UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

Master's thesis

Allan Nouri

The Term Structure of Interest Rates:

Introducing Monetary Policy Regimes at the Lower Bound

Supervisor: Professor Anders Rahbek

ECTS points: 30

Date of Submission: September 10, 2020

Keystrokes: 138190

ABSTRACT

Historically, the European Central Bank (hereafter, the "ECB") has used short-term nominal interest rates in the euro area as an indicator of the monetary policy stance. The ECB can influence short-term nominal interest rates in the euro area as the ECB is the monopoly supplier of the euro monetary base. Cash guarantees a zero nominal interest rate which imposes a lower bound on nominal interest rates. As short-term nominal interest rates reached the lower bound, the ECB's conventional interest rate policy became ineffective. The ECB responded by using unconventional monetary policy tools. However, short-term nominal interest rates do not fully reflect the complete monetary policy stance when the ECB conducts unconventional monetary policy.

The purpose of this thesis is to produce an indicator of the ECB's monetary policy stance when unconventional monetary policy is conducted. This thesis seeks to answer the following questions: *i. How would the short rate evolve in the euro area if the lower bound did not exist? ii. How can the short rate summarize the monetary policy stance when the ECB pursues unconventional monetary policy?* This thesis proposes a novel *regime-switching term structure model* (hereafter, "RSTSM"). The RSTSM is an extension of the traditional *Gaussian affine term structure model* (hereafter, "GATSM"). The novelty of the RSTSM is that the model introduces regime shifts in the short rate process, i.e., the process for the nominal interest rate with the shortest maturity. The regimes are unobserved but classified by the two monetary policy regimes; a *no* lower bound regime and a lower bound regime.

This thesis finds that *i.* the RSTSM can guide how the short rate would have evolved in the euro area if the lower bound did not exist, and *ii.* the RSTSM can summarize the ECB's actual monetary policy stance quite well when unconventional monetary policy is conducted. Compared to the GATSM, the RSTSM offers a superior fit at the short end of the forward curve during the lower bound period. In the RSTSM, the regimes evolve in accordance with the ECB's monetary policy. However, the GATSM and RSTSM are misspecified, indicating that neither the GATSM nor the RSTSM is consistently estimated. Several extensions are proposed to deal with the poor empirical performance of the models.

Contents

List of Tables

List of Figures

List of Abbreviations

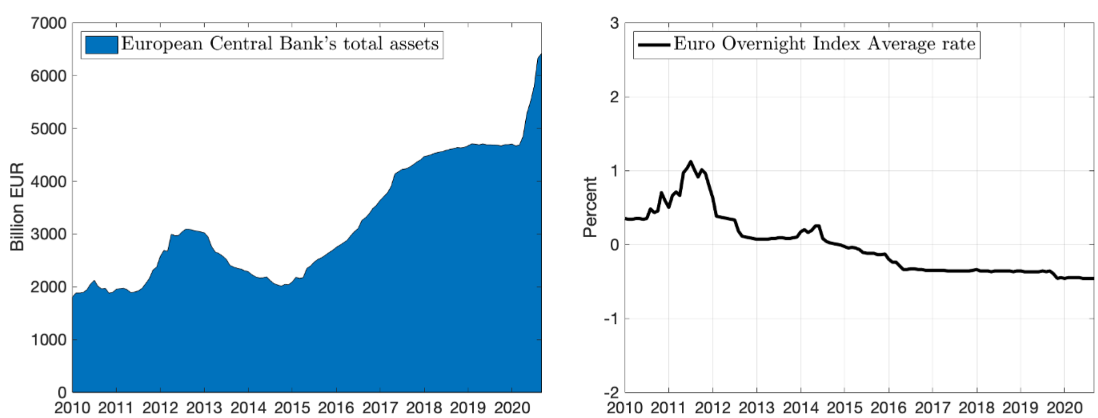
ECB	**E**uropean **C**entral **B**ank
EONIA	**E**uro **O**vernight **I**ndex **A**verage
€STR	**E**uro **S**hort-**T**erm **R**ate
GATSM	**G**aussian **A**ffine **T**erm **S**tructure **M**odel
RSTSM	**R**egime-**S**witching **T**erm **S**tructure **M**odel
LB	**L**ower **B**ound regime
NB	**N**o lower **B**ound regime

I. Introduction

A decade with historically low nominal interest rates has almost passed. Consequently, unconventional monetary policy measures have become a permanent and significant part of central banks' monetary policy toolbox. The extraordinary low nominal interest rates have forced a substantial shift in central banks' monetary policy. Cash guarantees a zero nominal interest rate which imposes a lower bound on nominal interest rates. As short-term nominal interest rates reached the lower bound, central banks' conventional interest rate policy became ineffective. Therefore, central banks have conducted unconventional monetary policy to sustain their basis of existence, i.e., to defend the course of the overall economy and to ensure price stability in particular.

Historically, short-term nominal interest rates have summarized central banks' monetary policy stance. However, merely observing short-term nominal interest rates does not fully reflect the complete monetary policy stance, as central banks have used unconventional monetary policy measures. For instance, in the euro area from 2010 to 2020, short-term nominal interest rates did not reflect that the European Central Bank conducted expansionary monetary policy through a massive expansion of its total assets, cf. figure 1.

Figure 1



Note: Own figures based on the European Central Bank Statistical Data Warehouse. The Euro Overnight Index Average rate is the average of overnight rates for unsecured interbank lending in the euro area, see ECB (2020g).

The purpose of this thesis is to produce an indicator of the European Central Bank's (hereafter, "ECB") monetary policy stance when unconventional monetary policy is conducted. This thesis seeks to answer the following questions: *i. How would the short rate evolve in the euro area if the lower bound did not exist? ii. How can the short rate summarize the monetary policy stance when the ECB pursues unconventional monetary policy?*

This thesis answers these questions by proposing a novel *regime-switching term structure model* (hereafter, "RSTSM"). The RSTSM is an extension of the traditional *Gaussian affine term structure model* (hereafter, "GATSM"), see, e.g., Singleton (2006). The RSTSM and GATSM both belong to the class of discrete-time affine term structure models. The novelty of the RSTSM is that the model introduces regime shifts in the short rate process, i.e., the process for the nominal interest rate with the shortest maturity. In the RSTSM, the regimes are unobserved but classified by the two monetary policy regimes; a *no* lower bound regime and a lower bound regime. The *no* lower bound regime is characterized by the ECB's conduct of conventional monetary policy, while the lower bound regime is characterized by the ECB's launch of unconventional monetary policy.

The paper is organized as follows: The next section describes the existing literature devoted to affine term structure models. Section II gives an introduction to central banks' monetary policy toolbox, the various monetary transmission channels and a description of the ECB's key policy rates. Section III presents the pricing methodology of zero-coupon bonds and a description of the GATSM and RSTSM. Subsequently, section IV sets the econometric methodology for the GATSM and RSTSM. Section V provides an empirical analysis of the GATSM and RSTSM based on the euro area forward rates. The empirical analysis suggests that *i.* the RSTSM can guide how the short rate would have evolved in the euro area if the lower bound did not exist, and *ii.* the RSTSM can summarize the ECB's actual monetary policy stance quite well when unconventional monetary policy is conducted. Section VI discusses the main findings of the empirical analysis and proposes several extensions to deal with the severe signs of misspecification in the GATSM and RSTSM. Finally, section VII concludes.

A. Literature

This section briefly outlines the literature related to term structure models. The term structure models considered in this thesis belong to the class of affine term structure models. Therefore, this section will focus on these models and their relation to monetary policy.

Among the most intensively studied models describing the evolution of interest rates are affine term structure models. Affine term structure models assume that interest rates are driven by an affine transformation of a set of state variables or factors. These models date back to the original single-factor affine term structure model introduced in Vasicek (1977). The single-factor affine term structure model in Vasicek (1977) incorporated mean-reverting behavior in the evolution of interest rates in accordance with the perception in monetary policy. However, the model in Vasicek (1977) assumed conditionally Gaussian distributed interest rates that did not preclude interest rates from being negative. Negative interest rates were controversial at the time. Cox (1985) extended the original model in Vasicek (1977) by a model that, among other things, prevented negative interest rates. The single-factor models in Vasicek (1977) and Cox (1985) were formulated in continuous time. The multi-factor affine term structure model was studied in continuous time in Duffie and Kan (1996). Backus, Foresi, and Telmer (1998) formulated these models in discrete time. The special case of a "no-arbitrage" Gaussian affine term structure model (hereafter, "GATSM") has been widely studied in the term structure literature due to its celebrated ability of tractable pricing, see Singleton (2006).

The GATSM assumes that the short rate is affine in a set of state variables or factors. These factors may be unobserved with Gaussian innovations, see e.g., Dai and Singleton (2000). In the GATSM, "no-arbitrage" implies cross-sectional restrictions on interest rates, which are characterized by risk aversion, see Singleton (2006). As the short rate is driven by the factors, the attitude towards risk is associated with the unpredictable variation in the factors. Therefore, the ingredients in the GATSM typically consist of a process for the short rate, factors and risk aversion. Altogether, these ingredients characterize a *pricing* *kernel* that is used for "no-arbitrage" pricing of, e.g., zero-coupon bonds. In monetary policy the GATSM is useful as the short rate typically signals the monetary policy stance. As unconventional monetary policy has been conducted, the short rate has not fully reflected the complete monetary policy stance. In this context, shadow rate term structure models have been widely adopted.

Among the most popular shadow rate term structure models are the model in Wu and Xia (2016). In Wu and Xia (2016), the short rate is assumed to be equal to the maximum of the lower bound and the shadow rate. The shadow rate is affine in a set of Gaussian factors similar to the short rate in the GATSM. Wu and Xia (2016) constructed the shadow rate using the insight in Black (1995). Intuitively, the shadow rate describes the evolution of the short rate if the lower bound did not exist. In monetary policy, the shadow rate is a useful indicator of the monetary policy stance, see Krippner (2014) and Bauer and Rudebusch (2016).

This thesis proposes a novel regime-switching term structure model (hereafter, "RSTSM"). The RSTSM extends the GATSM by introducing regime shifts in the process for the short rate. The regimes are unobserved but characterized by the conducted monetary policy. The RSTSM assumes that the short rate is piece-wise affine in a set of state variables or factors dictated by the regimes. The regimes constitute a *no* lower bound regime and a lower bound regime, where different monetary policy tools are used in each regime. Unlike the GATSM, the RSTSM classifies the short rate into monetary policy regimes.

Studies of regime-switching term structure models have increased rapidly as more sophisticated methods for estimation have been discovered, see Singleton (2006). Gray (1996), Bansal and Zhou (2002), Dai, Singleton, and Yang (2007) and Ang, Bekaert, and Wei (2008) etc. have studied term structure models with regime shifts. Typically, the regimes are unobserved and related to asymmetric movements in interest rates. For instance, Dai et al. (2007) and Ang et al. (2008) extend the GATSM by introducing regimes that are related to business cycles. The RSTSM is slightly different as the regimes are related to the conducted monetary policy.

II. Monetary policy and its transmission channels

Traditionally, monetary policy has been considered as the process by which central banks utilize their monopoly power over the monetary base to meet certain objectives, see Friedman (1968). For instance, the European Central Bank's (hereafter, "ECB") primary objective is to maintain price stability in the euro area; the ECB defines price stability as an annual increase in the Harmonised Index of Consumer Prices slightly below 2 pct., see ECB (2020c). Benefits of price stability include a nominal anchor for the economy, transparency of monetary policy and guidance for market participants' perception of price levels. Central banks set their official interest rates to influence short-term nominal interest rates via the market for reserves.[i] For instance, the ECB can influence short-term nominal interest rates in the euro area as the ECB is the monopoly supplier of the euro monetary base. Central banks can further implement alternative instruments to achieve their objectives.

Section II.A and II.B present the monetary policy instruments and explain theoretically how monetary policy eventually affects the overall economy. Section II.C describes the historical evolution of the ECB's official interest rates and the reference rate for short-term nominal interest rates in the euro area.

A. Conventional monetary policy

This section introduces the conventional monetary policy instruments that the central bank implements to accommodate its objectives. The conventional monetary policy instruments include open market operations, standing facilities and reserve requirements. Altogether, the conventional monetary policy instruments influence the monetary base and thereby affect short-term nominal interest rates. The monetary base consists of currency in circulation and reserves, i.e., counterparties' deposits at the central bank.

[i]In the market for reserves, the targeted short-term nominal interest rate reflects the market-clearing price of reserves, i.e., the equilibrium price at which the demand equals the supply of reserves, see, e.g., T. Keister (2008) for a theory of monetary policy in the reserves market.

Open market operations refer to the trades of financial assets the central bank executes in exchange for reserves. As reserves are convertible to currency, these trades will affect the monetary base and thereby influence short-term nominal interest rates. Typically, the trades include repurchase agreements and collateralized loans with counterparties satisfying certain eligibility criteria. Through open market operations, the central bank can saturate the market with liquidity and mitigate the effects of unexpected liquidity fluctuations on interest rates, e.g., by fine-tuning operations. Furthermore, the central bank utilizes open market operations to signal its monetary policy stance, see, e.g., Walsh (2003) and Romer (2012).

The standing facilities refer to the overnight deposit and marginal lending facilities offered by the central bank. For instance, counterparties can borrow at the marginal lending rate to cover short-term liquidity needs and deposit funds at the central bank against the deposit rate. The marginal lending rate is higher than the deposit rate, and together they form a channel or corridor. The central bank can use the standing facilities to influence the demand for reserves and thereby affect short-term nominal interest rates, see T. Keister (2008).

Reserve requirement refers to the minimum amount of reserves held by the counterparties and other financial institutions. The central bank sets and regulates the reserve requirement. Therefore, the central bank can freeze or unlock funds to increase or decrease the overall demand for reserves, thereby influencing short-term nominal interest rates, see, e.g., Bernanke and Gertler (1995).

The conventional monetary policy instruments influence short-term nominal interest rates directly and indirectly, which affects the course of the overall economy in the short run. The effects of conventional monetary policy on the overall economy are typically explained theoretically by the monetary transmission channels or mechanism. In practice, the monetary transmission channels are somewhat ambiguous as opposite secondary effects may offset the desired effects, see e.g., Walsh (2003) and Romer (2012).

Ireland (2005) and Walsh (2003) state that the key assumption of an efficient monetary transmission is financial frictions[ii]. Under this assumption, prices are not about to adjust immediately and proportionally to any policy-induced changes in the monetary base. Consequently, the classical dichotomy and the Fischer hypothesis do not hold.[iii] The monetary transmission channels include the interest rate channel, asset price channel and credit channel. Each of these channels will be briefly described below.

The interest rate channel was especially studied in the Keynesian school of economic thought, which has been the standard view over 50 years, see Mishkin (1996). The traditional view is that policy-induced changes in short-term nominal interest rates affect long-term nominal interest rates through the expectations hypothesis of the term structure of interest rates, see, e.g., Ireland (2005). Under the assumption of financial frictions, movements in the long-term nominal interest rate affect the long-term *real* interest rate. The long-term real interest rate is relevant for saving or investment and consumption decisions. In aggregate, these decisions will form the supply and demand in goods and labor markets, which will finally result in different levels of activity and inflation.

The asset price channel contributes primarily to the monetarists' view on the transmission mechanism, see Friedman (1968). Policy-induced changes in short-term nominal interest rates affect asset prices across a broad range of classes. Fluctuations in asset prices affect households' real balances of wealth. Households adjust consumption spending according to their real balances. For instance, households desire more consumption spending as the real balances of wealth increase, also known as the Pigou effect, see Mishkin (1996). The asset price channel is further described by Tobin's q-theory introduced in Tobin (1969). Specifically, Tobin (1969) related firms' investment decisions to Tobin's q-ratio, i.e., the ratio between a stock's market value and replacement value. Policy-induced changes in short-term nominal interest rates will affect stock prices, influencing Tobin's q-ratio for a given

[ii] In Villa (2016) financial frictions in the banking sector are related to moral hazard and adverse selection.

[iii] Classical dichotomy implies that nominal variables, e.g., nominal interest rates, do not influence real variables, e.g., output and employment, see Patinkin (1989). Fischer hypothesis implies that the expected inflation rate adjusts proportionally to any change in the nominal interest rate, leaving the real interest rate unaffected, see Fisher (1930)

replacement value. In the aggregate across all firms, aggregate investments, and therefore, the overall activity will be affected. Policy-induced changes in short-term nominal interest rates may also directly affect the inflation rate through the foreign exchange market; cross-border capital flows influence the exchange and inflation rate. The exchange rate further influence the overall activity through net exports, see Ireland (2005).

The credit channel is typically described by the bank lending and balance sheet channel, see Walsh (2003). Policy-induced changes in short-term nominal interest rates will affect banks' interest rate margin. The interest rate margin will affect investment decisions. The balance sheet channel relates firms' financing conditions to monetary policy actions. For instance, changes in nominal interest rates will affect listed firms' market value, future cash flows and collaterals. Altogether, this will translate into tighter or looser financing conditions. In aggregate, investments will be affected, which will further influence activity and inflation.

Conventional monetary policy works quite effectively through the various monetary transmission channels, see e.g., Walsh (2003), and Romer (2012). However, as short-term nominal interest rates are stuck at the lower bound, the economy is said to be caught in a *liquidity trap*, rendering the conventional monetary policy instruments ineffective.

Liquidity trap

A liquidity trap is an economic situation characterized by low short-term nominal interest rates and high saving rates. In such situations, cash or physical currency provides a lower bound for the nominal interest rate. As cash pays a nominal interest of zero, a further cut in short-term nominal interest rates may trigger economic agents to hoard cash. The occurrence of the lower bound is significantly related to the low levels of the *neutral real interest rate* or $\mathbf{r}^{\star}$. Wicksell (1907) introduced the neutral real interest rate defined as the short-term real interest rate consistent with maximum activity, full employment and stable prices. The discussion section in VI covers a further elaboration on the neutral real interest rate within monetary policy.

In a liquidity trap with low activity and deep deflation concerns, expansionary conventional monetary policy is questionable, see Romer (2012). Therefore, the central bank extends its monetary policy toolbox with unconventional monetary policy tools, see Walsh (2003).

B. Unconventional monetary policy

This section introduces the unconventional monetary policy actions that the central bank utilizes on top of its conventional monetary policy. The unconventional monetary policy tools include forward guidance and quantitative easing. These tools affect long-term nominal interest rates and thereby they influence the course of the overall economy. The transmission of unconventional monetary policy is slightly different from the transmission of conventional monetary policy in section II.A.

Forward guidance is an unconventional monetary policy instrument that aims to influence market expectations of future nominal interest rates. Through forward guidance, the central bank communicates the future level of short-term nominal interest rates. This publicly available information influences market expectations of future long-term nominal interest rates through the expectations hypothesis. The expectations hypothesis states that current and expected future short-term nominal interest rates pin down long-term nominal interest rates, see Walsh (2003). Forward guidance requires that the central bank is highly credible and equipped with tools that significantly affect market expectations, see T. Keister (2008).

Quantitative easing is an unconventional monetary policy action that drops high volumes of liquidity in the financial markets. Quantitative easing suggests that the central bank purchases financial assets with long maturity on a large-scale basis. Consequently, the demand and the prices of these particular assets increase, see, e.g., Bauer and Rudebusch (2014). As asset prices and nominal interest rates are negatively related, quantitative easing will effectively reduce long-term nominal interest rates and inject liquidity into the financial system, see Gagnon et. al. (2011). Typically, the financial assets include government bonds, mortgage loans and other relatively safe financial assets. The central bank finances its large-scale acquisitions through the issuance of new reserves, see Joyce et. al. (2011). Therefore, financial institutions can liquidate assets quickly by selling them to the central bank in exchange for reserves. Such trades will directly increase the financial institutions' liquidity on their balance sheet. Through the transmission channels, the trades will indirectly ease financial conditions and enhance economic activity and inflation.

The transmission channels through which unconventional monetary policy operates are theoretically explained by the signaling channel, portfolio re-balancing channel and liquidity premia channel. Altogether, these channels continue in the asset price channel which affects the overall activity and inflation. The asset price channel was introduced in section II.A.

The signaling channel describes the effect of unconventional monetary policy announcements on market expectations, see Bauer and Rudebusch (2014). Signaling takes place between the announcements and until the actions are carried out. As the announcements signal the future levels of short-term nominal interest rates, the market expectations of long-term nominal interest rates adjust through the expectations hypothesis. Consequently, asset prices adjust appropriately. The signal is somehow credible or at least committed by the central bank due to its high exposure to interest rate risk, e.g., from quantitative easing actions.

The portfolio re-balancing channel is primarily related to quantitative easing. As the central bank withdraws a large amount of safe assets from the market, portfolios may re-balance into other assets. For instance, if the central bank demands a large-scale amount of safe assets in its quantitative easing program, prices of safe assets and close substitutes will increase, see Joyce et. al. (2011). Ceteris paribus, portfolios consisting of the involving assets will deliver lower returns for the same risk level. Re-balancing the portfolio into a larger amount of risky assets to compensate for the lower return of safe assets induces risk-taking behavior. This search-for-yield will result in higher asset prices of risky assets.

The liquidity premia channel describes the effects of unconventional monetary policy on asset prices. During quantitative easing, the central bank's significant demand for financial assets improves market functioning, especially when market liquidity dries up. An improved market functioning may reduce the liquidity premia and increase asset prices as the relevant assets are less costly to liquidate in otherwise illiquid times, see, e.g., Gagnon et. al. (2011). This liquidity premia channel or market functioning channel is mostly present at the early stages of quantitative easing, where the central bank's demand for certain assets is activated, see Joyce et. al. (2011).

The signaling channel, portfolio re-balancing channel and liquidity premia channel continue in the asset price channel introduced in section II.A. Ultimately, the unconventional monetary policy instruments influence the overall activity and inflation.

C. *Official interest rates and the proxy for the short rate*

This section introduces the European Central Bank's (hereafter, "ECB") official interest rates and its reference rate for the euro area. The ECB's current reference rate is the Euro Overnight Index Average (hereafter, "EONIA"), see ECB (2020g).[iv] The reference rate is also the suggested proxy for the short rate in the euro area, see ECB (2020g).

The ECB operates with three official interest rates; the main refinancing operations rate, the marginal lending rate and the deposit rate, see ECB (2020c). On a weak basis, counterparties can borrow at the main refinancing operations rate to cover liquidity needs. On an overnight basis, counterparties can deposit at the ECB's deposit rate or borrow at the ECB's marginal lending rate. These three interest rates constitute the ECB's official interest rates set every six weeks as a part of its monetary policy, see ECB (2020e).

The ECB indirectly uses the EONIA rate as an operational target to signal its monetary policy stance. The EONIA rate is viewed as the first step of the transmission mechanism, see Bindseil (2014). Therefore, the ECB seeks to keep the EONIA rate in close distance to the official interest rates. Note, the EONIA rate is not a direct instrument of the ECB. Instead, it is the average of overnight rates for unsecured interbank lending in the euro area.

[iv] In September 2018, the ECB announced a new reference rate for euro area, see ECB (2020g). The new reference rate will be addressed and discussed in section VI.B.

Figure 2 on page 18 presents the evolution of the ECB's official interest rates and the EONIA rate from January 2005 to March 2020. Few characteristics can be highlighted from this figure. First, all interest rates take a sharp drop in the global financial crisis in 2007-08. In the wake of this crisis, all interest rates exhibit a downward trend, except for the sovereign debt crisis reaction in June 2011. The ECB reacted to high inflation in the euro area partly driven by tensions in the sovereign debt market and political turmoil in the oil markets, see also ECB (2020d). In figure 2, it is clear that the marginal lending rate is a cap of the interest rates, while the deposit rate sets a floor. Together, they constitute a corridor through which the main refinancing operation rate and the EONIA rate evolve.

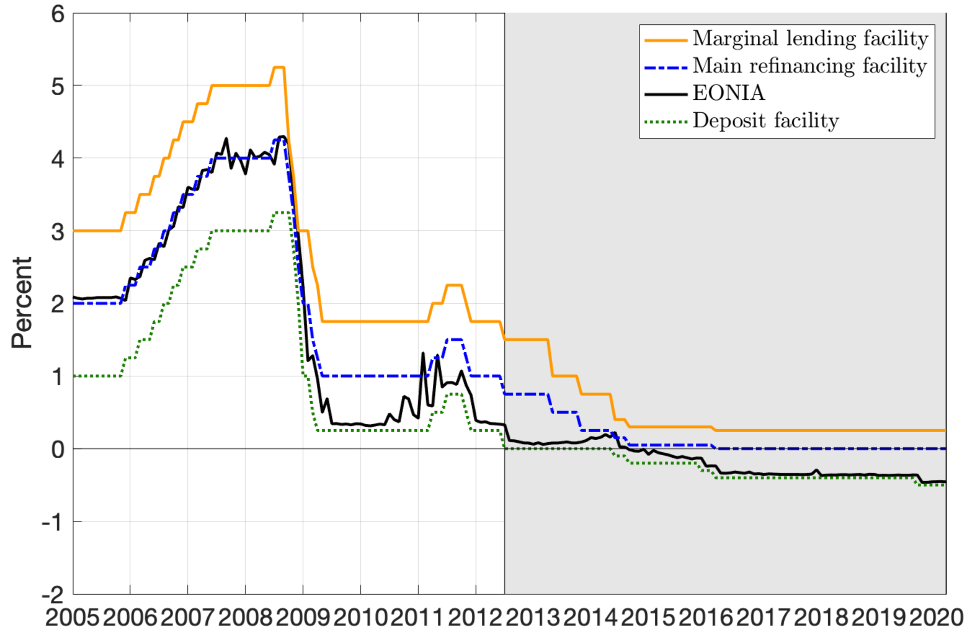
The shaded area in figure 2 corresponds to the lower bound period from July 2012. The lower bound on interest rates is assumed to be the minimum of the ECB's deposit rate and zero, which is assumed in the absence of storage costs, see Kortela (2016). In practice, the lower bound assumption is still open for discussion, see Rogoff (2017).[v]

Figure 2 shows that the EONIA rate has evolved in close distance to the ECB's official interest rates during the period from January 2005 to March 2020. In the lower bound period, i.e., the shaded area in figure 2, the EONIA rate has evolved at persistently low levels, not necessarily reflecting the ECB's actual monetary policy stance.

In figure 2, the ECB's conduct of monetary policy may be roughly divided into two monetary policy regimes; a no lower bound regime and a lower bound regime. The ECB implements its conventional monetary policy tools in the no lower bound regime, while the ECB further launches unconventional monetary tools in the lower bound regime. Clearly, figure 2 shows that the EONIA rate does not serve as an indicator of the monetary policy stance in the lower bound regime when the ECB conducts unconventional monetary policy.

[v] Section VI.A provides a discussion of the lower bound on interest rates.

Figure 2. The ECB's official interest rates and the EONIA rate



Note: European Central Bank's (ECB's) official interest rates and the Euro Overnight Index Average (EONIA) rate. The shaded area corresponds to the lower bound period from July 2012.

III. Term structure modeling

This section introduces the methodology for pricing zero-coupon bonds that is used to construct the term structure models. This section presents the traditional *Gaussian affine term structure model* (hereafter, "GATSM") and the *regime-switching term structure model* (hereafter, "RSTSM"). The GATSM and RSTSM belong to the class of affine term structure models which assume that the term structure of interest rates is driven by an affine transformation of a set of state variables or factors. The pricing methodology rules out arbitrage opportunities in the GATSM and RSTSM, resulting in cross-sectional restrictions on interest rates, i.e., across maturities at each point in time.

A. *Bond pricing*

This section starts with a quite general framework for pricing any risky asset based on Hansen and Richard (1987) and Singleton (2006) followed by an application to zero-coupon bonds. Subsequently, the yield and the forward rate of a zero-coupon bond are defined which will be used to construct the term structure models in the following sections.

The objective is to price any risky sequence of cash flows at time t for $t = 1, ..., T$. Let m_{t+1} denote the value or the payoff of a risky asset at time $t+1$ to be priced at time t. The payoff m_{t+1} for this particular asset is assumed to be in the information set of economic agents $\mathcal{I}_{t+1}$. The payoff space is given by $\mathbf{M}_{t+1} = \{m_{t+1} \in \mathcal{I}_{t+1} : \mathbb{E}^{\mathbb{P}}(m_{t+1}^2|\mathcal{I}_t) < \infty\}$[vi], where $\mathbb{P} \in [0,1]$ denotes the physical or data-generating probability measure with $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\mathbf{M}) = 1$.[vii] The price at time t of the asset delivering the payoff $m_{t+1} \in \mathbf{M}_{t+1}$ is given by the pricing function $\mathcal{P}_t : m_{t+1} \to \mathcal{I}_t$[viii].

Definition III.1. *Following Definition 2.4 in Hansen and Richard (1987), the pricing function $\mathcal{P}_t$ has no arbitrage opportunities on $\mathbf{M}$ if for any payoff $m_{t+1} \in \mathbf{M}_{t+1}$ for which $\mathbb{P}(m_{t+1} \geq 0) = 1$, $\mathbb{P}(\{\mathcal{P}_t \leq 0 \ \cap \ m_{t+1} > 0\}) = 0$.*

[vi] Theorem A.1 in Hansen and Richard (1987) show that the payoff space $\mathbf{M}_{t+1}$ is conditionally complete.
[vii] Following Singleton (2006) and Hansen and Richard (1987); $\mathbf{M} = \{\mathbf{M}_t\}_{t=1}^T$.
[viii] As in Assumption 2.2-2.4 Hansen and Richard (1987), it is assumed that the pricing function satisfies: i) value additivity, ii) conditional continuity at zero payoffs and iii) $\exists \ m_{t+1}^\star \in \mathbf{M}_{t+1} : \mathbb{P}(\mathcal{P}_t(m_{t+1}^\star) = 0) = 0$.

Definition III.1 implies that the asset must be traded at a positive price at time t, if the same asset delivers a non-negative payoff with positive probability at time $t+1$. Under certain regularity conditions, Hansen and Richard (1987) show that definition III.1 is equivalent to the existence of a special payoff or a *pricing kernel* denoted by $\mathcal{M}_{t+1} \in \mathbf{M}_{t+1}$ that is strictly positive $\mathbb{P}(\mathcal{M}_{t+1} > 0) = 1$ and satisfies:

$$\mathcal{P}_t = \mathbb{E}^{\mathbb{P}}(\mathcal{M}_{t+1} m_{t+1} | \mathcal{I}_t), \quad m_{t+1} \in \mathbf{M}_{t+1} \tag{1}$$

see Lemma 2.3 and the associated regularity conditions in Hansen and Richard (1987). In this context, the pricing kernel $\mathcal{M}_{t+1}$ is given in "reduced-form" in the sense that there is no direct or "structural" link to economic agents' preferences. If $R_{t+1} \equiv m_{t+1}/\mathcal{P}_t \in \mathbf{M}_{t+1}$ denotes the gross return on the asset, equation (1) implies that returns are those payoffs with unit price:

$$1 = \mathbb{E}^{\mathbb{P}}(\mathcal{M}_{t+1} R_{t+1} | \mathcal{I}_t), \quad R_{t+1} \in \mathbf{M}_{t+1} \tag{2}$$

Equation (2) is a version of the well-known fundamental theorem of asset pricing, see Cochrane (2001) and Singleton (2006). Equation (2) applies to any asset including zero-coupon bonds. As the term structure models will be based on prices of zero-coupon bonds with various maturities, it is convenient to introduce the concept of a zero-coupon bond and especially how the pricing equation (2) applies.

A zero-coupon bond is a fixed-income instrument that does not make periodic interest payments or so-called coupons. Instead, a zero-coupon bond is redeemed for its principal amount or face value at maturity. Therefore, a zero-coupon bond generates a deterministic cash flow in the absence of trading costs, taxes, default risk, etc. Let $\mathcal{P}_t^n$ denote the price of a n-period zero-coupon bond that contains a unit payoff for $n = 1, ..., p$. The gross return is given by $R_{t+1}^n \equiv \mathcal{P}_{t+1}^{n-1}/\mathcal{P}_t^n$. Here, it is used that a n-period zero-coupon bond at time t matures in

$n-1$ at time $t+1$. Equation (2) implies that the no-arbitrage price at time t is given by:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1}\mathcal{P}_{t+1}^{n-1}) \tag{3}$$

where $\mathbb{E}_t^{\mathbb{P}}(\,\cdot\,) = \mathbb{E}^{\mathbb{P}}(\,\cdot\,|\mathcal{I}_t)$. Intuitively, the price of a n-period zero-coupon bond at time t is given by the expected discounted price at time $t+1$ conditional on the information available at time t. A special case is the risk-free zero-coupon bond, see the definition below.

Definition III.2. *The one-period continuously compounded yield or risk-free rate r_t solves $1 = \mathcal{P}_t^1 \cdot \exp\{r_t\}$, or equivalently, it is given by: $r_t = -\log\mathcal{P}_t^1$.*

Definition III.2 implies that the one-period gross return is given by $R_t^1 = \exp\{r_t\}$. Therefore, the price of the one-period zero-coupon bond at time t is given by:

$$\mathcal{P}_t^1 = \exp\{-r_t\} = \mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})$$

The pricing kernel and the risk-free rate are fundamentally related and together they distinguish the physical and risk-neutral measures. Typically, bond pricing proceeds under the risk-neutral measure, i.e., pricing *as if* economic agents are neutral towards risk. Under the assumption of no arbitrage opportunities, there exists an equivalent martingale measure or risk-neutral measure $\mathbb{Q} \sim \mathbb{P}$, where $\mathbb{Q}$ is equivalent to $\mathbb{P}$ if they assign zero probabilities to the same events, see Singleton (2006). A so-called *change of measure* allows bond pricing under the physical measure in equation (3) to proceed under the risk-neutral measure. The change of measure involves certain steps which will be explicitly derived in the following.

Let $f_t^{\mathbb{P}}$ and $f_t^{\mathbb{Q}}$ denote the probability density function under the physical $\mathbb{P}$ and risk-neutral $\mathbb{Q}$ measures conditional on the information available at time t. Rewrite the price in equation (3) and use that $\exp\{-r_t\}/\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1}) = 1$ to obtain:

$$\begin{aligned}
\mathcal{P}_t^n &= \mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1}\mathcal{P}_{t+1}^{n-1}) \\
\mathcal{P}_t^n &= \int \mathcal{M}_{t+1}\mathcal{P}_{t+1}^{n-1} f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) d(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) \\
\mathcal{P}_t^n &= \int \frac{\exp\{-r_t\}}{\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})} \mathcal{M}_{t+1}\mathcal{P}_{t+1}^{n-1} f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) d(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) \\
\mathcal{P}_t^n &= \int \exp\{-r_t\}\mathcal{P}_{t+1}^{n-1} \underbrace{\frac{\mathcal{M}_{t+1}}{\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})} f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})}_{\equiv f_t^{\mathbb{Q}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})} d(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) \\
\mathcal{P}_t^n &= \int \exp\{-r_t\}\mathcal{P}_{t+1}^{n-1} f_t^{\mathbb{Q}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) d(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) \\
\mathcal{P}_t^n &= \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\mathcal{P}_{t+1}^{n-1})
\end{aligned} \tag{4}$$

where $f_t^{\mathbb{Q}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1}) \equiv \frac{\mathcal{M}_{t+1}}{\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})} f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})$ is a proper density function as the term $\mathcal{M}_{t+1}/\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})$ integrates to one. In equation (4), $\mathbb{E}_t^{\mathbb{Q}}(\cdot) = \mathbb{E}^{\mathbb{Q}}(\cdot|\mathcal{I}_t)$ denotes the conditional expectation at time t under the risk-neutral measure $\mathbb{Q}$. These steps show that bond pricing can proceed under the risk-neutral measure by adjusting $f_t^{\mathbb{P}}$ by $\mathcal{M}_{t+1}/\mathbb{E}_t^{\mathbb{P}}(\mathcal{M}_{t+1})$. Bond pricing under the risk-neutral measure is quite attractive as the price in equation (4) is typically easier to calculate compared to the price in equation (3). Knowledge of the risk-neutral conditional probability density function $f_t^{\mathbb{Q}}$ alone may be sufficient for pricing but insufficient for estimation as estimation is based on bond prices that are observed under the physical or data-generating probability measure. For both pricing and estimation, it is convenient to derive the relation between these measures explicitly. In this context, the pricing kernel can be seen as a transformation of the physical and risk-neutral measures:

$$\mathcal{M}_{t+1} = \exp\{-r_t\}\frac{f_t^{\mathbb{Q}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})}{f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})} = \exp\{-r_t\}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1} \tag{5}$$

where the term $(d\mathbb{Q}/d\mathbb{P})_{t+1} = f_t^{\mathbb{Q}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})/f_t^{\mathbb{P}}(\mathcal{M}_{t+1}, \mathcal{P}_{t+1}^{n-1})$ denotes the Radon-Nikodym derivative of the risk-neutral measure $\mathbb{Q}$ with respect to the physical measure $\mathbb{P}$. The Radon-

Nikodym derivative essentially determines the transformation between $f_t^{\mathbb{Q}}$ and $f_t^{\mathbb{P}}$. In the term structure models that will be introduced in the sections III.B-III.C, the models assume a specific process for the short rate r_t and the Radon-Nikodym derivative $(d\mathbb{Q}/d\mathbb{P})_{t+1}$, which implicitly assume a pricing kernel $\mathcal{M}_{t+1}$ process through equation (5).

In equation (4), the price of the n-period zero-coupon bond is recursive in the sense that the equation can be rewritten to:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\mathbb{E}_{t+1}^{\mathbb{Q}}(\exp\{-r_{t+1}\}\mathcal{P}_{t+2}^{n-2}))$$

As the information set expands gradually over time, i.e., $\mathcal{I}_t \subset \mathcal{I}_{t+1}$, law of iterated expectation implies that $\mathbb{E}_t^{\mathbb{Q}}(\mathbb{E}_{t+1}^{\mathbb{Q}}(\,\cdot\,)) = \mathbb{E}_t^{\mathbb{Q}}(\,\cdot\,)$. Therefore, the equation is given by:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\exp\{-r_{t+1}\}\mathcal{P}_{t+2}^{n-2})$$

These recursions continue until the date of maturity $t+n$ where the bond is redeemed for its face value of 1. By recursive substitution, the price at time t is given by:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\exp\{-r_{t+1}\}\cdots\exp\{-r_{t+n-1}\}) = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-\sum_{i=0}^{n-1} r_{t+i}\}) \qquad (6)$$

Equation (6) is a version of the *expectations hypothesis*, i.e., the price of longer term zero-coupon bonds at time t is given by the current and expected future short rates. The expectations hypothesis holds only under the risk-neutral measure. Fama (1984) and Cochrane (2001) explain the failure of the expectations hypothesis under the physical measure for various reasons including risk premia.

The price of the zero-coupon bond in equation (6) can be used to construct the *yield* or *forward rate* of the zero-coupon bond. These concepts will be defined below.

Definition III.3. *The n-period continuously compounded yield solves* $\mathcal{P}_t^n \exp\{n\mathcal{Y}_{t,n}\} = 1$. *The yield is the interest rate that justifies the price of a zero-coupon bond and it can be seen as the internal rate of return on the zero-coupon bond. Given the price* $\mathcal{P}_t^n$*, the n-period continuously compounded yield is given by:*

$$\mathcal{Y}_t^n = -\frac{1}{n}\log\mathcal{P}_t^n \tag{7}$$

Note, the one-period yield is equal to the short rate or risk-free rate: $\mathcal{Y}_t^1 = -\log\mathcal{P}_t^1 = r_t$.

Definition III.4. *The one-period continuously compounded forward rate of the n-period zero-coupon bond is denoted by* $\mathcal{F}_t^{n\to n+1}$*, where the arrow indicates that the zero-coupon bond contracted at time t lives from n to n+1. Under the assumption of no arbitrage opportunities, Cochrane (2001) and Hull (2014) show that the one-period forward rate satisfies:*

$$\exp\{(n+1)\mathcal{Y}_t^{n+1}\} = \exp\{n\mathcal{Y}_t^n\}\exp\{\mathcal{F}_t^{n\to n+1}\}$$

In the absence of arbitrage opportunities, it should not be possible to synthesize a $(n+1)$*-period zero-coupon bond by combining n-period and one-period zero-coupon bonds purchased and contracted at time t. By taking the logarithm of this equation, the one-period continuously compounded forward rate is given by:*

$$\mathcal{F}_t^{n\to n+1} = (n+1)\mathcal{Y}_t^{n+1} - n\mathcal{Y}_t^n \tag{8}$$

For the ease of notation, the one-period forward rate will be denoted by $\mathcal{F}_t^n \equiv \mathcal{F}_t^{n\to n+1}$.

These concepts are used in the term structure models introduced in the sections III.B-III.C.

B. Gaussian Affine Term Structure Model (GATSM)

This section introduces the discrete-time version of the *Gaussian affine term structure model* (hereafter, "GATSM"). The GATSM belongs to the class of affine term structure models, where it is assumed that the short rate is affine in a set of state variables or factors. The GATSM was originally formulated in continuous time, see, e.g., Vasicek (1977), Cox (1985) and Duffie and Kan (1996). This section presents the discrete-time version of the GATSM, see, e.g., Backus et al. (1998), Ang and Piazzesi (2003) and Le, Singleton, and Dai (2010). The pricing tractability of the GATSM has contributed to its well-known reputation in the term structure literature, see e.g., Singleton (2006) and Wu and Xia (2016). The GATSM will be derived under the assumption of no arbitrage opportunities similar to Duffie and Kan (1996), which essentially relies on the insights covered in section III.A.

This section begins with a complete description of the GATSM.

Definition III.5. GATSM *Under the physical measure $\mathbb{P}$, the Gaussian affine term structure model (GATSM) is defined by the following assumptions for time $t = 1, ..., T$ and maturities $n = 1, ..., p$:*

1. *The short rate, r_t, is affine in the k state variables, $\mathbf{X}_t$, i.e.,*

$$r_t = \delta_0 + \boldsymbol{\delta}_1' \mathbf{X}_t, \tag{9}$$

   where δ_0 is a scalar and $\boldsymbol{\delta}_1$ is a $k \times 1$ vector of parameters.

2. *The state variables are unobserved or latent and follow a first-order vector autoregression with Gaussian innovations:*

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}} \mathbf{X}_{t-1} + \boldsymbol{\Sigma} \boldsymbol{\epsilon}_t^{\mathbb{P}}, \quad \boldsymbol{\epsilon}_t^{\mathbb{P}} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k) \tag{10}$$

   where $\boldsymbol{\mu}^{\mathbb{P}}$ is $k \times 1$, $\boldsymbol{\Phi}^{\mathbb{P}}$ is $k \times k$, $\boldsymbol{\Sigma}$ is $k \times k$ and $\mathbf{X}_0$ is given.

3. The assumption of no-arbitrage opportunities implies the existence of a strictly positive pricing kernel. The pricing kernel follows:

$$\mathcal{M}_{t+1} = \exp\{-r_t\}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}}\Big)_{t+1} \tag{11}$$

where $(d\mathbb{Q}/d\mathbb{P})_{t+1}$ *denotes the Radon-Nikodym derivative of the risk-neutral measure* $\mathbb{Q}$ *with respect to the physical measure* $\mathbb{P}$*. Here, it is assumed that the Radon-Nikodym derivative is strictly positive and follows:*

$$\Big(\frac{d\mathbb{Q}}{d\mathbb{P}}\Big)_{t+1} = \exp\{-\frac{1}{2}\mathbf{\Lambda}_t'\mathbf{\Lambda}_t - \mathbf{\Lambda}_t'\boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}\}, \tag{12}$$

where $\mathbf{\Lambda}_t$ *is a* $k \times 1$ *vector of market prices of risk associated with the Gaussian innovations in the state variables* $\boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}$*, and it follows:*

$$\mathbf{\Lambda}_t = \lambda_0 + \boldsymbol{\lambda}_1'\mathbf{X}_t, \tag{13}$$

where λ_0 *is* $k \times 1$ *and* $\boldsymbol{\lambda}_1'$ *is* $k \times k$*.*

1., 2. and 3. implicitly define the pricing kernel. The term structure of interest rates is determined by 1. the affine relation between the short rate and the unobserved state variables, 2. the evolution of the state variables and 3. the attitude towards risk associated with Gaussian innovations in the state variables.

The GATSM presented in definition III.5 is the *multi-factor* version of the GATSM with k factors or state variables. As in Ang et al. (2008) and Wu and Xia (2016), the *three-factor* GATSM with $k = 3$ factors will be implemented in the empirical analysis in section V.

The market prices of risk associated with the shocks in the state variables distinguish the physical measure from the risk-neutral measure, see Duffee (2002). In the GATSM, there is no direct link to agents' preferences, but indirectly through the assumed process for the pricing kernel. The process in equation (12) ensures that the term structure of the interest rates is affine under both measures, see e.g., Rahbek and Nielsen (2011). The pricing kernel can also be economically justified in a representative agent model, see, e.g., Backus et al. (1998) and Singleton (2006).

In definition III.5, the unobserved state variables or factors are assumed to follow a first-order vector autoregression with *i.i.d.* Gaussian errors. This implies that the state variables are conditionally Gaussian distributed, i.e., $\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathcal{N}(\boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_t, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$. The corresponding conditional probability density function is therefore given by:

$$
\begin{aligned}
f^{\mathbb{P}}(\mathbf{X}_{t+1}|\mathbf{X}_t) &= |2\pi\boldsymbol{\Sigma}\boldsymbol{\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{X}_{t+1} - \boldsymbol{\mu}^{\mathbb{P}} - \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_t)'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\mathbf{X}_{t+1} - \boldsymbol{\mu}^{\mathbb{P}} - \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_t)\} \\
&= (2\pi)^{-\frac{k}{2}}|\boldsymbol{\Sigma}\boldsymbol{\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\epsilon}^{\mathbb{P}\prime}_{t+1}\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}\} \qquad (14)
\end{aligned}
$$

Note, $\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1} = \boldsymbol{\Sigma}^{-1}(\mathbf{X}_{t+1} - \boldsymbol{\mu}^{\mathbb{P}} - \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_t)$ by definition, see definition III.5. As the state variables are defined under the physical measure, the probability density function $f^{\mathbb{P}}$ is provided under the physical measure. By using the Radon-Nikodym derivative in definition III.5, the probability density function $f^{\mathbb{P}}$ can be transformed to the probability density function under the risk-neutral measure. As bond pricing proceeds under the risk-neutral measure, knowledge of the risk-neutral process for the state variables is required[ix]. This result in a change of measure, which will be presented below.

As it is assumed in definition III.5 that the term structure is driven by the state variables, the Radon–Nikodym derivative in section III.A is $(d\mathbb{Q}/d\mathbb{P})_{t+1} = f^{\mathbb{Q}}(\mathbf{X}_{t+1}|\mathbf{X}_t)/f^{\mathbb{P}}(\mathbf{X}_{t+1}|\mathbf{X}_t)$. Equation (12) and (14) imply that every term in this expression is known except for the risk-neutral density function $f^{\mathbb{Q}}(\mathbf{X}_{t+1}|\mathbf{X}_t)$. Therefore, the risk-neutral density function can be derived as follows:

[ix] Bond pricing could also proceed under the physical measure. However, this involves heavier notation, see e.g., appendix A in Ang and Piazzesi (2003)

$$
\begin{aligned}
f^{\mathbb{Q}}(\mathbf{X}_{t+1}|\mathbf{X}_t) &= f^{\mathbb{P}}(\mathbf{X}_{t+1}|\mathbf{X}_t)\Big(\frac{d\mathbb{Q}}{d\mathbb{P}}\Big)_{t+1} \\
&= \underbrace{(2\pi)^{-\frac{k}{2}}|\mathbf{\Sigma\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\epsilon}^{\mathbb{P}\prime}_{t+1}\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}\}}_{=f^{\mathbb{P}}(\mathbf{X}_{t+1}|\mathbf{X}_t)}\underbrace{\exp\{-\frac{1}{2}\mathbf{\Lambda}_t'\mathbf{\Lambda}_t-\mathbf{\Lambda}_t'\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}\}}_{=(\frac{d\mathbb{Q}}{d\mathbb{P}})_{t+1}} \\
&= (2\pi)^{-\frac{k}{2}}|\mathbf{\Sigma\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\epsilon}^{\mathbb{P}\prime}_{t+1}\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}+\mathbf{\Lambda}_t'\mathbf{\Lambda}+2\mathbf{\Lambda}_t'\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1})\} \\
&= (2\pi)^{-\frac{k}{2}}|\mathbf{\Sigma\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}+\mathbf{\Lambda}_t)'(\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}+\mathbf{\Lambda}_t)\}
\end{aligned}
$$

Define $\boldsymbol{\epsilon}^{\mathbb{Q}}_{t+1}=\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}+\mathbf{\Lambda}_t$ and insert to obtain:

$$
f^{\mathbb{Q}}(\mathbf{X}_{t+1}|\mathbf{X}_t) = (2\pi)^{-\frac{k}{2}}|\mathbf{\Sigma\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\epsilon}^{\mathbb{Q}\prime}_{t+1}\boldsymbol{\epsilon}^{\mathbb{Q}}_{t+1}\}
$$

In definition III.5, $\boldsymbol{\epsilon}^{\mathbb{P}}_{t+1}=\mathbf{\Sigma}^{-1}(\mathbf{X}_{t+1}-\boldsymbol{\mu}^{\mathbb{P}}-\mathbf{\Phi}^{\mathbb{P}}\mathbf{X}_t)$ and $\mathbf{\Lambda}_t=\lambda_0+\boldsymbol{\lambda}_1'\mathbf{X}_t$, implying that $\boldsymbol{\epsilon}^{\mathbb{Q}}_{t+1}=\mathbf{\Sigma}^{-1}\big(\mathbf{X}_{t+1}-(\boldsymbol{\mu}^{\mathbb{P}}-\mathbf{\Sigma}\lambda_0)-(\mathbf{\Phi}^{\mathbb{P}}-\mathbf{\Sigma}\boldsymbol{\lambda}_1')\mathbf{X}_t\big)$. Define $\boldsymbol{\mu}^{\mathbb{Q}}=\boldsymbol{\mu}^{\mathbb{P}}-\mathbf{\Sigma}\lambda_0$ and $\mathbf{\Phi}^{\mathbb{Q}}=\mathbf{\Phi}^{\mathbb{P}}-\mathbf{\Sigma}\boldsymbol{\lambda}_1'$. Insert the expressions to obtain:

$$
f^{\mathbb{Q}}(\mathbf{X}_{t+1}|\mathbf{X}_t) = (2\pi)^{-\frac{k}{2}}|\mathbf{\Sigma\Sigma}'|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{X}_{t+1}-\boldsymbol{\mu}^{\mathbb{Q}}-\mathbf{\Phi}^{\mathbb{Q}}\mathbf{X}_t)'(\mathbf{\Sigma\Sigma}')^{-1}(\mathbf{X}_{t+1}-\boldsymbol{\mu}^{\mathbb{Q}}-\mathbf{\Phi}^{\mathbb{Q}}\mathbf{X}_t)\}
$$

Under the risk-neutral measure $\mathbb{Q}$, the process for the state variables is given by:

$$
\mathbf{X}_t=\boldsymbol{\mu}^{\mathbb{Q}}+\mathbf{\Phi}^{\mathbb{Q}}\mathbf{X}_{t-1}+\mathbf{\Sigma}\boldsymbol{\epsilon}^{\mathbb{Q}}_t, \quad \boldsymbol{\epsilon}^{\mathbb{Q}}_t\overset{iid}{\sim}\mathcal{N}(\mathbf{0},\mathbf{I}_k) \tag{15}
$$

for time $t=1,...,T$ and $\mathbf{X}_0$ given. Equation (15) is a first-order vector autoregression with Gaussian innovations. In other words, the change of measure results in the same source of underlying risk associated with the state variables but with a mean-correction. This concludes the change of measure.

The risk-neutral process for the state variables in equation (15) will be used for pricing zero-coupon bonds. Following Duffie and Kan (1996), it is assumed that prices of zero-coupon bonds are driven by the state variables in an *exponential-affine* fashion, which implies that *log*-prices are affine in the state variables. Bond pricing in the GATSM will be presented in the result below.

Result III.1. Bond pricing in the GATSM *Let $\mathcal{P}_t^n \equiv \mathcal{P}_t^n(\mathbf{X}_t)$ denote the no-arbitrage price of a n-period zero coupon bond at time $t = 1, ..., T$ for maturities $n = 1, ..., p$ driven by the state variables in equation (15). Following Duffie and Kan (1996), the price of the zero-coupon bond is exponential-affine in the state variables*:

$$\mathcal{P}_t^n = \exp\{A_n + \mathbf{B}_n'\mathbf{X}_t\}, \tag{16}$$

where A_n is a scalar and $\mathbf{B}_n$ is $k \times 1$. In this representation, the state variables $\mathbf{X}_t$ are also known as the pricing factors. A_n and $\mathbf{B}_n'$ are the corresponding factor loadings that essentially determine how the state variables load into the term structure. Under the assumption of no-arbitrage opportunities, the bond prices satisfy the following:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\mathcal{P}_{t+1}^{n-1}) \tag{17}$$

see, e.g. equation (4) in section III.A. Here, $\mathbb{E}_t^{\mathbb{Q}}$ is the expectation under the risk-neutral measure $\mathbb{Q}$ conditional on the information available at time t denoted by $\mathcal{I}_t$. The information set is generated by the state variables such that $\mathcal{I}_t = \{\mathbf{X}_t, ..., \mathbf{X}_0\}$.

Equation (17) and (16) result in the following cross-sectional restrictions:

$$A_n = A_{n-1} + \mathbf{B}_{n-1}'\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}_{n-1}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1} - \delta_0 \tag{18}$$

$$\mathbf{B}_n' = \mathbf{B}_{n-1}'\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}_1' \tag{19}$$

This is essentially ordinary difference equations which recursions are initiated at $A_0 = 0$ and $\mathbf{B}_0' = \mathbf{0}$. This corresponds to the no-arbitrage conditions: $A_1 = -\delta_0$ and $\mathbf{B}_1' = -\boldsymbol{\delta}_1'$, i.e., equal to the parameters governing the short rate process presented in definition III.5.

Proof. Result III.1 can be derived by applying the standard calculations in the term structure literature, see e.g., Ang and Piazzesi (2003). See the derivations in appendix IX.A □

Result III.1 implies that the yield and forward rate of the zero-coupon bond are affine in the state variables. The yield of a zero-coupon bond was introduced in definition III.3 in section III.A. The continuously compounded yield of the n-period zero-coupon bond at time t denoted by $\mathcal{Y}_t^n$ is given by:

$$\mathcal{Y}_t^n = -\frac{1}{n}\log\mathcal{P}_t^n$$

As the price $\mathcal{P}_t^n$ is exponentially-affine in the state variables, the yield is given by:

$$\mathcal{Y}_t^n = a_n + \mathbf{b}_n'\mathbf{X}_t$$

with $a_n = -\frac{A_n}{n}$ and $\mathbf{b}_n' = -\frac{\mathbf{B}_n'}{n}$. The initial conditions of A_n and $\mathbf{B}_n'$ imply that the yield of the one-period zero-coupon bond is equal to the short rate, i.e., $\mathcal{Y}_t^1 = \delta_0 + \boldsymbol{\delta}_1'\mathbf{X}_t \equiv r_t$. The yield of the zero-coupon bond is used to obtain the forward rate.

In definition III.4 in section III.A, the forward rate was expressed in terms of the yields. The continuously compounded forward rate of the one-period zero-coupon bond commencing at time $t+n$ denoted by $\mathcal{F}_t^n$ was given by:

$$\mathcal{F}_t^n = (n+1)\mathcal{Y}_t^{n+1} - n\mathcal{Y}_t^n$$

By using the expression for the yields, the forward rate is given by:

$$\mathcal{F}_t^n = \alpha_n + \boldsymbol{\beta}_n'\mathbf{X}_t \tag{20}$$

where the factor loadings are given by:

$$\alpha_n = A_n - A_{n+1} = \delta_0 - \mathbf{B}_n'\boldsymbol{\mu}^{\mathbb{Q}} - \frac{1}{2}\mathbf{B}_n'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_n \tag{21}$$

$$\boldsymbol{\beta}_n' = \mathbf{B}_n' - \mathbf{B}_{n+1}' = \boldsymbol{\delta}_1' + \mathbf{B}_n' - \mathbf{B}_n'\boldsymbol{\Phi}^{\mathbb{Q}} \tag{22}$$

for time $t = 1, ..., T$ and maturities $n = 1, ..., p$.

In the empirical analysis in section V, the GATSM will be estimated based on 1-month forward rates of zero-coupon bonds with maturities in $n = \{3, 6, 12, 24, 60, 84, 120\}$ months and with $k = 3$ unobserved state variables or factors. Therefore, it is convenient to stack the forward rates and factor loadings in the following way:

$$
\begin{aligned}
\mathbf{A} &= \Big(\alpha_3, \alpha_6, \alpha_{12}, \alpha_{24}, \alpha_{60}, \alpha_{84}, \alpha_{120}\Big)' \\
\mathbf{B}' &= \Big(\boldsymbol{\beta}_3', \boldsymbol{\beta}_6', \boldsymbol{\beta}_{12}', \boldsymbol{\beta}_{24}', \boldsymbol{\beta}_{60}', \boldsymbol{\beta}_{84}', \boldsymbol{\beta}_{120}'\Big)' \\
\mathbf{F}_t &= \Big(\mathcal{F}_t^3, \mathcal{F}_t^6, \mathcal{F}_t^{12}, \mathcal{F}_t^{24}, \mathcal{F}_t^{60}, \mathcal{F}_t^{84}, \mathcal{F}_{t,}^{120}\Big)'
\end{aligned}
$$

and the resulting GATSM becomes:

$$
\mathbf{F}_t = \mathbf{A} + \mathbf{B}'\mathbf{X}_t, \tag{23}
$$

for time $t = 1, ..., T$ and $\mathbf{X}_0$ fixed. Equation (23) provides a basis for the econometric methodology introduced in section IV. This concludes the derivation of the GATSM. The next section will introduce the regime-switching term structure model.

C. Regime-Switching Term Structure Model (RSTSM)

The Gaussian affine term structure model (hereafter, "GATSM") has been widely applied for modeling the term structure of interest rates, see e.g., Singleton (2006). However, several papers in the term structure literature, including Gray (1996), Bansal and Zhou (2002), Dai et al. (2007), Ang et al. (2008), Kim and Singleton (2012) and Bikbov and Chernov (2013), find that interest rates evolve in so-called regimes. Typically, the regimes are unobserved but somehow characterized by asymmetric movements in interest rates, e.g., the regimes in Dai et al. (2007) and Ang et al. (2008) are characterized by business cycles.

This section introduces regime shifts in the GATSM presented in section III.B. The resulting model is called the *regime-switching term structure model* (hereafter, "RSTSM"). Specifically, the RSTSM introduces unobserved monetary policy regimes in the process for the short rate, where the regimes constitute a *no* lower bound regime and a lower bound regime for the short rate. This implies that the RSTSM is equivalent to the GATSM, except for the assumed process for the short rate. This will become clear in the complete definition of the RSTSM provided below.

Definition III.6. RSTSM *Under the physical measure $\mathbb{P}$, the regime-switching term structure model (RSTSM) is defined by the following assumptions for time $t = 1, ..., T$ and maturities $n = 1, ..., p$:*

1. *The short rate, r_t, is piece-wise affine in the k state variables, $\mathbf{X}_t$, i.e.,*

$$r_t = \delta_0(s_t) + \boldsymbol{\delta}_1'(s_t)\mathbf{X}_t, \tag{24}$$

where $\delta_0(s_t)$ is a scalar and $\boldsymbol{\delta}_1(s_t)$ is a $k \times 1$ vector of parameters. Here, the parameters are dependent on the regimes, s_t, defined below. The assumed process for the short rate allows the state variables to load differently into the short rate across regimes.

2. *The regimes are unobserved but characterized by a no lower bound regime (NB) and a lower bound regime (LB). The regimes, $s_t \in \mathcal{S}$ with $\mathcal{S} = \{NB, LB\}$, evolve according to an unobserved 2-state Markov chain with a transitional probability matrix given by:*

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{NB,NB} & \pi_{LB,NB} \\ \pi_{NB,LB} & \pi_{LB,LB} \end{pmatrix} \tag{25}$$

such that $\pi_{i,j}$ is the likelihood of $s_t = j$ given that the last regime was $s_{t-1} = i$, where $i, j \in \mathcal{S}$ with $\sum_{j \in \mathcal{S}} \pi_{i,j} = 1$. This specification of the regimes was especially studied in Hamilton (1994). The assumed process for the regimes allows the behavior of the regimes to be persistent, which is suitable in this context. Note, the regime probabilities denoted by $\pi_{i,j}$ are assumed to be identical under the physical and risk-neutral measures. Intuitively, this rules out the possibility that agents require risk compensation associated with unpredictable regime shifts, see e.g., Ang et al. (2008).

3. Similar to the GATSM presented in section III.B, the state variables are unobserved or latent and follow a first-order vector autoregression with Gaussian innovations:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{P}}, \quad \boldsymbol{\epsilon}_t^{\mathbb{P}} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k) \tag{26}$$

where $\boldsymbol{\mu}^{\mathbb{P}}$ is $k \times 1$, $\boldsymbol{\Phi}^{\mathbb{P}}$ is $k \times k$, $\boldsymbol{\Sigma}$ is $k \times k$ and $\mathbf{X}_0$ is given. Following Ang et al. (2008), it is assumed that the regimes evolve independently of the Gaussian innovations $\boldsymbol{\epsilon}_t^{\mathbb{P}}$.

4. Similar to the GATSM presented in section III.B, the pricing kernel follows:

$$\mathcal{M}_{t+1} = \exp\{-r_t\}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1} \tag{27}$$

where $(d\mathbb{Q}/d\mathbb{P})_{t+1}$ denotes the Radon-Nikodym derivative of the risk-neutral measure $\mathbb{Q}$ with respect to the physical measure $\mathbb{P}$. Similar to the GATSM, it is assumed that the Radon-Nikodym derivative follows:

$$\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1} = \exp\{-\frac{1}{2}\boldsymbol{\Lambda}_t'\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t'\boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}\}, \tag{28}$$

where $\boldsymbol{\Lambda}_t$ is a $k \times 1$ vector of market prices of risk associated with the innovations in the state variables, $\boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}$, and it follows:

$$\boldsymbol{\Lambda}_t = \lambda_0 + \boldsymbol{\lambda}_1'\mathbf{X}_t \tag{29}$$

where λ_0 is $k \times 1$ and $\boldsymbol{\lambda}_1'$ is $k \times k$.

1., 2., 3. and 4. implicitly define the pricing kernel. In the RSTSM, the term structure of interest rates is determined by 1. the piece-wise affine relation between the short rate and the unobserved state variables dictated by the regimes in 2., 3. the evolution of the state variables and 4. the attitude towards risk associated with Gaussian innovations in the state variables.

The RSTSM introduced in definition III.6 is the *multi-factor* version of the RSTSM with k factors or state variables. As in the GATSM, the *three-factor* RSTSM with $k = 3$ factors will be implemented in the empirical analysis in section V.

In equation (28) in definition III.6, it is assumed that the market prices of risk are solely associated with the unpredictable variation in the state variables. However, one may argue that the market prices of risk are also related to the regimes. For instance, Dai et al. (2007) consider a different specification of market prices of risk, where agents require risk premia for being exposed to unpredictable regime shifts. Alternatively, Ang et al. (2008) examine regime shifts in the market prices of risk, but with identical regime probabilities under the physical and risk-neutral measures as in definition III.6. The specifications in Dai et al. (2007) and Ang et al. (2008) can be seen as possible extensions of equation (28) in definition III.6. Furthermore, the regime probabilities denoted by $\pi_{i,j}$ for $i, j \in \mathcal{S} = \{NB, LB\}$ are assumed to be constant across time. This formulation is also used in Ang et al. (2008). However, Dai et al. (2007) consider time-varying regime probabilities under the physical measure but constant regime probabilities under the risk-neutral measure to ensure tractable bond prices. These model extensions will be reviewed in the discussion section VI.

There is a significant difference between the RSTSM presented in definition III.6 and the models proposed in Dai et al. (2007) and Ang et al. (2008). Unlike the RSTSM in defintion III.6, Dai et al. (2007) and Ang et al. (2008) introduce regime shifts in the process for the state variables so that the state variables follow a first-order regime-switching vector autoregression with Gaussian innovations. The difference is due to the purpose; the regimes in Dai et al. (2007) and Ang et al. (2008) are related to business cycles, while the regimes in the RSTSM in definition III.6 are related to monetary policy.

In definition III.6, the state variables evolve independently of the regimes. Therefore the state variables are still conditionally Gaussian distributed, i.e., $\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathcal{N}(\boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_t, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$. Furthermore, the assumed process for the Radon-Nikodym derivative in definition III.6 is identical to the process in the GATSM. Consequently, the change of measure follows one-to-one the steps in the GATSM in section III.B. To avoid a complete repetition, the risk-neutral process of the state variables are presented without derivations:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{Q}}, \quad \boldsymbol{\epsilon}_t^{\mathbb{Q}} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k) \tag{30}$$

where $\boldsymbol{\mu}^{\mathbb{Q}} = \boldsymbol{\mu}^{\mathbb{P}} - \boldsymbol{\Sigma}\lambda_0$ and $\boldsymbol{\Phi}^{\mathbb{Q}} = \boldsymbol{\Phi}^{\mathbb{P}} - \boldsymbol{\Sigma}\boldsymbol{\lambda}_1'$, for $t = 1, ..., T$ and $\mathbf{X}_0$ fixed. Equation (30) is used for bond pricing in the RSTSM. Following Dai et al. (2007) and Ang et al. (2008), the price of a zero-coupon bond is assumed to be exponential-affine in the state variables. Unlike the GATSM, the factor loadings are regime-dependent due to the regimes in the short rate process in definition III.6. Consequently, bond pricing in the RSTSM does not exist in closed-form and approximations similar to Bansal and Zhou (2002) are used.[x] This will become clear in the result below.

Result III.2. Bond pricing in the RSTSM *Let $\mathcal{P}_t^n \equiv \mathcal{P}_t^n(\mathbf{X}_t)$ denote the no-arbitrage price of a n-period zero coupon bond at time t for time $t = 1, ..., T$ and maturities $n = 1, ..., p$. Following Ang et al. (2008) and Bansal and Zhou (2002), the price is exponential-affine in the state variables or pricing factors with regime-dependent loadings:*

$$\mathcal{P}_t^n = \exp\{A_n(s_t) + \mathbf{B}_n'(s_t)\mathbf{X}_t\}, \tag{31}$$

where $A_n(s_t)$ is a scalar and $\mathbf{B}_n(s_t)$ is $k \times 1$. Here, s_t dictates the regimes introduced in definition III.6, while the pricing factors $\mathbf{X}_t$ evolve according to equation (30). Under the risk-neutral measure, the no-arbitrage bond price satisfies the following[xi]:

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\mathcal{P}_{t+1}^{n-1}) \tag{32}$$

where the short rate $r_t \equiv r_t(s_t, \mathbf{X}_t)$ is piece-wise affine in the factors $\mathbf{X}_t$ in equation (30) dictated by the regimes s_t in definition III.6. In equation (32), the expectation is under the risk-

[x] Bansal and Zhou (2002) propose a first-order approximation to the solution of equation (31). Hördahl and Tristani (2019) find that the approximation error is smaller than 0.1 basis points for zero-coupon bonds with maturity up to 18 months.

[xi] Equation (4) in section III.A derives the pricing equation.

neutral measure $\mathbb{Q}$ and conditional on the information available at time t denoted by $\mathcal{I}_t$. The information set is generated by the state variables and the regimes $\mathcal{I}_t = \{\mathbf{X}_t, s_t, ..., \mathbf{X}_0, s_0\}$. First-order approximations proposed in Bansal and Zhou (2002) lead to the following no-arbitrage restrictions:

$$A_n(i) = \sum_{j \in \mathcal{S}} \pi_{i,j} \Big(A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j) - \delta_0(i) \Big) \tag{33}$$

and

$$\mathbf{B}'_n(i) = \sum_{j \in \mathcal{S}} \pi_{i,j} \Big(\mathbf{B}'_{n-1}(j)\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i) \Big) \tag{34}$$

for $i, j \in \mathcal{S} = \{NB, LB\}$. The recursions are initiated at $A_1(i) = -\delta_0(i)$ and $\mathbf{B}'_1(i) = -\boldsymbol{\delta}'_1(i)$, i.e., equal to the parameters governing the short rate process presented in definition III.6. The no-arbitrage restrictions in equation (33) and (34) can be seen as a regime-weighted average of the loadings, see e.g., Bansal and Zhou (2002) and Hördahl and Tristani (2019).

Proof. Result III.2 can be derived by applying the standard calculations in the term structure literature, see e.g., Ang et al. (2008), along with the approximations proposed in Bansal and Zhou (2002). See the derivations in appendix IX.B □

Result III.2 implies that the yield and forward rate of the zero-coupon bond are affine in the state variables. The yield of a zero-coupon bond was introduced in definition III.3 in section III.A. The continuously compounded yield of the n-period zero-coupon bond at time t denoted by $\mathcal{Y}_t^n$ is given by:

$$\mathcal{Y}_t^n = a_n(s_t) + \mathbf{b}'_n(s_t)\mathbf{X}_t$$

with $a_n(s_t) = -\frac{A_n(s_t)}{n}$ and $\mathbf{b}'_n(s_t) = -\frac{\mathbf{B}'_n(s_t)}{n}$. The initial conditions in result III.2 imply that the one-period ahead yield is equal to the short rate $\mathcal{Y}_t^1 = \delta_0(s_t) + \boldsymbol{\delta}'_1(s_t)\mathbf{X}_t \equiv r_t$.

The forward rate of a zero-coupon bond was introduced in definition III.3 in section III.A.

The continuously compounded one-period forward rate at time t denoted by $\mathcal{F}_t^n$ is given by:

$$\mathcal{F}_t^n = \alpha_n(s_t) + \boldsymbol{\beta}_n'(s_t)\mathbf{X}_t \tag{35}$$

with $\alpha_n = A_n(s_t) - A_{n+1}(s_t)$ and $\boldsymbol{\beta}_n'(s_t) = \mathbf{B}_n'(s_t) - \mathbf{B}_{n+1}'(s_t)$.

In the empirical analysis in section V, the RSTSM will be based on one-month forward rates on zero-coupon bonds with maturities in $n = \{3, 6, 12, 24, 60, 84, 120\}$ months and with $k = 3$ unobserved state variables. Therefore, the RSTSM is given by:

$$\mathbf{F}_t = \mathbf{A}(s_t) + \mathbf{B}'(s_t)\mathbf{X}_t \tag{36}$$

where

$$\begin{aligned}
\mathbf{A}(s_t) &= \Big(\alpha_3(s_t), \alpha_6(s_t), \alpha_{12}(s_t), \alpha_{24}(s_t), \alpha_{60}(s_t), \alpha_{84}(s_t), \alpha_{120}(s_t)\Big)' \\
\mathbf{B}'(s_t) &= \Big(\boldsymbol{\beta}_3'(s_t), \boldsymbol{\beta}_6'(s_t), \boldsymbol{\beta}_{12}'(s_t), \boldsymbol{\beta}_{24}'(s_t), \boldsymbol{\beta}_{60}'(s_t), \boldsymbol{\beta}_{84}'(s_t), \boldsymbol{\beta}_{120}'(s_t)\Big)' \\
\mathbf{F}_t &= \Big(\mathcal{F}_t^3, \mathcal{F}_t^6, \mathcal{F}_t^{12}, \mathcal{F}_t^{24}, \mathcal{F}_t^{60}, \mathcal{F}_t^{84}, \mathcal{F}_{t,}^{120}\Big)'
\end{aligned}$$

for $t = 1, ..., T$ and $(\mathbf{X}_0, s_0)$ fixed. The next section describes how estimation of the GATSM in equation (23) and the RSTSM in equation (36) proceed, respectively.

IV. Econometric methodology

This section introduces the econometric methodology for the Gaussian affine term structure model (hereafter, "GATSM") and the regime-switching term structure model (hereafter, "RSTSM") introduced in the sections III.B-III.C. Estimation of the term structure models entails a dual problem of state and parameter estimation. In the GATSM, the estimation problem is to obtain estimated values of the parameters and state variables. In the RSTSM, the estimation problem is to obtain estimated values of the parameters, state variables and regimes. The estimation strategy is to combine maximum likelihood estimation along with filtering methods. Maximum likelihood estimation is used in both term structure models to obtain the parameter values. The Kalman filter is used to obtain the state variables in GATSM and the Kim filter is used to obtain the state variables and regimes in the RSTSM.

In the following, state estimation refers to estimation of the state variables and regimes, while parameter estimation refers to estimation of the parameters in the models.

This section sets the scene by deriving the state space representation of the GATSM and the RSTSM. Subsequently, parameter estimation is presented by a brief introduction to the principle of maximum likelihood estimation. Finally, the filtering methods used for state estimation are presented in the same order as the term structure models were derived in the sections III.B-III.C, i.e., the GATSM followed by the RSTSM.

A. State space representations

This section introduces the state space representation of the GATSM and RSTSM. The state space representation is a system of equations that describes the evolution of observed variables in terms of unobserved variables. In this context, the forward rates are observed, while the state variables and regimes are unobserved.

In the GATSM, the forward rates are given by an affine transformation of the state variables.[xii] This relation implies that the unobserved state variables can be indirectly obtained through the observed forward rates which in turn requires that the total number of forward rates available for estimation are identical to the number of state variables. Typically, the total number of forward rates available for estimation exceeds the number of state variables. A common estimation strategy for handling this issue is to assume that a part of or *all* forward rates are priced with errors, see, e.g., Singleton (2006). Following Wu and Xia (2016), it is assumed that all forward rates in the GATSM and RSTSM are priced with independent and identically distributed (*i.i.d.*) Gaussian errors, i.e., $\boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. p indicates that all the forward rates are priced with errors. This is essentially an ad hoc assumption, but it implies that filtering methods are applicable in state estimation, see Singleton (2006).[xiii] Typically, the errors are motivated by some sort of pricing errors in the market e.g., valuation techniques, market liquidity, bid-ask spreads etc. The state space representation of the GATSM and RSTSM are presented in the definitions below.

Definition IV.1. State space representation of the GATSM

Under the physical measure $\mathbb{P}$, the state space representation of the Gaussian affine term structure model (GATSM) is given by:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{P}} \qquad \text{(State equation)}$$

$$\mathbf{F}_t = \mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t \qquad \text{(Measurement equation)}$$

for $t = 1, ..., T$, $(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ and with $\mathbf{X}_0$ fixed. Here, the unobserved state variables or factors $\mathbf{X}_t$ are $k \times 1$ such that the parameters are $\boldsymbol{\mu}^{\mathbb{P}}$ $k \times 1$, $\boldsymbol{\Phi}^{\mathbb{P}}$ $k \times k$ and $\boldsymbol{\Sigma}$ $k \times k$. The forward rates $\mathbf{F}_t$ are $p \times 1$, the factor loadings are $\mathbf{A}$ $p \times 1$, $\mathbf{B}'$ $p \times k$ and finally $\boldsymbol{\Omega}$ $p \times p$. Note, the factor loadings depend on the parameters δ_0 1×1, $\boldsymbol{\mu}^{\mathbb{Q}}$ $k \times 1$, $\boldsymbol{\Sigma}$ $k \times k$, $\boldsymbol{\delta}_1$ $k \times 1$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$ $k \times k$, see, e.g., the equations (21)-(22) in section III.B.

[xii] In the RSTSM, the forward rates are piece-wise affine in the state variables due to the regimes.

[xiii] The assumption is further suitable in this context as filtering methods are used in state estimation of the RSTSM, which means that filtering methods are used in state estimation of both the GATSM and RSTSM.

Definition IV.2. State space representation of the RSTSM

Under the physical measure $\mathbb{P}$, the state space representation of the regime-switching term structure model (RSTSM) is given by:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{P}} \qquad \text{(State equation)}$$

$$\mathbf{F}_t = \mathbf{A}(s_t) + \mathbf{B}'(s_t)\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t \qquad \text{(Measurement equation)}$$

for $t = 1, ..., T$, $(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ *and with* $(\mathbf{X}_0, s_0)$ *fixed. The unobserved state variables* $\mathbf{X}_t$ *are* $k \times 1$ *with the parameters* $\boldsymbol{\mu}^{\mathbb{P}}$ $k \times 1$, $\boldsymbol{\Phi}^{\mathbb{P}}$ $k \times k$ *and* $\boldsymbol{\Sigma}$ $k \times k$. *The forward rates* $\mathbf{F}_t$ *are* $p \times 1$ *with* $\mathbf{A}$ $p \times 1$, $\mathbf{B}'$ $p \times k$ *and* $\boldsymbol{\Omega}$ $p \times p$. *The factor loadings, i.e.,* $\mathbf{A}$ *and* $\mathbf{B}'$, *depend on the parameters* π_{s_{t-1},s_t} 1×1, $\delta_0(s_t)$ 1×1, $\boldsymbol{\mu}^{\mathbb{Q}}$ $k \times 1$, $\boldsymbol{\Sigma}$ $k \times k$, $\boldsymbol{\delta}_1(s_t)$ $k \times 1$ *and* $\boldsymbol{\Phi}^{\mathbb{Q}}$ $k \times k$, *see, e.g., the equations (33)-(34) in section III.C. The regimes constitute a no lower bound regime (NB) and a lower bound regime (LB) such that* $s_t \in \mathcal{S} = \{NB, LB\}$ *and evolve according to an unobserved 2-state Markov chain with transitional probability matrix:*

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_{NB,NB} & \pi_{LB,NB} \\ \pi_{NB,LB} & \pi_{LB,LB} \end{pmatrix} \tag{37}$$

$\pi_{i,j}$ *is the likelihood of* $s_t = j$ *given that the last regime was* $s_{t-1} = i$, *where* $i, j \in \mathcal{S} = \{NB, LB\}$, *with* $\sum_{j \in \mathcal{S}} \pi_{i,j} = 1$.

The GATSM and RSTSM summarized by the system of equations in the definitions IV.1-IV.2 are used in the subsequent sections.

B. Maximum likelihood estimation

This section introduces parameter estimation using maximum likelihood estimation. Maximum likelihood estimation is a method to estimate the parameters in a given model by maximizing its corresponding likelihood function. Intuitively, maximum likelihood estimation seeks to find the parameters that maximize the likelihood of observing the actual sample. The parameters that solve the maximization problem are the maximum likelihood estimates.

The first step in maximum likelihood estimation is to derive the likelihood function for the given model. Knowledge of the probability density functions of the forward rates is required to derive the likelihood function for the two models. In the definitions IV.1-IV.2, the state space representations were defined under the assumption of *i.i.d.* Gaussian errors. Under this assumption and the initial values, the joint process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ is multivariate Gaussian distributed for the GATSM. For the RSTSM, the joint process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ is multivariate Gaussian conditional on the regimes. Therefore, the process for the forward rates $\{\mathbf{F}_t\}_{t=1}^T$ and the conditional process $\{\mathbf{F}_t|\mathbf{F}_{t-1}, ..., \mathbf{F}_1\}_{t=2}^T$ will also be Gaussian distributed for the GATSM and conditionally Gaussian distributed for the RSTSM. The Gaussian distribution is completely specified by its mean and variance, which implies that the likelihood function will be given by the conditional mean and variance of the forward rates. As the forward rates are observed under the physical measure $\mathbb{P}$, the likelihood function will be based on the probability density function under this measure.

Let $f^{\mathbb{P}}$ denote the probability density function under the physical measure, let all the parameters be collected in a vector $\boldsymbol{\theta}$ and let the notation $\mathbf{F}_{1:T}$ denote the forward rates from 1 to T, i.e., $\mathbf{F}_1, ..., \mathbf{F}_T$. The likelihood function in terms of the observed forward rates is given by:

$$L_T(\boldsymbol{\theta}) = f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_{1:T}) = f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_1) \times \prod_{t=2}^{T} f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}), \tag{38}$$

where the likelihood function $L_T(\boldsymbol{\theta})$ is a function of the parameters $\boldsymbol{\theta}$ given the sample T. The first term $f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_1)$ will be based on the initialization such that $\mathbf{F}_1$ is given by $\mathbf{X}_1$ in the GATSM and $\mathbf{F}_1$ is given by $(\mathbf{X}_1, s_1)$ in the RSTSM. The exact form of the term $f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1})$ depends on the term structure model, but it is completely given by the conditional mean and variance of the forward rates. These conditional moments are obtained by the filtering methods introduced subsequently in the sections IV.C-IV.D.

In application, it is convenient to consider the log-likelihood function, $\mathcal{L}_T(\boldsymbol{\theta}) = \log L_T(\boldsymbol{\theta})$, since it leads to a likelihood function of a sum rather than a product of $f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1})$. This adds numerical stability in application as sums are more stable compared to products, see, e.g., Cameron and Trivedi (2005). Based on the log-likelihood function, the maximum likelihood estimator solves the following maximization problem:

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\ \mathcal{L}_T(\boldsymbol{\theta}),$$

where $\boldsymbol{\Theta}$ denotes the parameter space. A well-known problem is that the parameters are not identified at the first onset.

The identification problem is that the parameters in the state equation only enter the measurement equation through products, see the definitions IV.1-IV.2. In other words, there exist several parameter values that produce the same likelihood value within the model resulting in an observationally equivalent structure. Consequently, the maximum likelihood estimator cannot uniquely identify the optimal parameters in the model. This identification issue is a well-known discussion in the term structure literature, see Dai and Singleton (2000), Joslin, Singleton, and Zhu (2010) and Hamilton and Wu (2012).

There are several ways in which identification of the parameters can be obtained. Typically, the parameters are restricted or fixed at some values, which is essentially an ad hoc assumption, but a way to obtain identification. For the GATSM and the RSTSM, the identifying restrictions in Dai and Singleton (2000) can be used.[xiv] The restrictions are imposed on the parameters in the process for the state variables. Following Dai and Singleton (2000) and Singleton (2006), the restrictions are given by:

$$\boldsymbol{\mu}^{\mathbb{P}} = \mathbf{0}, \quad \boldsymbol{\Phi}^{\mathbb{P}} \text{ and } \boldsymbol{\Phi}^{\mathbb{Q}} \text{ lower triangular}, \quad \boldsymbol{\Sigma} = \mathbf{I}_k \tag{39}$$

These restrictions are used to prevent the state variables from shifting, rotating and scaling. This identification scheme is suitable as the restrictions in (39) allow the parameters in the

[xiv] Wu and Xia (2016) use a more restrictive identification scheme proposed in Joslin et al. (2010), where the pricing errors of the forward rates are restricted to be the same. The identification scheme used here is less restrictive as the pricing errors of the forward rates are allowed to be different.

process for the short rate to be treated as free parameters.[xv] Furthermore, the identifying restrictions effectively restrict the parameters of the market prices of risk λ_0 and $\boldsymbol{\lambda}_1'$. This emerges directly from $\lambda_0 = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{\mathbb{P}} - \boldsymbol{\mu}^{\mathbb{Q}})$ and $\boldsymbol{\lambda}_1' = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Phi}^{\mathbb{P}} - \boldsymbol{\Phi}^{\mathbb{Q}})$. As the parameters of the market prices of risk are notoriously difficult to estimate, see, e.g., Ang et al. (2008) and Wu and Xia (2016), $\boldsymbol{\mu}^{\mathbb{Q}}$, $\boldsymbol{\Phi}^{\mathbb{Q}}$ and $\boldsymbol{\Phi}^{\mathbb{P}}$ will be estimated such that both λ_0 and $\boldsymbol{\lambda}_1'$ are implicitly determined. The resulting parameters to be estimated are 47 in the GATSM and 53 in the RSTSM for $k = 3$ state variables and $p = 7$ forward rates.

Consistency and asymptotic Gaussianity require that certain regularity conditions are met. The regularity conditions depend on the filtering methods used. Therefore, it is natural to revisit the regularity conditions in continuation of the filtering methods in the next sections.

C. *The Kalman filter and the GATSM*

This section introduces the filtering method used to obtain the state variables in the GATSM. In definition IV.1, the state space representation of the GATSM was presented, and it was given by:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{P}} \qquad \text{(State equation)}$$

$$\mathbf{F}_t = \mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t \qquad \text{(Measurement equation)}$$

for $t = 1, ..., T$ with $(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ and with $\mathbf{X}_0$ fixed. Here, the estimation problem is to compute the unobserved state variables given the observed forward rates.

The estimation strategy is to apply a Kalman filter to the GATSM summarized by the state space representation above. The Kalman filter was introduced in Kalman (1960). The Kalman filter is a recursive algorithm that produces estimates of unobserved variables based on a series of observed variables. In this context, the Kalman filter is used to produce estimates of the conditional mean and variance of the forward rates. As the conditional moments

[xv]Theoretically, the identifying restrictions should not affect the relative estimates of the parameters. In application, Hamilton and Wu (2012) find that different identification schemes lead to different relative estimates of the parameters for various reasons including the numerical optimization.

of the forward rates depend on the unobserved state variables, the Kalman filter is applied to obtain the state variables. The conditional moments are used in parameter estimation covered in section IV.B.

The state variables will be computed under the physical measure $\mathbb{P}$ as the forward rates are observed under this measure. The following notation will be used:

$$\mathbf{X}_{t|\tau} = \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:\tau}) \qquad \mathbf{P}^{\mathbf{X}}_{t|\tau} = \mathbb{V}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:\tau})$$
$$\mathbf{F}_{t|\tau} = \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:\tau}) \qquad \mathbf{P}^{\mathbf{F}}_{t|\tau} = \mathbb{V}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:\tau})$$

for $t = 1, ..., T$. $1:\tau$ denotes the series from 1 to $\tau \leq t$. The Kalman filter is given below.

Result IV.1. Kalman filter *The Kalman filter of the GATSM presented in definition IV.1 with initial conditions* $\mathbf{X}_{0|0}$ *and* $\mathbf{P}^{\mathbf{X}}_{0|0}$ *fixed is as follows:*
Prediction step:

$$\mathbf{X}_{t|t-1} = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1|t-1} \tag{40}$$
$$\mathbf{P}^{\mathbf{X}}_{t|t-1} = \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{P}^{\mathbf{X}}_{t-1|t-1}\boldsymbol{\Phi}^{\mathbb{P}\prime} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \tag{41}$$

and update step:

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t(\mathbf{F}_t - \mathbf{A} - \mathbf{B}'\mathbf{X}_{t|t-1}) \tag{42}$$
$$\mathbf{P}^{\mathbf{X}}_{t|t} = (\mathbf{I}_k - \mathbf{K}_t\mathbf{B}')\mathbf{P}^{\mathbf{X}}_{t|t-1} \tag{43}$$

with Kalman gain:

$$\mathbf{K}_t \equiv \mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B}(\mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B} + \boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1} \tag{44}$$

for $t = 1, ..., T$.

Proof. See derivations in appendix X.A. □

The Kalman filter runs through $t = 1, ..., T$ but requires initialization. The initialization is to set the initial values $\mathbf{X}_{0|0}$ and $\mathbf{P}^{\mathbf{X}}_{0|0}$. There are several ways in which to initialize the algorithm. For instance, if the process $\{\mathbf{X}_t\}_{t=1}^{T}$ is stationary, $\mathbf{X}_{0|0}$ $\mathbf{P}^{\mathbf{X}}_{0|0}$ can be drawn from the stationary distribution, i.e., $\mathbf{X}_{0|0} = \left(\mathbf{I}_k - \mathbf{\Phi}^{\mathbb{P}}\right)^{-1}\boldsymbol{\mu}^{\mathbb{P}}$ and $\mathbf{P}^{\mathbf{X}}_{0|0} = \left(\mathbf{I}_{k^2} - \mathbf{\Phi}^{\mathbb{P}} \otimes \mathbf{\Phi}^{\mathbb{P}}\right)^{-1}\text{vec}\big(\mathbf{\Sigma\Sigma}'\big)$, see Lütkepohl (2007). Alternatively, one could set uninformative values, i.e., $\mathbf{X}_{0|0} = \mathbf{0}$ and with $\mathbf{P}^{\mathbf{X}}_{0|0}$ sufficiently large such that $(\mathbf{P}^{\mathbf{X}}_{0|0})^{-1} \to \mathbf{0}$. Too uninformative values will affect the first couple of iterations in the algorithm which may therefore be discarded if the sample size is sufficiently large. In practice, one may use several starting values to assess its importance.

The first step in the Kalman filter is the prediction step. In the prediction step, the algorithm estimates the current state variables along with their errors based on the previous observed forward rates. The second step is the update step. In the update step, the algorithm updates the estimates by using a weighted average of the next observed forward rates and the predicted state variables. The updated state variables are also known as the filtered state variables. The Kalman gain controls the relative weight given to the observed forward rates and the predicted state variables. If the forward rates are observed with poor accuracy, the Kalman gain will be low. A low Kalman gain means that the filtered state variables will be close to the predicted state variables. If the observed forward rates are highly informative, the Kalman gain will be relatively large, and the Kalman filter will put more weight on the recent observed forward rates.

From the Kalman filter, the conditional mean and variance of the forward rates can be directly obtained for $t = 1, ..., T$ by:

$$\mathbf{F}_{t|t-1} = \mathbf{A} + \mathbf{B}'\mathbf{X}_{t|t-1}$$
$$\mathbf{P}^{\mathbf{F}}_{t|t-1} = \mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B} + \mathbf{\Omega\Omega}'$$

Here, it is clear that the conditional moments of the forward rates depend on the predicted state variables, which are calculated recursively in the Kalman filter. Given the conditional moments of the forward rates, the log-likelihood function can be derived. The parameters in the GATSM are given by $\boldsymbol{\theta} = (\boldsymbol{\mu}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{P}}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \delta_0, \boldsymbol{\delta}_1')'$ see definition IV.1 in section IV.A. As the conditional process $\{\mathbf{F}_t|\mathbf{F}_{t-1}, ..., \mathbf{F}_1\}_{t=2}^T$ is Gaussian distributed, its probability density function is given by:

$$f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}) = (2\pi)^{-\frac{p}{2}}|\mathbf{P}_{t|t-1}^{\mathbf{F}}|^{-\frac{1}{2}} \times \exp\{-\frac{1}{2}(\mathbf{F}_t - \mathbf{F}_{t|t-1})'(\mathbf{P}_{t|t-1}^{\mathbf{F}})^{-1}(\mathbf{F}_t - \mathbf{F}_{t|t-1})\} \tag{45}$$

and the resulting log-likelihood function is given up to constants by:

$$\mathcal{L}_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=2}^{T}\Big\{\log|\mathbf{P}_{t|t-1}^{\mathbf{F}}| + (\mathbf{F}_t - \mathbf{F}_{t|t-1})'(\mathbf{P}_{t|t-1}^{\mathbf{F}})^{-1}(\mathbf{F}_t - \mathbf{F}_{t|t-1})\Big\} \tag{46}$$

This log-likelihood function is used to estimate the parameters in the GATSM. If the assumption $(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ applies, the Kalman filter gives the maximum likelihood estimator. Otherwise, the Kalman filter gives the quasi maximum likelihood estimator, which has a different asymptotic variance than the usual maximum likelihood estimator, see, e.g., Harvey (1991) and Rahbek and Pedersen (2019a). The estimator is consistent and asymptotically Gaussian distributed, if certain regularity conditions are met. The regularity conditions include that the errors $\{(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)'\}_{t=1}^T$ are martingale differences with finite fourth order moments and that the forward rates $\{\mathbf{F}_t\}_{t=1}^T$ are stationary and weakly mixing, see Dunsmuir (1979) and Shumway and Stoffer (2000). As the forward rates depend on the state variables, the process $\{\mathbf{X}_t\}_{t=1}^T$ must be stationary, which requires that the eigenvalues of $\boldsymbol{\Phi}^{\mathbb{P}}$ are less than one in absolute value, see also Shumway and Stoffer (2000) for a further discussion. For a sufficiently large sample size, consistency ensures that the maximum likelihood estimator will converge in probability to the true parameters, while asymptotic Gaussianity ensures that the maximum likelihood estimator will converge in distribution to a Gaussian distribution.

D. The Kim filter and the RSTSM

This section introduces the filtering method used to obtain the unobserved state variables and regimes in the RSTSM. As will be explained in this section, the filtering method used in the GATSM becomes computational intractable when regimes are introduced. Therefore, a slightly different filtering method is presented in this section.

For convenience, the state space representation of the RSTSM presented in definition IV.2 in section IV.A is stated again:

$$\mathbf{X}_t = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t^{\mathbb{P}} \qquad \text{(State equation)}$$

$$\mathbf{F}_t = \mathbf{A}(s_t) + \mathbf{B}'(s_t)\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t \qquad \text{(Measurement equation)}$$

for $t = 1, ..., T$ with $(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ and with $(\mathbf{X}_0, s_0)$ fixed. The regimes constitute a *no* lower bound regime (NB) and a lower bound regime (LB) such that $s_t \in \mathcal{S} = \{NB, LB\}$. The regimes evolve according to an unobserved 2-state Markov chain with transitional probability matrix:

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_{NB,NB} & \pi_{LB,NB} \\ \pi_{NB,LB} & \pi_{LB,LB} \end{pmatrix} \tag{47}$$

$\pi_{i,j}$ is the likelihood of $s_t = j$ given that the last regime was $s_{t-1} = i$, where $i, j \in \mathcal{S} = \{NB, LB\}$, with $\sum_{j\in\mathcal{S}} \pi_{i,j} = 1$. This representation is also associated with a Markov-switching Gaussian state space model, see Kim (1994). Here, the estimation problem is to compute the unobserved state variables and regimes based on the observed forward rates.

The Kalman filter used in the GATSM is not used in the RSTSM for mainly two reasons. First, the regimes in the RSTSM make it difficult to evaluate the likelihood function of the observed forward rates as the regimes are unobserved. Second, every single regime path needs to be evaluated in the Kalman filter, which makes the Kalman filter computational intractable, see Kim and Nelson (1999). Therefore, the Kalman filter is modified such that it can produce estimates of the state variables and still be computational tractable for a moderate sample size.

The estimation strategy is to apply a Kim filter to obtain the state variables and regimes in the RSTSM.[xvi] The Kim filter was introduced in Kim (1994) and can be seen as a combina-

[xvi] Kim and Nelson (1999) propose another estimation strategy using Gibbs sampling. The Kim filter used here is more appropriate as it is an immediate extension of the Kalman filter that is already covered.

tion of a Kalman filter and a Hamilton filter. The Kalman filter produces estimates of the the state variables, while the Hamilton filter produces estimates of the regimes, see Hamilton (1989) and Hamilton (1994).

The idea of the Kim filter is to collapse the filtered state variables such that the number of evaluations increases linearly in time rather than exponentially. Specifically, the output of the Hamilton filter is used as an input in the Kalman filter to make it operable. This interaction relies on approximations, but Kim and Kang (2019) find that the approximation error is close to non-existent for highly persistent regimes. However, one should still keep in mind that this filtering method involves approximations, so the likelihood function of the RSTSM will also be a result of approximations.

Throughout this section, the following notation will be used for the conditional moments:

$$\mathbf{X}_{t|\tau}^{(i,j)} = \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:\tau}, s_t = j, s_{t-1} = i) \qquad \mathbf{P}_{t|\tau}^{\mathbf{X}(i,j)} = \mathbb{V}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:\tau}, s_t = j, s_{t-1} = i)$$
$$\mathbf{F}_{t|\tau}^{(i,j)} = \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:\tau}, s_t = j, s_{t-1} = i) \qquad \mathbf{P}_{t|\tau}^{\mathbf{F}(i,j)} = \mathbb{V}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:\tau}, s_t = j, s_{t-1} = i)$$

and the following notation for the regime probabilities:

$$\mathcal{Q}_{t|\tau}^{(i,j)} = \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:\tau}) \qquad \mathcal{Q}_{t|\tau}^{(j)} = \mathbb{P}_{\boldsymbol{\theta}}(s_t = j|\mathbf{F}_{1:\tau})$$

where $i, j \in \mathcal{S} = \{NB, LB\}$. The superscript (i, j) indicates that the conditional moments are regime-specific. In other words, there will be $2 \times 2 = 4$ combinations for each conditional moment. The Kim filter seeks to produce estimates of these regime-specific conditional moments, which can be used in parameter estimation covered in section IV.B. The Kim filter is presented on the next page.

Result IV.2. Kim filter *The Kim filter of the RSTSM presented in definition IV.D in section IV.A with initial conditions* $\mathbf{X}_{0|0}^{(i)}$, $\mathbf{P}_{0|0}^{\mathbf{X}(i)}$ *and* $\mathcal{Q}_{0|0}^{(i)}$ *fixed is as follows:*

Hamilton filter

Prediction step:

$$\mathcal{Q}_{t|t-1}^{(i,j)} = \pi_{i,j}\mathcal{Q}_{t-1|t-1}^{(i)} \tag{48}$$

and update step:

$$\mathcal{Q}_{t|t}^{(i,j)} = \frac{f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)}{f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1})}\mathcal{Q}_{t|t-1}^{(i,j)} \tag{49}$$

$$\mathcal{Q}_{t|t}^{(j)} = \sum_{i\in\mathcal{S}}\mathcal{Q}_{t|t}^{(i,j)} \tag{50}$$

Kalman filter

Prediction step:

$$\mathbf{X}_{t|t-1}^{(i)} = \boldsymbol{\mu}^{\mathbb{P}} + \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{X}_{t-1|t-1}^{(i)} \tag{51}$$

$$\mathbf{P}_{t|t-1}^{\mathbf{X}(i)} = \boldsymbol{\Phi}^{\mathbb{P}}\mathbf{P}_{t-1|t-1}^{\mathbf{X}(i)}\boldsymbol{\Phi}^{\mathbb{P}\prime} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \tag{52}$$

and update step:

$$\mathbf{X}_{t|t}^{(i,j)} = \mathbf{X}_{t|t-1}^{(i)} + \mathbf{K}_t^{(i,j)}(\mathbf{F}_t - \mathbf{A}(j) - \mathbf{B}'(j)\mathbf{X}_{t|t-1}^{(i)}) \tag{53}$$

$$\mathbf{P}_{t|t}^{\mathbf{X}(i,j)} = (\mathbf{I}_k - \mathbf{K}_t^{(i,j)}\mathbf{B}'(j))\mathbf{P}_{t|t-1}^{\mathbf{X}(i)} \tag{54}$$

with Kalman gain:

$$\mathbf{K}_t^{(i,j)} \equiv \mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\mathbf{B}(j)(\mathbf{B}'(j)\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\mathbf{B}(j) + \boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1} \tag{55}$$

and finally with the following approximations:

$$\mathbf{X}_{t|t}^{(j)} = (\mathcal{Q}_{t|t}^{(j)})^{-1} \sum_{i \in \mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)} \mathbf{X}_{t|t}^{(i,j)} \tag{56}$$

$$\mathbf{X}_{t|t} = \sum_{j \in \mathcal{S}} \mathcal{Q}_{t|t}^{(j)} \mathbf{X}_{t|t}^{(j)} \tag{57}$$

$$\mathbf{P}_{t|t}^{\mathbf{X}(j)} = (\mathcal{Q}_{t|t}^{(j)})^{-1} \sum_{i \in \mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)} (\mathbf{P}_{t|t}^{\mathbf{X}(i,j)} + (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})(\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})') \tag{58}$$

$$\mathbf{P}_{t|t}^{\mathbf{X}} = \sum_{j \in \mathcal{S}} \mathcal{Q}_{t|t}^{(j)} \mathbf{P}_{t|t}^{\mathbf{X}(j)} \tag{59}$$

for $t = 1, ..., T$ *and* $i, j \in \mathcal{S} = \{NB, LB\}$*.*

Proof. See derivations in appendix X.B based on Kim (1994), Hamilton (1989), Frühwirth-Schnatter (2006), Rahbek and Lange (2009). For a top-down view on the Kim filter derived in result IV.2, visit the flowchart presented in figure 3 on page 53. □

The Kim filter loops through $t = 1, ..., T$ but requires initialization. If the processes $\{\mathbf{X}_t\}_{t=1}^T$ and $\{s_t\}_{t=1}^T$ are stationary, $\mathbf{X}_{0|0}^{(i)}$, $\mathbf{P}_{0|0}^{\mathbf{X}(i)}$ and $\mathcal{Q}_{0|0}^{(i)}$ can be drawn from the stationary distribution, i.e., $\mathbf{X}_{0|0}^{(i)} = (\mathbf{I}_k - \mathbf{\Phi}^{\mathbb{P}})^{-1} \boldsymbol{\mu}^{\mathbb{P}}$, $\mathbf{P}_{0|0}^{\mathbf{X}(i)} = (\mathbf{I}_{k^2} - \mathbf{\Phi}^{\mathbb{P}} \otimes \mathbf{\Phi}^{\mathbb{P}})^{-1} \text{vec}(\mathbf{\Sigma}\mathbf{\Sigma}')$ and $\mathcal{Q}_{0|0}^{(i)}$ equals the invariant probability in $\bar{\boldsymbol{\pi}}$. Here, $\bar{\boldsymbol{\pi}}$ is given by $\bar{\boldsymbol{\pi}} = (\mathbb{P}(s_t = NB), \mathbb{P}(s_t = LB))'$ satisfying $(1\ 1)\bar{\boldsymbol{\pi}} = 1$ and $\bar{\boldsymbol{\pi}} = \mathbf{\Pi}\bar{\boldsymbol{\pi}}$. This is also suggested in Kim (1994). Alternatively, one could utilize uninformative values, e.g., by setting $\mathbf{X}_{0|0}^{(i)} = \mathbf{0}$, $\mathbf{P}_{0|0}^{\mathbf{X}(i)}$ sufficiently large and $\mathcal{Q}_{0|0}^{(i)} = 0.95$. In application, several starting values are selected to assess the importance.

From the Kim filter, the regime-specific conditional moments of the forward rates can be directly obtained by:

$$\mathbf{F}_{t|t-1}^{(i,j)} = \mathbf{A}(j) + \mathbf{B}'(j)\mathbf{X}_{t|t-1}^{(i)}$$
$$\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)} = \mathbf{B}'(j)\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\mathbf{B}(j) + \mathbf{\Omega}\mathbf{\Omega}'$$

for $t = 1, ..., T$ and $i, j \in \mathcal{S} = \{NB, LB\}$. The regime-specific conditional moments of the forward rates depend on the predicted state variables and regimes both approximated in the Kim filter. Given the regime-specific conditional moments of the forward rates, the log-likelihood function can be derived or approximated. The parameters in the RSTSM are given by $\boldsymbol{\theta} = (\boldsymbol{\mu}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{P}}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \delta_0^{\text{NB}}, \delta_0^{\text{LB}}, \boldsymbol{\delta}_1^{\text{NB}'}, \boldsymbol{\delta}_1^{\text{LB}'}, \pi_{\text{NB,NB}}, \pi_{\text{LB,LB}})'$, see definition IV.2 in section IV.A. The probability density function of the process $\{\mathbf{F}_t|\mathbf{F}_{t-1}, ..., \mathbf{F}_1\}_{t=2}^T$ is given by:

$$f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}) = \sum_{i,j\in\mathcal{S}} f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t, s_t = j, s_{t-1} = i|\mathbf{F}_{1:t-1}) = \sum_{i,j\in\mathcal{S}} f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)\,\mathcal{Q}_{t|t-1}^{(i,j)}$$

where law of total probability is used and $\mathcal{Q}_{t|t-1}^{(i,j)} = \mathbb{P}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:t-1})$ by definition. Here, it is important to stress that the process $\{\mathbf{F}_t|\mathbf{F}_{t-1}, ..., \mathbf{F}_1\}_{t=2}^T$ is non-Gaussian distributed, but the process conditional on the regimes $\{\mathbf{F}_t|\mathbf{F}_{t-1}, ..., \mathbf{F}_1, s_t = j, s_{t-1} = i\}_{t=2}^T$ is Gaussian distributed with conditional mean and variance $\mathbf{F}_{t|t-1}^{(i,j)}$ and $\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)}$. Hence, it is straightforward to show that the regime-specific probability density function is given by:

$$f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i) = (2\pi)^{-\frac{p}{2}}|\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)}|^{-\frac{1}{2}} \times \exp\{-\frac{1}{2}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)})'(\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)})^{-1}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)})\}$$

Here, there will be exactly 4 regime-specific probability density functions for each i and j in $\mathcal{S} = \{NB, LB\}$. The resulting log-likelihood function is given up to constants by:

$$\mathcal{L}_t(\boldsymbol{\theta}) = \sum_{t=2}^{T} \log\Big(\sum_{i,j\in\mathcal{S}} |\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)})'(\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)})^{-1}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)})\}\,\mathcal{Q}_{t|t-1}^{(i,j)}\Big) \quad (60)$$

for $t = 1, ..., T$ and i, j in $\mathcal{S} = \{NB, LB\}$. This log-likelihood function is used to estimate the parameters in the RSTSM. The maximum likelihood estimator maximizes this function with respect to the parameters in the RSTSM. Under certain regularity conditions, the estimator is consistent and asymptotically Gaussian distributed. The regularity conditions include that the errors $\{(\boldsymbol{\epsilon}_t^{\mathbb{P}}, \boldsymbol{\eta}_t)'\}_{t=1}^T$ are martingale differences with finite fourth order moments, the forward rates $\{\mathbf{F}_t\}_{t=1}^T$ and the regimes $\{s_t\}_{t=1}^T$ are stationary and weakly mixing, see Shumway and Stoffer (2000), Hamilton and Raj (2002) and Frühwirth-Schnatter (2006). The process $\{\mathbf{X}_t\}_{t=1}^T$ is stationary provided that the eigenvalues of $\boldsymbol{\Phi}^{\mathbb{P}}$ are less than one in absolute value, see, e.g., Shumway and Stoffer (2000). The process $\{s_t\}_{t=1}^T$ is stationary and weakly mixing, if the underlying Markov chain is irreducible and aperiodic, see, e.g., Hamilton (1994). The

Markov chain is irreducible, if there is no absorbing regime such that likelihood of staying in each regime is less than one, i.e., $\pi_{\text{NB,NB}} < 1$ and $\pi_{\text{LB,LB}} < 1$. Given irreducibility, the Markov chain is aperiodic if condition $\pi_{\text{NB,NB}} + \pi_{\text{LB,LB}} > 0$ is met, see Hamilton (1994).

Smoothed inference

The Kim filter presented in result IV.2 produces filtered values of the state variables and regimes. The filtered values are based on the current and the past observed forward rates. The regimes have a somewhat structural interpretation as they classify the monetary policy regimes. Therefore, it might be useful to conduct smoothed inference on the unobserved regimes based on all the observed forward rates.

Once the Kim filter and maximum likelihood estimation are executed, smoothed inference on the regimes can be obtained. Kim and Nelson (1999) and Hamilton (1994) propose a backward-algorithm to obtain these values. The smoothed values are calculated by backward iteration for time $t = T-1, ..., 1$. The algorithm is initialized by the final iteration in the Kim filter. The exact smoothing procedure is as follows.

Result IV.3. *The smoothed values of the regimes in the RSTSM with initial condition $\mathcal{Q}^{(j)}_{T|T}$ from the Kim filter in result IV.2 are given by:*

$$\mathcal{Q}^{(i)}_{t|T} \approx \mathcal{Q}^{(i)}_{t|t} \sum_{j \in \mathcal{S}} \left\{ \frac{\pi_{i,j}\, \mathcal{Q}^{(j)}_{t+1|T}}{\mathcal{Q}^{(j)}_{t+1|t}} \right\} \tag{61}$$

for $t = T-1, ..., 1$ and $i, j \in \mathcal{S} = \{NB, LB\}$.

Proof. See the brief derivation in appendix X.B.1 □

Here, $\mathcal{Q}^{(j)}_{t+1|t}$ is the predicted values of the regimes, $\mathcal{Q}^{(i)}_{t|t}$ is filtered values of the regimes and $\pi_{i,j}$ is the transition probability. All these quantities are obtained by maximum likelihood estimation combined with the Kim filter. The smoothed values of the regimes are computed by iterating backward in equation (61).

Figure 3. Flowchart of the Kim filter in the RSTSM

STEP 1: INITIALIZATION

$$\mathbf{X}_{0|0}^{(i)}, \quad \mathbf{P}_{0|0}^{\mathbf{X}(i)}, \quad \mathcal{Q}_{0|0}^{(i)} \quad \text{for } i,j \in \mathcal{S} = \{NB, LB\}$$

$t = 1$

STEP 2: PREDICTION

<u>KALMAN FILTER: STATE VARIABLES</u>

$$\mathbf{X}_{t|t-1}^{(i)}, \quad \mathbf{P}_{t|t-1}^{\mathbf{X}(i)}, \quad \mathbf{K}_{t}^{(i,j)}$$

<u>HAMILTON FILTER: REGIMES</u>

$$\mathcal{Q}_{t|t-1}^{(i,j)}$$

$$\Downarrow$$

<u>FORWARD RATES</u>

$$\mathbf{F}_{t|t-1}^{(i,j)}, \quad \mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)}$$

$$f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})$$

$t := t + 1$

STEP 3: UPDATE

<u>KALMAN FILTER: STATE VARIABLES</u>

$$\mathbf{X}_{t|t}^{(i,j)}, \quad \mathbf{P}_{t|t}^{\mathbf{X}(i,j)} \quad \overset{\text{approx.}}{\Rightarrow} \quad \mathbf{X}_{t|t}^{(j)}, \quad \mathbf{P}_{t|t}^{\mathbf{X}(j)}$$

<u>HAMILTON FILTER: REGIMES</u>

$$\mathcal{Q}_{t|t}^{(i,j)}, \quad \mathcal{Q}_{t|t}^{(j)}$$

for $t = T$

STEP 4: MAXIMUM LIKELIHOOD ESTIMATION

$$\mathcal{L}_T(\boldsymbol{\theta}) = \sum_{t=2}^{T} \log f_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})$$

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmax}}\ \mathcal{L}_T(\boldsymbol{\theta})$$

Note: See result IV.2 and appendix X.B.

V. Empirical analysis

This section conducts an empirical analysis of the Gaussian affine term structure model (hereafter, "GATSM") and the regime-switching term structure model (hereafter, "RSTSM"). Both term structure models were presented in section III.B-III.C and the estimation techniques were introduced in section IV.C-IV.D. The empirical analysis is based on one-month euro area forward rates on zero-coupon bonds from January 2005 to March 2020 collected from the European Central Bank (hereafter, "ECB").

This section begins with a description of the euro area forward rates followed by a brief outline of the implementation in MATLAB. Subsequently, the estimation results are presented, and the main findings for the GATSM and RSTSM are compared and interpreted. Several misspecification tests are performed to validate the underlying assumptions in the models. Finally, the findings in the RSTSM are related to the ECB's actual monetary policy.

A. Euro area forward rates

This section presents descriptive statistics of the euro area forward rates used in the empirical analysis of the GATSM and RSTSM. The forward rates are constructed from spot prices on AAA-rated euro area central government zero-coupon bonds collected from the ECB, see ECB (2020b). The forward rates are in annualized percentage at the monthly frequency with end-of-month observations from January 2005 to March 2020. The selected maturities are 1, 2, 5, 7 and 10 years as well as 3 and 6 months to capture the short end of the term structure. These maturities are similar to those considered in Wu and Xia (2016).

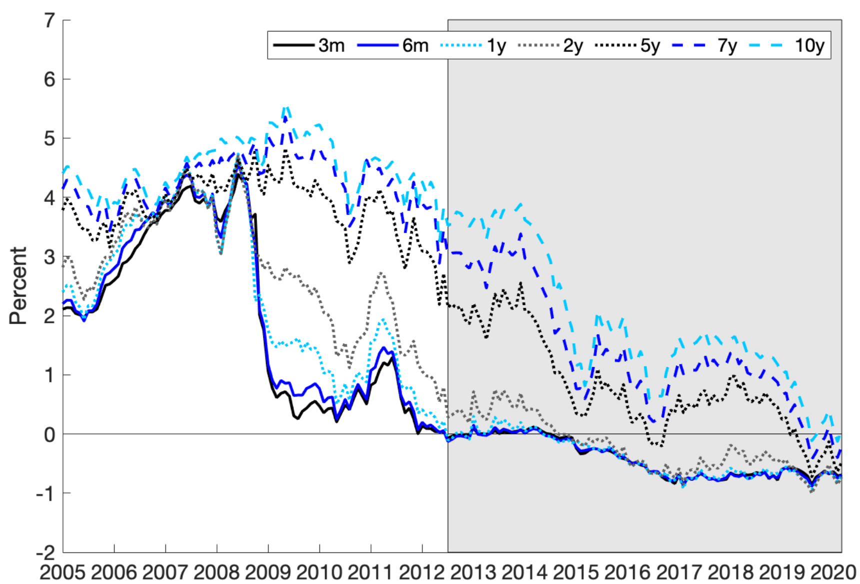
One-month euro area forward rates are presented in figure 4. In figure 4, forward rates with maturities up to 2 years take a sharp drop at the global financial crisis in 2007-08. In this period, the ECB cut the official interest rates to historically low levels.[xvii] In the wake of the crisis, longer-term forward rates exhibit a clear downward trend, while short-term forward rates have evolved at persistently low levels during the lower bound period. The

[xvii] Figure 2 in section II.C shows the ECB's reaction to the global financial crisis in 2007-08.

lower bound period corresponds to the shaded area in figure 4. Between 2014 and 2015, the ECB's launch of unconventional monetary policy is associated with a steep decline in longer-term forward rates. Since January 2020, almost all forward rates are in the negative territory.

Figure 4 shows that forward rates with relatively close maturities tend to move in tandem over time, indicating a factor structure. In particular, a short-, medium- and long-term factor may capture a significant part of the variation in the forward rates, which motivates the fact that few factors can model the term structure, see Ang et al. (2008) and Wu and Xia (2016). The shaded area in figure 4, i.e, the lower bound period, suggests that shorter-term forward rates evolve in so-called regimes. The regimes seem to be closely related to the ECB's monetary policy, i.e., conventional monetary policy in the *no* lower bound regime and unconventional monetary policy in the lower bound regime. This observation substantiates the importance of including regimes in the term structure model.

Figure 4. Euro area forward rates from January 2005 to March 2020



Note: 1-month forward rates with maturities in 3 and 6 months, 1, 2, 5, 7, 10 years covering the period from January 2005 to March 2020. The shaded area represents the lower bound period from July 2012.

Table I reports descriptive statistics of the forward rates in annualized percentage. Few stylized facts stand out. The average forward curve is upward sloping. Apart from short-term forward rates, the standard deviation or the volatility is decreasing in maturity. The low volatilities associated with short-term forward rates are seriously affected by the length of the lower bound period in the sample, which results in somewhat lower volatilities. The longer-term forward rates are left-skewed, while shorter-term forward rates are right-skewed compared to a Gaussian distribution. All forward rates have less kurtosis than a Gaussian distribution, indicating less extreme realizations than those from a Gaussian distribution. Altogether, these observations speak against that the forward rates are unconditionally Gaussian distributed and bring into question whether the forward rates are conditionally Gaussian distributed. These early-stage signs of misspecification will be considered in section V.B.3 when several misspecification tests are carried out.

Table I Descriptive statistics of euro area forward rates

	3M	**6M**	**1Y**	**2Y**	**5Y**	**7Y**	**10Y**
Mean	0.7579	0.8133	0.9381	1.2455	2.3006	2.7993	3.1606
Std. dev.	1.6005	1.6385	1.6656	1.6842	1.6737	1.6473	1.5968
Skewness	1.0160	0.9603	0.7768	0.3957	-0.1772	-0.3481	-0.4601
Kurtosis	2.6116	2.5241	2.2581	1.7032	1.4922	1.6436	1.7759
Min.	-0.8613	-0.8864	-0.9423	-1.0018	-0.7238	-0.4295	-0.1285
Max.	4.3889	4.5713	4.7352	4.7140	4.8283	5.3602	5.6114

Note: 1-month forward rates with maturities in 3 and 6 months, 1, 2, 5, 7, 10 years covering the period from January 2005 to March 2020 collected from the ECB, see ECB (2020b).

Table II presents different levels of autocorrelation in the forward rates. Specifically, the autocorrelation is for lag 1, 2, 3, 6 and 12 months. All the forward rates are highly persistent by the 1- and 12-month autocorrelation. The autocorrelation in the forward rates is slightly decreasing in maturity, indicating that shorter-term forward rates are more persistent than longer-term forward rates. Highly persistent forward rates may imply a persistent factor structure, which can introduce technical challenges. In general, the length of the lower bound period in the sample may increase the forward rates' overall persistence.

Table II Autocorrelation of euro area forward rates

	3M	**6M**	**1Y**	**2Y**	**5Y**	**7Y**	**10Y**
Lag 1	0.9894	0.9900	0.9876	0.9864	0.9809	0.9792	0.9777
Lag 2	0.9765	0.9762	0.9721	0.9711	0.9622	0.9579	0.9545
Lag 3	0.9585	0.9588	0.9555	0.9554	0.9460	0.9395	0.9337
Lag 6	0.8977	0.9014	0.9074	0.9153	0.8990	0.8847	0.8716
Lag 12	0.7706	0.7907	0.8223	0.8459	0.8162	0.7907	0.7695

Note: 1-month forward rates with maturities in 3 and 6 months, 1, 2, 5, 7, 10 years covering the period from January 2005 to March 2020 collected from the ECB, see ECB (2020b).

Finally, table III presents the correlation between euro area forward rates with different maturities. In general, forward rates with relatively close maturities tend to move in tandem over time. The high correlation between forward rates with different maturities motivates the fact that the forward rates exhibit a strong factor structure. The correlation is decreasing in maturity, indicating that factors possibly depend on different maturities.

Table III Correlation between euro area forward rates

	3M	**6M**	**1Y**	**2Y**	**5Y**	**7Y**	**10Y**
3M	1						
6M	0.9977	1					
1Y	0.9814	0.9913	1				
2Y	0.9346	0.9504	0.9796	1			
5Y	0.7828	0.8014	0.8523	0.9339	1		
7Y	0.7164	0.7349	0.7897	0.8845	0.9910	1	
10Y	0.6797	0.6972	0.7523	0.8501	0.9736	0.9943	1

Note: 1-month forward rates with maturities in 3 and 6 months, 1, 2, 5, 7, 10 years covering the period from January 2005 to March 2020 collected from the ECB, see ECB (2020b).

B. Estimation

This section presents the main findings of the GATSM and RSTSM based on the euro forward rates introduced in section V.A. First, this section briefly describes the implementation of the GATSM and RSTSM in MATLAB. Subsequently, the estimation results for the GATSM and RSTSM are presented and interpreted. Finally, several misspecification tests are reported to assess the underlying assumptions in the GATSM and RSTSM.

B.1. Implementation in MATLAB

The GATSM, as well as the RSTSM, are estimated by using MATLAB. The econometric methodology was provided in section IV. The parameters in the GATSM and RSTSM are estimated by maximum likelihood estimation. In the GATSM, the unobserved state variables are obtained by applying the Kalman filter. In the RSTSM, the unobserved state variables and regimes are computed by using the Kim filter.

As the euro area forward rates presented in section V.A are given in annualized percentage, the parameters of the state variables that enter the no-arbitrage recursions are scaled by $12 \cdot 100 = 1200$ in implementation of the GATSM and RSTSM in MATLAB. The no-arbitrage recursions are defined for forward rates in monthly percentage, so the parameters' scaling improves the convergence speed significantly, see also Wu and Xia (2016).

In MATLAB, all codes have been written from scratch as no existing solutions or packages were available.[xviii] Several checks for no coding mistakes have been conducted. For example, forward rates have been simulated from the models, and the resulting optimizations lead to the data-generating values. Furthermore, the results in Wu and Xia (2016) were replicated for the GATSM to ensure the robustness of the code.[xix] The GATSM and RSTSM were estimated using a rolling window.

[xviii] The codes can be executed without specific packages. All codes are written by myself.

[xix] The data in Wu and Xia (2016) were provided on their homepage: `https://sites.google.com/view/jingcynthiawu/research`. Estimation results for the GATSM were replicated under the identification scheme proposed in Wu and Xia (2016) and Joslin et al. (2010).

Estimation of the GATSM and RSTSM requires initialization. In the implementation, the starting values for the parameters, state variables and regimes turned out to have a significant impact on the optimal estimates. Different starting values lead to different estimates, while poor starting values never reached convergence. These numerical issues are well-known in the term structure literature, see e.g., Ang et al. (2008) and Hamilton and Wu (2012). Several strategies have been considered to circumvent numerical issues. A massive swarm of starting values was used to increase convergence. However, the speed of convergence was *very* slow, and it was not clear whether parameters have visited the whole parameter space during the iterations. In MATLAB, different solvers have been used; unrestricted and restricted gradient-based methods and gradient-free methods.[xx] The results were robust towards different solvers. Local convergence was achieved, and the highest likelihood-value was chosen. There is no guarantee that the local maximum of the likelihood function is also a *global* maximum. This should be kept in mind throughout the analysis.

All MATLAB files are provided in appendix XI.G.

B.2. Main findings

This section presents the estimation results of the GATSM and RSTSM based on the euro area forward rates introduced in section V.A. First, parameter estimates of the GATSM and RSTSM are reported. The regimes in the RSTSM are examined, followed by an analysis of the factors, including an interpretation of how they load into the forward curve. Finally, the average forward curves for the GATSM and RSTSM are presented and compared.

Table IV on page 62 reports maximum likelihood estimates with robust standard errors for the GATSM and RSTSM. The log-likelihood value in the GATSM and RSTSM is 1383.8 and 1482.3, respectively. This difference may reflect the additional flexibility of fitting the forward rates in the RSTSM compared to the GATSM.

[xx]In MATLAB, the solvers are `FMINUNC`, `FMINCON` and `FMINSEARCH`. `FMINUNC` and `FMINCON` use the BFGS-algorithm. `FMINSEARCH` uses the NM-algorithm.

In table IV, the estimated parameters of the GATSM and RSTSM are quite similar, except for the estimated parameters of the short rate, i.e., $\delta_0, \boldsymbol{\delta_1'}$ and $\delta_0^{\text{NB}}, \delta_0^{\text{LB}}, \boldsymbol{\delta_1'}^{\text{NB}}, \boldsymbol{\delta_1'}^{\text{LB}}$. The magnitude of the short rate parameters is not directly comparable due to the difference in the estimated values for the state variables or factors in the GATSM and RSTSM.

In table IV, the estimated risk-neutral and physical parameters of the factors, i.e., $\boldsymbol{\Phi}^{\mathbb{P}}$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$, indicate highly persistent dynamics for both the GATSM and RSTSM. This can be seen from the estimated eigenvalues in $\boldsymbol{\Phi}^{\mathbb{P}}$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$. The eigenvalues are close to one. The eigenvalue associated with the third factor is above one, indicating that the third factor is slightly explosive under the physical measure. Explosive state dynamics are typically found in estimation of the GATSM, see, e.g., Dai et al. (2007). Consequently, the standard errors are not consistently estimated as the asymptotic distribution of the estimator does not necessarily hold in the GATSM and RSTSM.

The high degree of persistence and even explosive behavior in the factors are somewhat constrained by economic theory. The economic theory dictates that forward rates on nominal zero-coupon bounds should not contain unit roots, since forward rates are restricted by the lower bound. Unit roots consist of random walk components that will cross the lower bound eventually. Therefore, the estimated parameters of the factors, i.e., $\boldsymbol{\Phi}^{\mathbb{P}}$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$ in table IV, are not in accordance with the theoretical perception of forward rates. The factors or state variables will be examined further when they are visualized subsequently.

Table IV further reports the estimated regime probabilities in the RSTSM, i.e., $\pi_{\text{NB, NB}}$ and $\pi_{\text{LB, LB}}$. "NB" and "LB" are abbreviations for the regimes; *no* lower bound regime (NB) and lower bound regime (LB). Based on the estimated values in table IV, the estimated transitional probability matrix is given by:

$$\boldsymbol{\Pi} = \begin{pmatrix} 0.9630 & 0.0065 \\ 0.0370 & 0.9935 \end{pmatrix}$$

In words, the likelihood of staying in the *no* lower bound regime is 0.9630, while the likelihood of staying in the lower bound regime is 0.9935. Therefore, the likelihood of escaping the lower bound regime is 0.0065. Note, this is not the unconditional likelihood of observing the *no* lower bound regime; rather, it is the conditional likelihood of observing the *no* lower bound regime given that the previous lower bound regime. Therefore, the regimes are highly persistent as expected; the sample contains solely one unambiguously shift in monetary policy regimes. Highly persistent regimes are somehow in accordance with monetary policy. Monetary policy is primarily based on price movements, which are quite persistent, see Ang et al. (2008). The slight difference in the estimated values of the regime probabilities may be due to the regimes' length in the sample, respectively.

Table IV Maximum likelihood estimates with robust standard errors

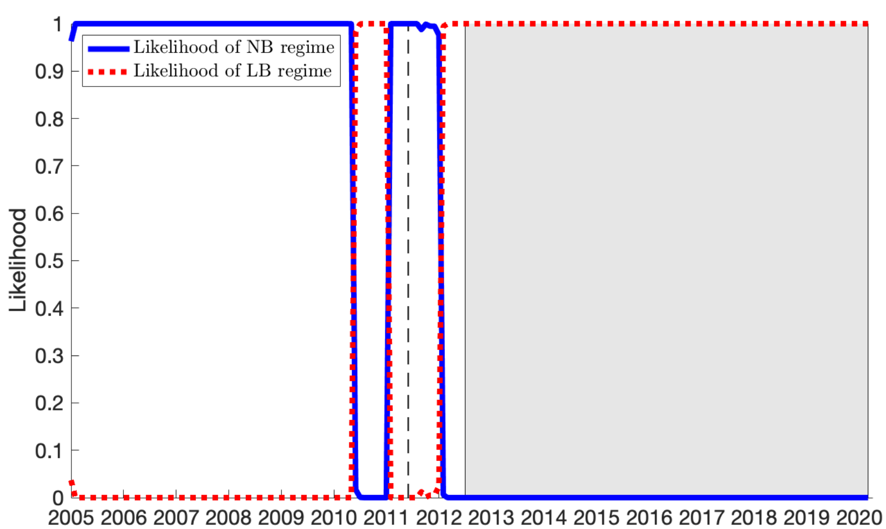
	GATSM			RSTSM		
	.9907 [.0000]	0	0	.9909 [.0197]	0	0
$\mathbf{\Phi}^{\mathbb{P}}$	.0129 [.0139]	.9574 [.0107]	0	.0168 [.0098]	.9831 [.0000]	0
	.0184 [.0094]	.0006 [.0160]	1.0028 [.0007]	-.0004 [.0573]	-.0052 [.0530]	1.0065 [.0204]
$\boldsymbol{\mu}^{\mathbb{Q}\prime}$	-.0075 [.0017]	.2066 [.0106]	.4187 [.0637]	-.0022 [.0080]	.2566 [.0876]	.1592 [.6782]
	.9990 [.0001]	0	0	.9979 [.0195]	0	0
$\mathbf{\Phi}^{\mathbb{Q}}$	.0127 [.0037]	.9717 [.0002]	0	-.0327 [.0003]	.9685 [.0024]	0
	-.0119 [.0182]	-.0909 [.0106]	.9725 [.0001]	-.0095 [.0240]	-.1121 [.0257]	.9693 [.0025]
δ_0		1.8540 [.1081]			—	
$\boldsymbol{\delta}_1'$	.0237 [.0024]	-.0036 [.0038]	-.1305 [.0046]	—	—	—
δ_0^{NB}		—			1.8251 [.0000]	
δ_0^{LB}		—			.8512 [.8862]	
$\boldsymbol{\delta}_1'^{\,\mathrm{NB}}$	—	—	—	-.0798 [.0194]	-.0345 [.0276]	-.2234 [.0000]
$\boldsymbol{\delta}_1'^{\,\mathrm{LB}}$	—	—	—	-.0042 [.0330]	.0554 [.0000]	-.0431 [.0000]
$\pi_{\mathrm{NB,\,NB}}$		—			.9630 [.0000]	
$\pi_{\mathrm{LB,\,LB}}$		—			.9935 [.0001]	
Log-likelihood		1383.8			1482.3	
Sample size		183			183	

Note: Maximum likelihood estimates for the GATSM and RSTSM with robust standard errors in brackets [·]. Sample from January 2005 to March 2020. The estimated values of the "pricing errors" in the forward rates are reported in table VI in appendix XI.A.

Figure 5 presents the smoothed regime probabilities in the RSTSM. The smoothed values are calculated by the backward smoothing algorithm introduced in section IV.D and derived in appendix X.B.1. In figure 5, the shaded area corresponds to the actual lower bound period from July 2012. The solid blue line indicates the likelihood of observing the *no* lower bound regime, while the red dotted line indicates the likelihood of observing the lower bound regime. Both probabilities are extracted from the entire history of the forward rates.

In figure 5, the smoothed regime probabilities deliver clear-cut predictions for the lower bound environment. This is especially evident from 2012. The vertical dashed line in June 2011 indicates the ECB's reaction to above-target inflation in the euro area.[xxi] The rise in inflation was partly driven by tensions in sovereign debt markets and political turmoil in the oil markets, see ECB (2020d). The smoothed regime probabilities in figure 5 seem to capture this extraordinary event quite well by indicating a shift from a lower bound regime to a *no* lower bound regime around 2011-12.[xxii]

Figure 5. Smoothed values of the regime probabilities in the RSTSM



Note: "NB" and "LB" are abbreviations for *no* lower bound regime and lower bound regime. The vertical dashed line represents the ECB's reaction to the rise in inflation, see ECB (2020d).

[xxi]The ECB increased the official interest rates, see, e.g., figure 2 in section II.C.
[xxii]Section VI.C discusses the limited regime shifts in the sample.

Figure 6 on page 65 presents the filtered values of the state variables or factors in the RSTSM. The filtered values of the factors in the GATSM are presented in figure 11 in appendix XI.C. The upper left panel in figure 6 visualizes the filtered values of the three factors in the RSTSM. All factors are highly persistent and resemble the reported results in table IV. The degree of persistence can be summarized by the eigenvalues of $\mathbf{\Phi}^{\mathbb{P}}$. As $\mathbf{\Phi}^{\mathbb{P}}$ is a lower triangular matrix, the eigenvalues are given by:

$$\text{eig}(\mathbf{\Phi}^{\mathbb{P}}) = (0.9909, 0.9831, 1.0065)'$$

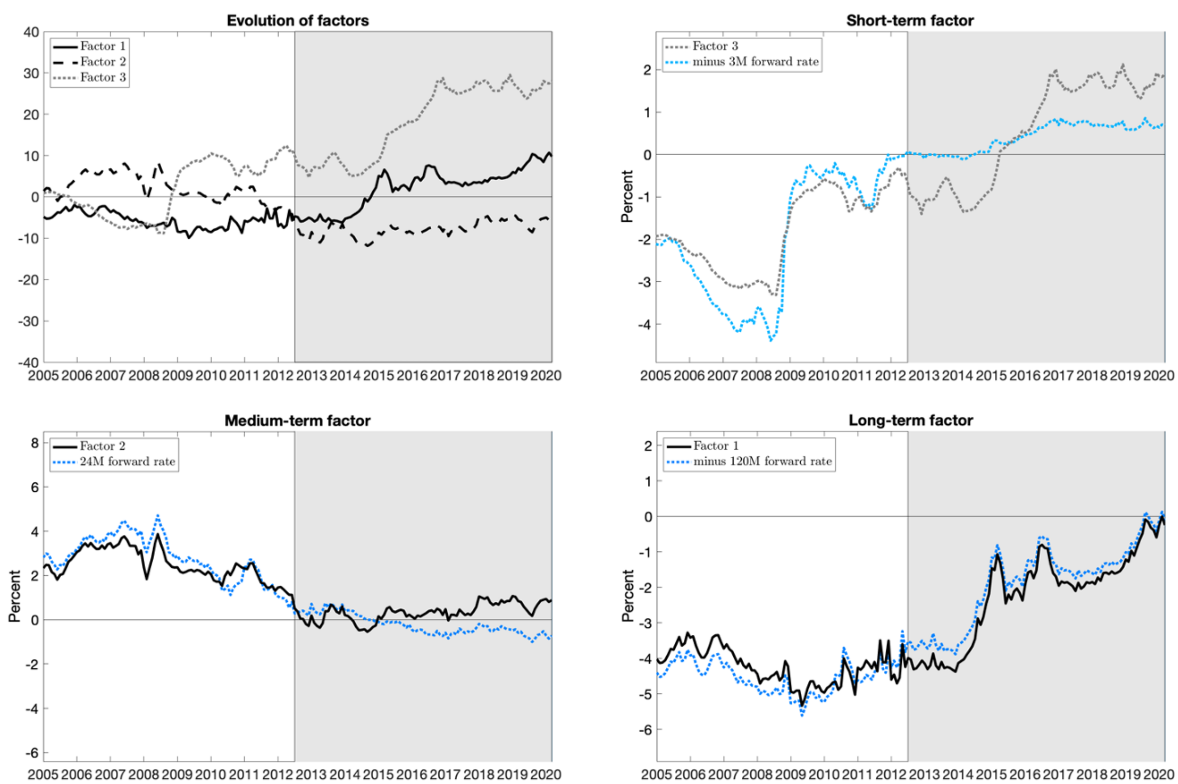
The eigenvalues associated with the two first factors are slightly below 1, indicating highly persistent mean-revering behavior. This is also evident from the upper left panel in figure 6, where the first factor appears to be relatively more persistent compared to the second factor. Importantly, the eigenvalue associated with the third factor is larger 1, indicating slightly explosive behavior. In the upper left panel in figure 6, the third factor differs from the other factors as the evolution does not exhibit mean-reverting behavior. Highly persistent or even slightly explosive behavior in the factors is a relatively common finding in the estimation of affine term structure models, see e.g., Dai et al. (2007) and Ang et al. (2008). As the factors are unobserved and filtered from observed forward rates, the factors reflect the forward rates' evolution. In section V.A, the euro area forward rates were highly persistent and exhibited somewhat structural shifts according to the ECB's monetary policy actions. Therefore, the process for the factors should reflect the same, e.g., by including regime shifts in $\mathbf{\Phi}^{\mathbb{P}}$. This may be seen as a natural extension of the RSTSM, which may handle the high degree of persistence and even explosive behavior.[xxiii]

As the factors or state variables are unobserved, they are statistically constructed and do not have an economic interpretation. However, the three factors typically reflect some stylized characteristics of the forward rates. In the upper-right panel in figure 6, the the third factor in the RSTSM is displayed against the negative 3-month forward rate. The third factor is scaled to match both the mean and range of the negative 3-month forward rate. The third

[xxiii]Section VI.C discusses further model extensions.

factor exhibits a linear correlation of 0.89 with the negative 3-month forward rate. Henceforth, the third factor will be referred to as a *short-term factor* as the third factor is mostly related to shorter-term forward rates. In the lower-left panel in figure 6, the second factor is plotted against the 24-month or 2-year forward rate. The close evolution of the second factor and the 24-month forward rate suggests that the second factor is a *medium-term factor*. Finally, the lower-right panel in figure 6 presents the first factor against the negative 120-month or 10-year forward rate. The first factor exhibits a strong linear correlation of 0.97 with the negative 120-month forward rate. Therefore, the first factor will be referred to as a *long-term factor*. The RSTSM and GATSM were estimated based on three factors. In figure 12 in appendix XI.D, the fitted forward rates in the GATSM and RSTSM are plotted against the actual euro area forward rates. All the forward rates in the GATSM and RSTSM are fitted quite well compared to the actual forward rates, indicating that three factors are sufficient to explain the variation in the euro area forward rates.

Figure 6. Filtered state variables or factors in the RSTSM

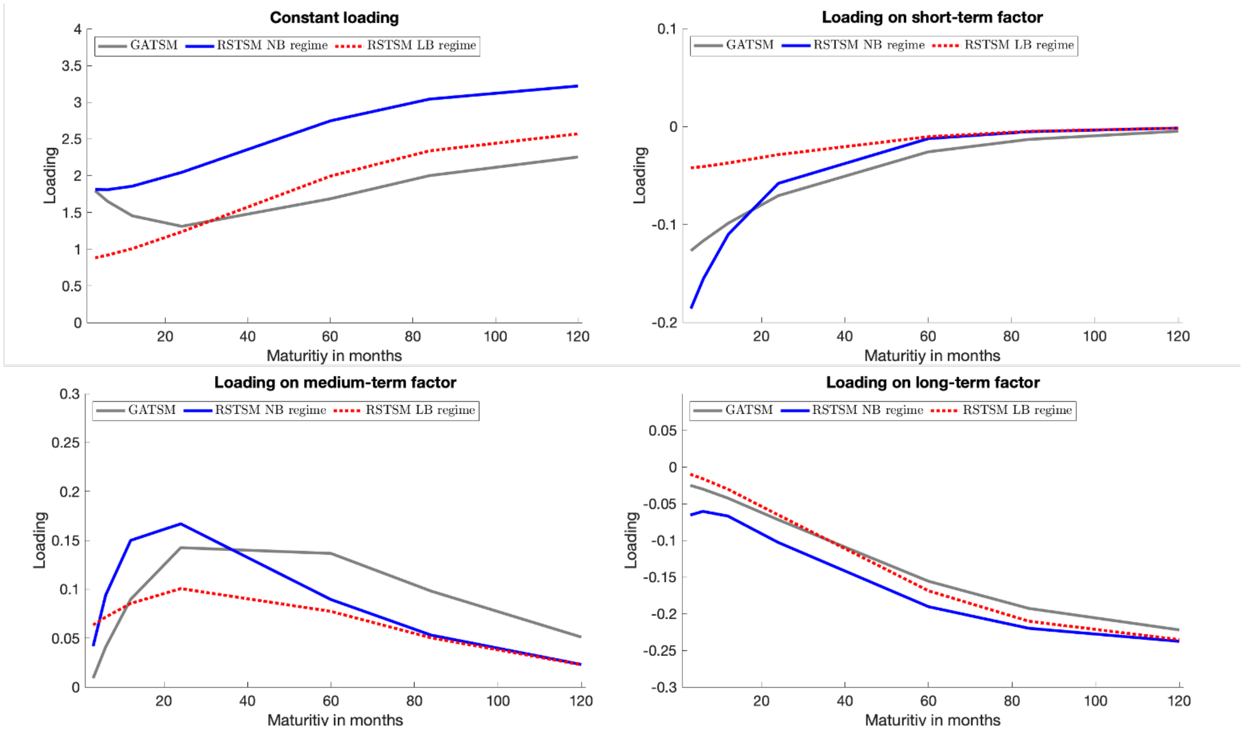


Note: Filtered values in the RSTSM are obtained by the Kim filter described in section IV.D. Short-, medium- and long-term factors are scaled to match the mean and range of the relevant forward rate.

Figure 7 presents the factor loadings in the GATSM and RSTSM. The factors loadings determine how the factors load into the term structure. The factor loadings correspond to $\mathbf{A}$ and $\mathbf{B}$ in equation (23)-(36) in section III.B-III.C. In the upper-left panel in figure 7, the estimated values of the constant loading are visualized for maturities from 1 to 120 months. The loading(s) on the constant(s) in the RSTSM is upward sloping with a relatively larger loading in the *no* lower bound regime. The constant loading in the GATSM and RSTSM differs, except for the constant loading on the short rate, i.e., the shortest maturity. In the upper-right panel in figure 7, the estimated loading on the short-term factor is displayed. In general, the loading on the short-term factor decreases in maturity, indicating a *slope* characteristic at the beginning of the term structure. The regime-dependence in the loading(s) on the short-term factor in the RSTSM is largest for short maturities and vanishes for maturities above 60 months or 5 years. In the lower-right panel in figure 7, the estimated loading on the medium-term factor is depicted. The loading on the medium-term factor is hump-shaped for both the GATSM and RSTSM, indicating a *curvature* characteristic with the largest loading in the middle of the term structure. The loading on the medium-term factor further contains *level* characteristic as the loading does not vanish completely at the start and end of the term structure. Finally, the lower-right panel in figure 7 exhibits the loading on the long-term factor. In general, the loading increases in maturity, indicating a *slope* characteristic at the end of the term structure.

The interpretation of the factor loadings is somewhat different from the interpretation in the standard principal component analysis. In the standard principal component analysis, the three factors typically constitute a level, slope and curvature, see, e.g., Diebold and Li (2003). Unlike the principal component approach, the factors in the GATSM and RSTSM are correlated by construction, since $\mathbf{\Phi}^{\mathbb{P}}$ is a lower triangular matrix which allows the factors to be correlated, see also Dai et al. (2007). Therefore, the factor loadings will not necessarily reflect level, slope and curvature characteristics.

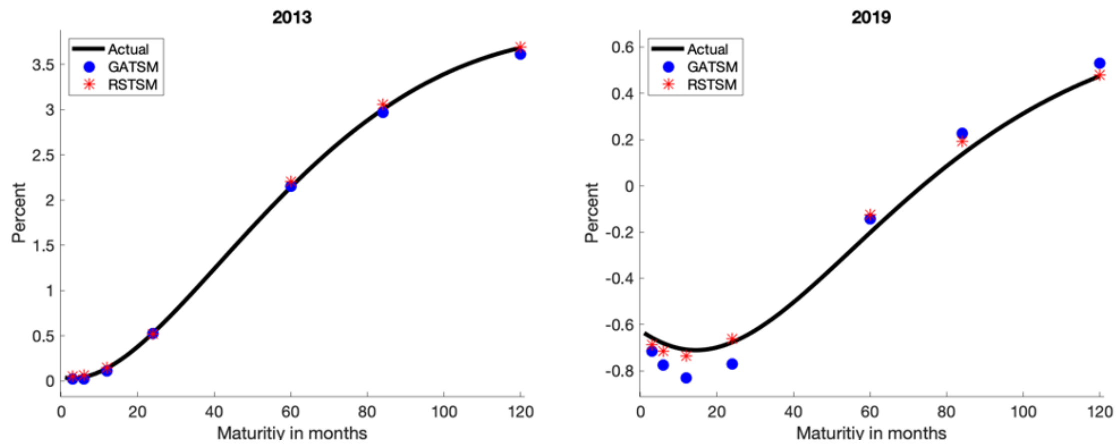
Figure 7. Estimated factor loadings in the GATSM and RSTSM



Note: "NB" and "LB" are abbreviations for *no* lower bound regime and lower bound regime. Estimated factor loadings in the GATSM and RSTSM corresponding to **A** and **B** in equation (23)-(36) in section III.B-III.C.

Figure 8 depicts average euro area forward curves against the average forward rates in the GATSM and RSTSM for 2013 and 2019. Both the GATSM and RSTSM fit the term structure of the forward rates quite well. The RSTSM exhibits superior performance at the short end of the average forward curve in 2019 compared to the GATSM. The superior performance compared to the GATSM may be due to the extra flexibility in RSTSM by including regime shifts in the process for the short rate. However, there is no such evidence for the average forward curve in 2013. Additional forward curves are presented in figure 13 in appendix XI.E. Based on these findings, the RSTSM seems to improve the cross-sectional fit of the forward rates, especially in the short end of the forward curve during the lower bound period.

Figure 8. Average euro area forward curves in 2013 and 2019



Note: Black curve: actual forward rates from ECB (2020b). Blue dots: fitted forward rates in the GATSM. Red dots: fitted forward rates in the RSTSM.

B.3. Misspecification

This section presents several misspecification tests to validate the underlying assumptions in the GATSM and RSTSM. Both the GATSM and RSTSM rely on the assumption of independent and identical distributed (*i.i.d.*) Gaussian errors. The assumption of *i.i.d.* Gaussian errors are especially used in the filtering methods presented in the sections IV.C-IV.D and further used to establish the asymptotic properties of the maximum likelihood estimator. Therefore, it is essential to validate the assumption of *i.i.d.* Gaussian errors in the GATSM and RSTSM. The misspecification tests will be based on the standardized residuals, i.e., the difference between the observed forward rates and the estimated forward rates scaled by the standard deviation. The misspecification tests include a test for Gaussian distributed residuals, a test for no autocorrelation in the residuals and finally a test for no autoregressive conditional heteroscedasticity (ARCH) effects in the residuals.

The Jarque-Bera test is used to examine whether the skewness and kurtosis of the standardized residuals are in accordance with those seen in a Gaussian distribution. Under the null hypothesis, i.e., the joint hypothesis that the skewness is 0 and the kurtosis is 3, the Jarque-Bera test is asymptotically χ^2-distributed with 2 degrees of freedom.

The test for autocorrelation in the standardized residuals is based on the Ljung-Box test. The null hypothesis is that there is no autocorrelation in the standardized residuals. Specifically, the number of lags included will be 10, so the Ljung-Box test is based on the overall autocorrelation in the first 10 lags of the standardized residuals. Under the null hypothesis, the Ljung-Box test is asymptotically χ^2-distributed with 10 degrees of freedom.

The so-called Lagrange multiplier test proposed in Engle (1982) is used to test for ARCH effects in the standardized residuals. As the test will be based on 5 lags, the null hypothesis is no ARCH effects in the first 5 lags of the standardized residuals. Under the null hypothesis, the Lagrange multiplier test is asymptotically χ^2-distributed with 5 degrees of freedom.

Table V presents the misspecification tests for both the GATSM and RSTSM. In general, table V reveals severe signs of misspecification in the GATSM and RSTSM. The Jarque-Bera test shows a clear sign of misspecification since the standardized residuals in the GATSM and RSTSM are not Gaussian distributed. The Jarque-Bera test-statistic relies on no autocorrelation and no ARCH effects in the standardized residuals. Therefore, the clear sign of misspecification may be influenced by other sources of misspecification. The Ljung-Box test shows that the GATSM and RSTSM suffer from autocorrelation in the standardized residuals. Autocorrelation in the standardized residuals implies that the parameters in the GATSM and RSTSM may be inconsistently estimated. Neither the GATSM nor the RSTSM passes the test for no ARCH effects in the standardized residuals. ARCH effects and non-Gaussian distributed residuals influence the asymptotic distribution of the estimator. Figure 14 and 15 in appendix XI.F present histograms, density- and QQ-plots of the standardized residuals. These figures primarily confirm that the distribution of the standardized residuals are skewed and with so-called fat tails compared to a standard Gaussian distribution. Based on the misspecification tests, the assumption of *i.i.d.* Gaussian errors is unlikely, and the empirical findings of the GATSM and RSTSM may be induced by errors. Several extensions of the GATSM and RSTSM will be proposed in section VI.C to circumvent the severe signs of misspecification.

Table V Misspecification tests

	GATSM			RSTSM		
	Gaussian	Autocorrelation	ARCH	Gaussian	Autocorrelation	ARCH
3M	61.60 [.001]	583.78 [.000]	75.59 [.000]	128.09 [.001]	454.06 [.000]	44.20 [.000]
6M	9.57 [.017]	781.53 [.000]	126.26 [.000]	11.18 [.012]	626.11 [.000]	102.47 [.000]
1Y	1.46 [.425]	710.92 [.000]	99.21 [.000]	63.02 [.001]	344.68 [.000]	16.87 [.005]
2Y	.019 [.500]	722.14 [.000]	105.73 [.000]	2050.09 [.001]	58.74 [.000]	.24 [.998]
5Y	138.8 [.001]	273.71 [.000]	7.49 [.189]	683.27 [.001]	70.99 [.000]	12.63 [.027]
7Y	20.89 [.002]	578.12 [.000]	81.75 [.000]	890.32 [.001]	105.40 [.000]	12.57 [.028]
10Y	5.79 [.048]	725.92 [.000]	100.15 [.000]	5.56 [.052]	718.05 [.000]	99.62 [.000]

Note: The misspecification tests are based on the standardized residuals. Test-statistics are reported with p-values in brackets $[\cdot]$.

In the RSTSM, the regimes was assumed to follow an unobserved 2-state Markov-switching variable s_t. Based on the maximum likelihood estimates reported in table IV, the estimated transition probability matrix is given by:

$$\mathbf{\Pi} = \begin{pmatrix} 0.9630 & 0.0065 \\ 0.0370 & 0.9935 \end{pmatrix}$$

Here, the likelihood of staying in the *no* lower bound regime (NB) is 0.9630 and the likelihood of staying in the lower bound regime (LB) is 0.9935. As the columns sum to 1, the likelihood of switching regime is given by the off-diagonal elements. According to the estimated values of the transition probability matrix, the regime variable s_t is irreducible. This can be verified by the condition $\pi_{\text{NB,NB}} = 0.9630 < 1$ and $\pi_{\text{LB,LB}} = 0.9935 < 1$. In other words, there is no absorbing regime as the likelihood of staying in either regime is less than 1. The regime variable s_t does further satisfy aperiodicity since $\pi_{\text{NB,NB}} + \pi_{\text{LB,LB}} = 0.9630 + 0.9935 > 0$. Together, Hamilton (1994) shows that the regime variable s_t is weakly mixing and geometric ergodic. Furthermore, Hamilton (1994) derives a closed-form solution of the unconditional regime probabilities. Based on formula in Hamilton (1994), the unconditional regime probabilities are given by:

$$\bar{\boldsymbol{\pi}} = \begin{pmatrix} 0.1496 \\ 0.8504 \end{pmatrix}$$

The unconditional regime probability of the *no* lower bound regime and the lower bound regime is 0.1496 and 0.8504, respectively. The unconditional regime probability guides the evolution of the regimes. Figure 10 in appendix XI.B visualizes the future evolution of the regimes conditional on the lower bound regime in March 2020. As in Hamilton (1994) and Rahbek and Pedersen (2019b), the 2-month ahead conditional regime probabilities are computed by $\mathbf{\Pi}^2$, where $\mathbf{\Pi}$ is based on the maximum likelihood estimates reported in table IV. This is repeated for 36 months or 3 years and the resulting conditional regime probabilities are plotted in figure 10 in appendix XI.B. The figure shows that the conditional regime probabilities approach the unconditional regime probabilities as the time horizon increases. In December 2021, the likelihood of staying in the lower bound regime is 0.9065, substantiating the lower bound regime is very persistent. Due to the limited regime shifts in the sample, the outcome may be somewhat arbitrary. Furthermore, the transition probability matrix may be inconsistently estimated, and therefore the descriptions provided here may be more interesting and informative than the outcome itself.

C. Findings in relation to monetary policy

This section relates the findings of the Gaussian affine term structure model (hereafter, "GATSM") and the regime-switching term structure model (hereafter, "RSTSM") to the European Central Bank's (hereafter, "ECB") monetary policy.

Historically, the ECB has used the Euro Overnight Index Average (hereafter, "EONIA") as a primary indicator of signaling the monetary policy stance. In the wake of the global financial crisis in 2007-08, the EONIA rate has been stuck at persistently low levels, not necessarily reflecting the ECB's actual monetary policy stance.

Figure 9 on page 74 presents the evolution of the EONIA rate and the estimated short rates in the GATSM and RSTSM from January 2005 to March 2020. In figure 9, the black line is the EONIA rate, the grey line is the estimated short rate in the GATSM, the blue line and red dotted line are the estimated short rates in the RSTSM in the *no* lower bound regime and in the lower bound regime, respectively.

During the lower bound period, see the shaded area in figure 9, the short rate in the *no* lower bound regime in the RSTSM represents the short rate if the lower bound did not exist. The EONIA rate and estimated short rate in the GATSM are stuck at persistently low levels as the lower bound becomes binding. In contrast, the short rate in the *no* lower bound regime in the RSTSM evolves as if the regime continues in the *no* lower bound regime. The short rate in the *no* lower bound regime in the RSTSM can be further analyzed in the context of the ECB's monetary actions during the lower bound period. The vertical dashed lines in figure 9 indicate the ECB's extraordinary monetary policy actions during the lower bound period, see also ECB (2020f). Each of these actions will be briefly explained and compared to the short rate evolution in the *no* lower bound regime in the RSTSM.

In July 2013, the ECB launched forward guidance to communicate its intention of keeping short-term interest rates at low levels for an extended period. As explained in section II.B, forward guidance influences longer-term interest rates. In figure 9, the short rate in the *no* lower bound regime increases following the announcement of forward guidance, indicating an implausible evolution.

In June and September 2014, the ECB launched extraordinary expansionary policy actions. These included a 0.2 percentage point cut in the deposit rate, the Targeted Longer-term Refinancing Operations (TLTRO I) and the asset purchase programs; third Covered Bond Purchase Programme (CBPP3) and Asset-backed Securities Purchase Programme (ABSPP). These actions attempt to decrease the whole term structure of interest rates; see the sections II.B-II.A for explanations. In figure 9, the short rate in the *no* lower bound regime decreases significantly, indicating a somehow plausible evolution.

In January 2015, the ECB's asset purchase programs were expanded to include Public Sector Purchase Programme (PSPP). In figure 9, January 2015 is associated with a further drop in the short rate in the *no* lower bound regime.

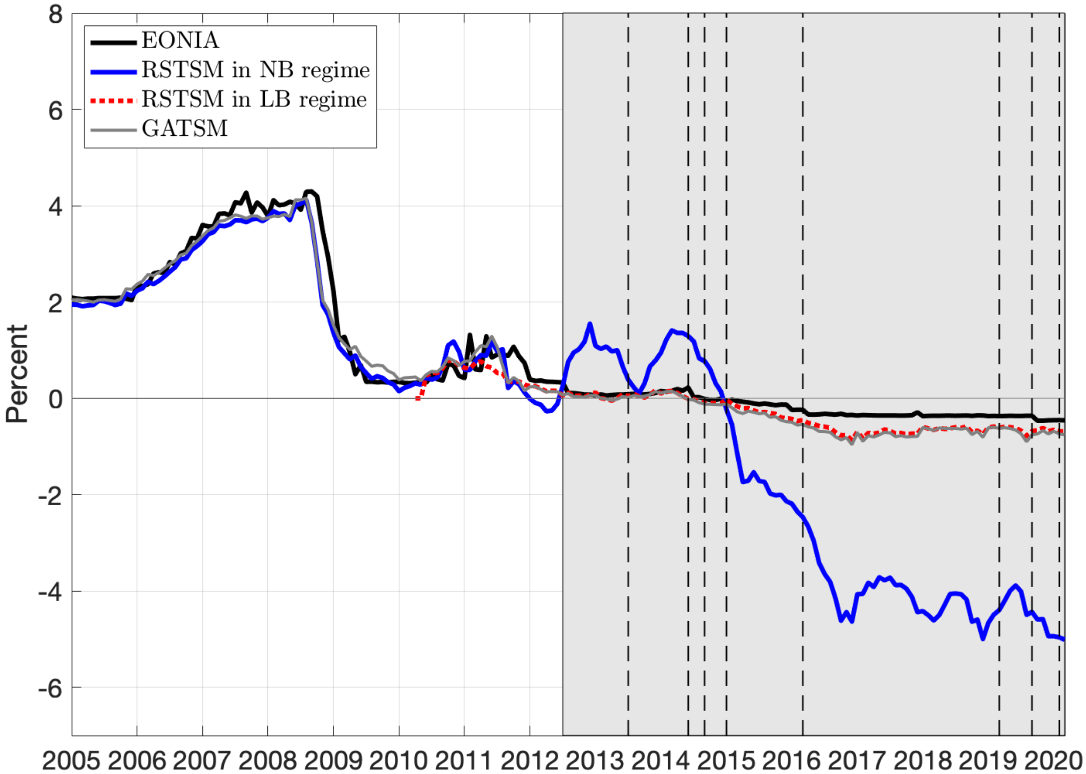
In March 2016, the ECB again launched extraordinary expansionary policy actions. These included a further cut of 0.2 percentage point in the despot rate, the second package of the TLTRO-series, i.e., TLTRO II, and asset purchase programs, including the Corporate Sector Purchase Programme (CSPP). In figure 9, March 2016 is associated with a large decline in the short rate in the *no* lower bound regime.

In March and September 2019, the ECB launched the third package of the TLTRO-series, i.e., TLTRO III, and decreased the deposit rate to the lowest rate of -0.5 pct. in the ECB's history. The short rate in the *no* lower bound regime in figure 9 decreases in this period, but this may be somewhat arbitrary as the evolution is quite volatile in the period 2017-19.

In March 2020, the ECB launched the Pandemic Emergency Purchase Programme (PEPP) to respond to the outbreak of the coronavirus. However, the effect is unreadable from figure 9 as March 2020 is the last observation in the sample.

Based on these findings, the estimated short rate in the *no* lower bound regime in the RSTSM summarizes the ECB's conduct of unconventional monetary policy quite well. Furthermore, the short rate in the RSTSM can be used as an indicator of the monetary policy stance, when other measures, e.g., the EONIA rate and the short rate in the GATSM, provide no guidance.

Figure 9. Indicators of the ECB's monetary policy stance



Note: "NB" and "LB" are abbreviations for *no* lower bound regime and lower bound regime. The black line is the Euro Overnight Index Average (EONIA) rate. The grey line is the estimated short rate in the GATSM. The blue dashed line is the estimated short rate in the RSTSM in the NB regime, and the red dotted line is the estimated short rate in the RSTSM in LB regime. The vertical dashed lines correspond to the European Central Bank's (ECB's) monetary policy actions, see ECB (2020f).

VI. Discussion

This section discusses the main findings in the empirical analysis. In particular, this section discusses how the findings for the short rate in the regime-switching term structure model (hereafter, "RSTSM") can be interpreted within monetary policy and how it can be used to assess the effect of unconventional monetary policy. Furthermore, this section addresses the empirical challenges that may arise when the European Central Bank's (hereafter, "ECB") existing reference rate is replaced with a new reference rate for the euro area. Finally, this section revisits the severe signs of misspecification in both the Gaussian affine term structure model (hereafter, "GATSM") and RSTSM. In this context, several model extensions are proposed.

A. Monetary policy — a new normal

In the wake of the global financial crisis in 2007-08, the ECB responded by cutting the official interest rates to historically low levels and launched several unconventional monetary policy actions to escape the lower bound constraint. These monetary policy initiatives effectively suppressed short-term nominal interest rates at the lower bound. The *new normal* in monetary policy is a *low for long* environment with low levels of short-term interest rates, low activity and below-target inflation, see Hartwell (2018). In this context, the Euro Overnight Index Average (hereafter, "EONIA") rate has been called into question as the EONIA rate may not reflect the ECB's actual monetary policy stance. Furthermore, it is still questionable how unconventional monetary policy actions affect the course of the overall economy.

Several studies, including Bernanke, Boivin, and Eliasz (2004), Swanson and Williams (2014) and Laine (2019), have examined the effects of monetary policy on the overall economy. These studies measure the effects of monetary policy in the euro area by using the EONIA rate as an indicator of the ECB's monetary policy stance. Laine (2019) finds much weaker effects of monetary policy as the ECB launched unconventional monetary policy actions. According to Laine (2019), the findings are highly influenced by the failure of the EONIA rate to capture unconventional monetary policy actions.

The findings in the empirical analysis in section V.C can enrich the models in e.g., Bernanke et al. (2004), Swanson and Williams (2014) and Laine (2019), by replacing the EONIA rate with the short rate in the RSTSM. As the evolution of the short rate in the RSTSM is not restricted by the lower bound, the short rate in the RSTSM can capture both conventional and unconventional monetary policy. A similar approach is considered in Wu and Xia (2016) that replaces the federal funds rate in Bernanke et al. (2004) with the shadow rate to examine the effects of the Federal Reserves' unconventional monetary policy on the overall economy.

In section V.C, the short rate in the RSTSM declined to extraordinarily low levels around -5 pct. from 2017 and forth. This level of the short rate in the RSTSM does not reflect the interest rate banks make short-term unsecured euro loans in the interbank market.[xxiv] The short rate in the RSTSM rather constitutes the short rate if the lower bound did not exist. In monetary policy, a short rate of -5 pct. is associated with a considerable expansionary policy. However, the short rate does not stand alone when central banks assess the monetary policy stance.

The ECB defines its monetary policy stance in *relative* terms. Under an efficient transmission channel, the ECB's official interest rates will influence the real interest rates. The ECB monitors its monetary policy stance by measuring the short-term real interest rate relative to the neutral real interest rate or $\mathbf{r}^{\star}$. If the real interest rate coincides with the neutral real interest rate, the monetary policy stance is said to be *neutral*. Similarly, if the real interest rate is below the neutral real interest rate, the monetary policy stance is *expansionary* and vice versa for a *contractionary* policy stance.

Wicksell (1907) introduced the neutral real interest rate defined as the short-term real interest rate consistent with maximum activity, full employment and stable prices. Monetary policy actively aims to guide the real interest rate towards the neutral real interest rate.

[xxiv] Unlike the short rate in the RSTSM, the EONIA rate is the average of overnight rates for unsecured interbank lending in the euro area, see ECB (2020g).

However, the neutral real interest rate is unobserved and estimated with uncertainty.

T. Laubach (2017) and C. Brand (2018) find a protracted fall in the neutral real interest rate in the euro area driven by demographic shifts, productivity slowdown and rising demand for safe, liquid assets. Specifically, T. Laubach (2017) and C. Brand (2018) find that the euro area neutral real interest rate has evolved around -2 pct. to -4 pct. from 2016 to 2019. In section V.C, the findings for the short rate in the RSTSM can be used to determine the ECB's monetary policy stance by measuring the short rate corrected for the expected inflation rate, relative to the low levels of the neutral real interest rate.

Throughout the empirical analysis in section V, the lower bound was assumed to be the minimum of the ECB's deposit rate and zero. This is consistent with the lower bound assumption in Kortela (2016), which is assumed in the absence of storage costs. In theory, the lower bound is given by the nominal interest rate that makes cash a perfect substitute for reserves. Kortela (2016) assumes that this level of the nominal interest is equal to zero. However, in practice, short-term nominal interest rates have evolved below this level without making cash preferable over reserves. Therefore, the assumption in Kortela (2016) is quite ambiguous. In contrast, Rogoff (2017) finds significant storage costs associated with holding cash, implying a lower bound below zero. Hence, the lower bound assumption in this paper and Kortela (2016) do not reflect the point where cash becomes a perfect substitute for reserves. Instead, it classifies a monetary policy regime in which unconventional monetary policy tools are implemented.

B. New reference rate for the euro area

This section addresses the challenges of the transition towards the new reference rate for the euro area; the Euro Short-Term Rate (€STR). The new reference rate is also the suggested proxy for the risk-free or short rate for the euro area, see ECB (2020g). At the operational level, the ECB uses the reference rate to signal its monetary policy stance as it typically reflects the first step of the transmission mechanism, see, e.g., Bindseil (2014).

In September 2018, a working group in the ECB constructed and recommended a new reference rate for the euro area, i.e., the €STR, see ECB (2020g). The previous reference rate, i.e., the EONIA rate[xxv], was called into question due to the tightened international regulations on reference rates, see IOSCO (2013). The EONIA rate was based on a voluntary statistical reporting system, where only a few financial counterparties contributed to the data collection. Consequently, the counterparties could collude on their data submissions to move the reference rate in a favorable way, which underpinned the lack of representativeness in the EONIA rate, see A. Schrimpf (2019). Additionally, the EONIA rate was calculated based on euro-transactions in the interbank market with a considerable low turnover.

Since October 2019, the ECB has published the new reference rate, i.e., the €STR, and has announced a complete phaseout of the EONIA rate in January 2022, see ECB (2020g). Until then, the EONIA rate will be pegged to the €STR plus a spread of 8.5 basis point.[xxvi] The ECB calculates the €STR based on observable transactions between systemic counterparties, including banks, money market funds, investment and pension funds, and non-financial corporations. The method and the sectoral coverage reflect the desired integrity and representativeness of the new reference rate, see A. Schrimpf (2019).

Monetary policy operates by steering short-term interest rates to trace the effects of policy-induced changes. In this context, a reliable and representative reference rate constitutes a significant part of the monetary policy transmission view, see ECB (2020g). Therefore, a smooth and complete transition towards the new reference rate is desirable.

The transition towards the new reference rate imposes several challenges for the counterparties using the reference rate in various products and contracts. An optimal exchange of information between the ECB and the counterparties is necessary to avoid financial risks and other potential distortions. How this works in practice remains to be seen.

[xxv] The Euro Overnight Index Average (EONIA) rate.
[xxvi] The spread reflects the methodology differences, see ECB (2020a)

In a term structure modeling framework, the new reference rate implies empirical challenges. First, the new reference rate is published from October 2019 and onwards. Based on a shorten sample, a comparison of the short rate and the new reference rate is questionable. Second, the new reference rate is based on a broad sectoral coverage in which not all counterparties have direct access to the ECB's deposit facility. Effectively, this means that the deposit rate does not further constitute an interest rate floor on the new reference rate, making the lower bound assumption further ambiguous. These challenges are crucial to address in the future work of this topic.

C. Shortcomings and next steps

This section discusses the underlying assumptions in the Gaussian affine term structure model (hereafter, "GATSM") and the regime-switching term structure model (hereafter, "RSTSM"). Several extensions of the term structure models are suggested to deal with the poor empirical performance of the models.

Section V.B.3 presented several misspecification tests for the GATSM and RSTSM. The misspecification tests were essentially testing the validity of the underlying assumptions in the term structure models. In the GATSM and RSTSM, the underlying assumptions are independent and identical distributed (*i.i.d.*) Gaussian errors. However, the findings in section V.B.3 imply that the assumptions of *i.i.d.* Gaussian errors are violated empirically. Consequently, the GATSM and RSTSM are inconsistently estimated in the sections V.B.2 and V.C.

The severe signs of misspecification are related to the postulated process for the unobserved state variables driven the forward rates. The state variables were assumed to follow a first-order vector autoregression, see e.g., equation (10) in section III.B. As the descriptive statistics of the forward rates in section V.A exhibited highly persistent forward rates, including solely one lag in the vector autoregressive process for the state variables may trigger autocorrelation in the errors. Furthermore, the forward rates in section V.A exhibited time-varying volatility, possibly dictated by the ECB's conduct of monetary policy. Both the GATSM and RSTSM assume constant conditional volatility in the process for the state variables, which may generate autoregressive conditional heteroskedasticity (ARCH) effects in the errors. These clear signs of misspecification in the GATSM and RSTSM may explain these models' poor empirical performance.

The GATSM and the RSTSM can be extended in several ways to circumvent the misspecification issues. One could postulate a higher-order vector autoregression for the unobserved state variables. Bond pricing will be unaffected if the vector autoregression is given in companion form, see, e.g., Singleton (2006). The specific order of the vector autoregression could be determined empirically by a general-to-specific model selection procedure, see Nielsen (2015). In the selection procedure, one would determine the vector autoregression order based on several criteria and tests. One could further relax the assumption of constant conditional volatility in the process for the state variables, e.g., by formulating a process for the conditional variance of the state variables, see Rahbek and Nielsen (2014). An immediate extension of the RSTSM is to introduce regime shifts in this process, see Ang et al. (2008) and Dai et al. (2007).

The RSTSM extended the GATSM by introducing regime shifts in the process for the short rate. The assumed process for the regimes was given by an unobserved 2-state Markov chain with constant transitional probabilities as in Ang et al. (2008). Furthermore, the RSTSM assumed that the market prices of risk associated with the shocks to the state variables were unaffected by the regimes. The assumption that the regimes follow a Markov chain was quite suitable for modeling persistent monetary policy regimes. However, the assumption of constant transitional probabilities is highly stylized as the monetary policy regimes depend on economic factors that vary across time.[xxvii] A further critique of the RSTSM is that the market prices of risk associated with shocks to the state variables are unaffected by the regimes. Typically, economic agents' attitudes towards risk are closely related to the conducted monetary policy. Therefore, one should introduce regime shifts in the process for the market prices of risk, e.g., by assuming the process in Dai et al. (2007).

[xxvii] One could propose time-varying transitional probabilities as in Dai et al. (2007).

The relatively small sample size considered in the empirical section V is also open for discussion. The sample spans from January 2005 to March 2020. During this period, the RSTSM solely detects a single regime shift. Ideally, the sample should contain several shifts to assess the empirical performance of the RSTSM. One could extend the empirical analysis by considering, e.g., a sample for U.S. forward rates. This is left for future work.

VII. Conclusion

The purpose of this thesis was to produce an indicator of the European Central Bank's (hereafter, "ECB") monetary policy stance when unconventional monetary policy is conducted. This thesis examined the following questions: *i. How would the short rate evolve in the euro area if the lower bound did not exist? ii. How can the short rate summarize the monetary policy stance when the ECB pursues unconventional monetary policy actions?*

This thesis answers these questions by proposing a novel regime-switching term structure model (hereafter, "RSTSM"). The RSTSM is an extension of the traditional Gaussian affine term structure model (hereafter, "GATSM"). The novelty of the RSTSM is that the model introduces regime shifts in the short rate process, i.e., the process for the nominal interest rate with the shortest maturity. In the RSTSM, the regimes are unobserved but classified by the two monetary policy regimes; a *no* lower bound regime and a lower bound regime.

This thesis conducts an empirical analysis of the GATSM and RSTSM based on euro area forward rates from January 2005 to March 2015. Maximum likelihood estimation and filtering methods are combined to obtain the parameters and state variables in the GATSM and the parameters, state variables and regimes in the RSTSM.

The findings in this thesis suggest that *i.* the RSTSM can guide how the short rate would have evolved in the euro area if the lower bound did not exist, and *ii.* the RSTSM can summarize the ECB's actual monetary policy stance quite well when unconventional monetary policy is conducted. Compared to the GATSM, the RSTSM offers a superior fit at the short end of the forward curve during the lower bound period. The regimes in the RSTSM are closely related to ECB's actual monetary policy. Finally, the RSTSM does not improve the severe signs of misspecification in the GATSM. Neither the GATSM nor the RSTSM is consistently estimated. Several model extensions are suggested to deal with the poor empirical performance of the models. These are left for future work.

VIII. References

A. Schrimpf, V. Sushko, 2019, Beyond libor: a primer on the new reference rates, Quarterly review, Bank for International Settlements.

Ang, Andrew, Geert Bekaert, and Min Wei, 2008, The term structure of real rates and expected inflation, *The Journal of Finance* 63, 797–849.

Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745–787.

Backus, David, Silverio Foresi, and Chris Telmer, 1998, Discrete-time models of bond pricing, Working Paper 6736, National Bureau of Economic Research.

Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *The Journal of Finance* 57, 1997–2043.

Bauer, Michael, and Glenn Rudebusch, 2016, Monetary policy expectations at the zero lower bound, *Journal of Money, Credit and Banking* 48, 1439–1465.

Bauer, Michael D., and Glenn D. Rudebusch, 2014, The Signaling Channel for Federal Reserve Bond Purchases, *International Journal of Central Banking* 10, 233–289.

Bernanke, Ben S, Jean Boivin, and Piotr Eliasz, 2004, Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach, Working Paper 10220, National Bureau of Economic Research.

Bernanke, Ben S, and Mark Gertler, 1995, Inside the black box: The credit channel of monetary policy transmission, *National Bureau of Economic Research* .

Bikbov, Ruslan, and Mikhail Chernov, 2013, Monetary policy regimes and the term structure of interest rates, *Journal of Econometrics* 174, 27 – 43.

Bindseil, Ulrich, 2014, *Monetary Policy Operations and the Financial System* (Oxford University Press).

Black, Fischer, 1995, Interest rates as options, *The Journal of Finance* 50, 1371–1376.

C. Brand, A. Penalver, M. Bielecki, 2018, The natural rate of interest: estimates, drivers, and challenges to monetary policy, Occasional paper, European Central Bank.

Cameron, A. Colin, and Pravin K. Trivedi, 2005, *Microeconometrics: Methods and Applications* (Cambridge University Press).

Cochrane, J. Howland, 2001, *Asset pricing* (Princeton University Press).

Cox, et al., John C., 1985, "a theory of the term structure of interest rates.", *Econometrica* 53, 385–407.

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *The Journal of Finance* 55, 1943–1978.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang, 2007, Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields, *The Review of Financial Studies* 20, 1669–1706.

Diebold, Francis X, and Canlin Li, 2003, Forecasting the term structure of government bond yields, Working Paper 10048, National Bureau of Economic Research.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405–443.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.

Dunsmuir, W., 1979, A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise, *The Annals of Statistics* 7, 490–506.

ECB, 2020a, *ECB provides a one-off spread between STR and EONIA (accessed August 21, 2020)*, https://www.ecb.europa.eu/press/pr/date/2019/html/ecb.pr190531~a3788de8f8.en.html.

ECB, 2020b, *Euro area yield curves (accessed August 20, 2020)*, https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html.

ECB, 2020c, *European Central Bank monetary policy strategy (accessed June 1, 2020)*, https://www.ecb.europa.eu/mopo/strategy/html/index.en.html.

ECB, 2020d, *The global financial crisis and the role of monetary policy (accessed July 5, 2020)*, https://www.ecb.europa.eu/press/key/date/2011/html/sp110924.en.html.

ECB, 2020e, *Official interest rates (accessed August 20, 2020)*, https://www.ecb.europa.eu/stats/policy_and_exchange_rates/key_ecb_interest_rates/html/index.en.html.

ECB, 2020f, *Past and future of the ECB monetary policy (accessed June 23, 2020)*, https://www.ecb.europa.eu/press/key/date/2018/html/ecb.sp180504.en.html.

ECB, 2020g, *Working group on euro risk-free rates (accessed August 10, 2020)*, https://www.ecb.europa.eu/paym/initiatives/interest_rate_benchmarks/WG_euro_risk-free_rates/html/index.en.html.

Engle, Robert F, 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987–1007.

Fama, Eugene F, 1984, Term premiums in bond returns, *Journal of Financial Economics* 13, 529 – 546.

Fisher, I., 1930, *The Theory of Interest as Determined by Impatience to Spend Income and Opportunity to Invest it*, Making of the modern world, part III (Macmillan).

Friedman, Milton, 1968, The role of monetary policy, *The American Economic Review* 58, 1–17.

Frühwirth-Schnatter, Sylvia, 2006, *Finite Mixture and Markov Switching Models*, volume 1 of *Springer Series in Statistics* (Springer New York).

Gagnon et. al., Joseph, 2011, The financial market effects of the federal reserve's large-scale asset purchases, *International Journal of Central Banking* 7, 3–43.

Gray, Stephen F., 1996, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42, 27 – 62.

Hamilton, James, and Baldev Raj, 2002, Advances in markov-switching models, *Empirical Economics* 27, 149–162.

Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.

Hamilton, James D., 1994, *Time Series Analysis*, first edition (Princeton University Press).

Hamilton, James D., and Jing Cynthia Wu, 2012, Identification and estimation of Gaussian affine term structure models, *Journal of Econometrics* 168, 315–331.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 517–613.

Hartwell, Christopher A., 2018, European monetary union reform and the new normal for monetary policy: Challenges and perspectives, Monetary Dialogue Directorate-General for Internal Policies, European Monetary Union.

Harvey, Andrew, 1991, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press).

Hull, John C., 2014, *Options, futures, and other derivatives*, 9th edition (Pearson Prentice Hall).

Hördahl, Peter, and Oreste Tristani, 2019, Modelling yields at the lower bound through regime shifts, BIS Working Papers 813, Bank for International Settlements.

IOSCO, 2013, Principles for financial benchmarks, Consultation report, International Organization of Securities Commissions.

Ireland, Peter, 2005, The monetary transmission mechanism, Working Papers 06-1, Federal Reserve Bank of Boston.

Joslin, Scott, Kenneth Singleton, and Haoxiang Zhu, 2010, A new perspective on gaussian dynamic term structure models, *Review of Financial Studies* 24, 926–970.

Joyce et. al., Michael, 2011, The financial market impact of quantitative easing in the united kingdom, *International Journal of Central Banking* 7, 113–161.

Kalman, R. E., 1960, A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering* 82, 35–45.

Kim, Chang-Jin, 1994, Dynamic linear models with markov-switching, *Journal of Econometrics* 60, 1 – 22.

Kim, Chang-Jin, and Charles Nelson, 1999, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, volume 1, first edition (The MIT Press).

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

Kim, Young Min, and Kyu Ho Kang, 2019, Likelihood inference for dynamic linear models with markov switching parameters: on the efficiency of the kim filter, *Econometric Reviews* 38, 1109–1130.

Kortela, T., 2016, A shadow rate model with time-varying lower bound of interest rates, Research Discussion Papers 19/2016, Bank of Finland.

Krippner, Leo, 2014, Measuring the stance of monetary policy in conventional and unconventional environments, CAMA Working Papers 2014-06, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Laine, Olli-Matti, 2019, The effect of the ecb's conventional monetary policy on the real economy: Favar-approach, *Empirical Economics* .

Le, Anh, Kenneth J. Singleton, and Qiang Dai, 2010, Discrete-Time Affine Term Structure Models with Generalized Market Prices of Risk, *The Review of Financial Studies* 23, 2184–2227.

Lütkepohl, H., 2007, *New Introduction to Multiple Time Series Analysis* (Springer Berlin Heidelberg).

Mishkin, Frederic S., 1996, The Channels of Monetary Transmission: Lessons for Monetary Policy, NBER Working Papers 5464, National Bureau of Economic Research, Inc.

Nielsen, Heino B., 2015, *Introduction to Likelihood-Based Estimation and Inference* (University of Copenhagen).

Patinkin, Don, 1989, *Neutrality of Money*, 273–287 (Palgrave Macmillan UK, London).

Rahbek, Anders, and Theis Lange, 2009, *An Introduction to Regime Switching Time Series Models, Handbook of Financial Time Series*, 871–887 (Springer Berlin Heidelberg).

Rahbek, Anders, and Heino Bohn Nielsen, 2011, Affine term-structure models in discrete time, *University of Copenhagen* In Lecture Notes for Financial Econometrics B, 1–15.

Rahbek, Anders, and Heino Bohn Nielsen, 2014, Unit root vector autoregression with volatility induced stationarity, *Journal of Empirical Finance* 29, 144 – 167.

Rahbek, Anders, and Rasmus S. Pedersen, 2019a, Qmle and the kalman filter, *University of Copenhagen* In Lecture Notes for Financial Econometrics A, 1–9.

Rahbek, Anders, and Rasmus S. Pedersen, 2019b, Stochastic volatility: An introduction, *University of Copenhagen* In Lecture Notes for Financial Econometrics A, 1–15.

Rogoff, K., 2017, Dealing with Monetary Paralysis at the Zero Bound, *Journal of Economic Perspectives* 31, 47–66.

Romer, David, 2012, *Advanced macroeconomics*, fourth edition (McGraw-Hill Irwin).

Shumway, Robert H., and David S. Stoffer, 2000, *Time Series Analysis and Its Applications* (Springer).

Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment* (Princeton University Press).

Swanson, Eric T., and John C. Williams, 2014, Measuring the effect of the zero lower bound on medium- and longer-term interest rates, *American Economic Review* 104, 3154–85.

T. Keister, J. McAndrews, A. Martin, 2008, Divorcing money from monetary policy, *Economic Policy Review* 14.

T. Laubach, J. C. Williams, K. Holston, 2017, Measuring the natural rate of interest: International trends and determinants, *Journal of International Economics* 108, S59 – S75, 39*th* Annual NBER International Seminar on Macroeconomics.

Tobin, James, 1969, A general equilibrium approach to monetary theory, *Journal of Money, Credit and Banking* 1, 15–29.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177 – 188.

Villa, Stefania, 2016, Financial frictions in the euro area and the united states: A bayesian assessment, *Macroeconomic Dynamics* 20, 1313–1340.

Walsh, Carl E., 2003, *Monetary Theory and Policy, 2nd Edition*, volume 1 of *MIT Press Books* (The MIT Press).

Wicksell, Knut, 1907, The influence of the rate of interest on prices, *History of Economic Thought Articles* 17, 213–220.

Wu, Jing Cynthia, and Fan Dora Xia, 2016, Measuring the macroeconomic impact of monetary policy at the zero lower bound, *Journal of Money, Credit and Banking* 48, 253–291.

IX. Appendix: Term structure modeling

A. Bond pricing in the GATSM

This section presents the derivations of result III.1. The idea is to combine equation (16) and equation (17) in result III.1, manipulate the terms and solve the expectations. A starting point is to insert equation (16) in (17):

$$\mathcal{P}_t^n = \mathbb{E}_t^{\mathbb{Q}}(\exp\{-r_t\}\exp\{A_{n-1} + \mathbf{B}'_{n-1}\mathbf{X}_{t+1}\})$$

The short rate is known at time t conditional on the information available at time t. To see this, it is useful to note that the state variables generate the information set and the short rate is completely determined by an affine transformation of the state variables, see e.g., definition III.5 in section III.B. The next step is to insert the process for the state variables in equation (15) in section III.B:

$$\begin{aligned}\mathcal{P}_t^n &= \exp\{-r_t + A_{n-1}\}\mathbb{E}_t^{\mathbb{Q}}(\exp\{\mathbf{B}'_{n-1}(\boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_t + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}})\}) \\ &= \exp\{-r_t + A_{n-1} + \mathbf{B}'_{n-1}\boldsymbol{\mu}^{\mathbb{Q}} + \mathbf{B}'_{n-1}\boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_t\}\mathbb{E}_t^{\mathbb{Q}}(\exp\{\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\})\end{aligned}$$

Here, all non-stochastic terms conditional on the information set at time t are taken out the expectation operator. Therefore, the underlying stochastic part in the bond price is the *i.i.d.* Gaussian innovations associated with the state variables. Under the assumption of $\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$, the exponential transformation of the Gaussian innovations means that the term $\exp\{\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\}$ is log-Gaussian distributed. The conditional expectation is given by:

$$\mathbb{E}_t^{\mathbb{Q}}(\exp\{\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\}) = \exp\{\frac{1}{2}\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}\}$$

Insert this expression in the bond price to get the following:

$$\mathcal{P}_t^n = \exp\{-r_t + A_{n-1} + \mathbf{B}'_{n-1}\boldsymbol{\mu}^{\mathbb{Q}} + \mathbf{B}'_{n-1}\boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_t + \frac{1}{2}\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}\}$$

Insert the short rate and collect terms to obtain:

$$\mathcal{P}_{t,n} = \exp\{A_{n-1} + \mathbf{B}'_{n-1}\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1} - \delta_0 + (\mathbf{B}'_{n-1}\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1)\mathbf{X}_t\}$$

The final step is to match the factor loadings in the following way:

$$\begin{aligned} A_n &= A_{n-1} + \mathbf{B}'_{n-1}\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1} - \delta_0 \\ \mathbf{B}'_n &= \mathbf{B}'_{n-1}\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1 \end{aligned}$$

This gives the bond pricing equation in result III.1 in section III.B:

$$\mathcal{P}^n_t = \exp\{A_n + \mathbf{B}'_n\mathbf{X}_t\},$$

The factor loadings follow ordinary difference equations which recursions are initiated at $A_0 = 0$ and $\mathbf{B}'_0 = \mathbf{0}$, see e.g., Duffie and Kan (1996) for a further discussion.

B. Bond pricing in the RSTSM

This section presents the derivations of result III.2. Result III.2 can be derived by utilizing the methods proposed in Bansal and Zhou (2002) and Ang et al. (2008). The first step is to let $s_t = i \in \mathcal{S} = \{NB, LB\}$ in equation (32):

$$\mathcal{P}^n_t = \mathbb{E}^{\mathbb{Q}}_t\big(\exp\{-r_t(i)\}\mathcal{P}^{n-1}_{t+1}\}\big)$$

Here, $\mathcal{P}^{n-1}_{t+1}$ depends on s_{t+1} via the exponential-affine relation in equation (31). Insert equation (31), let $s_{t+1} = j \in \mathcal{S} = \{NB, LB\}$ and use law of total probability to obtain:

$$\mathcal{P}^n_t = \sum_{j\in\mathcal{S}} \pi_{i,j}\mathbb{E}^{\mathbb{Q}}_t\big(\exp\{-r_t(i)\}\exp\{A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\mathbf{X}_{t+1}\}\big)$$

Insert the state variables given in equation (30) and take all terms that are measurable at time t out the expectation:

$$\begin{aligned}\mathcal{P}_t^n &= \sum_{j\in\mathcal{S}} \pi_{i,j}\exp\{-r_t(i) + A_{n-1}(j)\}\mathbb{E}_t^{\mathbb{Q}}\big(\exp\{\mathbf{B}'_{n-1}(j)(\boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_t + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}})\}\big) \\ &= \sum_{j\in\mathcal{S}} \pi_{i,j}\exp\{-r_t(i) + A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \mathbf{B}'_{n-1}\boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{X}_t\}\mathbb{E}_t^{\mathbb{Q}}\big(\exp\{\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\}\big)\end{aligned}$$

Under the assumption of $\boldsymbol{\epsilon}_t^{\mathbb{Q}} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$, the term $\exp\{\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\}$ is log-Gaussian distributed and its conditional expectation is given by:

$$\mathbb{E}_t^{\mathbb{Q}}\big(\exp\{\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\epsilon}_{t+1}^{\mathbb{Q}}\}\big) = \exp\{\frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j)\}$$

Insert this and the short rate for $s_t = i$ to obtain the following:

$$\mathcal{P}_t^n = \sum_{j\in\mathcal{S}} \pi_{i,j}\exp\{A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j) - \delta_0(i) + \big(\mathbf{B}'_{n-1}(j)\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i)\big)\mathbf{X}_t\}$$

At this stage, it is convenient to manipulate the terms in the following way:

$$1 = \sum_{j\in\mathcal{S}} \pi_{i,j}\exp\{A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j) - \delta_0(i) + \big(\mathbf{B}'_{n-1}(j)\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i)\big)\mathbf{X}_t - \log\big(\mathcal{P}_t^n\big)\}$$

The exponential is approximated around zero, i.e., $\exp(z) \simeq 1 + z$[xxviii]. Use this to obtain:

$$0 \simeq \sum_{j\in\mathcal{S}} \pi_{i,j}\Big(A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j) - \delta_0(i) + \big(\mathbf{B}'_{n-1}(j)\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i)\big)\mathbf{X}_t\Big) - \log\big(\mathcal{P}_t^n\big)$$

The equation is rewritten to the following:

$$\mathcal{P}_{t,n} \simeq \exp\{\sum_{j\in\mathcal{S}} \pi_{i,j}\Big(A_{n-1}(j) + \mathbf{B}'_{n-1}(j)\boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}'_{n-1}(j)\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{n-1}(j) - \delta_0(i) + \big(\mathbf{B}'_{n-1}(j)\boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i)\big)\mathbf{X}_t\Big)\}$$

Finally, the factor loadings are matched in the following way:

$$\mathcal{P}_{t,n} = \exp\{A_n(i) + \mathbf{B}'_n(i)\mathbf{X}_t\}$$

[xxviii] This is the first-order approximation proposed in Bansal and Zhou (2002). Note, $\sum_{j\in\mathcal{S}} \pi_{i,j} = 1$.

with

$$
\begin{aligned}
A_n(i) &= \sum_{j \in \mathcal{S}} \pi_{i,j} \Big(A_{n-1}(j) + \mathbf{B}'_{n-1}(j) \boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2} \mathbf{B}'_{n-1}(j) \boldsymbol{\Sigma} \boldsymbol{\Sigma}' \mathbf{B}_{n-1}(j) - \delta_0(i) \Big) \\
\mathbf{B}'_n(i) &= \sum_{j \in \mathcal{S}} \pi_{i,j} \Big(\mathbf{B}'_{n-1}(j) \boldsymbol{\Phi}^{\mathbb{Q}} - \boldsymbol{\delta}'_1(i) \Big)
\end{aligned}
$$

The recursions are initiated at $A_1(i) = \delta_0(i)$ and $\mathbf{B}'_1(i) = \boldsymbol{\delta}'_1(i)$. This concludes the no-arbitrage bond pricing in the RSTSM.

X. Appendix: Econometric methodology

A. Derivation of the Kalman filter

This section derives the Kalman filter that is used for estimating the state variables in the GATSM. The Kalman filter produces estimates of the conditional mean and variance of the state variables. These moments are determined recursively based on the assumptions and dynamics of the GATSM. The recursions require initialization, which was provided in section IV.C. The estimated conditional moments of the state variables are used for deriving the conditional mean and variance of the observed forward rates. The conditional moments of the forward rates are used for parameter estimation of the GATSM. All derivations below are based on Kalman (1960), Frühwirth-Schnatter (2006) and Rahbek and Pedersen (2019a).

The Kalman filter proceeds in two steps. The first step is the prediction step. For the ease of notation, let $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^{\mathbb{P}}$, $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}^{\mathbb{P}}$ and $\boldsymbol{\epsilon}_t \equiv \boldsymbol{\epsilon}_t^{\mathbb{P}}$.

Prediction

In the prediction step, the Kalman filter uses all the past available information from the observed forward rates to determine the unobserved state variables. The conditional mean of the state variables based on the past observed forward rates is given by:

$$
\begin{aligned}
\mathbf{X}_{t|t-1} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t-1}) \\
&= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t|\mathbf{F}_{1:t-1}) \\
&= \boldsymbol{\mu} + \boldsymbol{\Phi}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_{t-1}|\mathbf{F}_{1:t-1})}_{=\,\mathbf{X}_{t-1|t-1}} + \boldsymbol{\Sigma}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t|\mathbf{F}_{1:t-1})}_{=\,\mathbf{0}} \\
&= \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1}
\end{aligned}
$$

and the prediction error is given by:

$$\begin{aligned}
\mathbf{P}^{\mathbf{X}}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t - \boldsymbol{\mu} - \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1})(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t - \boldsymbol{\mu} - \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\boldsymbol{\Phi}\mathbf{X}_{t-1}\mathbf{X}'_{t-1}\boldsymbol{\Phi}' - \boldsymbol{\Phi}\mathbf{X}_{t-1}\mathbf{X}'_{t-1|t-1}\boldsymbol{\Phi}' - \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1}\mathbf{X}'_{t-1}\boldsymbol{\Phi}' + \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1}\mathbf{X}'_{t-1|t-1}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t\boldsymbol{\Sigma}'|\mathbf{F}_{1:t-1}\Big) \\
&= \boldsymbol{\Phi}\,\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1})(\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1})'|\mathbf{F}_{1:t-1}\Big)}_{=\,\mathbf{P}^{\mathbf{X}}_{t|t-1}}\,\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \\
&= \boldsymbol{\Phi}\mathbf{P}^{\mathbf{X}}_{t|t-1}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\boldsymbol{\Sigma}'
\end{aligned}$$

where the properties of the error term were used repeatedly, i.e., $\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}_t) = \mathbf{0}$, $\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t) = \mathbf{I}_k$ and finally $\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}_t\mathbf{X}'_t) = \mathbf{0}$ as $\boldsymbol{\epsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$.

The conditional moments of the state variables can be used to predict the observed forward rates conditional on the past forward rates. The conditional mean of the observed forward rates is given by:

$$\begin{aligned}
\mathbf{F}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t|\mathbf{F}_{1:t-1}) \\
&= \mathbf{A} + \mathbf{B}'\,\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:t-1})}_{=\,\mathbf{X}_{t|t-1}} + \boldsymbol{\Omega}\,\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\eta}_t|\mathbf{F}_{1:t-1})}_{=\,\mathbf{0}} \\
&= \mathbf{A} + \mathbf{B}'\mathbf{X}_{t|t-1}
\end{aligned}$$

and the conditional variance is given by:

$$\begin{aligned}
\mathbf{P}^{\mathbf{F}}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{F}_t - \mathbf{F}_{t|t-1})(\mathbf{F}_t - \mathbf{F}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t - \mathbf{A} - \mathbf{B}'\mathbf{X}_{t|t-1})(\mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t - \mathbf{A} - \mathbf{B}'\mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\mathbf{B}'\mathbf{X}_t\mathbf{X}'_t\mathbf{B} - \mathbf{B}'\mathbf{X}_t\mathbf{X}'_{t|t-1}\mathbf{B} - \mathbf{B}'\mathbf{X}_{t|t-1}\mathbf{X}'_t\mathbf{B} + \mathbf{B}'\mathbf{X}_{t|t-1}\mathbf{X}'_{t|t-1}\mathbf{B} + \boldsymbol{\Omega}\boldsymbol{\eta}_t\boldsymbol{\eta}'_t\boldsymbol{\Omega}'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbf{B}'\,\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big)}_{=\,\mathbf{P}^{\mathbf{X}}_{t|t-1}}\,\mathbf{B} + \boldsymbol{\Omega}\boldsymbol{\Omega}' \\
&= \mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B} + \boldsymbol{\Omega}\boldsymbol{\Omega}'
\end{aligned}$$

The covariance between the state variables and forward rates is given by:

$$
\begin{aligned}
\mathbf{P}^{\mathbf{FX}}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{F}_t - \mathbf{F}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{A} + \mathbf{B}'\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t - \mathbf{A} - \mathbf{B}'\mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{B}'(\mathbf{X}_t - \mathbf{X}_{t|t-1}) + \boldsymbol{\Omega}\boldsymbol{\eta}_t)(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\mathbf{B}'(\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})' + \boldsymbol{\Omega}\boldsymbol{\eta}_t(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big) \\
&= \mathbf{B}' \underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big)}_{=\,\mathbf{P}^{\mathbf{X}}_{t|t-1}} + \boldsymbol{\Omega}\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\boldsymbol{\eta}_t(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}\Big)}_{=\,\mathbf{0}} \\
&= \mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1}
\end{aligned}
$$

or, equivalently,

$$
\mathbf{P}^{\mathbf{XF}}_{t|t-1} = \mathbf{P}^{\mathbf{FX}\prime}_{t|t-1} = (\mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1})' = \mathbf{P}^{\mathbf{X}\prime}_{t|t-1}\mathbf{B} = \mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B},
$$

where it is used that $(\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ so $\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\eta}_t\mathbf{X}'_t) = \mathbf{0}$, and $\mathbf{P}^{\mathbf{X}\prime}_{t|t-1} = \mathbf{P}^{\mathbf{X}}_{t|t-1}$ due to symmetry.

The forward rates are conditionally Gaussian, so the conditional distribution is completely given by the mean and variance. These moments are used for constructing the likelihood function of the GATSM see equation (46) in section IV.C. The next step in the Kalman filter is to compute the updated or filtered values of the state variables.

Update

The update step in the Kalman filter relies on the assumption of Gaussian errors $(\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$. Under this assumption, the joint process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ is multivariate Gaussian. Therefore, the distribution of the process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ conditional on $\mathbf{F}_{1:t-1}$ is given by:

$$
(\mathbf{X}_t, \mathbf{F}_t)'|\mathbf{F}_{1:t-1} \;\sim\; \mathcal{N}\left(\begin{pmatrix} \mathbf{X}_{t|t-1} \\ \mathbf{F}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{P}^{\mathbf{X}}_{t|t-1} & \mathbf{P}^{\mathbf{XF}}_{t|t-1} \\ \mathbf{P}^{\mathbf{FX}}_{t|t-1} & \mathbf{P}^{\mathbf{F}}_{t|t-1} \end{pmatrix}\right)
$$

In the update step, the predicted state variables are updated by using a weighted average of the next observed forward rates and the predicted state variables. Therefore, the goal is to update $\mathbf{X}_{t|t-1}$ to $\mathbf{X}_{t|t}$, where $\mathbf{X}_{t|t}$ is also known as the filtered state variables. By using the properties of a multivariate Gaussian distribution, the updated or filtered state variables are given by:

$$\begin{aligned}
\mathbf{X}_{t|t} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:t}) \\
&= \mathbf{X}_{t|t-1} + \underbrace{\mathbf{P}^{\mathbf{XF}}_{t|t-1}(\mathbf{P}^{\mathbf{F}}_{t|t-1})^{-1}}_{\equiv\, \mathbf{K}_t}(\mathbf{F}_t - \mathbf{F}_{t|t-1}) \\
&= \mathbf{X}_{t|t-1} + \mathbf{K}_t(\mathbf{F}_t - \mathbf{F}_{t|t-1})
\end{aligned}$$

and the conditional variance of the filtered state variables is given by:

$$\begin{aligned}
\mathbf{P}^{\mathbf{X}}_{t|t} &= \mathbb{V}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:t}) \\
&= \mathbf{P}^{\mathbf{X}}_{t|t-1} - \mathbf{P}^{\mathbf{XF}}_{t|t-1}(\mathbf{P}^{\mathbf{F}}_{t|t-1})^{-1}\mathbf{P}^{\mathbf{XF}\prime}_{t|t-1} \\
&= \mathbf{P}^{\mathbf{X}}_{t|t-1} - \mathbf{K}_t\mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1} \\
&= (\mathbf{I}_k - \mathbf{K}_t\mathbf{B}')\mathbf{P}^{\mathbf{X}}_{t|t-1}
\end{aligned}$$

Here, $\mathbf{K}_t$ defines the Kalman gain or the linear regression coefficient, and it is given by:

$$\mathbf{K}_t = \mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B}(\mathbf{B}'\mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B} + \boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1}$$

The Kalman gain can be interpreted as the relative weight between the observed forward rates and the predicted state variables. The weight is determined by the precision of the observed forward rates and prediction of the state variables. This concludes the derivation of the Kalman filter.

B. Derivation of the Kim filter

This section derives the Kim filter in full details. The Kim filter is used to obtain the unobserved state variables and regimes. The Kim filter can be seen as a combination of the

Kalman filter and the Hamtilon filter. Therefore, it is natural to derive the filters separately. The obtained state variables and regimes from the Kim filter are used later for estimating the parameters in RSTSM. For a detailed explanation see section IV.D.

The derivations below are based on Kim (1994), Hamilton (1989), Frühwirth-Schnatter (2006), Rahbek and Lange (2009). For the ease of notation, let $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^{\mathbb{P}}$, $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}^{\mathbb{P}}$ and $\boldsymbol{\epsilon}_t \equiv \boldsymbol{\epsilon}_t^{\mathbb{P}}$.

Hamilton filter

The predicted and filtered values of the unobserved regimes are calculated from the Hamilton filter. The first step is to derive the prediction step.

Prediction

The unobserved regimes follow a first-order Markov chain. Therefore, the current regime is solely dependent on the last regime. The prediction step calculates the quantity $\mathcal{Q}_{t|t-1}^{(i,j)} = \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:t-1})$. The prediction in the Hamilton filter is given by:

$$
\begin{aligned}
\mathcal{Q}_{t|t-1}^{(i,j)} &= \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:t-1}) \\
&= \mathbb{P}_{\boldsymbol{\theta}}(s_t = j|\mathbf{F}_{1:t-1}, s_{t-1} = i)\mathbb{P}_{\boldsymbol{\theta}}(s_{t-1} = i|\mathbf{F}_{1:t-1}) \\
&= \underbrace{\mathbb{P}_{\boldsymbol{\theta}}(s_t = j|s_{t-1} = i)}_{=\ \pi_{i,j}} \underbrace{\mathbb{P}_{\boldsymbol{\theta}}(s_{t-1} = i|\mathbf{F}_{1:t-1})}_{=\ \mathcal{Q}_{t-1|t-1}^{(i)}} \\
&= \pi_{i,j}\mathcal{Q}_{t-1|t-1}^{(i)}
\end{aligned}
$$

The quantity $\mathcal{Q}_{t-1|t-1}^{(i)}$ is the updated or filtered value. The update step is explained below.

Update

In the update step, the Hamilton filter uses Bayes' theorem to update the predicted values. The update step is calculated simultaneously with the Kalman filter. For convenience, the derivations below treat the input from the Kalman filter as given. The update is given by:

$$\begin{aligned}
\mathcal{Q}_{t|t}^{(i,j)} &= \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i | \mathbf{F}_{1:t}) \\
&= \frac{f_{\boldsymbol{\theta}}(\mathbf{F}_t, s_t = j, s_{t-1} = i | \mathbf{F}_{1:t-1})}{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})} \\
&= \frac{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)}{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})} \underbrace{\mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i | \mathbf{F}_{1:t-1})}_{= \mathcal{Q}_{t|t-1}^{(i,j)}} \\
&= \frac{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)}{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})} \mathcal{Q}_{t|t-1}^{(i,j)}
\end{aligned}$$

Thanks to Bayes' theorem. The fraction $\frac{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)}{f_{\boldsymbol{\theta}}(\mathbf{F}_t | \mathbf{F}_{1:t-1})}$ is obtained by the Kalman filter. These will be derived in a moment see equations (62) and (63). The final step in the Hamilton filter is to obtain the current regime i.e. $\mathcal{Q}_{t|t}^{(j)} = \mathbb{P}_{\boldsymbol{\theta}}(s_t = j | \mathbf{F}_{1:t})$. This is step uses law of total probability:

$$\begin{aligned}
\mathcal{Q}_{t|t}^{(j)} &= \mathbb{P}_{\boldsymbol{\theta}}(s_t = j | \mathbf{F}_{1:t}) \\
&= \sum_{i \in \mathcal{S}} \underbrace{\mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i | \mathbf{F}_{1:t})}_{= \mathcal{Q}_{t|t}^{(i,j)}} \\
&= \sum_{i \in \mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}
\end{aligned}$$

for $t = 1, ..., T$ and $i, j \in \mathcal{S} = \{NB, LB\}$. This makes the Hamilton filter quite easy to implement and computationally fast to execute in application. Now, it remains to derive the Kalman filter.

Kalman filter

The predicted and filtered values of the state variables are obtained by the Kalman filter. The calculations follow the Kalman filter applied to the GATSM, but the difference is that the regimes are now included. The derivations below are based on Kim (1994), Kim and Nelson (1999) and Frühwirth-Schnatter (2006). The first step is to derive the prediction step.

Prediction

The prediction step produces estimates of the conditional mean and variance of the state variables based on the observed forward rates and the current and the previous regimes. The predicted state variables are given by:

$$\begin{aligned}
\mathbf{X}_{t|t-1}^{(i)} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t-1}, s_{t-1} = i) \\
&= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t|\mathbf{F}_{1:t-1}, s_{t-1} = i) \\
&= \boldsymbol{\mu} + \boldsymbol{\Phi}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_{t-1}|\mathbf{F}_{1:t-1}, s_{t-1} = i)}_{=\mathbf{X}_{t-1|t-1}^{(i)}} + \boldsymbol{\Sigma}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t|\mathbf{F}_{1:t-1}, s_{t-1} = i)}_{=\mathbf{0}} \\
&= \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{X}_{t-1|t-1}^{(i)}
\end{aligned}$$

with the following prediction error:

$$\begin{aligned}
\mathbf{P}_{t|t-1}^{\mathbf{X}(i)} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1}^{(i)})(\mathbf{X}_t - \mathbf{X}_{t|t-1}^{(i)})'|\mathbf{F}_{1:t-1}, s_{t-1} = i\Big) \\
&= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}\Big((\boldsymbol{\Phi}(\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1}^{(i)}) + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t)(\boldsymbol{\Phi}(\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1}^{(i)}) + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t)'|\mathbf{F}_{1:t-1}, s_{t-1} = i\Big) \\
&= \boldsymbol{\Phi}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}\Big((\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1}^{(i)})(\mathbf{X}_{t-1} - \mathbf{X}_{t-1|t-1}^{(i)})'|\mathbf{F}_{1:t-1}, s_{t-1} = i\Big)}_{=\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathbf{F}_{1:t-1}, s_{t-1} = i)}_{=\mathbf{I}_k}\boldsymbol{\Sigma}' \\
&= \boldsymbol{\Phi}\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\boldsymbol{\Sigma}'
\end{aligned}$$

where the properties of the error term were used repeatedly i.e. $\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t) = \mathbf{0}$, $\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \mathbf{I}_k$ and finally $\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\boldsymbol{\epsilon}_t\mathbf{X}_t') = \mathbf{0}$ as $\boldsymbol{\epsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$.

The predicted conditional mean and variance of the state variables are then used to estimate the conditional mean and variance of the current forward rates based on the previous observed forward rates, the current and last regimes. The prediction is given by:

$$\begin{aligned}
\mathbf{F}^{(i,j)}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{A}(j) + \mathbf{B}'(j)\mathbf{X}_t + \boldsymbol{\Omega}\boldsymbol{\eta}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i) \\
&= \mathbf{A}(j) + \mathbf{B}'(j)\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i) + \boldsymbol{\Omega}\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\eta}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)}_{=\,\mathbf{0}} \\
&= \mathbf{A}(j) + \mathbf{B}'(j)\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{F}_{1:t-1}, s_{t-1} = i)}_{=\,\mathbf{X}^{(i)}_{t|t-1}} \\
&= \mathbf{A}(j) + \mathbf{B}'(j)\mathbf{X}^{(i)}_{t|t-1}
\end{aligned}$$

with the following prediction error:

$$\begin{aligned}
\mathbf{P}^{\mathbf{F}(i,j)}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{F}_t - \mathbf{F}^{(i,j)}_{t|t-1})(\mathbf{F}_t - \mathbf{F}^{(i,j)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{B}'(j)(\mathbf{X}_t - \mathbf{X}_{t|t-1}) + \boldsymbol{\Omega}\boldsymbol{\eta}_t)(\mathbf{B}'(j)(\mathbf{X}_t - \mathbf{X}_{t|t-1}) + \boldsymbol{\Omega}\boldsymbol{\eta}_t)'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) \\
&= \mathbf{B}'(j)\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big)\mathbf{B}(j) + \boldsymbol{\Omega}\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)\boldsymbol{\Omega}' \\
&= \mathbf{B}'(j)\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_{t-1} = i\Big)}_{=\,\mathbf{P}^{\mathbf{X}}_{t|t-1}}\mathbf{B}(j) + \boldsymbol{\Omega}\boldsymbol{\Omega}' \\
&= \mathbf{B}'(j)\mathbf{P}^{\mathbf{X}}_{t|t-1}\mathbf{B}(j) + \boldsymbol{\Omega}\boldsymbol{\Omega}'
\end{aligned}$$

The covariance between the predicted state variables and the forward rates is given by:

$$\begin{aligned}
\mathbf{P}^{\mathbf{FX}(i,j)}_{t|t-1} &= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{F}_t - \mathbf{F}^{(i,j)}_{t|t-1})(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{B}'(j)(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1}) + \boldsymbol{\Omega}\boldsymbol{\eta}_t)(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) \\
&= \mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\mathbf{B}'(j)(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})' + \boldsymbol{\Omega}\boldsymbol{\eta}_t(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) \\
&= \mathbf{B}'(j)\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big) + \boldsymbol{\Omega}\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big(\boldsymbol{\eta}_t(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big)}_{=\,\mathbf{0}} \\
&= \mathbf{B}'(j)\underbrace{\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}\Big((\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})(\mathbf{X}_t - \mathbf{X}^{(i)}_{t|t-1})'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i\Big)}_{=\,\mathbf{P}^{\mathbf{X}(i)}_{t|t-1}} \\
&= \mathbf{B}'(j)\mathbf{P}^{\mathbf{X}(i)}_{t|t-1}
\end{aligned}$$

or, equivalently,

$$\mathbf{P}^{\mathbf{FX}(i,j)} = \mathbf{P}^{\mathbf{FX}(i,j)\prime}_{t|t-1} = (\mathbf{B}'(j)\mathbf{P}^{\mathbf{X}(i)}_{t|t-1})' = \mathbf{P}^{\mathbf{X}(i)\prime}_{t|t-1}\mathbf{B}(j) = \mathbf{P}^{\mathbf{X}(i)}_{t|t-1}\mathbf{B}(j)$$

where it used that $(\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$ so $\mathbb{E}^{\mathbb{P}}_{\boldsymbol{\theta}}(\boldsymbol{\eta}_t \mathbf{X}_t') = \mathbf{0}$, and $\mathbf{P}^{\mathbf{X}(i)\prime}_{t|t-1} = \mathbf{P}^{\mathbf{X}(i)}_{t|t-1}$ due to symmetry.

The forward rates are conditionally Gaussian distributed implying that the conditional distribution of the forward rates are completely given by the conditional mean and variance. The conditional probability density function is given by:

$$f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}) = \sum_{i,j\in\mathcal{S}} f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i)\mathcal{Q}^{(i,j)}_{t|t-1} \tag{62}$$

with

$$f^{\mathbb{P}}_{\boldsymbol{\theta}}(\mathbf{F}_t|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i) = (2\pi)^{-\frac{p}{2}}|\mathbf{P}^{\mathbf{F}(i,j)}_{t|t-1}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{F}_t - \mathbf{F}^{(i,j)}_{t|t-1})'(\mathbf{P}^{\mathbf{F}(i,j)}_{t|t-1})^{-1}(\mathbf{F}_t - \mathbf{F}^{(i,j)}_{t|t-1})\} \tag{63}$$

The conditional probability density function of the forward rates is used for updating the unobserved regimes in the Hamilton filter. Furthermore, the likelihood function of the RSTSM in section IV.D in equation (60) is based on these equations. The next step is to derive the update step.

Update

The update step relies on the assumption of Gaussian errors $(\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$. Under this assumption and conditional on the current and last regimes, the the joint process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ is multivariate Gaussian. The distribution of the process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^T$ conditional on $\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i$ is given by:

$$(\mathbf{X}_t, \mathbf{F}_t)'|\mathbf{F}_{1:t-1}, s_t = j, s_{t-1} = i \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{X}^{(i)}_{t|t-1}\\ \mathbf{F}^{(i,j)}_{t|t-1}\end{pmatrix}, \begin{pmatrix}\mathbf{P}^{\mathbf{X}(i)}_{t|t-1} & \mathbf{P}^{\mathbf{XF}(i,j)}_{t|t-1}\\ \mathbf{P}^{\mathbf{FX}(i,j)}_{t|t-1} & \mathbf{P}^{\mathbf{F}(i,j)}_{t|t-1}\end{pmatrix}\right)$$

which is based on the conditional moments derived from the prediction step. This conditional distribution can be used to update the state variables. By using the properties of a multivariate Gaussian distribution, the updated or filtered state variables are given by:

$$
\begin{aligned}
\mathbf{X}_{t|t}^{(i,j)} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i) \\
&= \mathbf{X}_{t|t-1}^{(i)} + \underbrace{\mathbf{P}_{t|t-1}^{\mathbf{XF}(i,j)}(\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)})^{-1}}_{\equiv\ \mathbf{K}_t^{(i,j)}}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)}) \\
&= \mathbf{X}_{t|t-1}^{(i)} + \mathbf{K}_t^{(i,j)}(\mathbf{F}_t - \mathbf{F}_{t|t-1}^{(i,j)})
\end{aligned}
$$

and with the following conditional variance:

$$
\begin{aligned}
\mathbf{P}_{t|t}^{\mathbf{X}(i,j)} &= \mathbb{V}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i) \\
&= \mathbf{P}_{t|t-1}^{\mathbf{X}(i)} - \mathbf{P}_{t|t-1}^{\mathbf{XF}(i,j)}(\mathbf{P}_{t|t-1}^{\mathbf{F}(i,j)})^{-1}\mathbf{P}_{t|t-1}^{\mathbf{XF}(\mathbf{i},\mathbf{j})\prime} \\
&= \mathbf{P}_{t|t-1}^{\mathbf{X}(i)} - \mathbf{K}_t^{(i,j)}\mathbf{B}'(j)\mathbf{P}_{t|t-1}^{\mathbf{X}(i)} \\
&= (\mathbf{I}_k - \mathbf{K}_t^{(i,j)}\mathbf{B}'(j))\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}
\end{aligned}
$$

Here, $\mathbf{K}_t^{(i,j)}$ defines the Kalman gain or the linear regression coefficient, and it is given by:

$$
\mathbf{K}_t^{(i,j)} = \mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\mathbf{B}(j)(\mathbf{B}'(j)\mathbf{P}_{t|t-1}^{\mathbf{X}(i)}\mathbf{B}(j) + \boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1}
$$

with the exact same intuition as in the Kalman filter in the GATSM see e.g. section IV.C. Typically, the derivation of the Kalman filter stops here, but the Kim filter uses approximations to collapse the filtered values of the state variables to make the Kalman filter operable for moderate sample sizes. The approximations are derived below, however, for a further explanation see Kim and Nelson (1999) and Kim (1994). The idea of the approximations is to reduce the evaluations from growing exponentially fast in time to linearly. To do so, the updated or filtered regimes are used. Specifically, the filtered regimes are weighted by the filtered state variables. The approximations for the filtered state variables are given by:

$$
\begin{aligned}
\mathbf{X}_{t|t}^{(j)} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t}, s_t = j) \\
&= \sum_{i\in\mathcal{S}} \mathbb{P}_{\boldsymbol{\theta}}(s_{t-1} = i|\mathbf{F}_{1:t}, s_t = j)\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i) \\
&= \frac{\sum_{i\in\mathcal{S}} \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:t})}{\mathbb{P}_{\boldsymbol{\theta}}(s_t = j|\mathbf{F}_{1:t})}\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}(\mathbf{X}_t|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i) \\
&\approx \frac{\sum_{i\in\mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}}{\mathcal{Q}_{t|t}^{(j)}}\mathbf{X}_{t|t}^{(i,j)}
\end{aligned}
$$

and the error is given by:

$$
\begin{aligned}
\mathbf{P}_{t|t}^{\mathbf{X}(j)} &= \mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})(\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})'|\mathbf{F}_{1:t}, s_t = j) \\
&= \frac{\sum_{i\in\mathcal{S}} \mathbb{P}_{\boldsymbol{\theta}}(s_t = j, s_{t-1} = i|\mathbf{F}_{1:t})}{\mathbb{P}_{\boldsymbol{\theta}}(s_t = j|\mathbf{F}_{1:t})}\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})(\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})'|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i) \\
&= \frac{\sum_{i\in\mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}}{\mathcal{Q}_{t|t}^{(j)}}\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})(\mathbf{X}_t - \mathbf{X}_{t|t}^{(j)})'|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i)
\end{aligned}
$$

Add and subtract $\mathbf{X}_{t|t}^{(i,j)}$ to get

$$
\mathbf{P}_{t|t}^{\mathbf{X}(j)} = \frac{\sum_{i\in\mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}}{\mathcal{Q}_{t|t}^{(j)}}\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}\Big(\Big((\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)}) - (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})\Big)\Big((\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)}) - (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})\Big)'|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i\Big)
$$

Multiply terms and rearrange to get

$$
\begin{aligned}
\mathbf{P}_{t|t}^{\mathbf{X}(j)} &= \frac{\sum_{i\in\mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}}{\mathcal{Q}_{t|t}^{(j)}} \times \ldots \\
&\Big(\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)})(\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)})'|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i)}_{\approx \mathbf{P}_{t|t}^{\mathbf{X}(i,j)}} - (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)})'|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i)}_{=\mathbf{0}} - \ldots \\
&\underbrace{\mathbb{E}_{\boldsymbol{\theta}}^{\mathbb{P}}((\mathbf{X}_t - \mathbf{X}_{t|t}^{(i,j)})|\mathbf{F}_{1:t}, s_t = j, s_{t-1} = i)}_{=\mathbf{0}}(\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})' + (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})(\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})'\Big)
\end{aligned}
$$

which reduces to

$$
\mathbf{P}_{t|t}^{\mathbf{X}(j)} \approx \frac{\sum_{i\in\mathcal{S}} \mathcal{Q}_{t|t}^{(i,j)}}{\mathcal{Q}_{t|t}^{(j)}}\Big(\mathbf{P}_{t|t}^{\mathbf{X}(i,j)} + (\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})(\mathbf{X}_{t|t}^{(j)} - \mathbf{X}_{t|t}^{(i,j)})'\Big)
$$

These calculations rely heavily on the distributional assumption imposed on the errors $(\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k+p})$. Note, the joint process $\{(\mathbf{X}_t, \mathbf{F}_t)\}_{t=1}^{T}$ is a mixture of Gaussian distributions, but Gaussian distributed conditional on the regimes. This concludes the derivation of the Kim filter.

B.1. Derivation of the backward smoothing algorithm

This section derives the backward algorithm that is used for obtaining the smoothed values of the unobserved regimes. The calculations use the stored filtered and predicted values of the regimes from the Kim filter. Therefore, smoothing presumes that the Kim filter has finished successfully.

Unlike filtering, smoothing uses all the information available from the observed forward rates. The goal is thus to obtain the values for $\mathcal{Q}_{t|T}^{(i)} = \mathbb{P}(s_t = i|\mathbf{F}_{1:T})$. Kim and Nelson (1999) and Hamilton (1994) propose a backward algorithm that can be used for obtaining the smoothed values of the regimes. The algorithm involves approximations which will be clear in a moment. The algorithm is backward in the sense that the iterations run through $t = T-1, ..., 1$ and for i, j in $\mathcal{S} = \{NB, LB\}$. First, $\mathcal{Q}_{t|T}^{(i)}$ is manipulated:

$$\begin{aligned}
\mathcal{Q}_{t|T}^{(i)} &= \mathbb{P}(s_t = i|\mathbf{F}_{1:T}) \\
&= \sum_{j\in\mathcal{S}} \mathbb{P}(s_{t+1} = j, s_t = i|\mathbf{F}_{1:T}) \\
&= \sum_{j\in\mathcal{S}} \mathbb{P}(s_t = i|s_{t+1} = j, \mathbf{F}_{1:T})\mathbb{P}(s_{t+1} = j|\mathbf{F}_{1:T})
\end{aligned}$$

Kim and Nelson (1999) propose the following approximation:

$$\mathbb{P}(s_t = i|s_{t+1} = j, \mathbf{F}_{1:T}) \approx \mathbb{P}(s_t = i|s_{t+1} = j, \mathbf{F}_{1:t})$$

Here, the the entire information about the forward rates are reduced to the current, i.e., from T to t. This involves an approximation as the regimes follow a first-order Markov chain see equation (5.30) on page 107 in Kim and Nelson (1999) for a detailed discussion. Next, the approximation is used in the previous expression:

$$
\begin{aligned}
\mathcal{Q}_{t|T}^{(i)} &\approx \sum_{j\in\mathcal{S}} \mathbb{P}(s_t = i|s_{t+1} = j, \mathbf{F}_{1:t})\mathbb{P}(s_{t+1} = j|\mathbf{F}_{1:T}) \\
&\approx \sum_{j\in\mathcal{S}} \frac{\mathbb{P}(s_t = i, s_{t+1} = j|\mathbf{F}_{1:t})\mathbb{P}(s_{t+1} = j|\mathbf{F}_{1:T})}{\mathbb{P}(s_{t+1} = j|\mathbf{F}_{1:t})} \\
&\approx \sum_{j\in\mathcal{S}} \frac{\mathbb{P}(s_t = i|\mathbf{F}_{1:t})\mathbb{P}(s_{t+1} = j|s_t = i)\mathbb{P}(s_t = j|\mathbf{F}_{1:T})}{\mathbb{P}(s_{t+1} = j|\mathbf{F}_{1:t})}
\end{aligned}
$$

Finally, it remains to insert definitions to obtain:

$$
\mathcal{Q}_{t|T}^{(i)} \approx \mathcal{Q}_{t|t}^{(i)} \sum_{j\in\mathcal{S}} \left\{ \frac{\pi_{i,j}\,\mathcal{Q}_{t|T}^{(j)}}{\mathcal{Q}_{t+1|t}^{(j)}} \right\} \tag{64}
$$

The backward recursions start from $\mathcal{Q}_{T|T}^{(j)}$ and the smoothed values are calculated by iterating backward in (64). This concludes the derivation of the smoothing algorithm.

XI. Appendix: Empirical analysis

A. *Estimated measurement errors in the GATSM and RSTSM*

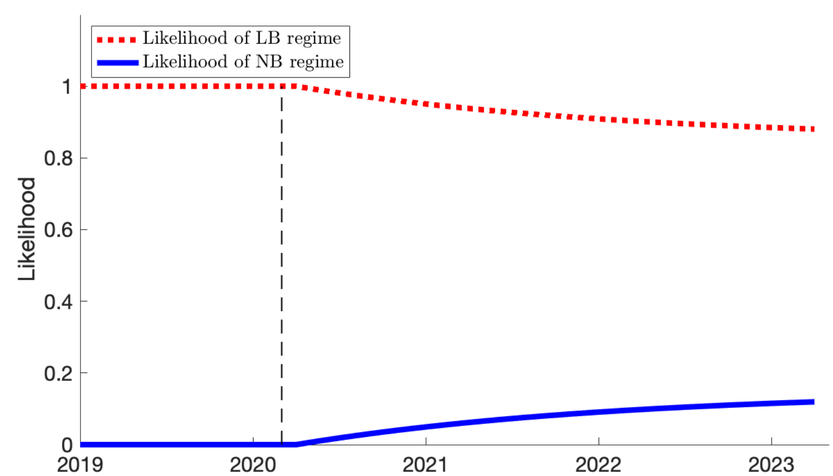
Table VI Maximum likelihood estimates of measurement errors

	GATSM								RSTSM						
	3M	**6M**	**1Y**	**2Y**	**5Y**	**7Y**	**10Y**		**3M**	**6M**	**1Y**	**2Y**	**5Y**	**7Y**	**10Y**
3M	.0000 [.0000]							**3M**	-.0002 [.0019]						
6M	.0001 [2.4178]	.0001 [1.8529]						**6M**	-.0148 [.0092]	-.0124 [.0091]					
1Y	.1062 [.0538]	.0987 [.0822]	.0332 [.0824]					**1Y**	-.0022 [.0226]	.0075 [.0095]	.0238 [.0002]				
2Y	-.1060 [.0103]	-.1730 [.0019]	-.1994 [.0027]	-.1118 [.0179]				**2Y**	-.1428 [.0005]	-.1375 [.0033]	-.0509 [.0000]	.0156 [.0037]			
5Y	.0774 [.0770]	.0601 [.0662]	.0174 [.0118]	.0455 [.0216]	.0299 [.0180]			**5Y**	.0334 [.0000]	.0304 [.0144]	.0308 [.0258]	.0712 [.0000]	.0421 [.0000]		
7Y	.0256 [.0238]	-.0018 [.0739]	-.0591 [.1614]	-.0500 [.1810]	.0834 [.0059]	.0690 [.0619]		**7Y**	-.0533 [.0000]	-.0140 [.0014]	.0449 [.0203]	.0465 [.0000]	.0578 [.0262]	.0490 [.0120]	
10Y	.1149 [.0930]	.1698 [.0848]	.1991 [.0063]	.1403 [.0399]	-.0727 [.1129]	-.1321 [.0966]	-.1037 [.0140]	**10Y**	.0709 [.0135]	.1098 [.0088]	.1321 [.0147]	.0343 [.0043]	-.1361 [.0000]	-.1113 [.0000]	-.0133 [.0010]

Note: Maximum likelihood estimates of the measurement error covariance matrix $\mathbf{\Omega}$, see definition IV.1-IV.2, with robust standard errors in brackets $[\cdot]$. Forward rates from January 2005 to March 2020.

B. *Future evolution of regimes in the RSTSM*

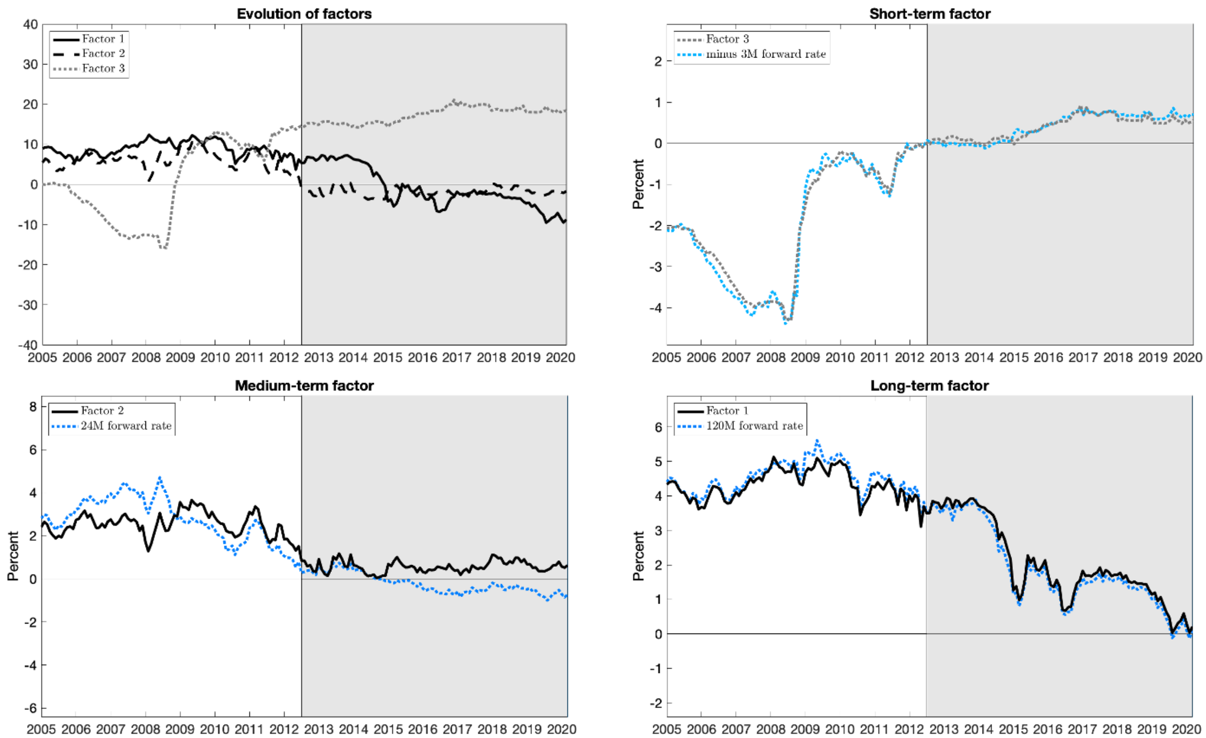
Figure 10. Evolution of regimes in the RSTSM from March 2020 to March 2023



Note: "NB" and "LB" are abbreviations for the regimes; *no* lower bound regime (NB) and lower bound regime (LB). Computations are based on the formulas in Hamilton (1994) and Rahbek and Pedersen (2019b). Vertical dashed line corresponds to March 2020.

C. Filtered state variables in the GATSM

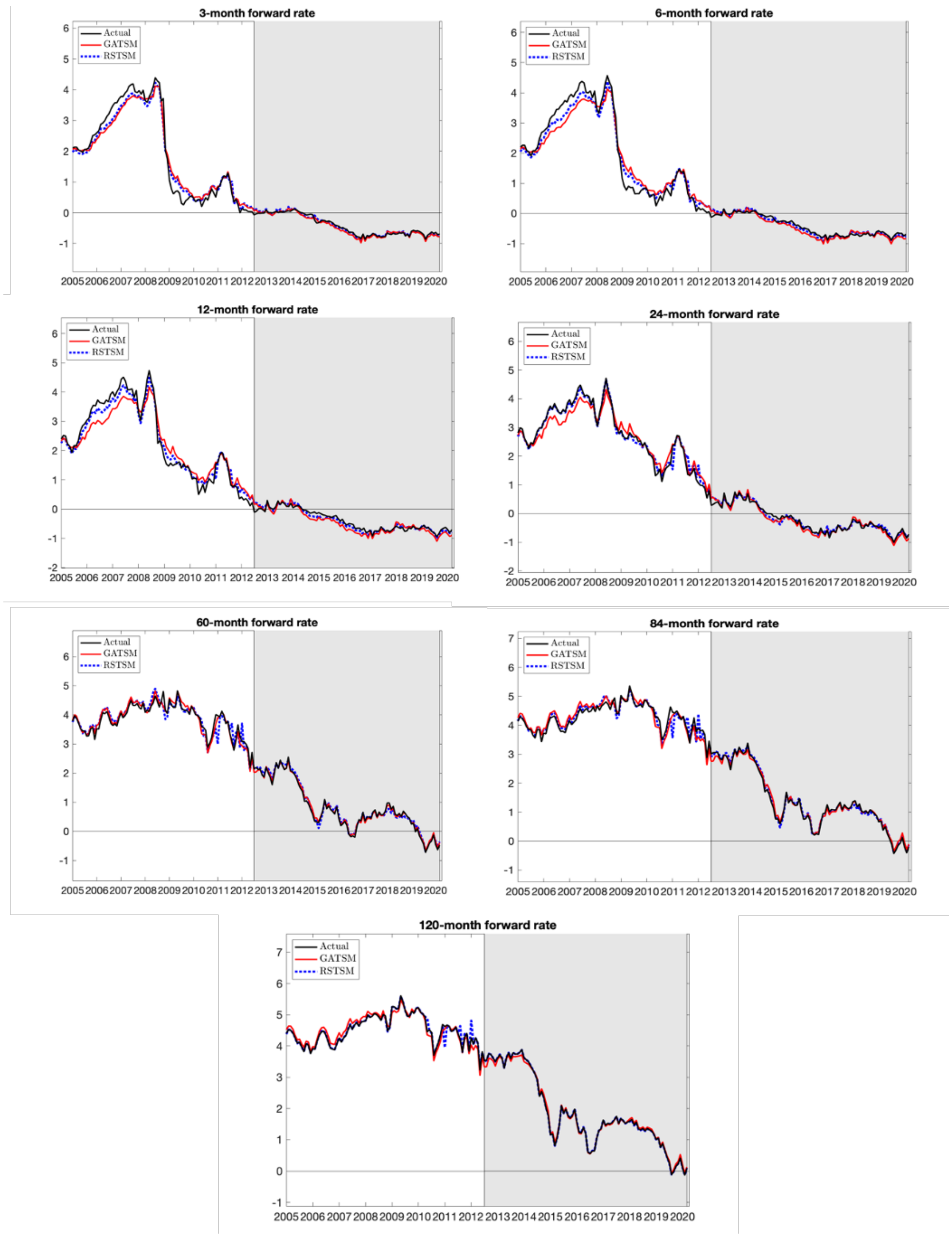
Figure 11. Filtered state variables or factors in the GATSM



Note: Filtered values in the GATSM are obtained by the Kalman filter described in section IV.C. Short-, medium- and long-term factors are scaled to match the mean and range of the forward rate.

D. *Actual versus fitted forward rates in the GATSM and RSTSM*

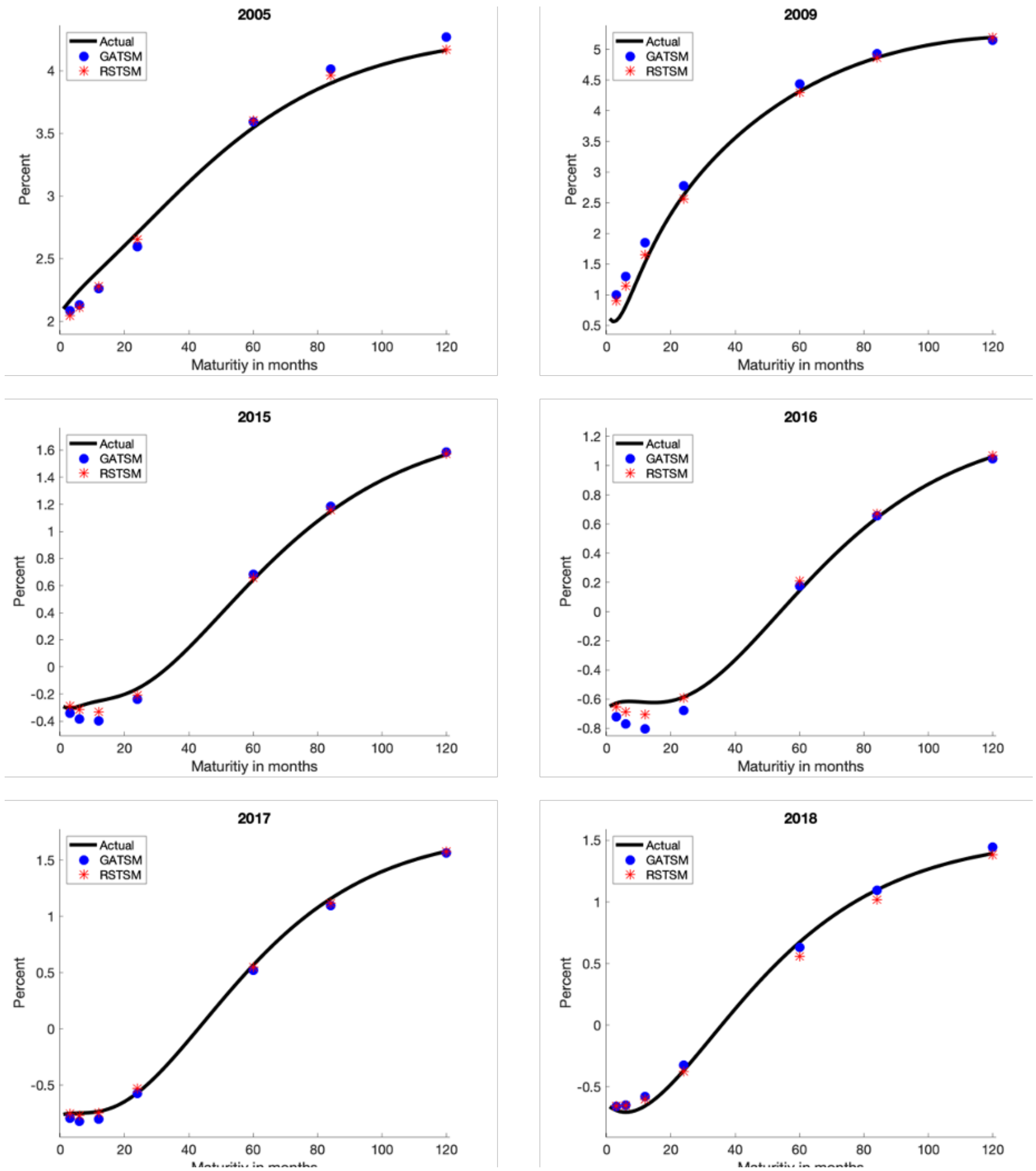
Figure 12. Actual and fitted forward rates from January 2005 to March 2020



Note: Actual euro area forward rates from the ECB, see ECB (2020b). Black line: actual forward rate. Red line: fitted forward rate in the GATSM. Blue dotted line: fitted forward rate in the RSTSM.

E. Average forward curves in the GATSM and RSTSM

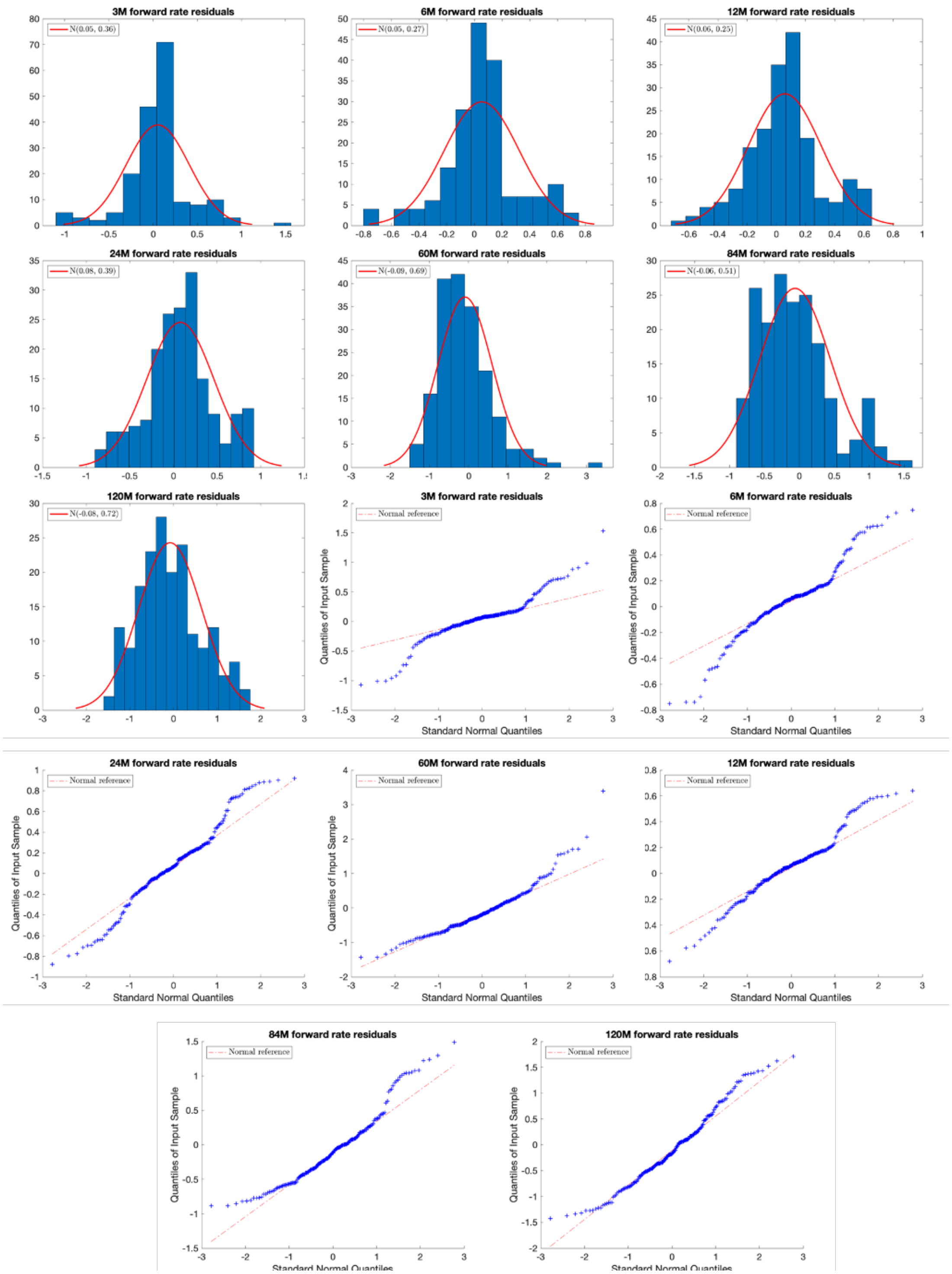
Figure 13. Average forward curves in the GATSM and RSTSM.



Note: Black curve: actual forward rates from ECB (2020b). Blue dots: fitted forward rates in the GATSM. Red dots: fitted forward rates in the RSTSM.

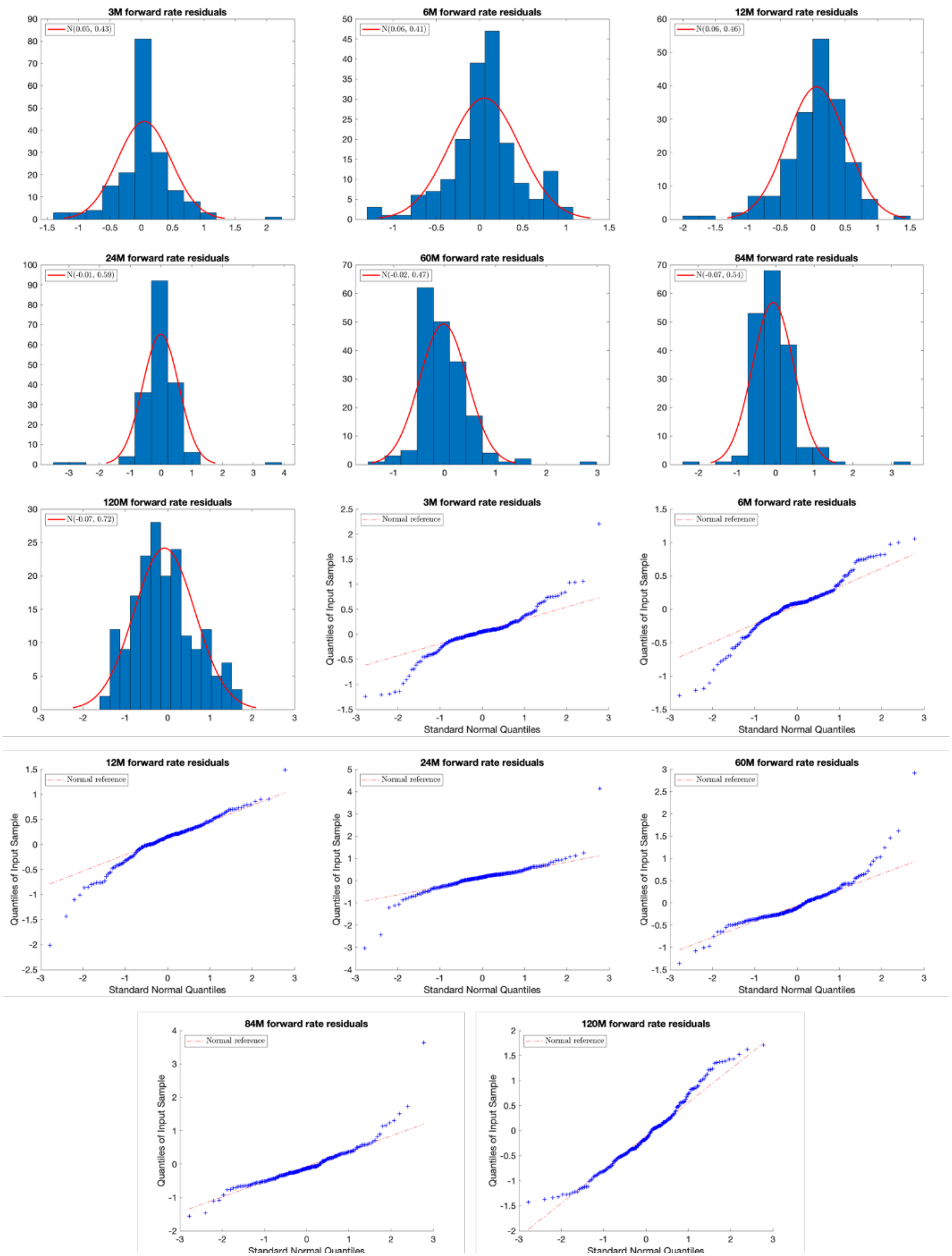
F. *Figures related to misspecification tests*

Figure 14. Visualizations of the standardized residuals in the GATSM



Note: The standardized residuals are obtained by the difference between the actual and fitted forward rates divided by the standard deviation.

Figure 15. Visualizations of the standardized residuals in the RSTSM



Note: The standardized residuals are obtained by the difference between the actual and fitted forward rates divided by the standard deviation.

G. MATLAB code

Table VII Description of the MATLAB-files

Folder	Link
`DATA`	https://www.dropbox.com/sh/9wj0tkb8k2f9i64/AABAk3-ICyd87SJ5PKNn2xy2a?dl=0
`FUNCTIONS`	https://www.dropbox.com/sh/8viieqbmqu8l27r/AAC6h_R0wpUputkf9vRYx1Txa?dl=0
`PARAMETERS`	https://www.dropbox.com/sh/znff5pimlsdjici/AAAjiafiHTftd_To7s4nMg5fa?dl=0
`MAIN`	https://www.dropbox.com/sh/v9un1yizbttcdlp/AAALOPUYE7yckNRsJxPHHz1da?dl=0
`FIGURES`	https://www.dropbox.com/sh/wf35pnfmw39o02a/AABlVLJCMwwzsYSh512N4W_Ja?dl=0

Folder	Description
`DATA`	Contains all data used in the empirical analysis section V and in the monetary policy section II.C.
`FUNCTIONS`	`KF_GATSM.m` and `KIM_RSTSM.m` are functions based on the econometric methodology in section IV.
	`gradient.m` and `hessian.m` are used to construct robust standard errors.
`PARAMETERS`	`gatsm.mat` and `rstsm.mat` are maximum likelihood parameter estimates.
`MAIN`	`gatsm.m` and `rstsm.m` estimate the models, report tables and plot figures in the empirical analysis section V.
`FIGURES`	Contains all figures in `png.`-format.

Note: In MATLAB, all codes have been written from scratch as no existing solutions or packages were available. The codes can be executed without specific packages. All codes are written by myself.