a set of latent factors driving term structure dynamics and additional macroeconomic factors augmenting the set of latent variables. Bi-directionality is guaranteed since the joint dynamic of the factor is derived. The framework allows to determine a unique factor process component for each macroeconomic variable included, which simplifies subsequent analysis. On the other side, this overview shows that macroeconomic variables already contain parts of the information in level, slope and curvature. In particular, it seems that the level of the term structure is related to inflation or inflation expectations and the slope of the term structure is related to the business cycle and monetary policy, respectively. In some examples, the authors completely renounced latent factors. With level and slope information already contained in output and inflation factors, a single additional latent

[6] See [Duf00].

factor might be sufficient.

The first two questions posed above are therefore answered: monetary policy works as a link between the economy and the term structure. Based upon the current macroeconomic situation, the term structure is set with the explicit goal to shape the future dynamics of the economy. The connection between term structures and macroeconomic variables is therefore necessarily bi-directional. Furthermore, we found that inflation data and economic activity are the major ingredients in a joint model. Additionally, as in the McCallum rule[7], monetary variables such as borrowed or unborrowed reserves or monetary aggregates may be used[8], particularly if we hold a monetarist point of view.

Another important factor for some countries is considered by Clarida, Gali and Gertler [CGG97]. These authors examine monetary policy for Germany, Japan, the U.S., the UK, France, and Italy. They find a set of leading central banks (US, Germany and Japan) which conducted monetary policy independently of other central banks based upon domestic considerations alone. The remaining central banks typically conducted monetary policy in response to Germany, something often called *German leadership hypothesis*. The hypothesis implies that the Taylor rule only works for leading central banks, whereas following central banks set their monetary policy according to leading banks rather than domestic macroeconomic variables as implied by the Taylor rule. In particular, this implies that the Taylor rule can not be applied for following central banks.

3.2.2 Indices or rates?

We will now discuss two approaches to include macroeconomic data into term structure models. They differ in that they are either index-based or rate-based. Many macroeconomic variables are available in both forms, an example would be GDP data, where the actual value of GDP can be seen as an index, whereas the GDP growth rate describes the

[7] The McCallum rule explains inflation with the growth of the money supply.

[8] *Monetary aggregates* are measures of the amount of currency outstanding. Currency in circulation and the vaults of depository institutions form the base aggregate. Broader aggregates contain additional money analogies with decreasing liquidity, for example demand deposits, savings deposits, time deposits and money market funds. *Bank reserves* are deposits of commercial banks at the central bank plus currency held in bank vaults. Given minimum reserve requirements, the amount of bank reserves determine the amount of credit the banking system has issued. On the other side, excess reserves, that is additional reserves beyond what is required by law, are a sign of distress in the banking sector. Excess reserves imply that instead of giving out new credit, the banking system hoards liquidity, as has been the case in the recent financial crisis.

These variables are of interest as they cover certain aspects of monetary policy uncovered by inflation and output growth as included in the Taylor rule. Note also that targeting monetary aggregates was a widespread approach to monetary policy until recently. To give an example, the monetary experiment 1979 to 1982 is typically assumed to be a result of the Fed targeting monetary aggregates. At least the ECB continued calling monetary aggregates an important aspect in monetary policy, whereas the Fed did not consider monetary aggregates explicitly and in fact ceased to issue data considering one of the broader measures.

(typically annualized) index change.

The first approach follows what is already proposed in the literature: we directly implement a macroeconomic rate as an observable state vector component[9]. Macroeconomic rates are reasonably described as mean reverting processes. To give an example, we want to implement the macroeconomic rate i_t as the $d+1$-th state vector component. Then, using the notation introduced in the Kalman filter, we simply follow the approach of Dewachter and Lyrio [DL03]

$$i_t := g_{n+1}(X_t; \theta) = X_t^{d+1} + \epsilon_t^{n+1}(\theta)$$

where g_{d+1} is the $(d+1)$-th measurement equation. The state vector therefore is assumed to consist of the observable component X_t^{d+1} and unobservable, latent factors. The joint dynamics of the state vector provides the bi-directional approach required. Estimation using the Kalman filter is fairly easy as the additional measurement equation is linear.

Note that although the macroeconomic rate i_t can be simulated continuously, it may not be interpreted as a "macro-shortrate". It does not describe the development of some macroeconomic variable over infinitesimally short time horizons, but the current growth rate of some macroeconomic variable over a given time period $[t_1, t_2]$. In particular, we may not integrate i_t over $[t_1, t_2]$ to derive index changes. Note also that typically only annualized rates are available, often published with a lag. Therefore, what we implement as i_t must be interpreted as a possibly lagged measurement of changes in some macroeconomic index over a given, past time span, typically one year. Annualization is due to smoothing the data. Note also that macroeconomic data may only be available at monthly frequencies, so that macroeconomic variables would enter high-frequency applications as piecewise-constant stochastic processes.

The second approach to include macroeconomic data uses the index rather than index changes. We can reasonably assume that for all macroeconomic indices a long-term "equilibrium" drift exists, which leads us to implement the index by

$$I_t = \exp\left(\bar{\mu}t + b + a^T X_t\right),$$

where $a^T X_t$ describes the deviation of the index from its long-term growth trend $\bar{\mu}$. The parameter b is required to standardize $a^T X_t$ in such a way that

$$E\left[b + a^T X_T | X_t\right] \to 0$$

for $T \to \infty$. This is necessary to ensure that the long-term drift of the index is solely described by $\bar{\mu}$. Analogously to the rate-based approach, we can define a to be the $d+1$-th unit vector $a := e_{d+1}$, hence only the $d+1$-th state vector component describes the

[9] See [DL03], and [WR04].

stochastic deviations of the index from its equilibrium trend growth. In this case, we can define $b := 0$ and $\mu_{d+1} = 0$ to ensure that the drift is only described by $\bar{\mu}$. In most cases, the index is directly observable, hence

$$g_{n+1}(X_t;\theta) := I_t = \exp\left(\bar{\mu}t + X_t^{d+1}\right)$$

which can be estimated using the Extended Kalman filter. Equivalently, we can use log-indices for estimation, hence

$$g_{n+1}(X_t;\theta) := \log I_t = \bar{\mu}t + X_t^{d+1},$$

which implements a linear measurement equation for the Kalman filter. Alternatively, we could also use rates for estimation. From a given index, the expected rate of index growth over a time period $[t,T]$ can be derived as follows

$$\begin{aligned} i_t &= E\left[\frac{1}{T-t}\log\left(\frac{I_T}{I_t}\right)\middle|\mathcal{F}_t\right] \\ &= \frac{1}{T-t}E\left[\log\left(\exp\left(\bar{\mu}T + a^T X_T - \bar{\mu}t - a^T X_t\right)\right)|\mathcal{F}_t\right] \\ &= \bar{\mu} + a^T\frac{E\left[X_T - X_t|\mathcal{F}_t\right]}{T-t}. \\ &= \bar{\mu} + a^T\frac{e^{-\kappa(T-t)}X_t + (1-e^{-\kappa(T-t)})\mu - X_t}{T-t} \\ &= \bar{\mu} + a^T(e^{-\kappa(T-t)} - 1)\frac{X_t - \mu}{T-t}. \end{aligned} \tag{3.20}$$

Such a measurement equation may be used with inflation and output forecasts. Historical rates can be derived analogously to define measurement equations using observed rates based on

$$i_t = \frac{1}{\tau}\log\left(\frac{I_t}{I_{t-\tau}}\right).$$

whereby $I_{t-\tau}$ depends on a previous state $X_{t-\tau}$. The time horizon τ can be chosen according to underlying data, for example τ is one year for annualized rates.

In general, this approach does not offer simple interpretations of state vectors and model parameters as in the rate-based approach. The stochastic factors driving the indices only describe deviations from the long-term equilibrium trend and not index dynamics as a whole. The rate-based approach requires long-term means μ of the rates, whereas the index-based approach requires equilibrium trends $\bar{\mu}$. The remaining parameters describing factor dynamics, correlation and bond prices are the same. The number of parameters required is the same for both approaches.

A major difference between the two approaches exists considering applications. Namely, we are able to derive easily prices of index-paying securities. We reinterpret the single dividend securities of section 3.1 with payoff functions

$$\exp\left(\bar{\mu}T + \left(\gamma^D\right)^T X_T\right),$$

as derivatives which pay the index at time T. As these prices were found to be arbitrage-free, estimation of the joint macro-finance model including index data allows to derive no-arbitrage prices for tradeable derivatives based on these indices. To give an example, an inflation-indexed coupon bond is a portfolio of index-paying derivatives $\Pi^I(t,T)$ for each coupon C at time t_i as well as the notional N at time t_n, hence

$$P^R(t,T) = \sum_{i=1}^{n} C\Pi^I(t,t_i) + N\Pi(t,t_n).$$

In many cases, inflation-indexed bonds also provide deflation protection of the notional, hence a put option with payoff $N(I_0 - I_T)^+$ at time T. As demonstrated in 3.1 in the cosh closed-form solutions for index-paying derivatives can be derived, whereas the Cairns model requires numerical integration. Only the put option might require a Monte Carlo approach.
In the same way, due to no-arbitrage of dividend paying securities, we are able to derive prices for derivatives whose payoff functions depend on these securities and hence on the indices. By the above derived formula for the rate, it is possible to derive caps and floors on inflation or economic growth rates, which we will later see may have applications in hedging extremal scenarios in life insurance.

In case of the Cairns and cosh model, both approaches presented can be implemented. The rate-based approach offers simple interpretations of the macro factors and the model parameters driving them. Both models allow for linear measurement equations and both models require equivalent numbers of model parameters. Considering term structure applications alone, the first approach is more intuitive. If we are interested in exotic macro-derivatives or inflation-indexed bonds, we would recommend using the second approach as it allows for closed-form solutions of index-paying derivatives in case of the cosh model and fast numerical solutions to index-paying derivatives in case of the Cairns model.

3.2.3 Stylized facts

In section 3.2.1, we already encountered some properties and hypotheses connecting the term structure and macroeconomic variables. In this section, we will examine what impact these properties have on the implementation of a joint model. As can be seen in the literature overview, in most cases term structure models are augmented with macroeconomic variables by assuming that some state vector components coincide with observable macroeconomic rates. We will now consider these approaches in light of certain stylized facts, namely

- time-homogeneity of the joint model
- the Fisher equation

- interest rate smoothing.

In the term structure models estimated so far, we implicitly assumed time-homogeneity of the dynamics. This is also assumed in most models presented in 3.2.1. However, according to Rudebusch and Wu [WR04], empirical evidence suggests that this relationship between the term structure and macroeconomic variables has changed. Their primary argument is that the transmission mechanism of monetary policy has changed. However, each regime switch in monetary policy should also imply a regime switch in the term structure. In particular, note that the infamous monetary experiment, which is widely attributed to a new approach of targeting monetary aggregates, is the standard example for a distinct regime in term structure data. To conclude, we can assume that a time-homogeneous monetary transition mechanism imposes the same problems as a pure term structure model with fixed long-term dynamics, hence time-inhomogeneity is no particular problem of joint macro-finance models but of long-term interest rate modeling in general.

The so called *Fisher equation*, developed in [Fis30], is one of the earliest hypothesis which link nominal interest rates and macroeconomic variables, in this case inflation. The Fisher equation states that there exists a long-range equilibrium between the nominal interest rate, real interest rates and expected inflation, namely

$$y(t,T) = y^R(t,T) + E_t\left[\pi(t,T)\right],$$

hence nominal interest rates equal the sum of the real interest rate and expected inflation. It is a stylized fact that real interest rates are highly persistent. If Fisher equation holds, the movements in interest rates will therefore predominantly reflect fluctuations in expected inflation, see[Mis93]. This fits very well the assumption that the level of the yield curve is closely related to expected inflation and generally high interest rates coincide with high inflation rates. The Fisher equation can be used to derive real interest rates in a framework which incorporates the inflation index. In case of the cosh model, we get, using the index-based approach,

$$\begin{aligned} y^R(t,T) &= y(t,T) - E_t\left[\pi(t,T)\right] \\ &= -\frac{\log\left(P(t,T)\right)}{T-t} - \bar{\mu}^I(T-t) - a^T\left(e^{-\kappa(T-t)} - 1\right)\frac{X_t^I - \mu^I}{T-t}. \end{aligned}$$

According to Rudebusch [Rud95], the federal reserve not only sets the target rate according to macroeconomic inflation, but also tries to *smooth* interest rate dynamics. Thus, monetary policy is implemented over the course of several meetings with gradual increases or decreases (but not both) in the target rate. Roberds and Whiteman [RW96] argue that interest rate smoothing is the main reason that the term spread can be used to forecast interest rate movements. The basic assumption is that interest rate smoothing prohibits surprises in the conduct of monetary policy.

Interest rate smoothing may be difficult to implement in a typical Markovian no-arbitrage factor model if lagged state factors are to be used. Furthermore, note that smoothing introduces asymmetric mean reversion behavior. In case of high or rising inflation and economic growth, the Fed tends to follow a gradual approach, slowly increasing the target rate. Interest rate cuts, however, tend to be considerably faster, the recent financial crisis being an example. Inclusion of macroeconomic variables might provide this asymmetric behavior, as the asymmetry coincides with different economic regimes. In particular, fast interest rate cuts coincide with a fast decrease in economic activity, whereas slow increases of the target rate coincide with increasing inflation.

3.2.4 Monetary policy rules

As discussed in 2.1, monetary policy is the defining link between macroeconomic variables and the term structure of interest rates. Monetary policy rules describe in a simplistic way how the central bank sets the target rate, hence how central bank behavior determines the term structure. This approach is based on current macroeconomic data and aims at influencing the future course of the economy as a whole. If monetary policy can be reasonably described by a simple rule, the implementation of the rule guarantees a realistic model of the mutual dependencies between macroeconomic variables and interest rates according to historical data. We will now present the Taylor rule.

The *Taylor rule* is a monetary policy rule proposed by [Tay93] which determines the target rate the monetary authority sets in reaction to the state of the economy. The current macroeconomic situation is described by real interest rates, current inflation and the deviations of both inflation and economic growth rates from their respective equilibrium growth trend. In formula,

$$r_t = \pi_t + r_t^* + a_\pi^*(\pi_t - \bar{\pi}) + a_u u_t$$

where r_t is the target rate, for example the federal funds rate in [Tay93]. π_t is the inflation rate, in Taylor's paper taken as the inflation rate over the previous four quarters as a proxy for expected inflation. This annualization smooths inflation rates. r_t^* is the equilibrium real interest rate. As real interest rates are very persistent, r_t^* is typically assumed as constant and in many implementations of the Taylor rule set to 2%. $\bar{\pi}$ is the desired rate of inflation, typically a long-term inflation target, which, assuming successful monetary policy, coincides with the long-term equilibrium trend in inflation. u_t is, according to Taylor's original paper, the percent deviation of real GDP growth from its target, measured by

$$u_t = 100\frac{U_t - U_t^*}{U_t^*}$$

where U_t is current real GDP and U_t^* is trend real GDP.

The rule implies a relatively high interest rate when inflation is above its target $\bar{\pi}$ or when the economy is overheating, with economic growth beyond the sustainable long-term trend $u_t > 0$. If these goals are in conflict, the rule provides guidance to policy makers on how to balance these competing considerations in setting an appropriate level for the interest rate. Although central banks typically do not explicitly follow the rule, analysis showed that the rule does a fairly accurate job of describing the conduct of monetary policy in the past, particularly in case of the Federal Reserve under the Greenspan chairmanship.

If we assume a term structure model which incorporates both (real) GDP and inflation in the rate-based approach, we can implement the Taylor rule as follows:

- The target rate can be interpreted as the (unobservable) shortrate r_t provided by the term structure model.
- π_t describes the inflation rate, implemented by the rate-based or index-based approaches, respectively. Following Taylor's original paper, π_t should be estimated from historical annual inflation rates.
- the real rate r_t^* can be derived using the Fisher equation, which however implies $r_t^* = r_t - \pi_t$. A widespread assumption determines $r_t^* := 2$.
- given successful monetary policy, the inflation target $\bar{\pi}$ equals the long-term mean μ^I of X_t^I.
- X_t^u describes current output growth. The long-term mean μ^u of output growth equals equilibrium output growth, hence $u_t := X_t^u - \mu^u$ determines deviation from equilibrium trend growth.

To summarize, the Taylor rule for the rate-based approach is given by

$$r_t \quad = \quad X_t^i + 0.02 + a_\pi^*(X_t^i - \mu^i) + a_u(X_t^u - \mu^u).$$

Alternatively, if we use the index-based approach, we can implement the Taylor rule as follows.

- The target rate can again be interpreted as the (unobservable) shortrate r_t as provided by the term structure model.
- The inflation rate π_t can be derived from historical data or, using forecasts, directly as expected inflation rate avoiding the approximation proposed by Taylor. Following (3.20), we get

  $$\pi_t = \bar{\mu}^I + a^T \frac{e^{-\kappa\tau} - 1}{\tau} X_t^I.$$

  where τ equals one year if annualized data is to be used. Obviously, $\bar{\mu}^I$ is the equilibrium inflation rate as the long-term mean of X_t^I is set to zero by assumption.

- For the equilibrium real interest rate r_t^*, we again recommend following the standard approach and setting $r_t^* := 2\%$.
- The output variable is defined by the index $U_t = \exp(\bar{\mu}^u t + X_t^u)$. Obviously, $\bar{\mu}^u$ provides the equilibrium growth trend, hence

$$\begin{aligned} u_t &:= 100\frac{U_t - U_t^*}{U_t} = 100\frac{\exp(\bar{\mu}^u t + X_t^u) - \exp(\bar{\mu}^u t)}{\exp(\bar{\mu}^u t + X_t^u)} \\ &= 100\left(1 - \exp(-X_t^u)\right). \end{aligned}$$

To summarize, the Taylor rule is given by

$$r_t = \bar{\mu}^I + 0.02 + a_\pi^*(\gamma^I)^T \left(e^{-\kappa^I \tau} - 1\right) X_t^I + a_u^* 100\left(1 - \exp\left(-X_t^u\right)\right).$$

In both approaches, the Taylor rule could be included into estimation by assuming that

$$\begin{aligned} g_{n+1}(X_t;\theta) &:= r_t(X_t;\theta) - (\pi_t(X_t;\theta) + r_t^*(X_t;\theta) + a_\pi^*(\theta)(\pi_t(X_t;\theta) - \bar{\pi}(\theta)) \\ &\quad + a_u(\theta)u_t(X_t;\theta)) + \epsilon_t^{n+1}(\theta) \end{aligned}$$

is observed as zero. Deviations of the rule are then interpreted as measurement error $\epsilon_t^{n+1}(\theta)$ with mean zero and zero cross- or autocorrelation, by assumption. As the Kalman filter minimizes the measurement error, the state vector is estimated implicitly in such a way that the Taylor rule holds on average. As Taylor already showed that this is the case based upon historical data, estimation only using term structure and relevant macroeconomic data implicitly contains the rule. Implementing the above described measurement equation however emphasizes the role of the Taylor rule within estimation, provides Kalman filter estimates of a_π^* and a_u^* as well as a time series of deviations from the Taylor rule.

Now if monetary policy determines the mutual dependencies between macroeconomic variables and the term structure, and policy rules are a simple yet efficient way to describe the conduct of monetary policy, we must assume that the policy rule holds into the future as well. In particular, we have to ensure that the monetary policy rule holds for simulated data. We will present a simple approach to this task.

The basic idea is to implement the deviations of the policy rule as an observable state vector component. As an example, we will take the Taylor rule and implement deviations of the Taylor rule as a mean reverting process with long-term mean zero. Output is implemented as depending on this rule-deviation variable and the remaining state variables. The Taylor rule describes how to set the target rate according to current inflation and economic activity. If the central bank keeps the target rate deliberately above the rule-implied rate, both economic activity and the inflation rate are below their respective equilibrium trends. On the other side, if the central bank sets the target rate below the rule-implied rate, both economic growth and inflation surpass their equilibrium trends. According to Taylor [Tay], the main reason for the 2007-2010 financial crisis was that the Fed kept the target rate below the level advocated by the Taylor rule. This promoted economic growth - but also the development of the house-price bubble. During the crisis following the burst of the house-price bubble, the target rate hit the zero lower bound. For the US, to give an example, the Taylor rule implied a necessary policy rate of -6%, hence keeping the target rate close to zero discouraged economic growth. Seemingly, deviations from the Taylor rule are costly no matter in which direction and therefore should be avoided. Furthermore, we see that deviations from the Taylor rule mean revert to zero, which justifies implementation of deviations from the rule as Ornstein-Uhlenbeck processes with zero long-term mean. We reinterpret the deviation of the Taylor rule as a state vector component

$$X_t^D := r_t(X_t;\theta) - [\pi_t(X_t;\theta) + r_t^*(X_t;\theta) + a_\pi^*(\theta)(\pi_t(X_t;\theta) - \bar{\pi}(\theta)) + a_u(\theta)u_t(X_t;\theta)]\,.$$

Now we have to restrict the degrees of freedom of the remaining variables to guarantee that the Taylor rule holds. Namely, we define a new measurement equation for economic growth based on the Taylor rule, hence

$$\begin{aligned} u_t(X_t;\theta) &= \frac{1}{a_u(\theta)}\left(X_t^D - r_t(X_t;\theta) + \pi_t(X_t;\theta) + r_t^*(X_t;\theta)\right.\\ &\quad \left. + a_\pi^*(\theta)(\pi_t(X_t;\theta) - \bar{\pi}(\theta))\right). \end{aligned}$$

This defines real GDP as a function of the state vector X_t including the inflation component X_t^I and the deviation factor X_t^D. By definition, interest rates and inflation rates are not restricted in their dynamics. GDP growth is stochastic and due to X_t^D not determined by interest rates and inflation alone. Furthermore, by definition, if X_t^D is strongly mean reverting with long-term mean zero, the Taylor rule holds on average for simulated data. Assuming that the central bank successfully follows the Taylor rule in the future, economic growth should develop as determined by the rule conditional on certain latent term structure factors and inflation.

3.2.5 Insurance Applications

We will now discuss possible benefits of including macroeconomic variables into term structure models from an insurance viewpoint. First, note that macroeconomic data should allow for more realistic models of term structure dynamics as central bank behavior and inflation expectations are major driving forces of the term structure. Particularly considering long-term dynamics, inflation and inflation expectations seem to be responsible for the variation in the long-end of the yield curve. We can also take as granted that forecasting power of the term structure model should increase and therefore time series properties as a whole are implemented more realistically due to macroeconomic data.

A second important contribution of macroeconomic variables lies in mutual dependencies of different financial markets. It is well known that return correlations of different financial markets are not constant in time. In particular, during times of economic distress reflected in macroeconomic variables, correlation between different financial markets tends to increase with adverse effects on diversification. As insurance products typically invest in multiple assets, all of which are subject to macroeconomic data, the inclusion of macroeconomic data provides a common stochastic factor to various financial markets. Macroeconomic data may therefore help implementing time varying return correlations.

Another point worth considering is simulation and hedging of extremal scenarios. Japan scenarios are most critical for life insurance companies. These scenarios are crucially dependent on macroeconomic variables, as a Japan scenario can only occur if both inflation and economic growth are substantially below their long-term trend and are expected to remain low. In such cases, monetary policy reaches the zero lower bound and the central bank has to apply alternative monetary instruments. These alternative instruments cause the very low and flat term structures of a Japan scenario. Consequently, the probability of a Japan scenario depends on the distribution of macroeconomic variables.
Hedging of insurance products by macroeconomic derivatives may be an interesting strategy. In case of a Japan-scenario, an inflation floor should provide a better protection than a floor on LIBOR rates. The recent experience showed that credit spreads between bond-implied rates and LIBOR rates may rise significantly in a Japan scenario caused by financial crisis. Consequently, during a Japan scenario, the payoff of a floor on bond-implied rates is significantly higher than the payoff of the available floor on LIBOR. Furthermore, note that in most cases, low short-term interest rates coincide with a high slope. In this case, hedging against a Japan scenario requires no payoff, as increasing duration of invested funds increases bond market returns. Since Japan scenarios are only possible if current inflation is extremely low, inflation floors would therefore provide a hedging instrument against Japan scenarios which does not suffer from the LIBOR-bond-yield spread. On the other side, a frequently encountered problem is that in times of crisis, correlation between financial markets increases and hence diversification effects decrease. Such times of crises typically coincide with recessions, in particular steep recessions and hence a steep decline in economic growth. This connection may be used to develop path-dependent derivatives based on economic growth which could be used to hedge against such crises, particularly the sudden loss of diversification.

Considering diversification again, Kothari and Shanken [KS04] examine asset allocation among stocks, inflation-indexed and nominal government bonds and a bank account. They find that substantial weight should be given to inflation-indexed bonds in an efficient portfolio. Consequently, inflation-indexed bonds provide diversification to bond and stock portfolios and may therefore be included into the investment policy of the insurance company. This requires the model to provide prices of inflation-indexed bonds.

Finally, there exist direct applications of macroeconomic variables for insurance products. First, macroeconomic variables may be of interest in product development, in particular indexed contracts. For once, instead of a guaranteed nominal return, life insurance companies and pension funds could guarantee real returns. On the other side, invalidity coverage may provide indexed payments in case of invalidity to cover inflation risk. In such cases, a joint model is necessary for risk management and pricing of indexed payments. Another direct application considers rational behavior based on the overall macroeconomic situation. In many insurance products, behavior of the insured is an important aspect of pricing, in particular cancellation is implemented in pricing. We can expect that changes in cancellation behavior should be related to macroeconomic variables. To give an example, we can assume that rising inflation leads to rising cancellation of older insurance contracts whose guaranteed return is below current inflation rates. On the other side, economic growth is positively correlated to disposable income and hence saving, which implies a positive correlation between economic growth and underwriting.

3.2.6 Summary

There is clear evidence in the literature that macroeconomic variables increase forecasting power of term structure models, hence macroeconomic variables improve time series properties of term structure models. We found that there exist clear bi-directional relation between macroeconomic variables and the term structure. In the literature, the joint model is implemented by assuming that the state factor driving the term structure model consists of unobservable, so-called latent factors and observable factors, which are linked to macroeconomic variables. We found two possibilities to implement observable state factors: either we assume the macroeconomic rates to be observable, and the underlying state vector to coincide with the rate directly, or alternatively we assume the state factor to describe deviations of index dynamics from long-term equilibrium trends. Both models imply similar measurement equations for estimation using the Kalman filter as well as equal numbers of model parameters. The index-based approach however allows for simple pricing of indexed-based derivatives, which could for example be used to price inflation-indexed bonds.

Another major finding considers monetary policy rules, in particular the Taylor rule. Monetary policy rules provide a simple implementation of the bi-directional link between term structure and macroeconomic variable. Historically, many (leading) central banks followed the Taylor rule in their conduct of monetary policy, hence historical macroeconomic variables and short-term interest rates follow dynamics as determined by the Taylor rule. Obviously, the Taylor rule should hold for simulated data by a joint macro-finance model as well, reflecting the assumption that the rule describes future monetary policy as well. We provide an easy idea how to implement the Taylor rule or more generally monetary policy rules into a joint macro-finance model which is to be estimated using the Kalman filter.

Finally, we discuss benefits of joint macro-finance models for insurance applications. Besides the general improvement of time series properties, macroeconomic variables might be useful in hedging extremal scenarios, since these typically coincide with certain macroeconomic situations. Macroeconomic variables might also be used to describe time varying correlation between bond and stock market returns. Finally, macroeconomic variables may be used to describe behavior of customers of insurance companies, for example regarding cancellation, which is often required in pricing.

Chapter 4

Conclusion

A basic task in mathematical finance lies in comparison of cash flows occurring at different points in time. Assuming stochastic interest rates, such a task requires an arbitrage-free financial model of term structure dynamics. Most term structure models presented in the literature so far were developed for banking applications. The major goal of this work is to examine so called state price density models on their applicability in insurance. A first contribution of this work is therefore the explicit insurance focus we take. We repeat estimation and implementation of the Cairns [Cai04a] model using the Extended Kalman filter already known in the literature. Additionally, we present the cosh model proposed by Rogers [Rog97] in its first estimation and implementation, again using the Extended Kalman filter. Comparing these models, our main result is that the cosh model may be used as a computationally efficient approximation to the otherwise superior Cairns model. A second contribution of this work is to provide ideas how to expand the pure bond market models to full investment models covering equity, government bond and inflation-indexed bond markets. As we show how to include macroeconomic variables as well as monetary policy rules, the cosh model may also be used as a macro-finance model in monetary policy applications as well as to examine the impact of macroeconomic variables on insurance products.

In section 2.1, we introduced a selection of criteria on term structure models and discussed their importance both for insurance and banking applications. The main difference hence lies in contractual time to maturity, risk factors included, particularly non-interest rate risk factors such as stock market risk or biometrical risks, and availability of market prices for comparison.

In section 2.2.1, we introduced the Rogers framework, which defines the state price density, and therefore a term structure model, by the specification of a state vector process and the choice of a function f with rather general properties. We discuss restrictions to the choice of f and the dynamics of X due to criteria on term structure models found in 2.1. In particular, we found that X should be a mean reverting process, $f(X_t)$ can not

be a martingale and $f(X_t)$ may only be a supermartingale by itself if certain additional conditions hold. Overall, we found that definition of s supermartingale based upon a mean reverting process is difficult. This motivated our examination of the cosh model, under which the state price density is not a supermartingale, since models with the state price density not being a supermartingale will likely dominate in applications of the Rogers model. Finally, we presented that the state price density approach is computationally superior to standard risk-neutral pricing in insurance applications due to often infrequent and irregular payments over very long time horizons.
In sections 2.2.5 and 2.2.6, we present the Cairns [Cai04a] and cosh [Rog97] models, respectively. Whereas both may be defined under the Rogers framework, the Cairns model originally was defined under the framework of Flesaker and Hughston [FH96]. A short comparison of these two approaches showed that the framework of Flesaker and Hughston requires definition of a martingale in dependence of a mean reverting state process, whereas Rogers requires definition of a supermartingale. Once the model is specified, the Rogers model requires a closed-form solution to the expectation of the state price density, whereas Flesaker and Hughston require closed-form solutions to the integral over the product of the chosen martingale and an additional function ϕ. As the later is typically more difficult to see, an implementation of the Rogers model is easier to be defined.
In section 2.2.7, we prove no-arbitrage for both models. Furthermore, we derive the physical measure in such a way that dynamics under the historical measure required for estimation are particularly simple.

In section 2.3, we introduce the extended Kalman filter. As the Rogers approach requires the choice of a state vector X, all Rogers models are factor models and hence the Kalman filter, or its extended form to cover for nonlinearities in the definition of yields, is a natural choice for estimation. In Section 2.3.3 we discuss estimation data for term structure models. For insurance applications, we recommend government bond-implied yields as a proxy to riskless interest rates. Macroeconomic data may improve long-term dynamics, interest rate derivative data may improve volatility fit. Whereas we find in section 3.2 that inclusion of macroeconomic data might indeed be easy, inclusion of derivative data first requires closed form solutions to interest rate derivatives and second most interest rate derivatives are based on Libor rates rather than government bond yields and therefore contain hitherto unconsidered risk factors.

Sections 2.3.4 and 2.3.5 present the implementations of the Extended Kalman filter for both models. Simulation exercises demonstrate the ability of the Extended Kalman filter to estimate the model parameters. We examine historical fit of the estimated model parameters by calculating mean absolute errors as well as cross-correlation and autocorrelation of the time series of residuals. We find that both models fit historical data remarkably well. Furthermore, in both cases residuals are highly autocorrelated and linked to curvature. Consequently, we estimated and examined three-factor models in section 2.3.7, which significantly improved historical fit, particularly with respect to curvature. In all cases, we found that the underlying state vector components are closely linked to the principal components of the term structure. In the two factor models one state vector component was highly correlated to long-end yields, whereas the other state vector component was highly correlated to an empirical proxy of the slope. In the three factor models, we found the same level factor. The other two state vector components described slope and curvature, yet in this case slope and curvature components were not clearly distinguishable. We concluded that for higher dimensions, the state vector still catches the higher principal components of term structure data, yet not with a single state vector component each. Consequently, models with dimension higher than 3 might have problems in simulating the data realistically, as the non-level components of the state vector were highly correlated.

Because long-end yields showed a falling trend and very low mean reversion throughout the data set used, the Extended Kalman filter underestimated the long-term mean of the level component of the state vector. Nevertheless, the high correlation of long-end yields and the level component allowed to specify the long-term mean of the level component. As described in section 2.3.8, we found a simple algorithm able to provide realistic estimates of the long-term mean of the state vector, thereby reducing the parameters to be estimated. As discussed in section 2.1, the coincidence of principal components and state vectors is often found in term structure models. Typically, however, short-end yields describe the level in interest rate models. Nevertheless, the high-end level factor should provide superior results in fitting the dynamics of long-end yields, since in most models long-end yields are a function of high-volatility and high-mean reversion short end factors and the model more or less deterministically reduces long-end volatility.

What is unique about the two models considered and likely for the whole Rogers framework under some rather mild conditions discussed in section 2.3.9 is that a single, distinguishable model parameter describes the asymptotic long rate, which is therefore constant, as in many other term structure models. In most models, the asymptotic long rate is a function of other model parameters, therefore introducing a long-end restriction into estimation that is largely ignored in the literature. The parameter α in the cosh and Cairns models allows for sensitivity analysis of the asymptotic long rate and exogenous specification, for example due to regulatory definition of an asymptotic long rate to be used in insurance as currently discussed in Germany. Two examples for exogenous specifications of α are provided.

Finally, section 2.4 concludes with a comparison of the Cairns and cosh model. The Cairns model provides a superior historical fit as well as superior forecasting power. Furthermore, it guarantees positive interest rates, which the cosh model can not. Nevertheless, the cosh model is by far superior computationally due to the integral in the definition of the state price density in the Cairns model requiring numerical integration. Since basic properties such as state vector dynamics, the long-end level, the constant asymptotic long rate, but also risk premia and forecasting power are very similar in both models, we can conclude that the cosh model provides a simple and fast approximation of the otherwise superior Cairns model.

In section 2.1, it was already discussed that insurance companies require term structure models not only for discounting liabilities, but also to simulate asset returns. Nevertheless, the term structure models presented may only provide returns of domestic government bonds. A major task therefore was to consistently extend the financial markets covered by the state price density models. The first, and most important, extension in section 3.1 considers equity. We first examined general stock pricing within the state price density approach. A main problem in implementation are dividend payments. Since dividends contribute a sizeable part to overall long-term stock returns, dividend payments should not be omitted. We found that the product of the stock price and the state price density $(S_t \varsigma_t)$ must be a supermartingale in all cases. If no dividends are payed, $(S_t \varsigma_t)$ is even a martingale.

We derived and implemented two approaches to include stock price data. First, in section 3.1.3, we use a dividend discount approach which interprets the stock price at time t as the value of all future dividends discounted to time t. Based upon economic considerations we presented a simple model for a discrete dividend process. The state price density was then used to discount the dividends, providing a stock market model which links the bond market as defined by the state price density and the stock market.

We found that the dividend discount model is unfeasible to be implemented in the Cairns framework due to computational limitations. For the cosh model, on the other side, it provides an arbitrage-free, implementable stock pricing framework. We showed how to implement an Extended Kalman filter for joint estimation of both interest rates and the stock price. Whereas the estimation approach still requires additional work, particularly considering the inclusion of dividend data into estimation, we were able to derive a parameter set of a three-factor model which provides a historical fit of the bond market similar to two-factor cosh models, but also fits the $S\&P500$ price index from 1985 to 2008 by a mean absolute error of merely 12 ticks.

The second stock market approach, presented in section 3.1.4, is based on the Black-Scholes stock market model with stochastic shortrate. Taking the shortrate as provided by the cosh model under the risk-neutral measure, we derived a closed form solution to the stochastic differential equation under the risk-neutral measure and under the physical measure. As this closed-form solution requires an integral of the shortrate from the initial time point until the current time, we recommended using return data rather than price data, decreasing the approximation problem. Again, we provided an Extended Kalman filter to jointly estimate bond and stock market dynamics. The Black-Scholes based approach was vastly superior computationally to the dividend discount approach, yet provided inferior historical fit. The approach allows for simple specification of correlations among the state vectors driving the market, which can also be partitioned into pure bond market, pure stock market and joint state vector components according to restrictions upon the parameter vectors σ and γ. We estimated an approach assuming general γ and a second approach restricting γ to get a pure stock market state vector component. Whereas a pure stock market component provides a near perfect fit of historical stock returns, Loglikelihood values are substantially lower than in the unrestricted case due to stock returns not being normal. The Black-Scholes approach therefore suffers from the same problems as the original Black-Scholes model. An important task for future research therefore is to examine the impact of stochastic volatility in insurance applications and implementing stochastic volatility in the joint model, which would enable us to examine dependencies between stochastic volatility in bond and stock markets.

Finally, we discussed the role of macroeconomic variables for term structure modeling. Again, the cosh model allows for a simple implementation of a macro-finance model. We also discuss the implementation of monetary policy rules as required for simulation purposes. Based upon results from the literature, we can conclude that introduction of macroeconomic data should improve long-term dynamics of the term structure, and in particular should improve term structure dynamics around the zero lower bound. The index-based approach presented allows for simple pricing formulae for inflation-indexed bonds. We also discuss possible applications of macroeconomic variables in insurance besides the improvement of term structure dynamics.

The work presented opens several directions for future research. For once, we argued that stochastic volatility should be of minor importance to insurance applications if the term structure model catches overall variability in the principal components. Nevertheless, this assumption should be tested empirically. Furthermore, the joint bond and stock market model provide interesting opportunities to examine the effect of interest rate risk on stock derivatives and the mutual dependencies of stochastic volatility in bond and stock market models.

Another line of future research should examine macroeconomic variables in insurance applications. Ideas already mentioned are guaranteed real returns, hedging by macroeconomic derivatives, investment into inflation-indexed bonds, examination of cross-asset correlation due to macroeconomic variables and modeling of customer behavior based upon these macroeconomic variables. Finally, note that a joint model which provides term structure dynamics, macroeconomic variables and stock market data should find interesting applications in monetary policy as well.

Finally, as already discussed in section 2.3.9, recent discussion of a regulatory approach considering the asymptotic long rate requires term structure models which allow for an analysis of the regulatory specifications as well as sensitivity analysis of insurance products on the asymptotic long rates. Both the Cairns and the cosh models are particularly applicable in these cases.

Appendix A

Appendix

A.1 The Ito-Doeblin formula

The Ito-Doeblin formula used throughout this work is taken from Oksendal [Oks06].

Theorem A.1.1. *Let* (X) *be a d-dimensional stochastic process with dynamics*

$$dX_t^{(i)} = u_i(t, X_t)dt + \sum_{j=1}^{d} v_{ij}(t, X_t)dZ_t^{(j)}$$

for all $i = 1, \dots, d$ *and* $Z = (Z_1, \dots, Z_d)$ *a d-dimensional Brownian motion under some measure* P *and*

$$P\left[\int_0^t |u_i(s, X_s)|ds < \infty \ \forall t \geq 0\right] = 1,$$

for all $i = 1, \dots, d$ *and* u_i *is* $\mathcal{F}_t = \sigma(B_s, s \leq t)$ *adapted. Furthermore*

$$P\left[\int_0^t |v_{ij}(s, X_s)^2|ds < \infty \ \forall t \geq 0\right] = 1,$$

for all $i = 1, \dots, d$*. If* f *is a twice continuously differentiable function from* $[0.\infty) \times \mathbb{R}^d$ *to* $\mathbb{R}^n$*, then the process* $f(t, X_t)$ *is again a stochastic integral whose component dynamics are given by*

$$\begin{aligned} df_k(X_t) = & \Bigg[\frac{\partial}{\partial t} f_k(t, X_t) + \sum_{i=1}^{d} u_i(t, X_t)\frac{\partial}{\partial x_i} f_k(t, X_t) \\ & + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\sigma(t, X_t)\sigma(t, X_t)^T\right)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f_k(t, X_t)\Bigg] dt + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} f_k(t, X_t)dZ_t^{(i)}. \end{aligned}$$

Proof. See Oksendal [Oks06], page 96. □

Note that for

$$dX_t^{(i)} = \kappa_i(\mu_i - X_t^{(i)})dt + \sum_{j=1}^{d} C_{ij}dZ_t^{(j)}$$

we have

$$P\left[\int_0^t |\kappa_i(\mu_i - X_s)|ds < \infty \ \forall t \geq 0\right] = 1$$

due to mean reversion of (X) and for all $i = 1, \ldots, d$ and u_i is $\mathcal{F}_t = \sigma(B_s, s \leq t)$ adapted. Furthermore

$$P\left[\int_0^t |C_{ij}^2|ds < \infty \ \forall t \geq 0\right] = 1,$$

since C_{ij} is a constant for all $i, j = 1, \ldots, d$ so that the Ito-Doeblin formula holds for all cases considered in this work.

A.2 The Dynkin formula

The Dynkin formula may be used to derive instantaneous forward rates of the Rogers model in section 2.2.1.

Theorem A.2.1. *Let f be twice continuously differentiable. Suppose τ is a stopping time with $E[\tau|\mathcal{F}_0] < \infty$. Then*

$$E[f(X_\tau)|X_0 = x] \quad = \quad f(x) + E\left[\int_0^\tau Gf(X_s)ds \middle| X_0 = x\right]$$

where the generator G *of an Ito-diffusion*

$$dX_t \quad = \quad \mu(X_t)dt + \sigma(X_t)dZ_t$$

is given by

$$Gf(x) \quad = \quad \sum_{i=1}^d \mu_i(x)\frac{\partial}{\partial x_i}f(x) + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \left(\sigma(x)\sigma(x)^T\right)_{ij}\frac{\partial^2}{\partial x_i \partial x_j}f(x).$$

Proof. See again Oksendal [Oks06], page 105. □

Corollary A.2.2. *Let f be twice continuously differentiable and G be the generator of an Ito-diffusion X. Then $E[f(X_t)|X_t = x]$ is differentiable with respect to t and*

$$\frac{\partial}{\partial t}E[f(X_t)|X_0 = x] \quad = \quad E\left[Gf(X_t)|\, X_0 = x\right].$$

Proof. Choosing $\tau = t$ in Dynkin's formula we see that $E[f(X_t)|X_t = x]$ is differentiable with respect to t and

$$\begin{aligned}\frac{\partial}{\partial t}E[f(X_t)|X_t = x] &= \lim_{h\to 0}\frac{E\left[\int_0^{t+h} Gf(X_s)ds\middle| X_0 = x\right] - E\left[\int_0^t Gf(X_s)ds\middle| X_0 = x\right]}{h} \\ &= E\left[Gf(X_s)ds|\, X_0 = x\right].\end{aligned}$$

□

Bibliography

[ACS99] ANDERSEN, Torben G. ; CHUNG, Hyung-Jin ; SORENSEN, Bent E.: Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. In: *Journal of Econometrics* (1999), Nr. 91, S. 61 – 87

[Alb07] ALBRECHT, Peter: Einige Überlegungen zur simultanen Modellierung von Aktienindex und Zinsstruktur. In: *Mannheimer Manuskripte zu Risikotheorie, Portfolio Management und Versicherungswirtschaft* (2007), Nr. 170

[AP03] ANG, Andrew ; PIAZZESI, Monika: A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. In: *Journal of Monetary Economics* 50 (2003), May, Nr. 4, S. 745–787

[Ape03] APEDJINOU, Kodjo M.: What Drives Interest Rate Swap Spreads? An Empirical Analysis of Structural Changes and Implications for Modeling the Dynamics of the Swap Term Structure. In: *Working paper* (2003)

[BB92] BERNANKE, Ben S. ; BLINDER, Alan S.: The Federal Funds Rate and the Channels of Monetary Transmission. In: *American Economic Review* 82 (1992), September, Nr. 4, S. 901–921

[BDT90] BLACK, Fischer ; DERMAN, Emanuel ; TOY, William: A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options. In: *Financial Analysts Journal* 46 (1990), Nr. 1, S. 33–39

[BEG97] BALDUZZI, Pierluigi ; ELTON, Edwin J. ; GREEN, T. C.: Economic News and the Yield Curve: Evidence from the U.S. Treasury Market / New York University, Leonard N. Stern School of Business. 1997 (98-005). – Working paper

[Bjö98] BJÖRK, Thomas: *Arbitrage Theory in Continuous Time.* Oxford University Press, 1998

[BK91] BLACK, Fischer ; KARASINSKI, Piotr: Bond and Option Pricing When Short Rates are Lognormal. In: *Financial Analysts Journal* (1991), Nr. July-August 1991, S. 52–59

[BM01] Brigo, Damiano ; Mercurio, Fabio: *Interest Rate Models: Theory and Practice.* Springer, 2001

[BMD10] Blanchard, Olivier ; Mauro, Paolo ; Dell'Ariccia, Giovanni: *Rethinking Macroeconomic Policy.* Version: January 2010. http://ssrn.com/abstract=1555117

[BRS04] Bernanke, Ben S. ; Reinhart, Vincent R. ; Sack, Brian P.: Monetary policy alternatives at the zero bound: an empirical assessment / Board of Governors of the Federal Reserve System (U.S.). 2004 (2004-48). – Finance and Economics Discussion Series

[BS73] Black, Fischer ; Scholes, Myron: The Pricing of Options and Corporate Liabilities. In: *The Journal of Political Economy* 81 (1973), Nr. 3, S. 637–654

[CA93] Campbell, John Y. ; Ammer, John: What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns. In: *The Journal of Finance* 48 (1993), Nr. 1, S. 3–37

[Cai04a] Cairns, Andrew J. G.: A Family Of Term-Structure Models For Long-Term Risk Management And Derivative Pricing. In: *Mathematical Finance* 14 (2004), Nr. 3, S. 415–444

[Cai04b] Cairns, Andrew J.: *Interest Rate Models: An Introduction.* Princeton University Press, 2004

[CDG02] Collin-Dufresne, Pierre ; Goldstein, Robert S.: Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. In: *The Journal of Finance* 57 (2002), Nr. 4, S. 1685–1730

[CGG97] Clarida, Richard ; Galí, Jordi ; Gertler, Mark: Monetary Policy Rules in Practice: Some International Evidence / C.E.P.R. 1997 (1750). – CEPR Discussion Papers

[(Ch02] (Chairman), Jim S.: Economic Benefits from U.S. Treasury Securities / Joint Economic Committee, United States Congress. 2002. – Joint Economic Committee Study

[Che03] Chernov, Mikhail: Empirical reverse engineering of the pricing kernel. In: *Journal of Econometrics* 116 (2003), Nr. 1-2, S. 329–364

[Chr01] Christiansen, Charlotte: Long Maturity Forward Rates / University of Aarhus, Aarhus School of Business. 2001 (01-12). – Finance Working Papers

[CIR85] Cox, John C. ; Ingersoll, Jonathan E. ; Ross, Stephen A.: A Theory of the Term Structure of Interest Rates. In: *Econometrica* 53 (1985), Nr. 2, S. 385–407

[CL02] Christiansen, Charlotte ; Lund, Jesper: Revisiting the shape of the yield curve: the effect of interest rate volatility / University of Aarhus, Aarhus School of Business, Department of Business Studies. 2002 (2002-3). – Finance Working Papers

[CS92] Chen, Ren-Raw ; Scott, Louis O.: Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure. In: *Review of Financial Studies* 5 (1992), Nr. 4, S. 613–36

[DIR96] Dybvig, Philip H. ; Ingersoll, Jonathan E. ; Ross, Stephen A.: Long Forward and Zero-Coupon Rates Can Never Fall. In: *Journal of Business* 69 (1996), January, Nr. 1, S. 1–25

[DK96] Duffie, D. ; Kan, R.: A Yield-Factor Model of Interest Rates. In: *Mathematical Finance* 6 (1996), Nr. 4, S. 379–406

[DL03] Dewachter, Hans ; Lyrio, Marco: Macro Factors and the Term Structure of Interest Rates / Katholieke Universiteit Leuven, Centrum voor Economische Studiën. 2003 (ces0304). – Discussion papers

[DRA04] Diebold, Francis X. ; Rudebusch, Glenn D. ; Aruoba, S. B.: The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach / National Bureau of Economic Research, Inc. 2004 (10616). – NBER Working Papers

[DS00] Dai, Qiang ; Singleton, Kenneth J.: Specification Analysis of Affine Term Structure Models. In: *Journal of Finance* 55 (2000), October, Nr. 5, S. 1943–1978

[DS01] Dai, Qiang ; Singleton, Kenneth J.: Expectation Puzzles, Time-varying Risk Premia, and Dynamic Models of the Term Structure / National Bureau of Economic Research, Inc. 2001 (8167). – NBER Working Papers

[DS04] Duffee, Gregory R. ; Stanton, Richard H.: Estimation of Dynamic Term Structure Models / Haas School of Business, University of California. 2004. – Working paper

[Dua04] Duarte, Jefferson: Evaluating an Alternative Risk Preference in Affine Term Structure Models. In: *Review of Financial Studies* 17 (2004), Nr. 2, S. 379–404

[Duf96] Duffie, Darrell: Special Repo Rates. In: *The Journal of Finance* 51 (1996), Nr. 2, S. 493–526

[Duf00] Duffee, Gregory R.: Term premia and interest rate forecasts in affine models. (2000), Nr. 2000-19

[EH91] Estrella, Arturo ; Hardouvelis, Gikas A.: The Term Structure as a Predictor of Real Economic Activity. In: *Journal of Finance* 46 (1991), June, Nr. 2, S. 555–76

[EM97] Estrella, Arturo ; Mishkin, Frederic S.: The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank. In: *European Economic Review* 41 (1997), July, Nr. 7, S. 1375–1401

[EM01] Evans, Charles L. ; Marshall, David: Economic determinants of the nominal treasury yield curve / Federal Reserve Bank of Chicago. 2001 (WP-01-16). – Working Paper Series

[Fam84] Fama, Eugene F.: Term premiums in bond returns. In: *Journal of Financial Economics* 13 (1984), December, Nr. 4, S. 529–546

[FGR03] Fan, Rong ; Gupta, Anurag ; Ritchken, Peter: Hedging in the Possible Presence of Unspanned Stochastic Volatility: Evidence from Swaption Markets. In: *The Journal of Finance* 58 (2003), Nr. 5, S. 2219–2248

[FH96] Flesaker, Bjorn ; Hughston, Lane: Positive Interest. In: *Risk* (1996), January, S. 46–49

[Fil09] Filipovic, Damir: *Term-Structure Models*. Springer, 2009

[Fis30] Fisher, Irving: *The Theory of Interest*. New York : The Macmillan Co., 1930

[Fis02] Fisher, Mark: Special repo rates: an introduction. In: *Economic Review* (2002), S. 27–43

[FMW03] Fischer, Tom ; May, Angelika ; Walther, Brigitte: Anpassung eines CIR-1-Modells zur Simulation der Zinsstrukturkurve. In: *Blätter der DGVFM* 26 (2003), November, Nr. 2, S. 193–206

[FMW04] Fischer, Tom ; May, Angelika ; Walther, Brigitte: Anpassung eines CIR-k-Modells zur Simulation der Zinsstrukturkurve. In: *Blaetter der DGVFM* 26 (2004), May, Nr. 3, S. 269–387

[GO97] Green, Richard C. ; Odegaard, Bernt A.: Are There Tax Effects in the Relative Pricing of U.S. Government Bonds? In: *Journal of Finance* 52 (1997), June, Nr. 2, S. 609–33

[Gor59] Gordon, M.J.: Dividends, Earnings, and Stock Prices. In: *The Review of Economics and Statistics* 41 (1959), Nr. 2, S. 99–105

[GR06] Graziano, Giuseppe D. ; Rogers, L.C.G: Hybrid Derivatives Pricing under the Potential Approach / University of Cambridge. 2006. – Working paper

[GSS03] Gürkaynak, Refet S. ; Sack, Brian ; Swanson, Eric: The excess sensitivity of long-term interest rates: evidence and implications for macroeconomic models / Federal Reserve Board. 2003 (2003-50). – Finance and Economics Discussion Series Working Paper

[GT96] Gallant, A. R. ; Tauchen, George: Which Moments to Match? In: *Econometric Theory* 12 (1996), Nr. 4, S. 657–681

[Har88] Hardouvelis, Gikas A.: The Predictive Power of the Term Structure During Recent Monetary Regimes. In: *The Journal of Finance* 43 (1988), Nr. 2, S. 339–356

[Har91] Harvey, Andrew C.: *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1991 (Cambridge Books 9780521405737)

[HKT02] Hubalek, Friedrich ; Klein, Irene ; Teichmann, Josef: A general proof of the Dybvig-Ingersoll-Ross-Theorem: Long forward rates can never fall. In: *Mathematical Finance* 12 (2002), October, Nr. 4, S. 447–451

[Hul00] Hull, John C.: *Options, Futures, and other Derivatives*. 4. Prentice Hall, 2000

[HW90] Hull, John ; White, Alan: Pricing Interest-Rate-Derivative Securities. In: *The Review of Financial Studies* 3 (1990), Nr. 4, S. 573–592

[HW01] Heidari, Massoud ; WU, Liuren: Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates? / Graduate School of Business, Fordham University. 2001 (0207013). – Working paper

[JLZ07] Jarrow, Robert ; Li, Haitao ; Zhao, Feng: Interest Rate Caps "Smile" Too! But Can the LIBOR Market Models Capture It? In: *Journal of Finance* 62 (2007), February, Nr. 1, S. 345–382

[Kel01] Kellerhals, B. P.: *Financial pricing models in continuous time and Kalman filtering*. Springer, Berlin, Deutschland, 2001 (Lecture notes in economics and mathematical systems). – 247 S. – ISBN 3–540–42364–8

[KK01] Korn, Ralf ; Korn, Elke: *Optionsbewertung und Portfolio-Optimmierung*. 2. Vieweg+Teubner, 2001

[KO07] Kim, Don H. ; Orphanides, Athanasios: The bond market term premium: what is it, and how can we measure it? In: *BIS Quarterly Review* (2007), June

[KP99] Kloeden, Peter E. ; Platen, Eckhard: *Stochastic Modelling and Applied Probability*. Bd. 23: *Numerical Solution of Stochastic Differential Equations*. Springer, 1999

[KS91] Karatzas, Ioannis ; Shreve, Steven E.: *Graduate Texts in Mathematics*. Bd. 113: *Brownian Motion and Stochastic Calculus*. 2. Springer, 1991

[KS04] Kothari, S. P. ; Shanken, Jay: Asset Allocation with Inflation-Protected Bonds. In: *Financial Analysts Journal* 60 (2004), Nr. 1, S. 54–70

[LLM02] Liu, Jun ; Longstaff, Francis A. ; Mandell, Ravit E.: The Market Price of Credit Risk: An Empirical Analysis of Interest Rate Swap Spreads / National Bureau of Economic Research, Inc. 2002 (8990). – NBER Working Papers

[Lon02] Longstaff, Francis A.: The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices / National Bureau of Economic Research, Inc. 2002 (9312). – NBER Working Papers

[LS91] Litterman, Robert ; Scheinkman, Jose: Common Factors Affecting Bond Returns. In: *Journal of Fixed Income* (1991), June

[LSCS01] Longstaff, Francis A. ; Santa-Clara, Pedro ; Schwartz, Eduardo S.: The Relative Valuation of Caps and Swaptions: Theory and Empirical Evidence. In: *Journal of Finance* 56 (2001), December, Nr. 6, S. 2067–2109

[Lut07] Lutz, Matthias: *Die Cairns-Familie von Zinsstrukturmodellen: Kalibrierung, Analyse und Anwendungen im Risikomanagement*, Universität Ulm, Diplomarbeit, 2007

[LW99] Leippold, Markus ; Wu, Liuren: The Potential Approach to Bond and Currency Pricing. (1999), March. `http://ssrn.com/abstract=155430`

[McC00] McCulloch, J. H.: Long Forward and Zero-Coupon Rates Indeed Can Never Fall, but Are Indeterminate: A Comment on Dybvig, Ingersoll and Ross / Ohio State University, Department of Economics. 2000 (00-12). – Working paper

[Mis93] Mishkin, Frederic S.: Is the Fisher Effect for Real? A Reexamination of the Relationship Between Inflation and Interest Rates / National Bureau of Economic Research, Inc. 1993 (3632). – NBER Working Papers

[MK93] McCulloch, J. H. ; Kwon, Heon-Chul ; Ohio State University (Hrsg.): *U.S. term structure data, 1947-1991*. Columbus, Ohio: Ohio State University, 1993. `http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm`

[MR05] Musiela, Marek ; Rutkowski, Marek: *Stochastic Modelling and Applied Probability*. Bd. 36: *Martingale Methods in Financial Modeling*. 2. Springer, 2005

[Oks06] OKSENDAL, Bernt: *Stochastic Differential Equations*. 3. Springer, 2006 (Universitext)

[Reb02] REBONATO, Riccardo: *Modern Pricing of Interest-Rate Derivatives: The Libor Market Model and Beyond*. Princeton University Press, 2002

[Rog97] ROGERS, L.C.G.: The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates. In: *Journal of Mathematical Finance* 7 (1997), Nr. 2, 157-176. `http://dx.doi.org/10.1111/1467-9965.00029`

[RSM04] REITZ, Stefan ; SCHWARZ, Willi ; MARTIN, Marcus R.: *Zinsderivate*. Vieweg, 2004

[Rud95] RUDEBUSCH, Glenn D.: Federal Reserve interest rate targeting, rational expectations, and the term structure / Federal Reserve Bank of San Francisco. 1995 (1995-02). – Working Papers in Applied Economic Theory

[Rut97] RUTKOWSKI, Marek: A note on the Flesaker-Hughston model of the term structure of interest rates. In: *Applied Mathematical Finance* (1997), S. 151–163

[RW96] ROBERDS, William ; WHITEMAN, Charles H.: Endogenous term premia and anomalies in the term structure of interest rates: explaining the predictability smile / Federal Reserve Bank of Atlanta. 1996 (1996-11). – Working paper

[Sad09] SADR, Amir: *Interest Rate Swaps and Their Derivatives: A Practitioner's Guide*. 1. John Wiley and Sons, 2009

[SB00] SCHAEFER, Stephen M. ; BROWN, Roger H.: Why Long Term Forward Rates (Almost) Always Slope Downwards / London Business School. 2000 (299). – Working paper

[Sch08] SCHULZE, Klaas: Asymptotic Maturity Behavior of the Term Structure / University of Bonn, Germany. 2008. – Bonn Econ Discussion Papers

[Shr04] SHREVE, Steven E.: *Stochastic Calculus for Finance II*. 1. Springer, 2004 (Springer Finance)

[SSW93] SUN, Tong sheng ; SUNDARESAN, Suresh ; WANG, Ching: Interest rate swaps: An empirical investigation. In: *Journal of Financial Economics* 34 (1993), August, Nr. 1, S. 77–99

[Tay] TAYLOR, John B.: The Financial Crisis and the Policy Responses: An Empirical Analysis of What Went Wrong / National Bureau of Economic Research, Inc (14631). – NBER Working Papers

[Tay93] Taylor, John B.: Discretion versus policy rules in practice. In: *Carnegie-Rochester Conference Series on Public Policy* 39 (1993), December, Nr. 1, S. 195–214

[TW08] Taylor, John B. ; Williams, John C.: A Black Swan in the Money Market / National Bureau of Economic Research, Inc. 2008 (13943). – NBER Working Papers

[Vas77] Vasicek, Oldrich: An equilibrium characterization of the term structure. In: *Journal of Financial Economics* 5 (1977), November, Nr. 2, S. 177–188

[Wil84] Wilkie, A.D.: Steps towards a comprehensive stochastic model / The Institute of Actuaries. 1984. – Research Discussion Paper

[Wil86] Wilkie, A.D.: A stochastic investment model for actuarial use. In: *Transactions of the Faculty of Actuaries* 39 (1986), S. 341–403

[Wil95] Wilkie, A.D.: More on a stochastic asset model for actuarial use. In: *British Actuarial Journal* 1 (1995), Nr. 5, S. 777–964

[WR04] Wu, Tao ; Rudebusch, Glenn: A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy / Society for Economic Dynamics. 2004 (104). – 2004 Meeting Papers

[Yao98] Yao, Yong: Term structure models: a perspective from the long rate / University of Groningen, Research Institute SOM (Systems, Organisations and Management). 1998 (98C25). – Research Report