UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Bayesian inference for term structure models

Thomas Correa e Silva Martins

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

ICMC USP SÃO CARLOS

Thomas Correa e Silva Martins

Bayesian inference for term structure models

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Michel Helcias Montoril
Co-advisor: Prof. Dr. Marcio Alves Diniz

USP – São Carlos
July 2022

Thomas Correa e Silva Martins

Inferência bayesiana para modelos da estrutura a termo

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Michel Helcias Montoril
Coorientador: Prof. Dr. Marcio Alves Diniz

USP – São Carlos
Julho de 2022

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Thomas Correa e Silva Martins, realizada em 09/06/2022.

Comissão Julgadora:

Prof. Dr. Michel Helcias Montoril (UFSCar)

Prof. Dr. Luiz Koodi Hotta (UNICAMP)

Prof. Dr. Marcio Poletti Laurini (USP)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

ABSTRACT

MARTINS, T. **Bayesian inference for term structure models**. 2022. 95 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

We explore recent advances in Bayesian methods in order to estimate the Vasicek, CIR and dynamic Nelson-Siegel (DNS) models for term structure of interest rates. The models are specified as state space time series. The main goal of this work is assessing and comparing the forecasting abilities of each model with respect to the observed data via mean absolute error. When estimated with synthetic simulated datasets, the models are able to successfully recover the latent vectors. As for the forecasting abilities, the multifactor models generally deliver the best predictions. The relevance of this work lies in integrating novel computational techniques for Bayesian inference with canonical models from the field of financial economics. Several aspects of both fields are discussed throughout the text.

Keywords: Affine interest rate models, Dynamic Nelson-Siegel, Bayesian inference, State space time series, Asset pricing.

RESUMO

MARTINS, T. **Inferência bayesiana para modelos da estrutura a termo**. 2022. 95 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Exploramos avanços recentes em métodos bayesianos para estimar os modelos de Vasicek, CIR e Nelson-Siegel dinâmico para a estrutura a termo da taxa de juros. Os modelos são especificados na forma de séries temporais de espaço de estados. O objetivo principal deste trabalho é analisar e comparar as habilidades de previsão de cada modelo em relação aos dados observados, por meio do desvio médio absoluto. Quando estimados com conjuntos de dados simulados sintéticos, os modelos conseguem recuperar os vetores latentes. Com relação às habilidades preditivas, os modelos multifatores geralmente realizam as melhores previsões. A relevância deste trabalho está em integrar novas técnicas computacionais para inferência bayesiana com modelos canônicos da área de economia financeira. Diversos aspectos de ambos os campos são discutidos ao longo do texto.

Palavras-chave: Modelos afins de taxas de juros, Nelson-Siegel dinâmico, Inferência bayesiana, Modelos espaço de estados, Precificação de ativos.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER

1

INTRODUCTION

The term structure of interest rates consists of a panel of yields for bonds with similar characteristics, and each of the bonds' expiration rates, or maturities. The graphical representation of the spot rates with each one of the maturities, at a given instant, is known as the yield curve. It is also common to refer to the term structure as "the yield curve". The term structure is a key object of study in economics, as it is formed through the transactions of bonds in the market and represents the market agents' expectations of the future.

As the longer and shorter maturity yields contain information about each other, this needs to be taken into account for statistical modelling of the term structure. Some of the most common techniques include specifying the panel of observed yields as a function of latent variables, known as factors. Models such as the ones by Vasicek (1977), Cox, Ingersoll and Ross (1985) or Diebold and Li (2006) can be written in state space form, and estimation can be carried out with appropriate statistical tools.

Bayesian estimation of state space models have become commonplace in the last years, especially with the development of computational tools such as Markov chain Monte Carlo. Practical advantages of these techniques, such as increased flexibility in specification of larger models, are becoming evident. Theoretical considerations are also relevant as a justification for using Bayesian methods. Also, some of the novel procedures in Bayesian modelling and computation, such as Hamiltonian Monte Carlo algorithms, can be useful in estimating financial time series models.

Our main goal in this work is using the Bayesian probabilistic framework to obtain posterior distributions of spot rate forecasts from term structure models. First, we estimate the models with simulated artificial data, in order to assess computational issues and parameter estimation, an approach justified by Gelman *et al.* (2020). Then, we fit the model with real data and show how to obtain the posterior distribution of forecasts, as well as computing a measure of error based on the absolute loss function.

Chapter 2 presents a literature review of yield curve modelling and some of the most common statistical models, as well as theoretical and practical aspects of Bayesian inference. Chapter 3 describes our methodology and statistical models, including the estimation techniques. Chapter 4 presents the results for estimation with simulated data, and observations about the estimated and true state variable vectors. Chapter 5 presents estimation results for real data, together with the model comparison via mean absolute error. Chapter 6 concludes with a brief discussion of the results, and some possible further directions for research.

CHAPTER

2

LITERATURE REVIEW

In this chapter we present a literature review of term structure modelling, as well as the definition of some terms used in financial economics. We then proceed in listing some of the most common term structure models. We cite some of the relations between the term structure and macroeconomic variables, justifying the relevance of term structure modelling to monetary and fiscal policy. In Section 2.2 we list both theoretical and practical aspects of Bayesian inference for statistical time series models.

2.1 Yield curve and asset pricing

2.1.1 The yield curve: spot and forward rates

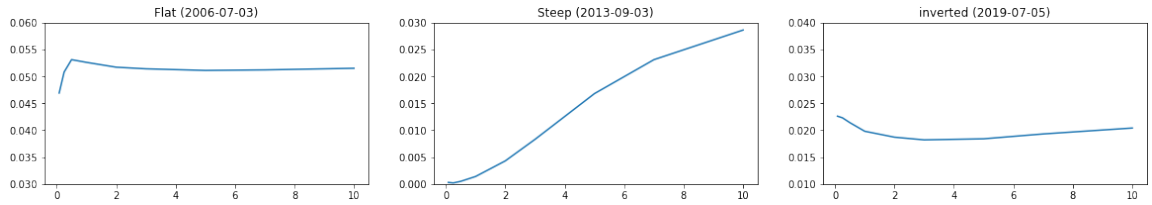
Modelling of the term structure of interest rates is of great relevance in areas such as economics and finance. Unlike the modelling of one stock price, where there is only one observed variable, in the case of term structure modelling the data consists of a panel of observed interest rates, each one with a different time to maturity.[1] The yield curve is the graphical representation of each rate with its respective maturity. A statistical model for the term structure needs to take into account the joint dynamics of all observed rates. The yield curve is commonly seen in three different shapes: flat, steep and inverted, as presented in Figure 1. A statistical model for the yield curve ideally should be able to replicate those patterns.

The yield curve can represent the yield to maturity (YTM) of coupon bonds, however it is also very common to work with the zero-coupon curve.[2] For the US case, the zero-coupon curve can be obtained through a process called stripping (FABOZZI, 2007, p. 81).

The yields on zero-coupon bonds are called spot rates, and the yield curve for the spot rates is the term structure of interest rates (FABOZZI, 2007, p. 82). As the spot rates are not

[1] The maturity is the date when the loaned amount is due to be paid back.
[2] The YTM is the rate sets a bond's cash flow at present value equal to its market price plus interest

Figure 1 – Three different yield curve shapes for US Treasury yield data



Note – The vertical axis contains the yields and the horizontal axis denotes time to maturity in years.

Source: Elaborated by the author.

always observable, statistical procedures such as bootstrapping can be used in order to construct the term structure from the observed curves (FABOZZI, 2007, p. 135).

The final type of yield curve is the forward curve, derived from the forward rates. A forward rate is a rate that can be "locked in" at a present instant for a transaction that will take place between two future instants (JAMES; WEBBER, 2000, p. 40). Forward rates are also not observable, so implied forward rates are obtained from the spot rates (FABOZZI, 2007, p. 148).

It is also possible to define spot and forward rates mathematically. A spot interest rate $R_S(t,T)$ is a rate observed at time t for a period that extends from t until T, $t < T$. A forward rate $R_F(t,T_1,T_2)$ is an interest rate locked in at instant t for a transaction that will supposedly take place between T_1 and T_2, with $t < T_1 < T_2$.

As it was mentioned before, yield curves are usually calculated for zero-coupon bonds, i.e., bonds that pay interest only at the maturity date.[3] Although the spot yield curve specifies only rates observed at the present instant t, implied forward rates are obtained by the following relation:

$$(1+R_S(t,T_2))^{T_2} = (1+R_S(t,T_1))^{T_1}(1+R_F(t,T_1,T_2))^{T_2-T_1}. \tag{2.1}$$

The assumption that current implied forward rates are unbiased estimates of future spot rates is known as the expectations hypothesis. If the expectations hypothesis is true, the forward rate in equation (2.1) will be the average of the spot rate between T_1 and T_2. In other words, holding either a long bond or a sequence of short bonds over the same time period should yield the same as the term premium (the excess return on long over short bonds) is constant over time (CAMPBELL, 2017).

Fama and Bliss (1987) test the empirical validity of the expectations hypothesis and conclude that the term premiums are not constant over time but related to the business cycle. They also conclude that current forward rates perform badly at forecasting short-term changes in interest rate. However, when the forecasting window is increased (over 1 year), the forward

[3] Also worth of notice is that the yield curve is commonly calculated for bonds assumed as free of credit and liquidity risks, usually government (sovereign debt) bonds.

rates are better at predicting changes, a fact the authors attribute to a mean-reverting tendency in interest rates.

Additional empirical tests of the expectations hypothesis include Froot (1989), which concludes that the hypothesis holds in the long-term but not in the short-term due to the time-varying term premium. Campbell and Shiller (1991) also arrive to a similar conclusion regarding the increased predictability of forward rates in time horizons larger than a year. Cochrane and Piazzesi (2005) forecast excess return on long bonds by means of a novel predictor, a linear combination of forward rates which is also connected to the business cycle, and attain high predictability. If the expectations hypothesis is perfectly valid, there is no need to model the yield curve, as all future (forward) rates can be inferred from the current spot rates.

2.1.2 Asset pricing: risk neutrality, complete markets and arbitrage

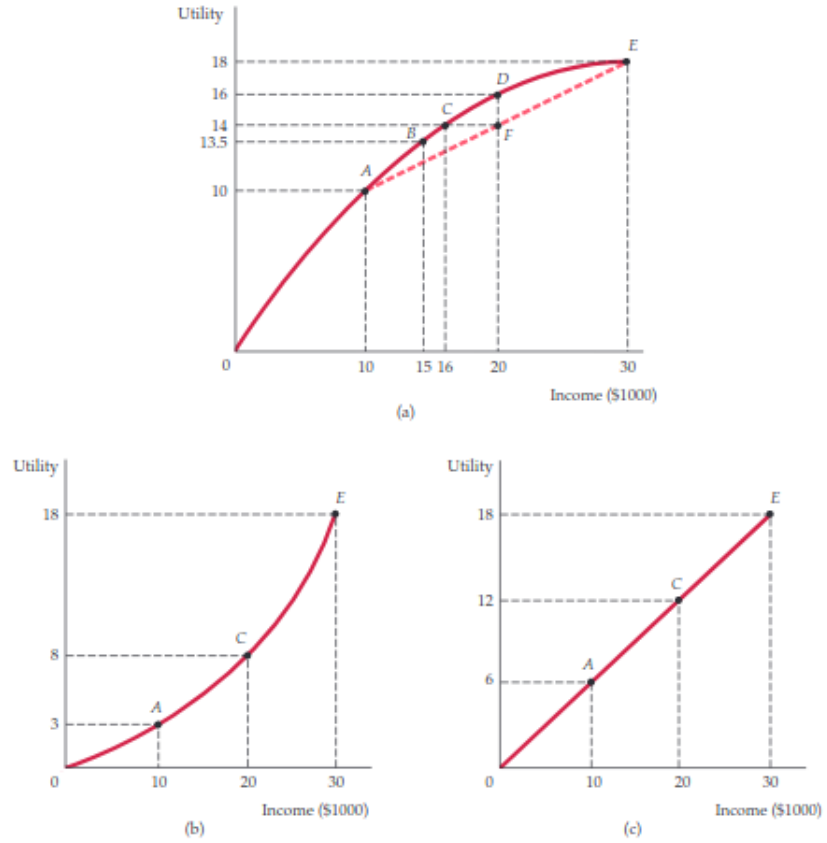
The main goal of asset pricing theory in financial economics is to study agents' behavior in trading financial assets on the market. As the future payoff of these assets involves uncertainty, an expected utility framework must be developed in order to study the behavior of agents in face of uncertainty and risk.

Expected utility theory is the benchmark framework in economics to study how agents make decisions when faced with random payoffs. The expected utility of the random payoff X is given by $E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx$, where $u(x)$ is the utility assigned to outcome x and $f(x)$ is the probability density function of X. In this context the random variable X can also be referred to as a lottery. The Von Neumann-Morgenstern utility theorem states that the agent will prefer lottery X to lottery Y if and only if $E[u(X)] > E[u(Y)]$. Each agents' preferences will define the functional forms of $u(x)$, and a very commonly used form is $u(x) = \log(x)$ because it can model risk aversion. Figure 2 represents three patterns for the utility of income functions: risk averse, risk seeking and risk neutral, respectively. As it can be seen, the marginal utility of income is diminishing for the risk averse, constant for the risk neutral and increasing for the risk seeking case.

Defining the set of every possible future states of the economy, contingent claims are securities that pay 1 if a certain state occurs and 0 for any other state. A market is said to be complete if every contingent claim can be replicated, making it possible for investors to "bet" on the occurrence of any of the future states (COCHRANE, 2009, p. 54).

The law of one price states that assets traded on the market with identical payoffs must have the same price (COCHRANE, 2009, p. 66). If this were not true, it would be possible to have a riskless profit by buying the same asset at the lower price and selling it at the higher price. Although it is common in non-technical language to denote "arbitrage" as violations of the law of one price, the word "arbitrage" in finance theory has a wider meaning (COCHRANE, 2009, p. 70).

Figure 2 – Utility of income functions for: (a) risk averse; (b) risk seeking; (c) risk neutral agents



Source: Pindyck and Rubinfeld (2017).

Arbitrage is defined as the opportunity to invest in assets with payoffs that are non-negative for all states and positive for at least one state with non-null probability, that is, obtain a return greater than the risk-free rate without taking on additional risk.[4] From there, the absence of arbitrage can be stated as, if one payoff dominates another, then its price is necessarily higher (COCHRANE, 2009, p. 70), a definition that might be helpful in drawing analogies with statistical decision theory (DEGROOT; SCHERVISH, 2012, p. 458).

Pricing an asset consists of discounting its future payoffs to present value, and the stochastic discount factor (SDF) is the random variable that connects both (COCHRANE, 2009, p. 16). The SDF is a function of investors' marginal utilities, which by themselves are functions of future consumption, a random variable. Therefore, the SDF is also a random variable, and asset pricing would require, at first, knowledge of the investors' marginal utility functions. As a matter of fact there is a more simple way of doing this calculation, known as risk-neutral pricing, which consists in obtaining a set of risk-neutral probabilities. The price of an asset in the risk-neutral world shall be equal to its expected future payoff discounted by the risk-free rate

[4] The payoff of an asset at instant T corresponds to its price at T plus any cash flow, e.g., dividends paid from the present instant t up to T.

(HULL, 2009, p. 300).[5]

The risk-neutral probabilities can only be obtained in a complete market with no arbitrage. A complete market implies a stochastic discount factor exists, and, also in the absence of arbitrage, that the SDF is positive and unique.

2.1.3 Affine term structure models

There are many possible approaches for statistical modelling of the term structure, and the affine models are among some of the most common models in the asset pricing literature. The term "affine model" can have many different meanings, therefore we shall stick to the one by Piazzesi (2010), which defines affine models as any arbitrage-free model where the observed bond yields are affine functions of a state vector. In mathematical notation, this means the logarithm of the bond price $P_t(\tau)$, with time to maturity τ, is

$$\ln P_t(\tau) = A(\tau) + B(\tau)' x_t, \tag{2.2}$$

which means $\ln P_t(\tau)$ is an affine function of a state vector x_t.

The price of a bond at maturity $\tau = 0$ is, by definition, equal to 1, so the price of a bond is given by

$$P_t = \mathrm{E}^* \left[\exp\left\{ -\int_t^{t+\tau} r_s ds \right\} \right], \tag{2.3}$$

and P_t is related to the spot yield $Y_t(\tau)$ of a bond through

$$Y_t(\tau) = -\frac{\ln P_t(\tau)}{\tau} = -\frac{A(\tau) + B(\tau)' x_t}{\tau},$$

where r_t is the short rate, that is, the rate as $\tau \to 0$, and E^* denotes that the expectation is taken under risk-neutral probabilities. The short rate shall be a function of the state vector, and, in affine term structure models, the function is affine by assumption (PIAZZESI, 2010). More precisely, the short rate is defined by

$$r_t = \delta_0 + \delta_1' x_t,$$

and, in particular cases where the short rate itself is the only factor, we have $\delta_0 = 0$ and $\delta_1 = 1$. The right-hand side of (2.3) is assumed to be of the form $\exp\{A(\tau) + B(\tau)' x_t\}$.

The second assumption behind affine term structure models is that the state vector x_t is an affine diffusion process under risk-neutral probabilities (PIAZZESI, 2010). Mathematically this can be written as

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \tag{2.4}$$

where $\mu_x(x_t)$ and $\sigma_x(x_t)$ are called, respectively, the drift and diffusion coefficients, and W_t is a Wiener process. A brief overview of Wiener processes is given in Appendix A. It is required, then, that both the drift and diffusion coefficients are affine.

[5] The future payoff expectation is calculated using the risk-neutral probabilities.

Duffie and Kan (1996) give a general definition of affine term structure models, assuming equation (2.4) is of the form

$$dx_t = (ax_t + b)dt + \Sigma S dW_t,$$

where the matrix S is a diagonal $n \times n$ matrix with ith diagonal element $v_i(x_t) = \sqrt{\alpha_i + \beta_i x_t}$, a and Σ are $n \times n$ matrices, b and β_i, for each i, are vectors of size n and α_i is, for each i, a scalar.

The coefficients $A(\tau)$ and $B(\tau)$ in equation (2.2) are obtained from a set of Riccati equations, and may or may not have closed forms, depending on the assumed process for the state vector (PIAZZESI, 2010). Following Lindström, Madsen and Nielsen (2015, p. 234), if we take the drift and diffusion coefficients with affine form, as in

$$\mu(x_t) = ax_t + b,$$
$$\sigma(x_t) = \sqrt{cx_t + d},$$

then the coefficients $A(\tau)$ and $B(\tau)$ shall solve the pair of ODEs

$$B'(\tau) = -aB(\tau) + \frac{1}{2}cB^2(\tau) - 1,$$
$$A'(\tau) = bB(\tau) - \frac{1}{2}dB^2(\tau).$$

Having defined affine term structure models, we now look into the question of choosing a functional form for the state vector. Rendleman and Bartter (1980) introduce a model where the short rate is the single factor and follows the process

$$dr_t = \mu r_t dt + \sigma r_t dW_t.$$

This kind of process does not show itself as an adequate model for the short rate. Since r_t is endogenously related to other macroeconomic variables, it presents the empirical property of mean-reversion. This means the short interest rate is unable to grow indefinitely, as opposed to, e.g., a stock price, and ends up reverting towards a long-term mean. In mathematical terms this means that when r_t is above its long-run mean its drift tends to be negative. Likewise, if r_t is below the long-term mean the drift tends to be positive (HULL, 2009, p. 730).

Vasicek (1977) looks into the case where the short rate is the only factor and follows the process

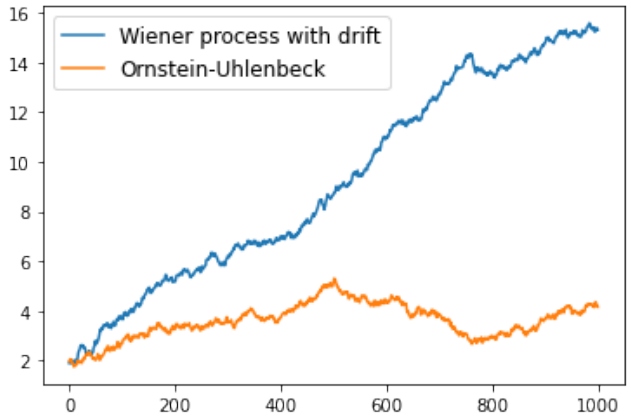
$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t, \tag{2.5}$$

where the parameters have the following interpretations: μ is the long-term mean to which the short rate reverses, κ is the speed of mean reversion and σ is the standard deviation of the logarithmic returns of r_t, i.e., the volatility. Figure 3 illustrates the difference between the short rate process in the Rendleman-Bartter and the Vasicek models.

For the issue of finding the functional forms for the coefficients $A(\tau)$ and $B(\tau)$ of the yield curve equation, the Vasicek model has a closed form solution. Recall that

$$Y_t(\tau) = -\frac{A(\tau) + B(\tau)r_t}{\tau}.$$

Figure 3 – Simulation of Rendleman-Bartter (Wiener process with drift) and Vasicek (Ornstein-Uhlenbeck) processes for the short rate



Note – Parameter values are: $\mu = 4$, $\sigma = 1$ and $\kappa = 1$ (for Vasicek).

Source: Elaborated by the author.

The expressions of $A(\tau)$ and $B(\tau)$ for the Vasicek model, obtained from Lindström, Madsen and Nielsen (2015, p. 236), are

$$B(\kappa,\tau) = \frac{1-\exp\{-\kappa\tau\}}{\kappa},$$
$$A(\kappa,\mu,\sigma,\tau) = \left(\frac{\sigma^2}{2\kappa^2}-\mu\right)\tau + \frac{1-\exp\{-\kappa\tau\}}{\kappa}\left(\mu-\frac{\sigma^2}{\kappa^2}\right) + \frac{\sigma^2}{4\kappa^3}(1-\exp\{-2\kappa\tau\}),$$

where κ, μ and σ are the parameters in equation (2.5).

One of the disadvantages of the Vasicek model is that it allows interest rates to assume a negative value, which can be undesirable, and also assumes volatility to be constant. Cox, Ingersoll and Ross (1985), henceforth CIR, with a slight modification to the Vasicek model, proposed the short rate process

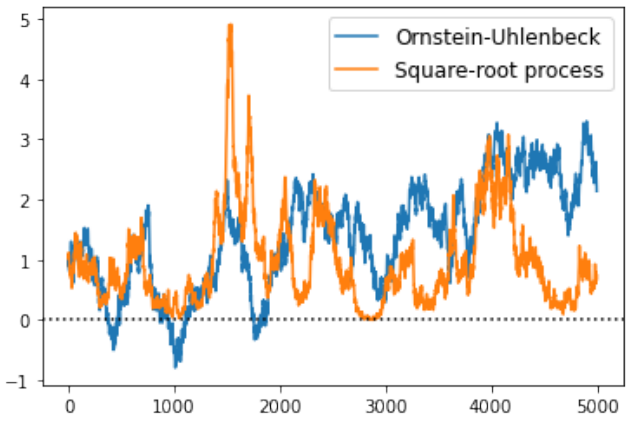
$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t.$$

By multiplying the diffusion coefficient by $\sqrt{r_t}$, interest rates cannot assume negative values while adding conditional heteroskedasticity structure, so long as the condition $2\kappa\mu > \sigma^2$ holds. Figure 4 illustrates the difference between the short rate processes of the Vasicek and CIR models. Following Lindström, Madsen and Nielsen (2015, p. 239), the coefficients $A(\tau)$ and

$B(\tau)$ have the form

$$
\begin{aligned}
Y_t(\tau) &= -\frac{A(\tau)+B(\tau)r_t}{\tau}, \\
\gamma &= \sqrt{\kappa^2+2\sigma^2}, \\
g(\tau) &= 2\gamma-(\kappa-\gamma)(\exp\{-\gamma\tau\}-1), \\
B(\tau) &= 2(\exp\{-\gamma\tau\}-1)/g(\tau), \\
A(\tau) &= \frac{2\kappa\mu}{\sigma^2}\ln\left[\frac{2\gamma\exp\{(\kappa-\gamma)\tau/2\}}{g(\tau)}\right].
\end{aligned}
$$

Figure 4 – Simulation of Vasicek (Ornstein-Uhlenbeck) and CIR (square-root) short rate processes



Note – Parameter values are: $\mu = 1$, $\kappa = 0.4$, $\sigma = 1$.

Source: Elaborated by the author.

2.1.4 Alternative approaches: the Nelson-Siegel model and dynamics

The empirical findings show that single-factor models often may not perform very well. Litterman and Scheinkman (1991) find that three-factor models generally suffice to explain practically all of the variation in the term structure.

A factor structure is present in situations where a high-dimensional object, e.g., the many interest rates across different maturities, can be well-explained by an underlying low-dimensional structure called factors (DIEBOLD; RUDEBUSCH, 2013). In the case of the Vasicek (1977) and Cox, Ingersoll and Ross (1985) models, there is only one factor: the short rate. Common factors cited in literature for the term structure of interest rates correspond to the level, slope and curvature of the curve (LITTERMAN; SCHEINKMAN, 1991). Factors can also be found by performing principal-component analysis, though this process might lead to spurious relationships when applied to non-stationary data (CRUMP; GOSPODINOV, 2019). As it was the case with Hull and White (1990), Balduzzi *et al.* (1996) and Duffie and Kan (1996), it is possible to specify affine models with more than one factor.

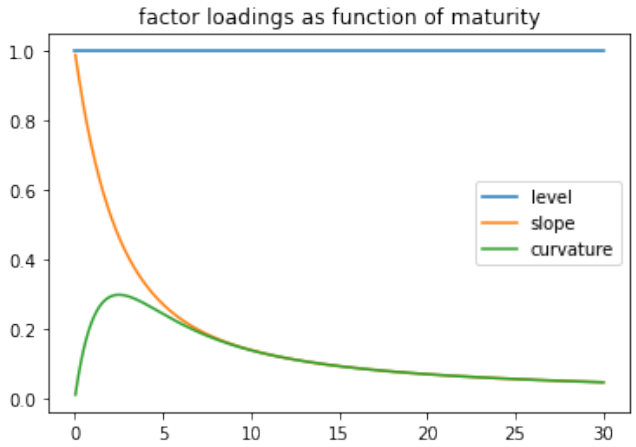
Nelson and Siegel (1987) propose another kind of model where a non-parametric function obtained through polynomial expansion is fitted to the yield curve at a single point in time. This model provides a good fit of the yield curve, but does not allow for time-varying factors. Diebold and Li (2006) turn the original Nelson-Siegel into the dynamic model

$$Y_t(\tau) = l_t + s_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + c_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right),$$

which has been named the dynamic Nelson-Siegel (DNS) model.

While in the original Nelson-Siegel the interest rate $y_t(\tau)$ is not a time series but actually a cross-section of observations of y at a fixed time period, the DNS writes the panel of interest rates $y_t(\tau)$ as a function of the latent factors l_t, s_t and c_t. The decay parameter λ can be treated as a constant to be calibrated from the data. This calibration procedure may be done by choosing an intermediate maturity, e.g., $\tau = 2.5$ years and maximizing the curvature loading $\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$ in λ. The resulting factor loadings as a function of maturity for the curvature loading attaining its maximum at $\tau = 2.5$ are presented in Figure 5.

Figure 5 – Factor loadings of the DNS model as a function of maturity



Note – Maturity is on the horizontal axis and λ is calibrated to the value of 0.7173.

Source: Elaborated by the author.

Christensen, Diebold and Rudebusch (2011) convert the dynamic Nelson-Siegel model into an arbitrage-free model building on the affine framework presented by Duffie and Kan (1996). The procedure consists in finding factor loadings which resemble the most those of the Duffie-Kan model. In the arbitrage-free Nelson-Siegel (AFNS) model the spot yields can be specified as

$$Y_t(\tau) = f_t^1 + \frac{1-e^{-\lambda\tau}}{\lambda\tau} f_t^2 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) f_t^3 - \frac{C(\tau)}{\tau},$$

where the three latent factors f^1, f^2 and f^3 have their dynamics under risk-neutral probabilities and $\frac{C(\tau)}{\tau}$ is an adjustment factor. Imposition of no-arbitrage can or cannot be useful throughout

different empirical applications (DIEBOLD; RUDEBUSCH, 2013). No-arbitrage restrictions can improve predictive performance when models are misspecified, but may not be relevant otherwise (JOSLIN; SINGLETON; ZHU, 2011). Imposing no-arbitrage restrictions, be it cross-sectional or dynamic, may not improve forecasting, according to Duffee (2011).

2.1.5 Role of macroeconomic variables in forecasting the yield curve

While macroeconomic theory seeks to understand the relation between observed yields and macroeconomic variables such as inflation, GDP or money supply, the main goal in finance is predictive modelling, often in no-arbitrage settings (JAMES; WEBBER, 2000, p. 271). However, there is much to be gained by combining both disciplines, as the short rate is an important variable for both finance and macroeconomics. This could result in better estimation of the short rate from a financial perspective by taking into account the central bank's response to inflation or output. Conversely, the understanding of future expectations of the economy might be improved by considering the relation between spot yields of longer and shorter maturities (DIEBOLD; RUDEBUSCH, 2013).

The relevance of macroeconomic variables in forecasting asset prices can be extended to other areas of finance other than yield curve modelling. Hahn and Lee (2006) find out the default and term spreads, which are known to be correlated with the business cycle, account for the variation in stock returns in the same way that the size (SMB) and book-to-value (HML) factors do.

Hybrid macro-finance term structure models can range from combining a generalized Vasicek model with the traditional IS-LM macroeconomic model (JAMES; WEBBER, 2000, p. 285) to augmenting the DNS or AFNS models with observed macroeconomic variables as additional factors (DIEBOLD; RUDEBUSCH, 2013). Bauer and Rudebusch (2017) find that macroeconomic variables related to monetary policy are useful in forecasting a portion of unspanned risk, i.e., uncorrelated with the other yield factors. There is a great potential in combining financial and macroeconomic models, not only for our goal of forecasting the term structure, but also for understanding of relationships between different macroeconomic variables.

2.1.6 The yield curve in macroeconomics: fiscal and monetary policy

The term structure is of great relevance to macroeconomics since the spread between long and short rates can provide insight regarding agents' future expectations of the economy and are helpful in forecasting macroeconomic variables e.g., economic activity or inflation. It can also be part of dynamic stochastic general equilibrium (DSGE) models.

Ang and Piazzesi (2003) estimate a vector autoregression model for the term structure combining three latent correlated AR(1) factors plus two observable macroeconomic variables, inflation and real activity, in a Taylor rule context for monetary policy. In the Taylor rule setting

for monetary policy, the policy interest rate is set as a weighted combination of the inflation deviation from its target and the output gap. Rudebusch and Wu (2003) also combine latent factors with macroeconomic observables in a Taylor rule context to describe the term structure, but impose no-arbitrage and write the two latent factors L_t and S_t as functions of the macroeconomic variables. The level factor L_t is as a weighted average of the nominal equilibrium rate i_t^* and its lagged value L_{t-1}, and the slope factor S_t is a weighted average of the central bank response through the Taylor rule $(g_\pi(\pi_t - L_t) + g_y y_t)$ and its lagged value S_{t-1}.[6]

As for fiscal policy, Dai and Philippon (2005) also work in a vector autoregression approach to model the yield curve, choosing as state space variables government deficits, inflation, real activity, federal funds rate and a latent factor. The authors also study the impact of deficit, spending and tax revenue shocks by impulse response functions.

Binsbergen *et al.* (2012) employ term structure data in a DSGE model with the goal of modelling the impact of rates on technology and preference parameters. The utility functions are of the Epstein-Zin form, and the authors estimate high values for both risk aversion and intertemporal elasticity of substitution parameters.[7]

2.2 Bayesian inference: from theory to practice

2.2.1 Theoretical aspects of Bayesian inference in predictive modelling

Some of the common reasons for why economists have traditionally shied away from Bayesian inference can be identified. Among those are the misconceptions that Bayesian inference is "subjective" (versus "objective" frequentist inference) or that Bayesian methods are "hard" (SIMS, 2010). This might not be the case for current-day macroeconomics, where Bayesian methods have become the standard (FERNÁNDEZ-VILLAVERDE, 2010), and the same can be said about empirical finance (JACQUIER; POLSON, 2012). However, the main point of concern for the use of Bayesian inference in economic and financial models seems to be that many economists think of Bayesian inference as just a set of practical tools, and not as a different interpretation of statistical inference (SIMS, 2010).

One of the greatest theoretical motivations for the use of Bayesian inference is the possibility of combining prior domain knowledge, in an explicit fashion, with the observed data (JACQUIER; POLSON, 2012). Perhaps the greatest example of this in finance is the Black-

[6] g_π and g_y are weights that represent how much the central bank values divergences in expected inflation or real activity and usually add up to 1.

[7] Epstein and Zin (1991) define a recursive utility function, where present utility is a function of present consumption and future expected utility. The intertemporal elasticity of substitution determines consumption growth with respect to the return rate on savings.

Litterman model (BLACK; LITTERMAN, 1990), which builds from the mean-variance portfolio optimization while allowing for prior subjective information on asset returns.

With the goal of predictive modelling in hand, it is worth noticing the characteristics of that goal, and how Bayesian reasoning might be helpful. The challenges posed by predictive modelling can differ from those of an inferential and/or causality perspective. One of these challenges is the bias-variance trade-off, as in classical statistical inference the bias and variance tolerance can radically differ from predictive modelling. Overfitting of models can be another concern (SHMUELI, 2010).

In machine learning, one of the ways to improve predictive performance of regression models is to introduce regularization, which is usually done in non-Bayesian settings by adding constraints on the regression objective function (SCHMIDT; FUNG; ROSALES, 2009). Ordinary least squares regression can be thought of an optimization problem, where the function to be minimized is the residual sum of squares. Common machine learning techniques such as ridge and LASSO introduce additional terms to the regression loss function, "shrinking" coefficients towards zero. The procedure can introduce more bias, but may also increase predictive performance and avoid overfitting (HELWIG, 2017). This can also help in variable selection.

As for the Bayesian case, the regularization can be imposed by informative prior distributions (POLSON; SOKOLOV, 2019), which is very straightforward in many MCMC algorithms. Another possible advantage of the Bayesian approach is that the posterior distribution contains all the information we need about the parameter after having observed the data. This decreases the reliance on standard summarizing statistics such as the mean, median or mode, although these can be obtained as Bayes estimators with specific loss functions. Common regularization procedures such as ridge or LASSO for regression models can be obtained in a Bayesian framework with certain priors on the coefficients, more specifically normal priors for ridge and Laplace priors for LASSO (TIBSHIRANI, 1996). Puelz (2018) gives a comprehensive overview of regularization methods in econometrics and finance, mostly from a Bayesian perspective, combining decision theory and model selection methods. As we can see, Bayesian inference provides a flexible way of imposing regularization, although Sims (2010) notes that Bayesian inference is not just a set of tools but a different approach to probabilistic reasoning.

Differences in frequentist and Bayesian reasoning can be traced back to epistemological concerns, and both approaches can contribute to statistical practice (BAYARRI; BERGER, 2004). Bayesian estimation has also become the dominant approach for estimation of DSGE models in macroeconomics, e.g., Smets and Wouters (2003), although for more practical than theoretical reasons. Macroeconomic data usually consist of quarterly time series of macroeconomic variables and the likelihood functions have many local maxima/minima and near-flat surfaces. Integrating the likelihood (times the prior) with computational methods, e.g., MCMC is much easier than optimizing the likelihood function (FERNÁNDEZ-VILLAVERDE, 2010).

2.2.2 Bayesian estimation of dynamic models

Time series, or dynamic models, have been a very prominent subfield of statistics. Specific estimation techniques, both likelihood-based and Bayesian, have been developed for models of that type. In the case of frequentist estimation, adaptations of the generalized method of moments (HANSEN; SINGLETON, 1982) or maximum-likelihood versions of the Kalman filter are commonly used. As for the Bayesian case, extensions of the Kalman filter, such as the forward-filtering backward-sampling (FFBS) smoothing algorithm (CARTER; KOHN, 1994) (FRÜHWIRTH-SCHNATTER, 1994) have become prominent. The FFBS algorithm has since been used to model financial data such as in stochastic volatility models (KIM; SHEPHARD; CHIB, 1998). A possible motivation for using Bayesian methods in time series analysis is that some likelihood methods, such as versions of the expectation maximization algorithm, might get "trapped" in local maxima for state space models (KANTAS *et al.*, 2015).

One of the main drawbacks of the FFBS algorithm is that, in order to obtain the joint distribution of the state vector, it relies on conditional sampling, which can become computationally expensive for longer time series, or higher dimensional data. Novelty methods for sampling of the state vector such as Chan and Jeliazkov (2009) and McCausland, Miller and Pelletier (2011) rely on the block diagonal structure of precision matrices in order to directly draw from the joint distribution of the state vector. Another concern regarding the FFBS algorithm is that it might generate biased estimates in nonlinear non-Gaussian settings (MORAL; DOUCET; SINGH, 2010).

Recent advances in Bayesian computation such as the Hamiltonian Monte Carlo (HMC) (BETANCOURT, 2017) and automatic differentiation variational inference (KUCUKELBIR *et al.*, 2017) can improve performance of Bayesian estimation. HMC seeks to improve upon the popular random-walk Metropolis algorithm with ideas from Hamiltonian mechanics. More specifically, HMC involves a transformation of the posterior parameter space into a bidimensional Hamiltonian phase space by the introduction of a momentum variable. The sampling occurs in the Hamiltonian phase space, and then marginalization in the momentum variable recovers the posterior distribution. The joint density of the phase space depends on the Hamiltonian function, which is the sum of kinetic and potential energy functions. While the potential energy function depends only on the target distribution, the kinetic energy depends on a mass matrix (BETANCOURT, 2017). Hoffman and Gelman (2014) describe the No U-Turn Sampling (NUTS) variant of HMC and applications, including a stochastic volatility model.

Automatic differentiation variational inference (ADVI) can be an alternative to Monte Carlo methods. It seeks to approximate the posterior through the minimization of the Kullback-Leibler (KL) divergence between the actual posterior and another approximating family of distributions. As there is no analytic form for the KL divergence, instead the evidence lower bound (ELBO) is maximized, as maximization of the ELBO is equivalent to minimizing the KL divergence. ADVI employs stochastic gradient ascent in order to maximize the ELBO

(KUCUKELBIR *et al.*, 2017), performing approximate inference when regular MCMC is unfeasible.

2.2.3 Bayesian inference in asset pricing and macroeconomics

The field of finance is highly dependent on statistical modelling, presenting several opportunities for the use of Bayesian inference to estimate model parameters and compute predictive densities. Applications range from market efficiency and return predictability tests to the use of Bayesian computation to estimate models with high computational cost, e.g., stochastic volatility models (JOHANNES; POLSON, 2010, p. 2).

Eraker (2001) presents a detailed study on the use of Bayesian inference and MCMC simulation for both affine interest rate and stochastic volatility models. The conclusion is that, besides high computational cost, MCMC algorithms show a good performance in comparison to maximum likelihood estimation for model parameters.

Ang, Dong and Piazzesi (2007) employ Bayesian estimation techniques for a macroeconomic model with Taylor rules for monetary policy, identifying monetary shocks through the affine no-arbitrage framework of Duffie and Kan (1996).[8] The authors impose priors in order to assure stationarity of the state-space and report the possibility of obtaining the full posterior distribution and increased tractability as the advantages of Bayesian estimation. Chib and Ergashev (2009) estimate a similar model with informative priors that result in an upward-sloping yield curve, justifying that choice with economic theory.

In the case of affine short-rate models, Gray (2005) applies Bayesian inference to the parameters of two models. The first one is the model from Chan *et al.* (1992), which is a general form that contains many well-known models as special cases. The second one is the canonical Vasicek (1977) model. The Chan *et al.* (1992) model is

$$dr_t = \alpha(\mu - r_t)dt + \sigma r_t^{\gamma} dW_t,$$

where the special case $\gamma = 0$ yields the Vasicek model, and $\gamma = 0.5$ results in the CIR model. The author notes that the Bayesian estimates compare favorably to the maximum likelihood counterparts.

Bayesian inference can also be applied to the DNS model. Laurini and Hotta (2010) start from the DNS model as presented in Diebold and Li (2006), also adding a stochastic volatility structure and time-varying decay parameter.[9] Bayesian estimation with informative priors allows reduced numerical instability. Caldeira, Laurini and Portugal (2010) also propose similar extensions to the Diebold-Li specification of the dynamic Nelson-Siegel model, such as the time-varying decay and stochastic volatility for the factor loadings. The authors notice improvements

[8] See Taylor (1993) for a better description of Taylor rules.
[9] Some of the modifications were introduced by Svensson (1994).

by the use of conditional heteroskedasticity. Das (2019) applies hierarchical Bayesian modeling and HMC sampling to the DNS model, obtaining a strong negative relationship between the level factor and Monte Carlo-simulated bond prices.

CHAPTER

3

METHODOLOGY AND MODELS

Our methodology consists of specifying five models, of which two are affine (Vasicek and CIR), as defined by Piazzesi (2010), and the others are versions of the dynamic Nelson-Siegel model from Diebold and Li (2006) in state space form. The Vasicek and CIR model are specified with the short rate process in continuous time. The observation equations are obtained mainly from Lindström, Madsen and Nielsen (2015) in pages 236 and 239, respectively, although some modifications present in Brigo and Mercurio (2001) are also employed. Following the approach by Johannes and Polson (2010, p. 54), we introduce some additional parameters for the observation equations where latent process parameters would appear. For the DNS model, we give three different specifications of the Nelson-Siegel decay parameter: as a calibrated constant; free parameter to be estimated; and as a dynamic parameter to be estimated. More details are provided in the following sections.

The choice of hyperparameter values is done by prior predictive checking, a procedure described in Gabry *et al.* (2019). This is done by guessing values for the hypeparameters and simulating data from the prior predictive distribution until the results look like data that can plausibly be observed. The procedure assigns some prior mass for hyperparameter values that generate extreme but plausible data sets, and no mass on completely implausible data sets. For the dataset we use in Chapter 5, which corresponds to the Brazilian yield curve from 2012 to 2020, it seems implausible that we would observe short rates above 0.3, or negative short rates. Therefore, this was taken into account in the priors selection procedure. The same goes for 1 year spot rates, for example.

Besides common probability distributions such as the Gaussian distribution, we also use less common distributions as priors for some parameters. These include the truncated normal and half-Cauchy distributions. The noncentral chi-squared distribution is also mentioned as the distribution for the short rate process under the CIR model. More details about these distributions can be found in Appendix A, with information about stochastic processes, such as the Wiener process.

The models were estimated using the Python programming language and the PyMC3 library (SALVATIER; WIECKI; FONNESBECK, 2016). The NUTS version of Hamiltonian Monte Carlo (HOFFMAN; GELMAN, 2014) is used for sampling from the posterior distributions, and the mass matrix of the Hamiltonian is initialized by automatic differentiation variational inference, or ADVI (KUCUKELBIR *et al.*, 2017). We simulate a chain of size 2000 with 1000 burn-in draws.

PyMC3 lets us implement the Vasicek and CIR differential equations (latent process) in the form of an Euler-Maruyama discretization (ELERIAN; CHIB; SHEPHARD, 2001), which we also do for the DNS factors. We specify $dt = \Delta t = \frac{1}{52}$. We shall evaluate the convergence of the simulated chains checking the density, trace and autocorrelation plots for each of the two chains for each model. The kernel density estimates (KDEs) for the marginal posterior distribution of the parameters are presented in Appendix B with the trace and autocorrelation plots.

As all models are specified in state space form, we add observation noise to each observation equation as well. In all models, we denote the random noise as η_t, which follows a normal distribution with mean zero and variance σ^2_{obs}. Treating both the latent vector and observation noise standard deviations as unknowns can be computationally expensive, so we calibrate and fix a value for σ_{obs}. Besides, stochastic singularity can also pose a problem to model estimation, which can be solved by adding measurement error to the observation equations (JOHANNES; POLSON *et al.*, 2003).

Schmidt, Krämer and Hennig (2021) analyze a problem of Bayesian inference in state space models with latent variables in differential equation form, in the context of epidemiological data. In their case, they apply the model with simulated data and set the variance of the latent process for the data 3 orders of magnitude greater than the measurement noise variance. Maybank, Bojak and Everitt (2017) study Bayesian inference for stochastic differential equations in a multidimensional neuroscience model, fixing some of the model parameters based on results of previous studies. In their model, the observation noise is fixed and two orders of magnitude smaller than the latent equation variances, also treated as known values. Within the dataset used in Chapter 5, the standard deviations of the percent changes of spot rates across maturities range from 0.012 (1 month) to 0.036 (3 years), so we deem reasonable to fix σ_{obs} to the value of 0.001.

3.1 Vasicek model

The model first introduced by Vasicek (1977) assumes the short rate follows an Ornstein-Uhlenbeck process. This choice ensures the short rate has the empirically observed mean-reverting property. Given its relevance as one of the benchmarks for term structure modelling, the Vasicek model seems like a good choice for a starting point.

The Vasicek model in state space form corresponds to

$$\begin{aligned} dr_t &= \kappa(\mu - r_t)dt + \sigma dW_t, \\ Y_t(\tau) &= -\frac{[A(b,a,\sigma,\tau) - B(b,\tau)r_t]}{\tau} + \eta_t, \eta_t \sim N(0, \sigma^2_{obs}), \\ B(b,\tau) &= \frac{1-\exp\{-b\tau\}}{b}, \\ A(b,a,\sigma,\tau) &= \left(\frac{\sigma^2}{2b^2} - a\right)\tau + \frac{1-\exp\{-b\tau\}}{b}\left(a - \frac{\sigma^2}{b^2}\right) + \frac{\sigma^2}{4b^3}(1-\exp\{-2b\tau\}), \end{aligned}$$

where $Y_t(\tau)$ is the observed spot rate for maturity τ, r_t is the short rate, W_t is the Wiener process i.e., $dW_t = \varepsilon\sqrt{dt}$, $\varepsilon \sim N(0,1)$. Our basis for the Vasicek model is the specification given in page 236 of Lindström, Madsen and Nielsen (2015), although modifications are made so that the coefficient $B(b,\tau)$ in the spot rate equation has a negative sign. This is usual in literature, as the sign of the coefficient does not preclude the affine structure. One example of this modification for the Vasicek model can be found on Brigo and Mercurio (2001, p. 59).

In the standard Vasicek model the parameters κ and μ also appear in the observation equation in place of b and a, respectively. However, in the Vasicek model the process for r_t occurs in a risk-neutral world, which means that modelling the market price of risk would be necessary for estimating the term structure with real data. This would require the introduction of an additional parameter in the state equation. Instead of doing this, we follow the approach by Johannes and Polson (2010) and introduce the parameters a and b to the observation equation as the real-world correspondents of μ and κ.

The parameters κ and μ appear in the state equation, a and b in the observation equation and $\sigma > 0$ in both state and observation equations. The long-term mean of the short rate is represented by μ, while κ can be interpreted as the mean speed at which the short rate reverts to its long-term mean. The standard deviation of the random normal shocks that affect the short rate, often called volatility, is denoted by σ. It will also affect the long-run level of the yield curve, as it enters the intercept of the observation equation. The parameters a and σ determine the long-run level of the yield curve, i.e., the level of the spot rates with higher maturities. The parameter b controls the decay of long-run rates in relation to shorter-maturity spot rates.

The prior probability distributions with hyperparameter values for the parameters are

$$\begin{aligned} \kappa &\sim TN_{(0,\frac{2}{\Delta t})}(0.7, 0.16), \\ \mu &\sim TN_{(0,\infty)}(0.08, 0.01), \\ \sigma &\sim HalfCauchy(0.03), \\ a &\sim N(0.1, 0.01), \\ b &\sim N(0.5, 0.25), \end{aligned} \tag{3.1}$$

where $TN_{(\alpha,\beta)}(\mu,\sigma^2)$ denotes the truncated normal distribution on interval (α,β) with mean μ and variance σ^2.

We truncate the κ Gaussian prior in order to impose stationarity and use regularizing priors for all parameters. We follow the recommendation by Polson and Scott (2012) to adopt half-Cauchy priors for the standard deviation parameters. The half-Cauchy prior has superior quadratic risk properties and better behavior in the presence of sparsity in comparison with, e.g., the inverse-gamma prior (POLSON; SCOTT, 2012). The variable η_t only corresponds to observation noise, which is added to the observation equation for, among other possible reasons, breaking a possible stochastic singularity (JOHANNES; POLSON, 2010, p. 55), but its standard deviation σ_{obs} has no empirical interpretation.

3.2 CIR model

The Cox, Ingersoll and Ross (1985) model aimed to solve some of the presumed issues with the Vasicek model, mainly allowing for negative short rates and having constant volatility. The CIR model assumes a square-root process for the short rate, ensuring it does not assume negative values. The short rate under the CIR model also has a conditional heteroskedasticity structure. It is worth mentioning that the short rate under the CIR model follows the same process as the log-volatility in the Heston (1993) stochastic volatility model.

The CIR model in state space form is

$$
\begin{aligned}
dr_t &= \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t, \\
Y_t(\tau) &= -\frac{A(\tau) - B(\tau)r_t}{\tau} + \eta_t, \\
\eta_t &\sim N(0, \sigma^2_{obs}), \\
\gamma &= \sqrt{b^2 + 2\sigma^2}, \\
g(\tau) &= 2\gamma + (b+\gamma)(\exp\{\gamma\tau\} - 1), \\
B(\tau) &= 2(\exp\{\gamma\tau\} - 1)/g(\tau), \\
A(\tau) &= \frac{2ba}{\sigma^2}\ln\left[\frac{2\gamma\exp\{(b+\gamma)\tau/2\}}{g(\tau)}\right].
\end{aligned}
$$

The interpretations of the state equation parameters κ and μ are similar to the Vasicek model in Section 3.1, as both models have the same drift coefficient. The parameter σ is a scaling factor for the volatility of the short rate, given that the volatility also depends on the square root of the short rate level. It also affects the decay of the observed spot rates as well, because it appears in the equation for γ. The parameters a and b also have similar interpretations as in the Vasicek model, so more details can be found in Section 3.1. We also use the approach from Johannes and Polson (2010, p. 54), treating the parameters in the state and observation equations, except for σ, as separate.

The main difference with the Vasicek model is that now $\sqrt{r_t}$ is multiplying the diffusion coefficient in the state equation. For the observation equations, we once again combine the specifications in Lindström, Madsen and Nielsen (2015, p. 239) and Brigo and Mercurio (2001, p. 66). This results in a coefficient $B(\tau)$ with a negative sign in the spot rate equation, as in the specification we use for the Vasicek model.

The short rate r_t has a noncentral chi-squared distribution under the CIR model (BRIGO; MERCURIO, 2001, p. 65). We do not use the exact transition densities for the stochastic simulation, but instead use an Euler-Maruyama discretization, as we do with the latent variables of the other estimated models. Results such as Higham and Mao (2005) and Alfonsi (2013) show that the Euler-Maruyama discretization displays strong convergence properties with a small enough time step.

The prior distributions for the corresponding parameters are

$$\begin{aligned}
\kappa &\sim TN_{(0,\frac{2}{\Delta t})}(0.7, 0.16),\\
\mu &\sim TN_{(0,\infty)}(0.08, 0.01),\\
\sigma &\sim HalfCauchy(0.1),\\
a &\sim N(0.1, 0.01),\\
b &\sim N(0.5, 1).
\end{aligned} \tag{3.2}$$

3.3 Dynamic Nelson-Siegel (DNS) model

The original Nelson and Siegel (1987) model is static, in the sense that it only estimates one equation of the yield curve at a fixed point in time. Diebold and Li (2006) modify the original Nelson-Siegel model by allowing factors to vary with time, creating a dynamic version that we will refer to as the dynamic Nelson-Siegel model, or DNS.

For the state space form of the DNS model, we specify three continuous time AR(1) latent factors, each corresponding to level, slope and curvature. These factors follow processes similar to the Vasicek short rate process. It is usual in the literature to estimate the factors in a vector autoregression structure, but we assume them to be independent for simplicity. The model is

$$\begin{aligned}
dl_t &= \kappa_l(\mu_l - l_t)dt + \sigma_l dW_{1t},\\
ds_t &= \kappa_s(\mu_s - s_t)dt + \sigma_s dW_{2t},\\
dc_t &= \kappa_c(\mu_c - c_t)dt + \sigma_c dW_{3t},\\
Y_t(\tau) &= l_t + s_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + c_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \eta_t, \eta_t \sim N(0, \sigma^2_{obs}),
\end{aligned}$$

where λ is a decay parameter which controls factor loadings and (W_{1t}, W_{2t}, W_{3t}) are independent Wiener processes.

The panel of observed spot rates for different maturities is modeled as a function of the three latent factors, l_t, s_t and c_t, each corresponding to level, slope and curvature. The factor loadings are parametrized so that the latent variables can be interpreted as level, slope and curvature factors in accordance with Litterman and Scheinkman (1991).

We chose prior distributions for the factor process parameters by the prior predictive check procedure. The prior distributions are

$$\begin{aligned}
\kappa_l &\sim TN_{(0,\frac{2}{\Delta t})}(1,1),\\
\kappa_s &\sim TN_{(0,\frac{2}{\Delta t})}(0.8,0.25),\\
\kappa_c &\sim TN_{(0,\frac{2}{\Delta t})}(0.5,0.16),\\
\mu_l &\sim TN_{(-0.5,\infty)}(0.12,0.0025),\\
\mu_s &\sim TN_{(-\infty,0.5)}(-0.04,0.0009),\\
\mu_c &\sim TN_{(-\infty,0.5)}(-0.05,0.0016),\\
\sigma_{l,s,c} &\sim HalfCauchy(0.015).
\end{aligned} \tag{3.3}$$

In the following subsections we give more detail into the three different versions we estimate of the DNS model with regards to the parameter λ.

3.3.1 Calibrated λ

For the calibrated λ version of the DNS model (DNS-C), the decay parameter λ is treated as a fixed quantity. The parameter is calibrated by maximizing λ in the curvature loading expression

$$\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

from the equation defining $Y_t(\tau)$, with τ chosen to represent an "intermediate" maturity. This is done in order to intermediate maturities for the curvature and emulate humped yield curves observed on the market. We pick the value of $\lambda = 1.2$ so that the curvature loading peaks around the intermediate one-year maturity.

3.3.2 Free λ

For the free λ version of the DNS model (DNS-F), we treat λ as a free parameter to be estimated. We consider λ the prior

$$\lambda \sim N(1.1,0,09).$$

3.3.3 Dynamic λ

In the third and final version of the DNS model (DNS-D), we treat λ as following a time series process over time. Therefore, we add a time subscript and write λ_t. We specify λ_t as the

random-walk process

$$\begin{aligned}\lambda_{t+1} &= \lambda_t + \varepsilon_t^{\lambda},\\ \varepsilon_t^{\lambda} &\sim N(0,\sigma_{\lambda}^2),\\ \sigma_{\lambda} &\sim HalfCauchy(1).\end{aligned}$$

CHAPTER

4

SIMULATED DATA

In this chapter each of the models shall be estimated for simulated term structure data. Estimation of the model with simulated data can be useful in assessing both possible computational issues and estimation of the latent variables, whose values for the real data are unknown (GELMAN *et al.*, 2020). Also, in an environment where we know the true parameter values for the data generating process, it becomes possible to evaluate the performance of the parameter estimates. Starting from the premise that the models seek to be a good enough approximation to the real data generating process (JOHANNES; POLSON *et al.*, 2003), we simulate data without the observation noise included in our statistical models.

The parameter estimates for each model shall be based on the absolute loss function, therefore the estimates will be equal to the posterior median for each parameter. More specifically, if we have the loss function $L(\theta, \delta(X))$ for the parameter θ and the estimator $\delta(X)$, the Bayes estimator shall be the value for $\delta(X)$ that minimizes the expected loss

$$\mathrm{E}[L(\theta, \delta(X)|X] = \int_{\Theta} L(\theta, \delta(X))P(\theta|X)d\theta,$$

and, for the case of the absolute loss function, i.e., $L(\theta, \delta(X)) = |\theta - \delta(X)|$, the Bayes estimator for real-valued parameters shall be the posterior median (DEGROOT; SCHERVISH, 2012, p. 411). Although other loss functions, such as the squared loss, could have also been used, the absolute loss tends to be more robust for outliers. As we use the absolute loss as the benchmark for prediction in Chapter 5, we also use it for parameter estimation for standardization purposes.

4.1 Vasicek model

We estimate the Vasicek model as presented in Section 3.1, with prior distributions specified in (3.1). In order to estimate the model and analyze its properties with respect to the data, we simulated data according to the model equations and the following parameter values:

$\kappa = 0.4$, $\mu = 0.05$, $\sigma = 0.02$, $a = 0.2$ and $b = 0.3$. The observation error standard deviation is fixed as $\sigma_{obs} = 0.001$.

We first simulated a $T = 440$ latent vector and then proceeded to calculate the spot yields from the observation equations. Since we wanted to reproduce weekly data, we set $\Delta t = \frac{1}{52}$ and 6 spot maturities, i.e., different values for τ: 0.17, 0.25, 0.5, 1, 3, 5.[1] We simulated the data in order to reproduce long-term behavior, i.e., ergodicity, so the starting value is not relevant.[2] The simulated data is presented in Figure 6. We then proceeded to estimate the model via HMC.

Table 1 presents the statistical summary for the Vasicek model. The median and standard deviations of the estimated posteriors are presented, with the lower and upper bounds of the 95% highest posterior density (HPD) interval, denoted by LB and UB, respectively. We present some additional MCMC diagnostics in Appendix B. The real parameter values are presented in the first column, with its respective parameter. Since Table 1 contains the posterior medians, standard deviations, HPD bounds and true values it is possible to assess how near the posterior medians and real values are to each other.

Table 1 shows that all true values used to simulate the data are contained within the 95% highest posterior density interval. The posterior median for μ is 0.11 standard deviation away from the assumed value in the simulation. For κ the distance between the posterior median and the true value is of 0.24 standard deviation, and κ of all parameters seems the less sensible to the data. The median for a, b and σ come close to the real values. The estimated short rate visually comes very close to the true simulated values, as presented in Figure 7. All of the 440 true values for the Vasicek short rate are contained in the 95% HPD interval.

Table 1 – Statistical summary for the Vasicek model - Simulated data

Parameter	Median	SD	LB	UB
$a = 0.200$	0.200	0.001	0.199	0.201
$b = 0.300$	0.300	0.002	0.297	0.303
$\kappa = 0.400$	0.460	0.251	0.019	0.917
$\mu = 0.050$	0.053	0.028	0.012	0.106
$\sigma = 0.020$	0.020	0.001	0.019	0.021

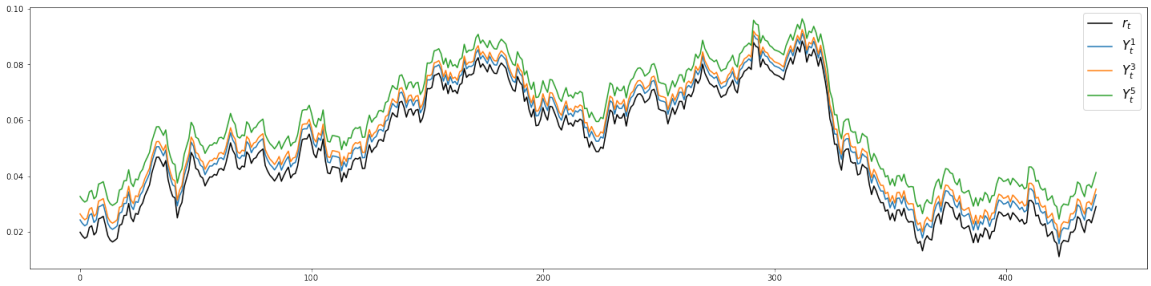
4.2 CIR model

The estimated CIR model is the one presented in Section 3.2, with the prior distributions from (3.2). We simulated the data and estimated the model in the same way done for the Vasicek model in the previous section. As the graph for the simulated data looks similar to the Vasicek

[1] The set of maturities correspond to time periods commonly observed on the market: 2 months, 3 months, 6 months, 1 year, 3 years and 5 years.

[2] The ergodicity behavior is obtained by simulating a series of size $2T = 880$ and discarding the first half i.e., the first 440 samples.

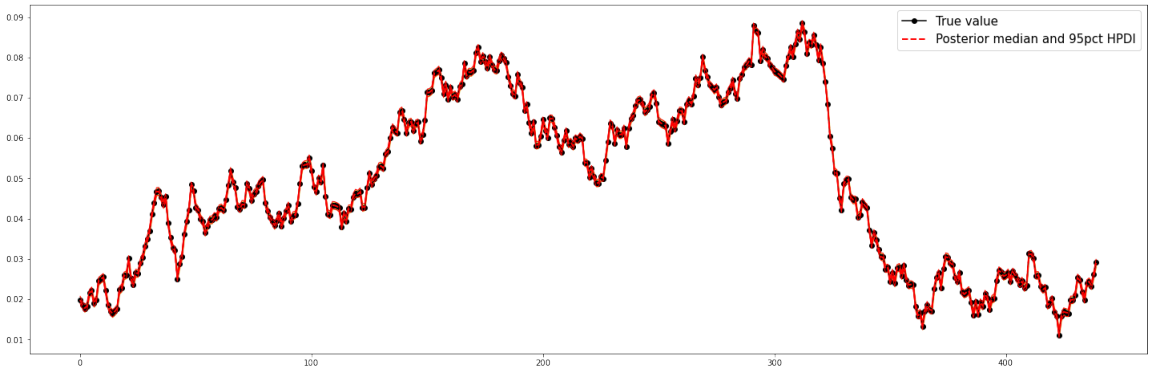
Figure 6 – Simulated data for the Vasicek model



Note – The short rate r_t is presented along with the observed spot rates for the maturities 1, 3 and 5.

Source: Elaborated by the author.

Figure 7 – Estimated short rate and 95% highest posterior density interval for the Vasicek model - Simulated data



Source: Elaborated by the author.

Note – The black line is the simulated state vector, and the red line the posterior median with the 95% highest posterior density interval.

simulated data in Figure 6, it is presented in Appendix B. We used the following values for the parameters: $\kappa = 0.4$, $\mu = 0.05$, $\sigma = 0.1$, $a = 0.2$ and $b = 0.3$. The observation error standard deviation was fixed as $\sigma_{obs} = 0.001$. The length of the data is also $T = 440$, the size of the discretization time step $\Delta t = \frac{1}{52}$ and the maturities 0.17, 0.25, 0.5, 1, 3 and 5.

The summary of the estimates for the CIR model is presented in Table 2. The posterior medians for parameters a and b are the same as the real values, as in the Vasicek model. The posterior mean for σ, however, is 0.33 SD away from its real value. The κ parameter seemed more responsive to the data and had its median close to the true value, with a distance of 0.125 standard deviation. The posterior median for μ had a distance of 0.06 SD to its real value, smaller than its equivalent for the Vasicek model. All true values for the parameters were contained in the 95% HPD intervals. Once again, the estimated short rate visually came very close to the true simulated values for the short rate, in a similar fashion to the Vasicek model. We observed that all of the 440 true values are contained in the 95% HPD interval for the short rate.

Table 2 – Statistical summary for the CIR model - Simulated data

Parameter	Median	SD	*LB*	*UB*
$a = 0.200$	0.200	0.001	0.198	0.201
$b = 0.300$	0.300	0.002	0.297	0.304
$\kappa = 0.400$	0.431	0.248	0.001	0.864
$\mu = 0.050$	0.052	0.034	0.001	0.120
$\sigma = 0.100$	0.099	0.003	0.092	0.106

4.3 DNS model

The DNS model is presented in Section 3.3, with the prior distributions given in (3.3). The three different specifications for the λ parameter mentioned in the previous section were tried, which means we estimated three different DNS models. We simulated one process for each of the three factors, and then proceeded to construct the observed spot rates from the observation equation.

The real parameter values were chosen in order to replicate the behavior of real term structures, with the level factor being positive and the slope and curvature factors being negative most of the time.[3] Therefore, the true value for the mean parameter of the level factor was positive ($\mu_l = 0.11$) and the assumed values for the mean of the slope and curvature factors were negative ($\mu_s, \mu_c = -0.03$). The true values for the mean-reversion speed parameters were: $\kappa_l = 0.8$, $\kappa_s = 0.6$, and $\kappa_c = 0.4$. As for the factor standard deviations, the assumed values were: $\sigma_l = 0.02$, $\sigma_s = 0.03$ and $\sigma_c = 0.04$. For the free λ model, λ had a true value of 1.2 and for the time-varying λ model σ_λ had a true value of 0.07. The observation error standard deviation was fixed as $\sigma_{obs} = 0.001$. The data once again had length $T = 440$ and $\Delta t = \frac{1}{52}$, and the spot rate had the same six different maturities: 0.17, 0.25, 0.5, 1, 3 and 5.

Simulated factors and spot rates for the fixed λ DNS model are presented in Figures 8 and 9 as an example. As the simulated factors and spot rates for the other two DNS models were similar, these are presented in Appendix B. Figure 10 illustrates the estimated DNS factors for the fixed λ DNS model. Once again, the estimated factors for the other two DNS models can be found in Appendix B.

Table 3 contains the statistical summary for the DNS model with fixed λ. The standard deviation parameters $\sigma_{l,s,c}$ had their posterior medians 1, 4 and 6 SD away from their real values, respectively. The assumed values fell out of the bounds for the 95% HPD interval, except for σ_l. The distributions for $\kappa_{l,s,c}$ seemed to learn less from the data, with wide 95% HPD intervals that all contained the true values. The mean parameters $\mu_{l,s,c}$ are the best in terms of estimation, as the posterior medians were closer to the assumed values (0.1, 0.56 and 0.9 SD, respectively). For the estimated latent factors, 4 points for the level factor and 14 points for the curvature factor fell

[3] It makes sense that the slope and curvature factors are only positive during periods where the yield curve is inverted, as this will cause shorter maturities to have higher yields than longer maturities.

outside the 95% HPD bounds.

Tables 4 and 5 present the statistical summaries for the DNS models with free and time-varying λ. For the free λ model, only the real values for σ_s and σ_c fell outside the 95% HPD interval. A total of 6 data points for the level factor and 14 for the curvature factor fell outside the 95% HPD interval bounds. As for the time-varying λ model, the same σ_s and σ_c parameters were the ones whose real values fell outside the 95% HPD interval bounds. The real value for σ_λ also fell outside of the 95% HPD bounds. We had 12 data points for the level factor, 10 for the slope factor and 11 for the curvature factor not contained in the 95% HPD intervals.

Table 3 – Statistical summary for the DNS model (calibrated λ) - Simulated data

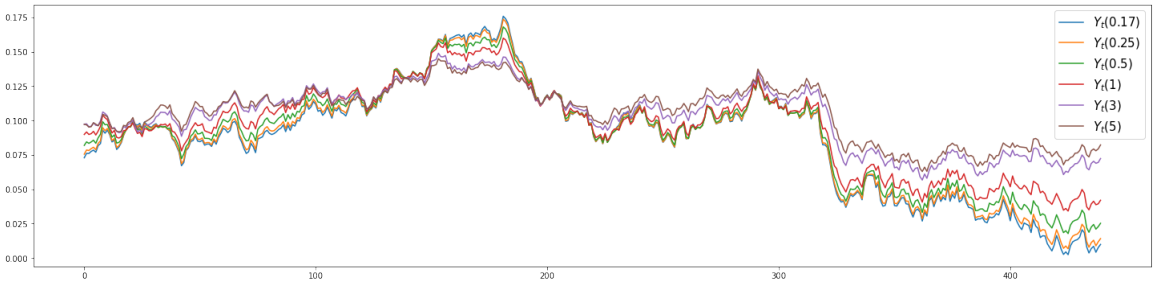
Parameter	Median	SD	*LB*	*UB*
$\kappa_l = 0.800$	0.494	0.316	0.005	1.071
$\mu_l = 0.110$	0.112	0.020	0.076	0.161
$\sigma_l = 0.020$	0.019	0.001	0.017	0.020
$\kappa_s = 0.600$	0.278	0.189	0.000	0.629
μ_s = -0.030	-0.043	0.023	-0.091	0.002
$\sigma_s = 0.030$	0.026	0.001	0.024	0.028
$\kappa_c = 0.400$	0.585	0.297	0.005	1.119
μ_c = -0.030	-0.012	0.020	-0.054	0.027
$\sigma_c = 0.040$	0.028	0.002	0.025	0.032

Figure 8 – Simulated factors for the DNS model (calibrated λ)

Note – The presented factors correspond to level, slope and curvature of the yield curve.

Source: Elaborated by the author.

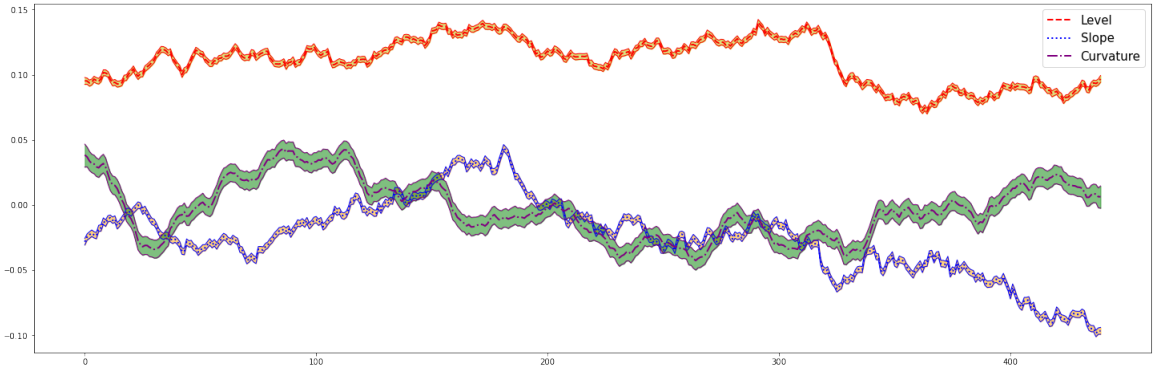
Figure 9 – Simulated spot yields for the DNS model (calibrated λ)



Note – The presented maturities are 0.17, 0.25, 0.5, 1, 3 and 5.

Source: Elaborated by the author.

Figure 10 – Estimated latent factors and 94% highest posterior density intervals for the DNS model (calibrated λ) - Simulated data



Source: Elaborated by the author.

Table 4 – Statistical summary for the DNS model (free λ) - Simulated data

Parameter	Median	SD	*LB*	*UB*
λ = 1.200	1.196	0.032	1.132	1.260
κ_l = 0.800	0.489	0.333	0.002	1.136
μ_l = 0.110	0.111	0.022	0.066	0.161
σ_l = 0.020	0.019	0.001	0.018	0.021
κ_s = 0.600	0.264	0.195	0.001	0.673
μ_s = -0.030	-0.043	0.023	-0.097	-0.001
σ_s = 0.030	0.026	0.001	0.024	0.028
κ_c = 0.400	0.616	0.293	0.053	1.150
μ_c = -0.030	-0.011	0.020	-0.059	0.020
σ_c = 0.040	0.028	0.002	0.024	0.032

Table 5 – Statistical summary for the DNS model (time-varying λ) - Simulated data

Parameter	Median	SD	*LB*	*UB*
σ_λ = 0.100	0.055	0.010	0.037	0.076
κ_l = 0.800	0.461	0.329	0.001	1.104
μ_l = 0.110	0.112	0.021	0.068	0.156
σ_l = 0.020	0.019	0.001	0.017	0.021
κ_s = 0.600	0.260	0.189	0.001	0.627
μ_s = -0.030	-0.043	0.023	-0.094	0.001
σ_s = 0.030	0.026	0.001	0.024	0.029
κ_c = 0.400	0.546	0.305	0.006	1.118
μ_c = -0.030	-0.012	0.021	-0.061	0.024
σ_c = 0.040	0.028	0.002	0.024	0.033

CHAPTER

5

REAL DATA APPLICATION

In this chapter we estimate the models with real term structure data from the Brazilian market. The data utilized in order to fit our models comes from NEFIN FEA-USP – Center for Research in Financial Economics of the Department of Economics of the University of São Paulo. It consists of the spot rate curve for Brazil, calculated from One-Day Interbank Deposit Futures contracts (DI rate) via flat-forward interpolation.[1] Although flat-forward interpolation is known to possibly introduce some issues, such as arbitrage opportunities (HAGAN; WEST, 2006), discussions about yield curve construction methods are beyond the scope of this work.

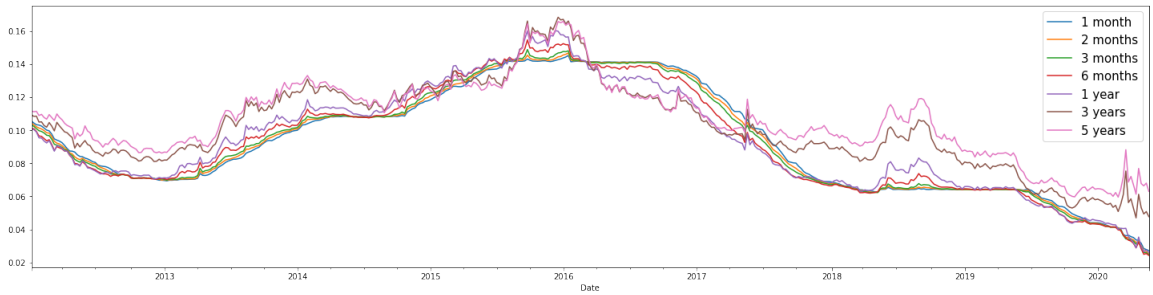
For model estimation we used a subsample which starts in January 2012 and ends on May 2020, with weekly observations ($\Delta t = \frac{1}{52}$), for a total sample size of $T = 438$. Our panel of spot rates included the following maturities: 2 months, 3 months, 6 months, 1 year, 3 years and 5 years. We use the priors mentioned in Chapter 3 for model estimation.

Figure 11 contains the graphical representation of the data. We can see that the rates do not go above 20%, neither below the 2% threshold, and, as mentioned in Chapter 3, this was accounted for in prior hyperparameter choice. Although we see both periods with steep and inverted yield curves, this period can be seen as quite stable for the Brazilian yield curve, if compared to the previous decades. The rates for short maturities are much less volatile than the ones for longer maturities.

It is of relevance that we estimate the proposed models with real market data. We learned from the previous chapter that the models are successfully able to estimate the latent vectors, since we worked in an environment where the true values were known. Because the data were simulated in order to look like real yield curves, we should also observe the same patterns regarding the convergence of the simulated samples. We already mentioned in Section 2.1.6 of the literature review how the term structure is relevant to macroeconomic policy. Forecasting of the yield curve is also crucial in financial markets, as it is used to price bonds that have their

[1] The data can be found at <http://nefin.com.br/data/spot_rate_curve.html>

Figure 11 – Brazilian yield curve from January 2012 to May 2020



Source: Elaborated by the author.

values as a function of future expected rates.

Besides model estimation, in this real data application section we can also use the estimated results to obtain predictions of future market data. Therefore, in this chapter we are also interested in evaluating the performance of the models for forecasting future observations. Our main measure of model forecasting ability is the rolling window mean absolute error (MAE), introduced in Section 5.4, and is based on the absolute loss function.

5.1 Vasicek model

We estimate the same model in Section 3.1, with the same prior distributions for the parameters, using the yield curve data described at the start of the current chapter. We fix $\sigma_{obs} = 0.001$. In the same style of the previous chapter, Table 6 presents the statistical summary for the Vasicek model with real term structure data. Medians for the parameters marginal posterior distributions, standard deviations and 95% HPD bounds (LB and UB).

Additional MCMC diagnostics such as Monte Carlo standard error (MCSE) and effective sample size (ESS) are presented in Appendix B. In terms of convergence of the simulated samples, no major issues were found for the Vasicek model. Marginal posterior distributions for the parameters, together with trace and autocorrelation plots, are presented in Appendix B.

Figure 12 shows the estimated short rate vector, with its posterior mean and 95% HPD intervals. The MCMC diagnostics presented in Appendix B show the absence of major problems regarding convergence of the simulated samples. The empirical interpretations of κ and μ is that r_t is pulled to the level of μ at rate κ (HULL, 2009, p. 731). As κ multiplies $(\mu - r_t)$, the gap between μ and r_t is "closed" at the rate $\hat{\kappa} = 0.087$ per interval Δt. The parameter μ is the level to where r_t tends to in the long term. This means in our estimated model r_t tends to 0.036 if we use the posterior median of μ as its estimate. The volatility σ is the standard deviation of the random shocks ε_t, times $\sqrt{\Delta t}$.

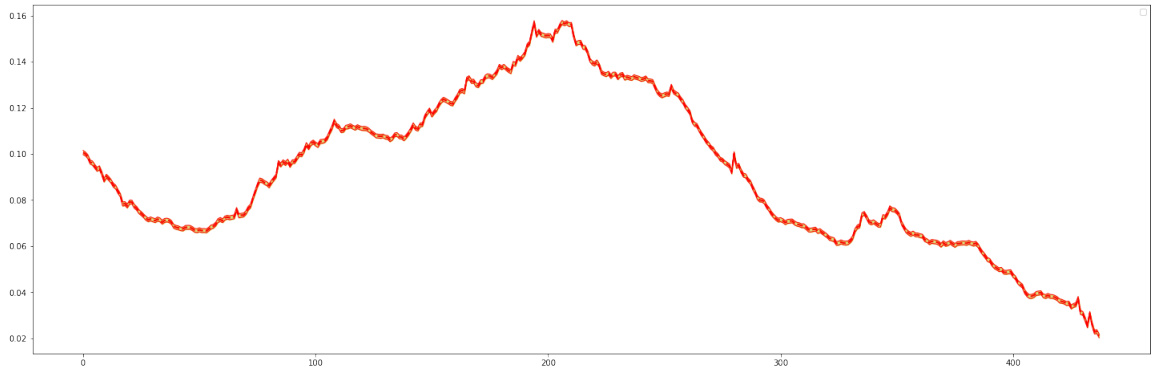
The observation equation parameters a and b determine the relationship between the short rate and the yield curve. The decay of shorter maturity rates away from the longer maturity

rates and towards the short rate is controlled by b. A higher value of b means the decay will be slower, and therefore a larger distance between shorter and longer spot rates. For a, as it only appears in the term $A(b,a,\sigma,\tau)$, it controls the level of the longer maturity rates. A higher value of a means the spot rates will be higher as τ grows. The parameter σ also affects the level of longer maturity spot rates, as a higher volatility implies longer rates will be higher.

Table 6 – Statistical summary for the Vasicek model - Real data

Parameter	Median	SD	*LB*	*UB*
a	0.138	<0.001	0.137	0.138
b	0.179	0.001	0.177	0.181
κ	0.087	0.053	0.001	0.183
μ	0.036	0.038	0.000	0.112
σ	0.012	<0.001	0.011	0.013

Figure 12 – Estimated short rate and 95% highest posterior density interval for the Vasicek model - Real data



Source: Elaborated by the author.

5.2 CIR model

For the CIR model, we estimated the model described in Section 3.2, with the same prior distributions and hyperparameter values, using the yield curve data. The observation noise σ_{obs} was fixed at 0.001. Table 7 presents the statistical summary for the CIR model. As with the Vasicek model, no major convergence issues were apparent through the MCMC diagnostics or the diagrams in Appendix B. The estimated short rate looks very much like the one for the Vasicek model presented in Figure 12, and so we omit the graph for the CIR short rate. The empirical interpretations for the CIR parameters are similar to the ones of the Vasicek model described in the previous section. The main difference is that the short rate volatility is affected by the square root of the level of the short rate, $\sqrt{r_t}$. Therefore, σ has a slightly different interpretation as in the Vasicek model.

Table 7 – Statistical summary for the CIR model - Real data

Parameter	Median	SD	*LB*	*UB*
a	0.138	<0.001	0.138	0.139
b	0.177	0.001	0.175	0.179
κ	0.110	0.063	0.000	0.220
μ	0.029	0.033	0.000	0.090
σ	0.042	0.002	0.039	0.045

5.3 DNS model

The three versions of the dynamic Nelson-Siegel model specified in Section 3.3 were estimated, with the aforementioned prior distribution and hyperparameter values. The parameter σ_{obs} was also fixed at 0.001.

Tables 8, 9 and 10 show statistical summaries for the three specifications of the DNS model. All three versions of the DNS model present good results in terms of autocorrelations and convergence, as diagnostics in Appendix B show. The latent factors for the calibrated and free λ models look very much the same, so only the factors for the free λ model are presented in Figure 13. For the dynamic λ model, the factors appear somewhat different, especially the curvature factor. The HPD bands are also wider for some time periods, as can be seen in Figure 14.

As the three factors have the same parametrization as the Vasicek short rate, the interpretation of the parameters is also similar. The decay parameter λ controls how the three factors are combined in order to obtain the yield curve. A higher λ means the slope factor will have a lesser effect on the longer rates, and the curvature loading will attain its maximum at shorter maturities.

Table 8 – Statistical summary for the DNS model (calibrated λ) - Real data

Parameter	Median	SD	*LB*	*UB*
κ_l	0.414	0.333	0.001	1.100
μ_l	0.110	0.027	0.046	0.163
σ_l	0.022	0.001	0.020	0.024
κ_s	0.372	0.227	0.002	0.792
μ_s	-0.035	0.020	-0.085	0.001
σ_s	0.023	0.001	0.022	0.025
κ_c	0.314	0.199	0.000	0.699
μ_c	-0.040	0.029	-0.100	0.019
σ_c	0.036	0.002	0.032	0.041

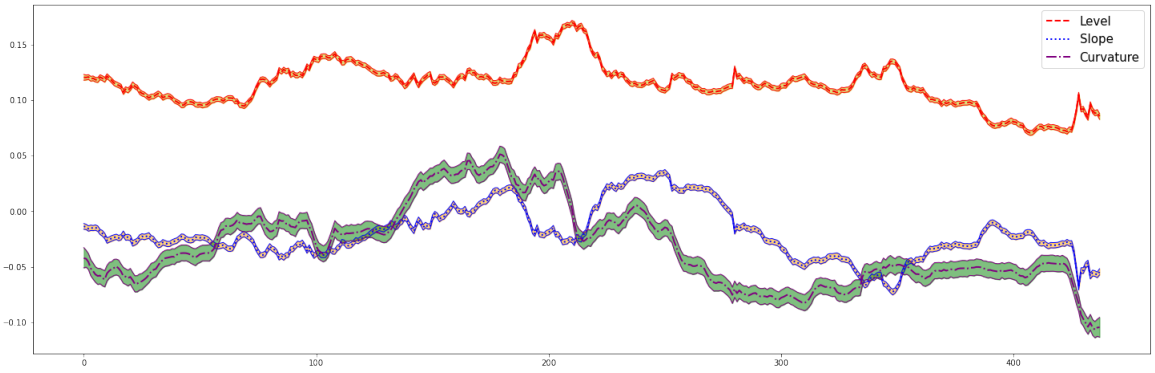
5.4 Comparison of forecasts for all models in rolling window framework

Our main goal for the model estimation with real term structure data is to evaluate the forecasting ability of each model and perform comparisons using a statistical error measure. Our

Table 9 – Statistical summary for the DNS model (free λ) - Real data

Parameter	Median	SD	*LB*	*UB*
λ	1.396	0.018	1.361	1.433
κ_l	0.424	0.316	0.000	1.051
μ_l	0.108	0.027	0.056	0.172
σ_l	0.022	0.001	0.021	0.024
κ_s	0.412	0.250	0.001	0.886
μ_s	-0.032	0.020	-0.079	0.000
σ_s	0.024	0.001	0.022	0.026
κ_c	0.274	0.190	0.001	0.645
μ_c	-0.042	0.031	-0.102	0.021
σ_c	0.033	0.002	0.029	0.037

Figure 13 – Estimated latent factors and 95% highest posterior density intervals for the DNS model (free λ) - Real data



Source: Elaborated by the author.

choice of measure is the mean absolute error (MAE), which is based on an absolute loss function. We calculate values for the MAE for the medians of predictions in a rolling window framework, described in the next paragraphs, and present them in Table 11.

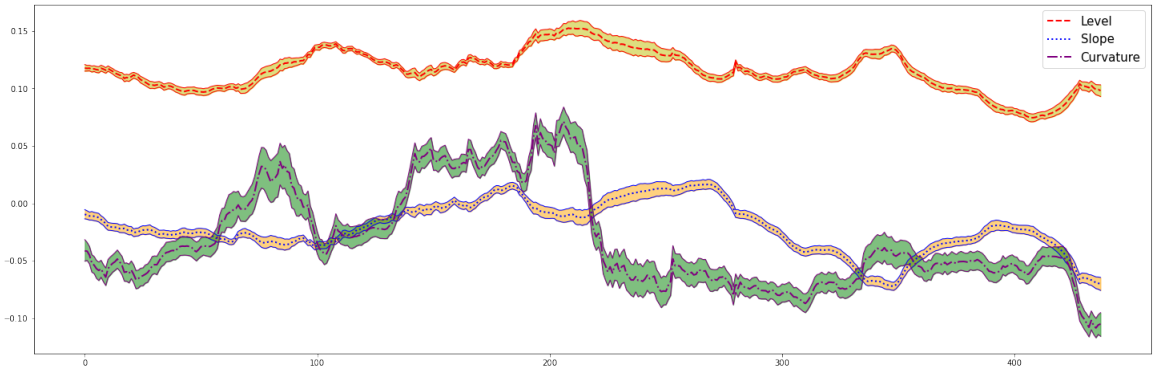
The rolling window framework employed in this section consists of estimating the models with observations ranging from k to $K = k + 400$, with k varying from 1 to 30. Each estimate produces draws $d = 1, \ldots, 2000$ for the parameters, which we use to obtain a total of 2000 predicted values for the spot rates at each maturity and horizon. We then proceed to compute the medians of the predicted spot rates across draws and compare those medians with the observed data points for each given horizon and maturity, using a statistical error measure. This is done for horizons $h = 1, \ldots, 8$, corresponding to the periods immediately following those used in estimation.

We first obtain predictions at each step of the MCMC algorithm, for all 30 windows, by using sampled values for each parameter to obtain predictions for the spot rates. We denote the predicted values for window (k, K), horizon h, maturity τ and draw d as $\hat{y}^{(\tau)}_{(K+h),d}$, which are obtained for each of the models. For example, for the Vasicek model the predicted value for

Table 10 – Statistical summary for the DNS model (time-varying λ) - Real data

Parameter	Median	SD	*LB*	*UB*
σ_λ	0.137	0.009	0.122	0.155
κ_l	0.184	0.173	0.000	0.550
μ_l	0.108	0.028	0.050	0.170
σ_l	0.013	0.001	0.012	0.015
κ_s	0.154	0.111	0.001	0.378
μ_s	-0.050	0.022	-0.100	-0.011
σ_s	0.013	0.001	0.012	0.015
κ_c	0.388	0.222	0.003	0.806
μ_c	-0.040	0.028	-0.098	0.016
σ_c	0.044	0.003	0.039	0.050

Figure 14 – Estimated latent factors and 95% highest posterior density intervals for the DNS model (time-varying λ) - Real data



Source: Elaborated by the author.

$\hat{y}^{(\tau)}_{(K+h),d}$ is obtained from

$$\begin{aligned}
\hat{r}_{(K+h),d} =& \hat{r}_{(K+h-1),d} + \hat{\kappa}_d(\hat{\mu}_d - \hat{r}_{(K+h-1),d})\Delta t + \hat{\sigma}_d \varepsilon_{1,t}\sqrt{\Delta t} \\
\hat{y}^{(\tau)}_{(K+h),d} =& -\frac{A(\hat{b}_d, \hat{a}_d, \hat{\sigma}_d, \tau) - B(\hat{b}_d, \tau)\hat{r}_{(K+h),d}}{\tau} + \sigma_{obs}\varepsilon_{2,t}, \\
\varepsilon_{1,t}, \varepsilon_{2,t} \sim& N(0,1),
\end{aligned}$$

where $\hat{\theta}_d$ denotes draw d for the parameter θ (which could be κ, μ etc.) estimated from the data in the window (k, K). We then proceed to calculate the medians for $\hat{y}^{(\tau)}_{(K+h),d}$ from the draws, denoting them by $\hat{y}^{(\tau)}_{med,(K+h)}$, and use them as our predictions for each window, horizon and maturity.

The next step is to compare the predictions to the observed values for each forecasting window, horizon and maturity. Our statistical error measure for this is the mean absolute error, or MAE. The estimate for the MAE is defined as

$$\widehat{MAE}^{(\tau)}_h = \frac{\sum_{K=401}^{430} |\hat{y}^{(\tau)}_{med,(K+h)} - y^{(\tau)}_{(K+h)}|}{30},$$

where $y^{(\tau)}_{(K+h)}$ is the observed value for horizon $h = 1,\ldots,8$. We shall have the mean of the forecasting error measures of the windows for all horizons, maturities and models.

The rolling window framework is a benchmark for time series forecasting, and its choice as our main measure of predictive ability seems justified. Schnaubelt (2019) finds that standard cross-validation methods can produce a large bias for time series models, and that forward-validation schemes, such as rolling window forecasts, are better suited for time series data.

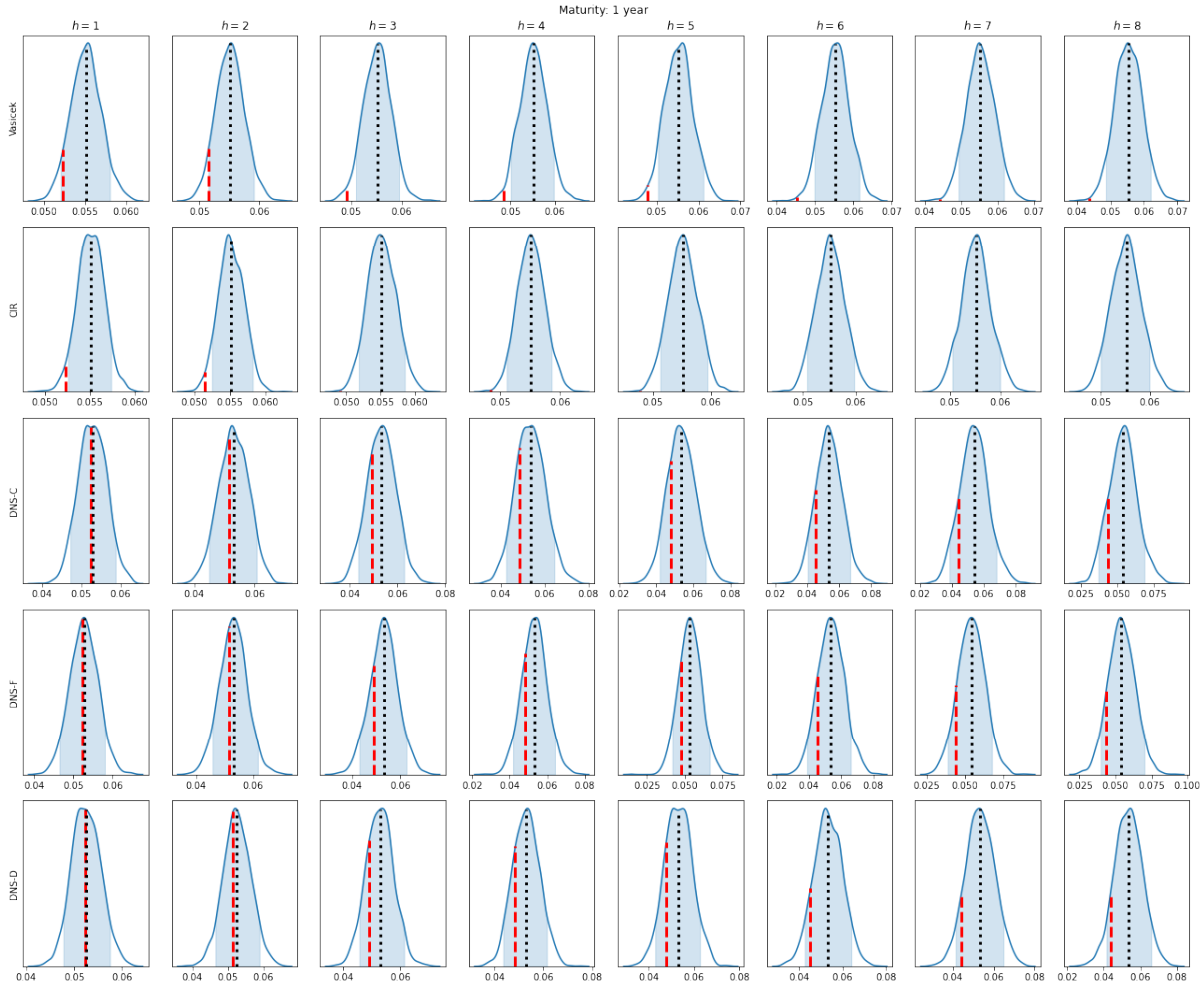
Table 11 presents the estimated mean absolute errors for all models across different maturities and horizons, with the lowest error for a given maturity and horizon highlighted in boldface. The results show that the DNS model with dynamic decay parameter (DNS-D) generally provides the better forecasts. The Vasicek and CIR models present the best forecasts for the shorter maturities (2, 3 and 6 months) with longer horizons. For the longer maturities (1, 3 and 5 years) the DNS models provide clearly better forecasts, with the DNS-D being the best and the other two coming very close. We also calculated alternative measures, based on squared and logarithmic deviations, which are presented in Appendix C. The fact that the DNS models are generally able to deliver better forecasts seems in accordance with Litterman and Scheinkman (1991), which states that three factors are able to explain most of the yield curve changes. Therefore, our conclusions reinforce the strength of these results, although with some caveats.

Figure 15 presents an example of the KDEs for the posterior distribution of the spot rates, in this case for the 1 year maturity spot rates obtained for the first rolling window. We present the posterior medians (dotted black line), 95% HPD intervals (shaded area) and real observed values (dashed red line). It is important to keep in mind that this is just one of the 30 rolling window forecasts, so the differences between posterior medians and real values shown in the figure might not be perfectly representative of the error measures in Table 11.

Table 11 – Comparison of mean absolute errors for the five models across different forecasting horizons and maturities - Rolling window forecast

Maturity	Model	Horizon							
		1	2	3	4	5	6	7	8
2 months	Vasicek	0.0047	0.0041	0.0036	0.0032	**0.0030**	**0.0030**	**0.0030**	**0.0032**
	CIR	0.0047	0.0041	0.0037	0.0033	**0.0030**	**0.0030**	**0.0030**	**0.0032**
	DNS-C	**0.0008**	0.0018	0.0027	0.0034	0.0044	0.0053	0.0062	0.0073
	DNS-F	0.0010	0.0019	0.0027	0.0035	0.0044	0.0055	0.0065	0.0075
	DNS-D	0.0009	**0.0016**	**0.0024**	**0.0031**	0.0039	0.0048	0.0056	0.0065
3 months	Vasicek	0.0036	0.0032	0.0029	**0.0026**	**0.0025**	**0.0027**	**0.0030**	0.0034
	CIR	0.0036	0.0031	0.0029	**0.0026**	0.0026	0.0028	**0.0030**	**0.0033**
	DNS-C	0.0009	0.0019	0.0028	0.0035	0.0045	0.0054	0.0063	0.0074
	DNS-F	0.0009	0.0018	0.0026	0.0034	0.0043	0.0053	0.0063	0.0073
	DNS-D	**0.0008**	**0.0016**	**0.0024**	0.0031	0.0039	0.0047	0.0056	0.0064
6 months	Vasicek	0.0014	**0.0016**	**0.0021**	**0.0025**	0.0031	0.0037	0.0043	0.0049
	CIR	0.0014	**0.0016**	**0.0021**	**0.0025**	**0.0030**	**0.0036**	**0.0041**	**0.0048**
	DNS-C	0.0011	0.0021	0.0030	0.0037	0.0046	0.0055	0.0063	0.0073
	DNS-F	0.0009	0.0017	0.0025	0.0033	0.0042	0.0051	0.0061	0.0070
	DNS-D	**0.0008**	**0.0016**	0.0024	0.0031	0.0039	0.0047	0.0055	0.0062
1 year	Vasicek	0.0028	0.0035	0.0042	0.0047	0.0053	0.0060	0.0065	0.0072
	CIR	0.0028	0.0034	0.0041	0.0046	0.0052	0.0058	0.0063	0.0070
	DNS-C	0.0014	0.0024	0.0033	0.0039	0.0047	0.0055	0.0062	0.0070
	DNS-F	0.0013	0.0022	0.0031	0.0038	0.0046	0.0054	0.0061	0.0069
	DNS-D	**0.0012**	**0.0021**	**0.0030**	**0.0036**	**0.0043**	**0.0050**	**0.0056**	**0.0062**
3 years	Vasicek	0.0055	0.0057	0.0061	0.0064	0.0063	0.0065	0.0066	0.0068
	CIR	0.0054	0.0056	0.0061	0.0063	0.0062	0.0064	0.0065	0.0066
	DNS-C	**0.0025**	**0.0035**	**0.0043**	**0.0048**	**0.0051**	**0.0051**	0.0055	0.0056
	DNS-F	0.0026	0.0037	0.0044	0.0049	0.0052	0.0053	0.0057	0.0059
	DNS-D	0.0027	**0.0035**	**0.0043**	0.0050	**0.0051**	**0.0051**	**0.0054**	**0.0055**
5 years	Vasicek	0.0093	0.0092	0.0095	0.0097	0.0094	0.0093	0.0091	0.0089
	CIR	0.0093	0.0092	0.0094	0.0097	0.0094	0.0092	0.0090	0.0088
	DNS-C	0.0028	0.0037	0.0047	0.0053	0.0054	0.0057	0.0057	0.0059
	DNS-F	**0.0027**	0.0035	0.0046	0.0053	0.0054	0.0058	0.0057	0.0059
	DNS-D	**0.0027**	**0.0034**	**0.0044**	**0.0052**	**0.0052**	**0.0056**	**0.0056**	**0.0056**

Figure 15 – Kernel density estimates of the posterior distributions for the 1 year spot rates across all models and horizons for the first rolling window



Source: Elaborated by the author.

CHAPTER

6

DISCUSSION AND FURTHER DIRECTIONS

In this chapter we briefly discuss the results presented in this work, together with some interpretations. We also list suggestions of possible directions for future research, taking into account our literature review in Chapter 2 and open research problems. Our suggestions focus on the interdisciplinary nature of the problem, taking into account novel research both in statistics and financial economics. All computational procedures utilized throughout this work may be made available and consulted upon request.

6.1 Discussion of the results

The main goal of this work was to estimate and compare the predictions of different models for the term structure of interest rates. We studied three different models: the affine models by Vasicek (1977) and Cox, Ingersoll and Ross (1985), and the dynamic Nelson-Siegel model of Diebold and Li (2006). For the last one, we provided three different specifications regarding the decay parameter λ. We used the Hamiltonian Monte Carlo algorithm for computational Bayesian inference, first with simulated data and then with real term structure data. The main goal of the simulated data section is to check how well the models and algorithms were able to estimate the latent variables. The estimates were compared with the true simulated values for the state vector that were also used for calculating the synthetic yield curve data used in estimation. As pointed out by Gelman *et al.* (2020), working in an environment where the true data to be estimated are known can be very useful for the assessment of statistical models. All of the five analyzed models were able to successfully estimate the true state vectors. We also present convergence diagnostics in Appendix B for the posterior chains, which indicate no major convergence issues were observed throughout most of the results. We then proceeded to estimate the model for the real yield curve data from the Brazilian market, ranging from 2012 to 2020.

With the estimates from real data, we evaluate model forecasting abilities in a rolling window framework with the mean absolute error as our benchmark measure. According to this

measure, the DNS model with time-varying decay parameter λ has provided the best forecasts for most of the horizons and maturities, although the affine models (Vasicek and CIR) performed better for shorter maturities with longer hozirons. For the longer maturities, the DNS models were noticeably better, with the time-varing λ DNS being the best and the other two being very close. The calibrated λ DNS provided the best forecasts for some maturities. Between the two affine models, the results were very close, with the Vasicek or the CIR delivering the better forecasts in different occasions. It is also worth mentioning that the data did not present short rates near zero, one of the main possible shortcomings of the Vasicek model that the CIR model seeks to solve.

Our results seem to justify the reputation of the dynamic Nelson-Siegel model from Diebold and Li (2006) as the main workhorse in term structure modelling. It is worth pointing out that treating the decay parameter λ as a time-varying process results in improvements from the original approach most of the time.

Our results show some support for the results from Litterman and Scheinkman (1991) that three factors provide the best description of the yield curve. We also conclude that working in a multi-model framework may prove beneficial for forecasting, as some models can perform better than others in different maturities or horizons. Tools such as Bayesian model averaging, for example, might also prove to be useful. The general framework for affine models presented in Duffie and Kan (1996) allows for multifactor models, so it may be possible to specify affine models that provide better forecasting results.

6.2 Further direction for research

6.2.1 Bayesian methods for time series forecasting

Directions for future research might include prior sensitivity analysis for the model parameters, especially for λ as a free parameter for the DNS model. Different specifications for the λ_t process in the dynamic scenario might also be the subject of future research. In particular, the prior predictive checking procedure by Gabry *et al.* (2019) can be an important part of the Bayesian workflow, as suggested by Gelman *et al.* (2020).

Bridging the breakthrough work across research areas might be an excellent further step for related research. For example, comparisons of the standard Bayesian econometric methods such as the Kalman filter, simulation smoothers (KOOP, 2003) and FFBS algorithms (CARTER; KOHN, 1994) with the HMC algorithm presented in this work for estimation of state space models. Novel methods in machine learning, such as Gaussian processes, are certainly of great interest either.

Bayesian state space modelling techniques may also be employed in estimating macroeconomic models that have the term spread, or even the entire yield curve, as an input. As explained

in Section 2.1.6, interactions between the term structure and macroeconomic policy making can go both ways. Future research may employ the Bayesian modelling approaches presented in this work for estimation of macro-finance models, and possibly as part of dynamic stochastic general equilibrium (DSGE) models.

6.2.2 Financial and macroeconomic theory

A stronger interaction between theory and empirical models may also be pursued with regards to financial theory and time series forecasting. In particular, the assumption of the no-arbitrage restrictions of Christensen, Diebold and Rudebusch (2011) for the dynamic Nelson-Siegel model and its possible effects on forecasting ability present one possible path for research. A more thorough study of the effect of such type of restriction on financial time series forecasting can prove to be fruitful in terms of research.

Although mentioned in Section 2.1.5, we did not explore any models that include macroeconomic variables, such as the inflation rate or GDP growth, in forecasting. Macro-finance models that decompose the term structure in macroeconomic factors resulting from monetary policy, e.g., Rudebusch and Wu (2003), can be estimated and compared with the tools described in this work.

The endogenous relationship between the yield curve and both monetary and fiscal policies may also be object of further investigation. Works that relate fiscal (DAI; PHILIPPON, 2005) or monetary policy variables (GALLMEYER; HOLLIFIELD; ZIN, 2005) to the term structure in a vector autoregression context already provide valuable contributions. Research such as Huse (2011) and Morales (2010) combine the DNS model from Diebold and Li (2006) with observed macroeconomic variables. Combining the knowledge of the ever-growing macro-finance theory with the cutting edge advances in statistical time series methods is a path to valuable research.

BIBLIOGRAPHY

ALFONSI, A. Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process. **Statistics & Probability Letters**, Elsevier, v. 83, n. 2, p. 602–607, 2013. Citation on page 41.

ANG, A.; DONG, S.; PIAZZESI, M. **No-arbitrage Taylor rules**. [S.l.], 2007. Citation on page 34.

ANG, A.; PIAZZESI, M. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. **Journal of Monetary economics**, Elsevier, v. 50, n. 4, p. 745–787, 2003. Citation on page 30.

BACHELIER, L. Théorie de la spéculation. In: **Annales scientifiques de l'École normale supérieure**. [S.l.: s.n.], 1900. v. 17, p. 21–86. Citation on page 76.

BALDUZZI, P.; DAS, S. R.; FORESI, S.; SUNDARAM, R. K. A simple approach to three-factor affine term structure models. **The Journal of Fixed Income**, v. 6, n. 3, p. 43–53, 1996. Citation on page 28.

BAUER, M. D.; RUDEBUSCH, G. D. Resolving the spanning puzzle in macro-finance term structure models. **Review of Finance**, Oxford University Press, v. 21, n. 2, p. 511–553, 2017. Citation on page 30.

BAYARRI, M. J.; BERGER, J. O. The interplay of Bayesian and frequentist analysis. **Statistical Science**, Institute of Mathematical Statistics, v. 19, n. 1, p. 58–80, 2004. Citation on page 32.

BETANCOURT, M. A conceptual introduction to Hamiltonian Monte Carlo. **arXiv preprint arXiv:1701.02434**, 2017. Citation on page 33.

BINSBERGEN, J. H. V.; FERNÁNDEZ-VILLAVERDE, J.; KOIJEN, R. S.; RUBIO-RAMÍREZ, J. The term structure of interest rates in a DSGE model with recursive preferences. **Journal of Monetary Economics**, Elsevier, v. 59, n. 7, p. 634–648, 2012. Citation on page 31.

BLACK, F.; LITTERMAN, R. Asset allocation: combining investor views with market equilibrium. **Goldman Sachs Fixed Income Research**, Goldman Sachs & Co., v. 115, 1990. Citation on page 32.

BLACK, F.; SCHOLES, M. The Pricing of Options and Corporate Liabilities. **Journal of Political Economy**, v. 81, n. 3, p. 637–654, May 1973. ISSN 0022-3808, 1537-534X. Available: <https://www.journals.uchicago.edu/doi/10.1086/260062>. Citation on page 76.

BRIGO, D.; MERCURIO, F. **Interest Rate Models Theory and Practice**. [S.l.]: Springer Berlin Heidelberg, 2001. (Springer Finance). ISBN 9783662045534. Citations on pages 37, 39, and 41.

BURKARDT, J. The truncated normal distribution. **Department of Scientific Computing Website, Florida State University**, v. 1, p. 35, 2014. Citation on page 77.

CALDEIRA, J. F.; LAURINI, M. P.; PORTUGAL, M. S. Bayesian inference applied to dynamic Nelson-Siegel model with stochastic volatility. **Brazilian Review of Econometrics**, v. 30, n. 1, p. 123–161, 2010. Citation on page 34.

CAMPBELL, J. Y. **Financial decisions and markets: a course in asset pricing**. 2017. ISBN: 9781400888221 OCLC: 1145891400. Citation on page 22.

CAMPBELL, J. Y.; SHILLER, R. J. Yield spreads and interest rate movements: A bird's eye view. **The Review of Economic Studies**, Wiley-Blackwell, v. 58, n. 3, p. 495–514, 1991. Citation on page 23.

CARTER, C. K.; KOHN, R. On Gibbs sampling for state space models. **Biometrika**, Oxford University Press, v. 81, n. 3, p. 541–553, 1994. Citations on pages 33 and 64.

CHAN, J. C.; JELIAZKOV, I. Efficient simulation and integrated likelihood estimation in state space models. **International Journal of Mathematical Modelling and Numerical Optimisation**, Inderscience Publishers, v. 1, n. 1-2, p. 101–120, 2009. Citation on page 33.

CHAN, K. C.; KAROLYI, G. A.; LONGSTAFF, F. A.; SANDERS, A. B. An empirical comparison of alternative models of the short-term interest rate. **The journal of finance**, Wiley Online Library, v. 47, n. 3, p. 1209–1227, 1992. Citation on page 34.

CHIB, S.; ERGASHEV, B. Analysis of multifactor affine yield curve models. **Journal of the American Statistical Association**, Taylor & Francis, v. 104, n. 488, p. 1324–1337, 2009. Citation on page 34.

CHRISTENSEN, J. H.; DIEBOLD, F. X.; RUDEBUSCH, G. D. The affine arbitrage-free class of Nelson–Siegel term structure models. **Journal of Econometrics**, Elsevier, v. 164, n. 1, p. 4–20, 2011. Citations on pages 29 and 65.

COCHRANE, J. H. **Asset pricing: Revised edition**. [S.l.]: Princeton University Press, 2009. Citations on pages 23, 24, and 75.

COCHRANE, J. H.; PIAZZESI, M. Bond risk premia. **American Economic Review**, v. 95, n. 1, p. 138–160, 2005. Citation on page 23.

CORDEIRO, G. M.; LEMONTE, A. J. The beta-half-Cauchy distribution. **Journal of Probability and Statistics**, Hindawi, v. 2011, 2011. Citation on page 77.

COX, J. C.; INGERSOLL, J. E.; ROSS, S. A. A theory of the term structure of interest rates. **Econometrica**, v. 53, n. 2, p. 385–407, 1985. Citations on pages 19, 27, 28, 40, and 63.

CRUMP, R. K.; GOSPODINOV, N. Deconstructing the yield curve. **FRB of New York Staff Report**, n. 884, 2019. Citation on page 28.

DAI, Q.; PHILIPPON, T. **Fiscal policy and the term structure of interest rates**. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 2005. Citations on pages 31 and 65.

DAS, S. Modeling Nelson–Siegel yield curve using Bayesian approach. In: **New Perspectives and Challenges in Econophysics and Sociophysics**. [S.l.]: Springer, 2019. p. 169–189. Citation on page 35.

DEGROOT, M.; SCHERVISH, M. **Probability and Statistics**. [S.l.]: Addison-Wesley, 2012. ISBN 9780321500465. Citations on pages 24 and 45.

DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. **Journal of econometrics**, Elsevier, v. 130, n. 2, p. 337–364, 2006. Citations on pages 19, 29, 34, 37, 41, 63, 64, and 65.

DIEBOLD, F. X.; RUDEBUSCH, G. D. **Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach**. [S.l.]: Princeton University Press, 2013. ISBN 9780691146805. Citations on pages 28 and 30.

DUFFEE, G. **Forecasting with the term structure: The role of no-arbitrage restrictions**. [S.l.], 2011. Citation on page 30.

DUFFIE, D.; KAN, R. A yield-factor model of interest rates. **Mathematical finance**, Wiley Online Library, v. 6, n. 4, p. 379–406, 1996. Citations on pages 26, 28, 29, 34, and 64.

ELERIAN, O.; CHIB, S.; SHEPHARD, N. Likelihood inference for discretely observed nonlinear diffusions. **Econometrica**, Wiley Online Library, v. 69, n. 4, p. 959–993, 2001. Citation on page 38.

EPSTEIN, L. G.; ZIN, S. E. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. **Journal of political Economy**, The University of Chicago Press, v. 99, n. 2, p. 263–286, 1991. Citation on page 31.

ERAKER, B. MCMC analysis of diffusion models with application to finance. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 19, n. 2, p. 177–191, 2001. Citation on page 34.

FABOZZI, F. J. **Fixed income analysis**. [S.l.]: John Wiley & Sons, 2007. Citations on pages 21 and 22.

FAMA, E. F.; BLISS, R. R. The information in long-maturity forward rates. **The American Economic Review**, p. 680–692, 1987. Citation on page 22.

FERNÁNDEZ-VILLAVERDE, J. The econometrics of DSGE models. **SERIEs: Journal of the Spanish Economic Association**, Asociación Española de Economía, v. 1, n. 1, p. 3–49, 2010. Citations on pages 31 and 32.

FROOT, K. A. New hope for the expectations hypothesis of the term structure of interest rates. **The Journal of Finance**, Wiley Online Library, v. 44, n. 2, p. 283–305, 1989. Citation on page 23.

FRÜHWIRTH-SCHNATTER, S. Data augmentation and dynamic linear models. **Journal of time series analysis**, Wiley Online Library, v. 15, n. 2, p. 183–202, 1994. Citation on page 33.

GABRY, J.; SIMPSON, D.; VEHTARI, A.; BETANCOURT, M.; GELMAN, A. Visualization in Bayesian workflow. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, v. 182, n. 2, p. 389–402, 2019. Citations on pages 37 and 64.

GALLMEYER, M. F.; HOLLIFIELD, B.; ZIN, S. E. Taylor rules, McCallum rules and the term structure of interest rates. **Journal of Monetary Economics**, Elsevier, v. 52, n. 5, p. 921–950, 2005. Citation on page 65.

GELMAN, A.; VEHTARI, A.; SIMPSON, D.; MARGOSSIAN, C. C.; CARPENTER, B.; YAO, Y.; KENNEDY, L.; GABRY, J.; BÜRKNER, P.-C.; MODRÁK, M. Bayesian workflow. **arXiv preprint arXiv:2011.01808**, 2020. Citations on pages 19, 45, 63, and 64.

GRAY, P. Bayesian estimation of short-rate models. **Australian Journal of Management**, SAGE Publications Sage UK: London, England, v. 30, n. 1, p. 1–22, 2005. Citation on page 34.

HAGAN, P. S.; WEST, G. Interpolation methods for curve construction. **Applied Mathematical Finance**, Taylor & Francis, v. 13, n. 2, p. 89–129, 2006. Citation on page 53.

HAHN, J.; LEE, H. Yield spreads as alternative risk factors for size and book-to-market. **Journal of Financial and Quantitative Analysis**, Cambridge University Press, v. 41, n. 2, p. 245–269, 2006. Citation on page 30.

HANSEN, L. P.; SINGLETON, K. J. Generalized instrumental variables estimation of nonlinear rational expectations models. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 1269–1286, 1982. Citation on page 33.

HELWIG, N. E. Adding bias to reduce variance in psychological results: A tutorial on penalized regression. **The Quantitative Methods for Psychology**, The Quantitative Methods for Psychology, v. 13, n. 1, p. 1–19, 2017. Citation on page 32.

HESTON, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. **The Review of Financial Studies**, Oxford University Press, v. 6, n. 2, p. 327–343, 1993. Citation on page 40.

HIGHAM, D. J.; MAO, X. Convergence of Monte Carlo simulations involving the mean-reverting square root process. **Journal of Computational Finance**, v. 8, n. 3, p. 35–61, 2005. Citation on page 41.

HOFFMAN, M. D.; GELMAN, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. **Journal of Machine Learning Research**, v. 15, n. 47, p. 1593–1623, 2014. Citations on pages 33 and 38.

HULL, J. **Options, futures and other derivatives**. 7th ed. ed. Upper Saddle River, NJ: Prentice Hall, 2009. ISBN 978-0-13-601586-4. Citations on pages 25, 26, 54, and 76.

HULL, J.; WHITE, A. Pricing interest-rate-derivative securities. **The review of financial studies**, Oxford University Press, v. 3, n. 4, p. 573–592, 1990. Citation on page 28.

HUSE, C. Term structure modelling with observable state variables. **Journal of Banking & Finance**, Elsevier, v. 35, n. 12, p. 3240–3252, 2011. Citation on page 65.

JACQUIER, E.; POLSON, N. G. **Asset allocation in finance: A Bayesian perspective**. [S.l.], 2012. Citation on page 31.

JAMES, J.; WEBBER, N. **Interest Rate Modelling**. [S.l.]: Wiley, 2000. (Wiley Series in Financial Engineering). ISBN 9780471975236. Citations on pages 22 and 30.

JOHANNES, M.; POLSON, N. MCMC methods for continuous-time financial econometrics. In: **Handbook of Financial Econometrics: Applications**. [S.l.]: Elsevier, 2010. p. 1–72. Citations on pages 34, 37, 39, and 40.

JOHANNES, M.; POLSON, N. *et al.* Iterative and recursive estimation in structural nonadaptive models: Comment. **Journal of Business & Economic Statistics**, American Statistical Association, v. 21, n. 4, p. 493–495, 2003. Citations on pages 38 and 45.

JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions, Volume 2**. [S.l.]: Wiley, 1995. (Wiley Series in Probability and Statistics). Citation on page 78.

JOSLIN, S.; SINGLETON, K. J.; ZHU, H. A new perspective on Gaussian dynamic term structure models. **The Review of Financial Studies**, Oxford University Press, v. 24, n. 3, p. 926–970, 2011. Citation on page 30.

KANTAS, N.; DOUCET, A.; SINGH, S. S.; MACIEJOWSKI, J.; CHOPIN, N. On particle methods for parameter estimation in state-space models. **Statistical science**, Institute of Mathematical Statistics, v. 30, n. 3, p. 328–351, 2015. Citation on page 33.

KIM, S.; SHEPHARD, N.; CHIB, S. Stochastic volatility: likelihood inference and comparison with ARCH models. **The review of economic studies**, Wiley-Blackwell, v. 65, n. 3, p. 361–393, 1998. Citation on page 33.

KOOP, G. **Bayesian Econometrics**. [S.l.]: Wiley, 2003. Citation on page 64.

KUCUKELBIR, A.; TRAN, D.; RANGANATH, R.; GELMAN, A.; BLEI, D. M. Automatic differentiation variational inference. **The Journal of Machine Learning Research**, JMLR. org, v. 18, n. 1, p. 430–474, 2017. Citations on pages 33, 34, and 38.

KUMAR, R.; CARROLL, C.; HARTIKAINEN, A.; MARTIN, O. ArviZ a unified library for exploratory analysis of Bayesian models in Python. **Journal of Open Source Software**, The Open Journal, v. 4, n. 33, p. 1143, 2019. Available: <https://doi.org/10.21105/joss.01143>. Citation on page 79.

LAURINI, M. P.; HOTTA, L. K. Bayesian extensions to Diebold-Li term structure model. **International Review of Financial Analysis**, Elsevier, v. 19, n. 5, p. 342–350, 2010. Citation on page 34.

LINDSTRÖM, E.; MADSEN, H.; NIELSEN, J. **Statistics for Finance**. [S.l.]: CRC Press, 2015. (Chapman & Hall/CRC Texts in Statistical Science). ISBN 9781482229004. Citations on pages 26, 27, 37, 39, and 41.

LITTERMAN, R.; SCHEINKMAN, J. Common factors affecting bond returns. **Journal of Fixed Income**, v. 1, n. 1, p. 54–61, 1991. Citations on pages 28, 42, 59, and 64.

MAYBANK, P.; BOJAK, I.; EVERITT, R. G. Fast approximate Bayesian inference for stable differential equation models. **arXiv preprint arXiv:1706.00689**, 2017. Citation on page 38.

MCCAUSLAND, W. J.; MILLER, S.; PELLETIER, D. Simulation smoothing for state–space models: A computational efficiency analysis. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 1, p. 199–212, 2011. Citation on page 33.

MORAL, P. D.; DOUCET, A.; SINGH, S. Forward smoothing using sequential Monte Carlo. **arXiv preprint arXiv:1012.5390**, 2010. Citation on page 33.

MORALES, M. The real yield curve and macroeconomic factors in the Chilean economy. **Applied Economics**, Taylor & Francis, v. 42, n. 27, p. 3533–3545, 2010. Citation on page 65.

NELSON, C. R.; SIEGEL, A. F. Parsimonious modeling of yield curves. **The Journal of Business**, University of Chicago Press, v. 60, n. 4, p. 473–489, 1987. Citations on pages 29 and 41.

PIAZZESI, M. Affine term structure models. In: **Handbook of financial econometrics: Tools and Techniques**. [S.l.]: Elsevier, 2010. p. 691–766. Citations on pages 25, 26, and 37.

PINDYCK, R.; RUBINFELD, D. **Microeconomics**. [S.l.]: Pearson, 2017. (Pearson series in economics). ISBN 9781292213316. Citation on page 24.

POLSON, N. G.; SCOTT, J. G. On the half-Cauchy prior for a global scale parameter. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 7, n. 4, p. 887–902, 2012. Citation on page 40.

POLSON, N. G.; SOKOLOV, V. Bayesian regularization: From Tikhonov to horseshoe. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley Online Library, v. 11, n. 4, p. e1463, 2019. Citation on page 32.

PUELZ, D. W. **Regularization in econometrics and finance**. Phd Thesis (PhD Thesis), 2018. Citation on page 32.

RENDLEMAN, R. J.; BARTTER, B. J. The pricing of options on debt securities. **Journal of Financial and Quantitative Analysis**, Cambridge University Press, v. 15, n. 1, p. 11–24, 1980. Citation on page 26.

RUDEBUSCH, G.; WU, T. A no-arbitrage model of the term structure and the macroeconomy. **manuscript, Federal Reserve Bank of San Francisco**, Citeseer, 2003. Citations on pages 31 and 65.

SALVATIER, J.; WIECKI, T. V.; FONNESBECK, C. Probabilistic programming in Python using PyMC3. **PeerJ Computer Science**, PeerJ Inc., v. 2, p. e55, 2016. Citation on page 38.

SCHMIDT, J.; KRÄMER, N.; HENNIG, P. A probabilistic state space model for joint inference from differential equations and data. **Advances in Neural Information Processing Systems**, v. 34, p. 12374–12385, 2021. Citation on page 38.

SCHMIDT, M.; FUNG, G.; ROSALES, R. Optimization methods for L1-regularization. **University of British Columbia, Technical Report TR-2009-19**, 2009. Citation on page 32.

SCHNAUBELT, M. **A comparison of machine learning model validation schemes for non-stationary time series data**. [S.l.], 2019. Citation on page 59.

SHMUELI, G. To explain or to predict? **Statistical science**, Institute of Mathematical Statistics, v. 25, n. 3, p. 289–310, 2010. Citation on page 32.

SIMS, C. Understanding non-Bayesians. **Unpublished chapter, Department of Economics, Princeton University**, 2010. Citations on pages 31 and 32.

SMETS, F.; WOUTERS, R. An estimated dynamic stochastic general equilibrium model of the euro area. **Journal of the European economic association**, Oxford University Press, v. 1, n. 5, p. 1123–1175, 2003. Citation on page 32.

SVENSSON, L. E. **Estimating and interpreting forward interest rates: Sweden 1992-1994**. [S.l.], 1994. Citation on page 34.

TAYLOR, J. B. Discretion versus policy rules in practice. In: ELSEVIER. **Carnegie-Rochester conference series on public policy**. [S.l.], 1993. v. 39, p. 195–214. Citation on page 34.

TIBSHIRANI, R. Regression shrinkage and selection via the LASSO. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996. Citation on page 32.

VASICEK, O. An equilibrium characterization of the term structure. **Journal of Financial Economics**, Elsevier, v. 5, n. 2, p. 177–188, 1977. Citations on pages 19, 26, 28, 34, 38, and 63.

VEHTARI, A.; GELMAN, A.; SIMPSON, D.; CARPENTER, B.; BÜRKNER, P.-C. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 1, p. 1–28, 2021. Citations on pages 79 and 80.

APPENDIX

A

ADDITIONAL STOCHASTIC PROCESSES AND PROBABILITY DISTRIBUTIONS

A.1 Wiener process

A Wiener process W_t, also called Brownian motion, is a stochastic process that can be seen as the continuous-time analogue of a random walk.[1] Although the Wiener process is continuous in time t, we define its properties in discrete time for a time interval Δ. Start by defining ε_t as following a standard normal distribution over discrete time. Then the stochastic process Z_t has independent and Gaussian increments if

$$Z_t - Z_{t-1} = \varepsilon_t,$$

where the variance of ε_t scales over time. Then, for a time interval Δ

$$Z_{t+\Delta} - Z_t \sim N(0, \Delta).$$

The notation dZ_t is used to represent $Z_{t+\Delta} - Z_t$ when Δ is arbitrarily small. As Z_t is not differentiable, dZ_t has a different interpretation in stochastic calculus than in standard calculus. As this is not the focus of this work, the relevant part is that dZ_t has the properties

$$\begin{aligned} \mathrm{E}(dZ_t) &= 0, \\ \mathrm{E}(dZ_t^2) &= dZ_t^2 = dt, \end{aligned}$$

where dt is the time difference for an arbitrarily small time interval (COCHRANE, 2009).

[1] Strictly speaking, Brownian motion refers to the physical movement of a particle which can be modelled mathematically through a Wiener process. It is, however, very common to use the name "Brownian motion" for the mathematical process itself, be it the discrete-time (random walk) or continuous-time (Wiener process) version.

A variable W_t follows a Wiener process if it has independent and Gaussian increments, which is equivalent to saying it follows two properties (HULL, 2009):

Property 1. The change ΔW_t during a time interval Δt is

$$\Delta W_t = \varepsilon_t \sqrt{\Delta t},$$

where ε_t follows the standard Normal distribution.

Property 2. The values of ΔW_t for two different intervals Δt are independent.

It is worth observing that W_t is both a Markovian process and a martingale, and also that

$$\begin{aligned} \mathrm{E}[W_t] &= 0, \\ \mathrm{Var}(W_t) &= \Delta t, \\ \mathrm{SD}(W_t) = \sqrt{\mathrm{Var}(W_t)} &= \sqrt{\Delta t}. \end{aligned}$$

The first mathematical formulation of the Brownian motion, as well as its first use in asset pricing, dates back to Bachelier (1900). The famous option pricing model by Black and Scholes (1973) assumes the price of the underlying asset follows a geometric Brownian motion. Continuous-time formulations of term structure models may also rely on the Wiener process in order to model random changes in variables.

The Wiener process is also part of the more elaborate diffusion processes. A diffusion process is defined as a stochastic process which solves a stochastic differential equation. For example, we can say the short rate r_t is a diffusion process under the Vasicek model because r_t satisfies

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t, \tag{A.1}$$

where equation (A.1) is a stochastic differential equation. More precisely, the general form for stochastic differential equations is

$$dx_t = \mu(t)dt + \sigma(t)dW_t,$$

where W_t is a Wiener process and the process x_t that solves the left-hand side is, therefore, a diffusion process.

A.2 Additional probability distributions

A.2.1 Truncated normal distribution

First, define the normal distribution with probability density function

$$\phi(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\},$$

with $\{\mu, x\} \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, and cumulative distribution function denoted by $\Phi(x; \mu, \sigma^2)$. Thus, we can obtain the truncated normal distribution by setting a truncation interval (a, b), assigning zero probability density to all values outside the interval and scaling the values so that the density integrates to 1 (BURKARDT, 2014).

More specifically, a random variable is said to follow a truncated normal distribution, or a normal distribution with mean $\overline{\mu}$ and variance $\overline{\sigma^2}$ truncated between a and b, if it has the probability density function

$$p(x; \overline{\mu}, \overline{\sigma}, a, b) = \begin{cases} 0, & \text{if } x \leq a \\ \frac{\phi(x; \overline{\mu}, \overline{\sigma}^2)}{\Phi(b; \overline{\mu}, \overline{\sigma}^2) - \Phi(a; \mu), \overline{\sigma}^2)}, & \text{if } a < x < b \\ 0, & \text{if } x \geq b \end{cases},$$

with $\phi(x; \mu, \sigma^2)$ denoting the PDF and $\Phi(x; \mu, \sigma^2)$ denoting the CDF of a (non-truncated) normal distribution with mean μ and variance σ^2. Note that $\overline{\mu}$ and $\overline{\sigma}^2$ do not refer to the mean and variance of the truncated normal distribution, but of the "parent" normal distribution (BURKARDT, 2014).

We now present expressions for the actual mean and variance of the truncated normal distribution, μ_{TN} and σ^2_{TN}. First, start by defining

$$\alpha = \frac{a - \overline{\mu}}{\overline{\sigma}},$$
$$\beta = \frac{b - \overline{\mu}}{\overline{\sigma}}.$$

Then, the mean and variance of the truncated normal can be written as

$$\mu_{TN} = \overline{\mu} - \overline{\sigma} \frac{\phi(\beta; 0, 1) - \phi(\alpha; 0, 1)}{\Phi(\beta; 0, 1) - \Phi(\alpha; 0, 1)},$$
$$\sigma^2_{TN} = \overline{\sigma}^2 \left[1 - \frac{\beta\phi(\beta; 0, 1) - \alpha\phi(\alpha; 0, 1)}{\Phi(\beta; 0, 1) - \Phi(\alpha; 0, 1)} - \left(\frac{\phi(\beta; 0, 1) - \phi(\alpha; 0, 1)}{\Phi(\beta; 0, 1) - \Phi(\alpha; 0, 1)} \right)^2 \right].$$

A.2.2 Half-Cauchy distribution

A random variable is said to follow a half-Cauchy distribution with parameter $\beta > 0$ if it has the probability density function

$$p(x; \beta) = \begin{cases} \frac{2}{\pi\beta[1 + (\frac{x}{\beta})^2]}, & \text{if } x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The half-Cauchy distribution has both undefined mean and variance. As the Cauchy distribution is symmetric around its location parameter, in this case taken to be zero, the half-Cauchy is obtained by truncating the Cauchy distribution on its left side (CORDEIRO; LEMONTE, 2011).

A.2.3 Noncentral chi-squared distribution

A random variable is said to follow a noncentral chi-squared distribution with degrees of freedom k and noncentrality parameter λ if it has the probability density function

$$p(x;k,\lambda) = \exp\{-(\lambda+x)/2\}\frac{1}{2}\left(\frac{x}{\lambda}\right)^{(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda x}),$$
$$I_a(y) = \left(\frac{1}{2}y\right)^a \sum_{j=1}^{\infty} \frac{(y^2/4)^j}{j!\Gamma(a+j+1)},$$

with support $\{x,k,\lambda\} \in \mathbb{R}_+$. The noncentral chi-squared distribution has mean $k+\lambda$ and variance $2(k+2\lambda)$ (JOHNSON; KOTZ; BALAKRISHNAN, 1995).

An additional way of characterizing the noncentral chi-squared distribution is by defining

$$Y = \sum_{i=1}^{K} X_i^2$$

as the sum of squared K independent Gaussian random variables with means μ_i and unit variance. Then Y shall follow a noncentral chi-squared distribution, with parameters K and $\lambda = \sum_{i=1}^{K} \mu_i^2$.

APPENDIX

B

MCMC CONVERGENCE DIAGNOSTICS AND ADDITIONAL PLOTS

Here we describe some MCMC convergence diagnostics for the models, and also present additional diagrams such as trace plots, kernel density estimates and autocorrelation diagrams for models in Chapters 4 and 5. The main reference for this section is Vehtari *et al.* (2021), and all of the statistics are computed via the ArviZ package for Python (KUMAR *et al.*, 2019).

B.1 Monte Carlo standard error

It is possible to evaluate the precision of the average $\overline{\theta}$ of S independent draws as an estimate of $\mathrm{E}[\theta|y]$ through the Monte Carlo standard error (MCSE). The MCSE can be computed as

$$MCSE = \sqrt{\mathrm{Var}(\overline{\theta})} = \sqrt{\frac{\mathrm{Var}(\theta|y)}{S}},$$

and this can be generalized to the posterior expectation of any function $g(\theta)$. In this work the MCSEs are computed from the effective sample sizes through the procedure described in Vehtari *et al.* (2021). We report the mean and standard deviation estimates for the MCSE.

B.2 Effective sample size

The effective sample size (ESS) is a measure of information contained in each sampling chain. A higher sampling autocorrelation means higher sampling uncertainty and therefore

smaller ESS. The ESS can be estimated as

$$\widehat{ESS} = \frac{S}{\hat{\tau}},$$
$$\hat{\tau} = -1 + 2\sum_{t'=0}^{K} \hat{P}_{t'},$$
$$\hat{P}_{t'} = \hat{\rho}_{2t'} + \hat{\rho}_{2t'+1},$$

where S is the number of samples and $\hat{\rho}_t$ are autocorrelations estimated at lag t via fast Fourier transform. K is the last integer for which $\hat{P}_K = \hat{\rho}_{2K} + \hat{\rho}_{2K+1}$ is positive.

Besides the mean and standard deviation estimates of ESS, Vehtari *et al.* (2021) propose two other versions: Bulk-ESS and Tail-ESS. Bulk-ESS is obtained by rank normalization of the draws in the chain, a procedure that consists of replacing drawn parameter values with rank normalized values. The rank normalized values are normal scores for pooled draws from all chains. Tail-ESS is the minimum of the effective sample sizes of the 5% and 95% quantiles. While Bulk-ESS is useful for assessing problems due to trend behavior of chains, Tail-ESS allows for diagnosing issues related to the scale of chains.

B.3 MCMC diagnostics for simulated and real data

Tables 12 to 16 present the MCMC diagnostics for each of the five models estimated with simulated data, in Chapter 4. For the first three models, we observe a large ESS for all parameters and small values for MCSE. This behavior is also observed for most of the parameters in the last two models, except for the free λ parameter and the standard deviation of the random walk λ_t, denoted by σ_λ. The MCSE's are still small, although a small bulk-ESS might indicate minor trend problems with the chains.

Tables 17 to 21 present the MCMC diagnostics for the five models estimated with real data, in Chapter 5. Overall we see the same patterns as in the simulated data. For the σ_λ parameter of the time-varying λ DNS model, estimation improves quite a lot, as we can see in the ESS counts or MCSE.

Table 12 – MCMC convergence for the Vasicek model - Simulated data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
a	<0.001	<0.001	1189	1189	1189	1233
b	<0.001	<0.001	1122	1122	1118	1227
κ	0.008	0.006	949	949	795	631
μ	0.001	0.001	630	630	750	566
σ	<0.001	<0.001	2116	2109	2128	1685

Table 13 – MCMC convergence for the CIR model - Simulated data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
a	<0.001	<0.001	1011	1011	1023	1124
b	<0.001	<0.001	966	966	969	1353
κ	0.010	0.007	663	663	611	588
μ	0.002	0.001	412	412	333	179
σ	<0.001	<0.001	1542	1531	1551	1263

Table 14 – MCMC convergence for the DNS model (calibrated λ) - Simulated data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
κ_l	0.011	0.008	777	777	701	998
μ_l	0.001	<0.001	1181	1172	1328	936
σ_l	<0.001	<0.001	2321	2321	2316	1701
κ_s	0.006	0.004	1119	1119	883	654
μ_s	<0.001	<0.001	2330	1785	2391	1081
σ_s	<0.001	<0.001	2098	2091	2109	1525
κ_c	0.010	0.007	885	885	832	638
μ_c	0.001	<0.001	1379	996	1546	1020
σ_c	<0.001	<0.001	542	539	550	900

Table 15 – MCMC convergence for the DNS model (free λ) - Simulated data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
λ	0.003	0.002	96	96	97	111
κ_l	0.012	0.008	833	833	725	645
μ_l	0.001	0.001	863	863	866	737
σ_l	<0.001	<0.001	2184	2184	2171	1590
κ_s	0.005	0.004	1415	1415	1102	714
μ_s	<0.001	<0.001	2758	2282	2779	1529
σ_s	<0.001	<0.001	2328	2328	2311	1586
κ_c	0.009	0.007	1009	1009	952	827
μ_c	0.001	<0.001	1535	967	1801	1207
σ_c	<0.001	<0.001	552	552	553	1056

B.4 Simulated data and estimated state vectors - Chapter 4

As the figures for the Vasicek simulated data and estimated short rate are already presented in Chapter 4, we proceed to analyze the remaining figures. The simulated data for the CIR model presented in Figure 16 looks overall very similar to the Vasicek simulated data presented in Figure 6 of Chapter 4.

For the DNS models, as the decay parameter λ does not enter the equation for the latent factors, the simulated factors for the free and time-varying λ versions of the DNS (Figures 17 and 19) are the same as the ones for the calibrated λ DNS presented in Figure 8 of Chapter 4.

Table 16 – MCMC convergence for the DNS model (time-varying λ) - Simulated data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
σ_λ	0.001	0.001	46	46	45	83
κ_l	0.011	0.008	879	879	688	567
μ_l	0.001	0.001	770	770	866	691
σ_l	<0.001	<0.001	1493	1489	1499	1385
κ_s	0.005	0.003	1686	1686	1287	997
μ_s	<0.001	<0.001	2356	1883	2400	1327
σ_s	<0.001	<0.001	2167	2167	2153	1670
κ_c	0.010	0.007	852	852	759	746
μ_c	0.001	0.001	1073	796	1142	1035
σ_c	<0.001	<0.001	299	296	303	450

Table 17 – MCMC convergence for the Vasicek model - Real data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
a	<0.001	<0.001	1164	1164	1167	1457
b	<0.001	<0.001	940	940	945	1370
κ	0.002	0.001	1247	1247	1063	788
μ	0.001	0.001	1214	1214	1033	913
σ	<0.001	<0.001	2456	2450	2444	1638

Table 18 – MCMC convergence for the CIR model - Real data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
a	<0.001	<0.001	2570	2569	2576	1618
b	<0.001	<0.001	2519	2519	2514	1464
κ	0.002	0.001	1143	1143	1023	805
μ	0.001	0.001	1057	1057	1192	1013
σ	<0.001	<0.001	3153	3102	3223	1553

The factor loadings that generate the spot yield curves are affected by different specifications of λ, but look similar overall. The spot yield curves for the free and dynamic λ DNS models, presented in Figures 18 and 21, respectively, do not look much different than the calibrated λ spot curve presented in Figure 9 of Chapter 4. The simulated λ_t for the dynamic λ DNS model is presented in Figure 20.

The estimated short rate for the CIR model, present in Figure 22, also looks similar to the Vasicek short rate presented in Figure 7 of Chapter 4. The estimated latent factors for the free λ DNS model presented in Figure 23 also look very similar to the fixed λ DNS latent factors presented in Figure 10 of Chapter 4. The factors for the dynamic λ DNS presented in Figure 24 look slightly different, with wider HPD intervals, especially for the slope and curvature factors.

Table 19 – MCMC convergence for the DNS model (calibrated λ) - Real data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
κ_l	0.013	0.009	709	709	633	888
μ_l	0.001	0.001	899	899	975	817
σ_l	<0.001	<0.001	2410	2410	2386	1819
κ_s	0.007	0.005	1001	1001	827	741
μ_s	<0.001	<0.001	2426	1719	2453	1301
σ_s	<0.001	<0.001	1991	1991	1995	1739
κ_c	0.005	0.004	1357	1357	1133	952
μ_c	0.001	0.001	2626	1715	2626	1504
σ_c	<0.001	<0.001	575	575	576	908

Table 20 – MCMC convergence for the DNS model (free λ) - Real data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
λ	0.001	0.001	153	153	153	406
κ_l	0.011	0.008	884	884	728	722
μ_l	0.001	<0.001	1476	1432	1603	1014
σ_l	<0.001	<0.001	2338	2338	2344	1737
κ_s	0.008	0.005	1046	1046	919	873
μ_s	<0.001	<0.001	2094	1234	2522	1217
σ_s	<0.001	<0.001	1985	1956	2040	1508
κ_c	0.005	0.004	1380	1380	1180	1101
μ_c	0.001	0.001	1965	1402	1996	1194
σ_c	<0.001	<0.001	696	696	690	1085

Table 21 – MCMC convergence for the DNS model (time-varying λ) - Real data

Parameter	$MCSE_\mu$	$MCSE_\sigma$	ESS_μ	ESS_σ	ESS_{bulk}	ESS_{tail}
σ_λ	<0.001	<0.001	705	705	708	1181
κ_l	0.006	0.004	785	785	654	996
μ_l	0.001	0.001	1650	1477	1696	1033
σ_l	<0.001	<0.001	1122	1122	1125	1249
κ_s	0.003	0.002	1502	1502	1117	919
μ_s	0.001	<0.001	1609	1407	1650	1363
σ_s	<0.001	<0.001	949	949	945	1425
κ_c	0.006	0.004	1300	1300	1167	1058
μ_c	0.001	0.001	1713	1569	1786	827
σ_c	<0.001	<0.001	849	848	847	1431

B.5 Marginal posterior distributions, trace diagrams and autocorrelation plots - Chapter 4

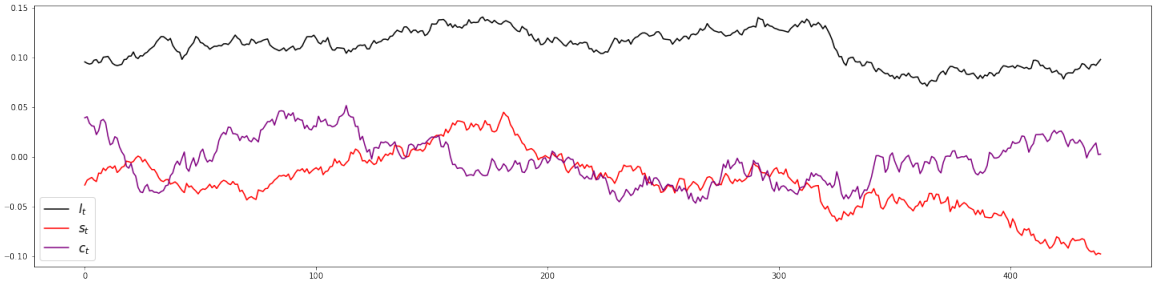
Figures 25 to 29 present the trace diagrams, kernel density estimates (KDEs) for the marginal posterior distributions and autocorrelation plots for the sampled chains for each of the parameters of the five models. The horizontal straight (black) lines on the trace plots are the true parameter values. For the KDE plots, the vertical dashed (red) lines are the true parameter

Figure 16 – Simulated data for the CIR model

Note – The short rate r_t is presented along with the observed spot rates for the maturities 1, 3 and 5.

Source: Elaborated by the author.

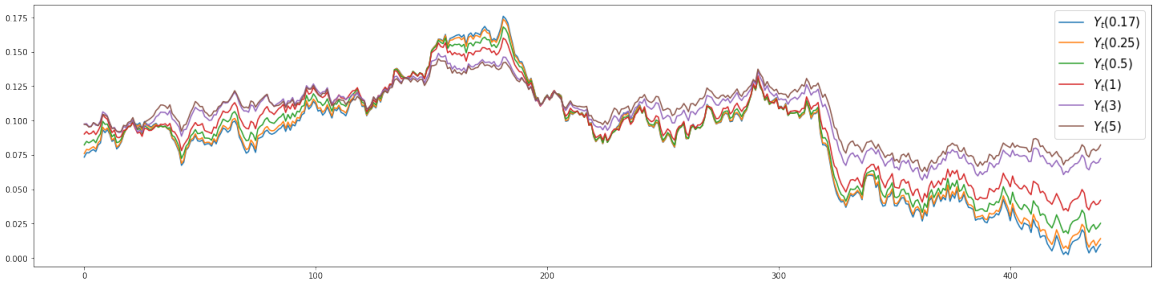
Figure 17 – Simulated factors for the DNS model (free λ)



Note – The presented factors correspond to level, slope and curvature of the yield curve.

Source: Elaborated by the author.

Figure 18 – Simulated spot yields for the DNS model (free λ)



Note – The presented maturities are 0.17, 0.25, 0.5, 1, 3 and 5.

Source: Elaborated by the author.

values, and the dotted (black) lines are the posterior medians for the sampled marginal posterior distributions.

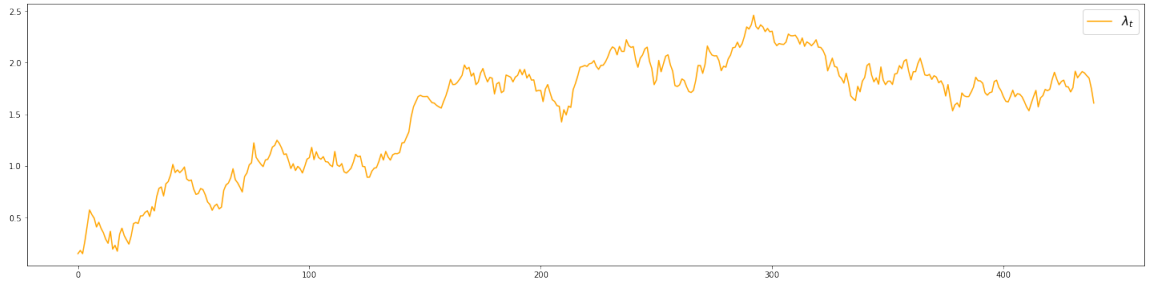
Throughout Chapter 4 we mentioned how much the medians of the posterior estimates for the parameters deviate from the parameters' true values, which can be verified in the KDE plots. As already mentioned in Chapter 4, none of the parameters for the five models show any major convergence issues, which can be verified in the trace and autocorrelation plots. Almost all

Figure 19 – Simulated factors for the DNS model (time-varying λ)

Note – The presented factors correspond to level, slope and curvature of the yield curve.

Source: Elaborated by the author.

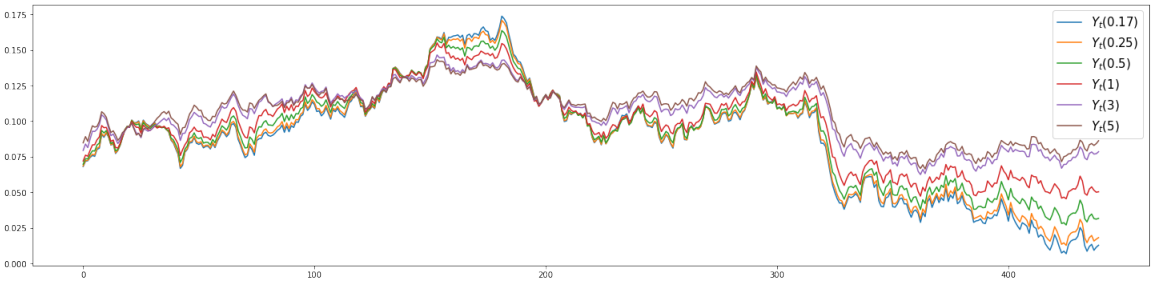
Figure 20 – Simulated λ_t for the DNS model (time-varying λ)



Note – The presented factors correspond to level, slope and curvature of the yield curve.

Source: Elaborated by the author.

Figure 21 – Simulated spot yields for the DNS model (time-varying λ)

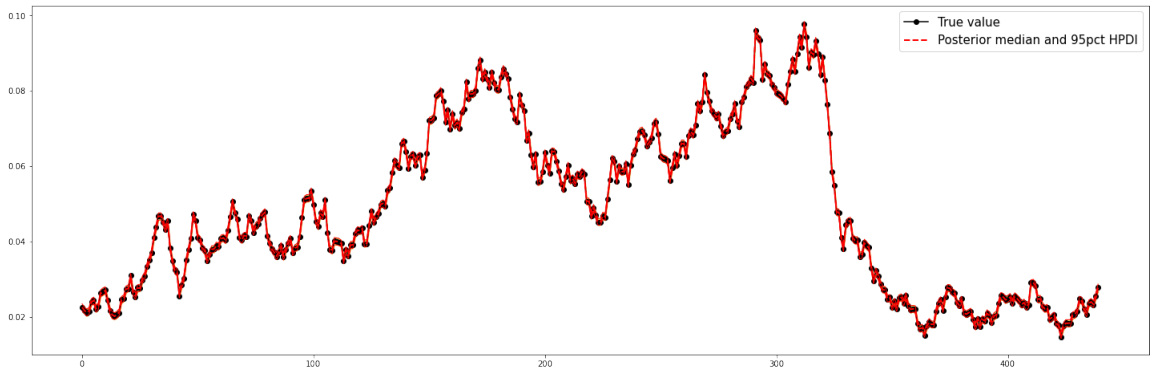


Note – The presented maturities are 0.17, 0.25, 0.5, 1, 3 and 5.

Source: Elaborated by the author.

of the parameters show fast decaying autocorrelation functions, indicating no major issues with chain convergence. The notable exception is the σ_λ parameter for the dynamic λ DNS model, with some persistently high autocorrelation.

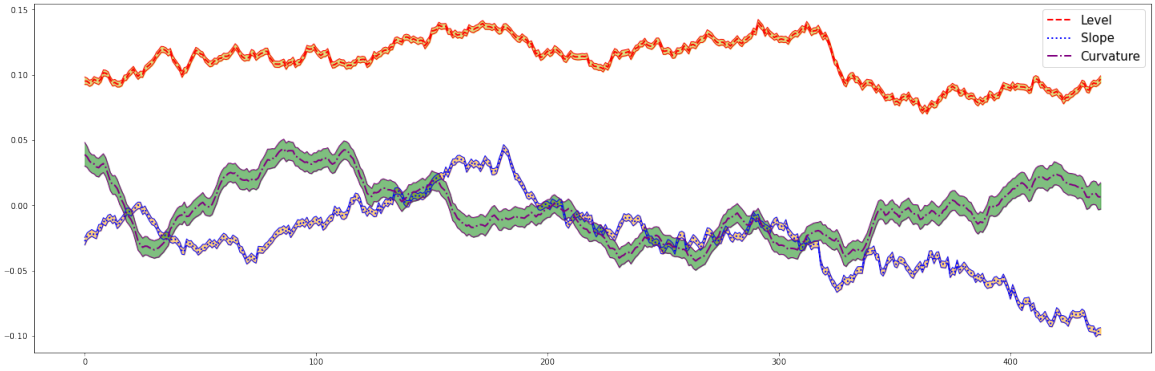
Figure 22 – Estimated short rate and 95% highest posterior density interval for the CIR model - Simulated data



Source: Elaborated by the author.

Note – The black line is the simulated state vector, and the red line the posterior median with the 95% highest posterior density interval.

Figure 23 – Estimated latent factors and 95% highest posterior density intervals for the DNS model (free λ) - Simulated data



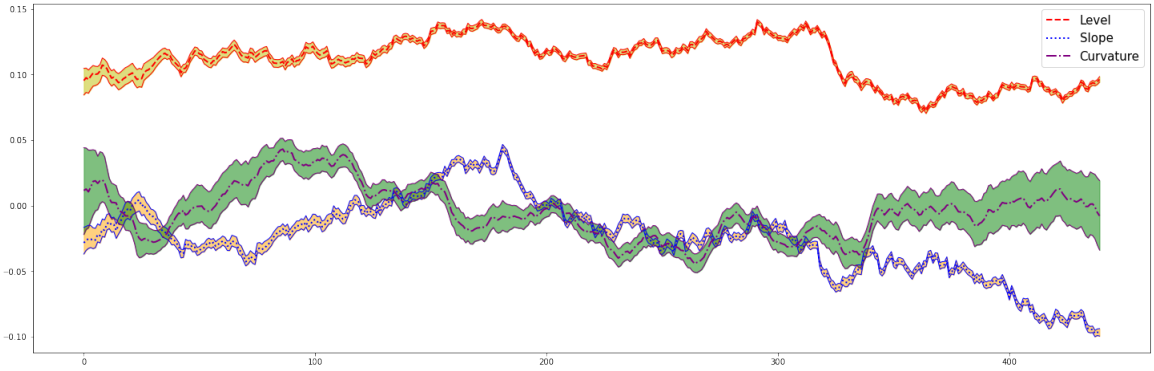
Source: Elaborated by the author.

B.6 Marginal posterior distributions, trace diagrams and autocorrelation plots - Chapter 5

Figures 30 to 34 show the trace diagrams, KDEs for the marginal posterior distributions and autocorrelation plots for the sampled chains for each of the parameters of the five models estimated with the dataset described in Chapter 5. While estimating the model with real data, we do not know the true parameter values. Thus, it is not possible to compare marginal posterior moments with the true values as it was the case with Chapter 4.

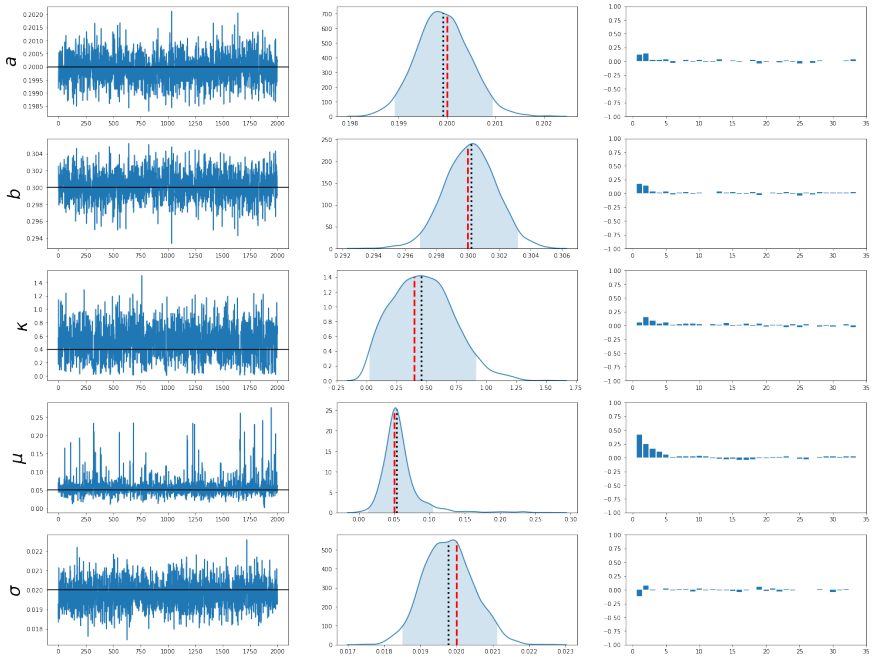
The KDEs for the two chains do not diverge very much from one another and the trace plots indicate no autocorrelation issues. Similarly to Chapter 4, the autocorrelations are fast decaying, therefore indicating no major chain convergence issues. The autocorrelation for σ_λ is significantly lower than for the simulated data model.

Figure 24 – Estimated latent factors and 95% highest posterior density intervals for the DNS model (time-varying λ) - Simulated data



Source: Elaborated by the author.

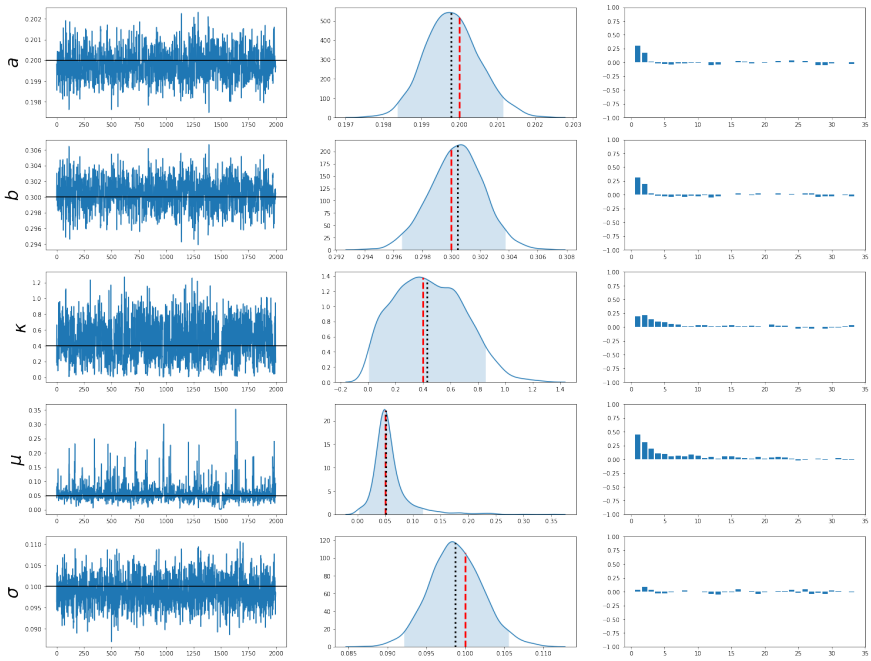
Figure 25 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the Vasicek model - Simulated data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

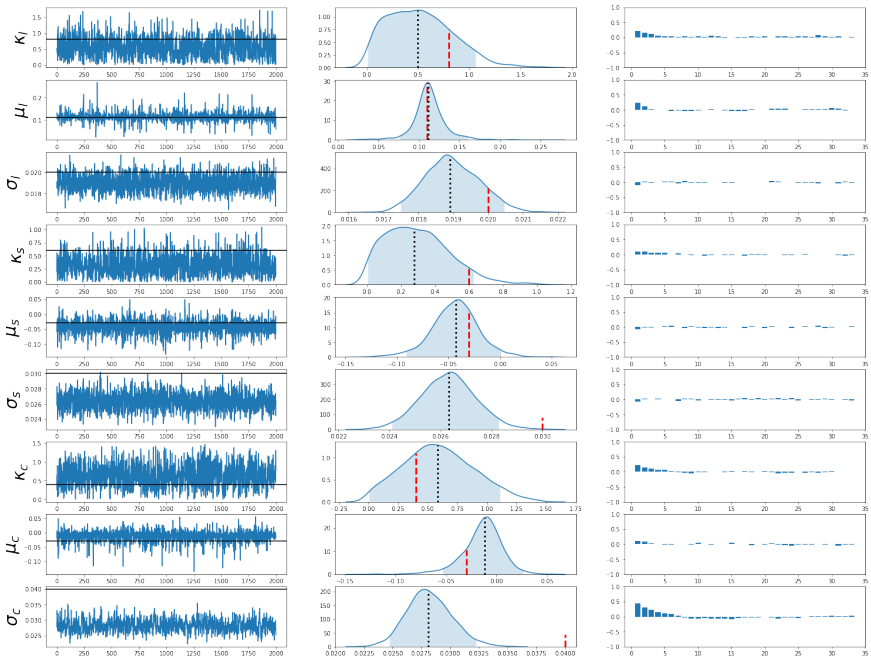
Figure 26 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the CIR model - Simulated data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

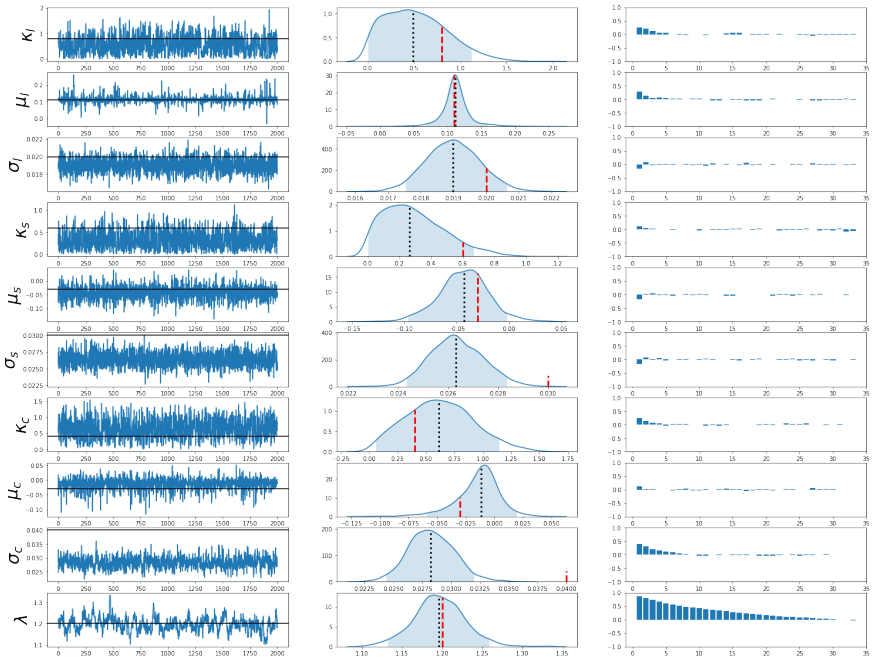
Figure 27 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (calibrated λ) - Simulated data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

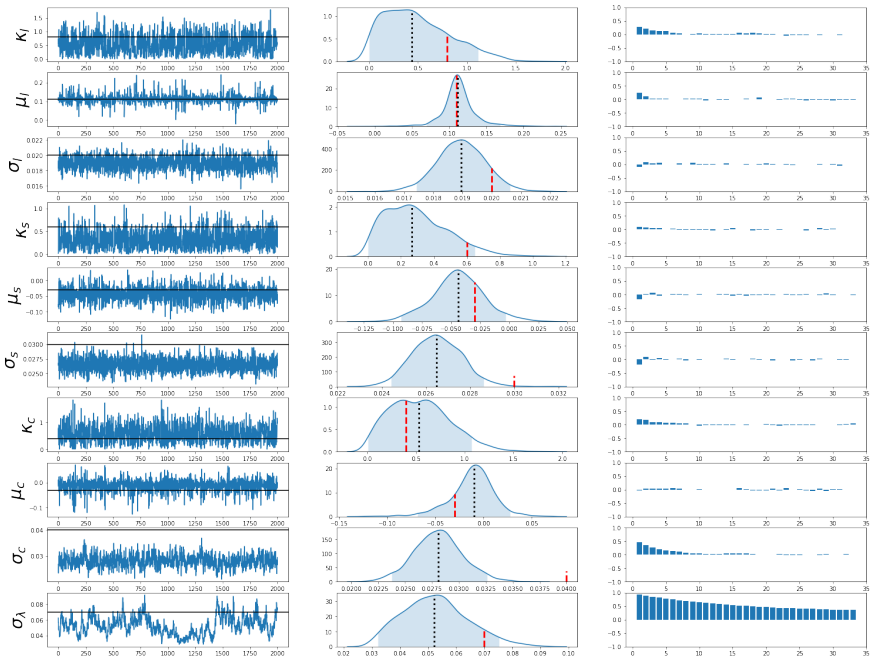
Figure 28 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (free λ) - Simulated data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

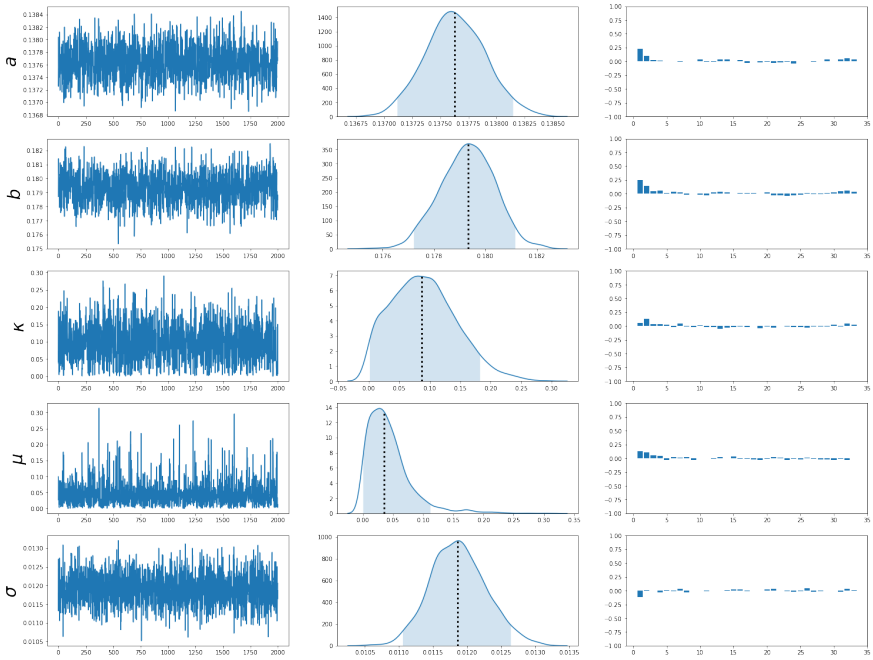
Figure 29 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (time-varying λ) - Simulated data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

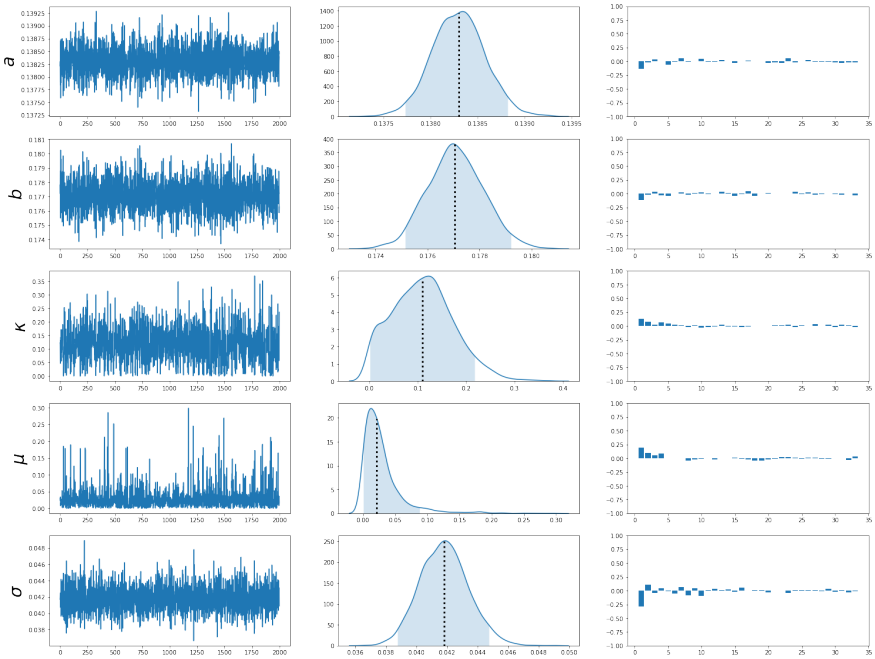
Figure 30 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the Vasicek model - Real data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

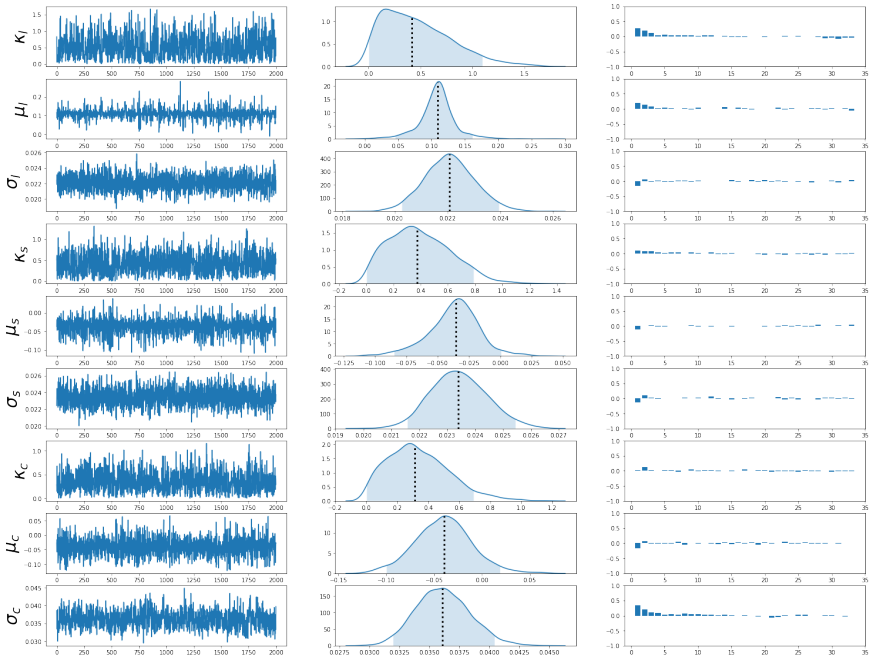
Figure 31 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the CIR model - Real data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

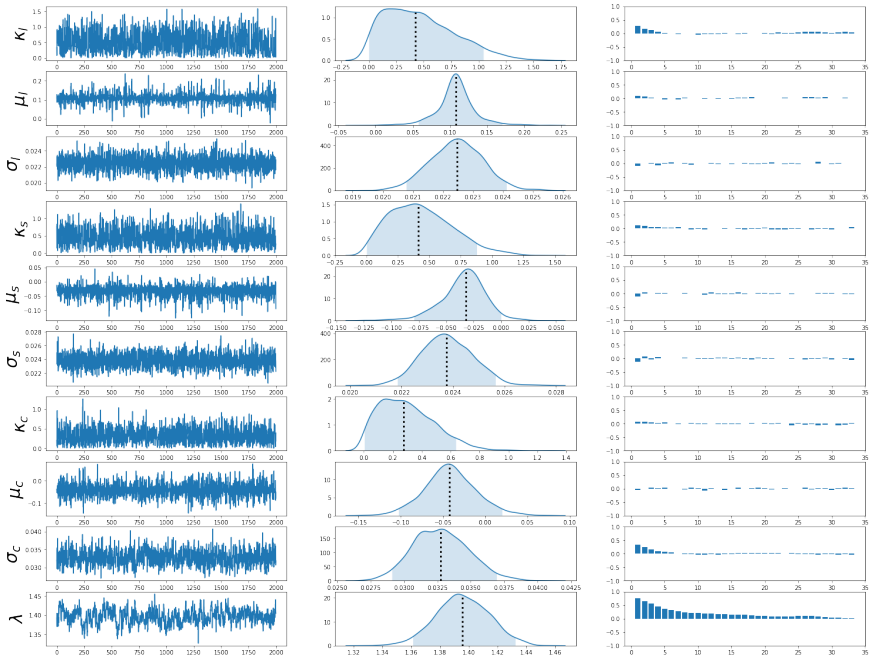
Figure 32 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (calibrated λ) - Real data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

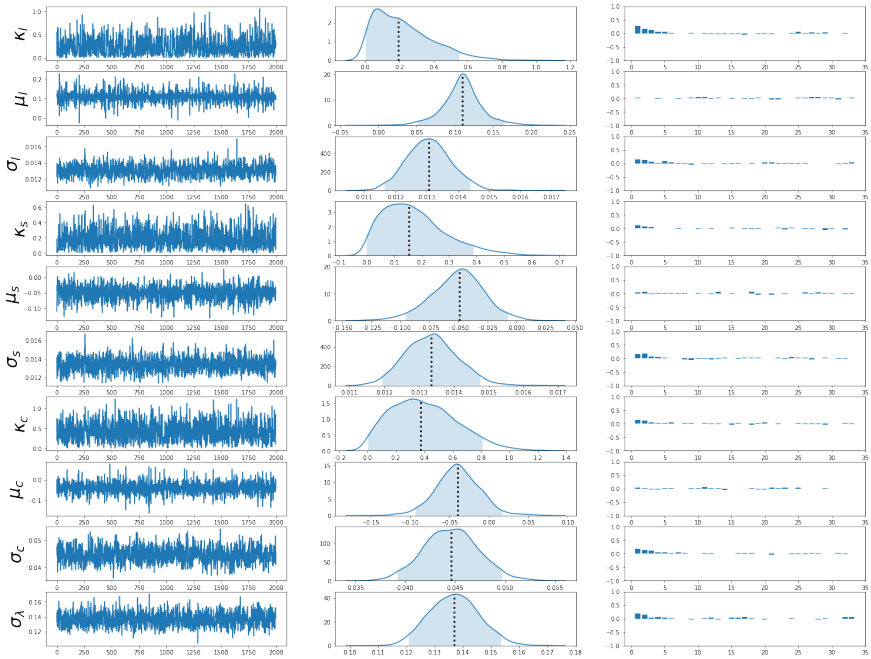
Figure 33 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (free λ) - Real data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

Figure 34 – Estimated trace diagrams, marginal posterior distributions and autocorrelation plots for the DNS model (time-varying λ) - Real data



Note – The three columns are the trace, KDE and autocorrelation for the sampled chain.

Source: Elaborated by the author.

APPENDIX

C

ADDITIONAL FORECASTING MEASURES

In this appendix we present two additional measures of statistical error for the rolling window forecast. The first one is the root mean squared error, or RMSE. The estimate of the RMSE is defined as

$$\widehat{RMSE}_h^{(\tau)} = \sqrt{\frac{\sum_{K=401}^{430} (\hat{y}_{med,(K+h)}^{(\tau)} - y_{(K+h)}^{(\tau)})^2}{30}},$$

with the same notation from Chapter 5. The second measure is the root mean squared logarithmic error, or RMSLE. The RSMLE estimate is obtained from

$$\widehat{RMSLE}_h^{(\tau)} = \sqrt{\frac{\sum_{K=401}^{430} \ln\left(\frac{1+y_{(K+h)}^{(\tau)}}{1+\hat{y}_{med,(K+h)}^{(\tau)}}\right)^2}{30}}.$$

Tables 22 and 23 contain, respectively, the RMSE and RMSLE for each model, maturity and time horizon. Similar conclusions to those of the Table 11 analysis can be drawn: the DNS-D model provides the best forecasts for most of the maturities. The two affine models perform better for shorter maturities and longer horizons, and the three DNS models perform better at longer maturities.

Table 22 – Comparison of root mean squared errors for the five models across different forecasting horizons and maturities - Rolling window forecast

Maturity	Model	Horizon							
		1	2	3	4	5	6	7	8
2 months	Vasicek	0.0047	0.0041	0.0036	0.0032	0.0030	0.0030	0.0030	0.0032
	CIR	0.0047	0.0041	0.0037	0.0033	0.0030	0.0030	0.0030	0.0032
	DNS-C	0.0008	0.0018	0.0027	0.0034	0.0044	0.0053	0.0062	0.0073
	DNS-F	0.0010	0.0019	0.0027	0.0035	0.0044	0.0055	0.0065	0.0075
	DNS-D	0.0009	0.0016	0.0024	0.0031	0.0039	0.0048	0.0056	0.0065
3 months	Vasicek	0.0036	0.0032	0.0029	0.0026	0.0025	0.0027	0.0030	0.0034
	CIR	0.0036	0.0031	0.0029	0.0026	0.0026	0.0028	0.0030	0.0033
	DNS-C	0.0009	0.0019	0.0028	0.0035	0.0045	0.0054	0.0063	0.0074
	DNS-F	0.0009	0.0018	0.0026	0.0034	0.0043	0.0053	0.0063	0.0073
	DNS-D	0.0008	0.0016	0.0024	0.0031	0.0039	0.0047	0.0056	0.0064
6 months	Vasicek	0.0014	0.0016	0.0021	0.0025	0.0031	0.0037	0.0043	0.0049
	CIR	0.0014	0.0016	0.0021	0.0025	0.0030	0.0036	0.0041	0.0048
	DNS-C	0.0011	0.0021	0.0030	0.0037	0.0046	0.0055	0.0063	0.0073
	DNS-F	0.0009	0.0017	0.0025	0.0033	0.0042	0.0051	0.0061	0.0070
	DNS-D	0.0008	0.0016	0.0024	0.0031	0.0039	0.0047	0.0055	0.0062
1 year	Vasicek	0.0028	0.0035	0.0042	0.0047	0.0053	0.0060	0.0065	0.0072
	CIR	0.0028	0.0034	0.0041	0.0046	0.0052	0.0058	0.0063	0.0070
	DNS-C	0.0014	0.0024	0.0033	0.0039	0.0047	0.0055	0.0062	0.0070
	DNS-F	0.0013	0.0022	0.0031	0.0038	0.0046	0.0054	0.0061	0.0069
	DNS-D	0.0012	0.0021	0.0030	0.0036	0.0043	0.0050	0.0056	0.0062
3 years	Vasicek	0.0055	0.0057	0.0061	0.0064	0.0063	0.0065	0.0066	0.0068
	CIR	0.0054	0.0056	0.0061	0.0063	0.0062	0.0064	0.0065	0.0066
	DNS-C	0.0025	0.0035	0.0043	0.0048	0.0051	0.0051	0.0055	0.0056
	DNS-F	0.0026	0.0037	0.0044	0.0049	0.0052	0.0053	0.0057	0.0059
	DNS-D	0.0027	0.0035	0.0043	0.0050	0.0051	0.0051	0.0054	0.0055
5 years	Vasicek	0.0093	0.0092	0.0095	0.0097	0.0094	0.0093	0.0091	0.0089
	CIR	0.0093	0.0092	0.0094	0.0097	0.0094	0.0092	0.0090	0.0088
	DNS-C	0.0028	0.0037	0.0047	0.0053	0.0054	0.0057	0.0057	0.0059
	DNS-F	0.0027	0.0035	0.0046	0.0053	0.0054	0.0058	0.0057	0.0059
	DNS-D	0.0027	0.0034	0.0044	0.0052	0.0052	0.0056	0.0056	0.0056

Table 23 – Comparison of root mean squared logarithmic error for the five models across different forecasting horizons and maturities - Rolling window forecast

Maturity	Model	Horizon							
		1	2	3	4	5	6	7	8
2 months	Vasicek	0.0045	0.0040	0.0035	0.0031	0.0029	0.0028	0.0029	0.0031
	CIR	0.0045	0.0039	0.0035	0.0031	0.0029	0.0029	0.0029	0.0031
	DNS-C	0.0008	0.0017	0.0025	0.0033	0.0042	0.0051	0.0060	0.0070
	DNS-F	0.0009	0.0018	0.0025	0.0034	0.0043	0.0053	0.0062	0.0072
	DNS-D	0.0008	0.0016	0.0023	0.0030	0.0038	0.0046	0.0054	0.0062
3 months	Vasicek	0.0035	0.0030	0.0027	0.0025	0.0024	0.0026	0.0029	0.0033
	CIR	0.0034	0.0030	0.0028	0.0025	0.0025	0.0027	0.0029	0.0032
	DNS-C	0.0009	0.0018	0.0027	0.0034	0.0043	0.0052	0.0061	0.0071
	DNS-F	0.0009	0.0017	0.0025	0.0033	0.0042	0.0051	0.0061	0.0070
	DNS-D	0.0008	0.0015	0.0023	0.0030	0.0037	0.0046	0.0054	0.0061
6 months	Vasicek	0.0013	0.0015	0.0020	0.0024	0.0029	0.0036	0.0041	0.0047
	CIR	0.0013	0.0016	0.0020	0.0024	0.0029	0.0035	0.0040	0.0046
	DNS-C	0.0011	0.0020	0.0029	0.0035	0.0044	0.0053	0.0061	0.0070
	DNS-F	0.0009	0.0016	0.0024	0.0031	0.0040	0.0049	0.0058	0.0067
	DNS-D	0.0008	0.0015	0.0023	0.0030	0.0037	0.0045	0.0052	0.0060
1 year	Vasicek	0.0027	0.0033	0.0040	0.0045	0.0051	0.0057	0.0062	0.0069
	CIR	0.0026	0.0033	0.0040	0.0044	0.0050	0.0056	0.0061	0.0067
	DNS-C	0.0013	0.0023	0.0032	0.0038	0.0045	0.0053	0.0059	0.0068
	DNS-F	0.0012	0.0021	0.0030	0.0037	0.0044	0.0051	0.0059	0.0066
	DNS-D	0.0011	0.0020	0.0028	0.0035	0.0041	0.0048	0.0054	0.0060
3 years	Vasicek	0.0052	0.0053	0.0058	0.0061	0.0060	0.0062	0.0062	0.0064
	CIR	0.0051	0.0053	0.0057	0.0060	0.0059	0.0061	0.0061	0.0063
	DNS-C	0.0023	0.0033	0.0040	0.0045	0.0048	0.0048	0.0052	0.0053
	DNS-F	0.0024	0.0035	0.0042	0.0046	0.0049	0.0050	0.0054	0.0056
	DNS-D	0.0025	0.0033	0.0040	0.0047	0.0048	0.0049	0.0051	0.0052
5 years	Vasicek	0.0087	0.0086	0.0089	0.0091	0.0088	0.0087	0.0085	0.0083
	CIR	0.0087	0.0086	0.0088	0.0091	0.0088	0.0086	0.0084	0.0082
	DNS-C	0.0027	0.0035	0.0044	0.0050	0.0051	0.0054	0.0054	0.0055
	DNS-F	0.0026	0.0033	0.0043	0.0050	0.0050	0.0054	0.0054	0.0055
	DNS-D	0.0025	0.0032	0.0041	0.0048	0.0048	0.0052	0.0052	0.0053


ICMC
USP
SÃO CARLOS