STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's Thesis in Economics
Academic Year 2017–2018

Term Structure Modeling near the Zero Lower Bound:

Regime Switching & Monetary Policy

Oliver Wilhelm Krek (41213)

Abstract
This thesis proposes a regime-switching extension to the well known autoregressive gamma and gamma-zero process nesting its linear counterpart. The affine term structure model based on the new process matches key stylized facts of interest rates during a zero lower bound period as well as in normal times. The regime structure uses interest rate option implied densities and builds on smooth transition regression trees, relating to potential Big Data as well as high frequency applications. Results for the US and EU show an improvement in the fit for the yield level, especially at the short end of the term structure.

Keywords: Affine Term Structure Model, Regime-Switching, Monetary Policy, Zero Lower Bound

JEL classification: C58, E43, E52, G12

Supervisor: Rickard Sandberg
Date submitted: 07.05.2018
Date examined: 28.05.2018
Discussant: Alice Dominici, Katherine Perez
Examiner: Mark Sanctuary

Acknowledgment

I would like to thank everybody for their critical input and feedback. Special thanks go to my supervisor Prof. Matthias Fengler for his extremely valuable inputs and comments in finding the topic and throughout writing this thesis as well as Prof. Dr. Rickard Sandberg for his continuous support and valuable discussions throughout the writing process. Last, I would like to thank Prof. Dr. Renne for his inputs and directions related to the estimation of the model.

Content

List of Figures

List of Tables

1 Introduction

The term structure of interest rates has long been a major topic for research in macroeconomics and finance, especially because of its practical importance. Since the seminal publication of the Vasicek model in 1977 much progress has been made from a theoretical as well as empirical perspective. Many different approaches for modeling the term structure have been proposed by researches over the years and much attention has been given to so-called factor models, where interest rates are assumed to be driven by a number of unobservable factors.

Besides the relevance for finance and research, interest rates are a corner stone of the economy and influence households as they do business decisions. This makes them important for society and hence play a major role in economic policy, especially monetary policy. Understanding their behavior is thus vital for conducting successful policy. During the second half of the 20th century, central banks around the world followed various procedures to conduct monetary policy, such as reserve targeting. Only in the late 20th century many central banks started to adopt what is called an inflation targeting procedure, see e.g. Svensson (2010), which involves adjusting the short term rate to inflation news.

The different chairmen and operating procedures of central banks have led to what is known as a monetary policy regime. Such changes in monetary policy can lead to a different behavior of the very short end of the term structure which is generally controlled by the central bank. As the long end of the yield curve is nothing else than a risk adjusted expectation of future short rates this can also lead to changes on the long end of the yield curve.

While monetary policy stayed relatively constant between the appointment of Paul Volcker in 1979 and the recent financial crisis, the latter in 2008 has led many central banks to a dramatic change in their policy by adopting a zero interest rate policy. In research as well as in practice this is commonly referred to as the Zero Lower Bond (ZLB) problem. On the policy side, as interest rates are generally constrained by zero from below, central banks are no longer able to react to a decrease in inflation with a decrease in the prevailing interest rate. Central banks thus started to use different policy tools, summarized under the term unconventional monetary policy.[1]

From a modeling point of view, the sheer existence of the lower bound became a major point of interest as well as an obstacle. Traditionally, so-called Gaussian factor models have been the workhorse model in this area. Theoretically they can have non-zero rates which makes them less applicable on the journey of interest rate modeling. Therefore, research over the past years has focused on zero-lower bound consistent models. Several solutions have been proposed

[1] See for example De Graeve and Lindé (2015) for an overview about the mechanisms of unconventional policy and Di Maggio and Kacperczyk (2017), Cook and Devereux (2016) or Hall (2016) for economic implications of the zero lower bound.

over the years, the most prominent being the Shadow-rate model already introduced by Black (1995). Other approaches include quadratic term structure models (QTSM) or models based on a square-root process.[2] A completely new approach and model was introduced by Monfort et al. (2017). In order to make their model zero lower bound consistent, they develop a new process called Autoregressive Gamma-zero (ARG-Zero) process. Their process is able to cope with three key term structure characteristics: consistency with non-negative yields, availability of closed-form bond pricing solutions and the ability to match extended zero rate periods.

To tackle different shapes of the yield curve due to monetary policy the literature has proposed so-called regime-switching models. These models go back to their introduction by Hamilton (1989). Those regimes are then often linked to economic conditions. Thus, when analyzing term structure models and monetary policy, one usually relies on macroeconomic indicators such as GDP, inflation or an output performance measure which are available at a monthly frequency. This makes them less usable in financial forecasting applications or in a high-frequency context. This raises the question whether there is a possibility to make a term structure model take into account the different monetary policy stances as well as making use of the more granular availability of financial data.

In this thesis, a regime-switching extension to the affine[3] ARG-Zero processes introduced by Monfort et al. (2017) is proposed and needed to overcome two issues. Namely, changes due to monetary policy and the high-frequency context of financial markets. The construction of the monetary regimes relies on financial data and thus can be applied to any frequency. The regime-switching extension builds on the theory of smooth-transition regression trees and is built directly into the process itself. The new process, which I call Regime-Switching ARG (RS-ARG) process nests the original ARG-Zero process as well as the ARG process introduced by Gourieroux and Jasiak (2006) while preserving the affine nature. Moreover, by using the regression tree specification, the new process is able to use any type of input data which can help to improve prediction, relating to possible big data applications. To my knowledge, this is the first combination of a regression tree approach with a latent factor model in the term structure literature.

This new process is subsequently applied to the US and EU as the sudden adoption of a zero interest rate policy following the financial and sovereign debt crisis most likely caused interest rates to exhibit a change in regime. Estimation results suggest that the two countries can be best described by a two regime structure, one featuring high and the other low yields. The 4-factor specification of the term structure model shows an extremely good performance in matching the level of yields with an exceptional improvement at the short end of the yield curve compared

[2] Also called Cox-Ingersoll-Ross type of process.

[3] Affine means linear.

to the linear counterpart. Besides the improvement on the level it provides a natural economic interpretation which links factor dynamics at each point in time to the current market perceived monetary policy stance.

The remainder of this thesis is organized as follows. Section 2 provides the economic background and rationale behind the structure of the model. Section 3 formally develops the new regime-switching extension of the ARG process and its associated term structure model. Section 4 describes the data at hand and the procedure used in obtaining risk-neutral densities. Section 5 presents the estimation strategy and results for the regime structure and term structure model. Section 6 discusses the validity and robustness of the model while Section 7 concludes. The appendix provides some technical details as well as additional Tables and Figures.

2 Economic Background

The following section will provide a brief overview of the economic and financial background related to the topics and research question stated in the previous section. First, the term structure of interest rates and monetary policy will be introduced along with their stylized facts. This will be followed by an overview about the most important results from asset pricing theory and a short overview of the literature about term structure models. The rationale behind regime-switching models will be presented and the section ends with an overview of the information content implied in option contracts.

2.1 The Term Structure of Interest Rates and Monetary Policy

The term structure of interest rates generally relies on a mathematical definition.[4] To better understand this concept it is necessary to introduce some notation such that the following concepts will be more easy to understand. First of all, the notation of a Zero Coupon Bond (ZCB) is one of the crucial concepts to understand. In general, a ZCB pays its holder one unit of account at its maturity T. This in contrast to a coupon bearing bond which pays its holder a coupon payment at fixed points in time. Throughout this thesis, t denotes the time today, T the maturity time and $h := T - t$ the time to maturity. As with any asset, each bond with residual maturity h has an associated price, which will be denoted by $P_t(h)$.[5]

Each price and bond has a continuously compounded yield, which is denoted by $R_t(h)$ and is the basic building block of the term structure of interest rates. The continuously compounded yield is defined as the constant rate at which the price today reaches 1 at maturity. Formally, this

[4] Throughout this subsection I will adopt the notation used in Lemke (2006).

[5] For the sake of simplicity, bond prices used throughout this thesis are assumed default-free.

reads as follows:

$$P_t(h)\exp(h \cdot R_t(h)) = 1,$$

which leads to the following closed form expression for the yield:

$$R_t(h) = -\frac{\ln(P_t(h))}{h}.$$

Furthermore, an additional important concept is the so-called instantaneous short rate. It is defined as the yield of a bond whose time to maturity is converging to zero. In the discrete time-setting adopted in this thesis, the short rate r_t is generally the shortest maturity available and will hence be the rate associated with a bond whose residual maturity is equal to one day. The term structure of interest rates can now be defined as a mapping between the time to maturity h and the corresponding yield (Lemke, 2006, p.7). Thus, the term structure can be seen as a function κ_t which assigns each time to maturity a yield:

$$\kappa_t : [0, T^*] \to \mathbb{R}, \quad h \to \kappa_t(h) = R_t(h)$$

where T^* can be seen as an upper bound $T^* < \infty$ to the time to maturity.
ZCBs can therefore be seen as the basic building block of the term structure and are clearly of particular interest in the term structure modeling literature. In practice however ZCB are not as frequent as one might guess. In fact, bonds traded today are mainly coupon bearing bonds and are further not available for each and every maturity possible. These practical limitations can sometimes be overcome by recognizing that the payments of the coupons can be replicated by a portfolio of zero coupon bonds. This approach provides a possible way to extract zero coupon bond data from traded securities.
Besides their theoretical interest, interest rates in general are of importance to the whole society. For example, interest rates enter in decisions of households, for instance when deciding whether to take up a loan to buy a house. Companies on the other hand take the whole term structure into account when making investment decisions. Furthermore, interest rates are the building block of many traded securities as they are involved in discounting. They also play an important role in areas like forecasting or risk management. This evidently shows why this area is of big relevance to theory as well as to practice.

The previously mentioned instantaneous short rate now provides the link to monetary policy. As central banks throughout the world generally control the overnight rate and hence the short rate of the market, the term structure of interest rates can be influenced by the central bank, at least this is what theory and empirical results tell. This is the reason why monetary policy is of interest to markets as it can drive the whole term structure and hence the price of many assets. Vice versa, it is important for the central bank to understand the dynamics of the term structure to conduct reasonable policy.

The monetary policy of a central bank is often subject to regulations. This means that the state or government determines which goals the central bank has to fulfill and is therefore not the same for each country. Nevertheless, one goal most central banks pursue is that of price stability. Most of the time this is formalized in form of a specific inflation target. This type of policy is commonly referred to as inflation targeting (see for example Svensson (2010)). While the Swiss National Bank (SNB) has formalized the goal of providing a sound economic environment, the Swedish Riksbank pursues the goal of providing an efficient payment system (Sveriges Riksbank (n.d.)). These different goals combined with the broad spectrum of country specific characteristics already indicates that monetary policy varies over countries and hence can impact the term structure of interest rates in various manners.

Mathematically speaking, this can be seen as some sort of non-linearity. Therefore, if a term structure model shall be able to explain the dynamics of many countries over several years, it has to be able to take into account such non-linearities arising from monetary policy.

2.1.1 Stylized Facts

Over the last decades several stylized facts of the term structure of interest rates have emerged. Figure 1 shows the evolution of the usual short rate proxies used by the market for Europe, the United Kingdom and the US, three of the most liquid and important financial markets. Short rate proxies are clearly linked to the economic conditions of each country/area. During the large economic expansion in all countries until the financial crisis all interest rates show a rising tendency. This behavior is in line with the common understanding of how interest rates affect inflation, see for example Taylor (1993). It is evidently clear that during the crisis the central banks around the world were forced to bring down the rates to stimulate the economy.

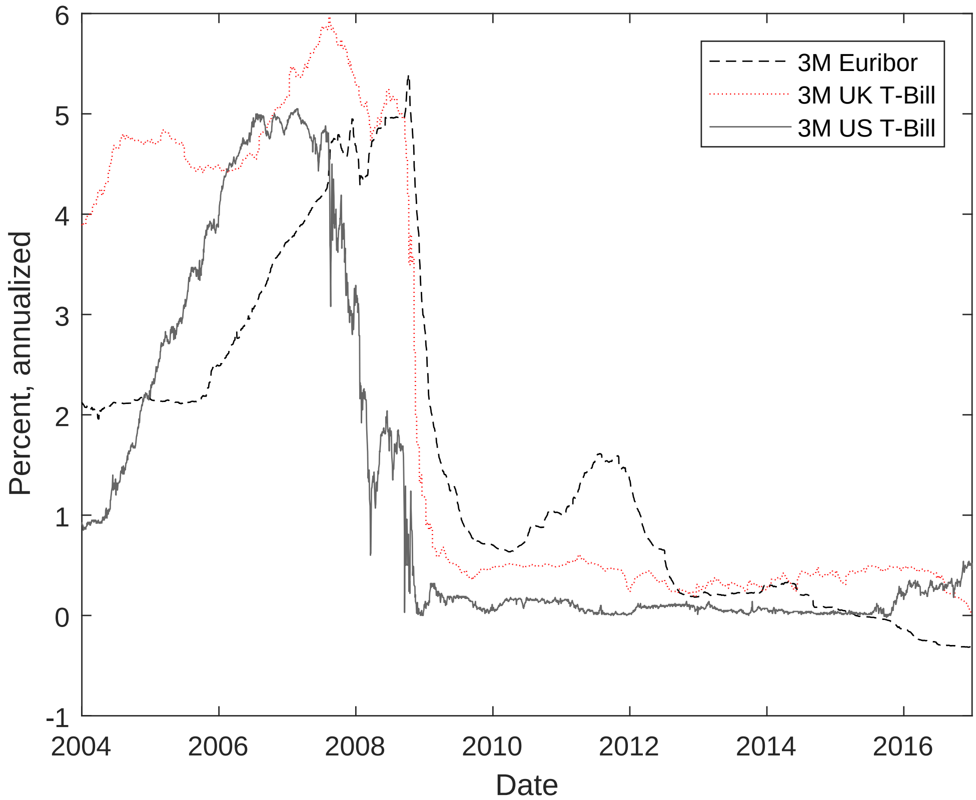


Figure 1: Short Rate Market Proxies

Notes: Data have been obtained from Thomson Reuters Datastream.
Source: Author's rendering of Thomson Reuters data (2017).

Table 1 shows summary statistics for the European zero coupon yields used throughout the thesis. One important stylized fact about yield curves is that they are generally upward sloping in the time to maturity, which can be seen from the mean value. Other possible shapes of the yield curve which frequently have been observed are downward sloping or hump-shaped. This pattern is also economically reasonable as investors generally require a risk compensation for an investment over a longer horizon. This risk premia has become more important in recent years as it became a target in unconventional monetary policy due to short rates being close to their zero lower bound, see for example De Graeve and Lindé (2015). Furthermore, the standard deviation is generally decreasing but not as evident as for the mean. This gives rise to the fact that interest rates are driven by some common factors. Regarding the kurtosis, all yields show a level of less than 3, so the probability of getting outliers is becoming smaller over time.
Last, another important stylized fact of interest rates is that they show an extremely high autocorrelation. Sometimes it is argued that interest rates behave like a unit root process which could cause additional obstacles in modeling interest rates. On the other hand, these high autocorrelations provide a good building block for modeling rates with processes that can feature high autocorrelation.

Table 1: Summary Statistics ECB ZCB Rates

Maturity	Mean	Std.	Skew.	Kurt.	Autocorr. (1st)
3M	0.9924	1.5311	0.7923	2.742	0.9928
1Y	1.0926	1.5909	0.7137	2.1415	0.9921
2Y	1.2466	1.5953	0.5211	1.9962	0.9906
5Y	1.7918	1.5406	0.0208	1.6380	0.9887
7Y	2.1352	1.4933	-0.1907	1.6284	0.9879
10Y	2.5334	1.4345	-0.3928	1.7316	0.9867

Notes: The value of the empirical moments correspond the yields expressed in annualized percentage points.

2.2 Asset Pricing Theory

The field of asset pricing has experienced substantial growth over the last decade. Generally, the basic goal of asset pricing is to find a way to price any possible financial claim. This already provides the link to the term structure literature. Yields, which are the inputs to the term structure are nothing else than sure future payoffs. Therefore, if zero coupon bonds could be prices correctly, then the yield curve can be modeled too. As interest rates are important for virtually any economic application, it is not surprising that many approaches have been developed over the last decades to model the term structure of interest rates.

Regarding the time dimension, asset pricing theory generally distinguishes between discrete and continuous time pricing model. While the former assumes that the information structure and prices arrive at discrete time intervals like days or months, the latter looks at this as a continuously evolving spectrum. The model that will be developed later on relies on the discrete time approach, for a unified treatment of asset pricing in continuous time see Shreve (2004).

Many of the models developed in asset pricing rely on some fundamental results. Before formally stating those results, the concepts arbitrage and the stochastic discount factor will be introduced as they are central to the whole theory.

Loosely speaking, arbitrage is the opportunity to make an almost surely gain without paying something now. Formally, one can define arbitrage as follows:

Definition 2.1. *No arbitrage (NA): A payoff space $\underline{X}$ and a pricing function $p(x)$ leave no arbitrage opportunities if every payoff x that is always nonnegative, $x \geq 0$ (almost surely), and positive, $x > 0$, with some positive probability, has positive price, $p(x) > 0$.*[6]

As money possesses a time-value, which the interest rate can be seen as, discounting future payment streams is used to derive their present value. This result holds true for sure payments in the future but contingent claims usually involve some sort of uncertainty about the exact payment. Furthermore, the economy is generally characterized by heterogeneous agents and these agents are mainly subject to a utility they derive from a future claim. For example, if an

[6]This definition corresponds to the definition in Cochrane (2005), p.61.

investor holds an asset that usually pays when the economy is in a bad state it might be more valuable to him such that he is willing to pay more for such an asset than for one that pays well when he is already of well. This means that when pricing contingent claims, one has to generalize the concept of a discount factor to take the uncertainty and willingness to pay into account.
To do so, the literature has developed the concept of the Stochastic Discount Factor (SDF). The stochastic part now reflects the fact that there is uncertainty and heterogeneity in the market. The SDF should then be able to price any asset. It is therefore not surprising that the center of research in asset pricing has been to identify the functional form of the SDF. On the other hand, as there are not many restrictions on the SDF, many different shapes have been proposed over last decades. Before briefly discussing the functional forms, the concepts of a risk-neutral probability and a change of measure have to be introduced as the model presented later on makes use of both concepts.
All possible future contingent claims are uncertain. There are some more risky and others less. In addition, as risk aversion differs among investors/agents they would not always agree on a fair price if they take into account their preferences. Therefore, one generally prices assets according to their risk-neutral price. This means that each asset which delivers the same expected return will have the same price. As pricing under uncertainty involves taking expectations, the distribution with respect to which the expectation is taking can differ from the risk-neutral one. One generally speaks of the real-world distribution $\mathbb{P}$ and the risk-neutral distribution $\mathbb{Q}$. In order to switch between these distributions, one employs a change of measure which reweighs the probabilities assigned to an outcome, such that some are more likely and others are less such that only the expected return matters for the price.
All the previously introduced concepts can be linked together, creating a flexible framework for asset pricing. The so-called "first fundamental theorem of asset pricing" establishes the following link between no-arbitrage and the risk-neutral distribution:

Theorem 2.1. *If a market model has a risk-neutral probability measure, then it does not admit arbitrage.*

Proof. See Shreve (2004), p.231. □

Furthermore, another core result from asset pricing links the SDF and no-arbitrage:

Theorem 2.2. *A market is arbitrage free iff there exists a positive SDF $M \in \mathcal{X}$ that prices all assets:*

$$p(x) = E[Mx],$$

for all $x \in \mathcal{M}$ and $P(M(s) > 0) > 0$.[7]

[7] $\mathcal{X}$ represents an abstract space which contains the SDF while $p(\cdot)$ represents the pricing function.

Proof. For a proof of this theorem see Cochrane (2005), p.68.[8] □

Due to those ties, two out of those three ingredients have to be specified and the third one follows as a by-product. Different modeling strategies have therefore been developed. Bertholon, Monfort, and Pegoraro (2008) for example provide a unified treatment of asset pricing in discrete time. For the model developed in the next section, the back modeling approach will be used. This means that specifying the risk-neutral dynamics and the SDF such that the historical or real world dynamics are obtained as a by-product.

The only thing that is left concerns the exact functional form of the SDF or the model. Over the years many different forms have been proposed. The reason for this is an additional result from asset pricing, which states that different functional forms of the SDF are equivalent. This has led to models in which the SDF has been derived from utility maximizing agents, and hence involves a ratio of marginal utilities, see for example Le, Singleton, and Dai (2010) who used habit-based preferences in their specification of the SDF.

Nevertheless, the most frequently used approach employs a so-called factor model. Those models assume that returns and prices are driven by a number of unobservable common factors. The reason for this is that it can be shown a linear factor model for the returns is equivalent to a SDF affine in those factors. Referring to the back modeling approach mentioned earlier, this means that one usually defines the dynamics of those factors under the RN-distribution and then determines the functional form of the SDF, which is then affine in those factors. Furthermore, this approach provides great-flexibility combined with tractable solutions which are two further reasons for its usefulness.

2.3 Term Structure Models

While range of proposed models covers many types of processes, the focus here is given to the biggest class of factor models, so-called affine models, where affine refers to a function that consists of a constant plus linear term. The model developed in the next section will also belong to this class. The name results from the fact that in those models the yields are an affine function of the underlying factors, formally:

$$R_t(h) = B_t(h) + A_t(h)'X_t.$$

The affine nature of the yield formula leads to the nice feature of obtaining closed-form solutions of bond-prices. This is one of the most appreciated characteristics of this class of model and clearly the reason why its the most used one. The previously introduced concept of no arbitrage

[8] The proof in Cochrane (2005) is for a slightly different wording of the theorem but the result is the same.

places some cross-section restriction on the yield curve. This as yields have to be risk-adjusted expectations of future short rate, Piazzesi (2010). The affine models are generally able to cope with these cross-section restrictions, especially multi-factor models seem to able to provide different dynamics for different maturities while still fulfilling the no-arbitrage restrictions.
Among the earliest examples are the models due to Vasicek (1977) for a discrete time model and the one proposed by Cox, Ingersoll, and Ross (1985) in continuous time. A milestone in the work of term structure modeling has been done by Duffie and Kan (1996) who provided a unified framework for affine factor models. A similar framework for the discrete-time counterpart has been developed by Le et al. (2010).
As term structure models simultaneously model the time-series dimension as well as the cross-section of yields they do have to fulfill quite a substantial amount of dynamics identified in empirical studies. This has led to various approaches in term structure modeling that aim at replicating some specific dynamics. J. H. Christensen, Lopez, and Rudebusch (2014) for example look more closely at the ability of no-arbitrage term structure models in modeling the yield curve volatility. Other examples that are concerned with the volatility dynamics include Longstaff and Schwartz (1992) or Trolle and Schwartz (2009). Besides volatility, the dynamics of the risk premia have been of major interest to researchers. While many term structure models provide a good fit for the whole yield curve, they quite often place restrictive assumptions on the risk premia. Therefore, many models like Monfort and Pegoraro (2007) have introduced models that allow for more flexible dynamics on the risk premia. Last, with macroeconomic data becoming more and more accurate and available, the macro-finance literature has seen quite substantial growth over last years, see for example Jardet, Monfort, and Pegoraro (2013) or Roussellet (2016) and Cochrane (2017) for an overview.
With the persistence of short rates close to the zero lower bound over the last years, the demand for term structure models that are able to replicate those dynamics has grown substantially. As some specific risk-neutral dynamics for the factors are generally assumed, the ability of model to replicate the dynamics around zero heavily relies upon the characteristics of the process chosen. In continuous time, the so-called square-root processes, which respect the zero lower bound, can be used. In discrete time a workhorse model for many years has been the Gaussian Affine Term Structure Model (GATSM). These models unfortunately have a serious flaw when it comes to modeling of rates around zero due to their innate ability of placing a non-zero probability on negative rates. Krippner (2012) provides an extension of the GATSM model more suitable for modeling zero rates to circumvent the issue. Another approach that has gained prominence over the last couple of years is the so-called shadow rate approach, see for example J. H. E. Christensen and Rudebusch (2016). This approach specifies the short-term rate as the maximum between the shadow rate and the lower bound, $r_t = max\{s_t, LB\}$. Thereby one is able to get a short rate

staying at zero while still allowing for conditional volatility on longer rates. Recently, Monfort et al. (2017) introduced a new type of affine model which respects the zero lower bound by modifying the non-central Gamma distribution in a way that it features an explicit probability mass for staying at zero. As the Gamma distribution is defined on the positive real line, there is no probability mass for getting negative rates. Extending the model to allow for negative rates depending on the state of the economy might be a possible future extension. This new process allows for closed-form pricing formulas, volatility on longer rates as well as the possibility to explicitly calculate lift-off probabilities.

2.4 Regimes and Structural Breaks

The way how monetary policy is conducted is subject to changes over time. On one hand clearly as the people who are in charge change over time and on the other hand changing economic conditions call for different policy measures. In the literature, such phenomena are known as regime switches or structural breaks, respectively. In the context of term structure models, the key player is clearly the central bank. As they generally control the short rate and as long-term rates are risk-neutral expectations of future short rates the shape of the yield curve is partially controlled by the central bank and therefore subject to changes in monetary policy. The way how monetary policy is conducted thereby introduces possible non-linear dynamics into the short rate process. To model these non-linearities, the literature has developed regime-switching models which allow the parameters of the model to change depending on the prevailing regime. Bikbov and Chernov (2013) on the other hand provide evidence that the whole term structure should be used when identifying changes in monetary policy regimes.

Regarding regime-switching models, the most widely used approach is due to Hamilton (1989) and the resulting model is called a Markov-Switching model. It is assumed that a hidden state variable determines the regimes as well as the transition probabilities. In the case of term structure modeling and bond pricing, Gourieroux, Monfort, Pegoraro, and Renne (2014) provide a unified framework for regime switching models.

From a mathematical perspective, a difference is drawn between hard and smooth splits. If a hard switch approach is used, then there is usually an indicator function that determines the dynamics while with smooth change, as the name suggests, the economy transitions over time between the regimes.

Another possible approach for constructing regime-switching models makes a connection to the statistical learning literature. This approach generally constructs what is called a regression tree where each leaf of the tree determines a regime. The convenient feature of this approach is that one usually does not have to determine the number of limiting regimes beforehand. This means that there is more flexibility in determining the true structure of the economy. Like for the Markov-Switching approach, hard or smooth splits can be modeled. Applications to term structure modeling can be seen in Audrino and Medeiros (2011) or Kameliya, Audrino, and De Giorgi (2014). Both papers provide evidence that regimes can be linked to business cycle conditions, which also is an indicator for monetary policy. This result also makes sense from an economic point of view as there is much agreement on the fact that the behavior of the economy is changing depending on the business cycle.

2.5 Options and their Information Content

Options are among the most important derivatives traded on financial markets. They give the holder the right to execute a certain action at a prespecified maturity date. The seminal work of Black and Scholes (1973) provided the groundwork for much of the research in the determination of option prices. Besides being interesting for the market on its own, they also contain a huge amount of information about possible future states of the economy.

I previously introduced the risk-neutral distribution and as options do nothing else than to represent a future contingent claim, they are also priced according to a risk-neutral distribution. An important result in option pricing is due to Breeden and Litzenberger (1978) and it shows that the second-order strike price derivative of the option pricing function is equal to the underlying risk-neutral distribution up to a scaling factor. Thus, by employing a technique which is explained in Section 4, one can back out the risk-neutral distribution which contains the probabilities the market assign to future outcomes.

Regarding interest rates, some of the most heavily traded options are Eurodollar and Euribor options. These options represent a future contingent claim on a policy rate future, usually the prevailing three month London Interbank Offered Rate (LIBOR) rate which closely follows the rate set by the respective central bank. As the price of the underlying future is indirectly linked to a policy rate, there is also the possibility to extract policy rate proxies from the futures price. Nevertheless, as the short rate is typically constrained by its zero lower bound, the underlying risk-neutral distribution becomes right-skewed as one moves towards the lower bound such that less probability mass is assigned to lower rates. This means that mode and mean from the distribution will differ which in turn means that the futures implied rate is not the most likely anymore. Bauer (2014) for example illustrates the issue from extracting possible future rates in the case of the United States.

Furthermore, having knowledge of the full implied density allows to gauge how uncertain the market is about a future rate. This should help to improve the fitted conditional volatility of yields which many term structure models have a difficulty with as documented in Cieslak and Povala (2016).

Thus, option implied densities contain a vast amount of information which can be used in order to improve the fit and forecasting ability of term structure models. Especially when considering the time-series dimension, the forward-looking nature of options provide a natural framework in this context.

2.6 Building a Model

The term structure of interest rates is the basic building block for many areas and as seen is heavily influenced by the respective monetary policy. It is therefore natural to take nonlinearities arising from different monetary policy regimes into account when trying to build a model that should capture the various shapes and dynamics identified. Thus, by combining the information embedded in options about monetary policy with a smooth transition regression tree approach should help to improve the fit of the term structure model while offering an economic interpretation about the dynamics.

3 A Regime Switching Zero Lower Bound Model

This section formally introduces the new RS-ARG process on which a regime-switching non-negative affine term-structure model (RS-NATSM) is built. The process and the term structure model rely heavily upon notations introduced by Monfort et al. (2017). The regime specification on the other hand builds on the smooth-transition regression literature for which Da Rosa, Veiga, and Medeiros (2008) is a good reference. A similar approach has been applied to term structure modeling by Audrino (2006), Audrino and Medeiros (2011) or Kameliya et al. (2014). In the following section adopting the notation used in Monfort et al. (2017) and Kameliya et al. (2014) helps tractability and their notation also provides a good framework to work out the model.
First, the regime-switching specification will be introduced as it provides the backbone of the later developed process and term structure model.

3.1 Smooth Transition Regression Trees

The previous section presented the economic motivation for using regimes in term structure modeling. The idea embedded in tree-structured smooth transition models is used to explicitly capture monetary policy dynamics. As in Audrino (2006) or Kameliya et al. (2014) the regimes are constructed by using a binary-tree in which every terminal node represents one regime. Each regime is therefore determined by a different threshold-structure in a large predictor space and characterized by regime-specific dynamics for the term structure.

3.1.1 Predictor Space

The predictor space is high-dimensional space which consists of all predictor variables which are selected to represent market expectations about future monetary policy. The variables selected are moments derived from option-implied risk-neutral densities on short-rate futures which have a forward-looking character. According the efficient market hypothesis, all available information is included in the observed market price. This is one key benefit of this approach as this means that the information structure is reflected in the shape of the implied density. In addition, as implied-densities are available at a high-frequency and at least daily, the model overcomes a particular shortcoming of many macro-finance models which have to rely on monthly data for e.g. inflation and are hence not usable in practice.
In what follows, the n-dimensional vector containing all predictor variables is denoted by $\boldsymbol{l}_t = (l_{1t}, l_{2t}, ..., l_{Nt}) \in \mathcal{G} \subseteq \mathbb{R}^N$ and the predictor space is denoted by $\mathcal{G}$.

3.1.2 Tree-Structure

To obtain a binary tree which represents the regime structure the predictor space $\mathcal{G}$ is partitioned into disjoint regions. The partition can formally be represented as:

$$\mathcal{P} = \mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_k,$$

$$\mathcal{G} = \bigcup_{j=1}^{k} \mathcal{R}_j,\ \mathcal{R}_i \cap \mathcal{R}_j = \emptyset,\ i \neq j$$

Each of the $\mathcal{R}_j$ represent one partition cell which has threshold structure define on it and each cell is also equipped with a term structure model that has specif local dynamics.
The root of the tree has position $j = 0$ and the parent nodes are indexed with a position j. Each parent generates a left- and a right-child node which are assigned position $2j + 1$ and $2j + 2$ respectively. As mentioned above, threshold values are used to determine the regimes and therefore each node is associated with a split variable $l_{s_j t} \in \boldsymbol{l}_t$, where $s_j \in \mathbb{S} = 1, 2, ..., N$. In addition, we denote by $\mathbb{J}$ the set of indexes of parent nodes and by $\mathbb{T}$ the set of indexes of terminal nodes. The cardinality of $|\mathbb{T}|$ denotes the number of limiting regimes.
The smooth transition approach used leads to transition probabilities between regimes. These transition probabilities represent the probability of changing the regime based on the current available information. Usually, a membership function is used to calculate those transition probabilities. Following Audrino and Medeiros (2011), Kameliya et al. (2014) and Da Rosa et al. (2008) we define the membership function as follows:

$$B_{\mathbb{J}i}(\boldsymbol{l}_t; \boldsymbol{\theta}) = \prod_{j \in \mathbb{J}} G(l_{s_j t}; \tau_j, c_j)^{(1-n_{i,j})(1+n_{i,j})} [1 - G(l_{s_j t}; \tau_j, c_j)]^{n_{i,j}(1+n_{i,j})/2},$$

with

$$n_{i,j} = \left\{ \begin{array}{l} -1 \text{ if the path to leaf } i \text{ does not include the parent node } j, \\ 0 \text{ if the path to leaf } i \text{ includes the right-child node of the parent node j}, \\ 1 \text{ if the path to leaf } i \text{ includes the left-child node of the parent node j}, \end{array} \right\}$$

The transition function $G(\cdot)$ can take on many functional forms but as it represents a probability, it is bounded between zero and one. The linear logistic function, as in Audrino and Medeiros (2011), Kameliya et al. (2014) and Da Rosa et al. (2008) is used subsequently to keep track and as it offers great flexibility:

$$G(l_{s_j t}; \tau_j, c_j) = \frac{1}{1 + e^{-\tau_j (l_{s_j t} - c_j)}},\ \tau_j \geq 0$$

This function takes on values between zero and one and as the probabilities over all regimes sum up to 1, the following must hold: $\sum_{i \in \mathbb{T}} B_{\mathbb{J}i}(\boldsymbol{l}_t; \theta_i) = 1$. It can be seen that the function involves one parameter, τ. This parameter specifies the sensitivity of the function to deviations from the threshold c_j. In fact, the function specifies the speed of transition between the regimes. If τ is small the function is less sensitive to deviations while when $\tau \to \infty$, the function becomes an indicator function.

3.2 RS-NATSM

Before presenting the regime switching structure in the process the baseline autoregressive Gamma-zero processes is introduced as its understanding is crucial. Therefore, the section relies heavily on the results developed in Monfort et al. (2017).

The ARG-Zero process builds on the Gamma-distribution $\gamma_\nu(\mu)$ which is a distribution defined on the positive real line and characterized by a shape parameter $\nu > 0$ and a scale parameter $\mu > 0$. Its closed form probability density function (pdf) is given by:

$$f_X(x; \nu, \mu) = \frac{\exp(-x/\mu)x^{\nu-1}}{\Gamma(\nu)\mu^\nu} \mathbb{1}_{\{x>0\}}.$$

Crucial for the application and the construction of the Gamma-zero distribution is the fact that $\gamma_\nu(\mu)$ converges in distribution to the so-called Dirac[9] mass at zero when ν goes to zero (Monfort et al., 2017, p.350). The Gamma distribution can be further extended to the non-central Gamma distribution. This distribution arises as a Poisson mixture of Gamma distributions, see Gourieroux and Jasiak (2006). Formally, given a Poisson random variable Z with positive parameter λ, the non-central Gamma distribution $\gamma_\nu(\lambda, \mu)$ is a mixture of $\gamma_{\nu+Z}(\mu)$ distributions, defined on the set of strictly positive real numbers and ν, λ and μ are strictly positive. In general, the pdf of a non-central Gamma distribution is very complicated but the associated Laplace transform is fairly simple and exists in closed form. For $X \sim \gamma_\nu(\lambda, \mu)$, one gets:

$$\varphi_X(u) = \mathbb{E}[\exp(uX)] = \exp\left[-\nu \log(1 - u\mu) + \lambda \frac{u\mu}{1 - u\mu}\right], \text{ for } u < \frac{1}{\mu}.$$

3.2.1 The ARG$_0$ Process

As mentioned above, and following Monfort et al. (2017), this distribution can be adapted to the case $\nu = 0$ if one sees $\gamma_0(\mu)$ as the Dirac distribution at zero. This, by definition, leads to the Gamma-zero distribution which has the desired property of having an explicit probability mass at zero.

[9] Also known as the Dirac delta function; it defines a function that is zero everywhere except at 0 and whose integral over the whole real line is equal to one.

Definition 3.1. [10] *Let X be a non-negative random variable. X follows a Gamma-zero distribution with parameters $\lambda > 0$ and $\mu > 0$, denoted $X \sim \gamma_0(\lambda, \mu)$, if its conditional distribution given $Z \sim \mathcal{P}(\lambda)$ is:*

$$X|Z \sim \gamma_Z(\mu).$$

The p.d.f. and the Laplace transform of X, respectively $f_X(x; \lambda, \mu)$ and $\varphi_X(u; \lambda, \mu)$, are given by:

$$f_X(x; \lambda, \mu) = \sum_{z=1}^{+\infty} \left[\frac{\exp(-x/\mu) x^{z-1}}{(z-1)! \mu^z} \times \frac{\exp(-\lambda)\lambda^z}{z!} \right] \mathbb{1}_{\{x>0\}} + \exp(-\lambda)\mathbb{1}_{\{x=0\}}$$
$$\varphi_X(u; \lambda, \mu) = \exp\left[\lambda \frac{u\mu}{(1-u\mu)}\right] \quad \text{for} \quad u < \frac{1}{\mu}.$$

As stated by Monfort et al. (2017) this new distribution has very desirable properties for zero lower-bound modeling as one has an explicit point mass at zero, $\mathbb{P}(X = 0) = exp(-\lambda)$. This property then carries over to the term structure modeling such that it is able to generate prolonged periods of 0 which is an empirical property observed over the last years in interest rate dynamics.

The associated discrete-time random process (X_t) is called Autoregressive Gamma-zero process and denoted by $\text{ARG}_0(\alpha, \beta, \mu)$, where $\alpha \geq 0, \beta \geq 0, \mu > 0$. Formally:

Definition 3.2. [11] *The random process (X_t) is a $ARG_0(\alpha, \beta, \mu)$ process of order one if the conditional distribution of X_{t+1}, given $\underline{X_t} = (X_t, X_{t-1}, ...)$, is the Gamma-zero distribution:*

$$(X_{t+1}|X_t) \sim \gamma_0(\alpha + \beta X_t, \mu) \quad \text{for} \quad \alpha \geq 0, \beta \geq 0, \mu > 0.$$

The conditional probability density function $f(X_{t+1}|X_t; \alpha, \beta, \mu)$ and the conditional Laplace transform $\varphi_{X,t}(u; \alpha, \beta, \mu)$ of the $ARG_0(\alpha, \beta, \mu)$ process are respectively given by:

$$\begin{aligned} f(X_{t+1}|X_t; \alpha, \beta, \mu) = &\sum_{z=1}^{+\infty} \left[\frac{exp(-X_{t+1}/\mu) X_{t+1}^{z-1}}{(z-1)! \mu^z} \times \frac{exp[-(\alpha + \beta X_t)](\alpha + \beta X_t)^z}{z!} \right] \mathbb{1}_{\{X_{t+1}>0\}} \\ &+ exp(-\alpha - \beta X_t)\mathbb{1}_{\{X_{t+1}=0\}} \end{aligned}$$

$$\begin{aligned} \varphi_{X,t}(u; \alpha, \beta, \mu) &:= \mathbb{E}[\exp(u X_{t+1})|\underline{X_t}] \\ &= \exp\left[\frac{u\mu}{1-u\mu}(\alpha + \beta X_t)\right], \quad \text{for} \quad u < \frac{1}{\mu}. \end{aligned} \tag{1}$$

Given the conditional Laplace transform in equation 1, one sees that the ARG_0 process is a $Car(1)$ process (see Darolles, Gourieroux, and Jasiak (2006)). Car processes generally have extremely nice properties such that they can be easily used to construct a term structure model.

[10] This definition corresponds to definition 2.1 in Monfort et al. (2017).
[11] This definition corresponds to definition 2.2 in Monfort et al. (2017).

In particular, these type of processes allow for closed-form pricing formulas which is of major interest in the term structure literature.[12]

3.2.2 The Regime-Switching ARG_0 Process

A new stochastic process that features regime-switching dynamics and builds on the previously developed framework of binary trees and the $\mathrm{ARG}_0(\alpha, \beta, \mu)$process can now be introduced. There are several ways to introduce regime switching dynamics, for example by having a regime-switching short-rate specification or having a different ARG process in each limiting regime. From a modeling perspective, one straightforward approach is to modify the Poisson variable used in the non-central Gamma distribution and thereby introduce the regime-switching dynamics on a process level. The Poisson variable is especially suited for this as it is well known that the sum of independent Poisson variables is again Poisson distributed, see for example Feller (1950). Each regime is then characterized by its own Poisson variable Z_i with positive parameter $\lambda_i = B_{\mathbb{J},i}(\boldsymbol{l};\boldsymbol{\theta})\Delta_i$. The Poisson parameter in each regime is therefore time-varying due to the weight from the membership function. The random variable $Z = \sum\limits_{i \in \mathbb{T}} Z_i$, which is now a sum of independent Poisson variables, is again Poisson distributed with intensity parameter equal to a regime-weighted sum of the regime-specific intensity. We thereby get a smooth transition parameter in our ARG_0 process which now features regime-switching characteristics.

Furthermore, this approach also allows for reasonable economic interpretations. For example, imagine that one regime is characterized by a low conditional mean and the other by a high one. The regime-limiting Poisson parameter, which is equal to $\lim\limits_{B_{\mathbb{J},i}(\cdot) \to 1} B_{\mathbb{J},i}\lambda_i = \Delta_i$, is high in one and low in the other regime. If one now transitions from the high to the low-level regime, the intensity parameter from the high one gets smaller, which can partially be interpreted as the central bank less frequently making changes in interest rates or the magnitude of the changes decreases, indicating only minor policy changes. This approach therefore naturally lends an economic interpretation which is not often the case with factor models.

The weighted Gamma-zero distribution and the associated random process, which are defined below, build on the outlined Poisson framework. Formally, we can define the weighted Gamma-zero distribution as follows:

Definition 3.3. *Let X be a non-negative random variable. X follows a weighted Gamma-zero distribution with parameters $\overline{\lambda} > 0$ and $\mu > 0$, denoted by $X \sim \gamma_0(\overline{\lambda}, \mu)$, if its conditional distribution given $Z|B_{\mathbb{J},i}(\boldsymbol{l};\theta) \sim \mathbb{P}(\overline{\lambda})$ and $B_{\mathbb{J},i}(\boldsymbol{l};\theta)$ is:*

$$X|Z \sim \gamma_Z(\mu),$$

[12] See for example Monfort and Renne (2014) for an application to defaultable bond pricing.

where:

$$\overline{\lambda} = \sum_{i \in \mathbb{T}} \lambda_i B_{\mathbb{J},i}(\boldsymbol{l}; \theta).$$

The p.d.f. and the Laplace transform of X, respectively $f_X(x; \overline{\lambda}, \mu)$ and $\varphi_X(u; \overline{\lambda}, \mu)$, are given by:

$$f_X(x; \overline{\lambda}, \overline{\mu}) = \sum_{z=1}^{+\infty} \left[\frac{\exp(-x/\mu) x^{z-1}}{(z-1)! \mu^z} \times \frac{\exp(-\overline{\lambda}) \overline{\lambda}^z}{z!} \right] \mathbb{1}_{\{x>0\}} + \exp(-\overline{\lambda}) \mathbb{1}_{\{x=0\}}$$

$$\varphi_X(u; \overline{\lambda}, \mu) = \exp \left[\overline{\lambda} \frac{u\mu}{1-u\mu} \right] \quad \textit{for} \quad u < \frac{1}{\mu}$$

The regime-switching extension conditions on the previously determined membership function and hence the time-varying Poisson parameter is conditionally deterministic and hence introduces no further stochastic patterns into the environment.

Definition 3.4. *Let us consider:*

(i) exogenously given transition probabilities $B_{\mathbb{J},i}(\boldsymbol{l}_t; \boldsymbol{\theta}))$;

The random process (X_t) is a Regime-Switching Autoregressive Gamma-zero, RS-ARG$_0(\overline{\alpha}_t, \overline{\beta}_t, \mu, B_{\mathbb{J},i}(\boldsymbol{l}_t; \boldsymbol{\theta}))$, of order one if the conditional distribution of X_{t+1}, given $\underline{X_t} = (X_t, X_{t-1}, ...)$, is the weighted Gamma-zero distribution:

$$(X_{t+1} | X_t) \sim \gamma_0(\overline{\alpha}_t + \overline{\beta}_t X_t, \mu) \quad \textit{for} \quad \overline{\alpha}_t \geq 0, \mu > 0, \overline{\beta}_t > 0,$$

where we have:

$$\overline{\alpha}_t = \sum_{i \in \mathbb{T}} \alpha_i B_{\mathbb{J}i}(\boldsymbol{l}_t; \boldsymbol{\theta}),$$

$$\overline{\beta}_t = \sum_{i \in \mathbb{T}} \beta_i B_{\mathbb{J}i}(\boldsymbol{l}_t; \boldsymbol{\theta}),$$

The conditional probability density function $f(X_{t+1}|X_t; \overline{\alpha}_t, \overline{\beta}_t, \mu)$ and the conditional Laplace transform $\varphi_{X,t}(u; \overline{\alpha}_t, \overline{\beta}_t, \mu)$ are respectively given by:

$$\begin{aligned} f(X_{t+1}|X_t; \overline{\alpha}_t, \overline{\beta}_t, \mu) = & \sum_{z=1}^{+\infty} \left[\frac{\exp(-X_{t+1}/\mu) X_{t+1}^{z-1}}{(z-1)! \mu^z} \times \frac{\exp[-(\overline{\alpha}_t + \overline{\beta}_t X_t)](\overline{\alpha}_t + \overline{\beta}_t X_t^z)}{z!} \right] \mathbb{1}_{\{X_{t+1}>0\}} \\ & + \exp(-\overline{\alpha}_t - \overline{\beta}_t X_t) \mathbb{1}_{\{X_{t+1}=0\}} \end{aligned}$$

$$\begin{aligned} \varphi_{X,t}(u; \overline{\alpha}_t, \overline{\beta}_t, \mu) & := \mathbb{E}[\exp(u X_{t+1}) | X_t] \\ & = \exp \left[\frac{u\mu}{1-u\mu} (\overline{\alpha}_t + \overline{\beta}_t X_t) \right] \quad \textit{for} \quad u < \frac{1}{\mu}. \end{aligned}$$

Conditional Moments of the RS-ARG$_0$ The derivation of the conditional moments follows along the lines of Monfort et al. (2017). The reason for this is that the new non-linear regime-switching process nests its linear counterpart, such that most of the results still hold. First note that the closed-form availability and the affine nature of the conditional Laplace transform allows for quick and easy computation of conditional and unconditional moments by using the conditional log-Laplace transform.[13]

Proposition 3.1. *Let (X_t) be a RS-ARG$_0$ process. The time-t conditional mean $\mathbb{E}_t(X_{t+1})$ and variance $\mathbb{V}_t(X_{t+1})$ of X_{t+1} given the past values are given by:*

$$\begin{aligned} \mathbb{E}_t(X_{t+1}) &= \overline{\alpha}_t\mu + \overline{\beta}_t\mu X_t = \overline{\alpha}_t\mu + \overline{\rho}_t X_t \\ \mathbb{V}_t(X_{t+1}) &= 2\mu^2\overline{\alpha}_t + 2\mu^2\overline{\beta}_t X_t = 2\mu^2\overline{\alpha}_t + 2\mu\overline{\rho}_t X_t, \end{aligned} \tag{2}$$

where $\overline{\rho}_t = \overline{\beta}_t\mu$ and represents the time-varying autoregressive coefficient.

Proof. The proof is a straightforward extension to time-varying parameters of the proof A.1 from Monfort et al. (2017) □

Corollary 3.1.1. *(X_t) has the following semi-strong AR(1) representation:*

$$X_{t+1} = \overline{\alpha}_t\mu + \overline{\rho}_t X_t + \sqrt{2\mu\mathbb{E}_t(X_{t+1})}\epsilon_{t+1},$$

where (ϵ_t) is a martingale difference sequence with a conditional variance equal to 1.

Proof. Straightforward adaption of Corollary 2.1.1 Monfort et al. (2017). □

The closed-form availability of the first two conditional moments directly leads to the semi-strong AR(1) representation. Furthermore, it is worth noting that this generalization of the ARG$_0$ process retains the key features of the non regime-switching counterpart. These are namely that the time-varying conditional variance is still proportional to the time-varying conditional mean and hence low level environments still feature low conditional volatility. Filipovic, Larsson, and Trolle (2017) have shown that if interest rates are close to their respective zero lower bound, they also feature low volatility. The regime-switching process is therefore able to cope with this empirical property while allowing for more volatile rates when one is far away from the lower bound. It also retains the lower bound for the conditional variance, $2\mu^2\overline{\alpha}_t$, which is now time-varying. The time-varying property is a desirable property, as it allows for an increase in the lower bound variance when the market starts to place a higher probability mass on lifting off and hence introduce more uncertainty into the market. Last, this generalization still allows for an easy estimation via Generalized Methods of Moments (GMM) or Quasi-maximum liklihood (QML).[14]

[13] Also called conditional cumulant generating function.

[14] Monte Carlo simulation suggests a quick convergence of the parameter estimates from either regime. The QML procedure is able to correctly identify the different limiting autoregressive characteristics.

3.3 The Extended RS-ARG$_0$ Process

The ARG$_\nu(\beta,\mu)$ from Gourieroux and Jasiak (2006), the ARG$_0(\alpha,\beta,\mu)$ and the RS-ARG$_0(\overline{\alpha}_t$, $\overline{\beta}_t,\mu)$ can all be seen as nested in a general class of regime-switching extended ARG$_\nu(\overline{\alpha}_t,\overline{\beta}_t,\mu)$ processes which now also includes a degree of freedom parameter ν that is not regime switching. Using the framework developed above, the process on which the latter regime-switching term structure model is built will now formally be introduced.

Definition 3.5. *Let us consider:*

(i) exogenously given transition probabilities; $B_{\mathbb{J},i}(\boldsymbol{l}_t;\boldsymbol{\theta}))$.

The univariate random process (X_t) *is an Extended Regime-Switching Autoregressive Gamma process, ERS-ARG*$_\nu(\overline{\alpha}_t,\overline{\beta}_t,\mu,B_{\mathbb{J},i}(\boldsymbol{l}_t;\boldsymbol{\theta}))$*, of order one if the conditional distribution of* X_{t+1}*, given* $\underline{X_t}=(X_t,X_{t-1},...)$*, is a non-centered Gamma distribution with time-varying parameters, such that:*

$$(X_{t+1}|X_t)\sim\gamma_\nu(\overline{\alpha}_t+\overline{\beta}_tX_t,\mu)$$

for $\overline{\alpha}_t\geq 0,\nu\geq 0,\mu>0,\overline{\beta}_t>0$.

The conditional probability density function and the Laplace transform of the ERS-ARG process are respectively given by:

$$\begin{aligned}
f_X(x;\nu,\overline{\alpha}_t,\overline{\beta}_t,\mu)=&\sum_{z=1}^{+\infty}\left[\frac{\exp(-X_{t+1}/\mu)X_{t+1}^{\nu+z-1}}{(\nu+z-1)\mu^{\nu+z}}\right.\\
&\left.\times\frac{\exp(-(\overline{\alpha}_t+\overline{\beta}_tX_t))(\overline{\alpha}_t+\overline{\beta}_tX_t)^z}{z!}\right]\mathbb{1}_{\{X_{t+1}>0\}}\\
&+\exp(-\overline{\alpha}_t-\overline{\beta}_tX_t)\mathbb{1}_{\{X_{t+1}=0,\nu=0\}}\\
\varphi_{X,t}(u;\nu,\overline{\alpha}_t,\overline{\beta}_t,\mu):=&\mathbb{E}\left[\exp(uX_{t+1})|\underline{X_t}\right]\\
=&\exp\left[\frac{u\mu}{1-u\mu}\overline{\beta}_tX_t+\overline{\alpha}_t\frac{u\mu}{1-u\mu}-\nu\log(1-u\mu)\right]\\
&\text{for}\quad u<\frac{1}{\mu}
\end{aligned}$$

Due to the closed-form availability of the conditional Laplace transform it is again very easy to work out the first two conditional moments. Note that the ERS-ARG as well as the RS-ARG process are both characterized by an exponential affine conditional Laplace transform and hence are Car(1) processes. Darolles et al. (2006) have shown that those processes are extremely nice to construct term structure models as they provide closed-form pricing equations. In addition, when setting $\overline{\alpha}_t$ equal to 0 and only one limiting regime exists, the classical ARG process from Gourieroux and Jasiak (2006) is obtained. Also, when we set ν equal to 0 we obtain the previously introduced RS-ARG$_0$ process and by only having one limiting regime, we obtain the simple ARG$_0$ process. Thus, the new ERS-ARG process nests all those processes. Moreover, the only condition is the existence of transition probabilities. This means that the functional form of the membership function can take on many forms which allow for great flexibility when it comes to computing the transition probabilities. In addition, the transition probabilities do not necessarily have to come from a membership function, the also could be exogenously given from, let's say, a first stage Probit-model.

A particular useful property of the new process is its ability to model various non-linearities during the zero lower-bound period. For example, while the non-regime switching process always has the same lower bound for the conditional variance, the regime-switching one can match varying patterns arising due to higher market uncertainty about future rates. The following section develops the regime switching affine term structure model which is based on the multivariate version of the RS-ARG$_0$ process, the so-called Vector RS-ARG$_0$ process (RS-VARG$_0$).

3.4 The Non-Negative Affine Regime Switching Term Structure Model

To construct the regime switching term structure model the dynamics of the latent state vector under the risk-neutral $\mathbb{Q}$-distribution will be directly specified. As mentioned before, the latent state vector will be based on the multivariate adaption of the previously introduced RS-ARG$_0$ process. Formally, this means that the n latent factors follow a vector ERS-ARG process (ERS-VARG). Furthermore, the unobservable short-rate between t and $t+1$ will be denoted by r_t.

Assumption 3.1. *Let us partition* $X_t = \left(X_t^{(1)\prime}, X_t^{(2)\prime}\right)'$, *where* $dim\left(X_t^{(1)}\right) = n_1$, $dim\left(X_t^{(2)}\right) = n_2$, *and* $n = n_1 + n_2$. *The risk-neutral distribution of* X_{t+1}, *conditionally on* $\underline{X_t}$, *is given by the product of the following conditional distributions:*

$$(X_{j,t+1}|\underline{X_t}) \overset{\mathbb{Q}}{\sim} \gamma_{\nu_j}\left(\overline{\alpha}_t^{\mathbb{Q}} + \overline{\beta}_t^{\mathbb{Q}\prime} X_t, \mu_j^{\mathbb{Q}}\right) \quad , \quad j \in \{1, ...n\},$$

where $\nu_j = 0$ *for any* $j \in \{1, ...n_1\}$, *while* $\nu_j \geq 0$ *if* $j \in \{n_1+1, ...n_2\}$; $\overline{\alpha}_t^{\mathbb{Q}} \geq 0, \mu_j^{\mathbb{Q}} > 0$ *and* $\overline{\beta}_t^{\mathbb{Q}}$ *is an n-dimensional vector of positive components. In other words, conditionally on* X_t, *the* n_1 *components of* $X_{t+1}^{(1)}$ *follow independent RS Gamma-zero distributions, while the* n_2 *components of* $X_{t+1}^{(2)}$ *follow independent ERS non-central Gamma-distributions.*[15]

As the ERS-ARG process possess the same properties as the extended ARG process from

[15] This assumption corresponds to assumption 1, page 352 in Monfort et al. (2017).

Monfort et al. (2017), the conditional Laplace transform of X_{t+1} given X_t is directly obtained given its conditional independence.

Proposition 3.2. *The risk-neutral Laplace transform of X_{t+1}, conditionally on $\underline{X_t}$, is given by:*

$$\begin{aligned}\varphi_t^{\mathbb{Q}}(u) &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\sum_{j=1}^{n} u_j X_{j,t+1}\right) |\underline{X_t}\right] \\ &= \exp\left[\sum_{j=1}^{n} a_{t,j}^{\mathbb{Q}}(u_j)' X_t + b_{t,j}^{\mathbb{Q}}(u_j)\right]\end{aligned}$$

where, for any $j \in \{1, ..., n\}$, we have:

$$\begin{aligned}a_{t,j}^{\mathbb{Q}}(u_j) &= \frac{u_j \mu_j^{\mathbb{Q}}}{1 - u_j \mu_j^{\mathbb{Q}}} \overline{\beta}_{j,t}^{\mathbb{Q}} \quad \text{and} \\ b_{t,j}^{\mathbb{Q}}(u_j) &= \frac{u_j \mu_j^{\mathbb{Q}}}{1 - u_j \mu_j^{\mathbb{Q}}} \overline{\alpha}_{j,t}^{\mathbb{Q}} - \nu_j \log(1 - u_j \mu_j^{\mathbb{Q}}).\end{aligned}$$

The process (X_t) is therefore a discrete-time affine (Car(1)) process.

Proof. Multivariate extension of the results from the univariate process. □

Corollary 3.2.1. *$\rho_j < 1$ for $j = 1, ...n$ is a sufficient condition for $\mathbb{Q}$-stationary of (X_t) given a lower-triangular $\beta_j^{\mathbb{Q}}$ matrix.*

Proof. See appendix A.3. □

Assumption 3.2. *The nominal short rate process (r_t) is given by a linear combination of the first n_1 components of X_t only, that is:*

$$r_t = \sum_{j=1}^{n_1} \delta_j X_{j,t} \tag{3}$$

where $\delta = [(\delta_j)'_{j=\{1,...,n_1\}}, 0_{n_2}]'$ has the first n_1 entries strictly positive, the remaining ones being equal to zero.[16]

It is worth noting that the regime-switching property form the ERS-ARG process carries over to the short-rate specification as it is a linear combination of conditionally independent RS-ARG$_0$ processes. Furthermore, in light of the recent negative interest rates, one could allow the short rate to become negative by setting $r_{min} \neq 0$ to the right hand side of equation (3).

[16]This assumption corresponds to Assumption 2, page 353, in Monfort et al. (2017).

Regarding computational ease, the conditional Laplace transform of equation (3.2) can be cast into matrix form to arrive at the following expression:

$$\varphi_t^{\mathbb{Q}} = \exp\left[\tilde{a}_t^{\mathbb{Q}}(u)'X_t + \tilde{b}_t^{\mathbb{Q}}(u)\right],$$

where:

$$\begin{aligned}
\tilde{a}_t^{\mathbb{Q}}(u) &= \overline{\beta}_t^{\mathbb{Q}}\left(\frac{u \odot \mu^{\mathbb{Q}}}{1 - u \odot \mu^{\mathbb{Q}}}\right) \\
\tilde{b}_t^{\mathbb{Q}}(u) &= \overline{\alpha}_t^{\mathbb{Q}\prime}\left(\frac{u \odot \mu^{\mathbb{Q}}}{1 - u \odot \mu^{\mathbb{Q}}}\right) - \nu' \log(1 - u \odot \mu^{\mathbb{Q}}) \\
\mu^{\mathbb{Q}} &= (\mu_1^{\mathbb{Q}}, ..., \mu_n^{\mathbb{Q}})', \quad \overline{\beta}_t^{\mathbb{Q}} = (\overline{\beta}_{1,t}^{\mathbb{Q}}, ..., \overline{\beta}_{n,t}^{\mathbb{Q}}), \\
\overline{\alpha}_t^{\mathbb{Q}} &= (\overline{\alpha}_{1,t}^{\mathbb{Q}}, ..., \overline{\alpha}_{n,t}^{\mathbb{Q}})', \quad \nu = (0, ...0, \nu_j n_1 + 1, ..., \nu_n)',
\end{aligned}$$

and where $\odot$ denotes the element-by-element product, Hadamard product, and with abuse of notations, the division and log operators work element-by-element when applied to vectors. The following proposition states the explicit zero-coupon bond pricing formulas:

Proposition 3.3. *If the n-dimensional state-vector (X_t) has a risk-neutral dynamics defined by Eq. 3.2 and if the short-term interest rate is defined as in Proposition 3.2, then the price at date t of the zero-coupon bond with residual maturity h, denoted $P_t(h)$, is given by:*

$$P_t(h) = \exp\left(A'_{t,h}X_t + B_{t,h}\right), \tag{4}$$

where A_h and B_h satisfy the following recursive equations:

$$\begin{aligned}
A_{t,h} &= -\delta + \tilde{a}_t^{\mathbb{Q}}(A_{t,h-1}) \\
&= -\delta + \overline{\beta}_t^{\mathbb{Q}}\left(\frac{A_{t,h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{t,h-1} \odot \mu^{\mathbb{Q}}}\right)
\end{aligned} \tag{5}$$

$$\begin{aligned}
B_{t,h} &= B_{t,h-1} + \tilde{b}_t^{\mathbb{Q}}(A_{t,h-1}) \\
&= B_{t,h-1} + \overline{\alpha}_t^{\mathbb{Q}\prime}\left(\frac{A_{t,h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{t,h-1} \odot \mu^{\mathbb{Q}}}\right) - \nu' \log(1 - A_{t,h-1} \odot \mu^{\mathbb{Q}})
\end{aligned} \tag{6}$$

with starting conditions $A_{t,0} = 0$ and $B_{t,0} = 0$. The date t continuously-compounded yield associated with a zero-coupon bond maturing in h periods is therefore given by the following non-negative affine function of X_t:

$$\begin{aligned}
R_t(h) &= \bar{A}'_{t,h}X_t + \bar{B}_{t,h}, \\
\bar{A}_{t,h} &= -\frac{1}{h}A_{t,h}, \quad \textit{and} \quad \bar{B}_{t,h} = -\frac{1}{h}B_{t,h}, \quad h \geq 1
\end{aligned} \tag{7}$$

Proof. See A.1 □

The no-arbitrage formula $R_t(h) = -\frac{1}{h}\log \mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t - ... - r_{t+h-1})]$ directly establishes the non-negativeness of the yield formulas. This as the short rate is specified as a linear combination of positive components and the yield is a positive combination of positive components and hence positive too.

3.4.1 The Historical $\mathbb{P}$-Dynamics

After having defined the risk-neutral $\mathbb{Q}$-dynamics, the dynamics of the state-vector under the real-world probability measure $\mathbb{P}$ have to be specified next. To do so, the one-period stochastic discount factor is assumed to be based on an exponentially affine change of measure which is common in the asset pricing literature. Formally, the change of measure is achieved by the following Radon-Nikodym derivative:

$$\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp[\theta' X_{t+1} - \psi_t^{\mathbb{Q}}(\theta)], \tag{8}$$

where $\psi_t^{\mathbb{Q}}(u) = \log \varphi_t^{\mathbb{Q}}(u)$ denotes the risk-neutral conditional Laplace transform of (X_t), and $\theta = (\theta_1, ..., \theta_n)'$ denotes the n-dimensional vector of market prices of risk factors. This change of measure directly leads to the following proposition about the historical distribution:

Proposition 3.4. *The historical distribution of (X_{t+1}), conditionally on $\underline{X_t}$, is given by the product of the conditional distributions:*

$$(X_{j,t+1}|\underline{X_t}) \overset{\mathbb{P}}{\sim} \gamma_{\nu_j}\left(\overline{\alpha}_t^{\mathbb{P}} + \overline{\beta}_t^{\mathbb{P}\prime} X_t, \mu_j^{\mathbb{P}}\right) \quad , \quad j \in \{1, ...n\},$$

where $\overline{\alpha}_{t,j}^{\mathbb{P}} \geq 0, \mu_j^{\mathbb{P}} > 0$, and $\overline{\beta}_{j,t}^{\mathbb{P}}$ is an n-dimensional vector of strictly positive components and the historical Laplace transform of X_{t+1}, given $\underline{X_t}$, is given by:

$$\varphi_t^{\mathbb{P}}(u) = \exp\left[\sum_{j=1}^{n} a_{t,j}^{\mathbb{P}}(u_j)' X_t + b_{t,j}^{\mathbb{P}}(u_j)\right],$$

where, for any $j \in \{1, ..., n\}$, we have:

$$\begin{aligned}
a_{t,j}^{\mathbb{P}}(u) &= \frac{u_j \mu_j^{\mathbb{P}}}{1 - u_j \mu_j^{\mathbb{P}}} \overline{\beta}_{j,t}^{\mathbb{P}} \\
b_{t,j}^{\mathbb{P}}(u_j) &= \frac{u_j \mu_j^{\mathbb{P}}}{1 - u_j \mu_j^{\mathbb{P}}} \overline{\alpha}_{j,t}^{\mathbb{P}} - \nu_j log(1 - u_j \mu_j^{\mathbb{P}}),
\end{aligned}$$

with

$$\overline{\alpha}^{\mathbb{P}}_{t,j} = \frac{\overline{\alpha}^{\mathbb{Q}}_{t,j}}{1-\theta_j \mu_j^{\mathbb{Q}}}, \quad \overline{\beta}^{\mathbb{P}}_{t,j} = \frac{1}{1-\theta_j \mu_j^{\mathbb{Q}}} \overline{\beta}^{\mathbb{Q}}_{j,t} \quad \textit{and}$$
$$\mu_j^{\mathbb{P}} = \frac{\mu_j^{\mathbb{Q}}}{1-\theta_j \mu_j^{\mathbb{Q}}}.$$

Proof. Straightforward adaption of the results from Monfort et al. (2017). □

Thus, compared to the linear counterpart, all coefficients are now time-varying convex combinations of the regime limiting values. The weights of the convex combination are the probabilities from the membership function. The distribution of the process at each point in time is therefore given by the most likely description of the extreme states of the economy or later on extreme states of monetary policy.

4 Data

To analyze the Regime-Switching Term Structure Model (RS-TSM), data for the predictor space as well as zero-coupon yields have been collected from various sources. For the predictor space, data on interest rate futures options for the Euro area and the US have been collected from Thomson Reuters Datastream. The Euro area options are commonly known as Euribor options and are traded on the LIFFE Intercontinental Exchange (ICE) in London, while one refers to Eurodollar options in the case of the US. First, a detailed overview of the implied density construction and some stylized facts about them will be presented. Following this, an overview about the respective yield curve data will be given.

4.1 Option Data

The options data set consists of daily data on Euribor options (EU area) and Eurodollar options (US) for the March quarterly cycle[17] and their respective underlying from 01.01.2007 to 31.12.2016. The underlying for the Euribor and Eurodollar are 3-month LIBOR rate denominated futures in the respective currency. The underlying rate can generally be seen as the closest market proxy to the short term rate of the respective monetary authority and the price of the futures is computed as 100 minus the respective LIBOR rate. On average there are 500 contracts available per year for each country. Both, Euribor and Eurodollar options, are American style options. The ICE in London uses a futures-style margining system, which means that they do not involve discounting, see R.-R. Chen and Scott (1993). Furthermore, it also means that the holder of the option has no opportunity costs of holding the option and hence one can neglect the early-exercise premium associated with American style options. The Eurodollar options, which are traded at the Chicago

[17]The March quarterly cycle includes the expiration month March, June, September and December.

Mercantile Exchange (CME) and also represent American style options are among the most heavily traded financial derivatives, C. Y.-H. Chen and Kuo (2014). Following Gebbia (2016), Eurodollar options are treated as European style options as the early-exercise premium does not influence the results in a substantial way.
Before deriving the risk-neutral densities, the raw option data has to be cleaned to match certain theoretical criteria such as no-arbitrage. Several levels of data filters are therefore applied which are all quite common in the literature.
First, all observations that are below the minimum tick-size of the exchange[18] are removed. Second, observations with seven days or less to maturity are removed from the sample due to expiration effects (see Lien and Li (2003) for expiration effects of options). Last, all observations that violate the arbitrage restrictions as in Buraschi, Trojani, and Vedolin (2014) are removed as well. These filters nearly capture all outliers such that the procedure runs fairly smooth in the empirical application.

4.1.1 Methodology Implied PDFs

As it is normal in asset pricing theory, contingent claims, such as those arising from options, are priced in a risk neutral setting. The price of a European option is equal to the discounted value of its expected payoff at expiration. For a European call option, this is formally:

$$\begin{aligned} C(K,t,T) &= e^{-rh} E_t^{\mathbb{Q}}[\max(F_T - K, 0)] \\ &= e^{-rh} \int_K^{\infty} (f-K) \pi_{F_T,t}^{\mathbb{Q}}(f) df, \end{aligned} \tag{9}$$

with F_T being the price of the underlying futures contract at expiration, $h = T - t$ the time to maturity, r the risk-free rate and $\pi_{F_T,t}(f)$ is the pdf of F_T at time t evaluated at the price f. There are generally two possible ways to construct the option implied pdfs, parametric techniques, like fitting a mixture of lognormal distributions as in Mirkov, Pozdeev, and Söderlind (2016), or non-parametric approaches like using polynomial fitting techniques. In related literature, where also implied pdfs from options on interest rate futures are constructed, a modified version of the non-parametric approach seems to be most widely used, which is why I adopt this approach in this thesis. For the non-parametric approach, one of the most important results is due to Breeden and Litzenberger (1978). They show that the market implied risk-neutral density is equal to the second order strike price derivative, scaled by some constant. Differentiating formula

[18]Minimum tick size Euribor: 0.005; Minimum Tick Size Eurodollar: 0.005.

(9) twice with respect to K yields:

$$\begin{aligned}
\frac{dC(K,t,T)}{dK} &= -e^{-r\tau}\int_K^\infty \pi_{F_T,t}(f)df \\
\frac{d^2C(K,t,T)}{dK^2} &= e^{-r\tau}\pi_{F_T,t}(K) \\
\pi_{F_T,t}(K) &= e^{-r\tau}\frac{d^2C(K,t,T)}{dK^2}
\end{aligned}$$

Given this insight one can construct the implied pdf by first constructing the call price function and then numerically differentiate twice. Nevertheless, this approach creates several obstacles as prices are only observed at discrete-time intervals and the pdf is a smooth function. Naturally, the smoothness is achieved by interpolating between data points. As underlying futures directly represent the short rate,[19] the strike prices of the underlying are subject to huge variation, given the decline to zero for the short term interest rates. Due to this reason, it is more convenient to do the interpolation in the implied volatility/delta surface as the deltas are bounded between zero and e^{-rh}. To do so, the procedure first introduced by Shimko (1993) is used where prices are transformed into implied volatilities and the strikes to deltas via the Black (1976) model for options on futures. To achieve the desired smoothness and curvature of the implied pdf, a cubic smoothing spline is used to interpolate between the data points at a high granularity (15'000 data points). It is a piecewise cubic polynomial and has many desirable properties when recovering the implied volatility surface, see for example Fengler (2009). The cubic smoothing spline offers a great deal of flexibility as it explicitly takes into account the balancing between the smoothness of the function and the errors of the fit that arise due to the fact that one wants to obtain a smooth function. The cubic smoothing spline $g(\cdot)$ is the result from the minimization of the following objective function:[20]

$$(1-\lambda)\sum_{i=1}^{n} w(i)[y(i)-g(x(i))]^2 + \lambda\int_x [g''(t)]^2 dt, \tag{10}$$

where x represents the deltas and y corresponds to the implied volatilities. Note that the cubic smoothing spline opitmization also takes into account specific weights, which corresponds to the Vega of the option. Bliss and Panigirtzoglou (2002) give a detailed about why using Vega as the weight in the optimization process. Last, the smoothing parameter λ determines the trade off between smoothness and fit. Specifically, in the application the parameter is set to 0.001 for most days. Due to the adoption of a zero interest rate policy and negative interest rates in the EU, the data starts to behave slightly differently around some days. In order to guarantee the

[19] The underlying futures are quoted as 100 minus prevailing rate.

[20] In the application, the MATLAB built in function csaps is used. The resulting pdfs show that the routine works as desired and produces reasonable estimates of the function.

positivity of fitted volatilities, the smoothing parameter is set to 0.0001 for certain days. Finally, after obtaining the smoothed values they are transformed back into strikes and prices and by numerical differentiation the implied pdfs are obtained.
As described in Section 2.5, the main goal of the implied pdfs is to construct proxies for the monetary policy stance of the market. It is generally assumed that the central bank controls the short-term rate and typically, one uses the 3-month rates as such a proxy. As the time to maturity of an option changes over time, they do not necessarily correspond to the expected short rate. So-called 90-day fixed horizon estimates of the implied pdf are therefore constructed by linearly interpolating the IV-delta function over time between the near and far-future contract, thus obtaining estimates of the 3-month risk-neutral implied pdf. As outlined in the beginning, one particular reason for using these implied pdfs is that the provide a much more comprehensive view of market expectations when rates are close to zero.

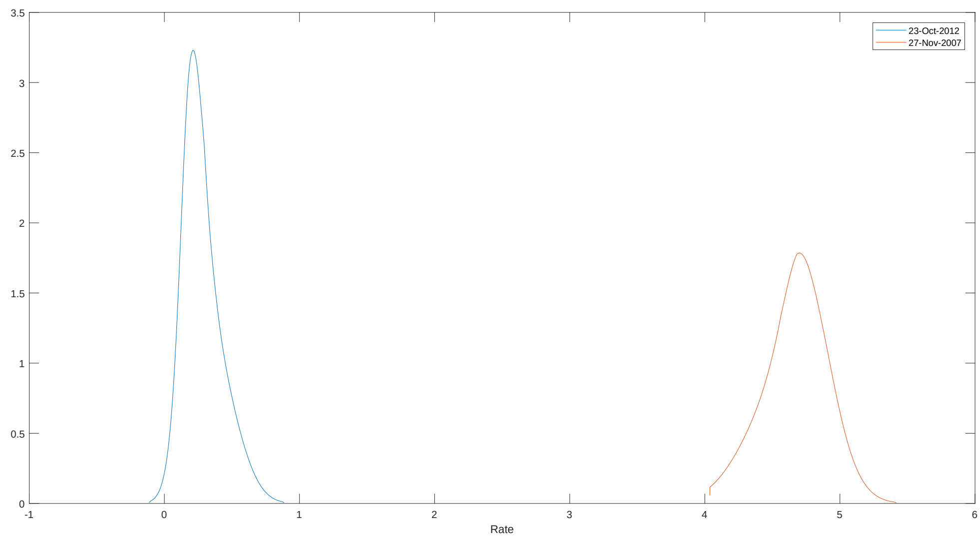


Figure 2: Option Implied PDFs

Figure 2 shows option implied densities derived from Euribor options for two different dates, October 2012 and November 2007. Clearly, the shape of the distribution differs substantially between these two days as the prevailing monetary policy stance differs as well. The shape of the distribution can be directly linked to current economic conditions. Vähämaa (2005) examine the implications the shape of the distribution in more detail and link it to market expectations. The November 2007 density corresponds to the beginning of the subprime crisis. As can be seen, the density is left-skewed as people expect the central bank to lower rates in order to prevent a recession. More probability mass is therefore placed on lower rates. The October 2012 density shows a right-skewed pattern and corresponds to the zero rate period. Again, it is evidently clear that the density reflects the current monetary policy. As rates were close to zero, the mode is also close to zero, but there is essentially no mass for future rates below zero, indisputably reflecting the zero lower bound. It can therefore be expected that using implied pdfs provide a much more accurate description of market expectations about monetary policy.

4.2 Interest Rate Data

Data for the whole term structure has been collected for the United States and the European Union. For the United States the maturities included range from 1 to 10 years. The data has been obtained from the Gürkaynak, Sack, and Wright (2007) database. This database has been widely used in term structure research about the U.S. such as in J. H. E. Christensen and Rudebusch (2016).

Figure 3 shows the zero coupon yields for the US for the sample period covering the time from 19. December 2006 until 12. September 2016. Thus, all major events which influenced markets around the world are included. In particular, the sample period covers a high-yield environment at the beginning and before the financial crisis, a transition phase as well as the whole zero lower bound period. Furthermore, at the end of the sample the US started to raise rates again. Therefore, the sample selected is optimal for testing a model that should be able to capture the various different monetary policy regimes as well as their transition.

The US data also reveals the presence of some possible trends. For example, while the shorter yields, such as the one and two year yields, show a rather sharp drop following the unfolding of the financial crisis in the level, the long term yields depict a rather smooth downward trend towards a much lower level. This evident trend makes the application of a linear model such as the non-regime switching model unreasonable. The reason for this is that in order to capture the trend, the procedure will most likely inflate the autoregressive parameters to one such that the process well get into the non-stationary region. Further evidence for this is the autocorrelation behavior of the interest rate data as can be seen in Table 2. The autocorrelation exceeds 0.997 for all maturities. This extreme persistent behavior which could lead to difficulties in the estimation stage. This would clearly spur the results for the linear model but also could cause potential problems for the regime-switching model.

Table 2: Summary Statistics ZCB Rates

			United States			
Maturity	1Y	2Y	4Y	5Y	7Y	10Y
Mean	0.9108	1.1133	1.6912	1.979	2.484	3.0405
Std.	1.3534	1.2456	1.0948	1.0525	1.0114	0.9968
Autocorr.	0.9978	0.9975	0.9971	0.997	0.997	0.9971
			European Union			
Maturity	3M	1Y	2Y	5Y	7Y	10Y
Mean	0.9924	1.0926	1.2466	1.7918	2.1352	2.5334
Std.	1.5211	1.5909	1.5953	1.5406	1.4933	1.4345
Autocorr.	0.9928	0.9921	0.9906	0.9887	0.9879	0.9867

Notes: Mean and standard deviation hold for yields expressed in annualized percentage points.

In addition, the period covered includes different shapes of the yield curve. At the beginning of the sample, the yield curve is u-shaped while it gets increasing after the sharp decline. The mean and standard deviation of the yields is also typical for interest data as has been mentioned at the beginning. Last, the presence of the rise at the end of the period makes the data appropriate for a forecasting application in order to test whether the model is able to capture the transition out of the low-yield environment.

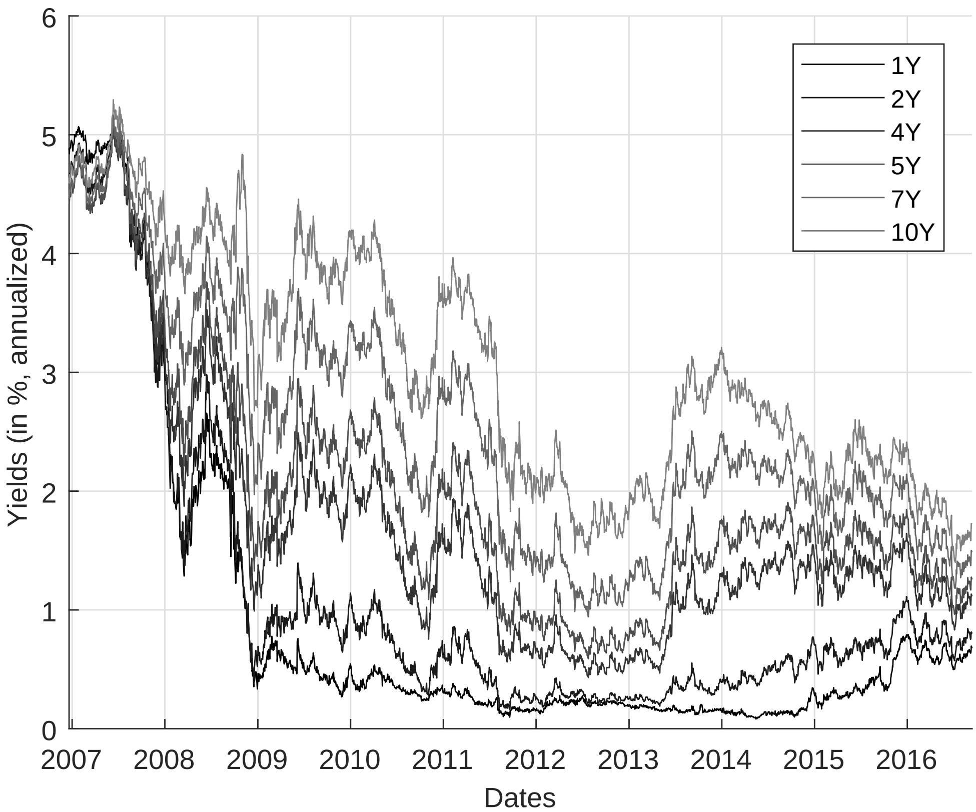


Figure 3: US Yields

Source: Author's rendering of Fed data (2017).

The interest rate data for the EU consists of daily yield observations for maturities consisting of 3 month, 1 year, 2 years, 5 years, 7 years and 10 years. These zero coupon data has been taken from ECB's statistical warehouse.[21] The selected sample spans from 19. December 2006 to 10. June 2014. The period ends before the introduction of negative interest rates by the ECB. Figure 4 presents the yields for the mentioned maturities and they exhibit a fairly similar behavior as their US counterpart. Contrary to the Fed's behavior, the ECB adopted a more monotonic decreasing such that the short rate nearly instantly dropped to 1% when they started to decrease rates in October 2008. The term structure is generally increasing over the whole sample period but different slopes can be observed. Before the financial crisis, the term structure was fairly flat with a steep increase at the very short end. This clearly changed over the sample period with the most monotonic slope being present after the rate drop. The slope seems to narrow down towards the end of the sample such that the term structure is now fairly flat at the beginning and starting to steepen towards the five and seven year yields. The reason for this seems to be that long-term yields did not experienced a large rate drop but rather show a downward trend. Thus, similar to the US, fitting a linear model to the sample period might not be appropriate. Also, comparing the standard deviation and autocorrelation for the sample clearly reveals the difference. While European rates are still fairly persistent over the sample, especially long-term rates showed much more variation, suggesting less issues in the estimation. In order for a term structure model to be an accurate description of the dynamics observed, it has to be able to match various kinds of yield curve shapes as well as flexible enough to match non-linearities observed over time. Summary statistics for the sample period are presented in Table 2.

[21] For detailed information about the construction of the zero coupon yields, see http://sdw.ecb.europa.eu/browseTable.do?node=9689726.

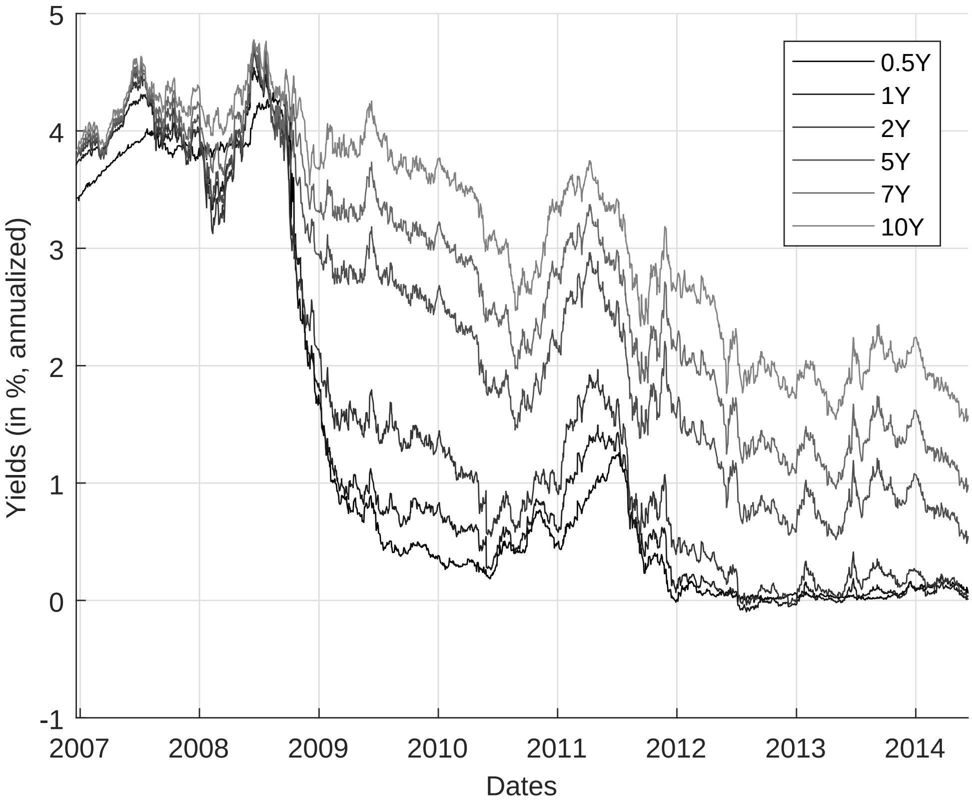


Figure 4: EU Yields

Source: Author's rendering of ECB data (2017).

Altogether, the sample period selected possesses different shapes of the yield curve while showing non-linearities which linear models have difficulty with to capture. This makes the periods suitable to show the performance of the non-linear regime-switching model. Moreover, the two areas had different kinds of monetary policy even though a similar behavior of the short rate is present. If the model presented in Section 3.2 is appropriate, the estimation procedure outlined in the next section should be able to consistently capture those differences and the overall term structure dynamics.

5 Estimation Strategy and Empirical Results from the RS-TSM

5.1 Estimation Strategy

As the new RS-ARG$_0$ nests the ARG$_0$ process of Monfort et al. (2017) and also possesses a weak AR(1) representation as presented in Subsection 3.2.2, it seems natural to use the linear Kalman filter (as in Monfort et al. (2017), de Jong (2000)) to derive an approximation of the likelihood function. As the basic properties of the process are similar to the nested one, the modification of the Kalman filter as introduced in Monfort et al. (2017) will be used here. The modified version of the Kalman filter, see A.2, allows for a deterministic time-variation in the yield prediction equation. To have a deterministic time-variation, the transition probabilities have to be known in advance. This means that the regime structure has be estimated in a first stage such that the term structure model can be estimated subsequently.

The RS-TSM can easily be casted into a linear state-space representation where we use the multivariate version of equation 2 as the transition equation and equation 7 as the measurement equation.

Formally, this leads to the following equations:

$$
\begin{aligned}
X_{t+1} &= \underbrace{\mu^{\mathbb{P}} \odot \overline{\alpha}_t^{\mathbb{P}}}_{m_t} + \underbrace{\mu^{\mathbb{P}} \odot \overline{\beta}_t^{\mathbb{P}\prime}}_{M_t} X_t \\
&+ \quad \underbrace{\{diag\left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(2\overline{\alpha}_t^{\mathbb{P}} + 2\overline{\beta}_t^{\mathbb{P}\prime} X_t\right)\right]\}^{1/2}}_{\Sigma_t^{1/2}} \epsilon_{t+1} \qquad (11) \\
&= m_t + M_t X_t + \Sigma_t^{1/2} \epsilon_{t+1}
\end{aligned}
$$

As shown in Section 4.2, the dataset consists of yield data for various maturities. Throughout the rest of this section, the vector of available maturities is denoted by $H_{US} = \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$ and $H_{EU} = \{0.4, 1, 2, 5, 7, 10\}$.[22] The day-count convention is 360-days per year. The measurement equation therefore reads:

$$
R_t(h) = \bar{B}_{t,h} + \bar{A}_{t,h} X_t + \sigma_R \eta_{R,h,t}, \quad h \in H_{EU/US}, \qquad (12)
$$

where σ_R denotes the standard deviation of the measurement noise. The measurement noise is added in order to break the stochastic singularity which arises as yields are a linear combination of a small number of factors. As in Monfort et al. (2017), it is assumed to be the same for all maturities. In addition, $\eta_{R,h,t}$ is assumed to be i.i.d. Gaussian white noise.

Putting this together, the following state-space representation of the system under consideration is obtained:

$$
\begin{cases}
X_{t+1} & = m_t + M_t X_t + \Sigma^{1/2} \epsilon_{t+1} \\
R_t & = \Gamma_{t,0} + \Gamma_{t,1} X_t + \Omega \eta_t,
\end{cases}
$$

where $\eta_t = (\eta'_{R,t})' \sim IIN(0, I)$. As usual in state-space models, the estimation is done via maximum likelihood estimation where the Kalman filter is used to derive an approximation of the likelihood function. Due to specific properties of the process, a modified version of the Kalman filter which is presented in Monfort et al. (2017) will be used as it has shown good properties to derive parameter estimates.

In order to obtain reasonable parameter estimates, the estimation procedure has to be quite

[22]The maturities in days are respectively $H_{US} = \{360, 720, 1080, 1440, 1800, 2160, 2520, 2880, 3600\}$ and $H_{EU} = \{90, 360, 720, 1800, 2520, 3600\}$.

robust as model involves several parameters. For example, a 4-factor specification with two limiting regimes requires the estimation of 33 parameters and the best performance is achieved with a BFGS algorithm. This means that the parameters will not be estimated directly, but rather a transformation to ensure positivity of the gamma-zero parameters and the delta values. More precise, $\theta = (\alpha^{\mathbb{P}}_{r,j}, \rho^{\mathbb{P}}_{r,j}, \rho^{\mathbb{Q}}_{2,j}, \beta^{\mathbb{P}}_{r,j\neq i}, \sigma_R)$, $r = 1, 2$. The rest of the parameters follow by identification. Special attention has to be devoted to the short rate. Particularly, to obtain a reasonable value for the short rate for the US, the delta is calibrated to match an initial short rate of 5.14%. This value has been obtained by using the slope between the one and two year yield and multiplying it by 1.1. The same procedure is adopted for the EU which is characterized by an initially upward sloping yield curve. The initial calibrated short rate is 3.321%. This specification does not correspond to the approach taken in Monfort et al. (2017) who impose an unconditional short rate. This is not possible here as the time-varying parameters do not allow for an unconditional short rate to hold for a whole sample period.

5.2 Regime Structure

The regime structure should resemble different monetary policy regimes. As the short rate is generally assumed to be controlled by the monetary authority it is naturally to assume structural breaks at the short rate. Therefore, the three month Euribor and T-Bill rate are assumed to follow a Regime-Switching ARG_0 process. To identify the regime limiting structure, an iterative algorithm which is presented in Audrino and Buehlmann (2001) is employed. The algorithm consists of two steps. First, a large tree is grown based on some criteria, here the negative log-likelihood. In a second step, the tree is pruned such that the optimal sub-tree is identified via an information criterion, for example with the Bayesian Information Criteria (BIC). The predictor space for the tree structure procedure consists of various measures obtained from option implied pdfs and thereby act as a market-based proxy for monetary policy. Using the notation from Section 3.1.1, the n-dimensional vector consisting of the predictor variables is $\boldsymbol{l}_t = (\mu_t, \sigma_t^2, \frac{\mu_{3,t}}{\sigma_t^3}, \frac{\mu_{4,t}}{\sigma_t^3}, Mode_t, \mu_t - Mode_t) \in \mathcal{G} \subseteq \mathbb{R}^4 \times \mathbb{R}^2_+$.

The algorithm used in the growing procedure generally works as follows: First, the non-regime process is fitted to the data to obtain an initial value for the negative log-likelihood. Subsequently, one partition cell $\mathcal{R}_{j*}$ in the existing partition $\mathcal{P}^{(0)}$ is split into $\mathcal{R}_{j*} = \mathcal{R}_{j*,left} \cup \mathcal{R}_{j*,right}$ at a specific threshold value. To search over all possible values from the predictor space would be computationally not feasible. Therefore, a grid search is performed where the grid points correspond to empirical α-quantiles of each predictor variable, where $\alpha = i/\text{mesh}$. In the empirical application, I choose mesh $= 6$ for the EU and mesh $= 8$ for the US. This leads to a total amount of 36, respectively 48 possible splitting values.

For each splitting threshold, the model is fitted by minimizing the negative log-likelihood while holding the parameters of the non-refined partition fixed and using the parameters from the previous iteration as starting values. This ensures that at the beginning of the optimization the log-likelihood value is the same as without the additional regimes such that the new specification fits at least as well as the new one. It should be noted that this is not a necessary procedure and the starting values for the linear model as starting values for the regime-structure. At each level of the tree, the specification which provides the largest decrease in the log-likelihood is used as the new optimal partition of the predictor space. This procedure is iterated up to a maximum tree-level. Due to the computational burden that arises as part of the time-varying parameters, the procedure is iterated up to a level of three. The reason is that for each additional regime the number of parameters that have to be estimated in the subsequent term structure model grows by $2n + n(n-1)/2$, where n represents the number of latent factors. With four factors this means 14 additional parameters for an additional regime. In addition, the estimation reveals that two regimes already provide an accurate description of the data at hand. As the grown tree clearly overfits the data, one usually searches for the optimal subtree by pruning such that the final specification provides a balance between fit and the number of regimes. Audrino and Buehlmann (2001) suggest to use a Quasi-Newton method in the estimation procedure. Several test have shown that the most stable results are obtained when using a BFGS algorithm in case of the US and a trust-region method for the EU.

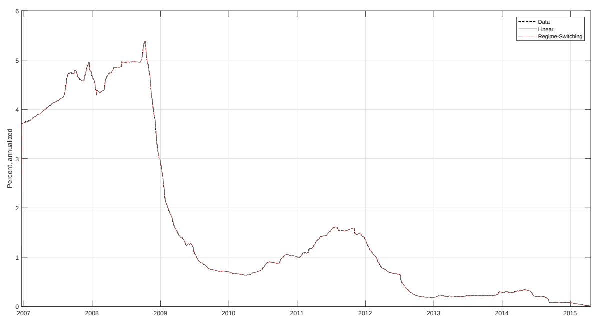
5.2.1 Regression Tree

The estimation procedure points towards the existence of two limiting regimes for the European and US monetary policy proxy. Figure 5 plots those proxies as well as their model implied counterparts. In both cases, the linear model already provides a very good fit for the data but the autoregressive coefficient for the linear model is extremely close to one in both cases such that it seems difficult to rule out a possible unit root. This result is not unexpected. As reasoned previously and given the autocorrelation pattern of the data, the autoregressive parameters must be close to one in order to provide a reasonable explanation of the data. However, the regime-switching calibration results is less persistent autoregressive parameters and thereby helps to rule out a possible unit rule while still being strongly persistent.

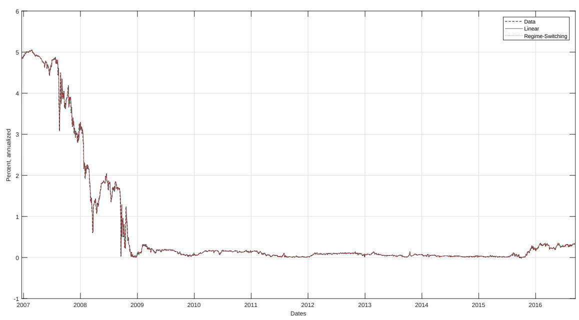
European Regime Structure The tree structure for the EU identifies one high and one low yield regime. The threshold variable which shows the largest improvement in the negative log-likelihood is the mean of the implied density with a threshold value of 0.8738 percent. This shows the impact of the financial crisis on monetary policy in Europe as rates fell under this threshold mid 2009, the aftermath of the crisis. The parameters identify the first regime as a low yield environment while the second one would correspond to normal times. Looking and the negative log-likelihood and BIC, the improvement is striking. The linear model clearly does not provide an accurate description of the data in this case.

The economic interpretation of the parameter estimates reported in Table 3 further support such a hypothesis, particularly the difference in the α parameter. The extremely low alpha parameter indicates a strong persistence of the zero lower bound if the economy enters the respective state. Furthermore, the sensitivity parameter τ for the EU is very high. This points towards a hard split type of regime structure which could also be generated by just using and indicator function instead of a smooth transition type of approach. Clearly, the associated transition probabilities presented in Figure B.1 show this evident behavior of the regime-switching ARG_0 process. Moreover, the reason for the hard-split can also be seen in the behavior of the three month Euribor rate. The ECB quickly and monotonically decreased the rates below the selected threshold, pointing towards a sudden change in monetary policy.

Moreover, it shows the ability of the approach and the model to identify nested structures in the sense that a hard split tree is always nested in a smooth one. The identified unconditional moments of the linear and non-linear model evidently show that the regime-switching model is the appropriate one. While one could argue that the unconditional mean of the linear model is an adequate description of the data, the variance is clearly far off. The regime-switching model however identifies much more plausible values for the unconditional behavior of interest rates.



(a) EU



(b) US

Figure 5: Regime-Switching Short Rate

US Regime Structure As expected, there exists one low and one high yield regime. The results in Table 3 clearly show the superior performance of the regime switching model. While the autoregressive parameters are fairly close to each other (R1: 0.94 vs. R2: 0.9), the α-values differ again substantially. The estimates point towards two important facts. First, the second regime is clearly a high yield regime with a ZLB persistence of approximately 30%. This is further supported by the unconditional mean of about 4.7% percent which is a reasonable value for the beginning of the 21st century. Second, the parameter from the first regime is virtually zero. This means a ZLB persistence of nearly 100% which is extraordinary. The most reasonable explanation for this is that the Fed left rates at zero for about 7 years (December 2008 - December 2015) while using forward guidance to conduct policy. They thus convinced the market that rates will stay at zero as long as macroeconomic indicators do not justify a rate hike. This also means that as soon as the economic outlook brightens and as the market expects the Fed to lift-off, the alpha value will increase again such that the LB is not binding anymore as it was during an extended period of time. Thus, the parameter values and the regime construction inevitably show the ability of the new process to match economic conditions by relying on rational market expectations.

Furthermore, Filipovic et al. (2017) point out the level-dependent conditional volatility characteristic of interest rates. This feature is easily matched by the RS-ARG_0 process on a conditional as well as unconditional level. The unconditional variance of the linear model is relatively high, four times higher than the mean. While this might make sense if a much longer time series of daily rates is looked at, it is definitely not in line with empirical evidence from past years. J. H. E. Christensen and Rudebusch (2017) for example speak in favor of a new "normal" low level environment which would also mean that a lower unconditional volatility has to be obtained. The regime-switching process on the other hand would definitely be able to match this property. As a smooth regime-structure is obtained during the estimation, the unconditional variance will never directly be 0, but much lower than for the linear model.

The threshold value identified corresponds to the mode of the distribution which, as reasoned previously, better reflects the market-based expectation of monetary policy during ZLB periods. The result is therefore in line with economic intuition due to the extensive ZLB period in the sample. A comparison to the EU results further justify the approach and validate the results as for the EU the mean seems to exhibit a stronger predictive power while the sample shows a much smaller ZLB period.

Last, the autoregressive parameter from the linear process is with 0.9989 extremely close to 1 which can be expected due to the non-linear behavior of the data. The regime-dependent parameters identify a more persistent pattern during the ZLB period but not as much as for the linear model. This clearly helps to rule out a unit-root linear process in favor of a regime-switching one.

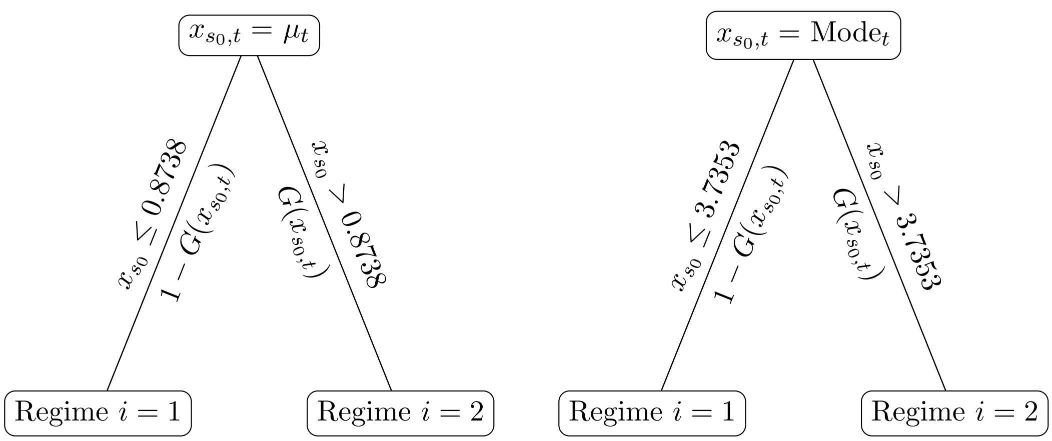
Summarizing, the results from the tree structure estimation clearly speak in favor of a regime-switching model to capture the dynamics of the short rates in the US and the EU. In addition, different moments from the option implied distribution get selected which the reason for can be seen in the unequal persistence of the ZLB period in the respective sample. Moreover, the

Table 3: 2-Regime Specification

	EU		US	
	Linear	2-Regime	Linear	2-Regime
α_1	0.2807	3.1e-05	1.244	$0^{\dagger}$
	(0.3118)	(0.312)	(0.0843)	(0.00)
α_2	-	91.738	-	193.08
		(3.08)		(9.884)
β_1	2.76e+03	1.959e+04	334.27	402.29
	(42.01)	(163.03)	(0.5)	(12.15)
β_2	-	1.965e+04	-	387.13
		(171.38)		(9.67)
μ	3.623e-04	5.0776e-05	0.003	0.0023
	(5.5e-06)	(4.41e-07)	(5.2e-06)	(5.63e-05)
τ	-	8.484e+03	-	1.36
		(0)		(0.026)
$\mathbb{E}_1$	1.22	3e-07	1.984	0
$\mathbb{E}_2$	-	2.26	-	4.69
$\mathbb{V}_1$	5.33	3e-09	3.17	0
$\mathbb{V}_2$	-	0.06	-	0.1195
log-lik.	-5960.6	**-7261.1**	-5361.2	**-5710.4**
BIC	11'898	**-14'386**	-10'699	**-11'374**

† The exact value is equal to 2.7e-22, which virtually is zero. Standard errors are computed from an outer product approximation of the Hessian and reported in parentheses.

different resulting transition probabilities show the approaches ability to differentiate between various types of monetary policy regimes as well as how they were enforced. It shows the extreme flexibility of the model and its reasonable empirical performance.



(a) EU Tree Structure

(b) US Tree Structure

5.3 Regime-Switching Interest Rate Dynamics

The empirical performance of the regime-switching term structure model on hand shows an extreme flexibility and superior performance and on the other hand helps to shed light on certain interest rate dynamics due to changes in monetary policy. The results correspond to a four factor specification with $n_1 = 1$ and $n_2 = 3$. The next section will look in more detail at the factor specification.

5.3.1 EU Term Structure

The estimation results for a fully parametrized regime switching model for the European Union pointed towards the non-significance of several parameters. These parameters have therefore been constrained to zero. Interestingly, the parameters that are found insignificant are the same under both regimes and furthermore the same as Monfort et al. (2017) found to be insignificant for Japanese Government Yields. Therefore, the interpretation of the coefficient matrices facilitates the same conclusions in the sense that $X_{j,t}$ Granger-causes $X_{i,t}$ if $j > i$ and zero still is not an absorbing state. Furthermore, the autoregressive matrices of the VAR(p) representation are quite persistent under both measures. This behavior has already been identified in the regime-structure and is therefore not surprising. The resulting estimates for the α parameter are clearly not as expected as the difference is marginal between the two regimes. On the other hand, as the first three factors generally have a 100% persistence, due to $\alpha = 0$, a high alpha parameter for the fourth factor is needed in order to ensure a possible lift-off.[23]

Further support for this conclusion can be seen in the filtered factor trajectories presented in Figure B.6. The first factor clearly resembles the path of the short rates and enters its ZLB period early 2012. The timing matches pretty well the date when the ECB first introduced zero interest rates (July 2012). The second and third factor also experience prolonged periods of zero during the sample. While the second factor shows a rather sudden drop like the first one, the third factor reveals a more gradual decline towards its zero state. The yield trajectories from Figure 4 clearly justify this movement. While the first and second factor are important for the dynamics of the short end, the third and fourth factor are needed to match the dynamics at the long end of the term structure. This facilitates an interpretation of the third factor as kind of a "level" factor for the long end. In contrary to the behavior of the first three factors, the fourth never enters a ZLB state but is rather fluctuating around some unconditional mean. Kim and Singleton (2012) noted in the case of Japan that long term rates show substantial conditional volatility while short-term rates are in a zero period. A similar pattern seems to hold for the EU and the slow moving downward-trend in the fourth factor would confirm the observation

[23] The lift-off refers to the event when the short term rate exits its respective lower bound.

made in Filipovic et al. (2017) that has been pointed out earlier. The necessity of slowly moving downward trending behavior becomes even more evident when looking at the fitted yields in Figure B.4.

A closer look at the exact factor loadings further helps to recover interest rate dynamics and also helps to shed light on the specific parts of the yield curve that have been impacted by the change in monetary policy. Figure 7 especially reveals the importance of the second and third factor to match the dynamics at different points in time and supports the hypothesis of an existing regime-change. The loadings of the first factor increase on a level basis but the shape does not change. This is expected as the values correspond to the levels of the estimated factors and hence a sharp drop in the level of the factor must be compensated for by an increase in the loading in order to generate some variation from the factor. The loadings on the fourth factor on the other hand do not change at all, supporting the reasoning that they are needed to generate the conditional volatility of long-term yields. In addition, the α_4 parameter reported in Table 4 is nearly the same under both regimes. This points towards a non regime-switching fourth factor which would be consistent with the factor loading and the α parameter values. Regarding the second and third factor, an enormous increase in the magnitude of the loading from the short to the long end seems to be evident. Given the factor trajectories mentioned earlier this makes absolutely sense as a much higher loading is needed in order to generate the short lived, relatively high, spikes in the rates that are observable in the data during the binding zero lower bound.

The estimated parameter values for the market price of risk factors show that a positive risk premium is associated with the first, second and fourth factor. If the interpretation that the third factor represents the slow moving level of the yields is accepted, then a negative risk premium can be justified as investors expect relatively stable long-term yields as they are confident in the central banks' ability to bring for example inflation back up. Nevertheless, the changing behavior of yields and different factor loadings as well as economic reasoning would call for more flexible market prices of risk. This would be an especially interesting direction for future research as it would provide more flexibility to the model and would further help to deepen our understanding of how investors perceive and price risk given certain economic conditions.

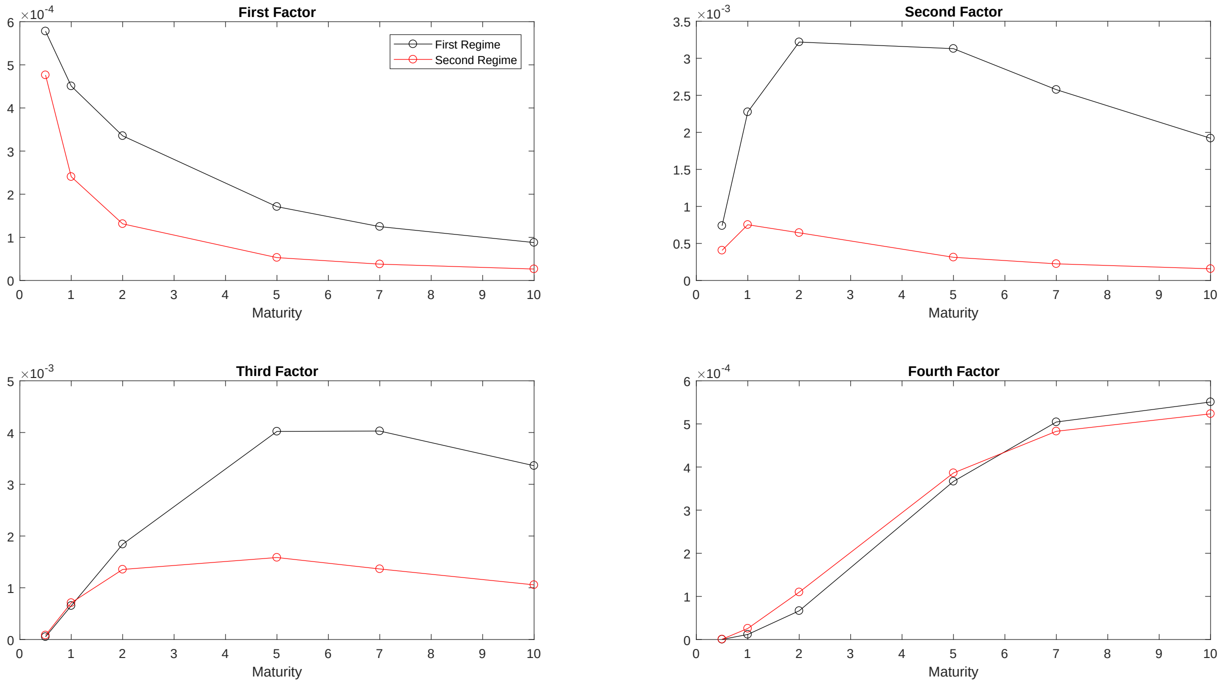


Figure 7: Factor Loadings EU Yields

Notes: The plotted factor loadings correspond to yields expressed in annualized percentage points.

Table 4: Parameter Estimates: EU

Regime 1						Regime 2					
	Value	*Std.*		Value	*Std.*		Value	*Std.*		Value	*Std.*
$\alpha_{4,1}^{\mathbb{P}}$	1.5611	*0.0887*	$\alpha_{4,1}^{\mathbb{Q}}$	1.5619	*0.0884*	$\alpha_{4,2}^{\mathbb{P}}$	1.6547	*0.0876*	$\alpha_{4,2}^{\mathbb{Q}}$	1.6566	*0.0872*
$\beta_{11,1}^{\mathbb{P}}$	0.9963	*0.0014*	$\beta_{11,1}^{\mathbb{Q}}$	0.9972	*7.6e-04*	$\beta_{11,2}^{\mathbb{P}}$	0.9917	*0.015*	$\beta_{11,2}^{\mathbb{Q}}$	0.9925	*6.84e-04*
$\beta_{22,1}^{\mathbb{P}}$	0.9986	*0.0011*	$\beta_{22,1}^{\mathbb{Q}}$	0.9988	*5.87e-04*	$\beta_{22,2}^{\mathbb{P}}$	0.9965	*0.0183*	$\beta_{22,2}^{\mathbb{Q}}$	0.9966	*5.4e-04*
$\beta_{33,1}^{\mathbb{P}}$	0.9995	*3.2e-04*	$\beta_{33,1}^{\mathbb{Q}}$	0.9993	*1.9e-04*	$\beta_{33,2}^{\mathbb{P}}$	0.9997	*0.0166*	$\beta_{33,2}^{\mathbb{Q}}$	0.9994	*1.49e-04*
$\beta_{44,1}^{\mathbb{P}}$	0.998	*0.0012*	$\beta_{44,1}^{\mathbb{Q}}$	0.9985	*6.72e-04*	$\beta_{44,2}^{\mathbb{P}}$	0.9985	*0.0322*	$\beta_{44,2}^{\mathbb{Q}}$	0.9991	
$\beta_{21,1}^{\mathbb{P}}$	0.0288	*0.0022*	$\beta_{21,1}^{\mathbb{Q}}$	0.0288	*0.0022*	$\beta_{21,2}^{\mathbb{P}}$	0.0192	*0.0015*	$\beta_{21,2}^{\mathbb{Q}}$	0.0192	*0.0023*
$\beta_{32,1}^{\mathbb{P}}$	0.0024	*2.23e-04*	$\beta_{32,1}^{\mathbb{Q}}$	0.0024	*2.22e-04*	$\beta_{32,2}^{\mathbb{P}}$	0.0066	*5.24e-04*	$\beta_{32,2}^{\mathbb{Q}}$	0.0066	*1.6e-04*
$\beta_{43,1}^{\mathbb{P}}$	2e-04	*1.65e-05*	$\beta_{43,1}^{\mathbb{Q}}$	2e-04	*1.65e-05*	$\beta_{43,2}^{\mathbb{P}}$	3.5e-04	*2.34e-05*	$\beta_{43,2}^{\mathbb{Q}}$	3.5e-04	*2.74e-06*
Constant Parameters											
θ_1	-8.49e-04	*7e-04*	θ_2	-1.37e-04	*5.44e-04*	θ_3	2.5e-04	*1.49e-04*	θ_4	-5.53e-04	*6.22e-04*
$\mu_1^{\mathbb{P}}$	1	-	$\mu_2^{\mathbb{P}}$	1	-	$\mu_3^{\mathbb{P}}$	1	-	$\mu_4^{\mathbb{P}}$	1	-
$\mu_1^{\mathbb{Q}}$	1.0008	*7e-04*	$\mu_2^{\mathbb{Q}}$	1.0001	*5.44e-04*	$\mu_3^{\mathbb{Q}}$	0.9997	*1.49e-04*	$\mu_4^{\mathbb{Q}}$	1.0006	*6.22e-04*
$\rho_{1,1}^{\mathbb{Q}}$	0.9934	*0.0032*	$\rho_{2,1}^{\mathbb{Q}}$	0.9968	*0.0015*	$\rho_{3,1}^{\mathbb{Q}}$	0.9992	*6.82e-04*	$\rho_{4,1}^{\mathbb{Q}}$	0.9996	*2.91e-04*
$\rho_{1,2}^{\mathbb{Q}}$	0.998	*3.09e-04*	$\rho_{2,2}^{\mathbb{Q}}$	0.9989	*1.59e-04*	$\rho_{3,2}^{\mathbb{Q}}$	0.999	*2.74e-05*	$\rho_{4,2}^{\mathbb{Q}}$	0.9991	*9.85e-06*
δ_1	0.0006	*1.97e-05*	σ_R	1.36e-04	*2.9e-07*						

Notes: The value of δ corresponds to a short rate in annualized percentage points and the measurement error corresponds to yields in annualized percentage points. The initial short rate has been calibrated to match approximately 5.14%. This is a reasonable value given the initial shape of the yield curve. Standard errors are computed from an outer product approximation of the Hessian.

5.3.2 US Term Structure

As for the EU, the regime-switching term structure model delivers interesting and good results. The optimization procedure seems to be unstable to a certain extent. The reason for this will most likely be the non-stationary behavior in one regime under $\mathbb{Q}$ which will be explained in more detail below. Nevertheless, the results presented here provide many useful insights which help to overcome this issue in future applications.

First, the $\alpha's$ clearly indicate the existence of a high and low yield regime. The level of the parameter values is overall at least twice as high in the second regime, reflecting the fact that a prolonged period of zero interest rates is not really likely in the second regime. Special attention has to be given towards the extreme persistence of the estimated parameters. In fact, the resulting parameter estimates for the autoregressive parameters in the first regime under $\mathbb{Q}$ show a non-stationary behavior while they are also above 0.999 in the second regime, further providing evidence for the extraordinary persistence of daily interest rates in the sample period. Nevertheless, even though the process is non-stationary in the first regime under $\mathbb{Q}$, the general process could still be stationary. Whether this holds or not depends on geometric ergodicity and an associated drift condition. However, this characteristic of the process is beyond the scope of this thesis.[24]

Even though the process is non-stationary, the filtered factor trajectories reveal important information about US interest rate dynamics over the last decade. Figure B.5 depicts the filtered factor estimates of the term structure model. As for the EU, the first and second factor experience extremely long and persistent periods of zero. This already offers one possible explanation for the non-stationary in the first regime as two factors virtually stayed constant over 6 years. Furthermore, the zero-period is fairly consistent with the zero interest period of the Fed. Also, the factors are able to get out of their zero state when the Fed starts with the liftoff. Likewise, the first factor pretty much corresponds to the short rate movements used to identify the regime-structure. The third factor experiences only a short period of zero which corresponds exactly to the time when the short rate starts to liftoff. Such a behavior also makes economically sense. If the four factors used in the estimation are seen as risk factors, a change in the long-term behavior of one of those underlying factors could make the market much more sensitive to this factor, such that over an extremely short period the third factor becomes irrelevant in generating fluctuations. Again, the fourth factor seems to be needed to match the conditional volatility pattern of long-term yields and thus fluctuates quite heavily over time. Contrary to the European results, the fourth factor experiences an extensive zero period during which medium and long term rates have been at their lowest level (mid 2012 to mid 2013). Also, Roussellet (2016) found

[24] A good reference about geometric ergodicity is given in Nummelin and Tweedie (1978).

that during this period the lift-off risk premium became negative, meaning the market expected a liftoff or it has been seen as a desirable event. Interestingly, this period closely corresponds to the time when European short term rates entered their respective zero lower bound as an answer to the sovereign debt crisis. This motivates the hypothesis that even though normally the state of the US spills over to other countries, the US has been explicitly vulnerable to external economic conditions during their zero interest rate period. This movement then leads to an interesting conjecture, namely that the fourth risk factor acts as a summary of external risk for the US or a term-premium residual.

Of particular interest are also the regime specific factor loadings for the US. While the loadings for the first three factors show the same behavior as the ones from the EU, the fourth one is completely different. Thus, for the US, also the fourth factor exhibits a non-linear behavior. If the fourth factor somehow represents external risk, the depicted loading could somehow make economically sense. As the US has been in a deep recession during the first regime, monetary policy became much more expansive to stimulate inflation and as rates dropped to zero, monetary policy was not able anymore to react to certain economic conditions with changes in the rate such that the economy became much more vulnerable to external movements.

The estimated market prices of risk show positive risk premia for factors two to four while the first one has negative risk premium. As the first factor pretty closely resembles the short rate, it seems that the market placed a totally different price on short rate movements. This could possibly be due to the extensive communication the Fed has relied upon over the last decade such that short rate movements became much more predictive.

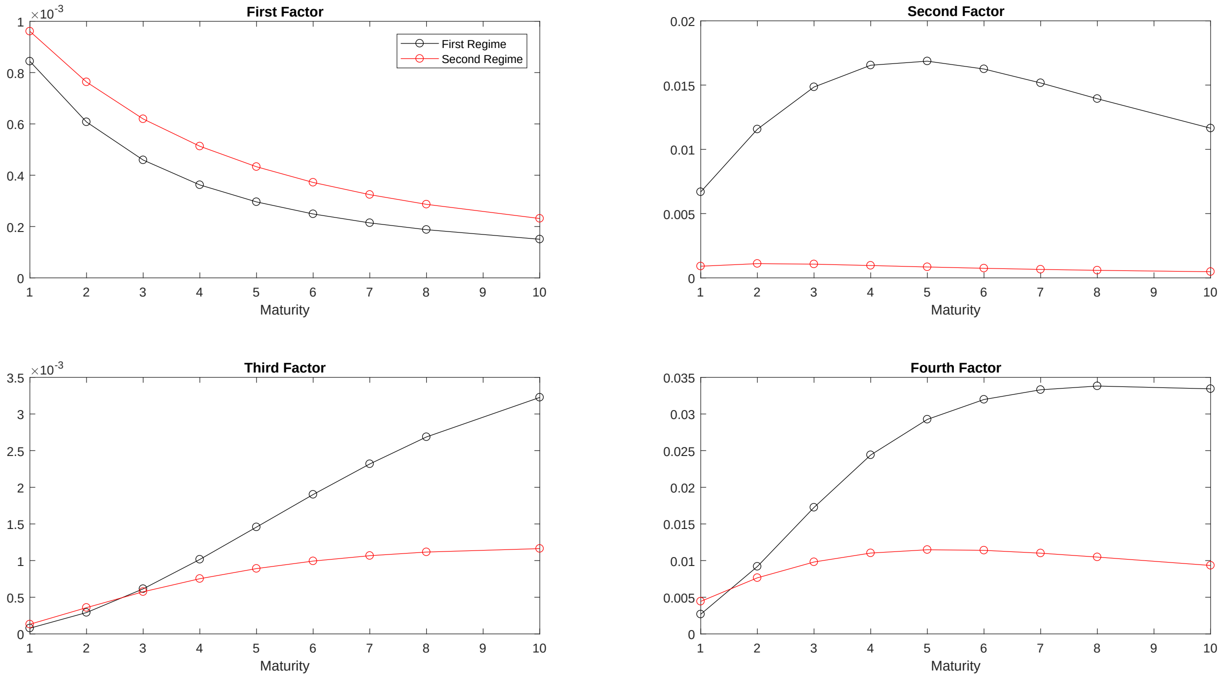


Figure 8: Factor Loadings US Yields

Notes: The plotted factor loadings correspond to yields expressed in annualized percentage points.

Overall, the results for the regime-switching term structure model are quite satisfactory. They are able to match the dynamics of the yield curve fairly closely while offering some economic interpretation. While the regime-switching specification for the EU provides stationary results, the same specification for the US results in a non-stationary $\mathbb{Q}$-distribution. This shows that even though the regime-switching specification is remarkably flexible, the model has its limitations. The reason for this will most likely be the extreme persistence of the binding zero lower bound. The autocorrelation patterns for the daily interest rates also suggest such a strong persistence causing a trend in the data that has to be matched. As this field of research is relatively new, not many explanations have been found and given yet. The observed persistence and pattern is of utmost importance as it most likely can be traced back to economic conditions such that models like the one developed in Roussellet (2016) will become crucial to understand the causes of this extreme persistence.

Table 5: Parameter Estimates: US

Regime 1						Regime 2					
	Value	*Std.*		Value	*Std.*		Value	*Std.*		Value	*Std.*
$\alpha^{\mathbb{P}}_{1,1}$	1.36e-04	*0.0232*	$\alpha^{\mathbb{Q}}_{1,1}$	1.36e-04	*0.0232*	$\alpha^{\mathbb{P}}_{1,2}$	0.0073	*0.266*	$\alpha^{\mathbb{Q}}_{1,2}$	0.0073	*0.2661*
$\alpha^{\mathbb{P}}_{2,1}$	9.96e-06	*0.0019*	$\alpha^{\mathbb{Q}}_{2,1}$	9.96e-06	*0.0019*	$\alpha^{\mathbb{P}}_{2,2}$	0.017	*0.0255*	$\alpha^{\mathbb{Q}}_{2,2}$	0.017	*0.0255*
$\alpha^{\mathbb{P}}_{3,1}$	0.0588	*0.0052*	$\alpha^{\mathbb{Q}}_{3,1}$	0.0588	*0.0052*	$\alpha^{\mathbb{P}}_{3,2}$	0.4624	*0.0611*	$\alpha^{\mathbb{Q}}_{3,2}$	0.4265	*0.0611*
$\alpha^{\mathbb{P}}_{4,1}$	0.0102	*0.0013*	$\alpha^{\mathbb{Q}}_{4,1}$	0.0102	*0.0013*	$\alpha^{\mathbb{P}}_{4,2}$	0.0224	*0.0074*	$\alpha^{\mathbb{Q}}_{4,2}$	0.0224	*0.0074*
$\beta^{\mathbb{P}}_{11,1}$	0.9986	*0.0012*	$\beta^{\mathbb{Q}}_{11,1}$	0.9981	*5.87e-04*	$\beta^{\mathbb{P}}_{11,2}$	0.9994	*0.047*	$\beta^{\mathbb{Q}}_{11,2}$	0.999	*6e-04*
$\beta^{\mathbb{P}}_{22,1}$	0.9999	*0.0038*	$\beta^{\mathbb{Q}}_{22,1}$	1.0003	*0.002*	$\beta^{\mathbb{P}}_{22,2}$	0.9958	*0059*	$\beta^{\mathbb{Q}}_{22,2}$	0.9962	*0.0019*
$\beta^{\mathbb{P}}_{33,1}$	0.9997	*0.0017*	$\beta^{\mathbb{Q}}_{33,1}$	1.0001	*8.52e-04*	$\beta^{\mathbb{P}}_{33,2}$	0.9993	*0.064*	$\beta^{\mathbb{Q}}_{33,2}$	0.9996	*8.5e-04*
$\beta^{\mathbb{P}}_{44,1}$	0.9978	*8.4e-04*	$\beta^{\mathbb{Q}}_{44,1}$	0.9981	*4.5e-04*	$\beta^{\mathbb{P}}_{44,2}$	0.9994	*0.03*	$\beta^{\mathbb{Q}}_{44,2}$	0.9997	*4.3e-04*
$\beta^{\mathbb{P}}_{21,1}$	0.0355	*0.0014*	$\beta^{\mathbb{Q}}_{21,1}$	0.0355	*0.0014*	$\beta^{\mathbb{P}}_{21,2}$	0.0071	*8.9e-04*	$\beta^{\mathbb{Q}}_{21,2}$	0.0071	*8.91e-04*
$\beta^{\mathbb{P}}_{31,1}$	3.7e-08	*2.4e-05*	$\beta^{\mathbb{Q}}_{31,1}$	3.7e-08	*2.4e-05*	$\beta^{\mathbb{P}}_{31,2}$	3.31e-05	*3.39e-04*	$\beta^{\mathbb{Q}}_{31,2}$	3.32e-05	*3.39e-04*
$\beta^{\mathbb{P}}_{41,1}$	1.22e-06	*5.8e-04*	$\beta^{\mathbb{Q}}_{41,1}$	1.22e-06	*5.7e-04*	$\beta^{\mathbb{P}}_{41,2}$	0.0234	*0.0047*	$\beta^{\mathbb{Q}}_{41,2}$	0.0234	*0.0047*
$\beta^{\mathbb{P}}_{32,1}$	9e-05	*2e-06*	$\beta^{\mathbb{Q}}_{32,1}$	9e-05	*2e-06*	$\beta^{\mathbb{P}}_{32,2}$	0.001	*3.64e-05*	$\beta^{\mathbb{Q}}_{32,2}$	0.001	*3.62e-05*
$\beta^{\mathbb{P}}_{42,1}$	0.0034	*1.4e-04*	$\beta^{\mathbb{Q}}_{42,1}$	0.0034	*1.37e-04*	$\beta^{\mathbb{P}}_{42,2}$	3e-04	*7.77e-04*	$\beta^{\mathbb{Q}}_{42,2}$	3e-04	*7.77e-04*
$\beta^{\mathbb{P}}_{43,1}$	0.0421	*0.005*	$\beta^{\mathbb{Q}}_{43,1}$	0.0421	*0.005*	$\beta^{\mathbb{P}}_{43,2}$	0.0041	*0.0038*	$\beta^{\mathbb{Q}}_{43,2}$	0.0041	*0.0038*
Constant Parameters											
θ_1	4.14e-04	*5.8e-04*	θ_2	-3.99e-04	*0.0019*	θ_3	-3.51e-04	*8.5e-04*	θ_4	-2.63e-04	*4.2e-04*
$\mu^{\mathbb{P}}_1$	1	-	$\mu^{\mathbb{P}}_2$	1	-	$\mu^{\mathbb{P}}_3$	1	-	$\mu^{\mathbb{P}}_4$	1	-
$\mu^{\mathbb{Q}}_1$	0.9996	*5.8e-04*	$\mu^{\mathbb{Q}}_2$	1.0004	*0.0019*	$\mu^{\mathbb{Q}}_3$	1.0004	*8.5e-04*	$\mu^{\mathbb{Q}}_4$	1.0003	*4.2e-04*
$\rho^{\mathbb{Q}}_{1,1}$	0.9977	*1.18e-04*	$\rho^{\mathbb{Q}}_{2,1}$	1.0007	*9.78e-05*	$\rho^{\mathbb{Q}}_{3,1}$	1.0004	*2.32e-05*	$\rho^{\mathbb{Q}}_{4,1}$	0.9983	*2.28e-04*
$\rho^{\mathbb{Q}}_{1,2}$	0.9985	*4.85e-04*	$\rho^{\mathbb{Q}}_{2,2}$	0.9966	*0.0014*	$\rho^{\mathbb{Q}}_{3,2}$	1-1e-05	*7.70e-05*	$\rho^{\mathbb{Q}}_{4,2}$	0.9999	*1.36e-04*
δ_1	0.0012	*4.56e-05*	σ_R	5.86e-05	*2.55e-07*						

Notes: The value of δ corresponds to a short rate in annualized percentage points and the measurement error corresponds to yields in annualized percentage points. The parameter estimates correspond to a fully parametrized model. Standard errors are computed from an outer product approximation of the Hessian.

6 Validation

In order to assess the validity of the model in general and the chosen four factor specification, the fit of the model is compared to a four factor linear version and to a regime-switching three factor model. The comparison is done on a Root Mean Squared Error (RMSE) basis of the pricing errors as well as on an assessment of the BIC. This makes especially sense for a comparison to the linear model as it is nested in the four-factor regime-switching model.

Table 6 presents the results for the various models for both areas. Looking at the US, it is evident that the regime-switching model performs superior to the other two models for every yield used in the estimation. The RMSE's do not exceed two basis points which show the incredible fit of the RS-TSM. The extreme improvement in the fit compared to the three factor regime-switching model shows that at least four factors are needed in order to match the dynamics of the yield curve. An extremely striking observation is the nearly 300% RMSE improvement for the one year yield when moving from a linear to a non-linear model. This also makes sense as the regime-switching dynamics are estimated on a short rate proxy and shorter maturity yields are expected to follow more closely the short rate.

The improvement on the BIC is also compelling, clearly speaking in favor of the regime-switching counterpart. Interestingly, the three factor regime-switching specification performs worse on a BIC basis for the US compared to the linear model as well as for most of the maturities. Even though the BIC suggests to rather go with the linear model, the improvement for the one year yield is worth mentioning. This clearly indicates that the chosen regime specification is indeed accurate for the US if one wants to model the short end of the yield curve.

As the regime-switching model is non-stationary it is worth looking at the dynamics under both distributions for the linear model. The results presented in Table C.2 show again an extreme persistence of the interest rate process in the risk-neutral world. Furthermore, as Jardet et al. (2013) and Kim and Singleton (2012) demonstrate, persistent processes face a downward bias in finite sample ML. This combined with potential standard errors clearly indicate that the linear model estimates would render a non-stationary model too. The model assessment results for the EU closely follow those of the US. The four factor regime-switching specification performs by far the best, also showing the lowest RMSE for nearly every maturity (the only exception is the 10 year rate). Again, one can see that the improvement at the short end of the yield curve is the most impressive, leading to the same conclusion as for the United States. The BIC again points towards a four factor regime-switching specification as the most accurate. As for the US, the regime-switching 3-factor specification is clearly the worst, again providing evidence that four factors are needed to capture the dynamics of the whole term structure.

Regarding the persistence problem identified in the US data, we can see that for the European

Table 6: Model Comparison

Panel (a): United States RMSE							
	Maturities 1Y	3Y	5Y	6Y	8Y	10Y	BIC
RS 4 factor	1.45	1.71	1.17	1.07	1.28	1.47	-567'240
RS 3 factor	3.16	2.92	1.84	2.23	2.38	2.34	-538'850
Lin 4 factor	3.8	2.61	1.8	2.05	1.97	2.53	-543'010
Panel (b): European Union RMSE							
	Maturities 0.5Y	1Y	2Y	5Y	7Y	10Y	BIC
RS 4 factor	1.57	2.62	3.38	1.88	1.57	2.09	-282'680
RS 3 factor	2.6	4.08	4.16	2.71	2.06	3.11	-277'800
Lin 4 factor	2.42	4.1	4.15	1.91	1.76	1.9	-279'610

Notes: The RMSE's for the yields are expressed in basis points.

case the regime structure helps to resolve the unit-root problem associated with the non-linear behavior of the data. While the autoregressive parameters for the linear model, see Table C.2, are extremely close to one under both distributions their regime-switching counterpart looks much more stable. This points towards the ability of the model to accommodate non-linear behavior in general. One reason for the resulting stationary process could again be that the zero state is not as persistent for the EU data as for the US.

Thus, based on a pricing error basis as well as on model selection criteria, the regime-switching model seems to be the most accurate and reasonable description of the data at hand. Besides overall indication, the fact that the linear one performs the worst for the short maturities clearly provides the strongest evidence in favor of regime-switching dynamics due to changes in monetary policy. In addition, it also indicates that the modeling procedure for the monetary policy proxy is accurate enough to capture changing market behavior.

Nevertheless, the fact that the improvement is not that striking for the long-end of the yield curve clearly rejects the expectation hypothesis and also points towards the limitation of conventional monetary policy of affecting long-term yield movements. On the other hand, if unconventional monetary policy had an effect on long-term yields then in order to match yield dynamics to monetary policy one would have to explicitly model the different policy instruments. This reasoning provides a good starting point for further improvement of the model at hand.

7 Conclusion

In this thesis, a new non-linear version of the well known important ARG process of Gourieroux and Jasiak (2006) and its zero complement of Monfort et al. (2017) in form of a regime-switching extension has been introduced. The subsequent empirical application to European and US term structure data shows a clear superior performance of the new process as interest rate dynamics have experienced a change in their behavior due to the change in monetary policy following the financial crisis.

These dynamics are captured by using market based expectations about future monetary policy from option implied densities on policy rate futures. The resulting regime structure differentiates between a high and low yield environment which confirms the hypothesis that interest rates behave differently depending on the state of the economy. Furthermore, the results point towards a similar behavior of interest rates in the US and EU. This reflects the global impact of the financial crisis on interest rates around the world. In addition, the usage of option implied densities makes this approach extremely viable for financial applications, especially in a forecasting context. This property of the approach is exceptionally effective as usually macroeconomic variables are used in this context, rendering the model unusable for short-term forecasting. Given the extreme persistence of daily interest rates, improving their forecasting performance has many possible use cases.

The elegant nature of the regime-switching extension preserves all properties of the nested linear model. In particular, this means that the affine pricing formulas are preserved and a linear Kalman filter can be used to approximate the likelihood function in the estimation stage. Furthermore, the resulting process still belongs to the Car(1) type of processes, which have proven to be extremely useful in discrete-time asset pricing.

While the fit of the model on a level basis shows a particularly satisfying performance, many areas remain unexplored as this would be out of scope of this thesis. Cieslak and Povala (2016) and Jacobs and Karoui (2009) have shown that estimation on only yield levels fail to match the second conditional moment in a satisfying manner. Thus, it would be of particular importance to look more closely at the volatility dynamics of the regime-switching model. Monfort et al. (2017) showed that due to the linearity of the state-space representation it can easily be augmented to include volatility measurement equations. The linear EGARCH proxies used in their empirical application can easily be replaced by a tree structured GARCH model which has been estimated in the context of interest rate modeling by Audrino and Medeiros (2011). Exploring the models ability to generate closed form liftoff formulas would be a second interesting point. As option implied densities are used to construct the regimes, the model implied lift-off probabilities should provide a more accurate description of actual lift-off probabilities. Roussellet (2016) shows that the market price of risk associated with a potential lift-off changed over time as well as its economic costs. The time-varying parameters of the model should help to match those dynamics. Given the regime structure, an application to derivatives pricing should prove extremely helpful as option moments directly characterize the dynamics. It would therefore be natural to apply this in an option pricing context.

While the regime-structure helped to improve upon the stationarity issues associated with a linear model, the US results show an enormous persistence in interest rates on a daily level. This is most likely due to extreme persistence of the ZLB. One possibility to handle this issue could be to include survey based forecasts in the model. Kim and Orphanides (2012) introduced this approach and showed that this helps to cope with the persistence of interest rate data.

Last, one possible extension would be to directly include observable variables in the Poisson specification. For example, it is well known that interest rates react to changes in inflation but as soon as the short rate enters the effective lower bound, it cannot really react, such that a regime-dependent parameter on inflation could help to reflect unconventional monetary policy in the model as well as possibly help overcoming the persistence issue identified in the US.

The vast amount of possible extensions and directions for future show the extreme flexibility and tractability the non-linear model has to offer. Especially in the spirit of macro-finance models, matching dynamics to economic conditions can not be neglected.

Glossary

ARG-Zero Autoregressive Gamma-zero

BIC Bayesian Information Criteria

Car Compund autoregressive process

CME Chicago Mercantile Exchange

ECB European Central Bank

GATSM Gaussian Affine Term Structure Model

GMM Generalized Methods of Moments

ICE Intercontinental Exchange

LIBOR London Interbank Offered Rate

pdf probability density function

QML Quasi-maximum liklihood

QTSM quadratic term structure models

RMSE Root Mean Squared Error

RS-ARG Regime-Switching ARG

RS-TSM Regime-Switching Term Structure Model

OTM Out-of-the-Money

SDF Stochastic Discount Factor

SNB Swiss National Bank

ZCB Zero Coupon Bond

ZLB Zero Lower Bond

8 References

Audrino, F. (2006). Tree-structured multiple regimes in interest rates. *Journal of Business & Economic Statistics*(3), 338–353.

Audrino, F., & Buehlmann, P. (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*(4), 727–744.

Audrino, F., & Medeiros, M. C. (2011). Modeling and forecasting short-term interest rates: the benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*(6), 999–1022.

Bauer, M. D. (2014). Options-based expectations of future policy rates. *FRBSF Economic Letter*.

Bertholon, H., Monfort, A., & Pegoraro, F. (2008). Econometric asset pricing modelling. *Journal of Financial Econometrics*, *6*(4), 407–458.

Bikbov, R., & Chernov, M. (2013). Monetary policy regimes and the term structure of interest rates. *Journal of Econometrics*, *174*, 27–43.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, *3*(1/2), 167–179.

Black, F. (1995). Interest rates as options. *The Journal of Finance*(5), 1371–1376.

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

Bliss, R. R., & Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking and Finance*, *26*, 381–422.

Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*(4), 621–651.

Buraschi, A., Trojani, F., & Vedolin, A. (2014). When uncertainty blows in the orchard: Comovement and equilibrium volatility risk premia. *Journal of Finance*, *69*(1), 101–137.

Chen, C. Y.-H., & Kuo, I.-D. (2014). Investor sentiment and interest rate volatility smile: Evidence from eurodollar options markets. *Review of Quantitative Finance and Accounting*, *43*(2), 367–391.

Chen, R.-R., & Scott, L. (1993). Pricing interest rate futures options with futures-style margining. *Journal of Futures Markets*, *13*(1), 15–22.

Christensen, J. H., Lopez, J. A., & Rudebusch, G. D. (2014). Can spanned term structure factors drive stochastic yield volatility?. *Working Paper Series*.

Christensen, J. H. E., & Rudebusch, G. D. (2016). Modeling yields at the zero lower bound: Are shadow rates the solution? In E. Hillebrand & S. J. Koopman (Eds.), *Dynamic factor* *models* (Advances in Econometrics ed., Vol. 35, pp. 75–125). Emerald Group Publishing Limited.

Christensen, J. H. E., & Rudebusch, G. D. (2017). New evidence for a lower new normal in interest rates. *FRBSF Economic Letter*, *2017*(17), 1–5.

Cieslak, A., & Povala, P. (2016). Information in the term structure of yield curve volatility. *Journal of Finance*, *71*(3), 1393–1436.

Cochrane, J. H. (2005). *Asset pricing.* Princeton, N.J. Princeton University Press, 2005.

Cochrane, J. H. (2017). Macro-finance. *REVIEW OF FINANCE*(3), 945–985.

Cook, D., & Devereux, M. B. (2016). Exchange rate flexibility under the zero lower bound. *Journal of International Economics*, *101*, 52–69.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*(2), 385–407.

Darolles, S., Gourieroux, C., & Jasiak, J. (2006). Structural laplace transform and compound autoregressive models. *Journal of Time Series Analysis*, *27*(4), 477–503.

Da Rosa, J. C., Veiga, A., & Medeiros, M. C. (2008). Tree-structured smooth transition regression models. *Computational Statistics and Data Analysis*, *52*, 2469–2488.

De Graeve, F., & Lindé, J. (2015). Effects of unconventional monetary policy: theory and evidence. *Sveriges Riksbank Economic Review*(1), 43–74.

de Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics*(3), 300–314.

Di Maggio, M., & Kacperczyk, M. (2017). The unintended consequences of the zero lower bound policy. *Journal of Financial Economics*, *123*, 59–80.

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

Feller, W. (1950). *An introduction to probability theory and its applications.* New York Wiley, 1950-1966.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, *9*(4), 417–428.

Filipovic, D., Larsson, M., & Trolle, A. B. (2017). Linear-rational term structure models. *Journal of Finance*, *72*(2), 655–704.

Gebbia, N. (2016). Option-implied libor rate expectations across currencies. *Working Papers – U.S. Federal Reserve Board's International Finance Discussion Papers*(1182), 1–44.

Gourieroux, C., & Jasiak, J. (2006). Autoregressive gamma processes. *Journal of Forecasting*, *25*(2), 129–152.

Gourieroux, C., Monfort, A., Pegoraro, F., & Renne, J.-P. (2014). Regime switching and bond pricing. *Journal of Financial Econometrics*, *12*(2), 237–277.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, *54*, 2291–2304.

Hall, R. E. (2016). Search-and-matching analysis of high unemployment caused by the zero lower bound. *Review of Economic Dynamics*, *19*(Special Issue in Honor of Dale Mortensen), 210–217.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*(2), 357–384.

Jacobs, K., & Karoui, L. (2009). Conditional volatility in affine term-structure models: Evidence from treasury and swap markets. *Journal of Financial Economics*, *91*, 288–318.

Jardet, C., Monfort, A., & Pegoraro, F. (2013). No-arbitrage near-cointegrated VAR(p) term structure models, term premia and GDP growth. *Journal of Banking and Finance*, *37*, 389–402.

Kameliya, F., Audrino, F., & De Giorgi, E. (2014). Monetary policy regimes: Implications for the yield curve and bond pricing. *Journal of Financial Economics*, *113*, 427–454.

Kim, D. H., & Orphanides, A. (2012). Term structure estimation with survey data on interest rate forecasts. *Journal of Financial & Quantitative Analysis*, *47*(1), 241–272.

Kim, D. H., & Singleton, K. J. (2012). Term structure models and the zero bound: An empirical investigation of japanese yields. *Journal of Econometrics*, *170*, 32–49.

Krippner, L. (2012). Modifying gaussian term structure models when interest rates are near the zero lower bound (this is a revised version of cama working paper 36/2011). *CAMA Working Papers*.

Le, A., Singleton, K. J., & Dai, Q. (2010). Discrete-time affine term structure models with generalized market prices of risk. *Review of Financial Studies*, *23*(5), 2184–2227.

Lemke, W. (2006). *Term structure modeling and estimation in a state space framework.* Berlin Springer, 2006.

Lien, D., & Li, Y. (2003). Options expiration effects and the role of individual share futuress contracts. *Journal of Futures Markets*, *23*(11), 1107–1118.

Longstaff, F. A., & Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance*(4), 1259–1282.

Mirkov, N., Pozdeev, I., & Söderlind, P. (2016). Toward removal of the swiss franc cap: Market expectations and verbal interventions. *Swiss National Bank Working Papers*(10), 1–31.

Monfort, A., & Pegoraro, F. (2007). Multi-lag term structure models with stochastic risk premia. *Working papers*.

Monfort, A., Pegoraro, F., Renne, J.-P., & Roussellet, G. (2017). Staying at zero with affine processes: An application to term structure modelling. *Journal of Econometrics*, *201*(Theoretical and Financial Economentrics: Essays in Honor of C. Gourieroux), 348–366.

Monfort, A., & Renne, J.-P. (2014). Compound autoregressive processes and defaultable bond pricing. In J. S. Chadha, A. Durr, M. Joyce, & L. Sarno (Eds.), *Developments in macro-finance yield curve modelling* (p. 141 - 168). Cambridge University Press.

Nummelin, E., & Tweedie, R. L. (1978). Geometric ergodicity and r-positivity for general markov chains. *The Annals of Probability*(3), 404–420.

Piazzesi, M. (2010). Chapter 12: Affine term structure models. *Handbook of Financial Econometrics, Vol 1*, 691–766.

Roussellet, G. (2016). Affine term structure modeling and macroeconomic risks at the zero lower bound (november 2, 2016). *Available at SSRN*. Retrieved from `https://ssrn.com/abstract=2863271`

Shimko, D. (1993). Bounds of probability. *Risk*, 33–37.

Shreve, S. E. (2004). *Stochastic calculus for finance II : continuous-time models.* New York Springer, 2004.

Svensson, L. E. (2010). Chapter 22: Inflation targeting. *Handbook of Monetary Economics*, *3*, 1237–1302.

Sveriges Riksbank. (n.d.). *The tasks of the riskbank.* (Retrieved on April 01, 2018, from https://www.riksbank.se/en-gb/about-the-riksbank/the-tasks-of-the-riksbank/)

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 195–214.

Trolle, A. B., & Schwartz, E. S. (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies*, *22*(5), 2007–2057.

Vähämaa, S. (2005). Option-implied asymmetries in bond market expectations around monetary policy actions of the ecb. *Journal of Economics and Business*, *57*, 23–38.

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

A Theoretical Appendix

A.1 Risk-neutral Laplace Transform and Closed-Form Yield Formula

Proof. of Proposition 3.3.
The proof follows along the lines of Monfort et al. (2017), page 364. Given that the one-period unobservable short rate is equal to $r_t = \delta' X_t$, where the n_1 components are different from zero, and assuming that the basic asset pricing equation 4 holds, we get the following result:

$$\begin{aligned}
P_t(h) &= \exp(A'_{t,h} X_t + B_{t,h}) \\
&= \mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r_t) \exp(A'_{t,h-1} X_{t+1} + B_{t,h-1}) \right] \\
&= \exp(-r_t + B_{t,h-1}) \mathbb{E}_t^{\mathbb{Q}} \left[\exp(A'_{t,h-1} X_{t+1}) \right] \\
&= \exp \left[B_{t,h-1} + \sum_{j=1}^{n} b_{j,t}^{\mathbb{Q}} (A_{j,t,h-1}) \right. \\
&= + \left. \left(\sum_{j=1}^{n} a_{j,t}^{\mathbb{Q}} (A_{j,t,h-1} - \delta) \right)' X_t \right]
\end{aligned}$$

The result for equations 5 and 6 follows immediately by identifying the coefficients and thereby prove the result. □

A.2 Modified Kalman Filter due to Monfort et al. (2017)

Given a vector of observed variables Y_t, the measurement equation is of the form:

$$Y_t = \Gamma_{t,0} + \Gamma_{t,1} X_t + \Omega_t^{1/2} \eta_t,$$

with η_t being a martingale difference sequence with zero-mean and identity variance-covariance matrix. The modified filter they present allows for deterministic time-variation in the matrices, which is particularly useful in my case due to the time-varying parameters of the process itself. One point which is different to the usual Kalman Filter is that the matrix $\Sigma(X_{t-1})$ from the transition equation depends on past values, and hence the Kalman filter isn't optimal anymore. To overcome the issue, the Kalman Filter is slightly modified but still employs a recursive forecasting and updating procedure. In particular, let us denote by $P_{t|\tau}$ the modified filter estimate of the conditional variance-covariance matrix of X_t, given $\underline{Y_\tau}$. This directly leads us to the following forecasting and updating equations:

$$\begin{cases} X_{t+1|t} & = m_t + M_t X_{t|t} \\ P_{t+1|t} & = M_t P_{t|t} M_t' + \Sigma_t(X_{t|t}) \\ Y_{t+1|t} & := \Gamma_{t+1,0} + \Gamma_{t+1,1} X_{t+1|t} \\ F_{t+1|t} & := \Gamma_{t+1,1} P_{t+1|t} \Gamma_{t+1,1}' + \Sigma_{t+1} \end{cases}$$

The updating is given by:

$$\begin{cases} X_{t+1|t+1} & = \left[X_{t+1|t} + P_{t+1|t} \Gamma_{t+1,1}' F_{t+1|t}^{-1} (Y_{t+1} - Y_{t+1|t}) \right]_+ \\ P_{t+1|t+1} & = P_{t+1|t} - Pt + 1|t \Gamma_{t+1,1}' F_{t+1|t}^{-1} \Gamma_{t+1,1} P_{t+1|t}, \end{cases}$$

where $[X]_+$ is a vector whose ith element is $\max(x_i, 0)$.
As initial values, I use the marginal moments at time t=0, so the day before the first estimation is done. The approximation of the quasi log-likelihood is given by:

$$-\frac{1}{2} \sum_{t=1}^{T} \left[\ln\{(2\pi)^{n_t} |F_{t|t-1}|\} + (Y_t - Y_{t|t-1})' F_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \right] \tag{13}$$

A.3 Proof of Corollary 3.2.1

Given Corollary 3.1.1 Monfort et al. (2017), the process is $\mathbb{Q}$-stationary if $\rho < 1$ given a lower triangular β matrix. At each point in time, β_t is a convex combination of β_j for $j = 1, ..., n$ as $\sum_{i=1}^{n} B_{\mathbb{J}i}(\boldsymbol{l}; \boldsymbol{\theta}) = 1$. ρ_t is bounded from above by $\max(\beta_j \mu)$, for $j = 1, ...n$ such that $\rho_t < 1$ if both regimes are stationary itself.

B Figure Appendix

B.1 Transition Probabilities

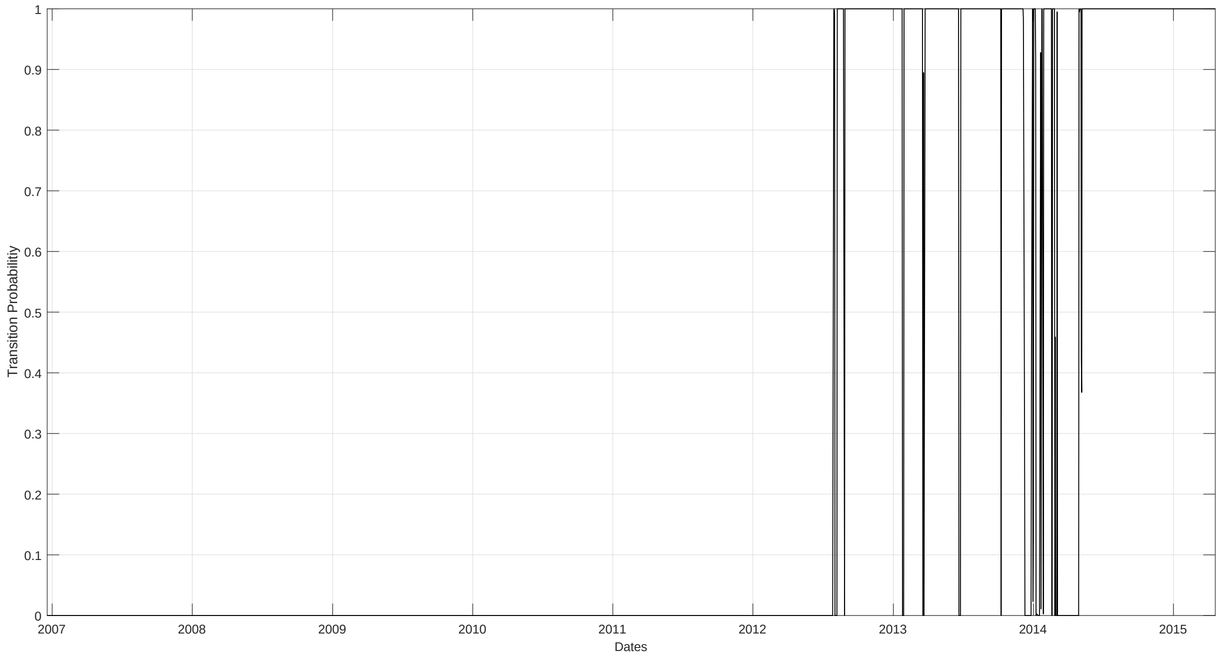


Figure B.1: EU Transition Probabilities

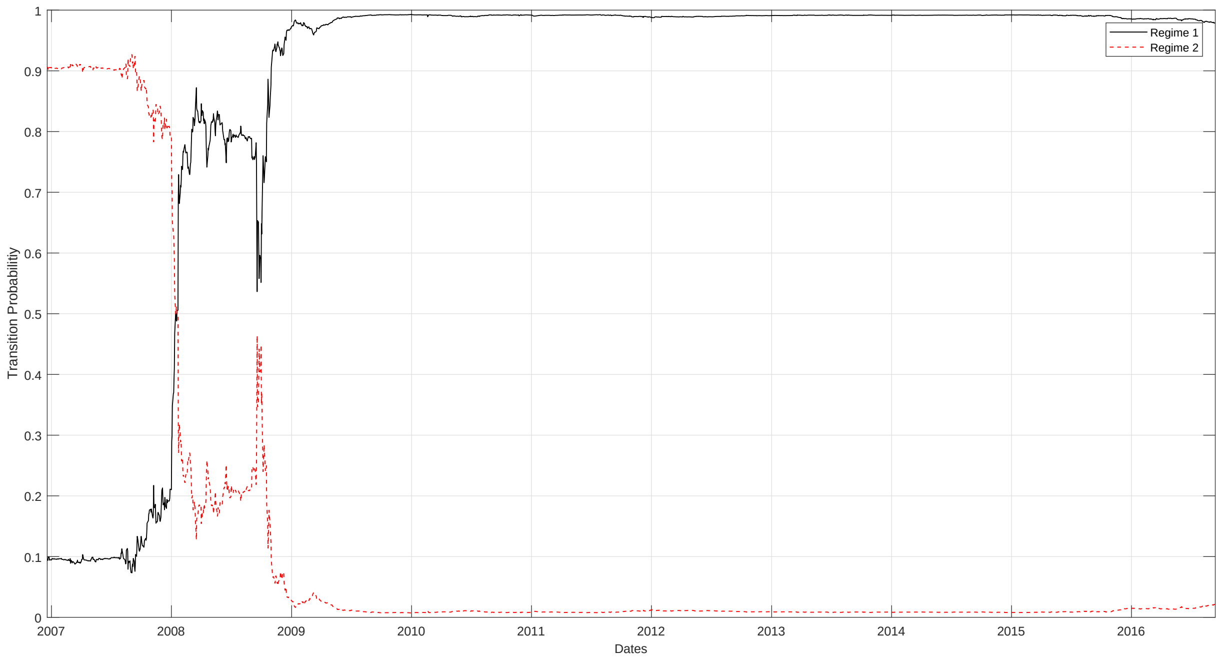


Figure B.2: US Transition Probabilities

B.2 Model Implied Yield Trajectories

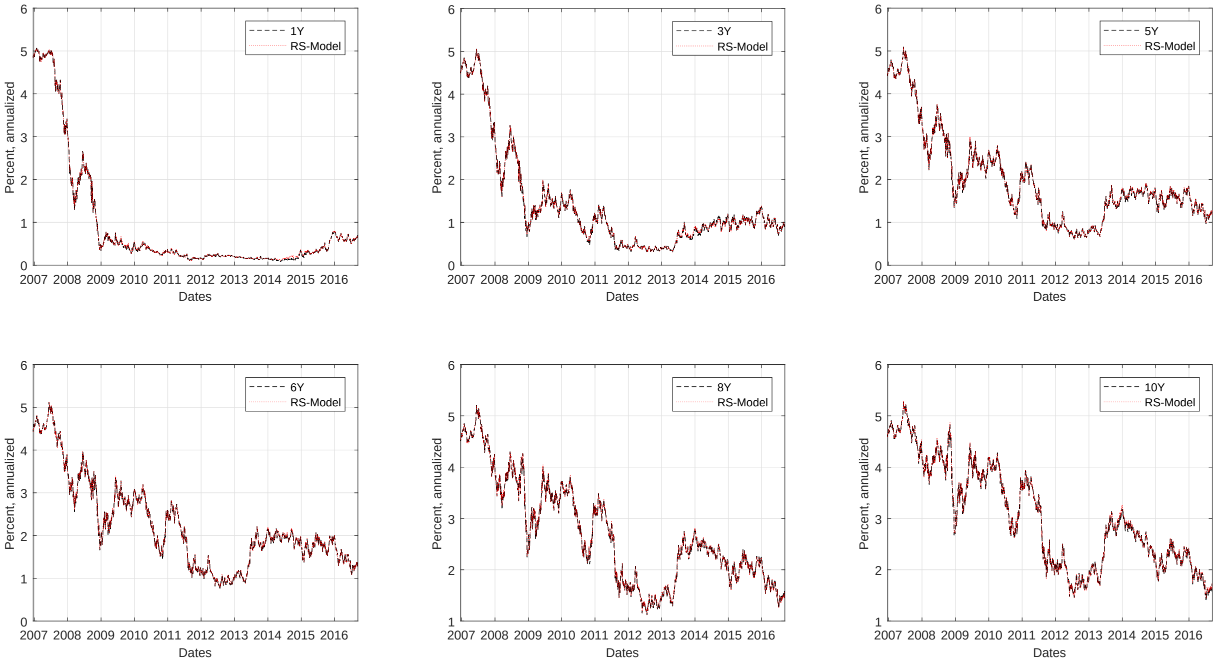


Figure B.3: US Model Implied Yields

Notes: Yields are expressed in annualized percentage points. Due to the extremely small measurement error, the difference between model implied yields and real data is nearly not visible.

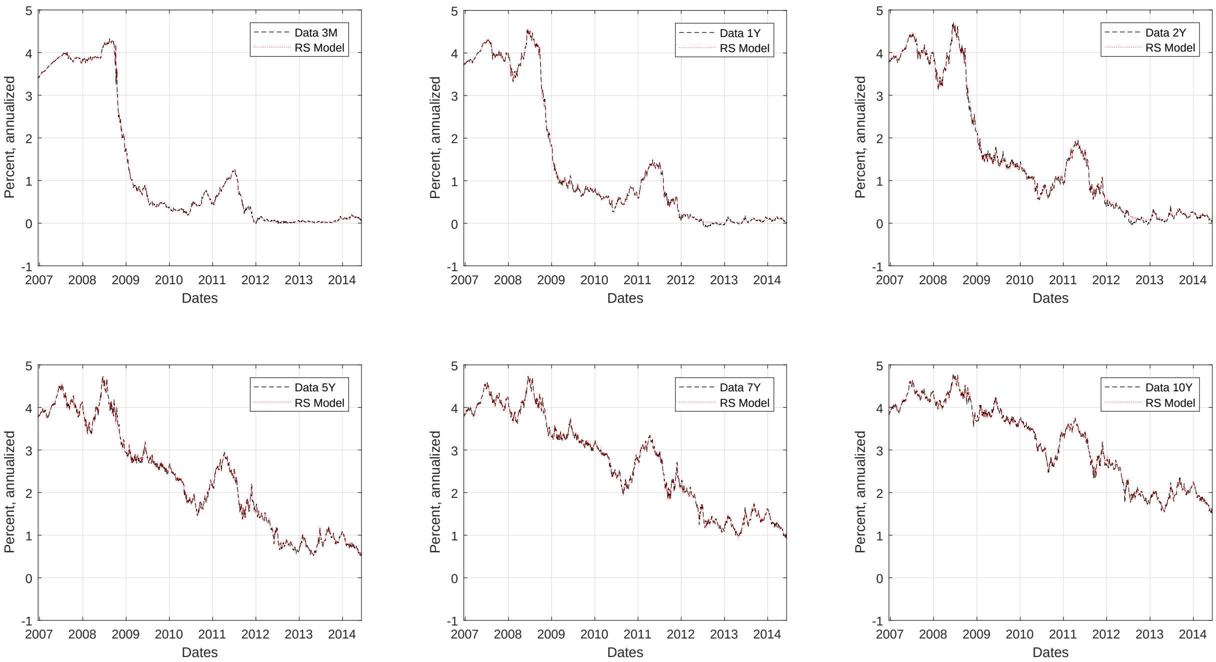


Figure B.4: EU Model Implied Yields

Notes: Yields are expressed in annualized percentage points. Due to the extremely small measurement error, the difference between model implied yields and real data is nearly not visible.

B.3 Factor Trajectories

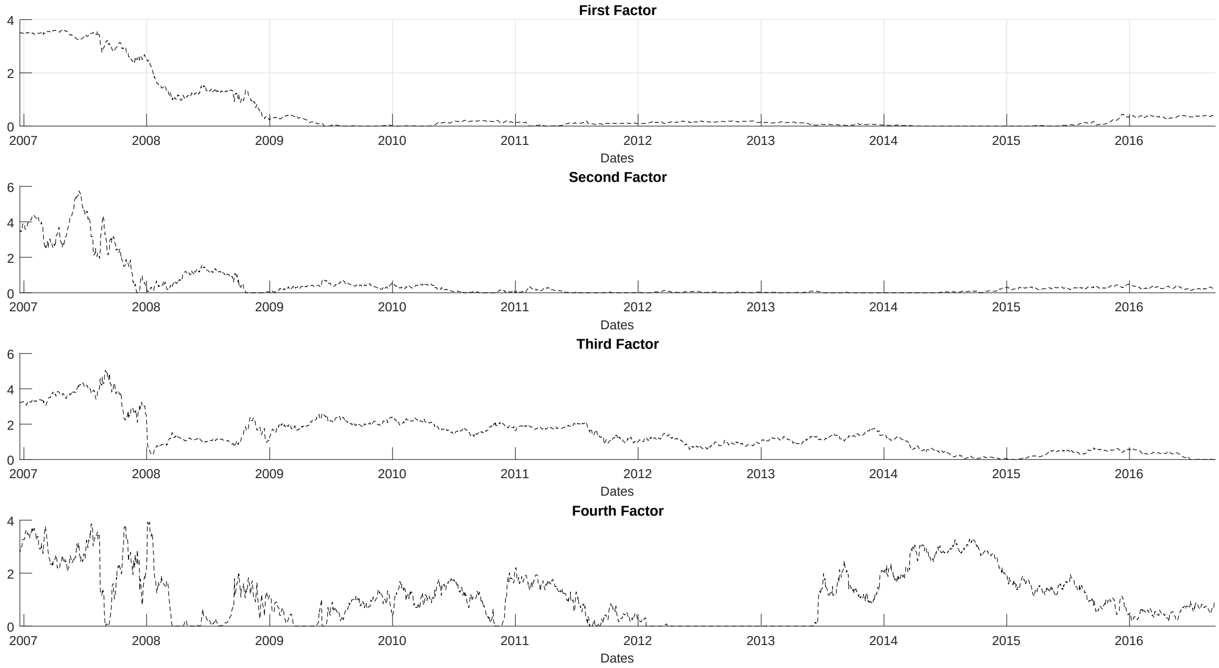


Figure B.5: Factor Trajectories US

Notes: The plotted factor trajectories are normalized with the standard deviation of the filtered factor values.

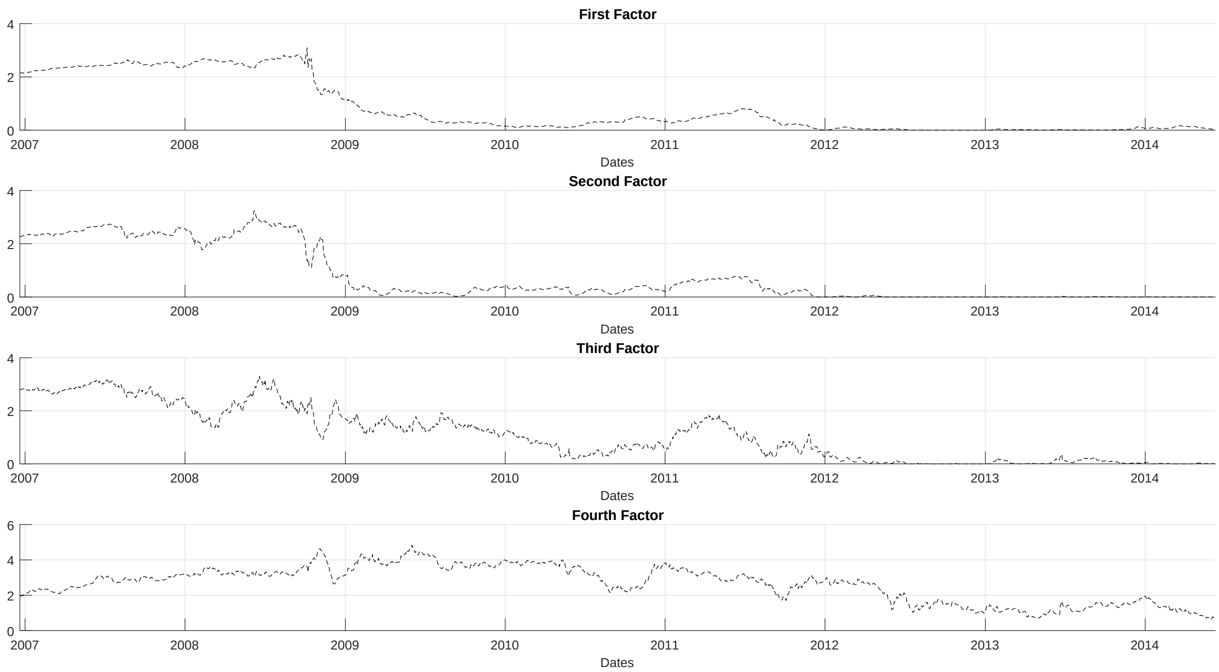


Figure B.6: Factor Trajectories EU

Notes: The plotted factor trajectories are normalized with the standard deviation of the filtered factor values.

C Table Appendix

C.1 Linear Model US

Table C.1: Parameter Estimates: Linear Model US

	Value	*Std.*		Value	*Std.*
$\alpha_1^{\mathbb{P}}$	3.54e-10	*3.19e-09*	$\alpha_1^{\mathbb{Q}}$	1.15e-15	*3.18e-09*
$\alpha_2^{\mathbb{P}}$	0.6779	*7.5653*	$\alpha_2^{\mathbb{Q}}$	0.69	*7.5648*
$\alpha_3^{\mathbb{P}}$	0.0022	*0.0207*	$\alpha_3^{\mathbb{Q}}$	0.0147	*0.0206*
$\alpha_4^{\mathbb{P}}$	5.11e-11	*3.42e-10*	$\alpha_4^{\mathbb{Q}}$	3.7e-10	*3.42e-10*
$\beta_{11}^{\mathbb{P}}$	0.9974	*0.00128*	$\beta_{11}^{\mathbb{Q}}$	0.9994	*0.0066*
$\beta_{22}^{\mathbb{P}}$	0.9999	*8.65e-04*	$\beta_{22}^{\mathbb{Q}}$	0.9996	*4.31e-04*
$\beta_{33}^{\mathbb{P}}$	0.9983	*0.0226*	$\beta_{33}^{\mathbb{Q}}$	0.9937	*0.0112*
$\beta_{44}^{\mathbb{P}}$	1-1e-05	*6.73e-07*	$\beta_{44}^{\mathbb{Q}}$	0.9999	*0.0012*
$\beta_{21}^{\mathbb{P}}$	5.43e-04	*0.0049*	$\beta_{21}^{\mathbb{Q}}$	3.35e-04	*0.0014*
$\beta_{31}^{\mathbb{P}}$	0.01	*0.1047*	$\beta_{31}^{\mathbb{Q}}$	0.0077	*0.0264*
$\beta_{41}^{\mathbb{P}}$	7.9e-12	*8.86e-11*	$\beta_{41}^{\mathbb{Q}}$	3.77-16	*2.09e-11*
$\beta_{32}^{\mathbb{P}}$	0.0522	*0.9898*	$\beta_{32}^{\mathbb{Q}}$	3.86e-10	*0.3335*
$\beta_{42}^{\mathbb{P}}$	6.58e-11	*1.02e-09*	$\beta_{42}^{\mathbb{Q}}$	6.88e-14	*4.21e-10*
$\beta_{43}^{\mathbb{P}}$	1.08e-04	*9.51e-04*	$\beta_{43}^{\mathbb{Q}}$	1.08e-04	*7.22e-04*
		Constant Parameters			
θ_1	-0.0022	*0.0067*	θ_2	-6.74e-05	*4.31e-04*
θ_3	-0.0017	*0.0113*	θ_4	1.12e-04	*0.0012*
$\mu_1^{\mathbb{P}}$	1	-	$\mu_2^{\mathbb{P}}$	1	-
$\mu_3^{\mathbb{P}}$	1	-	$\mu_4^{\mathbb{P}}$	1	-
$\mu_1^{\mathbb{Q}}$	1.0022	*0.0067*	$\mu_2^{\mathbb{Q}}$	1.0001	*4.31e-04*
$\mu_3^{\mathbb{Q}}$	1.0017	*0.0113*	$\mu_4^{\mathbb{Q}}$	0.9999	*0.0012*
$\rho_1^{\mathbb{Q}}$	0.9996	*0.0038*	$\rho_2^{\mathbb{Q}}$	1-1e-09	*2.04e-08*
$\rho_3^{\mathbb{Q}}$	1-1e-09	*4.52e-08*	$\rho_4^{\mathbb{Q}}$	0.9998	*0.0024*
δ_1	0.0013	*0.0196*	σ_R	1.7e-04	*1.63e-04*

Notes: The value of δ corresponds to a short rate in annualized percentage points and the measurement error corresponds to yields in annualized percentage points. The delta is calibrated to match an unconditional short rate of 0.747%. This value has been calculated to match the average yield curve. The parameter results correspond to a fully parametrized model with 19 estimated parameters. The reported standard errors are calculated from the bootstrapped variance with $N = 100$.

C.2 Linear Model EU

Table C.2: Parameter Estimates: Linear Model EU

	Value	*Std.*		Value	*Std.*
$\alpha_1^{\mathbb{P}}$	8.68e-06	*2.74e-05*	$\alpha_1^{\mathbb{Q}}$	8.69e-06	*2.73e-05*
$\alpha_2^{\mathbb{P}}$	8e-12	*3.05-15*	$\alpha_2^{\mathbb{Q}}$	8e-12	*2.13e-13*
$\alpha_3^{\mathbb{P}}$	3.37e-04	*0.0014*	$\alpha_3^{\mathbb{Q}}$	3.37e-04	*0.0014*
$\alpha_4^{\mathbb{P}}$	0.5913	*0.3099*	$\alpha_4^{\mathbb{Q}}$	0.5927	*0.309*
$\beta_{11}^{\mathbb{P}}$	0.9957	*0.0067*	$\beta_{11}^{\mathbb{Q}}$	0.9966	*0.0048*
$\beta_{22}^{\mathbb{P}}$	0.9996	*4.14e-04*	$\beta_{22}^{\mathbb{Q}}$	0.9994	*0.0266*
$\beta_{33}^{\mathbb{P}}$	1-1e-06	*1.65-05*	$\beta_{33}^{\mathbb{Q}}$	1-5e-06	*8.27e-06*
$\beta_{44}^{\mathbb{P}}$	0.9946	*0.0058*	$\beta_{44}^{\mathbb{Q}}$	0.9968	*0.0031*
$\beta_{21}^{\mathbb{P}}$	0.017	*0.0329*	$\beta_{21}^{\mathbb{Q}}$	0.017	*0.0263*
$\beta_{31}^{\mathbb{P}}$	6e-09	*2.01e-08*	$\beta_{31}^{\mathbb{Q}}$	6e-09	*9.27e-09*
$\beta_{41}^{\mathbb{P}}$	8.83e-10	*5.02e-11*	$\beta_{41}^{\mathbb{Q}}$	8.85e-10	*1.36e-09*
$\beta_{32}^{\mathbb{P}}$	3.44e-04	*3.44e-04*	$\beta_{32}^{\mathbb{Q}}$	3.44e-04	*4.03e-04*
$\beta_{42}^{\mathbb{P}}$	0.0022	*0.0024*	$\beta_{42}^{\mathbb{Q}}$	0.0022	*0.0026*
$\beta_{43}^{\mathbb{P}}$	2.36e-06	*3.95e-06*	$\beta_{43}^{\mathbb{Q}}$	2.37e-06	*3.57e-06*
		Constant Parameters			
θ_1	-9.55e-04	*0.0049*	θ_2	2.01e-04	*0.0266*
θ_3	-5.42e-08	*8.27e-06*	θ_4	-0.0023	*0.0031*
$\mu_1^{\mathbb{P}}$	1	-	$\mu_2^{\mathbb{P}}$	1	-
$\mu_3^{\mathbb{P}}$	1	-	$\mu_4^{\mathbb{P}}$	1	-
$\mu_1^{\mathbb{Q}}$	1.001	*0.0049*	$\mu_2^{\mathbb{Q}}$	0.9998	*0.0266*
$\mu_3^{\mathbb{Q}}$	1	*8.27e-06*	$\mu_4^{\mathbb{Q}}$	1.0023	*0.0031*
$\rho_1^{\mathbb{Q}}$	0.9976	*0.0071*	$\rho_2^{\mathbb{Q}}$	0.9992	*0.0532*
$\rho_3^{\mathbb{Q}}$	1-1e-07	*5.96e-07*	$\rho_4^{\mathbb{Q}}$	0.9991	*0.0022*
δ_1	0.0004	*0.0033*	σ_R	0.0002	*2.65e-05*

Notes: The value of δ corresponds to a short rate in annualized percentage points and the measurement error corresponds to yields in annualized percentage points. The delta is calibrated to match an unconditional short rate of 1.1736%. This value has been calculated to match the average yield curve. The reported standard errors are calculated from the bootstrapped variance with $N = 100$.