DYNAMIC TERM STRUCTURE MODELING UNDER MACROECONOMIC-DRIVEN REGIME SHIFTS

by

Emile A.L.J. van Elen† (74.30.15)

B.Sc. Tilburg University 2010

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Quantitative Finance and Actuarial Science

Tilburg School of Economics and Management

Tilburg University, the Netherlands

Supervisors

Prof.dr. R.J. Mahieu

Prof.dr. J.M. Schumacher

August 8, 2012

Abstract

This thesis proposes a regime-switching extension of the dynamic Nelson-Siegel and Björk-Christensen term structure models. A first-order Markov process with time-varying transition probabilities governs the switches. The transition probabilities, in turn, are driven by three leading economic indicators. The models are estimated through an application of the Kim filter using U.S. zero-coupon yields. As such, two distinct regimes are identified which are interpreted as a regime of normal economic activity and a regime of aberrant activity. The regime-switching models are benchmarked against the random walk and compete with various existing, single-regime models in both an in-sample and out-of-sample study. It follows that the single-regime models and their regime-switching counterparts fit the data equally well. In addition, the regime-switching models provide superior forecasts in some instances, particularly at the short end of the yield curve. Furthermore, inclusion of macroeconomic data is beneficial to interest rate forecasting at forecasting horizons of six months and beyond.

Key words: Term structure; Nelson-Siegel model; Regime switching; Time-varying transition probabilities; Kim filter; Forecasting

Contents

Acknowledgments

I would like to express my sincere gratitude to my thesis supervisor, prof.dr. R.J. Mahieu, for the continuing support and confidence he has given me to successfully complete this project. His extended knowledge on a whole array of topics, innovative mindset, enduring patience and, above all, his kindness have been a true blessing. Thank you.

I would like to extend my appreciation to the chairman of my graduation committee, prof.dr. J.M. Schumacher. His unbounded dedication to academia and profound sense of humor made his classes particularly pleasant to attend. In addition, I will not forget that it were his efforts that enabled me to spend a semester at University of South Carolina.

I acknowledge the help rendered by prof.dr. R.R. Bliss of Wake Forest University in his willingness to share his interest rate data.

I am highly indebted to my family and in particular my parents, Ton and Anneke. I owe them more than I can say and I deeply thank them for their unconditional support. I thank my sister, Eva, as well. Her facetiousness has brought me many great moments of relief during this project.

My sincere thanks are also due to my dear friend, Stefan Damen. Not only has he offered much-needed comments on preliminary versions of this thesis, also has he proven to always be there for me when requested upon. Working with him, throughout our entire studies, has been truly great and I hope the future enables us to do so once more.

Last, but most certainly not least, I am grateful to my beloved girlfriend, Ireen Christiaans. I am confident our future together will be just as bright as it has been during the past nine years.

Thank you all.

1 Introduction

An adverse macroeconomic climate, lack of trust and political impotence are just a few of the frequently-mentioned factors that are currently endangering the existence of Europe's monetary union. Given some other major developments prior to the unfolding of the Eurozone issues - in particular the 2007 U.S. housing bubble and the credit crunch - investors remain wary and tend to stay away from anything that even smacks of risk. At the beginning of 2012, for instance, investors demanded a premium in excess of thirty percent on the 10-year bonds of troubled Greece over the safely regarded 10-year German bunds. For comparison, the spread oscillated around two percent in 2009. The limited number of economists that foresaw such sizable market movements - not intended as an objurgation of any kind - exemplifies the highly unpredictable nature of interest rates. This is particularly uncomforting to e.g., bond portfolio managers, pension funds, homeowners and risk managers to whom (accurately predicted) interest rates are of key importance. A great deal of research is hence concerned with fitting and forecasting the yield curve. This thesis aims to contribute to that by proposing a regime-switching term structure model and subsequently benchmarking it against various competing models to evaluate fitting capabilities as well as out-of-sample forecasting accuracy.

1.1 Brief literature review

During the last three decades or so major advancements have been made in the field of term structure modeling. Despite the efforts, however, no superior model exists as is the case with e.g., option pricing theory where the Black and Scholes (1973) model became the generally accepted standard. Rather, the existing term structure literature has diverged into three strands.

One strand takes a theoretical approach and rests on the principles of the general equilibrium framework. Seminal contributions include Vasicek (1977) and Cox, Ingersoll and Ross (1985) who both propose models that are functions of the instantaneous short rate. These so-called affine models particularly gained in popularity due to Duffie and Kan (1996) who generalized this class of models. For forecasting purposes affine models turn out to be less suited, as shown by Duffee (2002). In addition, estimation of these models is generally problematic "because of the existence of numerous likelihood maxima that have essentially identical fit to the data but very different implications for economic behavior" (Christensen, Diebold and Rudebusch, 2011).

A second strand of literature, which adapts only the basics of the equilibrium framework, aims to impose no-arbitrage restrictions - imperative to derivatives pricing - by optimally fitting the empirical data from a cross-sectional perspective. A prominent contribution in that respect is e.g., Hull and White (1990) who propose to extend Vasicek's model with time-varying parameters. Because of the cross-sectional focus these no-arbitrage models generally tend to produce inaccurate forecasts of future interest rates as well which is e.g., shown by Duffee (2011).

A third strand of literature builds on the seminal work of Nelson and Siegel (1987) (hereafter NS) who introduce a parsimonious three-factor model that turns out to fit the yield curve remarkably well. Four- and five-factor extensions have subsequently been proposed by e.g., Svensson (1994) and Björk and Christensen (1999) (hereafter BC). Interest in this particular class increased substantially after Diebold and Li (2006) reinterpreted the original NS model as a dynamic factor model. They showed that their dynamic extension forecasts the future yield curve more accurate than various competing models - including a random walk - at multiple forecasting horizons. De Pooter (2007) studies in-sample and out-of-sample performance of several Nelson-Siegel type of models. Whereas he finds the model of Svensson (1994) to show the best in-sample fit (albeit marginally better than the BC model), he finds the four-factor BC model to provide the most accurate interest rate forecasts at various forecasting horizons. Recently, Christensen et al. (2011) proposed an arbitrage-free NS model and as such their work bridges the gap between theory and practice. Numerous studies, however, show that no-arbitrage restrictions - also with regard to the NS class - are actually irrelevant for forecasting purposes; see for instance, Coroneo, Nyholm and Vidova-Koleva (2008), Duffee (2011) and Joslin, Singleton and Zhu (2011).

During the last decade the field of term structure modeling further developed by linking the yield curve to the macroeconomy (the models of which are referred to as macro-finance models going forward). Most prominently, Ang and Piazzesi (2003), who consider macroeconomic data in conjunction with a three-factor affine model, find inclusion of macroeconomic variables to be beneficial to interest rate forecasting. A notable study within the NS class in that respect is Diebold, Rudebusch and Aruoba (2006) who extend the state vector with various economic variables. Other interesting contributions, on the basis of Nelson-Siegel type of models as well, are e.g., De Pooter, Ravazzolo and Van Dijk (2010) and Huse (2011).

Due to its intuitive appeal the field of term structure modeling also reached out to models that account for a multi-regime setting. That is to say, the shape of the yield curve may change materially for a prolonged period of time e.g., due to fiscal policy. Seminal contributions are Gray (1996), Ang and Bekaert (2002) and Bansal and Zhou (2002) who all model the short rate as a regime-switching process. These studies - and many others e.g., Dai, Singleton and Yang (2007) - find strong evidence for the presence of multiple regimes in term structures of interest rates.

Recent studies - see Zantedeschi, Damien and Polson (2011) for an overview - have looked into the possibilities of combining regime-switching models with macro-finance models. Zhu and Rahman (2009), for instance, successfully extend the NS model to a regime-switching model and, similar to Diebold et al. (2006), add various macroeconomic variables to the state vector. As such, they find "significant bidirectional linkages between the yield curve and economic activity.” Bernadell, Coche and Nyholm (2005) extend the dynamic NS model by assuming that the constant of the slope factor follows a first-order Markov-switching process. In particular, a state transition in their model occurs if certain subjective cut-off values of macroeconomic indicators are exceeded. They find their regime-switching model to significantly outperform the single-regime NS models, particularly at forecasting horizons beyond 24 months.

1.2 Research description

By virtue of the recent encouraging advancements in the field of term structure modeling, this thesis aims to contribute to the literature by extending the model of Bernadell et al. (2005) in two directions. First, I extend their model by adopting a time-varying transition probability (hereafter TVTP) matrix that is driven by leading macroeconomic indicators. As such, regime switches are endogenous as opposed to the exogenous, superimposed thresholds of the economic indicators as in Bernadell et al. (2005). The explicit link that arises between the yield curve and the macroeconomy makes this macro-finance model particularly suited to e.g., aid strategic investment decisions. That is, on the basis of subjective beliefs about the future state of the economy investors are able to construct forecasts of the future term structure (Huse, 2011).

Second, I also deploy this time-varying regime-switching extension to the four-factor Björk-Christensen model. Interest in the BC model is particularly motivated by the encouraging out-of-sample results of De Pooter (2007), as outlined above. To study the effect of the TVTPs and, ultimately, the effect of the macroeconomic variables, I also consider these NS and BC switching extensions with fixed transition probabilities (hereafter FTPs). In-sample and out-of-sample performance of these particular model specifications has not been examined before and could hence be an interesting direction to explore.

I use the Kim (1994) filtering procedure to estimate the regime-switching models. This procedure, which heavily leans on the pioneering work of Kalman (1960) and Hamilton (1989, 1994), entails a maximization of an approximate log-likelihood function. I consider various competing models including a random walk and the single-regime NS and BC models with both autoregressive- and vector autoregressive-specified state vectors. Following De Pooter (2007), I estimate the single-regime models using a two-step procedure as well as a one-step procedure. In so doing, the effect of the estimation procedure of the single-regime models can be examined as well.

The random walk aside, a total of twelve term structure models is thus considered: two regime-switching yields-only (FTP) models, two regime-switching macro-finance (TVTP) models, four single-regime two-step models and four single-regime one-step models. Given the nested nature of some of these models, I test their significance by virtue of the likelihood ratio test (LRT). I use this test as well to obtain some degree of statistical confidence of the presence of multiple regimes.

In light of the above, the central research question of this thesis is:

How does in-sample and out-of-sample performance of the proposed regime-switching extensions of the Nelson-Siegel and Björk-Christensen models compare to the performance of existing, single-regime term structure models?

The data comprise U.S. zero-coupon yields spanning the period from January, 1962 up until and including December, 2008. Time-varying transition probabilities of the macro-finance models are based on three leading macroeconomic indicators related to GDP, inflation and housing starts.

With regard to in-sample fitting capabilities, I particularly follow De Pooter (2007) by reporting various error statistics at selected maturities and by fitting the yield curve at several dates. I study out-of-sample performance - at forecasting horizons of 1, 6, 12 and 24 months - by simulating potential future yield movements on the basis of boostrapped historical yields. To preserve the typical dynamic evolution of the historical yields, I use the Moving Block Bootstrap method of Künsch (1989) to bootstrap blocks of yields of length 24 months rather than single months. Using (trace) root mean squared prediction errors and techniques from Diebold and Mariano (1995), I evaluate out-of-sample performance. In addition, I perform a robustness check to evaluate the sensitivity of the block length.

The main results i.e., results related to the regime-switching models, are as follows. The regime-switching models identify two intuitively interpretable, distinct regimes. The first regime corresponds to a regime with an upward sloping yield curve and may hence be interpreted as a regime where economic activity is normal. The second regime shows a flattish yield curve which matches with a central bank's active, intervening policy of an increased (short-term) rate to e.g., control inflation. As such, this regime constitutes a period of aberrant economic activity. Filtered state probabilities, generally, match reasonably well with NBER recession periods. The time-varying transition probability denoting the probability to stay within the aberrant regime is decreasing with time during NBER recessions, which is intuitively appealing as well. The identified regimes are persistent in the sense that (average) transition probabilities are above 90 percent for all models.

LRT results, first and foremost, imply that the regime-switching yields-only models (i.e., those characterized by FTPs) are rejected in favor of their macro-finance counterparts (i.e., those based on TVTPs) at all conventional significance levels. Second, the four-factor BC model is strongly favored in lieu of the three-factor NS model. In addition, the results suggest that a second regime is existent indeed.

Error statistics indicate that in-sample fit of the single-regime and multi-regime models is to a large extent equal. With regard to out-of-sample performance, all four regime-switching models are able to outperform a random walk at various maturities and forecasting horizons. Outperformance generally increases when the forecasting horizons lengthens and particularly occurs at the short end of the yield curve. No model, however, including the benchmark models, consistently outperforms the random walk. I find that inclusion of macroeconomic data is beneficial to term structure forecasting at horizons of six months and beyond. Numerous times, the yields-only regime-switching models as well as their macro-finance counterparts provide the most accurate forecasts of all twelve models. This encouraging result implies it is worthwhile to fit and forecast the term structure of interest rates on the basis of regime-switching (macro-finance) models.

The remainder of the thesis is structured as follows. Section 2 introduces notation and elaborates on some of the concepts related to term structures of interest rates. Section 3 discusses the class of Nelson-Siegel models, the regime-switching extensions are presented and motivated in Section 4. Section 5 presents the data. I outline the estimation procedure in Section 6, which also reports corresponding estimation results and likelihood ratio test results. Section 7 evaluates in-sample fit and Section 8 examines out-of-sample performance. Concluding remarks and some directions for future research are offered in Section 9.

2 Concepts and notations

This section introduces the interest rate framework and notations that are used throughout the thesis. I briefly discuss various methods that allow for the construction of zero-coupon yields and motivate why the unsmoothed Fama-Bliss estimation method is preferred.

2.1 Term structure of interest rates

Term structures of interest rates describe the cross-sectional relation between bonds with a certain maturity and the corresponding interest rate. The term structure may be represented by e.g., the yield curve, the discount curve and the forward curve. Whenever the term structure is represented by either one of those curves, other representations may be deduced from the relations below.[1]

Assume continuous compounding throughout the section. At a given time t, let $P_t(\tau_i)$ denote the price of a discount bond with maturity time τ_i for $i \in \{1, 2, \ldots, N\}$. That is, the present value of an asset at time t that pays 1 at maturity, which is τ_i periods ahead. Let $y_t(\tau_i)$ denote the corresponding zero-coupon nominal interest rate i.e., the yield to maturity. The relation between the yield curve and the discount curve can be described by

$$P_t(\tau_i) = e^{-y_t(\tau_i)\tau_i} \quad \Longleftrightarrow \quad y_t(\tau_i) = -\frac{1}{\tau_i}\log\big(P_t(\tau_i)\big). \tag{2.1}$$

The short rate, denoted by r_t, may be defined as an annualized interest rate for an infinitesimally small time window. That is,

$$r_t \equiv y_t(0) = \lim_{\tau_i \to 0} y_t(\tau_i).$$

In practice, the three-month rate turns out to be a better proxy for the short rate than e.g., overnight loans because such loans (i.e., loans with very short maturities) are affected by factors that yield curve models, in general, do not aim to cover (Schumacher, 2011).

Given the discount or yield curve, one may find an expression for the forward curve, $f_t(\tau_i)$ using

$$f_t(\tau_i) = \frac{-P_t'(\tau_i)}{P_t(\tau_i)} = y_t(\tau_i) + y_t'(\tau_i)\tau_i.$$

The relation between the yield curve and the forward curve is given by

$$y_t(\tau_i) = \frac{1}{\tau_i}\int_0^{\tau_i} f_t(u)du, \tag{2.2}$$

[1] For an extensive discussion on the numerous representations of the term structure of interest rates, refer to e.g., Svensson (1994) and Schumacher (2011).

which suggests that the yield curve may be interpreted as a weighted average of forward rates with different maturities.

2.2 Unsmoothed Fama-Bliss yields

The zero-coupon term structure representations in (2.1) - (2.2) yield the advantage that it allows for a direct comparison of term structures of e.g., a country over the course of time or of different countries/companies at the same time. Indeed, term structures based on coupon-bearing bonds are sensitive to so-called 'coupon effects': Caks (1977) as cited in De Pooter (2007) points out that bonds that have different coupon rates but are identical otherwise may have different yields. One initial challenge that arises with the representations in (2.1) - (2.2) is that zero-coupon yields are generally unobservable in financial markets for maturities of over twelve months. Therefore, one has to resort to estimation techniques to derive a full representation of the zero-coupon yield curve.

One such estimation technique rests on parametric (or function-based) estimation, most notably those on the basis of the Nelson-Siegel model and extensions thereof. Since this thesis is centered around Nelson-Siegel type of models, I choose not to estimate zero-coupon yields using this estimation method to prevent potentially biased results. Another strand of methods rests on spline methods. For instance, McCulloch (1975) proposes to estimate the discount curve using cubic splines. The method turns out, however, to poorly fit flattish curves resulting from a diverging discount curve at longer maturities. This caveat is overcome by Vasicek and Fong (1982) who apply an exponential spline. Unfortunately, this technique suffers from another serious drawback, for forward rates are not strictly positive in some cases. Fama and Bliss (1987) introduce a third method that does not suffer from these issues. In particular,

> *"Their method sequentially constructs the forward rates necessary to price successively longer-maturity bonds, often called an 'unsmoothed Fama-Bliss' forward rates, and then constructs 'unsmoothed Fama-Bliss yields' by averaging the appropriate unsmoothed Fama-Bliss forward rates."* (Diebold and Li, 2006)

Assuming a constant forward rate between different maturities, the method of Fama and Bliss (1987) computes prices for all included bonds. Numerous studies e.g., Diebold and Li (2006), De Pooter (2007), Koopman, Mallee and Van der Wel (2010) and Huse (2011) hence make use this estimation technique, as does this thesis.

3 Nelson-Siegel models

Starting with the original formulation of Nelson and Siegel (1987), this section discusses several Nelson-Siegel type of models. I discuss a reformulated version as proposed by Diebold and Li (2006), giving the model factors an intuitive interpretation. Moreover, I touch upon the four-factor Björk and Christensen (1999) extension and the Dynamic Nelson-Siegel model in state-space form. The section ends with a discussion on absence of arbitrage in relation to the class of Nelson-Siegel models.

3.1 Classical Nelson-Siegel model

Nelson and Siegel (1987) step aside from traditional financial theory and aim to introduce

> *"A simple, parsimonious model that is flexible enough to represent the range of shapes generally associated with yield curves: monotonic, humped, and S shaped."* (Nelson and Siegel, 1987)

Recall from Section 2 the maturity time, τ_i for $i = 1, 2, \ldots, N$. Then, the Nelson-Siegel forward curve is given by

$$f(\tau_i) = b_1 + b_2 e^{-\lambda\tau_i} + b_3 \lambda\tau_i e^{-\lambda\tau_i}. \tag{3.1}$$

Here, $b_1,\ b_2,\ b_3$ and λ are the model's parameters. Through an application of integration equation (2.2) one obtains the following expression of the yield curve:

$$y(\tau_i) = b_1 + b_2 \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - b_3 e^{-\lambda\tau_i}. \tag{3.2}$$

The model presented in (3.2) turns out to fit the observed term structure remarkably well. As a result, a diversity of financial institutions such as investment banks, pension funds and central banks use the Nelson-Siegel model, or an extension thereof, to model the term structure of interest rates. For instance, De Pooter (2007) indicates that nine out of thirteen central banks that report their term structure estimations to the Bank of International Settlements (BIS) use a Nelson-Siegel type of model.[1]

Despite the model's success from a cross-sectional point of view, the Nelson-Siegel model does a poor job when it comes to out-of-sample time series forecasts because in practice $b_1,\ b_2,\ b_3$ and λ tend to vary over time. Diebold and Li (2006) suggest to overcome this caveat by dynamically extending the model i.e., they introduce a model that allows the model's parameters to change at each time t.

[1]Moreover, Coroneo et al. (2008) state that the European Central Bank models term structures using an extended Nelson-Siegel model.

Diebold and Li (2006), moreover, suggest a rearrangement of the factors of the model in (3.2) which gives the factors a clear and distinct economic interpretation. Allowing now for time-dependency as discussed above, Diebold and Li (2006) suggest to let $b_{1,t} = \beta_{1,t}$, $b_{2,t} = \beta_{2,t} + \beta_{3,t}$ and $b_{3,t} = \beta_{3,t}$. In so doing, (3.2) may be written as

$$\begin{aligned} y_t(\tau_i) &= \beta_{1,t} + \left(\beta_{2,t} + \beta_{3,t}\right)\frac{1 - e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} - \beta_{3,t} e^{-\lambda_t \tau_i} \\ &= \beta_{1,t} + \beta_{2,t}\frac{1 - e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} + \beta_{3,t}\left(\frac{1 - e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} - e^{-\lambda_t \tau_i}\right). \end{aligned} \tag{3.3}$$

The model in (3.3) is referred to as the Dynamic Nelson-Siegel (DNS) model. $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ can be interpreted as a level factor, a slope factor and a curvature factor, respectively. To see this, first consider Figure 3.1 which shows how a given term structure can be broken down into these three factors.

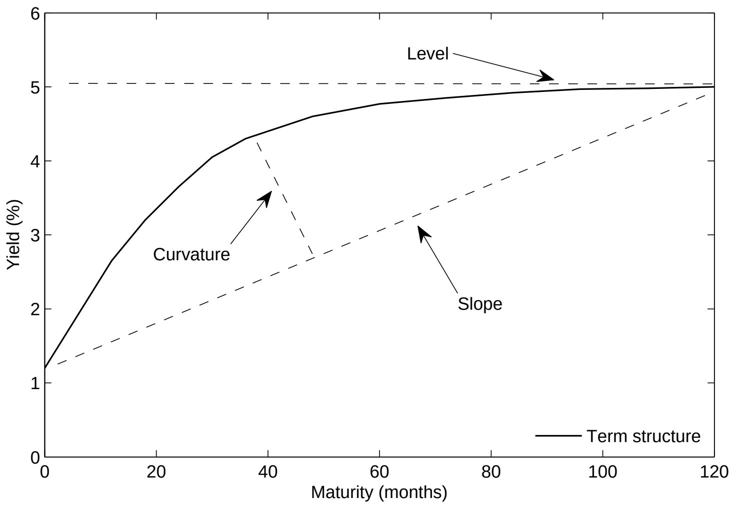

Figure 3.1: **Example yield curve broken down into three factors.**

Note that the loading on $\beta_{1,t}$ equals 1, a constant independent of maturity. Hence, $\beta_{1,t}$ affects the term structure equally for different maturities, meaning it can be interpreted as a long-term (level) factor indeed. The loading on $\beta_{2,t}$, $(1 - e^{-\lambda_t \tau_i})/\lambda_t \tau_i$, approaches 0 when $\tau_i \to \infty$ and 1 when $\tau_i \to 0$. Therefore, one might say that $\beta_{2,t}$ affects the term structure primarily in the short run i.e., it affects the slope of the curve. $\beta_{3,t}$'s loading equals $\left(1 - e^{-\lambda_t \tau_i}\right)/\lambda_t \tau - e^{-\lambda_t \tau_i}$, which, as a function of maturity, starts at a level of 0, gradually increases for some time, only to return to 0 later on. Therefore, $\beta_{3,t}$ affects the yield curve in the medium term primarily; Diebold and Li (2006) hence refer to $\beta_{3,t}$ as a curvature factor.

Figure 3.2 shows the loadings on $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ as a function of maturity from which the limiting behavior of the model can easily be deduced:

$$\lim_{\tau_i \to 0} y_t(\tau_i) = r_t = \beta_{1,t} + \beta_{2,t}, \qquad \lim_{\tau \to \infty} y_t(\tau_i) = l_t = \beta_{1,t}.$$

Note that r_t and l_t may be interpreted as a short rate and a long rate, respectively.

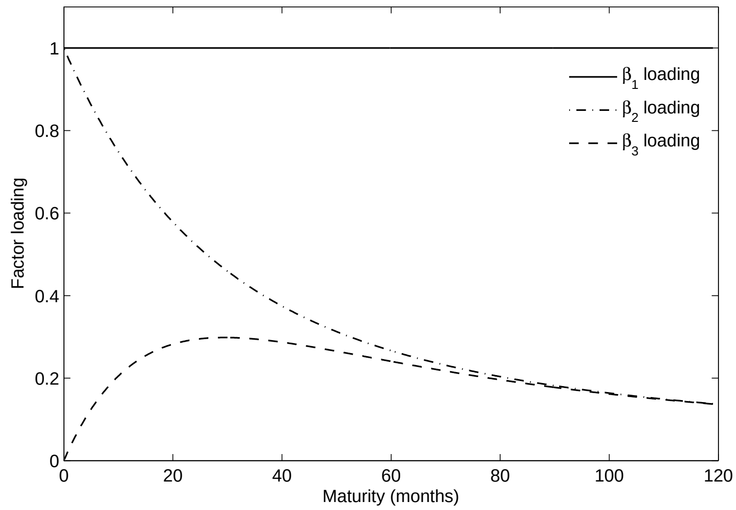

Figure 3.2: **Factor loadings on the parameters of the Nelson-Siegel model ($\boldsymbol{\lambda = 0.0609}$).**

The fourth parameter, λ_t indicates the rate of exponential decay. That is, a slow decay of the yield curve corresponds to a small value for λ_t, resulting in a better fit of the yield curve at long maturities (the converse statement also holds true). λ_t, moreover, determines the maximum location of the loading on the curvature factor, $\beta_{3,t}$. Diebold and Li (2006) argue that this parameter might as well be taken as a constant with little degradation of fit. In so doing, the estimation procedure of $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ is simplified because (3.3) reduces to a linear regression. In particular, Diebold and Li (2006) fix the decay parameter at 0.0609 with maturities measured in months, their argumentation for which is as follows. Since λ_t determines the maximum location of the loading on the curvature (medium-term) factor, Diebold and Li (2006) consider the maximization on typical medium-term maturities: two and three years (i.e., 24 and 36 months, respectively). By taking the average of this, 30 months, they find that the value for λ_t that maximizes the loading on $\beta_{3,t}$ at 30 months equals 0.0609.[2] Koopman et al. (2010), on the contrary, treat the decay parameter as a fourth latent (time-varying) model factor and find this to significantly increase in-sample fit. In this thesis I follow Diebold and Li (2006) by assuming a fixed decay parameter going forward.[3]

[2]Based on unsmoothed Fama-Bliss zero yields constructed using end-of-month price quotes for U.S. treasuries spanning the period from January, 1985 up until and including December, 2000.

[3]As such, I drop the subscript t for the decay parameter, λ.

3.2 Dynamic Nelson-Siegel model in state-space form

In addition to the suggestion to dynamically extend the Nelson-Siegel model, Diebold and Li (2006) propose to model the evolution of $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ using two benchmark models: a first-order vector autoregressive process for the three factors (i.e., VAR(1)) and three univariate autoregressive processes (i.e., AR(1)). Note that, whenever the factors follow an autoregressive process, the model forms a state-space system. Following Diebold et al. (2006) in particular, the state-space model with a VAR(1) representation of the factors at time $t = 1, 2, \dots, T$ is given by

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix}, \tag{3.4}$$

$$\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{pmatrix} + \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \upsilon_{1,t} \\ \upsilon_{2,t} \\ \upsilon_{3,t} \end{pmatrix}. \tag{3.5}$$

In vector notation, the state-space equations (3.4) - (3.5) are esthetically more appealing:

$$y_t = H\beta_t + \epsilon_t, \tag{3.6}$$

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + \upsilon_t, \tag{3.7}$$

where y_t is a vector of dimension $N \times 1$ and β_t a vector of dimension 3×1.[4]

Depending on the estimation method, assumptions have to be made with regard to the error terms. E.g., the Kalman filter (more on which in due course) assumes the disturbances, ϵ_t and υ_t in (3.6) and (3.7) to be white Gaussian noise and to be mutually uncorrelated:

$$\begin{pmatrix} \epsilon_t \\ \upsilon_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix} \right]. \tag{3.8}$$

Moreover, it is assumed that the error terms are orthogonal to the initial state vector:

$$\mathbb{E}(\beta_0 \epsilon_t') = 0, \qquad \mathbb{E}(\beta_0 \upsilon_t') = 0. \tag{3.9}$$

Throughout the thesis I assume the matrix R to be diagonal, implying that yields are uncorrelated for different maturities.[5] Assumptions about the diagonality of Q (i.e., whether the factors are allowed to be correlated) are often driven by computational feasibility. That is,

[4] In the more general case, assume β_t to have dimension $K \times 1$.

[5] Diebold et al. (2006) remind the reader this is a common assumption when estimating no-arbitrage term structure models.

an AR(1) model is often preferred over a VAR(1) model to prevent the number of parameters from growing too large. Moreover, it is sometimes argued (e.g., Diebold and Li (2006)) that allowing for factor correlation results in poor forecasts of economic variables. A VAR(1) model, on the other hand, allows for additional model flexibility. For benchmark purposes I include both AR(1) and VAR(1) models.

The state-space formulation of the DNS model as formulated above is actually quite flexible. One could, for instance, extend the number of state equation factors to four or even five. Frequently cited extensions in that respect are the Svensson (1994) and the Björk and Christensen (1999) models, the latter of which is to be discussed next.

3.3 Four-factor Björk-Christensen model

De Pooter (2007) compares a variety of Nelson-Siegel type of models and concludes that a fourth factor not only increases flexibility but also adds to the model's forecasting power. In particular, he finds the four-factor Björk-Christensen model to outperform other models in an out-of-sample study.[6] As such, the BC model is an interesting model to study in conjunction with the three-factor NS model.

Björk and Christensen (1999) suggest to add an additional factor to the NS forward curve from (3.1) as follows:

$$f_t(\tau_i) = \beta_{1,t} + \beta_{2,t} e^{-\lambda \tau_i} + \beta_{3,t} \lambda \tau_i e^{-\lambda \tau_i} + \beta_{4,t} e^{-2\lambda \tau_i}. \tag{3.10}$$

An advantageous economic implication of this forward expression is that it is consistent with the Hull and White (1990) extension of the Vasicek (1977) model, whereas (3.1) is not.[7]

One obtains an expression of the corresponding yield curve by integrating (3.10) according to (2.2):

$$y_t(\tau_i) = \beta_{1,t} + \beta_{2,t} \frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} + \beta_{3,t} \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} - e^{-\lambda \tau_i} \right) + \beta_{4,t} \frac{1 - e^{-2\lambda \tau_i}}{2\lambda \tau_i}.$$

Recall how the loading on $\beta_{2,t}$, the slope factor, affects the short run of the yield curve and note the similarity of the loadings on $\beta_{2,t}$ and $\beta_{4,t}$. Indeed, the difference in loadings merely stems from a scaling by a factor 2 in the exponent and the denominator, implying that the loading on the fourth factor decays to 0 at a faster rate. This behavior is depicted in Figure 3.3. Therefore, $\beta_{4,t}$ may be interpreted as a second slope factor and as such the BC model captures the term structure's slope by some combination of $\beta_{2,t}$ and $\beta_{4,t}$.

[6] De Pooter (2007), moreover, reports that the BC model outperforms the three-factor NS model in sample.

[7] Let $\mathcal{M}$ be a term structure model (e.g., the Hull and White (1990) model) and let $\mathcal{C}$ be a parametrized class of forward curves (e.g., the Nelson-Siegel class of forward curves). Björk and Christensen (1999) define the pair $(\mathcal{M}, \mathcal{C})$ to be *consistent* if all forward curves that can be produced by the model $\mathcal{M}$ are contained within $\mathcal{C}$, given the initial forward curve is within $\mathcal{C}$.

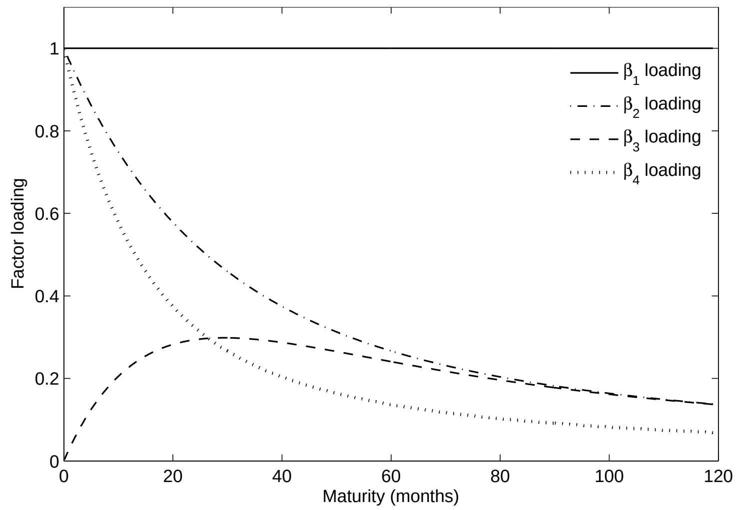

Figure 3.3: **Factor loadings on the parameters of the four-factor Björk-Christensen model ($\lambda = 0.0609$).**

From Figure 3.3 it can also be seen that the instantaneous short rate, r_t at a given time t now depends on three parameters:

$$\lim_{\tau_i \to 0} y_t(\tau_i) = \beta_{1,t} + \beta_{2,t} + \beta_{4,t}.$$

The long rate, l_t of the Nelson-Siegel and the four-factor Börk-Christensen models are the same.

The state-space formulation of (3.6) - (3.7) remains valid for the BC model if the underlying matrices are adapted accordingly. In particular, $\beta_t = \left(\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t}\right)^T$ and

$$H = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} & \frac{1-e^{-2\lambda\tau_1}}{2\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} & \frac{1-e^{-2\lambda\tau_2}}{2\lambda\tau_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} & \frac{1-e^{-2\lambda\tau_N}}{2\lambda\tau_N} \end{pmatrix}.$$

Restrictions on the errors terms as formulated in e.g., (3.8) and (3.9) are to be imposed as well.

3.4 Arbitrage-Free Dynamic Nelson-Siegel model

The class of Nelson-Siegel models gained in popularity for the reasons discussed above i.e., the models fit well cross-sectionally, they can be estimated in a straightforward manner and the factors have sound economic interpretations. Moreover, NS type of models are relatively accurate when it comes to forecasting as studied by e.g., Diebold and Li (2006) and De Pooter (2007). Despite these merits, however, the model, by construction, does not impose the necessary restrictions on the parameters that would prevent the model from arbitrage opportunities. This observation, as shown by e.g., Björk and Christensen (1999) implies that the dynamic evolution of the model's parameters under the risk neutral measure is not necessarily consistent with the transformation of those parameters to construct the yield curve under the empirical measure.

Recently, Christensen et al. (2011) proposed an adjusted Nelson-Siegel model and proved that their model does satisfy the no-arbitrage conditions. This model, the Arbitrage-Free Nelson-Siegel (AFNS) model, is characterized by a correction term which is only dependent of maturity. Christensen et al. (2011) show that the correction term has little impact on the shape of the yield curve for short maturities. The correction does increase with maturity but remains small: it is circa 50 (90) basis points for a model with independent (correlated) factors at a 30-year maturity, the longest maturity considered by Christensen et al. (2011).

The relevance of the exclusion of arbitrage opportunities to term structure forecasting (and hence to this thesis) has been extensively studied. Most notably, Duffee (2011) finds that cross-sectional restrictions resulting from the no-arbitrage property are not helpful in a forecasting setting. If cross-sectional restrictions are applied, however, then they can be used to impose dynamic restrictions on the compensation of risk. Duffee (2011) finds the dynamic restrictions to be useful but hastens to conclude that they can be imposed without the no-arbitrage property as well. In particular,

> *"The restrictions on a VAR implied by an arbitrage-free Gaussian dynamic term structure model cannot be rejected against the alternative of an unrestricted VAR."* (Duffee (2011) as cited in Huse (2011))

Christensen et al. (2011) are somewhat less definite and argue that forecasting gains could potentially materialize from no-arbitrage restrictions. They, however, also state that their

> *"Evidence [regarding out-of-sample performance] is much less conclusive than for in-sample fit."* (Christensen et al., 2011)

Joslin et al. (2011) observe that imposing no-arbitrage restrictions is irrelevant for forecasting purposes and that forecasts of the AFNS factors are equivalent to forecasts of parameters of an unrestricted VAR(1). Coroneo et al. (2008) show that the parameters of the Nelson-Siegel model are not statistically different from those obtained from the AFNS formulation at the 95% confidence level. They also find that the original Nelson-Siegel model performs equally well in an out-of-sample forecasting study as its no-arbitrage counterpart.[8]

Given these findings and bearing the aim of this thesis in mind, I proceed without imposing no-arbitrage conditions.

[8] Coroneo et al. (2008) base their findings on a working paper version of Christensen et al. (2011), published in 2007 by the Federal Reserve Bank of San Francisco (Working paper No. 2007-20).

4 Regime-switching term structure models

Being an integral part of the thesis, this section proposes extensions of the dynamic NS and BC models by allowing for a regime switch in the state equations. The models presented in this section have a direct link - and hence an intuitive interpretation - to (different states of) the economy. Therefore, the section also aims to motivate the particular model choices, which are mainly driven by standard economic theory.

Two of the proposed models endogenously link the yield curve's shape to the economy. This particularly materializes by adopting a time-varying transition probability matrix based on various macroeconomic variables. The concept of linking the economy to the yield curve is not new in itself. Zantedeschi et al. (2011) list an overview of the main macro-finance models that have been presented in the literature. All these models, original and contributive in their owns ways, take different approaches from the approach considered in this thesis.

The section first proceeds with a discussion on the regime-switching setting that is applied. The regime-switching yields-only models, which are based on fixed transition probabilities, are subsequently introduced. The models on the basis of time-varying transition probabilities, i.e., the regime-switching macro-finance models, are introduced thereafter.

4.1 A two-regime setting: tractable and economically sound

The notion of the existence of multiple regimes in an economy is intuitively appealing. Distinctive regimes that readily come to mind involve different states of the economy and as such regimes are intimately related to the business cycle. One might think of e.g., significant changes in the slope of the yield curve or changes in debt capital markets' volatility. The case in favor of regime switching finds support in numerous empirical studies. Amongst others, Ang and Bekaert (2002), Bansal and Zhou (2002) and Dai et al. (2007) find strong evidence that multiple regimes are existent in term structures. Some even argue, e.g., Zhu and Rahman (2009), that different regimes are a stylized fact of the term structure.

In this vein, I consider a two-state setting where one regime is characterized by a normal level of economic activity and another where economic activity is aberrant.[1] One could allow for more than two regimes; such model setup, however, would not only lack a sound economic interpretation, it would also seriously jeopardize the model's tractability. That is, it would imply a substantial increase in the number of parameters, making it computationally intense and time consuming to estimate all parameters involved. Statistical significance of the estimates is likely to decrease as well.

Regime switching can be implemented in different ways e.g., by the construction of an independent switching model where it is assumed that transitioning to a certain regime is independent of previous regimes. This thesis, however, considers a more advanced regime

[1] Section 6.6 tests the significance of the second regime.

switch - Markov switching - which allows for dependency on previous states. In particular, I focus on a first-order Markov-switching process, meaning that the probability to be in a certain state at time t depends only on the state at time $t-1$ i.e., the previous state. On a technical note, let state $S_t \in \{1, 2\}$ be a discrete hidden (i.e., unobserved) variable at time t. The transition probability matrix, P that governs the variable S_t is given by

$$P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}. \tag{4.1}$$

Here, $p_{ij} = \mathbb{P}[S_t = j \mid S_{t-1} = i, S_{t-2} = h, \ldots] = \mathbb{P}[S_t = j \mid S_{t-1} = i]$ i.e., the probability to switch from state i to state j. In the two-state case, it holds that $p_{i1} + p_{i2} = 1$ for $i \in \{1, 2\}$ since, w.p. 1, one either stays in the current state or switches to the other. This reduces the number of parameters of P to be estimated to two instead of four. The fixed transition probability matrix in (4.1) is to be used to extend the single-regime NS and BC models to regime-switching yields-only models, which will be discussed next.

4.2 Regime-switching yields-only model

4.2.1 Model formulation

Although one could incorporate switching elements in both the observation and the state equation, it is sensible to switch within the state equation only. This stems from observing that once the loading matrix, H is estimated, the yield curve is fully determined by the model's factors (i.e., the state equation). Put differently, shaping the yield curve comes down to a transformation of the factors once the loading matrix is determined.

Set in state-space form, consider the observation equation of the NS model at time $t = 1, 2, \ldots, T$ as in (3.6):

$$y_t = H\beta_t + \epsilon_t.$$

Let S_t denote state dependency and define the regime-switching state equation by

$$\beta_t = \tilde{\mu}_{S_t} + F\beta_{t-1} + \nu_t, \tag{4.2}$$

where

$$\tilde{\mu}_{S_t} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_{2,S_t} \\ \tilde{\mu}_3 \end{pmatrix}, \qquad F = \begin{pmatrix} f_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f_{33} \end{pmatrix}. \tag{4.3}$$

That is, the state vector is assumed to switch in the mean of $\beta_{2,t}$ (i.e., the slope factor) only, as proposed by Bernadell et al. (2005).[2] The level and curvature factors follow AR(1) processes. The error terms, ϵ_t and υ_t, are assumed to follow normal distributions and to be mutually uncorrelated:

$$\begin{pmatrix} \epsilon_t \\ \upsilon_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix}\right]. \tag{4.4}$$

Here, both R and Q are assumed diagonal. Furthermore, it is assumed that the first-order Markov-switching process, S_t adopts the transition matrix in (4.1).

Incorporating this particular regime switch in the Björk-Christensen model yields another interesting model, which I consider as well. I assume that $\beta_{4,t}$ also follows an AR(1) process; (4.3) thus becomes

$$\tilde{\mu}_{S_t} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_{2,S_t} \\ \tilde{\mu}_3 \\ \tilde{\mu}_4 \end{pmatrix}, \qquad F = \begin{pmatrix} f_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & f_{33} & 0 \\ 0 & 0 & 0 & f_{44} \end{pmatrix}.$$

Note that, in so doing, the vector notation of the state-space formulation does not change. Underlying assumptions with regard to the error terms are similar to those of the NS regime-switching formulation.

4.2.2 Model motivation

The representation of the state equation in (4.2) is mainly driven by economic theory in general and the Taylor principle in particular. Taylor (1993) as cited in Bernadell et al. (2005) argues how central banks should respond to macroeconomic changes like increased inflation or lower than expected GDP growth. In a regime characterized by high inflation, for instance, a central bank should act on this by increasing its policy rate (which is a short term rate). As a result, the slope of the yield curve is to decrease, corresponding to a flattish (or even inverse) yield curve. In another regime, where inflationary levels are 'in control' and where GDP growth is anywhere between modest to high, a central bank will either lower the policy rate to give a positive impulse to the economy or change it only marginally, if at all. In this regime the yield curve is hence expected to be (moderately) upward sloping.

This principle rests on the notion that the long rate (i.e., the yield curve's level) is relatively more stable than the short rate. The Fisher decomposition of nominal yields supports this notion. That is, the nominal interest rate approximately equals the sum of the expected

[2] It is noted Bernadell et al. (2005) actually propose a three-state model, where the different regimes are a priori and exogenously determined by subjective cut-off levels of economic activity (GDP) and inflation (CPI). Since this thesis aims to switch between regimes endogenously, I proceed using two states as motivated in the introduction of this section.

real interest rate and the inflation rate. In the long run the expected real rate equals the growth of the economy. In the short run the expected real rate is driven by many other factors than just the economy's growth, implying a less stable rate. The rate in the long run is hence expected to be more stable indeed (Bernadell et al., 2005). The notion, moreover, finds empirical support in actual yield curve data as short rates turn out to be more volatile than long rates; this will be further discussed in Section 5.

Put short, the model specification as such gives rise to two economic regimes. The first regime is characterized by a 'relatively normal' economy in the sense that GDP is (slowly) increasing and, accordingly, the yield curve is upward sloping. A second regime may be classified as 'aberrant' which may e.g., constitute a period of abnormal inflation rates.

4.3 Regime-switching macro-finance model with time-varying transition probabilities

4.3.1 Model formulation

Relaxing the assumption of fixed transition probabilities and allowing for time-varying transition probabilities instead further increases model flexibility. More importantly, it also enables one to endogenously govern the discrete hidden variable, S_t on the basis of additional explanatory variables e.g., leading economic indicators. In so doing, the time-varying transition probability matrix governing S_t becomes

$$P_t = \begin{pmatrix} p_{11,t} & p_{21,t} \\ p_{12,t} & p_{22,t} \end{pmatrix} = \begin{pmatrix} p_{11,t} & 1 - p_{22,t} \\ 1 - p_{11,t} & p_{22,t} \end{pmatrix}, \tag{4.5}$$

where $p_{ij,t} = \mathbb{P}[S_t = j \mid S_{t-1} = i, z_{t-1}]$ and where $z_{t-1} = (1, z_{1,t-1}, \ldots, z_{k,t-1})'$, a vector containing k additional explanatory variables. This thesis follows Diebold, Lee and Weinbach (1994) and Filardo (1994) who propose a logistic function of the k variables.[3] That is,

$$\mathbb{P}[S_t = j \mid S_{t-1} = j, z_{t-1}] = \frac{\exp(a_j \cdot z_{t-1})}{1 + \exp(a_j \cdot z_{t-1})}, \qquad j \in \{1, 2\}. \tag{4.6}$$

Note that $a_{S_t} = (a^0_{S_t}, a^1_{S_t}, \ldots, a^k_{S_t})$, a vector of constants of size $(k+1) \times 1$ for $S_t \in \{1, 2\}$. These constants, which are subject to estimation, determine the weights of the variables collected in z_{t-1}. Its first element, $a^0_{S_t}$ is not affiliated with an explanatory variable i.e., it is a constant that provides additional fit. When $a^1_{S_t} = a^2_{S_t} = \ldots = a^k_{S_t} = 0$ one obtains the FTP model and as such the TVTP models nest their FTP counterparts.

The state-space representation of the yield curve from Section 4.2 remains in effect; I consider the NS model as well as the BC model. Since these particular regime-switching

[3] The logistic transformation ensures that transition probabilities lie between zero and one. Alternatively, one could deploy a probit function as in e.g., Gray (1996).

macro-finance models have not been proposed in the literature before it will be interesting to examine their in-sample and out-of-sample performance.

4.3.2 Model motivation

Modeling time-varying transition probabilities is intuitively appealing for at least three reasons. First, assume that, without loss of generality, regimes have a finite duration and that a potential regime-switch may take place at each time period. Given a currently active regime, the probability this regime remains the active regime in e.g., twenty time periods is lower than the probability this regime remains active in e.g., only two periods. This line of reasoning particularly makes sense from an economic point of view: in an ongoing economic crisis the probability that the crisis lasts is likely to decrease with time, implying indeed that TVTPs should not be excluded a priori.

Second, a currently prevailing interest regime might well transition to another regime due to the unfolding of new (public) information. One might particularly think of prolonged changes of leading, influential economic indicators such as consumer confidence, inflation and GDP. That is to say, central banks may alter their policy rates for extended periods of time on the basis of a materially different state of the economy; responses from financial markets to the unfolding information can be assumed to influence the shape of the yield curve as well. Hence, an explicit link arises between the macroeconomy, the yield curve and transition probabilities of interest rate regimes.

Third, regime-switching macro-finance models enable one to examine the sensitivity of the yield curve to economic indicators. In particular, one can construct a prediction of the future yield curve on the basis of forecast economic indicators. This might, for instance, be relevant to strategic investors

> *"Who are interested in forecasting bond prices and might have a better idea of the expected state of the economy than the expected state of the yield curve. Moreover, it is also of value to financial authorities, as a tool to asses financial stability."* (Huse, 2011)[4]

A successful implementation of a regime-switching model where TVTPs are driven by economic variables is due to Filardo (1994). He finds the Hamilton model with FTPs to be a poor forecaster of NBER recession cycles. Therefore, Filardo (1994) proposes the usage of TVTPs as in (4.6), where a large set of leading (macro)economic indicators is included. In so doing, he links the business cycle in a two-regime setting (growth and recession) to the macroeconomy and finds a high correlation between NBER recessionary periods and the inferred state probabilities.

[4] Huse (2011) employs a single-regime macro-finance NS term structure model. His reasoning, for that matter, also applies to the regime-switching macro-finance models proposed in this thesis.

Recent studies reveal that inclusion of macroeconomic data in itself results in more accurate interest rate forecasts as well.[5] Most notably, Ang and Piazzesi (2003) examine a three-factor macro-finance affine term structure model and conclude that inclusion of economic data increases predictive accuracy. Moreover, Diebold et al. (2006), who extend the state vector of an NS model with economic data, find strong evidence that changes in the macroeconomy affect the future term structure. De Pooter et al. (2010) (who propose combining yields-only models with macro-finance models) and Huse (2011) also find that macro-finance models are better forecasters in an out-of-sample study.

Now the regime-switching models are introduced, they have to be estimated first before in-sample and out-of-sample performance can be examined. An introduction of the data that are used to obtain parameter estimates is, therefore, lined up next.

[5]The great majority of those studies are based on single-regime models as well.

5 Data

THE ZERO-COUPON INTEREST RATE data and their characteristics are discussed in this section. Macroeconomic data, used to infer time-varying transition probabilities, are presented as well.

5.1 Zero-coupon yields

The thesis uses unsmoothed end-of-month U.S. zero-coupon yields, spanning the period from January, 1962 up until and including December, 2008 to estimate the models.[1] All yields are continuously compounded and presented on an annualized basis. Yield data are constructed using forward rates through an application of (2.2). These forward rates are extracted from the Center for Research in Security Prices (CRSP) and are based on filtered prices (average bid-ask) on U.S. bonds using the Fama and Bliss (1987) bootstrap method.

I follow Diebold and Li (2006) and De Pooter (2007) by cross-sectionally restricting the number of maturities to $N = 17$. In particular, I consider yields with a maturity of $\tau_i = 3$, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. The period 1962:1 - 2008:12 corresponds to $T = 564$ observations. Combined with the 17 maturities, this amounts to a total of 9,588 data points.[2]

Figure 5.1 presents a three-dimensional plot of the zero-coupon data. The figure shows a large variation of yields over the course of time, largely due to central banking policy and major economic events. The unprecedented high rates during the Volcker period in the early 1980s, for instance, are quite different form the lower level yields some ten years before. Also clearly visible is the sharp decline in interest rates after the burst of the dot-com bubble. Note how the extremely low interest rates from 2007 onwards reflect the current challenging economic environment, initiated by the U.S. housing bubble and succeeded by a global credit crunch.

Table 5.1, reporting descriptive statistics of the zero-coupon yields, reveals several stylized facts that are typically associated with interest rate data. It follows, for instance, that the average yield curve is upward sloping and that short rates are more volatile than long rates. Moreover, the short end of the curve seems to be less persistent (i.e., lower autocorrelation) than the long end. On a more general note, yields are skewed to the right, regardless the maturity. This gives rise to a probability distribution having more probability mass in the right tale. Furthermore, yields are leptokurtic which suggests that the yield distribution is characterized by thick tales.

[1] I kindly thank prof. Robert R. Bliss for sharing this data.

[2] To guarantee coverage of the entire yield curve I interpolate a small percentage (341 of 9,588 observations i.e., 3.56%) of the dataset. Interpolation is only required at yields with maturities of 84, 96, 108 and 120 months and only before 1971. Amongst others, Bernadell et al. (2005) revert to similar (interpolation) techniques to complete their data. Note that a complete dataset facilitates estimation.

To get a feel for the yield curve's factors the table also reports descriptive statistics for the level, slope and curvature. Following Diebold and Li (2006), I define the level as the 10-year yield, the slope as the 10-year yield minus the 3-month yield and the curvature as two times the 2-year yield minus the sum of the 3-month and 10-year yields. Note how the stylized facts match the interpretation of the NS and BC model factors. That is, recall the limiting behavior of the NS and BC models as discussed in Section 3.[3] The stylized fact that short rates tend to be more volatile appears in the factors through the dependency of the short rates on both $\beta_{1,t}$ and $\beta_{2,t}$, whereas the long rates only depend on $\beta_{1,t}$. A similar reasoning may be applied to the yield curve's persistency: the stylized fact that long rates are more persistent than short rates comes about by observing from Table 5.1 that the yield curve's level (interpreted as $\beta_{1,t}$), is the most persistent amongst the defined level, slope and curvature.

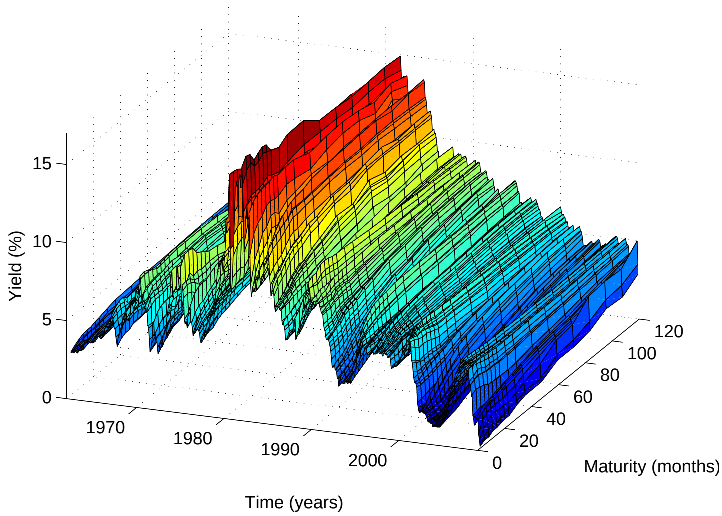

Figure 5.1: **3D plot of U.S. zero-coupon yields.**

Notes: The figure presents a three-dimensional plot of the end-of-month unsmoothed U.S. zero-coupon yields (expressed in percentages). The sample runs from January, 1962 up to and including December, 2008 (564 observations).

[3]The reasoning that follows is based on the NS model. However, since the long-run limiting behavior of the NS and BC models are the same, the story goes for both models.

Figure 5.2 graphs the median yield curve and the 5th, 25th, 75th and 95th percentiles. Judging from the 95th percentile, the graph confirms the observation that the yield curve is skewed to the right. Also note the gap between the 5th and 25th percentiles and observe it is not as wide as the gap between the 75th and 95th percentiles.

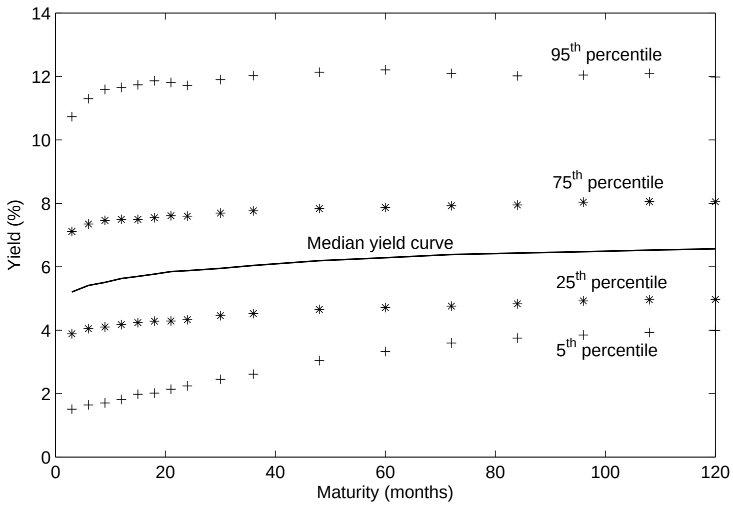

Figure 5.2: **Median yield curve along with several percentiles.**

Notes: See Figure 5.1 for details.

Table 5.1: **Descriptive statistics of U.S. zero-coupon yields.**

Maturity (months)	Mean	Std. dev.	Median	Minimum	Maximum	Skewness	Kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3	5.655	2.815	5.205	0.041	16.019	1.018	4.684	0.978	0.744	0.508
6	5.859	2.842	5.413	0.254	16.481	0.962	4.439	0.979	0.759	0.535
9	5.963	2.842	5.508	0.218	16.393	0.918	4.281	0.979	0.769	0.556
12	6.041	2.805	5.631	0.410	15.822	0.854	4.068	0.979	0.774	0.571
15	6.101	2.796	5.696	0.407	16.042	0.818	3.934	0.980	0.785	0.595
18	6.160	2.789	5.771	0.476	16.229	0.830	3.930	0.981	0.791	0.608
21	6.216	2.780	5.849	0.545	16.177	0.843	3.923	0.981	0.796	0.620
24	6.241	2.738	5.879	0.538	15.649	0.831	3.842	0.981	0.797	0.629
30	6.332	2.682	5.948	0.837	15.397	0.800	3.696	0.981	0.806	0.650
36	6.412	2.654	6.041	1.000	15.765	0.841	3.752	0.982	0.813	0.664
48	6.549	2.603	6.193	1.029	15.820	0.856	3.671	0.982	0.822	0.686
60	6.632	2.556	6.284	1.576	15.005	0.880	3.559	0.984	0.837	0.708
72	6.741	2.536	6.386	1.577	14.979	0.885	3.517	0.984	0.847	0.725
84	6.804	2.479	6.433	1.786	14.974	0.909	3.556	0.984	0.842	0.724
96	6.881	2.456	6.476	2.335	14.935	0.907	3.461	0.986	0.858	0.743
108	6.927	2.449	6.525	2.248	15.017	0.933	3.531	0.986	0.859	0.744
120 (level)	6.927	2.392	6.566	2.935	14.925	0.945	3.578	0.986	0.849	0.738
Slope	1.271	1.351	1.226	-3.505	4.542	-0.257	3.045	0.938	0.417	0.018
Curvature	-0.099	0.763	-0.029	-2.761	3.169	-0.226	4.281	0.843	0.365	0.159

Notes: The table presents descriptive statistics (expressed in percentages) of end-of-month unsmoothed U.S. zero-coupon yields. The sample runs from January, 1962 up to and including December, 2008 (564 observations). $\hat{\rho}(i)$ denotes sample autocorrelation with a time lag of i months. The table also reports the yield curve's level, slope and curvature; refer to the text for their exact definitions.

5.2 Macroeconomic indicators

Inclusion of macroeconomic data allows one to endogenously link the macroeconomy to the term structure. For this to be effective it is imperative the 'proper' macroeconomic indicators are considered, which could e.g., be guaranteed by including a large quantity of economic variables. Unfortunately, doing so would also imply a substantial increase in the number of parameters and would hence be detrimental to the estimation procedure. Therefore, I limit the number of economic indicators to three. One could argue that a mere three indicators do not capture the full dynamics of the economy. However, they should be able to give a rough indication and as such it suits the purpose of the thesis.[4]

On the premises of standard economic theory and existing literature e.g., Huse (2011), I include indicators concerned with inflation, economic growth and the housing market. I obtain data from the Federal Reserve Bank of St. Louis' database, FRED.[5] For the sake of consistency, I consider annualized statistics on a monthly basis over the same sample period as the unsmoothed Fama-Bliss yields i.e., from January, 1962 up until and including December, 2008. To indicate inflation I consider year-on-year (y-o-y) changes of the monthly U.S. Consumer Price Index (CPI). More specifically, I consider the CPI for all urban customers for all items (`CPIAUCSL`). I consider y-o-y changes of real GDP to indicate economic growth (`GDPC1`). This seasonally and inflation-adjusted statistic is given on a quarterly basis. Therefore, I follow Bernadell et al. (2005) and assume equal GDP figures for the months within each quarter. To capture movements in the housing market I consider y-o-y changes of the seasonally adjusted monthly statistic `HOUST`. It reports the number of new privately owned U.S. housing units started in a given month. Given the 2007 housing crisis, it is of particular interest to include this indicator.

Figure 5.3 depicts time series of monthly y-o-y changes of the three leading macroeconomic indicators. It is interesting to observe the unprecedented inflationary levels in the early 1980s and how the FED under the ruling of Paul Volcker successfully combated it. What is also interesting is that the business cycle of economic growth becomes readily apparent from the GDP series. The vertical axes reveal that housing starts fluctuate much more than the other two indicators. It is notable if not disturbing to see how housing starts remained stable for a decade, only to come down during the 2007 burst of the U.S. housing bubble.

Table 5.2 reports corresponding descriptive statistics. It follows that the y-o-y changes of all three statistics are positive on average; negative changes also occur, to be seen from the minimums. The table, moreover, corroborates that housing starts tends to fluctuate most. Autocorrelations at different displacements reveal that CPI is the most persistent indicator.

[4]In fact, a great amount of research is concerned with the identification of key economic indicators. See, for instance, Stock and Watson (1989) who identify the commercial paper spread as a key indicator for business cycle forecasting. Identification of key indicators for term structure fitting and forecasting purposes is an interesting topic of research. It is, however, also beyond the scope of this thesis.

[5]Federal Reserve Economic Data. See `http://research.stlouisfed.org/fred2/`.

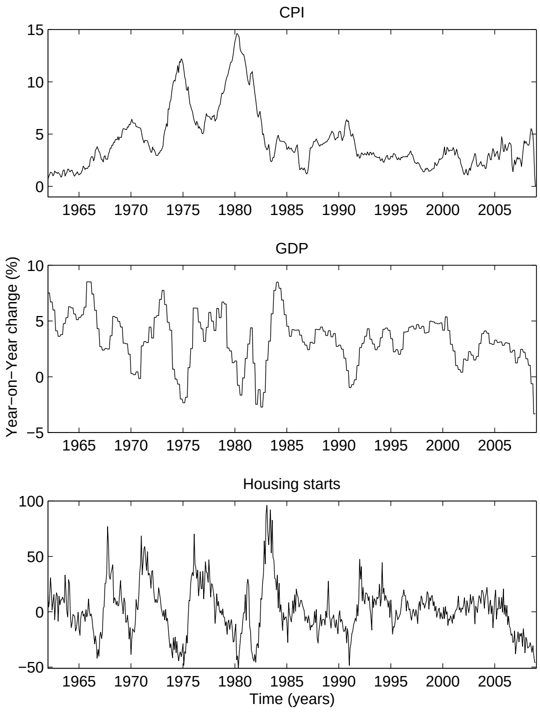

Figure 5.3: **Time series of leading U.S. macroeconomic indicators.**

Notes: The figure presents time series (year-on-year percentage change) of three leading macroeconomic indicators. The monthly sample runs from January, 1962 up to and including December, 2008 (564 observations).

Table 5.2: **Descriptive statistics of leading U.S. macroeconomic indicators.**

	Mean	Std. dev.	Median	Minimum	Maximum
CPI	4.328	2.873	3.486	-0.021	14.592
GDP	3.301	2.282	3.438	-3.321	8.510
Housing starts	1.668	22.737	0.751	-50.586	96.189
	Skewness	Kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
CPI	1.515	4.944	0.988	0.755	0.473
GDP	-0.365	3.303	0.947	0.160	-0.111
Housing starts	0.654	4.582	0.856	-0.002	-0.164

Notes: The table reports descriptive statistics (year-on-year percentage change) of three leading macroeconomic indicators. The monthly sample runs from January, 1962 up to and including December, 2008 (564 observations). $\hat{\rho}(i)$ denotes sample autocorrelation with a time lag of i months.

6 Parameter estimation

This section presents estimation procedures of the regime-switching models and of competing, single-regime models. Estimation results - including time series of filtered state probabilities and time-varying transition probabilities - are reported as well. Significance of various models and the second regime is evaluated using likelihood ratio tests.

6.1 General outline of estimation procedure

There exist different methods as to how the class of Nelson-Siegel term structure models may be estimated. A common two-step approach, due to Diebold and Li (2006), applies Least Squares in the first step to cross-sectionally estimate the measurement equation's parameters; the second step concerns the estimation of a time series model of the state equation. A somewhat more tedious method is a one-step approach. It is due to Diebold et al. (2006) and uses the Kalman (1960) filter. De Pooter (2007) finds this one-step procedure to show better out-of-sample results than the two-step procedure. Moreover,

> *"The simultaneous estimation of all parameters [i.e., the one-step procedure] produces correct inference via standard theory."* (Diebold et al., 2006)

The Kalman filter can be extended such that regime-switching models in state-space form may be estimated as well. Hence, the one-step estimation procedure is strongly preferred. Estimation of regime-switching models in general rests on the principles of the Hamilton (1989, 1994) filter. Kim and Nelson (1999) apply the Kalman and Hamilton filters accordingly and extensively outline a procedure - the Kim (1994) filter - to estimate regime-switching models in a state-space environment. I apply that procedure to estimate the regime-switching models.

Kim and Nelson (1999) actually present different methods as to how regime-switching models can be estimated. For instance, one could revert to Gibbs sampling - a Bayesian approach - as in e.g., Zhu and Rahman (2009). Bayesian statistical estimation methods, however, are generally based on prior knowledge or beliefs about the parameters. That is, it is not preferred, in general, to limit parameter flexibility a priori to a subjective belief. Furthermore, Bayesian posterior densities may be subject to high degrees of complexity and may become intractable as such. In addition, research has shown, see e.g., Kim and Schmidt (2000), that Bayesian estimation results in conjunction with non-informative priors do not substantially differ from classical estimation results. In light of the above, I apply the classic estimation approach by directly maximizing the (approximate) log-likelihood function.[1]

Due to the models' highly parametrized character, sensible initial parameters are of key importance to the one-step procedure. Failure to do to may result in suboptimal solutions.

[1] The Bayesian approach is, however, a welcome alternative in case the number of observations is limited, as argued in Zhu and Rahman (2009). Fortunately, this thesis' dataset comprises sufficient observations, making it worthwhile to resort to the more traditional (i.e., classic) methodologies indeed.

Therefore, I do consider the Diebold and Li (2006) two-step estimation procedure and use the resulting estimates to initiate one-step estimation procedures of the single-regime and regime-switching models. In addition, I deploy the two-step models as competitor models to the regime-switching models. As a sidetrack, it will be interesting as well to look into potential in-sample and out-of-sample differences between the two estimation methods.

To maintain overview it is worthwhile to briefly summarize the models (and their corresponding estimation methods) discussed thus far. This is done in Table 6.1. All these models will be estimated and eventually judged on their in-sample and out-of-sample performances. Within the table, NS2-VAR, for instance, refers to the Nelson-Siegel model that is estimated in two steps and allows the factors to be correlated. The Björk-Christensen model estimated using the Kalman filter and with a restricted state equation is identified as BC1-AR. The model labeled as BC1-RS-X refers to the regime-switching macro-finance (i.e., based on time-varying transition probabilities) Björk-Christensen model.

Table 6.1: **Overview of presented term structure models.**

Label	Model description	Section
NS2-AR	2-step estimation of NS model, AR(1) for βs	6.2
NS2-VAR	2-step estimation of NS model, VAR(1) for βs	6.2
BC2-AR	2-step estimation of BC model, AR(1) for βs	6.2
BC2-VAR	2-step estimation of BC model, VAR(1) for βs	6.2
NS1-AR	1-step estimation (Kalman) of NS model, AR(1) for βs	6.3
NS1-VAR	1-step estimation (Kalman) of NS model, VAR(1) for βs	6.3
BC1-AR	1-step estimation (Kalman) of BC model, AR(1) for βs	6.3
BC1-VAR	1-step estimation (Kalman) of BC model, VAR(1) for βs	6.3
NS1-RS	1-step estimation (Kim) of regime-switching NS model	6.4
BC1-RS	1-step estimation (Kim) of regime-switching BC model	6.4
NS1-RS-X	1-step estimation (Kim) of regime-switching NS model, TVTPs	6.5
BC1-RS-X	1-step estimation (Kim) of regime-switching BC model, TVTPs	6.5

The section proceeds with a discussion on how estimation techniques are implemented for each model, followed by a presentation of the actual estimates. In that vein, I start with the two-step estimation of the NS and BC models. Next, the one-step models without regime switching are discussed. Subsequently, the section moves on to the estimation of the regime-switching models and presentation of corresponding filtered state probabilities and time-varying transition probabilities. The section concludes by presenting likelihood ratio tests to evaluate significance of the different models and the second regime. All models are estimated with Matlab.[2]

[2]Matlab 7.12.0.635 (R2011a). Implementation of the estimation procedures of the regime-switching models based in part on Nyholm (2008).

6.2 Single-regime yields-only models: two-step procedure

6.2.1 Step one: observation equation

As discussed in Section 3.1, I follow Diebold and Li (2006) by fixing the decay parameter, λ at 0.0609. Therefore, the first step of the two-step estimation procedure comes down to an OLS regression. The regression is performed on the observation equation presented in (3.6). Recall that the covariance matrix R of this observation equation is assumed diagonal at all times. For both the NS2 and BC2 model, the regression is performed for each observation in the sample (i.e., 564 regressions are performed with NS2 and 564 regressions with BC2). Matlab identifies the regression parameters by assuming zero-correlation between the error term and the explanatory variables.

Table B.1 in the appendix presents statistics of the estimates that result from the OLS regression. The table shows that $\beta_{1,t}$ remains strictly positive for both NS2 and BC2, as opposed to the other factors. Whereas $\beta_{1,t}$ is of the same magnitude across the two models, for the other factors this is not quite true. Apart from a different mean, the standard deviation of the factors of the BC2 model is higher and judging from the minimums and maximums, the BC2 factors are somewhat more 'spiky'.

For both models, autocorrelations with a 1-month time lag of all factors are high. When the time lag increases to 12 months, all factors but $\beta_{1,t}$ have an autocorrelation of less than 0.5. That is, $\beta_{1,t}$ turns out to be most persistent factor, which indeed should be the case given the limiting behavior of the models. Correlation between factors is low for NS2, which is also observed by Diebold and Li (2006). Correlation of the NS2 factors with the defined level, slope and curvature are close to unity, implying the three Nelson-Siegel factors, $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$, may indeed be interpreted as a level (L_t), slope (S_t) and curvature (C_t) factor, respectively. Diebold and Li (2006) find these correlations to be $\rho(\hat{\beta}_{1,t}, L_t) = 0.97$, $\rho(\hat{\beta}_{2,t}, S_t) = -0.99$ and $\rho(\hat{\beta}_{3,t}, C_t) = 0.99$.

Correlation between BC2's $\beta_{2,t}$ and $\beta_{4,t}$ is high, which is, given the similarity of their loadings, in the line of expectation. These two factors turn out to be substantially correlated with the curvature factor, $\beta_{3,t}$ as well. Correlation of BC2's level factor with the other factors is comparable to the NS2 case i.e., the correlation is low. With the exception of $\beta_{1,t}$, correlation of the Björk-Christensen factors with the yield factors seems somewhat lower than in the case of the three-factor NS2 model.

6.2.2 Step two: state equation

The estimated factors from the first step are now used as input for the second step, the estimation of the autoregressive state equations. In accordance with the discussion in Section 3.2 about the specification of the state equation, I model both an AR(1) and VAR(1). Therefore, a total of four models are estimated via the two-step procedure: NS2-VAR, BC2-VAR, NS2-

AR and BC2-AR. It is assumed the disturbances, υ_t follow (multivariate)normal distributions with mean zero and (co)variance matrix Q.

Table B.4 presents an overview of the parameter estimates. For all four models, estimates on the diagonal of F are significant at the 5%-level. Whereas NS2-VAR, NS2-AR and BC2-AR turn out to be stationary, the autoregressive coefficient of $\beta_{2,t}$ of BC2-VAR is estimated to equal 1.070 and hence BC2-VAR seems non-stationary with the two-step estimation procedure.[3] Elements of the constant vector, $\tilde{\mu}$ are significant in most cases. An exception to that is the constant associated with $\beta_{2,t}$ in the VAR specification of NS2 as well as BC2, and the constant of NS2-AR's curvature factor.

6.3 Single-regime yields-only models: one-step procedure

As with the two-step procedure, I estimate four models with the one-step estimation procedure: NS1-VAR, BC1-VAR, NS1-AR and BC1-AR.

6.3.1 Estimation procedure

In the one-step estimation procedure - based on the Kalman filter - I apply maximum likelihood on a function that follows from the prediction-error decomposition of the likelihood. That is, the log-likelihood function, $l(\theta)$ to be maximized with respect to the unknown parameters, $\theta = (\theta_1, \ldots, \theta_s)$ is given by

$$l(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \ln(2\pi|f_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{T} \eta'_{t|t-1} f^{-1}_{t|t-1} \eta_{t|t-1}.$$

Here, $\eta_{t|t-1} = y_t - y_{t|t-1}$, which is an $N \times 1$ vector of yield prediction errors where $y_{t|t-1} = \mathbb{E}[y_t \mid \Psi_{t-1}]$, the forecast of y_t given information up to time $t-1$ collected in Ψ_{t-1}. $f_{t|t-1} = \mathbb{E}[\eta^2_{t|t-1}]$, an $N \times N$ matrix which corresponds to the variance of the prediction errors. The operators $|X|$ and X^{-1} refer to the determinant and inverse of a matrix X, respectively. The parameter set is $\theta = (\lambda, \tilde{\mu}, F, R, Q)$. Note this now includes λ and that it is assumed fixed for all t once more. Assumptions about the disturbances, ϵ_t and υ_t are given in (3.8) and (3.9). For an elaborate discussion on the Kalman filtering procedure, refer to Appendix A.1.

When one allows for a non-diagonal Q the number of parameters to be estimated significantly increases. For instance, the BC1-VAR model requires estimation of no less than 48 parameters (i.e., the decay parameter (λ), four constants in $\tilde{\mu}$, sixteen coefficients in F, seventeen - the number of maturities - variances in R, four variances and six co-variances in Q). Similarly, the NS1-VAR model requires estimation of 36 parameters. The number of parameters of NS1-AR equals 27 and for BC1-AR the number is 30. Indeed, the one-step

[3] An autoregressive process is said to be *stationary* if the absolute value of each autoregressive coefficient (i.e., the root) is strictly smaller than 1.

estimation of these models is challenging, but not impossible.

The initial parameter configuration is as follows. I start λ at 0.0609, the value found by Diebold and Li (2006). Estimates of the second step from the two-step procedure of either the VAR or AR representation of the state equation are used for $\tilde{\mu}$ and F. The variance parameters in R are set to 1, those in Q are set to their respective two-step procedure estimates. Covariance terms in Q start from the two-step estimates as well. The Kalman filter is initialized using the unconditional mean and unconditional covariance matrix of the state vector, expressions of which are given by (A.4) and (A.5), respectively. Positivity of the variance terms in R and Q is ensured by setting a lower bound of zero for these parameters.

The log-likelihood is maximized using the interior-point algorithm. The Hessian, used for the computation of standard deviations, is estimated using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The termination tolerance between two consecutive iterations equals 10^{-6}.

6.3.2 Estimation results

Panels A and B of Table B.2 list the one-step factor statistics of the VAR(1) and AR(1) specifications, respectively. The results are quite comparable to the two-step results. In particular, factors statistics of the Nelson-Siegel models do not differ much. For the Björk-Christensen models, it holds that the averages and corresponding standard deviations of the slope factors, $\beta_{2,t}$ and $\beta_{4,t}$, are somewhat different. Moreover, one notices the minimums (maximums) of the one-step BC models to be higher (lower). That is to say, it seems the one-step estimation procedure smoothes out the factor estimates. This is corroborated by Figure C.1 which simultaneously shows time series of the factors from the one-step and two-step procedures. De Pooter (2007) finds the one-step factors to be smoothed as well.

One-step factor statistics and correlations of NS1-VAR and NS1-AR are much alike. Differences between BC1-VAR and BC-AR are larger but still diminutive. In particular, statistics of the three BC factors, $\beta_{2,t}$, $\beta_{3,t}$ and $\beta_{4,t}$ differ for the two specifications of the state equation. The same goes for the correlations of these factors with the other factors and the level, slope and curvature.

The estimate of the decay parameter, $\hat{\lambda}$ equals 0.0561, 0.0578, 0.0562 and 0.0702 for NS1-VAR, BC1-VAR, NS1-AR and BC1-AR, respectively. All decay parameters are highly significant. Estimates of the state equation - presented in Table B.5 - are comparable to those found using the two-step procedure. The most notable difference is that all models are now stationary. A second difference is the order of magnitude of the variance terms of Q of BC1-AR relative to BC2-AR. This might be the result of the simultaneous consideration of both yield dimensions (cross sectional and time series). Other, to some extent smaller differences relate to the significance of certain parameters in the transition equation, F and constant vector, $\tilde{\mu}$.

Given the stationary character and correct inference of the one-step estimates as outlined in Section 6.1, the one-step estimates are generally preferred in lieu of their two-step counterparts.

6.4 Regime-switching yields-only models

6.4.1 Estimation procedure

Parameters of the regime-switching models are again estimated by maximizing a log-likelihood function, but now through an application of the Kim filter. Without loss of generality, let M denote the number of states and define $\mathbb{P}[S_t = j \mid \Psi_{t-1}]$ the probability to be in regime j at time t conditional on information available at time $t-1$. Moreover, let $f(y_t \mid S_t = j, \Psi_{t-1})$ be the density function of y_t given regime j at time t and information Ψ_{t-1}. Collect the state probabilities and densities of all M regimes in the $M \times 1$ vectors $\pi_{t|t-1}$ and D_t, respectively. That is,

$$\pi_{t|t-1} = \begin{pmatrix} \mathbb{P}[S_t = 1 \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = j \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = M \mid \Psi_{t-1}] \end{pmatrix}, \qquad D_t = \begin{pmatrix} f(y_t \mid S_t = 1, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = j, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = M, \Psi_{t-1}) \end{pmatrix}. \tag{6.1}$$

The (approximate) log-likelihood to be maximized w.r.t. $\theta = (\theta_1, \ldots, \theta_s)$ is then given by

$$l(\theta) = \sum_{t=1}^{T} \ln \left[\mathbb{1}'(\pi_{t|t-1} \odot D_t) \right], \tag{6.2}$$

where $\mathbb{1}$ represents an $M \times 1$ vector of ones and the symbol $\odot$ denotes element-by-element multiplication. Assumptions about the error terms are stated in (4.4).

The iterative Kim filtering algorithm estimates a regime-switching state-space model by taking three consecutive steps at each time t: the Kalman filtering step, the Hamilton filtering step and the collapsing step. The discussion in Appendix A.2 extensively elaborates on these steps in a general context. Here, partly based on Bernadell et al. (2005), I present how the most important parts of the three steps come about with the model setup from Section 4.2.

Let $\beta_{t|t} = \mathbb{E}[\beta_t \mid \Psi_t]$, the expectation of β_t given information up to and including time t. Similarly, let $\beta^j_{t|t} = \mathbb{E}[\beta_t \mid \Psi_t, S_t = j]$, the expectation of β_t given $S_t = j$ and information up to and including time t. Kim (1994) proposes to estimate β_t by a weighted average of $\beta^j_{t|t}$ for each regime j:

$$\beta_{t|t} = \pi'_{t|t} \beta^j_{t|t}. \tag{6.3}$$

The Kim filter, in particular, foresees in finding an estimate of $\beta^j_{t|t}$ for all j. To that end, the first step (i.e., the Kalman filter) computes for all i, j the conditional state vector, $\beta^{(i,j)}_{t|t} = \mathbb{E}[\beta_t \mid \Psi_t, S_t = j, S_{t-1} = i]$. As such a total of M^2 of these $\beta^{(i,j)}_{t|t}$s are computed. The Kalman filter is constructed in such a way that at time $t-1$ an initial prediction of $\beta^{(i,j)}_t$ is made, denoted by $\beta^{(i,j)}_{t|t-1}$. Since regime dependency of my model setup enters through the constant of the state equation, the initial Kalman prediction, $\beta^{(i,j)}_{t|t-1}$ for each i, j becomes

$$\beta^{(i,j)}_{t|t-1} = \tilde{\mu}_j + F\beta^i_{t-1|t-1}. \tag{6.4}$$

This implies that the Kalman prediction errors, captured in η, become regime dependent as well:

$$\eta^{(i,j)}_{t|t-1} = y_t - y^j_{t|t-1} = y_t - H\beta^{(i,j)}_{t|t-1}. \tag{6.5}$$

The second step of the Kalman filter, predicting the covariance matrix $P_{t|t-1} = \mathbb{E}[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})' \mid \Psi_{t-1}]$ of β_t one step ahead, is unaffected by the setup and is hence given by

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q.$$

Therefore, the estimate of the variance of the prediction errors is unaffected too:

$$f_{t|t-1} = HP_{t|t-1}H' + R. \tag{6.6}$$

At time t, updated estimates of the covariance matrix and the regime-dependent state vectors are, respectively, computed as

$$\begin{aligned} P_{t|t} &= P_{t|t-1} - P_{t|t-1}H'f^{-1}_{t|t-1}HP_{t|t-1}, \\ \beta^{(i,j)}_{t|t} &= \beta^{(i,j)}_{t|t-1} + P_{t|t-1}H'f^{-1}_{t|t-1}\eta^{(i,j)}_{t|t-1}, \end{aligned} \tag{6.7}$$

which completes the Kalman filter. From (6.3) it can be seen one is particularly interested in $\beta^j_{t|t}$ (and ultimately $\beta_{t|t}$) rather than $\beta^{(i,j)}_{t|t}$. The step from $\beta^{(i,j)}_{t|t}$ to $\beta^j_{t|t}$ involves an approximation (the third step), but before one undertakes this step particularly note from (6.3) one is ought to compute the state probabilities, $\pi_{t|t}$ as well. This is where the second step (i.e., the Hamilton filter) comes into play. Hamilton (1989) proposes to predict at time $t-1$ the regime probabilities of time t by

$$\pi_{t|t-1} = P\pi_{t-1|t-1}. \tag{6.8}$$

Here, P denotes the transition probability matrix of size $M \times M$ as in e.g., (4.1) and $\pi_{t-1|t-1}$

is given.[4]

At time t, the predicted state probabilities are updated using the conditional densities collected in D_t, which are computed on the basis of the Kalman filtered estimation results. In particular, the expressions in (6.5) and (6.6) are used to calculate the (reduced) conditional density of y_t for each i, j.[5] That is,

$$\begin{aligned} f\big(y_t \mid \Psi_{t-1}, S_t = j, S_{t-1} = i\big) &= (2\pi)^{-\frac{T}{2}} |f_{t|t-1}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(\eta_{t|t-1}^{(i,j)})'(f_{t|t-1})^{-1}\eta_{t|t-1}^{(i,j)}\Big) \\ &\propto |f_{t|t-1}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(\eta_{t|t-1}^{(i,j)})'(f_{t|t-1})^{-1}\eta_{t|t-1}^{(i,j)}\Big). \end{aligned}$$

Note that, when this expression is summed over all i and multiplied by the corresponding state probability, one finds the j-th element of D_t as in (6.1).[6] As such, the updated estimate of the state probabilities becomes

$$\pi_{t|t} = \frac{(\pi_{t|t-1} \odot D_t)}{\mathbb{1}'(\pi_{t|t-1} \odot D_t)}. \tag{6.9}$$

To preserve tractability the third step of the Kim filter collapses the M^2 filtered state vectors, $\beta_{t|t}^{(i,j)}$ from (6.4) into M approximated vectors, $\beta_{t|t}^{j}$. Using the joint probability of S_t and S_{t-1} this approximation is given by

$$\beta_{t|t}^{j} = \frac{\sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t]\beta_{t|t}^{(i,j)}}{\mathbb{P}[S_t = j \mid \Psi_t]}, \quad j = \{1, 2, \ldots, M\}. \tag{6.10}$$

For all j, $\beta_{t|t}^{j}$ is subsequently substituted into (6.4), which completes the iteration.[7]

After completion of the algorithm (6.10) is plugged into (6.3), yielding an estimate of the unobserved state vector, β_t for all t. Note that the log-likelihood function (6.2) conveniently follows from the denominator of (6.9). The parameter set is now given by $\theta = (\lambda, \tilde{\mu}, F, R, Q, P)$.

Due to the diagonality of R and Q, the total number of parameters to be estimated stays within reasonable bounds. For the NS1-RS model, there is the decay parameter (λ), four constants in $\tilde{\mu}$ (two of which for the slope factor), two coefficients in F (since f_{22} is assumed zero), seventeen variances in R, three variances in Q and two transition probabilities

[4] It is noted the matrix in (4.1) applies to the case where $M = 2$.

[5] Reduced in the sense that the constant term $(2\pi)^{-\frac{T}{2}}$ is scaled to unity. This facilitates optimization because for large T the original constant approaches zero.

[6] The symbols P and f are overloaded. The former, P may refer to either the transition probability matrix or to the covariance matrix of β_t. The latter, f, refers either to the variance of the predictions errors or to the density function. One can determine the meaning of the symbols by verifying whether the subscript includes a pipe character (|), in which case P and f refer to the covariance matrix of β_t and the variance of the prediction errors, respectively. That being said, I leave it to reader to determine the intention of the respective symbols in the remainder of the thesis.

[7] As said, refer to Appendix A.2 for a more extensive discussion on the Kim filter. The appendix e.g., also discusses how the Hamilton filter allows for the construction of the joint probabilities used in (6.10).

in P; a grand total of 29 parameters. For BC1-RS the number is 32 since three additional parameters arise in $\tilde{\mu}, F$ and Q. The initial parameter configuration for λ, R and Q is similar to the configuration of Section 4.2. State equation parameters in $\tilde{\mu}$ and F are set to the AR estimates of the two-step procedure, with the exception of f_{22} which is assumed zero at all times. Furthermore, $\tilde{\mu}_{2,1}$ is initialized at the two-step estimate and $\tilde{\mu}_{2,2}$ at two plus the two-step estimate. This is done to ensure different estimates for $\tilde{\mu}_{2,S_t}$ for each regime j. Transition probabilities p_{11} and p_{22} are both set at 0.90, state probabilities start from the unconditional probabilities of S_t as given in (A.13). The filtering procedure itself is initialized using the unconditional mean and unconditional covariance matrix of the state vector. The interior-point algorithm alongside the BFGS method for standard errors is used again with a termination tolerance of 10^{-6}. Positivity of R and Q is ensured by imposing a lower bound of zero.

6.4.2 Estimation results

From Panel A of Table B.3 the fortunate observation can be made that factor statistics are to a high degree comparable to the factor estimates of the single-regime models. Whereas this specifically holds true for the Nelson-Siegel models, the Björk-Christensen models do show some deviation in the slope factors, $\beta_{2,t}$ and $\beta_{4,t}$. This can be seen by observing that the minimums (maximums) of BC1-RS are higher (lower), implying that BC's $\beta_{2,t}$ and $\beta_{4,t}$ tend to fluctuate less in the regime-switching case. This is corroborated by lower values of the corresponding standard deviations. In fact, from time series of the factors in Figure C.2 it becomes apparent that particularly BC's $\beta_{2,t}$ fluctuates between two levels with little disturbance. The figure also reveals that time series of the factors of the regime-switching and single-regime Nelson-Siegel models are quite similar. The same goes for BC's $\beta_{1,t}$ and (to a somewhat lesser extent) for BC's $\beta_{3,t}$.

For both models the level factor, $\beta_{1,t}$ remains the most persistent factor. Correlation between the Nelson-Siegel factors and correlation of these factors with the level, slope and curvature are comparable to the single-regime case. Regarding the Björk-Christensen model, it seems the regime-switching character of $\beta_{2,t}$ affects the correlation of this factor with $\beta_{3,t}$ and $\beta_{4,t}$, as both correlations are now positive. With the exception of $\beta_{1,t}$, factor correlation of BC1-RS with the level, slope and curvature differs as well. It is particularly interesting to observe that correlation between the slope and BC1-RS's $\beta_{4,t}$ is now much stronger than in the single-regime case. This matches with the theoretical interpretation of $\beta_{4,t}$ as a second slope factor, as discussed in Section 3.3.

Table B.6 presents estimates of the state equation, decay parameter and transition matrix, P. It follows that both NS1-RS and BC1-RS are stationary and $\hat{\lambda}$ equals 0.0571 and 0.0762, respectively. For both models, the switching constant of the slope factor, $\tilde{\mu}_{2,S_t}$ shows two distinct estimates that are both highly significant. Estimates of $\tilde{\mu}_{2,S_t}$ of the NS model,

$\tilde{\mu}_{2,1} = -3.295$ and $\tilde{\mu}_{2,2} = -0.205$, strongly support the model specification and motivation from Section 4.2. That is to say, the estimate of $\tilde{\mu}_{2,1}$ produces an upward sloping term structure and hence corresponds to a regime of relatively normal economic activity (i.e., regime 1). The estimate of $\tilde{\mu}_{2,2}$ matches with a flattish curve, which is typically observed in a period of aberrant economic activity (i.e., regime 2). The same line of reasoning holds for the estimates of BC's $\tilde{\mu}_{2,S_t}$, where $\tilde{\mu}_{2,1} = 1.490$ (upward sloping) and $\tilde{\mu}_{2,2} = 2.792$ (flattish).

The estimated transition matrix, $\hat{P}$ shows that transition probabilities of the active states are close to unity. This is reassuring because it indicates that the identified regimes are persistent and that the model supports the data (Bernadell et al., 2005). Figures 6.1 and 6.2 show time series of the filtered state probabilities of the aberrant regime (regime 2) of NS1-RS and BC1-RS, respectively.[8] It follows, in general, that switching from one state to the other only occurs after a prolonged period of time. That is to say, once the yield curve enters a regime it tends to stay in that regime for some time, only to switch to the other after numerous years. Therefore, the results link naturally to the evolution of the business cycle.

The shaded areas overlaying the figures are NBER recession periods. Focusing first on Figure 6.1, it can be seen that filtered state probabilities fit reasonably well with the NBER recessions. This corroborates with the economic interpretation of the estimates of $\tilde{\mu}_{2,S_t}$ as outlined above. For instance, the model clearly 'catches' the 1975 and 1990 recessions. Also, it tends to distinguish between the two recessions in the early 1980s, as can be seen from the slightly lower probability of regime 2 right after the end of the 1980 recession. It is interesting to see that the model identifies the credit crunch, the period from 2007 onwards, as a 'normal' regime (i.e., regime 1), which contrasts the NBER data. The reason for this might be that the model setup is such that an aberrant regime corresponds to a high policy rate, whereas the 2007 recession is characterized by low interest rates.

State probabilities of the BC1-RS model also match reasonably well with the NBER data, as can be seen from Figure 6.2. As such, regime 2 fits the recessions best which is again in line with the interpretation of the switching constant of the slope factor. The model identifies the 1975 and 1990 recessions and also tends to switch in the early 1980s. Where NS1-RS failed to do so, the most recent period of economic contraction (2007) is now identified. Judging from the spikiness of the state probabilities it seems that BC1-RS is slightly less persistent.

[8] That is, the figures plot equation (6.9).

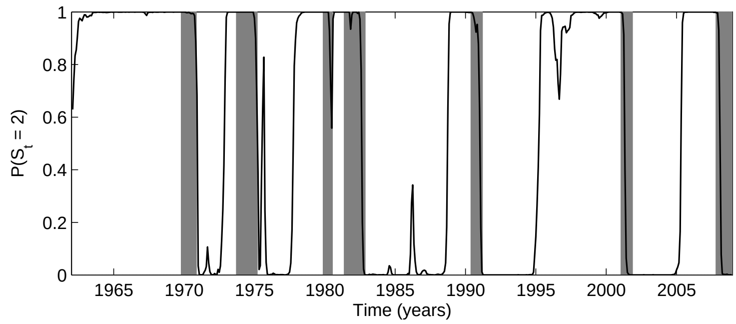

Figure 6.1: **Filtered state probability of the NS1-RS model.**

Notes: The solid line corresponds to the filtered state probability of regime 2, the regime affiliated with aberrant economic activity. Shaded areas denote NBER recession periods.

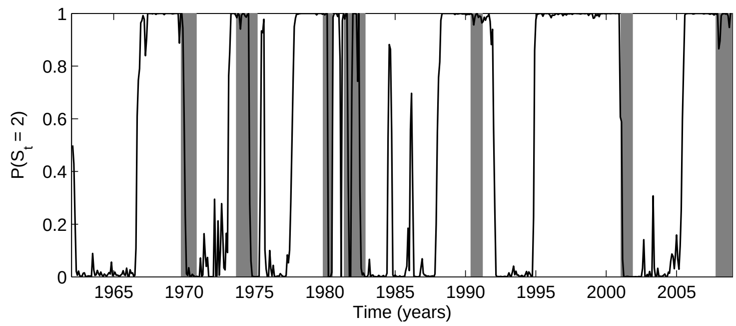

Figure 6.2: **Filtered state probability of the BC1-RS model.**

Notes: Refer to Figure 6.1 for details.

Figure 6.3 shows the average observed and fitted term structure for the full sample and for the two regimes. Here, the yield data are split in order to fit the regime-dependent curves: for each month t, the yields of that month are assigned to regime j if $\mathbb{P}[S_t = j \mid \Psi_t] > 0.5$. For NS1-RS this is done using the filtered state probabilities from Figure 6.1, for BC1-RS the probabilities of Figure 6.2 are used. Figure 6.3 confirms that one regime is identified by an upward sloping curve and another by a flattish curve. It follows that the average curve of the full sample can be interpreted as some combination of the regime-dependent curves. In particular, for the Björk-Christensen model the average yield curve nicely falls between the two regime-dependent curves.

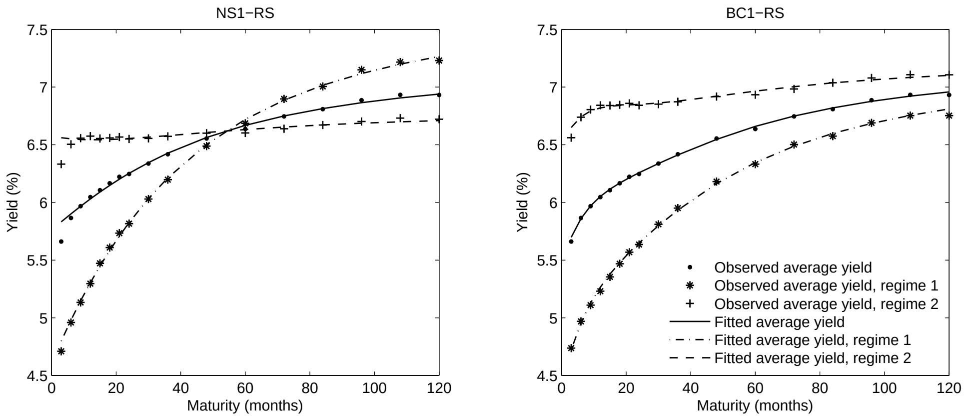

Figure 6.3: **Average fit of the regime-switching yields-only models.**

Notes: The left graph shows the average fit of the three-factor (Nelson-Siegel) yields-only regime-switching model (**NS1-RS**), the right graph presents its four-factor (Björk-Christensen) counterpart (**BC1-RS**). Solid lines correspond to the average fit of the full sample (1962:1 - 2008:12). Dashed lines correspond to average fitted regime-dependent yield curves. Dots, asterisks and crosses denote observed yields. See the text for further details.

6.4.3 Alternative model specifications

Instead of the constant of the slope factor, $\tilde{\mu}_{2,S_t}$ that governs the regime switch one could also govern the switch through other or additional parameters. Empirical investigation, however, suggests that most of these alternative model specifications come with some serious drawbacks.

One could, for instance, govern the switch through the constant associated with the level factor, $\beta_{1,t}$. Empirical results turn out to be poor, however, since switching hardly occurs. Moreover, such model specification is rather hard to back by solid economic theory. Alternatively, one could co-switch all constants at once, but again empirical results are not reassuring. In particular, state probabilities tend to fluctuate quite intense, implying that regimes are not very persistent which is intuitively unappealing. Estimates turn out to be insignificant as well. The same goes when one simultaneously switches BC's $\beta_{2,t}$ and $\beta_{4,t}$.

Bernadell et al. (2005) argue that including an autoregressive coefficient for the slope factor (i.e., to relax the assumption $f_{22} = 0$) results in hardly any state transitions. I test this claim by recalibrating the switching models where f_{22} is now estimated alongside the other parameters. With the Björk-Christensen model I indeed find that switching takes place just once for a period of less than twelve months. With the Nelson-Siegel model multiple state transitions do occur but results are far from persistent. As such, I concur with Bernadell et al. (2005) on the model setup where f_{22} is restricted.

Other studies - e.g., Zhu and Rahman (2009) - govern a switch through the variance-covariance matrix of the state equation, Q. This is an intuitively appealing model specification since two regimes may be identified as a tranquil regime and a turbulent regime. However, it can be argued that such switch is not of much added value in an out-of-sample study since ahead forecasts are hardly affected by this setup.

6.5 Regime-switching macro-finance models with time-varying transition probabilities

6.5.1 Estimation procedure

Estimating regime-switching models with time-varying transition probabilities is in essence not very different from the fixed transition probability case. Once again, estimates are obtained by maximizing the approximate log-likelihood function in (6.2), which is now unconditionally dependent of y_t as well as the additional explanatory variables, z_t. That is, the densities collected in D_t are now joint densities of y_t and z_t. This is a potential estimation problem because it implies that the parameters of y_t and z_t should be jointly estimated. Since parameter estimates of the latter process are irrelevant (i.e., they are nuisance parameters), it is beneficial to bypass this joint estimation and resort to the traditional procedure instead (i.e., estimate y_t only). Filardo (1998) provides a sufficient condition that, when the condition is met, this is justified and the Hamilton filtering procedure from Section 6.4 may hence be applied. In particular, the condition allows for the factorization of the joint log-likelihood function into concentrated log-likelihood functions.[9] In so doing, the parameters of z_t are concentrated out of the joint log-likelihood when the condition is met.

Condition 6.1 (Filardo, 1998)**.** *The k additional explanatory variables collected in z_t that enter the transition probability functions are contemporaneously conditionally uncorrelated with the discrete hidden state, S_t.*

Numerous studies make use of lagged informative indicators and assume Condition 6.1 holds. Although direct verification of the condition is not possible in many cases - as is the case here - (due to the hidden character of S_t), using lagged variables is reasonable as long as the variables in z_{t-1} are predetermined with respect to the unobserved state, S_t (Filardo, 1998). Therefore, I follow this practice as can be seen from (4.6) where the leading indicators of known time $t-1$ are used (i.e., z_{t-1}) to estimate the transition probabilities at time t. Note particularly that in so doing no future economic information is used for the computation of the TVTPs. Note, moreover, that it is imperative the indicators are observable indeed whenever the quantity of interest is. The macroeconomic variables I consider are in fact observable when

[9]Without loss of generality, divide the unknown parameter set, θ into two subsets, θ_1 and θ_2. Following Filardo (1998), a *concentrated* log-likelihood function, $l(\theta_i),\ i = \{1, 2\}$ is a function that satisfies $l(\hat{\theta}_1, \hat{\theta}_2) = \max_{\theta \in \Theta} l(\theta) = \max_{\theta_1 \in \Theta_1} l(\theta_1) + \max_{\theta_2 \in \Theta_2} l(\theta_2)$. This property yields the advantage that, irrespective of θ_2, consistent estimates of θ_1 are found by maximizing its concentrated log-likelihood function, $l(\theta_1)$. As such, the parameters of y_t and S_t are collected in θ_1, those of z_t in θ_2. See Filardo (1998) for further details.

the yield data are and there is hence no need to further lag the three leading indicators.[10] All in all, it is reasonable to assume Condition 6.1 holds true and without restating all steps, I thus proceed with the estimation procedure from Section 6.4. It is obvious yet important to note, however, that the prediction equation in (6.8) is now given by

$$\pi_{t|t-1} = P_t \pi_{t-1|t-1},$$

where the time-dependent transition matrix, P_t is defined as in (4.5).

The parameter set of the log-likelihood function consists of $\theta = (\lambda, \tilde{\mu}, F, R, Q, a_1, a_2)$. For the NS1-RS-X model this implies θ comprises 35 parameters: one decay parameter, four constants in $\tilde{\mu}$, two coefficients in F, seventeen variances in R, three variances in Q, four transition probability-related constants in a_1 and four in a_2. Similarly, BC1-RS-X requires estimation of 38 parameters. The initial parameter configuration is the same as the one in Section 6.4 with the exception of the transition probabilities. That is, the coefficients of a_1 and a_2 start at zero, implying initial transition probabilities of 0.5. Once more, I apply the interior-point algorithm with a termination tolerance of 10^{-6}. I use the BFGS method to compute standard errors and positivity of R and Q is ensured by imposing a lower bound of zero.

6.5.2 Estimation results

Table B.3, Panel B reports factor statistics of the regime-switching macro-finance models. The table reveals that factor statistics and correlations of these macro-finance models are to a high degree comparable to the regime-switching yields-only models as presented in Panel A. That is to say, differences mostly occur in the decimals. Figure C.2, showing time series of the factors, corroborates the above as differences between the factors are hardly noticeable. The most striking difference is the somewhat deviating path of the switching slope factor, $\beta_{2,t}$ of the BC model.

Estimates of the state equation, decay parameter and TVTPs are reported in Table B.7. Again, they are similar to the regime-switching yields-only case. The decay parameter, λ equals 0.0571 and 0.0763 for NS1-RS-X and BC1-RS-X, respectively. The switching constants of NS1-RS-X are $\tilde{\mu}_{2,1} = 3.307$ and $\tilde{\mu}_{2,1} = -0.215$, for BC1-RS-X the estimation yielded $\tilde{\mu}_{2,1} = 1.596$ and $\tilde{\mu}_{2,1} = 2.870$. These constants again strongly support the model setup:

[10] That is, the Monthly CRSP U.S. Treasury Database from which the forward rates are extracted to construct the unsmoothed Fama-Bliss yields is updated annually. FRED's inflation and housing indicators are updated monthly and the GDP indicator quarterly (all with a lag of circa six weeks). Therefore, these leading indicators are observable when the forward rates are observable. Given the low updating frequency of the CRSP forward rates, it might be worthwhile to resort to alternative yield databases for non-academic purposes. Because of the six week time lag of the FRED database, one could argue the variables should be lagged further in that case. Of course, leading indicators may be obtained from other sources which might be updated faster. From a quick-and-dirty estimation experiment I come to conclude that estimates with TVTPs based on a two-period lag differ only marginally from those based on a one-period lag as presented in the thesis.

the constants of regime 1 are in accordance with a normal economic regime, those of regime 2 with an aberrant regime. Ten of the sixteen estimated transition probability constants are significant at the five-percent level. It is interesting to observe that both housing starts constants of BC1-RS-X are significant at the five-percent level whereas the null cannot be rejected for NS1-RS-X. Similarly, the NS1-RS-X constants of GDP are both significant (five percent), which cannot be said for BC1-RS-X.

Figure 6.4 depicts TVTPs of the NS1-RS-X model. In spite of the insignificance of some of the transition probability estimates, the time-varying transition probabilities themselves support the NBER recession periods surprisingly well. This can specifically be seen from the bottom figure, which depicts the probability that given the previous period was an aberrant regime (e.g., a recession), the current period is again a recession. One observes that during an NBER recession this probability decreases with time, implying that during a recession the probability to get out of it increases with time. Right after an NBER recession has ended this probability increases again. Recall from Section 4.3 this behavior was actually reasoned to be case. Therefore, it is fair to conclude that the assumption of time-varying transition probabilities is more realistic indeed than the assumption of fixed transition probabilities. Note that, fortunately, a mere three leading indicators suit their purpose - giving a rough indication of the state of the economy - quite well. The average transition probability of p_{11} equals circa 0.97 and for p_{22} the average probability is 0.96. That is, the regimes are persistent as was the case with the regime-switching yields-only models. The filtered state probability of NS1-RS-X of regime 2 is presented in Figure 6.5. Given the great similarity between the estimates of NS1-RS and NS1-RS-X it comes as no surprise that filtered state probabilities of these models are to a large extent the same as well.

Figure 6.6 presents time-varying transition probabilities of the regime-switching macro-finance BC model. It is fortunate to observe that the bottom figure shows a similar pattern as the one from NS1-RS-X. The upper figure is slightly different as the transition probability seems to fluctuate somewhat more intense. With an average probability of p_{11} of approximately 0.93 the model still strongly supports the data, however. The average transition probability of regime 2 is reassuring as well: $p_{22} = 0.95$. Figure 6.7 presents the filtered probability of state 2 of BC1-RS-X, which is quite similar to the probability corresponding to BC1-RS. Both Björk-Christensen switching models identify the 2007 housing crisis, something the Nelson-Siegel switching models do not.

Similar to Section 4.2, I assign the observed yields at time t to regime j if $\mathbb{P}[S_t = j \mid \Psi_t] > 0.5$. In so doing, it follows from Figure 6.8 that the average fitted curves resemble the average curves of the yields-only switching models (Figure 6.3). This is particularly true for the curves of NS1-RS-X and NS1-RS. More precisely, NS1-RS-X has an average fit that is again crossed by the two individually average fitted regimes. It can also be seen that the nice interpretation of the Björk-Christensen model is preserved when TVTPs are included: the average fitted curve falls between the fitted curves of the two distinct regimes.

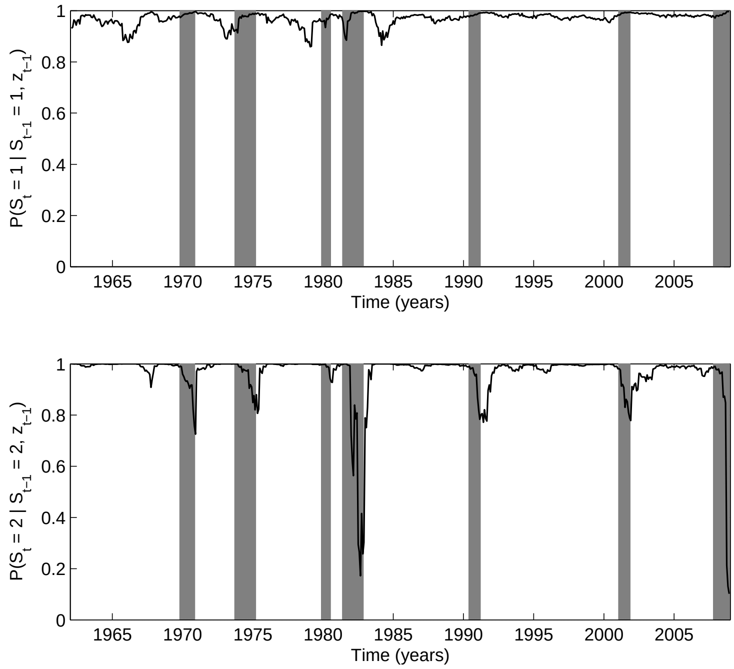

Figure 6.4: **Time-varying transition probabilities of the NS1-RS-X model.**

Notes: Solid lines denote time-varying transition probabilities. Regime 1 (upper graph) corresponds to a regime of normal economic activity, regime 2 (bottom graph) to a regime of aberrant activity. Shaded areas indicate NBER recessions.

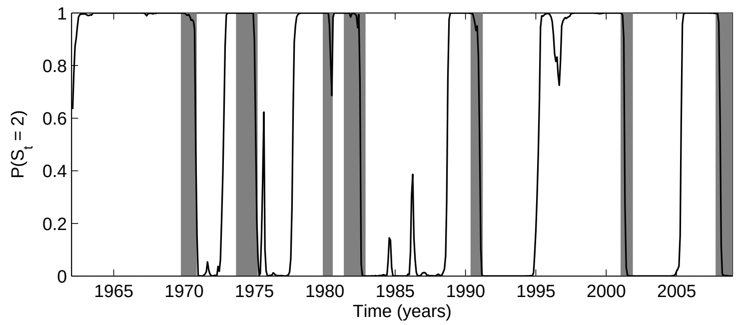

Figure 6.5: **Filtered state probability of the NS1-RS-X model.**

Notes: Refer to Figure 6.1 for details.

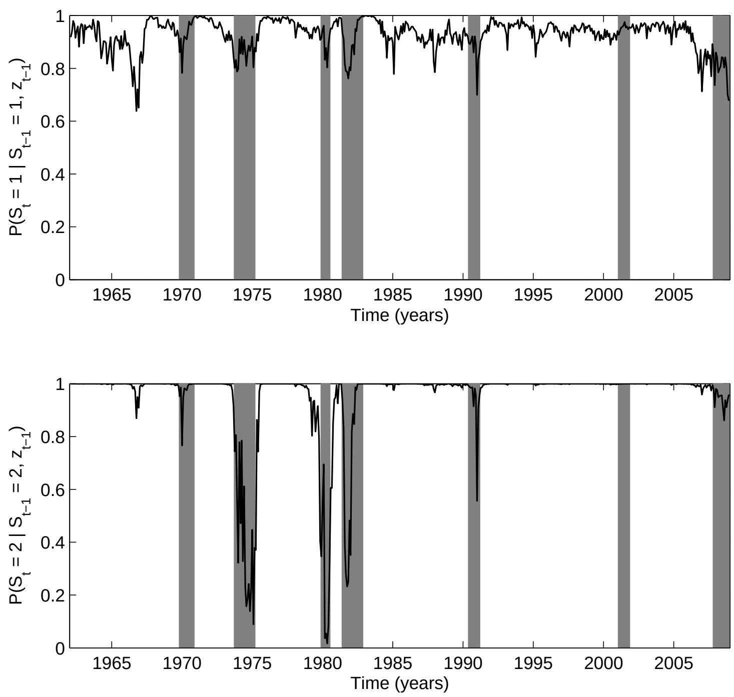

Figure 6.6: **Time-varying transition probabilities of the BC1-RS-X model.**

Notes: Refer to Figure 6.4 for details.

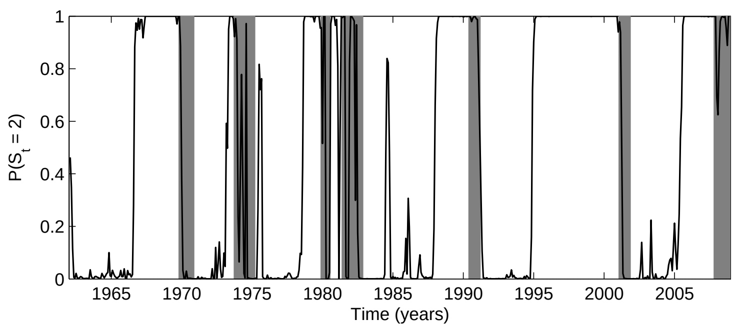

Figure 6.7: **Filtered state probability of the BC1-RS-X model.**

Notes: Refer to Figure 6.1 for details.

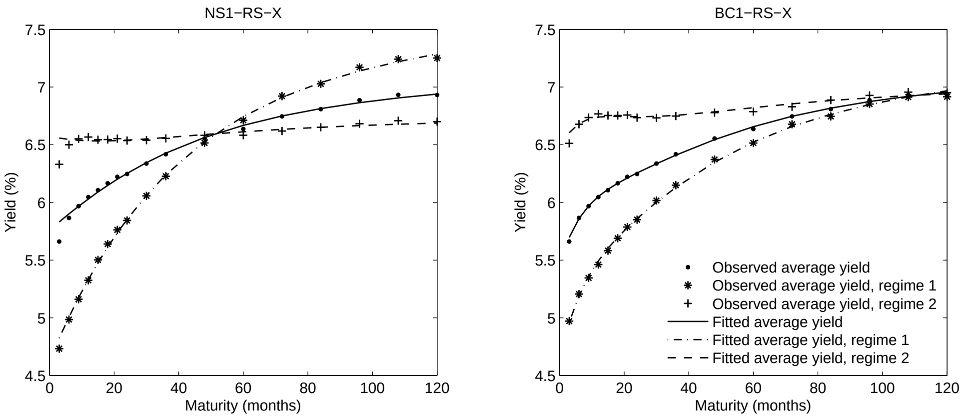

Figure 6.8: **Average fit of the regime-switching macro-finance models.**

Notes: The left graph shows the average fit of the three-factor (Nelson-Siegel) regime-switching macro-finance model (**NS1-RS-X**), the right graph presents its four-factor (Björk-Christensen) counterpart (**BC1-RS-X**). See Figure 6.3 for additional details.

6.6 Evaluation of the likelihood

6.6.1 Likelihood ratio test

Given the nested nature of many of the above term structure models the likelihood ratio test (LRT) is a convenient way to test the statistical significance of the extended models. Most prominently, it enables one to test if the assumption of time-varying transition probabilities is significantly different from the assumption of fixed probabilities. In addition, the LRT gives some direction into the significance of the second regime. One can also test the significance of the additional factor ($\beta_{4,t}$) and it can be used to test whether the AR(1) benchmark models are rejected in favor of their VAR(1) counterparts.

Recall the unknown parameter set $\theta = (\theta_1, \ldots, \theta_s)$ and let $r < s$ be the number of parameter restrictions. That is, r is the difference in the number of parameters of the unrestricted model and the restricted (null) model. The hypothesis to statistically test two nested models is then given by

$$\mathrm{H}_0 : (\theta_1, \ldots, \theta_r) = (\theta_{1,0}, \ldots, \theta_{r,0}) \qquad \text{vs.} \qquad \mathrm{H}_1 : (\theta_1, \ldots, \theta_r) \neq (\theta_{1,0}, \ldots, \theta_{r,0}).$$

The test statistic is $2\Big(l(\hat{\theta} \mid y) - l(\hat{\theta}_0 \mid y)\Big)$, which is approximately χ^2-distributed with r degrees of freedom. Recall l is the log-likelihood function; $\hat{\theta}$ denotes the maximum likelihood estimate and $\hat{\theta}_0$ denotes the estimate given H_0 is true. The approximate likelihood ratio test of size α is to reject the null hypothesis if

$$2\Big(l(\hat{\theta} \mid y) - l(\hat{\theta}_0 \mid y)\Big) \geq \chi^2_{1-\alpha}(r). \tag{6.11}$$

Panel A of Table B.8 reports log-likelihood values of the single-regime one-step models and the regime-switching models from which LRTs can be constructed.

6.6.2 Three-factor Nelson-Siegel vs. four-factor Björk-Christensen

Panel B presents LRT statistics and corresponding distributions and p-values between the single-regime models. The tests NS1-AR vs. BC1-AR and NS1-VAR vs. BC1-VAR test the significance of the additional factor, $\beta_{4,t}$. With respective LRT statistics of 984.20 and 1847.32, they both strongly reject the null of the three-factor model in favor of the four-factor model at all conventional significance levels. This is corroborated by the upper two tests (NS1-RS vs. BC1-RS and NS1-RS-X vs. BC1-RS-X) in Panel D, which presents LRT tests between the regime-switching models.

6.6.3 AR(1) vs. VAR(1)

Somewhat less relevant to the central research question but still interesting are the tests NS1-AR vs. NS1-VAR and BC1-AR vs. BC1-VAR in Panel B. They both reveal that the null of an AR(1)-specified state vector is rejected in favor of its VAR(1) counterpart at a level far below one percent. The bottom test, NS1-AR vs. BC1-VAR, tests the least flexible single-regime model against the most flexible model. That is, 21 restrictions are imposed and as such it simultaneously tests the fourth factor and the VAR(1) specification. The test statistic is 1894.12 and with a p-value approaching zero the null is strongly rejected once more.

6.6.4 Single regime vs. multiple regimes

Given the integral presence of regime-switching models in the thesis it is of high interest to test the existence of a second regime. Since the factors of the regime-switching models from Sections 4.2 and 4.3 follow AR(1) processes (with the exception of the switching factor, $\beta_{2,t}$), the most obvious tests for this purpose are NS1-AR vs. NS1-RS and BC1-AR vs. BC1-RS. Unfortunately, practical testing issues arise because the regime-switching models do not fully nest the single-regime models due to the imposed restrictions on the autoregressive coefficient matrix, F (i.e., $f_{22} = 0$). Furthermore, the parameters of the second regime are not identified under the null of a single-regime model. That is, the transition probabilities (p_{11} and p_{22}) and the constant of $\beta_{2,t}$ under regime 2 ($\tilde{\mu}_{2,2}$) are not identified. Hence, one can no longer assume the test statistic to be approximately χ^2-distributed under the null and the traditional LRT test from (6.11) is, therefore, strictly speaking, not valid. To overcome this caveat Hansen (1992, 1996) proposes a standardized likelihood ratio test for nested regime-switching models by concentrating the identified parameters out of the likelihood function. The resulting concentrated likelihood is then optimized with respect to the unidentified parameters by means of a grid search. Garcia (1998) as cited in Gelman and Wilfling (2009), however, points out that the standardized LRT is computationally onerous, even for models with a relatively limited amount of parameters. The non-nesting nature of my models forms an additional difficulty. Put short, applying the standardized LRT to my models is arduous; I will leave this issue for future research.

To get a certain degree of confidence about the significance of the second regime anyway, numerous studies (e.g., Gray (1996) and Wilfling (2009) and references therein) resort to the traditional LRT (6.11) instead. This common procedure finds empirical support in Gelman and Wilfling (2009) who apply Markov-switching GARCH models to examine (the volatility of) stock prices of takeover targets. In particular, they evaluate the finite-sample distribution of their LRT statistic by means of bootstrapping techniques and find the results to be "fairly similar" to a χ^2 distribution with the degrees of freedom equal to the number of parameter restrictions i.e., to the traditional LRT. By virtue of this encouraging result and the implicit precedent from previous studies as outlined above, I apply the LRT (6.11) as well and reemphasize that the distribution under the null can no longer be assumed χ^2. That being said, the test statistic of the test NS1-AR vs. NS1-RS (BC1-AR vs. BC1-RS) in Panel C is 164.06 (345.76), which, given the size of the statistics, provides a certain degree of confidence indeed that the single-regime models are rejected in favor of the regime-switching models.

6.6.5 Fixed transition probabilities vs. time-varying transition probabilities

Note that the tests between regime-switching models in Panel D are in fact valid because all parameters are now identified under the null. NS1-RS vs. NS1-RS-X and BC1-RS vs. BC1-RS-X test the assumption of fixed transition probabilities versus the assumption of endogenously macroeconomic driven time-varying transition probabilities. With $\chi^2(6)$-distributed test statistics of 19.60 and 35.24 and corresponding p-values of 0.003 and $3.87 \cdot 10^{-6}$, respectively it follows that the null of FTPs is rejected for both models at the usual levels. This is a reassuring result, because it implies that it is imperative to include macroeconomic data and as such the macro-finance models are favored over their yields-only counterparts. The bottom test of Panel D tests the most restrictive regime-switching model against the most flexible regime-switching model. In so doing, one imposes nine parameters restrictions. The huge statistic of 1201.14 implies that the extensive BC1-RS-X model - characterized by TVTPs and the additional factor - is highly favored over the restrictive NS1-RS model.

The LRT results are comforting in the sense that the data support the extended, regime-switching models. In particular, the conclusion that the FTP models are strongly rejected in favor of the TVTP models is encouraging. As noted above, however, the LRTs also reject the restrictive single-regime models in favor of the more conventional extensions e.g., the additional factor or the VAR specification of the state vector. As such, it will be interesting to further study how the different models perform in sample, particularly at different maturities. This will hence be discussed next.

7 In-sample performance

This section evaluates in-sample fit on the basis of several loss functions and maturities. Given the potential of additional flexibility of the more parametrized models, it is reasonable to expect these models to show better in-sample performance. LRT results from the previous section support this idea. In particular, the BC models are expected to have a better fit than the NS models and the same goes for the VAR(1) models versus the AR(1) models. As such, it is of particular interest to see how the regime-switching models fit within the sample at varying maturities and how they compare to the competing, single-regime models.

Without going into too much detail, it is interesting as well to study in-sample differences of the two-step and one-step estimation procedures. In that light, it is important to remark that it could be argued to

> *"Only focus on the fit from step one of the two-step estimation procedure due to the fact that the one-step procedure potentially also uses (future) time-series information which is unavailable if we want to fit the term structure at a given point in time."* (De Pooter, 2007)

Although a valid point is made here, one could question whether it is a sufficient argument to exclude in-sample fitting evaluation of the one-step models from the outset. Consider, for instance, the case where one aims to fit the currently prevailing yield curve. Since, by definition, future yields are not observable (and hence only historical data can be used) it will be valid to deploy the one-step models. In another case where one estimates the one-step model up until the current period and wishes to fit the term structure of e.g., a year ago, future yields could potentially be used indeed. This is, however, irrelevant as the goal after all is cross-sectionally fitting the yield curve the best way possible. In any case, it is worthwhile to at least consider in-sample performance of the one-step models. This does not take away that the two-step procedures and one-step procedures are quite different. That is to say, one should bear in mind that estimation differences exist while evaluating in-sample fit.

7.1 In-sample fitting results

7.1.1 Error statistics

For consistency's sake I compute similar loss functions of the fitting errors as De Pooter (2007): standard deviation, root mean squared error, mean absolute error, minimum error and maximum error. Table 7.1 reports these error statistics in basis points for selected maturities. For each maturity and each function, the bold number indicates the best performing model.

All models turn out to fit the data remarkably well and it is encouraging to find that differences between the regime-switching models and competing, single-regime models are relatively small. With mean errors varying anywhere between -4.16 and 17.06 basis points, overall in-sample fit is quite good indeed.

The different maturities reveal that the models produce the largest error statistics at the short end of the curve, particularly at the 3-month rate. Errors at the far end of the curve (i.e., 10-year) are smaller but they are, in turn, higher than those in the middle e.g., the 2-year yields. The BC2 model turns out to be the winner when it comes to fitting the 3-month and 10-year yields, as this model reports superior error statistics for all five error measures for both maturities. 1-year rates are fitted very well also by BC2, which is again the winner with four of five measures. Interestingly, BC1-VAR performs particularly well at the 6-month level. Medium-term yields (2-year and 5-year) do not have a clear winner; NS1-VAR, BC1-VAR and BC1-RS-X are the top performers in this category.

Judging from the bold numbers the BC models win in 25 out of 30 cases. The five times that the NS models provide superior in-sample fit is mostly due to NS1-VAR. This is corroborated by the mean errors which show that (particularly the 3-month) errors of NS are non-negligibly higher than those of BC. Hence, I conclude that the BC models show better in-sample fit than the NS models, which justifies the additional (fourth) factor. De Pooter (2007) finds similar results and, moreover, shows that an additional factor provides for much more in-sample fit than does a non-prefixed decay parameter, λ that is to be estimated alongside the factors.[1]

Error statistics of NS1-AR and NS1-VAR are nearly identical. Those of BC1-VAR seem only marginally better than those of BC1-AR; in some instances error statistics of the latter model are actually higher than those of the former. Although the LRT rejected the AR specification in favor of the VAR specification, on the basis of these error statistics it seems, however, that additional in-sample fit due to the more flexible VAR-specified state vector is not very substantial.

Differences between the two estimation procedures of the single-regime models are small. The two-step procedure, however, seems to perform better at the tails of the curve i.e., at the short and long rates. This can, for instance, be seen from the minimum and maximum errors. The difference might well be due to the observation from Section 6.3 that one-step estimates are smoothed out compared to their two-step counterparts. That is to say, the observation that the one-step estimates, in general, attain less extreme values might explain the dilution of in-sample fit. Apparently, potential in-sample benefits of the one-step models (i.e., potential usage of future data) fail to materialize in practice.

It is fortunate to observe that loss functions of the regime-switching models are much in line with the (mainly one-step) single-regime models. In fact, the regime-switching models are amongst the top performers numerous times and in one instance the macro-finance model BC1-RS-X is the best performing model. Therefore, I conclude that the single-regime models and regime-switching models fit the yield data equally well within the sample.

[1] De Pooter (2007) finds the (adjusted) four-factor Svensson models, characterized by a second decay parameter, to show the best in-sample performance, albeit only marginally better than the BC models.

Table 7.1: **In-sample fit: error statistics (basis points).**

	NS2	BC2	NS1-AR	NS1-VAR	BC1-AR	BC1-VAR	NS1-RS	BC1-RS	NS1-RS-X	BC1-RS-X
	Mean error									
3m	8.58	2.37	17.01	16.96	4.58	7.65	17.06	3.71	17.06	3.56
6m	-2.53	-2.71	3.54	3.49	-0.60	-0.07	3.56	-0.55	3.55	-0.66
1y	-4.16	-0.40	-1.59	-1.62	0.37	-0.25	-1.59	0.60	-1.60	0.51
2y	1.61	2.03	0.56	0.53	1.13	1.49	0.58	1.06	0.57	0.92
5y	4.57	0.75	3.13	3.09	1.53	1.03	3.17	2.22	3.15	2.01
10y	-0.84	-3.17	0.86	0.84	3.08	3.89	0.74	2.57	0.73	2.42
	Standard deviation									
3m	13.57	**6.21**	27.10	27.01	18.44	18.91	27.00	22.19	26.99	22.37
6m	6.41	6.46	10.04	9.97	2.73	**0.36**	10.05	3.87	10.04	3.96
1y	10.98	**7.79**	8.71	8.70	9.95	9.77	8.63	10.78	8.63	10.67
2y	6.28	6.30	5.80	**5.77**	5.85	6.20	5.80	5.82	5.80	5.81
5y	8.30	6.55	7.38	7.39	7.11	**5.63**	7.36	7.86	7.35	7.82
10y	12.33	**9.41**	13.44	13.46	14.17	12.70	13.55	15.43	13.55	15.42
	Root mean squared error									
3m	16.06	**6.65**	32.00	31.89	19.00	20.39	31.94	22.49	31.93	22.65
6m	6.89	7.00	10.64	10.57	2.80	**0.36**	10.66	3.91	10.65	4.01
1y	11.74	**7.80**	8.85	8.85	9.96	9.77	8.78	10.80	8.78	10.68
2y	6.49	6.62	5.82	**5.79**	5.96	6.37	5.83	5.92	5.82	5.89
5y	9.48	6.58	8.02	8.01	7.28	**5.72**	8.01	8.17	8.00	8.08
10y	12.36	**9.93**	13.47	13.49	14.50	13.28	13.57	15.64	13.57	15.61
	Mean absolute error									
3m	11.62	**4.61**	22.86	22.83	13.55	14.22	22.89	16.59	22.89	16.75
6m	4.81	4.88	7.70	7.68	2.01	**0.25**	7.72	2.87	7.72	2.95
1y	8.83	**5.68**	6.56	6.55	7.26	7.12	6.53	8.17	6.53	8.04
2y	4.45	4.50	4.03	**4.00**	4.07	4.38	4.02	4.03	4.02	4.02
5y	7.46	4.72	6.12	6.09	5.33	**4.13**	6.12	6.06	6.11	5.95
10y	8.48	**6.01**	8.92	8.92	9.16	7.96	8.99	10.44	8.99	10.44
	Minimum error									
3m	-29.47	**-17.65**	-50.39	-55.73	-55.02	-60.41	-49.53	-55.45	-49.52	-54.76
6m	-31.26	-31.16	-36.08	-36.24	-16.05	**-2.26**	-36.66	-17.48	-36.63	-17.19
1y	-44.75	**-32.34**	-33.10	-33.08	-36.27	-38.90	-32.90	-37.25	-32.90	-36.63
2y	-33.32	-32.09	-32.70	-32.71	-32.22	-34.11	-33.10	-32.12	-33.07	**-31.82**
5y	-28.48	-30.09	-27.59	-27.51	-29.99	**-23.94**	-28.09	-29.00	-28.06	-28.87
10y	-62.52	**-34.59**	-67.22	-66.17	-71.24	-54.05	-67.28	-81.20	-67.37	-81.70

Continued on next page.

Table 7.1 – **continued**

	NS2	BC2	NS1-AR	NS1-VAR	BC1-AR	BC1-VAR	NS1-RS	BC1-RS	NS1-RS-X	BC1-RS-X
					Maximum error					
3m	73.53	**36.45**	164.61	165.41	111.68	101.08	160.03	116.96	160.00	117.90
6m	24.93	25.27	46.28	45.71	10.49	**1.24**	45.92	16.44	45.92	16.76
1y	**30.19**	40.48	36.41	36.00	40.86	49.61	35.83	39.31	35.85	39.82
2y	32.80	35.96	24.71	**24.43**	30.03	33.47	24.54	30.50	24.53	30.34
5y	46.36	**25.71**	42.82	43.06	36.95	27.29	41.85	42.91	41.90	42.78
10y	73.63	**69.54**	83.38	83.59	96.45	94.19	83.97	92.15	83.96	92.03
					$\hat{\rho}_1$					
3m	0.752	0.661	0.735	0.737	0.671	0.640	0.731	0.757	0.731	0.761
6m	0.491	0.498	0.543	0.546	0.294	0.106	0.539	0.408	0.540	0.419
1y	0.587	0.433	0.461	0.456	0.475	0.454	0.467	0.556	0.466	0.549
2y	0.467	0.459	0.450	0.448	0.451	0.448	0.451	0.457	0.450	0.453
5y	0.706	0.616	0.668	0.668	0.650	0.577	0.669	0.689	0.669	0.685
10y	0.706	0.641	0.683	0.685	0.703	0.680	0.688	0.733	0.689	0.733
					$\hat{\rho}_{12}$					
3m	0.413	0.503	0.378	0.378	0.459	0.497	0.375	0.519	0.375	0.521
6m	0.401	0.405	0.235	0.234	0.179	0.132	0.231	0.211	0.233	0.218
1y	0.338	0.310	0.301	0.295	0.313	0.311	0.307	0.353	0.306	0.349
2y	0.118	0.107	0.112	0.111	0.125	0.140	0.115	0.146	0.114	0.143
5y	-0.021	-0.018	-0.057	-0.058	-0.061	-0.080	-0.062	-0.022	-0.063	-0.023
10y	0.182	0.136	0.160	0.157	0.194	0.163	0.170	0.274	0.171	0.275

Notes: The table presents error statistics (in basis points) of the term structure models for selected maturities (first column) on the full sample (1962:1 - 2008:12). For each maturity and for selected evaluation measures, the number in **bold** corresponds to the best performing model. Autocorrelations with a time lag of 1 and 12 months are denoted by $\hat{\rho}_1$ and $\hat{\rho}_{12}$, respectively. For the two-step models (NS2 and BC2) the decay parameter, λ is fixed at 0.0609. For the other models, λ is estimated alongside the other parameters. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

7.1.2 Visualizing fitting capabilities

Figure 7.1 depicts in-sample fit of the average upward-sloping yield curve on a model-by-model basis. The figure corroborates that the BC models seem to fit a little better than the NS models, which can specifically be seen at the short end of the yield curve. Differences between NS models are hardly visible and the same holds for differences between the BC models.

Fitting capabilities become somewhat more apparent when zooming in on specific dates, as can be seen in Figures C.3 and C.4 in the appendix. All models show a similar fit when the curve is upward sloping (e.g., April, 1995), but the downward sloping curve of May, 1989 is fitted differently. Whereas two-step models and one-step NS models show a near-flat downward sloping line, one-step BC models show a hump in the short end. The hump-shaped curve of October, 1978 seems to be best fitted by BC2, as this model catches the 3-month rate reasonably well while all other models fail to do so. Moreover, BC2 is the only model to catch both humps in the October, 2007 curve; the other models only fit the second hump. These two figures are in line with results from Table 7.1 in the sense that all models fit remarkably well. Furthermore, BC2 seems to fit the more exotic yields curves somewhat better than the other models, although differences are marginal at most.

In conclusion, the following statements can be made with regard to in-sample fit. BC models, in general, fit the data better than NS models. No additional fit is gained from the more flexible VAR(1) specification, the AR(1) specification performs equally well. No in-sample benefits materialize from the one-step estimation procedure. Most notably, it can be concluded that the single-regime and regime-switching models fit the data equally well. Encouraged by these decent in-sample results the next section evaluates out-of-sample performance in a simulation-based setting.

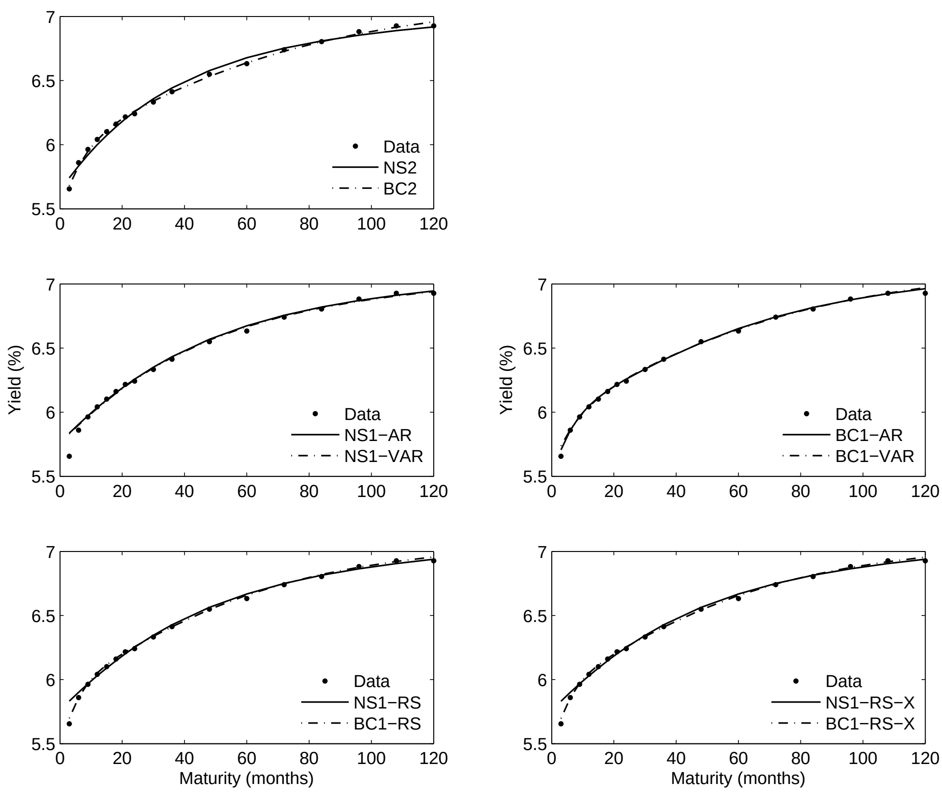

Figure 7.1: **Average fitted term structure.**

Notes: The figure shows the average fit on the full sample (1962:1 - 2008:12) of all term structure models. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

8 Out-of-sample forecasting

At this stage the question remains how the different models - particularly the regime-switching models - perform out of sample. Therefore, the two central questions of this section are: 'how do the regime-switching models perform out of sample relative to various benchmark models?' and 'what is the effect of macroeconomic data on forecasting accuracy?'. Furthermore, the section examines whether the additional BC factor increases out-of-sample performance. After all, the three-factor model might well perform better due to potential overfitting. Other issues that will be addressed relate to the effects on forecasting accuracy of the VAR(1) specification and the estimation procedure of the single-regime models.

8.1 Forecasting procedure: a simulation-based approach

To ensure sensible model parameters I use the full sample of 564 monthly observations to calibrate the models. I subsequently construct, for each model m, predictions of future yields at $h = 1, 6, 12$ and 24-month forecasting horizons. I study out-of-sample performance by simulating potential future yields and comparing them to the yield predictions. Following Bernadell et al. (2005), I construct these potential future yields by resampling the historical zero-coupon yield data. In order to preserve the time series properties of the historical data, I use the moving block bootstrap (MBB) method as proposed by Künsch (1989). To this end, I compute monthly absolute differences of the term structure evolution of which there are 563 in total. Setting the block length to 24 months - the largest forecast horizon - I hence construct 540 blocks. From these blocks I randomly draw $B = 500$ samples, with replacement. The simulation starts at the last observed time T (December, 2008). Refer to Appendix A.3 for additional details of the MBB.

Given the unprecedented low yields at the starting point of the out-of-sample study, quite a few sampled paths will result in negative yields. A negative yield (or a yield of zero percent for that matter), particularly on bonds with longer maturities, is unlikely to occur in practice, however. To overcome this issue I impose a lower bound on the future yield curve of a half times the minimum observed yield for each maturity of the full sample (1962:1 - 2008:12). In so doing, the lower bound is economically sound in the sense it is close to the historically low yields of 2011 and 2012.[1]

Figure 8.1 presents a three-dimensional plot of the simulated yields. All rates start from their respective December, 2008 levels and evolve according to the bootstrapped paths. It can be seen from the bottom plane (5^{th} percentile) that in some cases interest rates remain low i.e., they are close to (or attain) the lower bound. In other cases yields go up to levels above five percent, to be seen from the upper plane (95^{th} percentile). Figure 8.2 depicts the median yield with respect to both the number of simulations and the number of simulated months.

[1] For an overview of historical U.S. interest rates refer, for instance, to the U.S. Department of the Treasury: `http://www.treasury.gov/resource-center/data-chart-center/interest-rates`.

The median yield curve - in particular at the long end - is not too far off of the starting point of the simulation. The figure, moreover, reveals that yields are anywhere between almost zero (5^{th} percentile) and a little under five percent (95^{th} percentile).

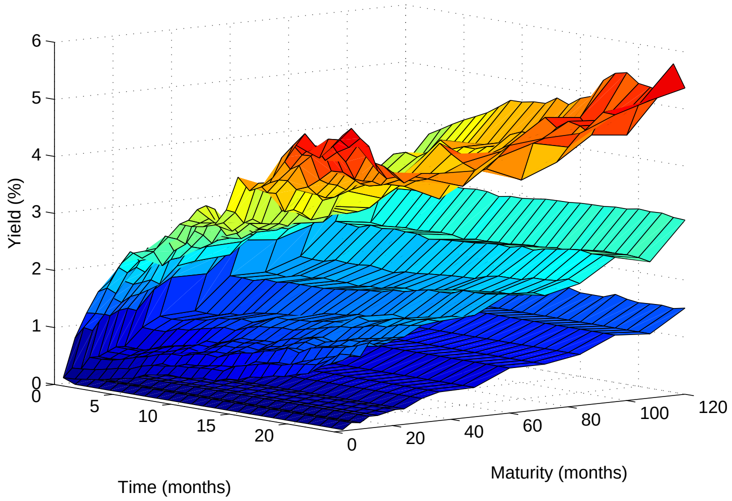

Figure 8.1: **Simulated evolution of yield curve.**

Notes: The figure presents the 5^{th}, 50^{th} and 95^{th} percentiles (corresponding to the lower, middle and upper planes, respectively) of 500 simulated yield curves. Simulated yields start at the last observed value (December, 2008) and are simulated 24 months ahead.

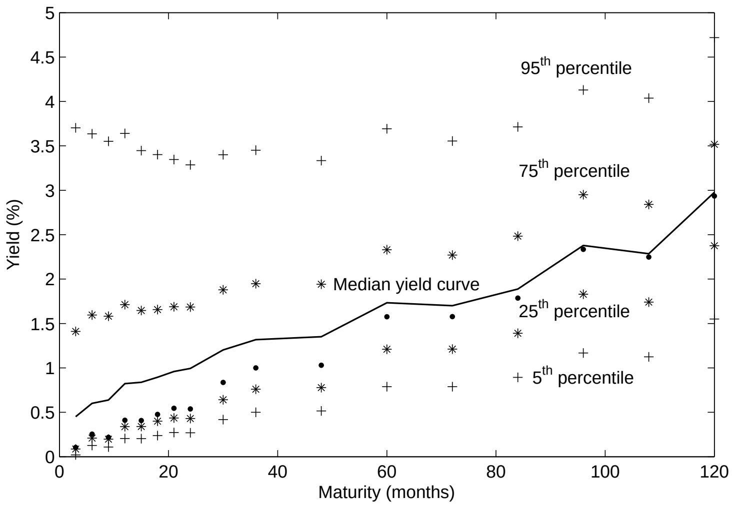

Figure 8.2: **Simulated median yield curve along with several percentiles.**

Notes: Percentiles calculated with respect to both the number of simulations (500) and the number of simulated months (24). Dots correspond to last observed yield (December, 2008).

Predictions of future yields are constructed by computing predictors for the factors of the state equations. Recall T corresponds to the last observed data point (December, 2008). For the single-regime models the h-month ahead vector of factor forecasts, $\hat{\beta}_{T+h}$ is found by iterating forward (3.7). That is,

$$\hat{\beta}_{T+h} = \left[I - \hat{F}^h\right]\left[I - \hat{F}\right]^{-1}\hat{\mu} + \hat{F}^h\beta_T, \tag{8.1}$$

where I denotes the identity matrix of size $K \times K$. State equation estimates, $\hat{F}$ and $\hat{\mu}$ have been presented before; they can be found in the tables in the appendix.

The non-switching AR(1) factors of the regime-switching models are predicted by (8.1) as well. In that case, I is a scalar of value 1. A predictor for the regime-switching yields-only factor, $\beta_{2,t}$ is constructed by taking its expectation at time $T + h$:

$$\begin{aligned}\hat{\beta}_{2,T+h} &= \mathbb{E}[\beta_{2,T+h}] \\ &= \mathbb{E}[\tilde{\mu}_{2,S_{T+h}} + \nu_{T+h}] \\ &= \mathbb{P}[S_{T+h} = 1 \mid \Psi_T] \cdot \hat{\mu}_{2,1} \quad + \quad \mathbb{P}[S_{T+h} = 2 \mid \Psi_T] \cdot \hat{\mu}_{2,2}.\end{aligned} \tag{8.2}$$

Hamilton (1994) proposes to estimate the state probabilities at time $T + h$ of the above equation by

$$\hat{\pi}_{T+h|T} = \hat{P}^h\hat{\pi}_{T|T}. \tag{8.3}$$

Here, $\hat{\pi}_{T|T}$ is found from (6.9) at time T and $\hat{P}$ denotes the estimated transition matrix. For the regime-switching macro-finance models the fixed transition probability matrix in (8.3) is replaced by its time-dependent counterpart i.e., it is replaced by $\hat{P}_T$, the transition matrix at time T.

With regard to the decay parameter (λ), I use the in-sample estimates for the one-step models and the Diebold and Li (2006) estimate of 0.0609 for the two-step models. h-month ahead yield predictions are now found by substituting (8.1) and/or (8.2) in the measurement equation:

$$\hat{y}_{T+h} = H\hat{\beta}_{T+h}.$$

Note that, since H is a function of $\hat{\lambda}$, it is in fact an estimate in itself.

8.2 Competing model: the random walk

It is known that in finance in general it is hard to consistently outperform a random walk (RW). In the case of term structures this is no different. Indeed, it is an intuitively appealing feature if a model is a better predictor than a 'drunk men's walk'. Therefore, I benchmark all term structure models against RW, which is given by

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \epsilon_t(\tau_i), \qquad \epsilon_t(\tau_i) \sim \mathcal{N}\Big(0, \sigma^2(\tau_i)\Big),$$

where $\sigma^2(\tau_i)$ captures the variance for each maturity τ_i, $i = 1, \ldots, N$. The h-month ahead forecast is given by $\hat{y}_{T+h}(\tau_i) = y_T(\tau_i)$, the yield at last observed time T.

8.3 Forecasting performance measures

To evaluate model performance in the simulation study, I follow De Pooter (2007) by reporting (Trace) Root Mean Squared Prediction Errors, or (T)RMSPEs. For all maturities, define the prediction error of bootstrap b and model m at time $T + h$ given information up to observed time T as

$$\varepsilon^m_{T+h|T,b}(\tau_i) \equiv \hat{y}^m_{T+h|T,b}(\tau_i) - y_{T+h}(\tau_i).$$

In so doing, the RMSPE, which reports statistics for each maturity, becomes

$$\text{RMSPE}_m(\tau_i) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\Big(\varepsilon^m_{T+h|T,b}(\tau_i)\Big)^2}.$$

The TRMSPE can be interpreted as the RMSPE averaged over all maturities and as such it captures a model's performance in a single number. To that end, I include all seventeen maturities. The statistic is given by

$$\text{TRMSPE}_m = \sqrt{\frac{1}{N}\frac{1}{B}\sum_{i=1}^{N}\sum_{b=1}^{B}\Big(\varepsilon^m_{T+h|T,b}(\tau_i)\Big)^2}.$$

In addition, I apply techniques from Diebold and Mariano (1995) to statistically test forecasting performance of the regime-switching models relative to the single-regime models. Under the null of the test proposed by Diebold and Mariano (1995), forecasting accuracy of two competing models is equal. To measure accuracy I now use another, closely related loss function, the mean squared prediction error (MSPE).[2] Formally, the Diebold-Mariano test is constructed on the basis of e.g., T_0 forecasts that are estimated at every time period using a moving window. Since I consider B bootstrapped blocks instead, the test that I apply deviates somewhat from the original test. That is to say, the formal test for equal predictive accuracy of two competing models, m_1 and m_2 becomes

$$\text{H}_0 : \mathbb{E}\Big[\big(\varepsilon^{m_1}_{T+h|T,b}\big)^2\Big] = \mathbb{E}\Big[\big(\varepsilon^{m_2}_{T+h|T,b}\big)^2\Big] \qquad \text{vs.} \qquad \text{H}_1 : \mathbb{E}\Big[\big(\varepsilon^{m_1}_{T+h|T,b}\big)^2\Big] \neq \mathbb{E}\Big[\big(\varepsilon^{m_2}_{T+h|T,b}\big)^2\Big].$$

[2] MSPE in conjunction with the Diebold-Mariano test is common practice. See e.g., Diebold and Li (2006).

Here, the notation for maturity (τ_i) is dropped in the interest of brevity. Given the loss differential, $d_b = \left(\varepsilon^{m_1}_{T+h|T,b}\right)^2 - \left(\varepsilon^{m_1}_{T+h|T,b}\right)^2$, the test statistic, S is

$$S = \frac{\bar{d}}{\left(\frac{1}{B}\hat{V}(\bar{d})\right)^{\frac{1}{2}}}, \quad \text{where} \quad \bar{d} = \frac{1}{B}\sum_{b=1}^{B} d_b, \qquad \hat{V}(\bar{d}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(d_b - \bar{d}\right)^2. \tag{8.4}$$

That is, $\bar{d}$ and $\hat{V}(\bar{d})$ denote the sample mean and variance, respectively.[3] Diebold and Mariano (1995) show the statistic, S is asymptotically standard normal distributed under the null. Therefore, an approximate two-sided test of size of α is to reject the null if

$$|S| > \Phi^{-1}(1 - \alpha/2),$$

where Φ^{-1} is the inverse of the cumulative distribution function of the standard normal distribution. It can be shown that the Diebold-Mariano test is only valid when two competing models, m_1 and m_2 are non-nested. Recall from Section 6.6 that the regime-switching models do not strictly nest the single-regime models implying it is safe indeed to apply the test.

I particularly deploy the above test to construct confidence intervals to graphically display the regime-switching models' out-of-sample performance relative to two single-regime benchmarks: NS1-VAR and BC1-VAR.

8.4 Forecasting results

Table 8.1 comprehensively reports simulation results for selected maturities at 1-month and 6-month forecasting horizons. Table 8.2 presents results at 12-month and 24-month horizons. The top rows of both tables list (T)RMSPEs of the random walk in basis points. It follows that RW's prediction errors are an increasing function of the forecasting horizon. This is a natural observation since the yield is likely to deviate more form the last observed value (i.e., the RW estimate) when the forecasting horizon increases. The other rows denote out-of-sample performance *relative* to the random walk. As such, numbers strictly smaller than 1 indicate outperformance of RW; they are denoted in bold.

What follows is an extensive description and interpretation of the simulation results. For completeness' sake, this includes an elaborate discussion on forecasting results of differences between single-regime models.

[3] Because prediction errors in the original Diebold-Mariano test use overlapping data, the loss differentials, d_t in that case are serially correlated for $h > 1$. To guarantee estimational consistency Diebold and Mariano (1995), therefore, actually used a different estimator of the variance. That is, they add a correction term to the estimator in (8.4): $\hat{V}(\bar{d}) = \hat{\zeta}_0 + 2\sum_{j=1}^{T_0-1}\hat{\zeta}_j$, where $\hat{\zeta}_j = T_0^{-1}\sum_{t=j+1}^{T_0}(d_t - \bar{d})(d_{t-j} - \bar{d})$. Since my loss differentials are constructed of bootstrapped data rather than of a time series, it is noted I do not take such correcting term into account. Besides, numerical differences between the two estimated variances in this case are low and as such it is safe to assume that statistical conclusions based on either of the two estimators to be quite alike.

8.4.1 1-month forecasting horizon

Panel A of Table 8.1 shows that root mean squared prediction errors of the random walk at a 1-month horizon are somewhere between 32 and 40 basis points. Beating the benchmark at this horizon turns out to be challenging, on average. That is to say, none of the models report a TRMSPE below 1. BC2-AR forecasts particularly poor with a TRMSPE of 1.80. The one-step Björk-Christensen AR model does not seem to deliver a great job either. NS2-AR and NS1-AR beat the drunk men's walk twice and once, respectively. With respective TRMSPEs of 1.21 and 1.19 their overall performance is still rather disappointing, however.

In general, it seems that predictive accuracy of most models - particularly the VAR and regime-switching models - decreases with maturity. This is also noted by De Pooter (2007) who, moreover, finds a reverse pattern for the AR models. My results, at least for $h = 1$, do not support this reverse pattern. This different outcome might be due to the sample period or the way the out-of-sample study is set up.[4]

The single-regime VAR models seem to forecast considerably better than their AR counterparts at the 1-month horizon. Except for BC2-VAR, all VAR models outperform RW at least at one maturity. Their traced RMSPEs are also considerably lower with values ranging between 1.06 and 1.11. This is in line with De Pooter (2007) who also finds the VAR models to be more accurate at this horizon (he finds this effect to be stronger for the two-step models).

With regard to the estimation procedure, I find the one-step models to yield better forecasts at the 1-month horizon, as can be seen from the TRMSPEs. This suggests that simultaneous inclusion of the factor and yield dynamics is imperative to term structure forecasting.

Results thus far are inconclusive about the potential forecasting gain of the fourth factor. That is, the BC-VAR and BC1-RS(-X) models report lower loss functions than the NS-VAR and NS1-RS(-X) models, respectively. This cannot be said for the models with an AR-specified state vector, however.

With TRMSPEs of 1.10 and 1.11 the regime-switching models forecast, on average, comparable to the VAR models. This is a promising outcome because the regime-switching models are in fact AR models and as such they forecast a great deal better than the single-regime AR models. What is reassuring as well is that the regime-switching models are the best predictors of the 3-month rate at the 1-month horizon. That is, BC1-RS, NS1-RS-X and BC1-RS-X outperform the random walk by 7%, 6% and 3%, respectively. NS2-VAR does also quite well with a 5% outperformance. The regime-switching models and single-regime VAR models tend to perform equally well at medium and long yields; however, the models are mostly unable to beat the random walk. The effect on predictive accuracy of the macroeconomic indicators is not clear-cut at the 1-month horizon as differences amongst the switching models are small.

[4] De Pooter (2007) uses unsmoothed Fama-Bliss U.S. zero-coupon yields constructed on the basis of end-of-month price quotes for U.S. treasuries spanning the period from January, 1984, up until and including December, 2003; maturities are the same. De Pooter (2007) assesses out-of-sample performance by recalibrating his models at each period in time using a moving window.

8.4.2 6-month forecasting horizon

Overall predictive accuracy increases noticeably when the forecasting horizon increases to six months, as can be seen from Panel B of Table 8.1. The TRMSPE of the random walk has increased to circa 89 basis points, which now seven of the twelve models are able to outperform. These seven include all four VAR models and three regime-switching models (NS1-RS reports a TRMSPE of 1.01). Overall, BC2-VAR and NS1-VAR are most accurate at this horizon as they both report an outperformance of 6%. Accuracy of the other five outperforming models follows closely, in particular the regime-switching macro-finance model BC1-RS-X with its 5% outperformance.

Zooming in on individual maturities, the random walk is primarily outperformed at short and medium yields. At the 3-month maturity the best performing models are NS1-VAR and BC1-RS with an outperformance of an impressive 14%. The other regime-switching models follow closely. 6-month and 1-year rates are best predicted by the regime-switching yields-only models. NS2-VAR and NS1-VAR forecast these maturities second best; other outperforming models (including the macro-finance models) are only marginally behind. The VAR models, in general, tend to produce the best medium-term predictions. It is fortunate to observe that the two regime-switching extensions of the three-factor NS model are most accurate when it comes to forecasting the long rate (10-year); the benchmark is not outperformed, however. Macroeconomic data seem to be beneficial to forecasting as TRMSPEs of the TVTP models are now lower than those of the FTP models.

Results at this horizon are somewhat less conclusive about the estimation procedure. That is, where at the 1-month horizon all one-step models outperformed their two-step counterparts, now BC2-VAR outperforms BC1-VAR. In all other instances the one-step models still beat their respective two-step variants. Again, forecasting accuracy of the single-regime AR models stays far behind the other models. In fact, they all report TRMSPEs that are worse than those at the 1-month horizon. Forecasting results of the effect of the fourth factor remain indistinct.

8.4.3 12-month and 24-month forecasting horizons

Table 8.2 shows that forecasting accuracy is improved further at longer horizons. Differences between the VAR models and regime-switching models at the 12-month horizon are quite small. For instance, the regime-switching BC1-RS model forecasts most accurate at the short end of the curve with approximately 18% outperformance, but the VAR models as well as the other regime-switching models are only marginally behind. A similar pattern can be seen at the 6-month rate where NS2-VAR and NS1-VAR are most accurate. Whereas 1-year rates are best predicted by NS2-VAR, NS1-VAR and BC1-RS-X, medium yields are forecast quite accurately by BC2-VAR and BC1-VAR.

Panel B reveals that the TRMSPE of the random walk at the 24-month prediction horizon amounts to approximately 173 basis points. The panel also reveals that BC1-VAR provides the most accurate forecasts overall at this horizon. Focusing on individual maturities, it is reassuring to observe that the regime-switching models are consistently amongst the top performers at the short end of the curve; outperformance of RW at these maturities amounts to 25%. It seems to become more common to outperform the random walk at the 10-year rate as five models (three of which regime-switching models) report RMSPEs below 1 at the 24-month forecasting horizon. With an outperformance of 2% the regime-switching macro-finance model BC1-RS-X forecasts most accurate at the long end.

The one-step models, in general, continue to provide more accurate predictions than the two-step models. Moreover, single-regime AR models remain strongly dominated by the VAR and multi-regime models. Out-of-sample results remain inconclusive about the effect of the additional (fourth) factor: whereas predictive accuracy increases for the VAR and multi-regime models, the opposite holds true for the single-regime models that assume an AR-specified state vector. This inconclusiveness actually differs from the outcomes of De Pooter (2007), who found the four-factor BC model - in particular with an AR-specified state vector - to outperform the other models. Again, this might well be the result of the sample period or the setup of the out-of-sample study, as outlined above.

At both 12- and 24-month forecasting horizons it follows that inclusion of leading economic indicators has a positive effect on overall forecasting accuracy: TRMSPEs of the two macro-finance models are lower than those of their respective regime-switching yields-only counterparts. It may hence be concluded that out-of-sample performance increases when regime-switching models take time-varying transition probabilities based on macroeconomic variables into account.

8.4.4 Focus on regime-switching models: Diebold-Mariano test

Given the integral presence of regime-switching models in the thesis, it is of specific interest to further look into their predictive accuracy. As outlined above, I hence additionally benchmark these models to two competing, single-regime term structure models: NS1-VAR and BC1-VAR. These benchmarks are preferred for at least three reasons. First, it is interesting to see how the newly proposed models directly compete against single-regime Nelson-Siegel and Björk-Christensen models. Second, NS1-VAR and BC1-VAR are one-step models - similar to the regime-switching models - and as such they are generally preferred in the existing literature (conform the discussion of Section 6.1). Third, it is clear from Tables 8.1 and 8.2 that the single-regime AR models are not quite suited to serve as a benchmark i.e., they are clearly dominated by their VAR counterparts as well as the regime-switching models.

Following Bernadell et al. (2005), Figures 8.3 and 8.4 display out-of-sample performance for selected maturities of the switching models relative to NS1-VAR and BC1-VAR, respectively. Solid lines denote the difference between mean squared prediction errors of a given switching model and the benchmark. Therefore, a number smaller than 0 means the regime-switching model outperforms the single-regime benchmark model. Dashed lines correspond to 95% confidence intervals which are constructed on the basis of the altered Diebold and Mariano (1995) test as was discussed in Section 8.3.

With the exception of NS1-RS-X, it follows from Figure 8.3 that the 3-month yield is most accurately predicted by the switching models at 1- and 24-month horizons. This result is significant for NS1-RS and BC1-RS. Results are not as clear-cut at the 1-year rate. Performance of the macro-finance models, for instance, is barely significant although BC1-RS-X beats the benchmark at the 24-month horizon. The graphs evidently show that the multi-regime models have a hard time beating NS1-VAR at the 5-year yield; the macro-finance models, however, seem to significantly outperform at the 24-month horizon. Turning attention to the 10-year rates, it can be seen NS1-RS and NS1-RS-X are the winners at short horizons; they fail to outperform at longer horizons. BC1-RS-X significantly outperforms at $h = 12$ and $h = 24$, albeit marginally.

To a certain extent similar results are obtained from Figure 8.4. All switching models but NS1-RS significantly predict the 3-month yield more accurate than the single-regime BC1-VAR model at 1- and 6-month horizons. Results for longer maturities are mostly insignificant with the exception of NS1-RS-X which is less accurate than the benchmark. Out-of-sample results for the 1-year rate are now less ambiguous in the sense that the multi-regime models mostly outperform the benchmark at 6- and 12-month horizons. This effect is stronger for the macro-finance models. Once more, the benchmark is the clear winner at the medium-term yield of 5-year. Judging from the vertical axes, however, the macro-finance models forecast this yield somewhat more accurate than the yields-only switching models. The tables have turned at the 10-year rate: for nearly all horizons the multi-regime models outperform BC1-VAR; the effect is not significant, however, at the 24-month horizon.

In conclusion, the following can be said with regard to the NS1-VAR benchmark. NS1-RS and NS1-RS-X tend to outperform at the tails of the yield curve, mainly at the 1-month horizon. At a 24-month horizon, NS1-RS (NS1-RS-X), moreover, beats the benchmark at the 3-month (5-year) rate. BC1-RS notably beats the benchmark at the short end of the curve for all horizons. BC1-RS-X tends to consistently outperform at the 24-month horizon.

Regarding the BC1-VAR benchmark the following concluding remarks can be made. NS1-RS is more accurate than the benchmark for the 3-month and 1-year yields at 1- and 6-month horizons. Moreover, the model outperforms the benchmark at the 10-year rate at 1-, 6-, and 12-month horizons. The same conclusion can be drawn for BS1-RS but significance seems to have increased. In fact, this conclusion also applies to the macro-finance models although the benchmark is now also outperformed at the 3-month and 1-year rates at the 12-month horizon. Except for the 5-year rate, the regime-switching macro-finance model BC1-RS-X seems to yield the most accurate predictions overall.

As such, Figures 8.3 and 8.4 show that neither the switching models nor the single-regime benchmarks consistently outperform one another. Hence, it is safe to say that in some cases the regime-switching models provide superior forecasts while in other instances the single-regime models are most accurate.

8.4.5 Summary of main out-of-sample results

Regarding out-of-sample performance as a whole it follows that, with the exception of the two BC-AR models, all term structure models are able to provide more accurate forecasts than a random walk at various forecasting horizons and maturities. Outperformance particularly occurs at short and medium yields and predictions become more accurate when the forecasting horizon increases. The latter might be due to notion that yield movements in the near future are more sensitive to various (e.g., political) factors which most models do not aim to account for (Schumacher, 2011). No model consistently outperforms the random walk (i.e., outperformance for all maturities and at all forecasting horizons).

Single-regime VAR models generally produce more precise yield forecasts than their AR counterparts. Furthermore, the one-step estimation procedure yields superior forecasts. Simulation results are inconclusive about the effect of the additional factor, $\beta_{4,t}$. That is to say, the fourth factor is beneficial to forecasting accuracy with the single-regime VAR and multi-regime models, but the opposite holds true for the single-regime AR models.

Regime-switching models provide the most accurate predictions multiple times and are amongst the best performing models in many instances. Moreover, inclusion of leading economic indicators has a positive effect on out-of-sample performance at forecasting horizons of six months and beyond. As such, it is worthwhile to jointly consider regime-switching models and existing, single-regime models when one's goal is to forecast the term structure of interest rates.

Table 8.1: **Out-of-sample forecasting results: 1-month and 6-month horizon.**

	Panel A: 1-month horizon								**Panel B: 6-month horizon**							
	TRMSPE	RMSPE							TRMSPE	RMSPE						
Maturity	**all**	**3m**	**6m**	**1y**	**3y**	**5y**	**7y**	**10y**	**all**	**3m**	**6m**	**1y**	**3y**	**5y**	**7y**	**10y**
RW	36.38	34.97	36.75	39.75	37.79	35.64	34.06	32.61	88.51	116.46	110.75	102.65	79.81	75.09	73.56	73.49
	Single-regime models															
NS2-AR	1.21	1.12	**0.97**	**0.97**	1.10	1.22	1.66	1.38	1.30	**0.98**	**0.98**	1.03	1.49	1.62	1.79	1.08
NS2-VAR	1.11	**0.95**	**0.97**	1.03	1.00	1.05	1.41	1.61	**0.96**	**0.87**	**0.88**	**0.87**	**0.98**	1.09	1.24	1.04
BC2-AR	1.80	1.77	1.75	1.74	1.50	1.52	2.19	1.00	2.35	1.99	2.07	2.19	2.59	2.54	2.66	1.66
BC2-VAR	1.07	1.01	1.00	1.00	1.08	1.02	1.21	1.54	**0.94**	**0.90**	**0.91**	**0.90**	**0.94**	**0.98**	1.05	1.16
NS1-AR	1.19	1.34	1.04	**0.98**	1.03	1.13	1.61	1.31	1.22	1.00	**0.98**	**0.99**	1.32	1.48	1.69	1.07
NS1-VAR	1.10	1.08	**0.95**	1.00	1.00	1.02	1.36	1.56	**0.94**	**0.86**	**0.87**	**0.88**	**0.95**	1.03	1.16	1.06
BC1-AR	1.23	1.04	1.08	1.11	1.04	1.11	1.64	1.26	1.50	1.29	1.34	1.39	1.54	1.61	1.83	1.13
BC1-VAR	1.06	**0.98**	1.00	1.00	1.08	1.04	1.20	1.50	**0.96**	**0.92**	**0.93**	**0.94**	**0.98**	1.00	1.01	1.21
	Regime-switching models															
NS1-RS	1.10	1.00	**0.95**	1.03	1.00	1.03	1.40	1.51	1.01	**0.87**	**0.87**	**0.87**	1.05	1.17	1.35	1.00
BC1-RS	1.11	**0.93**	**0.97**	1.00	1.00	1.04	1.37	1.71	**0.97**	**0.86**	**0.87**	**0.86**	1.00	1.11	1.24	1.04
NS1-RS-X	1.10	**0.94**	**0.99**	1.09	1.01	1.01	1.37	1.53	**0.97**	**0.88**	**0.90**	**0.89**	**0.97**	1.10	1.27	1.01
BC1-RS-X	1.11	**0.97**	1.03	1.06	1.00	1.01	1.32	1.75	**0.95**	**0.90**	**0.89**	**0.88**	**0.95**	1.05	1.18	1.07

Notes: The table presents 1-month (**Panel A**) and 6-month (**Panel B**) ahead out-of-sample forecasts of the estimated models for selected maturities. Forecasts are evaluated with (Trace) Root Mean Squared Prediction Errors - (T)RMSPEs. The top row reports (T)RMSPEs of the Random Walk (RW) in basis points. The other rows contain (T)RMSPEs *relative* to the random walk. Therefore, numbers smaller than 1 indicate outperformance of the random walk; they are presented in bold. The full sample (1962:1 - 2008:12) is used for parameter calibration. Future yields are simulated using the moving block bootstrap method of Künsch (1989); 500 blocks with a block-length of 24 months are drawn randomly, with replacement. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

Table 8.2: **Out-of-sample forecasting results: 12-month and 24-month horizon.**

	Panel A: 12-month horizon								**Panel B: 24-month horizon**							
	TRMSPE	RMSPE							TRMSPE	RMSPE						
Maturity	**all**	**3m**	**6m**	**1y**	**3y**	**5y**	**7y**	**10y**	**all**	**3m**	**6m**	**1y**	**3y**	**5y**	**7y**	**10y**
RW	119.61	151.38	147.72	141.19	109.45	97.79	94.19	98.28	173.15	223.59	220.49	210.27	158.34	137.66	129.05	127.21
	Single-regime models															
NS2-AR	1.41	1.05	1.06	1.12	1.61	1.81	1.97	1.29	1.31	**0.97**	**0.98**	1.03	1.48	1.71	1.87	1.48
NS2-VAR	**0.94**	**0.83**	**0.82**	**0.82**	**0.97**	1.12	1.26	1.00	**0.89**	**0.77**	**0.76**	**0.76**	**0.92**	1.08	1.21	1.03
BC2-AR	2.17	1.79	1.83	1.92	2.40	2.54	2.68	1.84	1.70	1.30	1.32	1.40	1.88	2.12	2.30	1.88
BC2-VAR	**0.88**	**0.83**	**0.83**	**0.83**	**0.89**	**0.98**	1.07	1.03	**0.82**	**0.75**	**0.75**	**0.74**	**0.83**	**0.95**	1.05	**0.99**
NS1-AR	1.33	1.06	1.04	1.07	1.47	1.69	1.88	1.27	1.27	**0.97**	**0.96**	1.00	1.40	1.63	1.81	1.45
NS1-VAR	**0.91**	**0.83**	**0.82**	**0.82**	**0.92**	1.04	1.15	1.01	**0.85**	**0.78**	**0.76**	**0.76**	**0.86**	**0.98**	1.08	**0.99**
BC1-AR	1.74	1.59	1.60	1.61	1.81	1.93	2.09	1.41	1.66	1.47	1.46	1.47	1.76	1.92	2.07	1.65
BC1-VAR	**0.89**	**0.84**	**0.85**	**0.86**	**0.90**	**0.96**	**0.99**	1.12	**0.79**	**0.75**	**0.75**	**0.75**	**0.79**	**0.88**	**0.93**	1.01
	Regime-switching models															
NS1-RS	1.01	**0.85**	**0.84**	**0.85**	1.07	1.23	1.38	1.01	**0.89**	**0.76**	**0.76**	**0.76**	**0.95**	1.09	1.21	1.02
BC1-RS	**0.96**	**0.82**	**0.83**	**0.84**	1.01	1.14	1.26	1.00	**0.86**	**0.75**	**0.75**	**0.76**	**0.90**	1.02	1.12	**0.99**
NS1-RS-X	**0.94**	**0.85**	**0.85**	**0.83**	**0.95**	1.11	1.28	1.00	**0.84**	**0.81**	**0.80**	**0.77**	**0.82**	**0.97**	1.10	**0.99**
BC1-RS-X	**0.92**	**0.83**	**0.83**	**0.82**	**0.94**	1.06	1.19	1.00	**0.83**	**0.76**	**0.75**	**0.74**	**0.85**	**0.97**	1.07	**0.98**

Notes: The table presents 12-month (**Panel A**) and 24-month (**Panel B**) ahead out-of-sample forecasts of the estimated models for selected maturities. Refer to Table 8.1 for a detailed description.

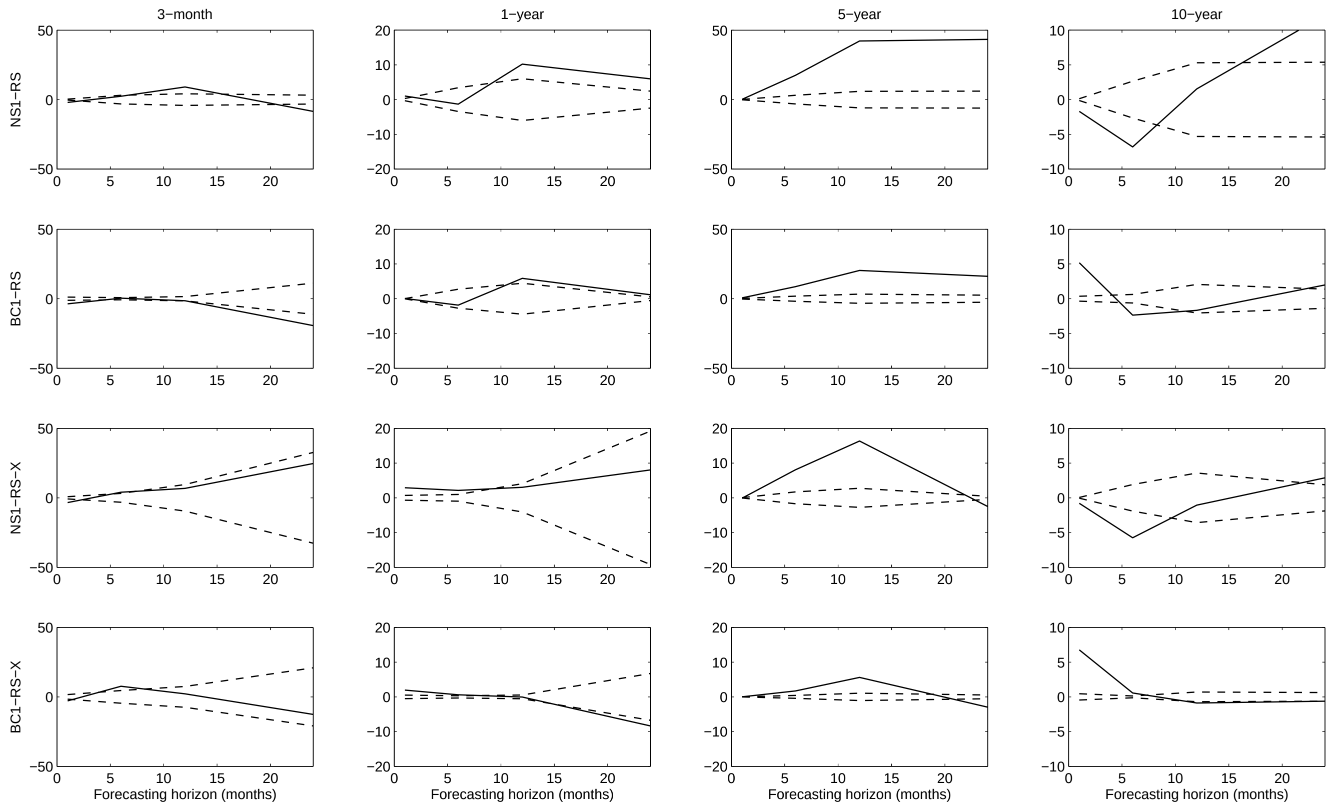

Figure 8.3: **Regime-switching models out-of-sample performance (basis points) - relative to NS1-VAR.**

Notes: The figure presents out-of-sample performance for selected maturities of the regime-switching models benchmarked against NS1-VAR. Solid lines denote the difference in basis points of the mean squared prediction errors of a particular regime-switching model and NS1-VAR. Therefore, values below 0 indicate the regime-switching model outperforms the benchmark. Dashed lines are 95% confidence intervals, which are computed using Diebold and Mariano (1995). The forecasting horizons are 1, 6, 12 and 24 months. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

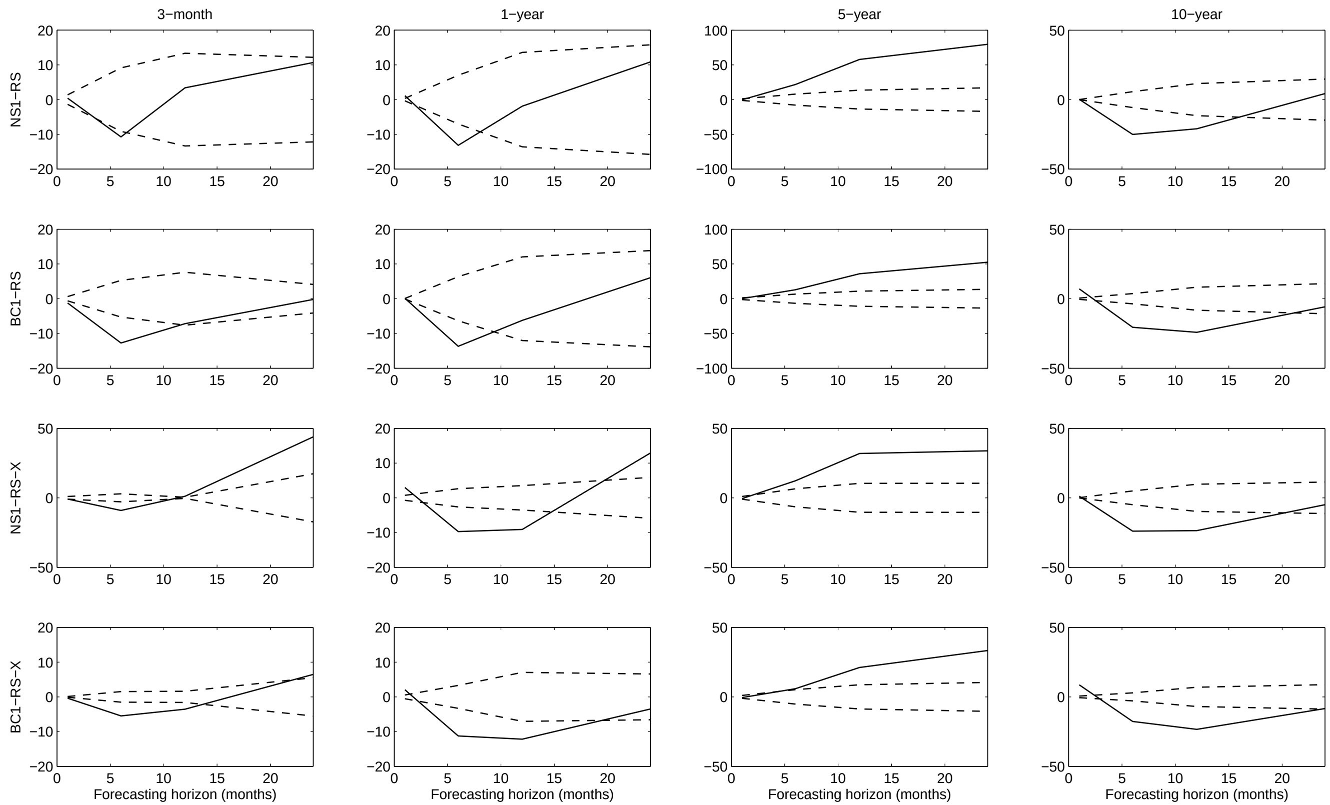

Figure 8.4: **Regime-switching models out-of-sample performance (basis points) - relative to BC1-VAR.**

Notes: The figure presents out-of-sample performance for selected maturities of the regime-switching models benchmarked against BC1-VAR. See Figure 8.3 for a detailed description.

8.5 Forecasting robustness check

Recall from Section 8.4 how 500 potential future yields with a length of 24 months have been constructed using the MBB. Indeed, as has been argued, in so doing one preserves the dynamic evolution of the historical yields. As a robustness check it is interesting, however, to study forecasting accuracy when this historic dynamic evolution is altered. Since it has been concluded that the term structure models are particularly able to outperform the random walk at longer forecasting horizons, I perform such check at the 24-month horizon.

Therefore, I construct potential future yields of length 24 months once more but now by concatenating two blocks of length 12 months. As such, the historical yield evolution is interrupted after 12 months only to be followed by another, unrelated (i.e., independent) path of historical yields of length 12 months. To achieve this, I independently apply the MBB two times, yielding 1000 blocks (i.e., two collections of 500 blocks) with a block length of 12 months each. Every pair of blocks that follows from the two collections is consequently concatenated such that 500 blocks of length 24 months are constructed.

On a technical note, let $\mathcal{B}_l^{i,b}$ denote a block of length l. Here, $b = \{1, \ldots, B = 500\}$ denotes the bootstrap and $i \in \{1, 2\}$ represents one of the two applications of the MBB. The two collections, aggregately comprising 1000 blocks, are thus concatenated as follows:

$$\Big\{ \underbrace{\Big[\mathcal{B}_{12}^{1,1},\ \mathcal{B}_{12}^{2,1}\Big]}_{\mathcal{B}_{24}^{*,1}},\ \underbrace{\Big[\mathcal{B}_{12}^{1,2},\ \mathcal{B}_{12}^{2,2}\Big]}_{\mathcal{B}_{24}^{*,1}},\ \ldots\ ,\ \underbrace{\Big[\mathcal{B}_{12}^{1,500},\ \mathcal{B}_{12}^{2,500}\Big]}_{\mathcal{B}_{24}^{*,500}} \Big\} = \Big\{ \mathcal{B}_{24}^{*,1},\ \mathcal{B}_{24}^{*,2},\ \ldots,\ \mathcal{B}_{24}^{*,500} \Big\}.$$

Here, $\mathcal{B}_{24}^{*,b}$ corresponds to concatenated bootstrap b of length 24 months i.e., a potential future yield evolution. The collection $\{\mathcal{B}_{24}^{*,1},\ \mathcal{B}_{24}^{*,2},\ \ldots,\ \mathcal{B}_{24}^{*,500}\}$ represents all 500 potential future yield movements.

Table 8.3 presents out-of-sample forecasting results based on the concatenated yields. Numbers in bold again indicate that the random walk is outperformed. These results may be directly compared to the results from Panel B of Table 8.2.

First of all, it can be seen that estimates of RW are quite similar. Forecasting results of the term structure models are comparable as well. That is to say, out-of-sample results turn out to be robust as differences are mostly small. In particular, models that outperformed the random walk do here too. Moreover, the regime-switching models remain amongst the top performing models, most notably for short-term yields.

It is interesting to observe that TRMSPEs of virtually all models are now somewhat lower. This effect, albeit still small in absolute terms, seems to be greatest for BC2-AR and BC1-AR where the difference in TRMSPE is approximately 0.5 in both instances. Looking at specific maturities, it follows that this difference mostly stems from gains at the short end of the curve. This effect is to a lesser extent also present for NS2-AR and NS1-AR. It is duly noted that the size of the effect might well be the result of the drawn sample.[5] For the other models, the lower reported RMSPEs are more or less equally spread across the different maturities. An exception to that is BC1-VAR, where differences are hardly existent in the first place.

I reemphasize that forecasting robustness results are much in line with the results from the actual forecasting study. As such, the fortunate observation can be made that conclusions from Section 8.4 about the models' forecasting accuracy remain in effect. This is particularly reassuring for the newly proposed regime-switching models as it was concluded that they are among the best performing models and in some instances even provide superior forecasts.

Table 8.3: **Forecasting robustness results: 24-month horizon.**

	TRMSPE	RMSPE						
Maturity	**all**	**3m**	**6m**	**1y**	**3y**	**5y**	**7y**	**10y**
RW	176.64	239.45	229.88	214.90	159.85	137.98	127.27	128.93
		Single-regime models						
NS2-AR	1.28	**0.91**	**0.94**	1.01	1.47	1.71	1.89	1.42
NS2-VAR	**0.87**	**0.75**	**0.75**	**0.75**	**0.93**	1.08	1.22	1.00
BC2-AR	1.65	1.20	1.25	1.36	1.87	2.12	2.34	1.81
BC2-VAR	**0.82**	**0.74**	**0.74**	**0.74**	**0.84**	**0.95**	1.06	**0.97**
NS1-AR	1.24	**0.91**	**0.93**	**0.98**	1.40	1.63	1.83	1.39
NS1-VAR	**0.83**	**0.76**	**0.75**	**0.75**	**0.87**	**0.98**	1.09	**0.97**
BC1-AR	1.61	1.35	1.38	1.43	1.75	1.91	2.09	1.58
BC1-VAR	**0.79**	**0.75**	**0.75**	**0.75**	**0.80**	**0.88**	**0.93**	1.01
		Regime-switching models						
NS1-RS	**0.88**	**0.75**	**0.75**	**0.76**	**0.95**	1.09	1.22	**0.99**
BC1-RS	**0.85**	**0.74**	**0.74**	**0.75**	**0.91**	1.02	1.13	**0.97**
NS1-RS-X	**0.84**	**0.81**	**0.80**	**0.77**	**0.83**	**0.97**	1.11	**0.97**
BC1-RS-X	**0.82**	**0.76**	**0.75**	**0.74**	**0.86**	**0.97**	1.08	**0.97**

Notes: The table presents 24-month ahead out-of-sample forecasts of the estimated models for selected maturities. Future yields are simulated by applying the moving block bootstrap twice. That is, two times 500 blocks of length 12 months are randomly drawn, with replacement. The 1000 resulting blocks are pairwise concatenated such that 500 blocks of length 24 months are constructed. As such, results in this table are a robustness check to the results of Panel B, Table 8.2. Refer to the text and Table 8.1 for further details.

[5] That is to say, the size of the effect might be lower/higher in another randomly drawn sample. This turns out to be the case indeed although differences are small here as well.

9 Conclusions

THIS SECTION CONCLUDES THE thesis by summarizing the approach that is taken and by summarizing the main empirical results and conclusions. In addition, possible directions for future research are offered.

9.1 Summary

This thesis proposes a regime-switching extension of the dynamic three-factor Nelson-Siegel and four-factor Björk-Christensen term structure models. These extended models are successfully estimated and consequently benchmarked against various single-regime models to evaluate in-sample fit and out-of-sample performance.

The proposed regime-switching models can be viewed as a twofold extension of Bernadell et al. (2005), who assume the constant of the slope factor, $\beta_{2,t}$ to follow a first-order Markov-switching process. First and foremost, their model is extended by assuming the Markov process is governed by a time-varying transition probability matrix that is driven on the basis of three leading macroeconomic indicators related to GDP, inflation and housing starts. Second, this time-varying transition probability extension is applied to the four-factor Björk-Christensen model. To examine the effect of the macroeconomic variables, the NS and BC switching models are also applied to the case where the Markov process is assumed to be governed by a fixed transition probability matrix. As such, a total of four regime-switching models are considered: two macro-finance models (i.e., based on TVTPs) and two yields-only models (i.e., based on FTPs).

Using zero-coupon U.S. yields (sample period: January, 1962 up until and including December, 2008), the Kim (1994) filter is used to estimate the regime-switching models. In so doing, all four models identify two distinct regimes that are readily interpretable. That is, the first regime is characterized by an upward sloping yield curve and hence corresponds to a regime of relatively normal economic activity. The second regime presents a flat yield curve and as such it is affiliated with an aberrant economic regime e.g., as a result of fiscal policy. Overall, filtered state probabilities tend to follow NBER recession periods reasonably well. In addition, the TVTP indicating the probability to stay in the aberrant regime is decreasing in times of NBER recessions which is intuitively appealing. Since the (average) transition probabilities of the active states of all switching models are in excess of 90 percent, it may be concluded that the identified regimes are persistent.

A total of eight competing, single-regime term structure models are considered. These competing models differ from each other in the way they are estimated (two-step vs. one-step), in the number of model factors (three-factor NS vs. four-factor BC) and in the specification of the state vector (AR vs. VAR).

Several statistical conclusions are drawn on the basis of the likelihood ratio test. Most notably, it is concluded that regime-switching yields-only (FTP) models are rejected in favor of regime-switching macro-finance (TVTP) models. This implies that overall fit increases when TVTPs based on macroeconomic data are modeled. Furthermore, LRT results suggest that a second regime is existent indeed. Results, moreover, strongly reject the three-factor NS model over the more flexible four-factor BC model; an AR-specified state vector is rejected in lieu of its VAR counterpart as well.

By computing multiple error statistics at various maturities, in-sample fit is examined further. As such, the reassuring conclusion is drawn that the (one-step) single-regime models and the regime-switching models fit the data equally well. In line with expectations, it is also concluded that an additional (fourth) model factor increases in-sample fit at most maturities. Furthermore, it is concluded from the maturity-dependent error statistics that estimational differences between single-regime models are small and that differences between single-regime state vectors (i.e., AR vs. VAR) do not substantially affect in-sample fit.

The thesis subsequently focuses on out-of-sample performance. To that end, each model computes an estimate of the future yield curve at 1, 6, 12 and 24 months forecasting horizons. These estimates are compared to potential future yields, which are constructed by bootstrapping historical yield data using the moving block bootstrap of Künsch (1989). As such, error statistics - more precisely, (Trace) Root Mean Squared Prediction Errors, (T)RMSPEs - can be constructed, which are compared to the (T)RMSPEs of a random walk.

Out-of-sample results are as follows. All regime-switching models and six out of eight single-regime models outperform the random walk at various maturities and forecasting horizons. These models are particularly able to outperform at short and medium yields. In general, outperformance increases when the forecasting horizon lengthens, although no model consistently outperforms the random walk.

It is concluded that a VAR-specified state vector and the one-step estimation procedure provide more accurate forecasts than an AR-specification and the two-step procedure, respectively. Forecasting results, however, are inconclusive about the effect of the additional, fourth factor.

Regarding predictive accuracy of the regime-switching models, it is concluded that these models forecast quite well. In particular, at the short end of the curve the switching models provide the most accurate forecasts numerous times. In many other instances the multi-regime models are amongst the best performing models. In addition, it is concluded that overall out-of-sample performance of the regime-switching macro-finance models exceeds out-of-sample performance of their yields-only counterparts at forecasting horizons of six months and beyond. This result corroborates with numerous related studies in the sense that inclusion of macroeconomic data is imperative to term structure forecasting.

The promising in-sample and out-of-sample results of the proposed regime-switching and macro-finance extensions as such might well further shift attention towards these kind of models to fit and forecast the term structure of interest rates.

9.2 Future research

The thesis raises a number of questions for future research. First, it is of interest to study the effect of the data sample on out-of-sample performance. In particular, given that the last observed yield coincides with an NBER recession, it cannot be excluded from the outset that the regime-switching models forecast differently when the last observed yield does not coincide with a recessionary period. In addition to that, it would be worthwhile to analyze subsample fitting and forecasting performance. It is noted, however, that if the amount of available yields is limited, such subsample analysis will most likely be impeded due to the highly parametrized character of some of the models.

Second, it would be contributive to asses out-of-sample performance by using a moving window and relate that to out-of-sample results of the moving block bootstrap. In so doing, one can look into potential biasing effects of the out-of-sample method itself.

Third, the question remains how the regime-switching macro-finance models perform as a strategic investment tool or as a stress-testing instrument. That is to say, a direction for future research is to study the sensitivity of the yield curve with respect to the macroeconomic variables. In that light, it will be interesting as well to examine the effect of different economic indicators. Or, for that matter, the effect of individual indicators e.g., the effect of an inflation-related indicator only.

Finally, the likelihood ratio test was used to provide some confidence of the presence of the second regime. Although this is a common approach, it is worthwhile to inquire into the possibilities of more accurate tests e.g., a standardized likelihood ratio test. If, for instance, such tests indicate the existence of more than two regimes it will be particularly interesting to adapt the model accordingly and evaluate in-sample fit and predictive accuracy on the basis of that model.

References

Ang, A. and Bekaert, G. (2002). Regime Switches in Interest Rates. *Journal of Business and Economic Statistics*, 20 : pp 163 - 182.

Ang, A. and Piazessi, M. (2003). A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50 : pp 745 - 787.

Bank for International Settlements (2005). Zero-Coupon Yield Curves: Technical Documentation. *Monetary and Economic Department. BIS papers No. 25.*

Bansal, R. and Zhou, H. (2002). Term Structure of Interest Rates with Regime Shifts. *Journal of Finance*, 57 : pp 1997 - 2043.

Bernadell, C., Coche, J. and Nyholm, K. (2005). Yield Curve Prediction for the Strategic Investor. *European Central Bank. Working paper No. 472.*

Björk, T. and Christensen, B. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 9 : pp 323 - 348.

Black, F. and Scholes, F. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81 : pp 637 - 654.

Bliss, R.R. (1996). Testing Term Structure Estimation Methods. *Federal Reserve Bank of Atlanta. Working paper No. 96-12a.*

Caks, J. (1977). The Coupon Effect on Yield to Maturity. *Journal of Finance*, 32 : pp 103 - 115.

Christensen, J.H.E., Diebold, F.X. and Rudebusch, G.d. (2011). The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. *Journal of Econometrics*, 164 : pp 4 - 20.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2008). How Arbitrage-Free is the Nelson-Siegel Model? *European Central Bank. Working paper No. 874.*

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rate. *Econometrica*, 53 : pp 385 - 407.

Dai, D., Singleton, K. and Yang, W. (2007). Regime shifts in a dynamic term structure model of US treasury bond yields. *Review of Financial Studies*, 20 : pp 1669 - 1706.

De Pooter, M. (2007). Examining the Nelson-Siegel Class of Term Structure Models. *Tinbergen Institute Discussion Paper. No. 2007-043/4.*

De Pooter, M., Ravazzolo, F. and Van Dijk, D. (2010). Term Structure Forecasting Using Macro Factors and Forecast Combination. *FRB International Finance Discussion Papers. No. 993.*

Diebold, F.X., Lee, J-H. and Weinbach, G.C. (1994). In: Hargreaves, C. (Ed.), Regime Switching with Time-Varying Transition Probabilities. Nonstationary Time Series Analysis and Cointegration. (Advanced Texts in Econometrics, Granger, C.W.J. and Mizon, G., eds.). *Oxford University Press* : pp 283 - 302.

Diebold, F.X. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 130 : pp 337 - 364.

Diebold, F.X. and Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13 : pp 253 - 263.

Diebold, F.X., Rudebusch, G.D. and Aruoba, B. (2006). The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach. *Journal of Econometrics*, 131 : pp 309 - 338.

Duffee, G.R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance*, 57 : pp 405 - 443.

Duffee, G.R. (2011). Forecasting with the term structure: The role of no-arbitrage restrictions. *Economics Working Paper Archive 576.* The Johns Hopkins University, Department of Economics.

Duffie, D. and Kan, R. (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6 : pp 379 - 406.

Fama, E. and Bliss, R. (1987). The Information in Long-Maturity Forward Rates. *American Economic Review*, 77 : pp 680 - 692.

Filardo, A.J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Statistics*, 12 : pp 299 - 308.

Filardo, A.J. (1998). Choosing Information Variables for Transition Probabilities In a Time-Varying Transition Probability Switching Model. *Federal Reserve Bank of Kansas City. RWP 98-09.*

Garcia, R. (1998). Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. *International Economic Review*, 39 : pp 763 - 788.

Gelman, S. and Wilfling, B. (2009). Markov-Switching in Target Stocks During Takeover Bids. *Journal of Empirical Finance*, 16 : pp 745 - 758.

Goldfeld, S.M. and Quandt, R.E. (1973). A Markov Model for Switching Regressions. *Journal of Econometrics*, 1 : pp 3 - 16.

Gordon, K. and Smith, A.F.M. (1988). Modeling and Monitoring Discontinuous Changes in Time Series. *In Bayesian Analysis of Time Series and Dynamic Linear Models, ed. J. C. Spall* : pp 359 - 392.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42 : pp 27 - 62.

Hamilton, J.D. (1989). A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57 : pp 357 - 384.

Hamilton, J.D. (1994). Time Series Analysis. *Princeton University Press.*

Hansen, B.E. (1992). The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics*, 7 : pp S61 - S82.

Hansen, B.E. (1996). Erratum: The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics*, 11 : pp 195 - 198.

Hull, J. and White, A. (1990). Pricing Interest-Rate-Derivative Securities. *The Review of Financial Studies*, 3 : pp 573 - 592.

Huse, C. (2011). Term Structure Modelling with Observable State Variables. *Journal of Banking & Finance*, 35 : pp 3240 - 3252.

Joslin, S., Singleton, K.J. and Zhu, H. (2011). A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies*, 24 : pp 926 - 970.

KALMAN, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering, Transactions of the ASME*, 82 (Series D) : pp 35 - 45.

KOOPMAN, S.J., MALLEE, M.I.P. AND VAN DER WEL, M. (2010). Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters *Journal of Business and Economic Statistics*, 28 : pp 329 - 343.

KIM, C.J. (1994). Dynamic Linear Models with Markov-Switching. *Journal of Econometrics*, 60 : pp 1 - 22.

KIM, C.J. AND NELSON, C.R. (1999). State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. *MIT Press.*

KIM, Y. AND SCHMIDT, P. (2000). A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data. *Journal of Productivity Analysis*, 14 : pp 91 - 118.

KÜNSCH, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics*, 17 : pp 1217 - 1241.

MCCULLOCH, J.H. (1975). The tax adjusted yield curve. *Journal of Finance*, 30 : pp 811 - 830.

NELSON, C.R. AND SIEGEL, A.F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, 60 : pp 473 - 489.

NYHOLM, K. (2008). Strategic Asset Allocation in Fixed-Income Markets: a Matlab-Based User's Guide. *Chichester: John Wiley & Sons Ltd.*

SCHUMACHER, J.M. (2011). Financial Models. *Tilburg University, Tilburg, the Netherlands.*

STOCK, J.H. AND WATSON, M.W. (1989). In: Blanchard, O.J. and Fisher, S. (Ed.), New Indexes of Coincident and Leading Economic Indicators. NBER Macroeconomics annual 1989. *MIT Press*, 4 : pp 351 - 393.

SVENSSON, L.E.O. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. *NBER Working Paper Series, 4871.*

TAYLOR, J.B. (1993). Discretion version policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39 : pp 195 - 214.

VASICEK, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5 : pp 177 - 188.

VASICEK, O.A. AND FONG, H.G. (1982). Term Structure Modeling using Exponential Splines. *Journal of Finance*, 37 : pp 339 - 348.

WILFLING, B. (2009). Volatility Regime-Switching in European Exchange Rates Prior to Monetary Unification. *Journal of International Money and Finance*, 28 : pp 240 - 270.

ZANTEDESCHI, D., DAMIEN P. AND POLSON, N.G. (2011). Predictive Macro-Finance with Dynamic Partition Models. *Journal of the American Statistical Association*, 106 : pp 427 - 439.

ZHU, X. AND RAHMAN, S. (2009). A Regime Switching Macro-Finance Model of the Term Structure. *Working Paper. Nanyang Technological University.*

Appendices

A Econometric techniques

A.1 Kalman filter

Kalman (1960) proposes a recursive algorithm, the Kalman filter, that may be used to estimate the parameters of a state-space model at time $t = 1, 2, \ldots, T$. Gains in computing speed contributed to the algorithm's popularity, which resulted in several extensions of the filter e.g., the Kalman-Bucy filter. Kalman's algorithm, first used by Diebold et al. (2006) in conjunction with Nelson-Siegel type of term structure models, finds application in many other fields e.g., the Apollo space program, navigation of cruise missiles and weather forecasting.[1]

In the light of recognizability, this section partly follows steps and notations from Kim and Nelson (1999). Consider the following general state-space framework at time $t = 1, 2, \ldots, T$. Measurement equation:

$$y_t = H_t \beta_t + A z_t + \epsilon_t. \tag{A.1}$$

Transition equation:

$$\beta_t = \tilde{\mu} + F \beta_{t-1} + \nu_t. \tag{A.2}$$

Assume the disturbances are white Gaussian noise and mutually uncorrelated:

$$\begin{pmatrix} \epsilon_t \\ \upsilon_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix}\right]. \tag{A.3}$$

Furthermore, assume y_t is a vector of size $N \times 1$ and β_t a vector of size $K \times 1$. Note the strong similarity of this framework to the Nelson-Siegel model in state-space form as presented in (3.6) - (3.7). Indeed, the NS representation is a special case of this framework.

The filtering procedure can be split in a *prediction* step and an *updating* step. In order to properly introduce the Kalman filter, however, first fix notation by defining:

- ▷ Ψ_t. Collection of all available information up to and including time t.
- ▷ $\beta_{t|t-1} = \mathbb{E}[\beta_t \mid \Psi_{t-1}]$. Expectation of β_t, conditional on the information set, Ψ_{t-1}.
- ▷ $\beta_{t|t} = \mathbb{E}[\beta_t \mid \Psi_t]$. Expectation of β_t, conditional on the information set, Ψ_t.
- ▷ $P_{t|t-1} = \mathbb{E}[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})' \mid \Psi_{t-1}]$. Covariance matrix of β_t, conditional on the information set, Ψ_{t-1}.
- ▷ $P_{t|t} = \mathbb{E}[(\beta_t - \beta_{t|t})(\beta_t - \beta_{t|t})' \mid \Psi_t]$. Covariance matrix of β_t, conditional on the information set, Ψ_t.
- ▷ $y_{t|t-1} = \mathbb{E}[y_t \mid \Psi_{t-1}] = H_t \beta_{t|t-1} + A z_t$. Forecast of y_t, conditional on the information set, Ψ_{t-1}.

[1] For other applications, refer to e.g., the Department of Computer Science at University of North Carolina - Chapel Hill: `http://www.cs.unc.edu/~welch/kalman/`.

▷ $\eta_{t|t-1} = y_t - y_{t|t-1}$. Prediction error.

▷ $f_{t|t-1} = \mathbb{E}[\eta^2_{t|t-1}]$. Variance of the prediction error.

Prediction step

The Kalman filtering procedure first constructs an estimate of β_t at the beginning of time t (i.e., based on information of time $t-1$), denoted by $\beta_{t|t-1}$. Similarly, the filter computes an initial prediction of the covariance matrix of β_t (i.e., $P_{t|t-1}$), which one uses in subsequent steps to update the estimate of the unobserved state vector, β_t. The two prediction steps are:

1. $\beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1}$,
2. $P_{t|t-1} = FP_{t-1|t-1}F' + Q$.

Updating step

At the end of time t (when y_t becomes observable), the prediction error, $\eta_{t|t-1}$ can be determined. This prediction error, in turn, is used to form a better estimate of the state vector. This materializes by correcting the initial estimate of β_t by a factor K, which is known as the Kalman gain. It can be shown that $K_t = P_{t|t-1}H'_t f^{-1}_{t|t-1}$ and as such it can interpreted as the weight assigned to information that became available in the period between $t-1$ and t.[2] The updating phase of the basic filter comprises the following four steps:

3. $\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t$,
4. $f_{t|t-1} = H_tP_{t|t-1}H'_t + R$,
5. $\beta_{t|t} = \beta_{t|t-1} + K_t\eta_{t|t-1}$,
6. $P_{t|t} = P_{t|t-1} - K_tH_tP_{t|t-1}$.

By iterating the six steps in the predication and updating steps for time $t = 1, 2, \ldots, T$, the Kalman filter produces a mean squared error estimate for the state vector, β_t at every time point t. Note the estimate at time t only uses information that is then available.[3]

Initialization of the filter

Starting values for β_t and P_t are required to initiate the Kalman filter. In general, β_t is assumed to be stationary, making the unconditional mean and covariance matrix of β_t a suitable choice for the starting values of β_t and P_t, respectively. That is,

$$\beta_{0|0} = (I - F)^{-1}\tilde{\mu}, \tag{A.4}$$

[2] It is noted this thesis does not provide a proof of the Kalman filter, refer to e.g., Hamilton (1989, 1994) and Kim and Nelson (1999) instead.

[3] A closely related procedure, the Kalman smoother, estimates the state vector at time t using information of the full sample. Most smoothing algorithms iterate backwards in time and as such they use the estimates that result from the Kalman filter. Because the Kalman smoother is not used in the thesis it is not further discussed here.

where it should be recalled that I denotes the identity matrix and X^{-1} the inverse of a matrix X. The covariance matrix may be initiated by

$$vec(P_{0|0}) = (I - F \otimes F)^{-1} vec(Q). \tag{A.5}$$

Here, $vec(A)$ is the vectorization of a matrix A and $A \otimes B$ the Kronecker product of the matrices A and B.

Estimation of the model's parameters

Recall that the parameter set is given by $\theta = (\theta_1, \ldots, \theta_s)$ and that $l(\theta)$ denotes the log-likelihood function. Using the prediction-error decomposition of the likelihood it can be shown the sample log-likelihood function that is to be maximized is given by

$$l(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \ln(2\pi|f_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{T} \eta'_{t|t-1} f^{-1}_{t|t-1} \eta_{t|t-1},$$

where $|X|$ refers to the determinant a matrix X.

A.2 Kim filter

Let S_t denote regime dependency and extend the framework in (A.1) - (A.3) by allowing for a regime switch in both the measurement and transition equation.[4]
Measurement equation:

$$y_t = H_{S_t}\beta_t + A_{S_t} z_t + \epsilon_t.$$

Transition equation:

$$\beta_t = \tilde{\mu}_{S_t} + F_{S_t}\beta_{t-1} + G_{S_t}\nu_t.$$

Underlying assumptions:

$$\begin{pmatrix} \epsilon_t \\ \upsilon_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R_{S_t} & 0 \\ 0 & Q_{S_t} \end{pmatrix}\right].$$

Kim (1994) proposes a recursive algorithm, the Kim filter, to estimate this state-space model. Each iteration of the algorithm consists of three consecutive steps: the Kalman filtering step, the Hamilton filtering step and the collapsing step.

Expanded Kalman filter

The estimation procedure of the above state-space system should now take information into consideration with respect to the unobserved first-order Markov-switching process, S_t. Therefore, one has to expand the notation of the original Kalman filter from Section A.1. Without loss of generality, let states, $(i, j) = 1, 2, \ldots, M$ and fix:

[4] The term structure models proposed in this thesis allow for a regime switch in the transition equation only. In the interest of a general presentation of the Hamilton filter, however, the measurement equation is allowed to switch as well. This is in line with Kim and Nelson (1999) whom this section closely follows.

▷ $\beta_{t|t-1}^{(i,j)} = \mathbb{E}[\beta_t \mid \Psi_{t-1}, S_t = j, S_{t-1} = i]$. Expectation of β_t, conditional on information up to and including time $t-1$, $S_t = j$ and $S_{t-1} = i$.

▷ $P_{t|t-1}^{(i,j)} = \mathbb{E}[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})' \mid \Psi_{t-1}, S_t = j, S_{t-1} = i]$. Covariance matrix of β_t, conditional on information up to and including time $t-1$, $S_t = j$ and $S_{t-1} = i$. This may be interpreted as a mean squared error matrix.

Given $S_t = j$ and $S_{t-1} = i$, the six recursive steps from the Kalman filter in Section A.1 are now to be extended to:

1. $\beta_{t|t-1}^{(i,j)} = \tilde{\mu}_j + F_j \beta_{t-1|t-1}^{i}$,
2. $P_{t|t-1}^{(i,j)} = F_j P_{t-1|t-1}^{i} F_j' + G_j Q_j G_j'$,
3. $\eta_{t|t-1}^{(i,j)} = y_t - H_j \beta_{t|t-1}^{(i,j)} - A_j z_t$,
4. $f_{t|t-1}^{(i,j)} = H_j P_{t|t-1}^{(i,j)} H_j' + R_j$,
5. $\beta_{t|t}^{(i,j)} = \beta_{t|t-1}^{(i,j)} + P_{t|t-1}^{(i,j)} H_j' \big[f_{t|t-1}^{(i,j)}\big]^{-1} \eta_{t|t-1}^{(i,j)}$,
6. $P_{t|t}^{(i,j)} = \Big(I - P_{t|t-1}^{(i,j)} H_j' \big[f_{t|t-1}^{(i,j)}\big]^{-1} H_j\Big) P_{t|t-1}^{(i,j)}$.

Collapsing the estimates

Amongst others, Gordon and Smith (1988) as cited in Kim and Nelson (1999) argue that this filtering algorithm rapidly grows out of bounds in the number of states. To see this, let the number of states be $M = 2$. In so doing, the number of cases to consider amounts to over 32,500 after $t = 15$ iterations already. One way to overcome this caveat, as proposed by Kim and Nelson (1999), is to collapse the posteriors $\beta_{t|t}^{(i,j)}$ and $P_{t|t}^{(i,j)}$ (both of which there are M^2) into $\beta_{t|t}^{j}$ and $P_{t|t}^{j}$ such that there are M collapsed posteriors. In particular, they propose the following *approximations* after each iteration for all j:

$$\beta_{t|t}^{j} = \frac{\sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t] \beta_{t|t}^{(i,j)}}{\mathbb{P}[S_t = j \mid \Psi_t]}, \tag{A.6}$$

$$P_{t|t}^{j} = \frac{\sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t] \big(P_{t|t}^{(i,j)} + (\beta_{t|t}^{j} - \beta_{t|t}^{(i,j)})(\beta_{t|t}^{j} - \beta_{t|t}^{(i,j)})'\big)}{\mathbb{P}[S_t = j \mid \Psi_t]}. \tag{A.7}$$

Note these expressions are fed into the first two steps of the expanded Kalman filter in the next iteration.

Hamilton filter

Hamilton (1989) introduces a scheme, the Hamilton filter, that may be used to make inference about the probabilistic terms in the right hand sides of (A.6) and (A.7). The Hamilton filter, which can be viewed as an extension of the work of Goldfeld and Quandt (1973), starts by predicting the joint probability of S_t and S_{t-1} at the beginning of the t-th iteration in the following way:

$$\mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_{t-1}] = \underbrace{\mathbb{P}[S_t = j \mid S_{t-1} = i]}_{=p_{ij}} \mathbb{P}[S_{t-1} = i \mid \Psi_{t-1}]. \tag{A.8}$$

Note that p_{ij} is an element of the transition probability matrix, P as in e.g., (4.1).

Second, the joint density (denoted by f) of y_t, S_t and S_{t-1} is considered, which in turn results in the marginal density of y_t. That is,

$$f(y_t, S_t = j, S_{t-1} = i \mid \Psi_{t-1}) = f(y_t \mid S_t = j, S_{t-1} = i, \Psi_{t-1})\mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_{t-1}],$$

where it can be shown that the conditional density is given by

$$f(y_t \mid S_t = j, S_{t-1} = i, \Psi_{t-1}) = (2\pi)^{-\frac{T}{2}} |f_{t|t-1}^{(i,j)}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\eta_{t|t-1}^{(i,j)\prime}(f_{t|t-1}^{(i,j)})^{-1}\eta_{t|t-1}^{(i,j)}\Big).$$

As such, the marginal density becomes

$$f(y_t \mid \Psi_{t-1}) = \sum_{i=1}^{M}\sum_{j=1}^{M} f(y_t \mid S_t = j, S_{t-1} = i, \Psi_{t-1})\mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_{t-1}]. \tag{A.9}$$

Third, when new information at the end of time t becomes available i.e., when y_t becomes observable, the Hamilton filter updates the predicted join probability from (A.8) as follows:

$$\mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t] = \frac{f(y_t, S_t = j, S_{t-1} = i \mid \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})}, \tag{A.10}$$

with

$$\mathbb{P}[S_t = j \mid \Psi_t] = \sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t].$$

Once the Kim filtering procedure is completed, an estimate for the state vector follows by taking a probability-weighted average of the regime-dependent state vectors, $\beta_{t|t}^{j}$:

$$\beta_{t|t} = \sum_{j=1}^{M} \mathbb{P}[S_t = j \mid \Psi_t]\beta_{t|t}^{j}.$$

Hamilton filter in vector notation

The above introduction of the Hamilton filter is presented on an element-by-element basis, as opposed to the matrix notation as in e.g., Hamilton (1994). Indeed, an element-by-element representation is intuitive and hence preferred as a first introduction. It is convenient, however, (e.g., from a programmer's point of view) to consider the Hamilton filter in vector notation as well. Recall the prediction equation in (A.8) and sum both sides over i:

$$\sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_{t-1}] = \sum_{i=1}^{M} \underbrace{\mathbb{P}[S_t = j \mid S_{t-1} = i]}_{=p_{ij}} \mathbb{P}[S_{t-1} = i \mid \Psi_{t-1}] \iff$$

$$\mathbb{P}[S_t = j \mid \Psi_{t-1}] = \sum_{i=1}^{M} \mathbb{P}[S_t = j \mid S_{t-1} = i]\mathbb{P}[S_{t-1} = i \mid \Psi_{t-1}]. \tag{A.11}$$

Considering now not just the predicted j-th state probability but the collection of all M elements, the LHS of (A.11) becomes an $M \times 1$ vector, denoted by $\pi_{t|t-1}$. Similarly, collect the state probabilities of the RHS in the $M \times 1$ vector $\pi_{t-1|t-1}$, and the corresponding transition probabilities in the $M \times M$ matrix P. In so doing, the prediction equation of the Hamilton filter may be written as

$$\begin{pmatrix} \mathbb{P}[S_t = 1 \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = j \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = M \mid \Psi_{t-1}] \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{j1} & \cdots & p_{M1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1j} & \cdots & p_{jj} & \cdots & p_{Mj} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1M} & \cdots & p_{jM} & \cdots & p_{MM} \end{pmatrix} \begin{pmatrix} \mathbb{P}[S_{t-1} = 1 \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_{t-1} = j \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_{t-1} = M \mid \Psi_{t-1}] \end{pmatrix}.$$

Or, equivalently,

$$\pi_{t|t-1} = P\pi_{t-1|t-1}.$$

Therefore, it follows that the prediction step of the element-by-element representation may be interpreted as a prediction for a fixed S_{t-1}, as opposed to the matrix notation that considers the prediction for all $S_{t-1} \in \{1, 2, \ldots, M\}$.

A similar argument can be made to show how the element-by-element representation of the updating step fits into matrix notation. To this end, recall the updating equation in (A.10) and sum both sides over i. Then, element j of the state probability vector, $\pi_{t|t}$ becomes

$$\sum_{i=1}^{M} \mathbb{P}[S_t = j, S_{t-1} = i \mid \Psi_t] = \sum_{i=1}^{M} \frac{f(y_t, S_t = j, S_{t-1} = i \mid \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})} \iff$$

$$\mathbb{P}[S_t = j \mid \Psi_t] = \frac{f(y_t, S_t = j \mid \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})} \iff$$

$$\vdots \qquad = \frac{\mathbb{P}[S_t = j \mid \Psi_{t-1}]f(y_t \mid S_t = j, \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})} \iff$$

$$\mathbb{P}[S_t = j \mid \Psi_t] = \frac{\mathbb{P}[S_t = j \mid \Psi_{t-1}]f(y_t \mid S_t = j, \Psi_{t-1})}{\sum_{i=1}^{M} \mathbb{P}[S_t = j \mid \Psi_{t-1}]f(y_t \mid S_t = j, \Psi_{t-1})}. \tag{A.12}$$

Considering now the collection of all M elements, the numerator of the RHS of (A.12) may be written as

$$\begin{pmatrix} \mathbb{P}[S_t = 1 \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = j \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = M \mid \Psi_{t-1}] \end{pmatrix} \odot \begin{pmatrix} f(y_t \mid S_t = 1, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = j, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = M, \Psi_{t-1}) \end{pmatrix}.$$

Or in shorthand notation as

$$\pi_{t|t-1} \odot D_t.$$

Likewise, the denominator becomes

$$\begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \end{pmatrix}_{1\times M} \left[\begin{pmatrix} \mathbb{P}[S_t = 1 \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = j \mid \Psi_{t-1}] \\ \vdots \\ \mathbb{P}[S_t = M \mid \Psi_{t-1}] \end{pmatrix} \odot \begin{pmatrix} f(y_t \mid S_t = 1, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = j, \Psi_{t-1}) \\ \vdots \\ f(y_t \mid S_t = M, \Psi_{t-1}) \end{pmatrix} \right],$$

which in shorthand can be written as

$$\mathbb{1}'(\pi_{t|t-1} \odot D_t).$$

Combining the result one readily finds the familiar expression of the updating equation as in e.g., Hamilton (1994):

$$\pi_{t|t} = \frac{(\pi_{t|t-1} \odot D_t)}{\mathbb{1}'(\pi_{t|t-1} \odot D_t)}.$$

Initialization and optimization of the Kim filter

Starting values for β_t and P_t, $\beta^j_{0|0}$ and $P^j_{0|0}$, respectively are obtained in a similar way as described in Section A.1. In the two-state case, starting values for $\mathbb{P}[S_0 = j \mid \Psi_0]$, $j \in \{1, 2\}$ may be obtained from the steady-state probabilities of S_t. That is, the unconditional[5] probability is well-known to be given by

$$\mathbb{P}[S_0 = j \mid \Psi_0] = \frac{1 - p_{(3-j)(3-j)}}{2 - p_{11} - p_{22}} \qquad \forall\, j \in \{1, 2\}. \tag{A.13}$$

Note that from the above relation it follows that the marginal density and hence the likelihood function is a function of the transition probabilities as well.

Since (A.9) computes the density of y_t conditional on Ψ_{t-1} (i.e., $f(y_t|\Psi_{t-1})$) the log-

[5] Unconditional on information about the states. Other information, collected in Ψ may be considered, however.

likelihood function, $l(\theta)$ is readily computed:

$$\begin{aligned} l(\theta) &= \sum_{t=1}^{T} \ln\left[f(y_t \mid \Psi_{t-1})\right] \\ &= \sum_{t=1}^{T} \ln\left[\mathbb{1}'(\pi_{t|t-1} \odot D_t)\right]. \end{aligned}$$

A nonlinear optimization method is required to maximize this log-likelihood function.

A.3 Moving block bootstrap

Bootstrapping techniques are commonly deployed to derive properties of a given estimator (e.g., the variance). I use it to construct potential future movements of the yield curve. Künsch (1989) proposes a technique - the moving block bootstrap (MBB) - that resamples blocks of historical data rather than single observations. As such, the MBB is particularly useful when ones wants to retain the typical evolution of the time series.

I apply the MBB of Künsch (1989) as follows. Recall the presence of T observed monthly yields, denoted by $\{y_1, \ldots, y_T\}$. Here, y_t is a $N \times 1$ vector where N denotes the number of maturities. Compute $T-1$ forward yield changes, $\Delta y_t = y_{t+1} - y_t$ and let the integer c denote the block length in months. In so doing, a total of $D = (T-1) - c + 1 = T - c$ overlapping blocks, $\mathcal{B}_1, \ldots, \mathcal{B}_D$ can be constructed:

$$\begin{aligned} \mathcal{B}_1 &= (\Delta y_1, \Delta y_2, \ldots, \Delta y_c) \\ \mathcal{B}_2 &= \quad (\Delta y_2, \Delta y_3, \ldots, \Delta y_{c+1}) \\ &\vdots \qquad\qquad\qquad\qquad \ddots \\ \mathcal{B}_{D-1} &= \qquad\qquad\qquad\qquad (\Delta y_{T-c-1}, \Delta y_{T-c}, \ldots, \Delta y_{T-2}) \\ \mathcal{B}_D &= \qquad\qquad\qquad\qquad\qquad (\Delta y_{T-c}, \Delta y_{T-c+1}, \ldots, \Delta y_{T-1}). \end{aligned}$$

Subsequently, draw an integer amount of B blocks at random from $\{\mathcal{B}_1, \ldots, \mathcal{B}_D\}$, with replacement. As such, all blocks in the collection have an equal probability of $\frac{1}{D}$ of being drawn. Let $\{\mathcal{B}_1^*, \ldots, \mathcal{B}_B^*\}$ denote the collection of randomly drawn blocks. Note that each block within this collection has a length of c months and that it represents a potential future evolution of the yield curve. The collection $\{\mathcal{B}_1^*, \ldots, \mathcal{B}_B^*\}$ of yield changes can now be used to construct potential future yields with the last observed yield of time T as starting point for each block.

For example, recall my dataset spans $T = 564$ yields, implying the construction of $T-1 = 563$ yield changes. Setting the block length to $c = 24$, the longest forecasting horizon, $D = 540$ blocks are thus constructed. From these 540 blocks I randomly sample $B = 500$ blocks, which are to be used in the simulation study.

B Tables

B.1 Factor statistics

Table B.1: **Factor statistics of the single-regime two-step models.**

	NS2				BC2		
	Summary statistics						
	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Mean	7.183	-1.542	-0.389	7.336	1.748	-2.848	-3.633
Std. dev.	2.310	1.846	1.993	2.316	6.397	4.257	6.263
Minimum	3.463	-5.616	-6.795	4.056	-15.638	-28.469	-32.027
Maximum	14.150	5.315	7.622	14.635	27.118	7.856	11.699
$\hat{\rho}_1$	0.986	0.953	0.856	0.982	0.806	0.780	0.759
$\hat{\rho}_{12}$	0.871	0.488	0.339	0.861	0.381	0.323	0.263
	Correlation estimated factors						
	β_1	β_2	β_3	β_1	β_2	β_3	β_4
β_1	1	-	-	1	-	-	-
β_2	-0.069	1	-	0.065	1	-	-
β_3	0.329	0.287	1	0.083	-0.824	1	-
β_4	-	-	-	-0.085	-0.960	0.890	1
	Correlation yield factors						
	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Level	0.983	0.081	0.459	0.986	0.062	0.172	-0.044
Slope	0.074	-0.984	-0.204	0.061	-0.375	0.007	0.103
Curvature	0.344	0.321	0.978	0.379	0.338	0.182	-0.277

Notes: The table presents factor statistics and correlations of the single-regime two-step models on the full sample (1962:1 - 2008:12). Decay parameter, λ fixed at 0.0609. $\hat{\rho}_j$ denotes autocorrelations with a time lag of j months. Refer to the text for the definitions of the level, slope and curvature. See Table 6.1 for an overview of which abbreviation corresponds to which model. Section 6 outlines the estimation procedure.

Table B.2: **Factor statistics of the single-regime one-step models.**

	Panel A: VAR(1) estimation results							**Panel B: AR(1) estimation results**						
	NS1-VAR			BC1-VAR				NS1-AR			BC1-AR			
	Summary statistics													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Mean	7.254	-1.496	-0.624	7.380	1.064	-2.485	-2.873	7.253	-1.495	-0.626	7.322	1.904	-3.064	-3.739
Std. dev.	2.300	1.904	1.820	2.341	5.259	3.387	5.241	2.301	1.906	1.823	2.348	3.139	2.381	2.547
Minimum	3.648	-5.779	-7.726	4.101	-13.295	-18.378	-25.138	3.671	-5.777	-7.682	3.689	-6.419	-14.228	-12.652
Maximum	14.266	5.661	5.318	14.506	24.336	6.590	16.952	14.249	5.670	5.308	14.469	12.774	3.410	2.453
$\hat{\rho}_1$	0.987	0.957	0.890	0.987	0.837	0.851	0.804	0.987	0.957	0.892	0.988	0.972	0.916	0.953
$\hat{\rho}_{12}$	0.875	0.513	0.320	0.873	0.332	0.240	0.213	0.875	0.513	0.321	0.876	0.547	0.303	0.339
	Correlation estimated factors													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
β_1	1	-	-	1	-	-	-	1	-	-	1	-	-	-
β_2	-0.047	1	-	0.156	1	-	-	-0.046	1	-	0.036	1	-	-
β_3	0.351	0.179	1	0.011	-0.758	1	-	0.345	0.181	1	0.209	-0.342	1	-
β_4	-	-	-	-0.190	-0.935	0.847	1	-	-	-	-0.105	-0.815	0.594	1
	Correlation yield factors													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Level	0.983	0.102	0.463	0.984	0.170	0.100	-0.152	0.983	0.104	0.458	0.988	0.111	0.293	-0.107
Slope	0.070	-0.968	-0.089	0.071	-0.326	-0.092	-0.0252	0.071	-0.969	-0.093	0.058	-0.614	-0.202	0.055
Curvature	0.348	0.369	0.916	0.384	0.461	0.138	-0.378	0.348	0.368	0.915	0.384	0.237	0.701	-0.110

Notes: The table presents factor statistics and correlations of the single-regime one-step models on the full sample (1962:1 - 2008:12). **Panel A** reports estimates of the VAR(1) models, **Panel B** of the AR(1) models. Decay parameter, λ estimated at 0.0561, 0.0578, 0.0562 and 0.0702 for NS1-VAR, BC1-VAR, NS1-AR and BC1-AR, respectively. Refer to Table B.1 for further details.

Table B.3: **Factor statistics of the regime-switching models.**

	Panel A: yields-only models							**Panel B: macro-finance models**						
	NS1-RS			BC1-RS				NS1-RS-X			BC1-RS-X			
	Summary statistics													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Mean	7.250	-1.489	-0.642	7.291	2.151	-3.218	-3.988	7.249	-1.489	-0.642	7.290	2.173	-3.232	-4.011
Std. dev.	2.309	1.902	1.817	2.358	0.637	2.439	1.684	2.309	1.902	1.817	2.357	0.628	2.448	1.685
Minimum	3.634	-5.771	-7.684	3.332	1.123	-11.362	-7.603	3.633	-5.771	-7.685	3.335	1.114	-11.372	-7.629
Maximum	14.284	5.672	5.256	14.427	3.138	4.117	2.414	14.285	5.672	5.265	14.423	3.241	4.097	2.278
$\hat{\rho}_1$	0.988	0.956	0.893	0.987	0.934	0.927	0.961	0.988	0.955	0.893	0.987	0.917	0.928	0.964
$\hat{\rho}_{12}$	0.875	0.511	0.322	0.869	0.445	0.477	0.491	0.875	0.511	0.322	0.869	0.445	0.473	0.497
	Correlation estimated factors													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
β_1	1	-	-	1	-	-	-	1	-	-	1	-	-	-
β_2	-0.050	1	-	-0.035	1	-	-	-0.050	1	-	-0.092	1	-	-
β_3	0.343	0.192	1	0.225	0.347	1	-	0.343	0.192	1	0.234	0.342	1	-
β_4	-	-	-	-0.110	0.315	0.646	1	-	-	-	-0.087	0.325	0.650	1
	Correlation yield factors													
	β_1	β_2	β_3	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_1	β_2	β_3	β_4
Level	0.983	0.098	0.455	0.986	0.060	0.365	0.022	0.983	0.098	0.455	0.986	0.003	0.374	0.045
Slope	0.070	-0.969	-0.110	0.073	-0.583	-0.669	-0.930	0.070	-0.969	-0.110	0.074	-0.590	-0.666	-0.930
Curvature	0.351	0.362	0.919	0.390	0.179	0.845	0.206	0.351	0.362	0.919	0.389	0.158	0.847	0.215

Notes: The table reports factor statistics and correlations of the regime-switching models on the full sample (1962:1 - 2008:12). **Panel A** shows estimates of the regime-switching yields-only models. **Panel B** presents estimates of the regime-switching macro-finance models. Decay parameter, λ estimated at 0.0571, 0.0762, 0.0571 and 0.0763 for NS1-RS, BC1-RS, NS1-RS-X and BC1-RS-X, respectively. Refer to Table B.1 for further details.

B.2 State equation estimation results

Table B.4: **Estimates of the state equation of the single-regime two-step models (NS2 and BC2).**

	Panel A: VAR(1) estimation results							**Panel B: AR(1) estimation results**						
	NS2			BC2				NS2			BC2			
	Autoregressive coefficient matrix, $\hat{F}$													
	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$
$\beta_{1,t}$	0.980**	0.027**	0.008	0.978**	0.039**	0.009	0.047**	0.980**	0	0	0.975**	0	0	0
	(0.007)	(0.009)	(0.009)	(0.008)	(0.010)	(0.010)	(0.014)	(0.007)			(0.008)			
$\beta_{2,t}$	-0.018	0.938**	0.042**	-0.009	1.070**	-0.039	0.303**	0	0.953**	0	0	0.807**	0	0
	(0.011)	(0.013)	(0.013)	(0.073)	(0.092)	(0.091)	(0.122)		(0.013)			(0.025)		
$\beta_{3,t}$	0.060**	0.039	0.829**	0.054	-0.061	0.889**	-0.134	0	0	0.861**	0	0	0.787**	0
	(0.020)	(0.025)	(0.024)	(0.051)	(0.065)	(0.063)	(0.085)			(0.022)			(0.026)	
$\beta_{4,t}$	-	-	-	-0.008	-0.152	0.084	0.562**	-	-	-	0	0	0	0.762**
				(0.079)	(0.100)	(0.098)	(0.132)							(0.027)
	Covariance matrix, $\hat{Q}$													
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$
$\beta_{1,t}$	0.139	-	-	0.174	-	-	-	0.142	-	-	0.187	-	-	-
$\beta_{2,t}$	-0.060	0.308	-	0.652	14.036	-	-	0	0.314	-	0	14.265	-	-
$\beta_{3,t}$	-0.051	0.022	1.023	-0.520	-9.037	6.897	-	0	0	1.042	0	0	6.983	-
$\beta_{4,t}$	-	-	-	-0.776	-15.023	9.902	16.437	-	-	-	0	0	0	16.513
	Constant vector, $\hat{\mu}$													
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$
$\tilde{\mu}$	0.196**	0.043	-0.446**	0.293**	0.940	-1.120**	-1.049*	0.147**	-0.078**	-0.060	0.188**	0.348**	-0.624**	-0.883**
	(0.055)	(0.082)	(0.149)	(0.064)	(0.577)	(0.404)	(0.624)	(0.052)	(0.031)	(0.044)	(0.060)	(0.165)	(0.134)	(0.198)

Notes: The table presents estimates of the state equation of the single-regime two-step models. **Panel A** reports estimates of a VAR(1) specification of the state equation, **Panel B** of an AR(1) specification. Standard errors, estimated using maximum likelihood, are in parenthesis. * (**) denotes significance at the 10%-level (5%-level). See Table 6.1 for an overview of which abbreviation corresponds to which model.

Table B.5: **Estimates of the state equation of the single-regime one-step models (NS1 and BC1).**

	Panel A: VAR(1) estimation results							**Panel B: AR(1) estimation results**						
	NS1			BC1				NS1			BC1			
	Autoregressive coefficient matrix, $\hat{F}$													
	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$
$\beta_{1,t}$	0.986**	0.027**	0.004	0.990**	0.028**	0.004	0.035**	0.980**	0	0	0.977**	0	0	0
	(0.006)	(0.008)	(0.009)	(0.006)	(0.008)	(0.008)	(0.010)	(0.003)			(0.003)			
$\beta_{2,t}$	-0.023**	0.952**	0.035**	0.045	0.960**	-0.070	0.174**	0	0.957**	0	0	0.970**	0	0
	(0.005)	(0.012)	(0.012)	(0.035)	(0.066)	(0.057)	(0.056)		(0.012)			(0.005)		
$\beta_{3,t}$	0.037**	0.008	0.880**	0.003	0.013	0.935**	-0.050	0	0	0.901**	0	0	0.935**	0
	(0.012)	(0.020)	(0.018)	(0.024)	(0.043)	(0.038)	(0.042)			(0.019)			(0.011)	
$\beta_{4,t}$	-	-	-	-0.078**	-0.018	0.123*	0.714**	-	-	-	0	0	0	0.945**
				(0.039)	(0.072)	(0.063)	(0.063)							(0.005)
	Covariance matrix, $\hat{Q}$													
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$
$\beta_{1,t}$	0.094	-	-	0.113	-	-	-	0.093	-	-	0.100	-	-	-
$\beta_{2,t}$	-0.034	0.301	-	0.348	9.146	-	-	0	0.300	-	0	0.322	-	-
$\beta_{3,t}$	-0.015	0.021	0.684	-0.244	-5.098	3.596	-	0	0	0.664	0	0	0.792	-
$\beta_{4,t}$	-	-	-	-0.419	-9.769	5.604	10.785	-	-	-	0	0	0	0.364
	Constant vector, $\hat{\mu}$													
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$
$\tilde{\mu}$	0.140**	0.113**	-0.331**	0.155**	0.040	-0.352*	0.069	0.143**	-0.068**	-0.068*	0.167**	0.064**	-0.210**	-0.212**
	(0.037)	(0.032)	(0.071)	(0.042)	(0.281)	(0.184)	(0.314)	(0.027)	(0.029)	(0.036)	(0.028)	(0.026)	(0.051)	(0.033)

Notes: The table presents estimates of the state equation of the single-regime one-step models. **Panel A** shows estimates of a VAR(1) specification of the state equation, **Panel B** of an AR(1) specification. Standard errors, estimated using the Hessian of the maximum likelihood function, are in parenthesis. * (**) denotes significance at the 10%-level (5%-level). See Table 6.1 for an overview of which abbreviation corresponds to which model.

Table B.6: **Estimates of the regime-switching yields-only models (NS1-RS and BC1-RS).**

Panel A: NS1-RS estimation results

	$\hat{F}$			$\tilde{\mu}_{S_t}$	
	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$S_t = 1$	$S_t = 2$
$\beta_{1,t}$	0.992**	0	0	0.047	-
	(0.004)			(0.032)	
$\beta_{2,t}$	0	0	0	-3.295**	-0.205**
				(0.086)	(0.037)
$\beta_{3,t}$	0	0	0.901**	-0.071**	-
			(0.017)	(0.029)	
$\beta_{4,t}$	-	-	-	-	-
	$\hat{Q}$			$\hat{P}$	
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	p_{11}	p_{22}
$\beta_{1,t}$	0.093	-	-	0.977**	0.981**
$\beta_{2,t}$	0	1.297	-	(0.009)	(0.007)
$\beta_{3,t}$	0	0	0.661	$\hat{\lambda}$	
$\beta_{4,t}$	-	-	-	0.0571**	
				(0.001)	

Panel B: BC1-RS estimation results

	$\hat{F}$				$\tilde{\mu}_{S_t}$	
	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$	$S_t = 1$	$S_t = 2$
$\beta_{1,t}$	0.993**	0	0	0	0.045	-
	(0.004)				(0.034)	
$\beta_{2,t}$	0	0	0	0	1.490**	2.792**
					(0.073)	(0.088)
$\beta_{3,t}$	0	0	0.923**	0	-0.254**	-
			(0.013)		(0.050)	
$\beta_{4,t}$	0	0	0	0.952**	-0.213**	-
				(0.007)	(0.030)	
	$\hat{Q}$				$\hat{P}$	
	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$	p_{11}	p_{22}
$\beta_{1,t}$	0.097	-	-	-	0.918**	0.969**
$\beta_{2,t}$	0	0.033	-	-	(0.013)	(0.010)
$\beta_{3,t}$	0	0	0.814	-	$\hat{\lambda}$	
$\beta_{4,t}$	0	0	0	0.273	0.0762**	
					(0.001)	

Notes: The table presents estimates of the regime-switching yields-only NS1-RS (**Panel A**) and BC1-RS (**Panel B**) models . Estimates for the constant of the slope factor ($\beta_{2,t}$) are regime-dependent. $\hat{P}$ corresponds to the transition matrix, where $\hat{p}_{ii} = \mathbb{P}[S_t = i | S_{t-1} = i]$, $S_t = i \in \{1, 2\}$. Standard errors, estimated using the Hessian of the maximum likelihood function, are in parenthesis. * (**) denotes significance at the 10%-level (5%-level). See Table 6.1 for an overview of which abbreviation corresponds to which model.

Table B.7: **Estimates of the regime-switching macro-finance models (NS1-RS-X and BC1-RS-X).**

Panel A: NS1-RS-X estimation results

	$\hat{F}$: $\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\tilde{\mu}_{S_t}$: $S_t = 1$	$S_t = 2$
$\beta_{1,t}$	0.992**	0	0	0.047	-
	(0.004)			(0.029)	
$\beta_{2,t}$	0	0	0	-3.307**	-0.215**
				(0.123)	(0.070)
$\beta_{3,t}$	0	0	0.901**	-0.071*	-
			(0.016)	(0.036)	
$\beta_{4,t}$	-	-	-	-	-

	$\hat{Q}$: $\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1,t}$	0.093	-	-
$\beta_{2,t}$	0	1.300	-
$\beta_{3,t}$	0	0	0.661
$\beta_{4,t}$	-	-	-

$\hat{P}_t$	$S_t = 1$	$S_t = 2$
$a^0_{S_t}$	5.303**	0.447**
	(0.465)	(0.113)
$a^1_{S_t}$	-0.099	0.271**
	(0.084)	(0.058)
$a^2_{S_t}$	-0.356**	1.055**
	(0.085)	(0.067)
$a^3_{S_t}$	0.016	-0.019
	(0.020)	(0.019)

$\hat{\lambda}$
0.0571**
(0.001)

Panel B: BC1-RS-X estimation results

	$\hat{F}$: $\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\beta_{4,t-1}$	$\tilde{\mu}_{S_t}$: $S_t = 1$	$S_t = 2$
$\beta_{1,t}$	0.992**	0	0	0	0.047	-
	(0.005)				(0.037)	
$\beta_{2,t}$	0	0	0	0	1.596**	2.870**
					(0.093)	(0.098)
$\beta_{3,t}$	0	0	0.940**	0	-0.206**	-
			(0.018)		(0.076)	
$\beta_{4,t}$	0	0	0	0.979**	-0.096**	-
				(0.009)	(0.049)	

	$\hat{Q}$: $\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\beta_{4,t}$
$\beta_{1,t}$	0.098	-	-	-
$\beta_{2,t}$	0	0.044	-	-
$\beta_{3,t}$	0	0	0.814	-
$\beta_{4,t}$	0	0	0	0.267

$\hat{P}_t$	$S_t = 1$	$S_t = 2$
$a^0_{S_t}$	2.798**	8.710**
	(0.411)	(1.193)
$a^1_{S_t}$	0.078	-0.462**
	(0.072)	(0.059)
$a^2_{S_t}$	-0.089	0.002
	(0.057)	(0.101)
$a^3_{S_t}$	0.051**	0.121**
	(0.014)	(0.024)

$\hat{\lambda}$
0.0763**
(0.001)

Notes: The table presents estimates of the regime-switching macro-finance NS1-RS-X (**Panel A**) and BC1-RS-X (**Panel B**) models. Estimates for the constant of the slope factor ($\beta_{2,t}$) are regime-dependent. $\hat{P}_t$ contains estimates of the logistic function that is used to compute the time-varying transition probabilities. In particular, $a^1_{S_t}, a^2_{S_t}$ and $a^3_{S_t}$ refer to CPI, GDP and housing starts for state S_t, respectively; $a^0_{S_t}$ is a constant. Standard errors, estimated using the Hessian of the maximum likelihood function, are in parenthesis. * (**) denotes significance at the 10%-level (5%-level). See Table 6.1 for an overview of which abbreviation corresponds to which model.

B.3 Log-likelihood values and likelihood ratio tests

Table B.8: **Log-likelihood values and likelihood ratio tests.**

Panel A: Log-likelihood

Model	Parameters	log-likelihood	Model	Parameters	log-likelihood
NS1-AR	27	14,266.45	NS1-RS	29	14,348.48
NS1-VAR	36	14,289.85	BC1-RS	32	14,931.43
BC1-AR	30	14,758.55	NS1-RS-X	35	14,358.28
BC1-VAR	48	15,213.51	BC1-RS-X	38	14,949.05

Panel B: Single-regime environment

Null	Alternative	Statistic	Distribution	p-value
NS1-AR	BC1-AR	984.20	$\chi^2(3)$	0
NS1-VAR	BC1-VAR	1847.32	$\chi^2(12)$	0
NS1-AR	NS1-VAR	46.80	$\chi^2(9)$	$4.28 \cdot 10^{-7}$
BC1-AR	BC1-VAR	909.92	$\chi^2(18)$	0
NS1-AR	BC1-VAR	1894.12	$\chi^2(21)$	0

Panel C: Testing for presence of regimes

Null	Alternative	Statistic	Distribution	p-value
NS1-AR	NS1-RS	164.06	$\chi^2(2)$	0
BC1-AR	BC1-RS	345.76	$\chi^2(2)$	0

Panel D: Two-regime environment

Null	Alternative	Statistic	Distribution	p-value
NS1-RS	BC1-RS	1165.90	$\chi^2(3)$	0
NS1-RS-X	BC1-RS-X	1181.54	$\chi^2(3)$	0
NS1-RS	NS1-RS-X	19.60	$\chi^2(6)$	0.003
BC1-RS	BC1-RS-X	35.24	$\chi^2(6)$	$3.87 \cdot 10^{-6}$
NS1-RS	BC1-RS-X	1201.14	$\chi^2(9)$	0

Notes: The table presents log-likelihood values (**Panel A**) and likelihood ratio test results (**Panels B - D**) of all one-step models. Entries in the column *Distribution* denote $\chi^2(r)$-probability distributions of the test statistics in the column *Statistic*; r denotes the degrees of freedom. Caution should be exercised in Panel C, where the distributions are approximate distributions, see the text for further details. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

C Figures

C.1 Time series of estimated factors

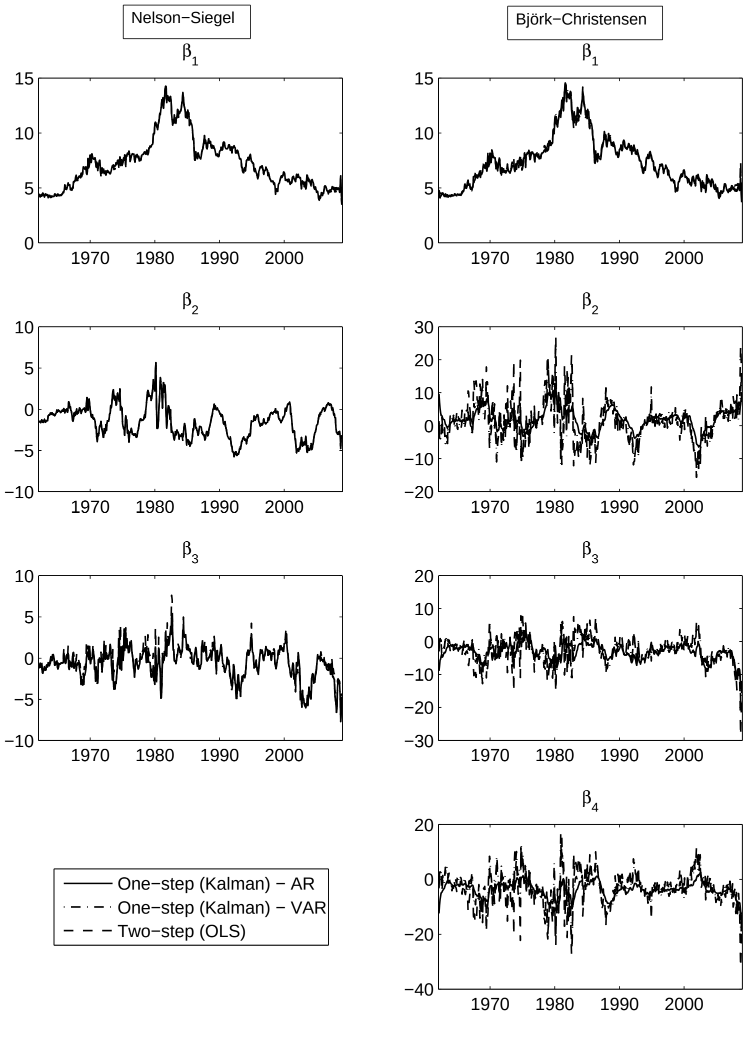

Figure C.1: **Time series of the factors of the single-regime models.**

Notes: The figure shows time series of the factors of the single-regime Nelson-Siegel and Björk-Christensen models. The time series cover the full sample period (1962:1 - 2008:12). Solid lines correspond to one-step (Kalman) AR(1) factors and dash-dot lines to VAR(1) factors. For those models, the decay parameter, λ is estimated alongside the factors and is assumed fixed for all t. Dashed lines correspond to two-step (OLS) estimated factors, where λ is fixed at 0.0609.

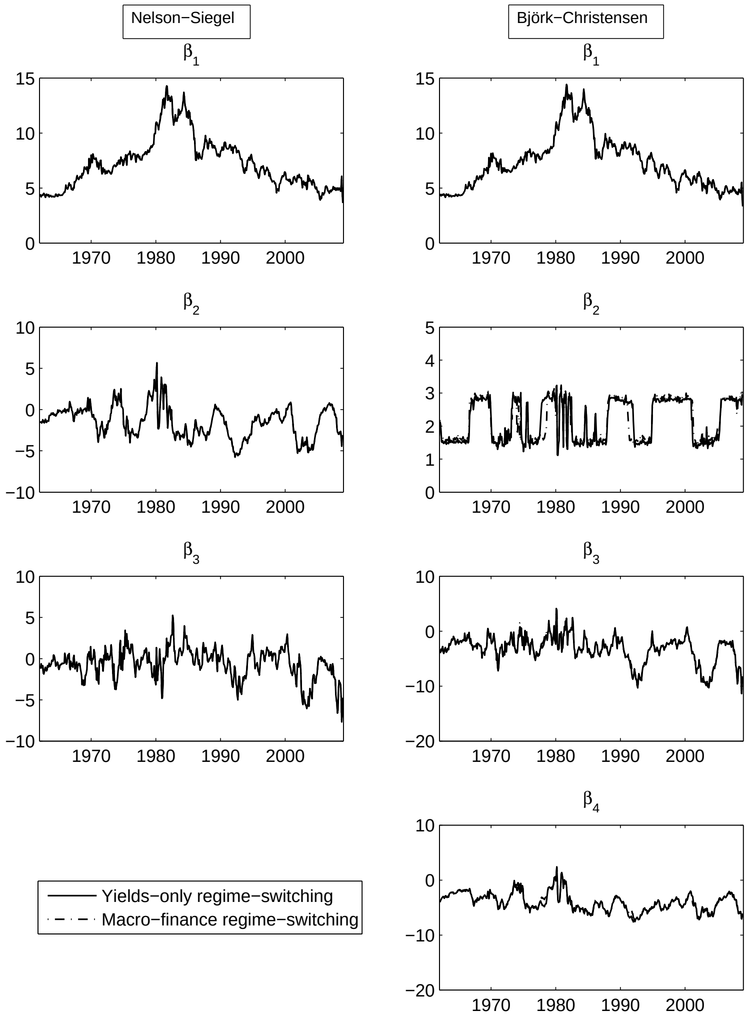

Figure C.2: **Time series of the factors of the regime-switching models.**

Notes: The figure shows time series of the factors of the regime-switching Nelson-Siegel and Björk-Christensen models. Regime-switching enters the models through the slope factor, $\beta_{2,t}$. The time series cover the full sample period (1962:1 - 2008:12). Solid lines correspond to regime-switching yields-only models (NS1-RS and BC1-RS). Dashed lines represent regime-switching macro-finance models (NS1-RS-X and BC1-RS-X). For all models, the decay parameter, λ is estimated alongside the factors and is assumed fixed for all t.

C.2 Fitted yields for selected months

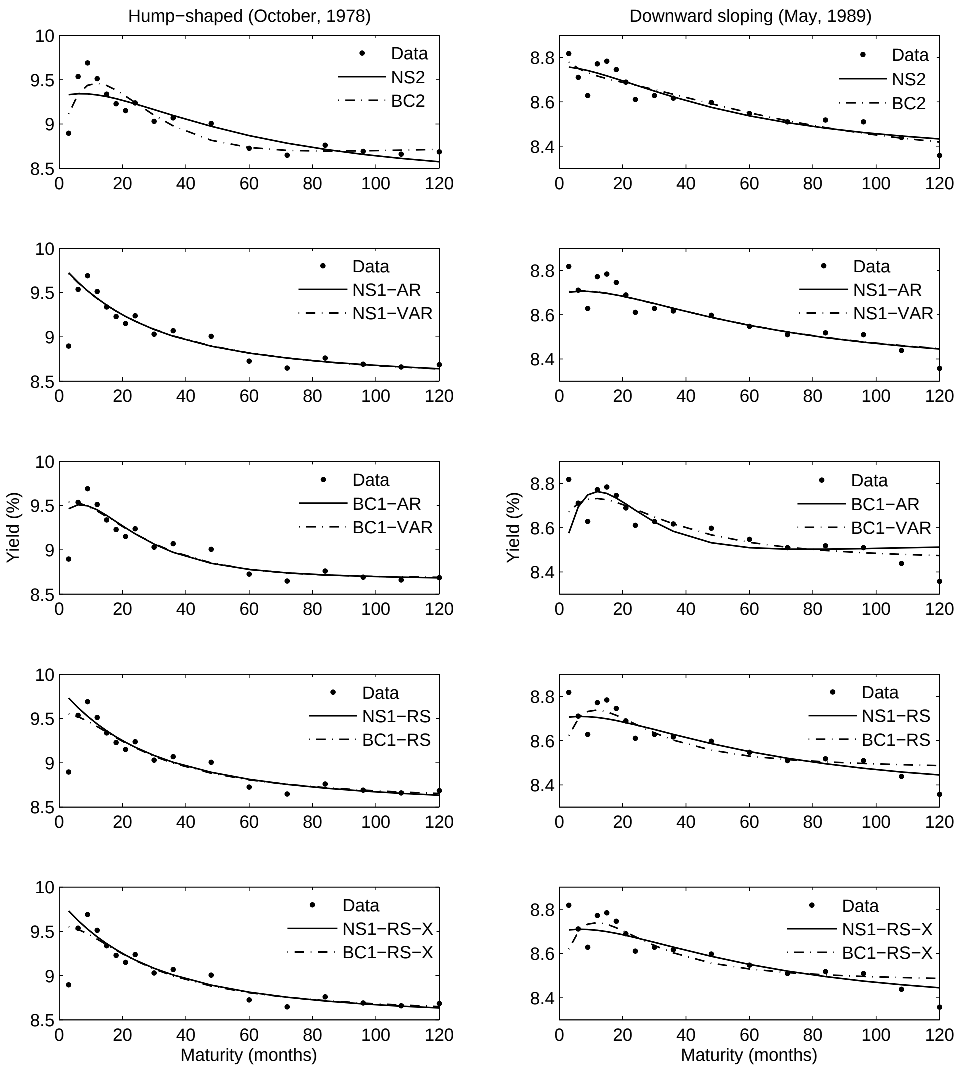

Figure C.3: **Fitted term structure for selected months (1).**

Notes: The figure shows the hump-shaped term structure of October, 1978 and the downward sloping curve of May, 1989 - observed yields are denoted by dots. Solid and dashed lines depict the term structure models' fit of these observed yields. Refer to Table 6.1 for an overview of which abbreviation corresponds to which model.

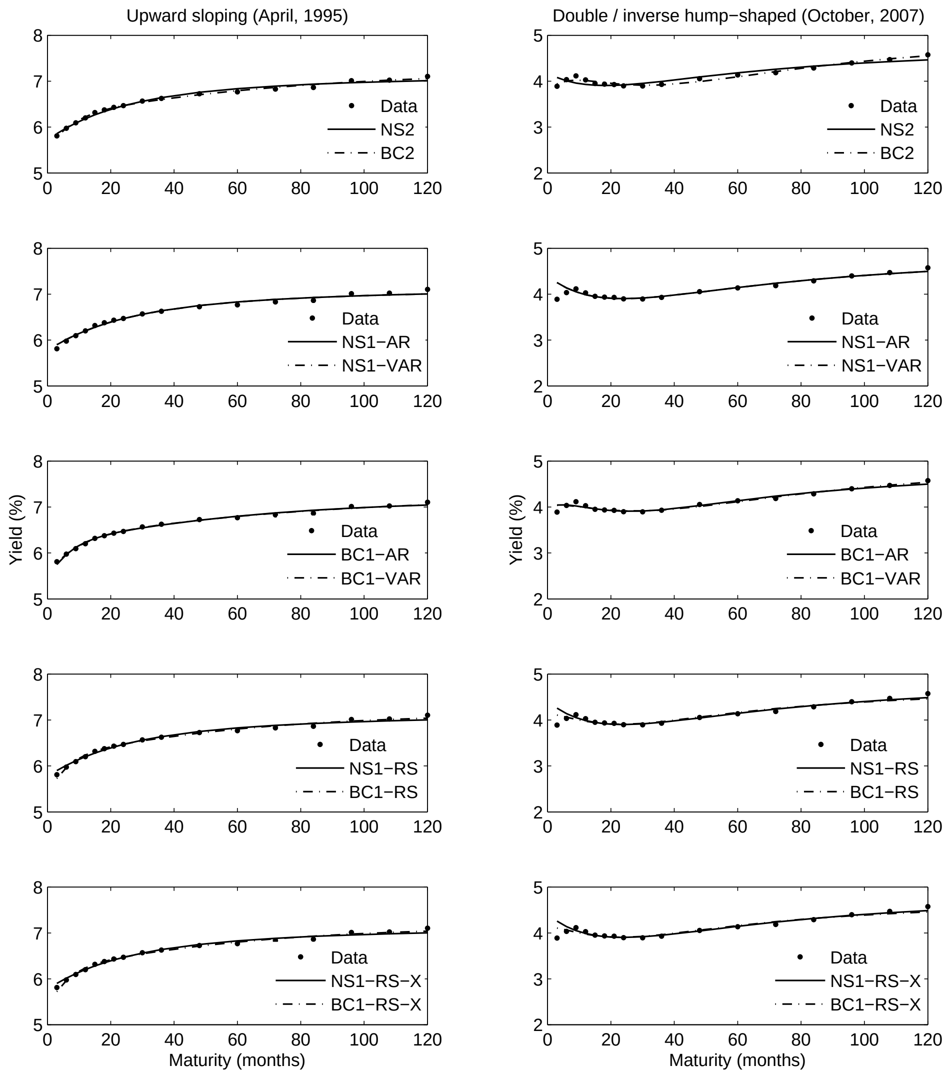

Figure C.4: **Fitted term structure for selected months (2).**

Notes: The figure shows the upward sloping term structure of April, 1995 and the double / inverse hump-shaped curve of October, 2007. See Figure C.3 for further details.