Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Departamento de Economia
Programa de Pós-Graduação em Economia — Área: Economia Aplicada

Cássio Roberto de Andrade Alves

Essays on macroeconometrics and financial econometrics: a Bayesian approach

Ensaios sobre macroeconometria e econometria de finanças: uma abordagem Bayesiana

Supervisor: Prof. Dr. Márcio Poletti Laurini

Ribeirão Preto
2023

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Prof. Dr. Fabio Augusto Reis Gomes
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Milton Barossi Filho
Chefe do Departamento de Economia

Prof. Dr. Luciano Nakabashi
Coordenador do Programa de Pós-Graduação em Economia - Área: Economia Aplicada

CÁSSIO ROBERTO DE ANDRADE ALVES

Essays on macroeconometrics and financial econometrics: a Bayesian approach

Ensaios sobre Macroeconometria e Econometria de Finanças: Uma Abordagem Bayesina

Tese apresentada ao Programa de Pós-Graduação em Economia — Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Doutor em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

Supervisor: Prof. Dr. Márcio Poletti Laurini

Ribeirão Preto
2023

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Alves, Cássio Roberto de Andrade Alves
Essays on macroeconometrics and financial econometrics: a Bayesian approach / Universidade de São Paulo – USP
Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Departamento de Economia
Programa de Pós-Graduação em Economia - Área: Economia Aplicada;
Supervisor: Prof. Dr. Márcio Poletti Laurini
Ribeirão Preto, 2023- 104 p. : il. 30cm

Tese de Doutorado – Universidade de São Paulo, 2023.

1. Econometria Bayesiana. 2. Macroeconomia. 3. Finanças. 4. Previsão. 5. Variáveis instrumentais. I. Orientador: Prof. Dr. Márcio Poletti Laurini. II. Universidade de São Paulo. III. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto.

Name: Cássio Roberto de Andrade Alves
Title: Essays on macroeconometrics and financial econometrics: a Bayesian approach

Tese apresentada ao Programa de Pós-Graduação em Economia — Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Doutor em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

Approved in:

Examination board:

Prof. Dr. Prof. Dr. Márcio Poletti Laurini(Supervisor)
FEA-RP/Universidade de São Paulo

Prof. Dr. Hedibert Freitas Lopes
INSPER

Prof. Dr. André Portela Santos
CUNEF Universidad

Prof. Dr. Guilherme Valle Moura
Universidade Federal de Santa Catarina

Prof. Dr. João Frois Caldeira
Universidade Federal de Santa Catarina

Prof. Dr. Jorge Alberto Achcar
DMS/Universidade de São Paulo

Acknowledgements

During my doctoral studies and the preparation of this thesis over the past four years, I have been fortunate to have had the support and inspiration of many brilliant individuals. Despite encountering some challenges along the way, having great people by my side has made the journey much less painful. I am truly grateful for everyone who has contributed in any way to my doctoral journey. This sentiment is deeply heartfelt.

I am thankful to my advisor, **Márcio Laurini**. Márcio is one of the most brilliant and friendly academics I ever know. Working with him during this time has allowed me to acquire a small fraction of his vast knowledge, which is of great value. Besides teaching me technical subjects in econometrics and computation, I could learn from his very productive style of working. He also was by my side during difficulties to find resources to develop this research.

I am especially thankful to my lovely girlfriend, **Claudia Daiane Batista Bettio**. I am grateful for your unwavering support throughout my Ph.D. journey. Your encouragement and understanding have been a constant source of strength for me, and I could not have accomplished this milestone without you by my side. Thank you for all the long discussions about my research, about academic life and for staying with me in the days of struggle and in the days of glory.

I express my gratitude to my beloved parents, **Luiz Carlos Alves** and M**irian Cristina de Andrade Alves**, for their constant love and support which have been an essential element throughout my life, including my doctoral journey. I thank you, Mom and Dad!

I extend my appreciation to my beloved brother, **Luiz Gustavo de Andrade Alves**, for his continuous motivation to push beyond my limits, which I deeply value. I am also grateful for our extensive conversations on various topics such as economics, computation, Bayesian econometrics, and many others.

I would also like to express my gratitude to all the professors who guided me throughout my doctoral journey. The knowledge they imparted to me holds immense value, and I genuinely appreciate their contributions. I would like to give special recognition to two professors in particular. Firstly, **Jefferson Bertolai**'s teachings on macroeconomics, mathematics, and programming have been transformative. Your classes have significantly altered my perspective on various subjects. Secondly, I want to acknowledge the profound influence of **Walter Belluzo** on my understanding of econometrics and effective teaching methodologies. Despite not being my direct instructor, collaborating with him has deeply influenced my approach to the teaching process.

I would also like to express my gratitude to the friendships I cultivated during my doctoral program, specifically **Udilmar Zabot**, **Bruno Caleman**, and **Gustavo Sung**.

I had the privilege to engage with them in various economic discussions, both virtually and in person. Their presence made a significant impact on my journey, and I am thankful for their friendship.

Several other friends also played a crucial role in my journey. **Aishameriane Venes**, **Bruno Schock**, **Denise Manfredini**, **Leilane Freitas**, **Paulo Victor Fonseca**, **Richard Schorrenberg**, and **Werley Cordeiro** were always available to review and provide feedback on my work. I express my gratitude to all of you!

Over the past four years, I had the privilege to attend various economics and econometrics conferences, where I received invaluable feedback on my presentations. I would like to extend my appreciation to **Guilherme Moura**, **Raul Riva**, **Vitor Passebom**, and **Eduardo Mendes** for their significant contributions.

I also would like to express my sincerest gratitude to **Kuruvilla Joseph Abraham**, whose invaluable support and collaboration were instrumental in the completion of this thesis. Joseph's expertise and insightful contributions significantly enriched the research and writing process.

I would also like to extend my heartfelt thanks to **Fernando Roriz**, **Luka Barbosa** and **Thomas Wu**, my esteemed heads, for their unwavering support and mentorship throughout the last part of my doctoral journey. Their guidance and encouragement have been crucial in shaping both my academic and professional growth. Their belief in my potential has been a constant source of motivation, and I sincerely appreciate the opportunities they have provided me.

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

Abstract

Alves, C. R. A. (2023) **Essays on macroeconometrics and financial econometrics: a Bayesian approach**. Doctoral Dissertation - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

The goal of this doctoral dissertation is to showcase the importance and applicability of Bayesian econometrics in the realm of financial and macroeconomic analysis. The thesis is formed by four independent essays in macroeconometrics and financial econometrics, in which the Bayesian estimation is the common factor of the four essays. In the first article, a heteroscedastic structure with jumps is added to a structural model to measure inflationary persistence in Brazil from 1995 to 2019. The general result is that including stochastic volatility with jumps reduces intrinsic inflation persistence. In the second article, the Brazilian Central Bank Communication is used to predict the yield curve. The results indicate that the Central Bank Communication, measured by the sentiment of the Central Bank, helps to predict the term structure of interest rate. In the third article we explore the the analysis of Central Bank Communication and yield curve in a in-sample perspective. The objective of this article is to investigate the bidirectional relation between the Brazilian Central Bank Communication and the yield curve. The results shown that the Central Bank Communication can shape yield curve curvature and slope, and that there is a strong relation between monetary authority communication and the curvature of the yield curve. Both article two and three present evidence that words of monetary authority impacts market players, making it a valuable instrument for monetary policy. The last article proposes an instrumental variable Bayesian shrinkage approach to estimate the Capital Asset Pricing Model (CAPM) using a large set of instruments. Using simulated data, the proposed approach reduces the bias of estimation, caused by measurement errors. In an empirical application, the proposed method delivered a different estimation from the traditional approach and this difference increases the explanatory power of the Capital Asset Pricing Model in explaining the variation in the average cross-sectional returns of assets.

Keywords: Inflation persistence, Monetary policy, Stochastic volatility, Yield Curve, Sentiment Analysis, Bayesian Estimation, Instrumental Variable, Shrinkage prior.

Resumo

Alves, C. R. A. (2023) **Ensaios sobre macroeconometria e econometria de finanças: uma abordagem Bayesiana**. Tese (Doutorado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

O objetivo desta dissertação de doutorado é mostrar a importância e aplicabilidade da econometria bayesiana no domínio da análise financeira e macroeconômica. A tese é formada por quatro ensaios independentes em macroeconometria e econometria financeira, em que a estimação Bayesiana é o fator comum dos quatro ensaios. No primeiro artigo, uma estrutura heterocedástica com saltos é adicionada a um modelo estrutural para medir a persistência inflacionária no Brasil de 1995 a 2019. O resultado geral é que a inclusão da volatilidade estocástica com saltos reduz a persistência intrínseca da inflação. No segundo artigo, a Comunicação do Banco Central do Brasil é utilizada para prever a curva de juros. Os resultados indicam que a Comunicação do Banco Central, medida pelo sentimento do Banco Central, ajuda a prever a estrutura a termo da taxa de juros. No terceiro artigo exploramos a análise da Comunicação do Banco Central e da curva de juros em uma perspectiva in-sample. O objetivo deste artigo é investigar a relação bidirecional entre a Comunicação do Banco Central do Brasil e a curva de juros. Os resultados mostraram que a Comunicação do Banco Central pode moldar a curvatura e a inclinação da curva de juros, e que existe uma forte relação entre a comunicação da autoridade monetária e a curvatura da curva de juros. Tanto o artigo dois quanto o terceiro apresentam evidências de que as palavras da autoridade monetária impactam os agentes do mercado, tornando-se um instrumento valioso para a política monetária. O último artigo propõe uma abordagem de encolhimento bayesiano de variável instrumental para estimar o Capital Asset Pricing Model (CAPM) usando um grande conjunto de instrumentos. Usando dados simulados, a abordagem proposta reduz o viés de estimativa, causado por erros de medição. Em uma aplicação empírica, o método proposto forneceu uma estimativa diferente da abordagem tradicional e essa diferença aumenta o poder explicativo do Capital Asset Pricing Model em explicar a variação nos retornos transversais médios dos ativos.

Palavras-chaves: Econometria Bayesiana; Finanças; Macroeconomia; Previsão; Variáveis instrumentais; Função de Resposta ao impulso .

List of Figures

List of Tables

Contents

1 General Introduction

Over the last decades, there have been significant transformations in the methods used by researchers to examine economic and financial time series. The increase in computer processing capacity has played a relevant role in these transitions. Researchers are able to analyze large models in terms of parameters to estimate more than ever before. These large models may appear either because of large data sets or because of the flexibility in the structure of the model, such as time-varying parameters. In the wake of computational improvements, Bayesian analysis has gained a highlighted role since it can estimate complex models using prior knowledge and updating posterior probabilities by iterative methods that usually require much computation resources.

Models involving a large number of parameters relative to the number of observations can be naturally dealt with by regularizing it via prior distribution. By combining prior knowledge with sample information, Bayesian analysis allows us to make inferences about parameters through analysis of their posterior distribution moments. The posterior distribution for complex models, such as high-dimensional and non-linear models, usually has no analytical form and requires simulation methods to estimate its moments. Indeed, this is the case in many applications in finance and macroeconomics, which frequently uses time-series data and is interested in latent states that depends on time, consequently leading to high-dimensional integrals.

The Monte Carlo (MC) method, first introduced by Metropolis and Ulam (1949), has played a crucial role in simulation methods for estimating posterior distributions and served as the foundation for modern Markov Chain Monte Carlo (MCMC) methods. While it could be argued that Monte Carlo was already in use before the twentieth century[1], Metropolis and Ulam (1949) were the pioneers to propose that statistical sampling could approximate a solution to a problem without an analytical solution. This idea led to the development of the Metropolis-Hastings (M-H) algorithm by Metropolis et al. (1953) and further refinements by Hastings (1970), which is now one of the most widely used MCMC algorithms for posterior distribution estimation.

The Metropolis-Hastings (M-H) algorithm is designed to generate a sequence of samples from a Markov Chain that converges to the target posterior distribution, which may not have a known analytical form or a simple direct way to sample from it. The algorithm uses an instrumental distribution to generate candidate samples and an accept-reject step to determine whether the candidate sample or the previous state is the next value of the Markov Chain. Under certain conditions, the Markov Chain generated by the M-H algorithm will converge to a stationary distribution equivalent to the target posterior distribution (For a formal and clear explanation of the mechanism underlying

[1] (HITCHCOCK, 2003, pp. 254)

M-H, please refer Chib and Greenberg (1995)). It makes the algorithm a flexible tool for estimating complex posterior distributions, with applications in various fields such as macroeconomics and finance.

The extraction of the volatility of time series, for instance, is a task of interest of both financial and macroeconomic fields (RUPPERT; MATTESON, 2011). Before the Bayesian analysis of stochastic volatility models, introduced by Jacquier, Polson and Rossi (2002), the literature had two options to extract volatility. The first was to estimate the parameters of a stochastic volatility model via methods of moments and then use the Kalman filtered (or smoothed) estimates of the volatility. The second approach is to approximate a non-linear state-space model implied by the stochastic volatility model by a normal-linear state-space model and then use quasi-maximum likelihood constructed with the prediction error decomposition obtained from Kalman Filter. Jacquier, Polson and Rossi (2002) proposed a Bayesian estimation of the stochastic volatility model that uses the M-H algorithm. The Bayesian solution proposed by these authors dominates both the method of moments and the quasi-maximum likelihood approaches.

While the Metropolis-Hastings (M-H) algorithm is a flexible and versatile method for estimating posterior distributions, it can become computationally expensive in high-dimensional parameter spaces. In such cases, the traditional M-H algorithm can require large chains to achieve stationary, which can be time-consuming and challenging, even with modern computational tools. To deal with this problem, the Markov Chain Monte Carlo (MCMC) literature has evolved to include new algorithms for posterior distribution sampling, including improved versions of the M-H algorithm.

For example, the Hamiltonian Monte Carlo (HMC) algorithm, also known as Hybrid Monte Carlo (DUANE et al., 1987; NEAL et al., 2011), is an efficient MCMC method that uses gradient information to guide the sampling process, allowing for faster convergence and more efficient exploration of the parameter space. HMC generally requires much fewer iterations to converge to the stationary distribution than M-H. The cost per iteration of HMC, however, is higher than that of M-H due to the inclusion of a numerical integrator used to integrate the differential equations that describe Hamiltonian dynamics. Another example of an MCMC algorithm is the Elliptical Slice Sampler (MURRAY; ADAMS; MACKAY, 2010), which is less flexible than the M-H algorithm but is better suited for certain types of distributions. This algorithm uses elliptical contours to define the proposal distribution, which can be more efficient than the random walk proposal distribution used in the M-H algorithm.

Overall, new and better posterior distribution sampling techniques have been discovered in the MCMC literature, giving researchers a variety of alternatives to select from based on the particular situation at hand. These methods have substantially broadened the use of Bayesian inference, making it possible to analyze increasingly intricate models and data sets. In this way, this thesis aims to investigate macroeconomic and financial issues using these new, advanced sampling techniques. The emergence of these innovative techniques has made it feasible to study aspects of macroeconomics that were previously too expensive or unattainable. This thesis utilizes these efficient MCMC methods to investigate fresh applications in the fields of finance and macroeconomics.

This thesis is formed by four self-contained essays in macroeconometrics and financial econometrics. The Bayesian estimation is the common factor of the three essays. The first one involves a macroeconomic subject: we analyze the inflation persistence taking into account the stochastic volatility of inflation with changes in level. The second and third essays treat a theme that is of interest to both macroeconomists and financial economists: we investigate the impact of the Brazilian Central Bank on the yield curve by examining both in-sample and out-of-sample effects. The fourth is related to a financial econometric issue: we study the estimation of asset price models in a data-rich environment, aiming to correct measurement errors present in such estimation.

The estimation process of the essays involves jumps in stochastic volatility, sparsity in data, and regularization tools. These characteristics can be easily dealt with in the Bayesian approach, using hierarchical models and MCMC methods. In the traditional Classical approach, such characteristics may be very difficult or even impossible to estimate, due to the difficulties in evaluating the likelihood function and moment conditions or because of the ill-behaved likelihood function. In this sense, the Bayesian approach plays an essential role in the estimation of such intricated models.

In Chapter 2, we present a new method for measuring inflation persistence. By controlling for conditional heteroscedasticity and allowing for jumps in volatility. We find that the Brazilian inflation persistence is lower compared to the traditional autoregressive techniques that ignore the changes in volatility and changes in the inflation target. This paper has been accepted in Economics Bulletin.

Chapter 3 presents evidence on how we can use text information extracted from the Central Bank documents to better predict the yield curve. Results indicate that, together with macroeconomic variables, the sentiment of the Central Bank helps to predict the yield curve. This paper was published in the Journal of Forecasting.

In Chapter 4, the focus is on the relationship between central bank communication and the yield curve, but with an in-sample perspective. The study evaluates how changes in central bank communication affect the yield curve and vice versa. The results indicate that changes in central bank communication affect the yield curve, and central bank communication also responds to changes in the shape of the yield curve. Furthermore, there is a strong association between central bank communication and the curvature of the yield curve.

Finally, Chapter 5 is interested in estimating the capital asset price model in the context of many instruments. To deal with the many instruments setting we use a Bayesian regularization technique. Using synthetic data, the results indicate that using many instruments achieves better estimates, in terms of bias and variance. In an empirical analysis, the study demonstrates that utilizing many instruments and the Bayesian regularization technique can enhance the explanatory power of expected cross-sectional returns.

2 Measuring inflation persistence under time-varying inflation target and stochastic volatility with jumps

Abstract

We analyze whether the presence of a time-varying inflation target and stochastic volatility affect inflation persistence. We estimated different autoregressive specifications for inflation with and without time-varying parameters. The results show that the inflation persistence diminishes when we consider time-varying inflation target and stochastic volatility with jumps. We conclude that neglecting the time variation in inflation target and inflation volatility results in an upward-biased estimation of persistence.

Keywords: Inflation persistence, Time-varying parameters, Stochastic volatility.

2.1 Introduction

How long inflation shocks last, and the magnitude of these shocks are measures that interest monetary policymakers. These quantities are related to inflation persistence and volatility, which can be obtained by estimating reduced-form models for inflation. Inflationary processes may present changes in the mean (or inflation target) and the variance (volatility) over time, and these changes may affect inflation persistence. Thus, the implications of the time-varying inflation target and time-varying volatility on the estimation of inflation persistence, if any, should be considered to give more accurate information to policymakers. This paper addresses these implications.

The unconditional expectation of inflation may change over time due to shifts in the inflation target, for instance. To capture these changes, researchers proposed the use of time-varying mean parameters to model inflation and measure its persistence (COGLEY; SARGENT, 2001; DOSSCHE; EVERAERT, 2005; BILICI; ÇEKIN, 2020). Their results showed evidence against constant mean for inflation. Furthermore, the variance of inflation may be higher or lower over time. According to the literature on modeling inflation, the volatility of shocks affecting inflation is also governed by a time-varying parameter (COGLEY; SARGENT, 2005; LAURINI; VIEIRA, 2013). Consequently, measures of inflation persistence based on the constant mean and homoskedastic models for inflation may be unreliable.

The literature on transitory and permanent decomposition in macroeconomic time series usually applies it for the first moment. Examples of this approach are the work by Stock and Watson (2007), which separates transitory and permanent components for inflation, and Krane (2011), which uses the decomposition for the transitory and permanent shocks to the output. The transitory and permanent decomposition of time series second moment is less common and has been little explored in the literature for inflation.

The volatility may vary due to institutional or structural changes, representing permanent shifts, while other sporadic variations represent transitory movements. The model of stochastic volatility with jumps introduced by Qu and Perron (2013) allows us to separate these two kinds of changes in the inflation volatility, just as the finance literature has done (see, e.g., Chaim and Laurini (2018)). Extracting these two components of inflation volatility gives us information about the magnitude of shocks affecting inflation.

Besides the inclusion of jumps in the inflation volatility, there is no evidence of the effects of stochastic volatility with jumps (SVWJ) and time-varying inflation target (TVIT) on inflation persistence. Thus, this paper aims to verify if the presence of a TVIT and SVWJ affect inflation persistence. To this end, we examine Brazilian inflation data from 1995 to 2020, which is an interesting case since it includes different levels in the mean and a relevant institutional change with the inflation target adoption in 1999.

The contribution of this paper is two-folded. First, we analyze how the inclusion of TVIT and SVWJ affects inflation persistence. Accordingly, we compare specifications with and without these characteristics. To model the TVIT, we use the approach of Dossche and Everaert (2005). Second, we verify if there exist permanent changes in inflation volatility by implementing the Qu and Perron (2013) approach to model volatility.

Our approach to decomposing the inflation time series is related to the literature initiated by Nelson and Plosser (1982), which argues that, in general, macroeconomic time series are better described by a non-stationary process. Later, this literature was extended by Perron (1989), which showed that if one considers a breaking point in the time series, the measure of persistence is affected. Our approach allows multiple breaks in the inflation time series by introducing a TVIT, and additionally, we introduce time-varying variance by introducing SVWJ. In this way, we show that, at least for inflation, these two features affect the persistence of the inflation time series.

The results showed that both TVIT and SVWJ affect inflation persistence. The specification that includes these two characteristics exhibits an intrinsic inflation persistence of 0.56 compared to 0.76 in the specification without them. Moreover, the SVWJ presents evidence of permanent shifts in volatility. These results point to the importance of considering a TVIT and heteroskedastic model to find an accurate measure of inflation persistence.

2.2 Modeling Inflation

The literature on inflation persistence usually models inflation using the k-order autoregressive process, AR(k). This approach allows us to extract a measure of intrinsic inflation persistence by summing the autoregressive coefficients (FUHRER, 2010). We adopt this approach and incorporate TVIT and SVWJ.

To introduce TVIT, we follow Dossche and Everaert (2005) and Kozicki and Tinsley (2005). The inflation is allowed to follow an AR(k) process around the inflation target perceived by the private agents, π_t^P:

$$\pi_t = \left(1 - \sum_{i=1}^{k} \varphi_i\right) \pi_t^P + \sum_{i=1}^{k} \varphi_i \pi_{t-i} + \sigma_t \nu_t, \quad \nu_t \sim \mathcal{N}(0,1), \tag{2.1}$$

where σ_t represents the standard deviation of shocks affecting inflation.

The perceived inflation target π_t^P evolves as a convex combination of the perceived inflation target in the previous period and the inflation target pursued by the Central Bank, π_t^T. That is:

$$\pi_{t+1}^P = (1-\delta)\pi_t^P + \delta \pi_{t+1}^T. \tag{2.2}$$

Private agents may obtain information about π_{t+1}^T by comparing the interest rate set by the Central Bank and their expectation about the interest rate (see Kozicki and Tinsley (2005), for details). The Central Bank inflation target follows a driftless random walk with innovation $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, which reflects, for instance, changes in Central Bank preferences. Using this assumption in equation (2.2), we obtain:

$$\pi_{t+1}^P = (2 - \delta)\pi_t^P + (\delta - 1)\pi_{t-1}^P + \delta\eta_{t+1}. \tag{2.3}$$

Equations (2.1) and (2.3) model the inflation around a TVIT. Moreover, the way we model the inflation target allows us to extract an expectation-based inflation persistence component, measured by $(1 - \delta)$. Note that if δ is close to one, then the private agents perfectly predict the Central Bank's inflation target, and there is no persistence effect due to expectations errors (DOSSCHE; EVERAERT, 2005).

Following Qu and Perron (2013), we decompose the log-variance as the sum of a transitory component, h_t, and a permanent component, μ_t, that is, $\log(\sigma_t^2) = h_t + \mu_t$, so that $\sigma_t = \exp(h_t/2 + \mu_t/2)$. The transitory component follows a stationary AR(1), while the permanent component is a compound binomial process:

$$h_t = \rho h_{t-1} + \sigma_h \varepsilon_{h,t}, \quad -1 < \rho < 1 \quad \varepsilon_{h,t} \sim \mathcal{N}(0,1), \tag{2.4}$$

$$\mu_t = \mu_{t-1} + d_t \sigma_w w_t, \quad w_t \sim \mathcal{N}(0,1), \text{ and } d_t \sim \text{Bernoulli}(p) \tag{2.5}$$

Equations (2.4)-(2.5) allow us to separate transitory changes from permanent shifts in level of stochastic volatility. These equations together with equations (2.1) and (2.3) form the complete model. Note that, this general model nested several specifications: *Model 1)* π_t^P and σ_t are constant; *Model 2)* only π_t^P is constant; *Model 3)* π_t^P varies over time, but σ_t is constant; and *Model 4)* both π_t^P and σ_t varies over time.

All models are estimated using Bayesian methods. Prior distributions are not presented here to save space but are available upon request. We use Markov Chain Monte Carlo procedures by combining Metropolis-Hastings and a threshold sampling scheme with an auxiliary variable to draw the posterior distribution of a permanent component of volatility, as described in Laurini, Mauad and Aiube (2016).

2.3 Data

We used the Brazilian monthly inflation measured by the IPCA (*Índice Nacional de Preços ao Consumidor Amplo* - broad national consumer price index) as the observable variable for inflation in the model presented in section 2.2. The period ranges from January 1995 to March 2021, including different levels for the inflation target and variance. The unconditional mean for this sample is 0.55% per month with a variance of 0.18. For the first half of observations, these sample moments are 0.63% and 0.27, while for the second half, they are 0.46% and 0.07

The autocorrelation function and partial autocorrelation function are interesting to determine the number of lags to use in the AR. Figure 2.1 displays these statistics, which motivate us to use an AR in the estimation.

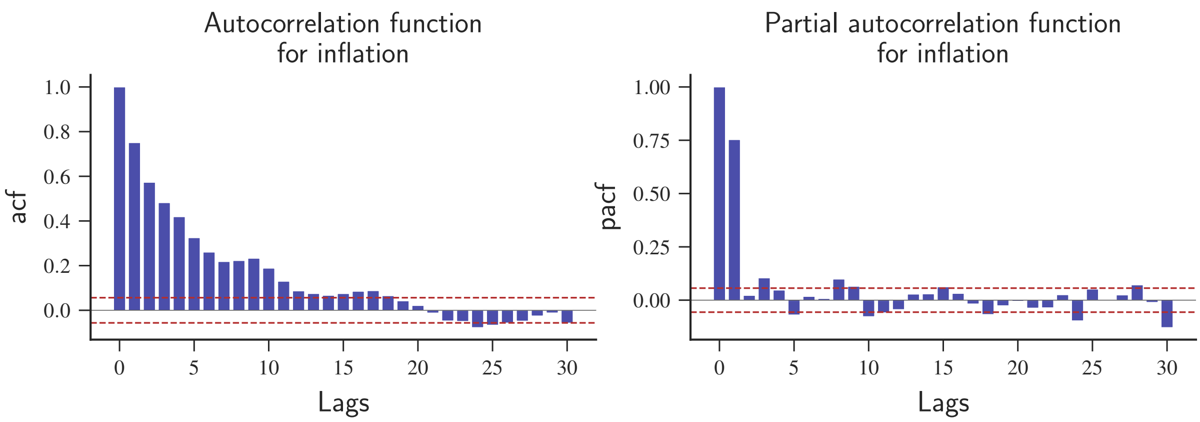

Figure 2.1 – Autocorrelation function (acf) and partial autocorrelation function (pacf).

2.4 Results and Discussion

Table 2.1 summarizes the posterior distribution of the models. Since we choose $k = 1$ lag for the AR for inflation, the parameter φ measures the intrinsic inflation persistence. For the models with a TVIT, we can obtain the expectation-based inflation persistence using the parameter δ.

Table 2.1 – Posterior distribution for all models: 25% quantile, mean and 75% quantile

	Model 1			Model 2			Model 3			Model 4		
	q25%	mean	q75%	q25%	mean	q75%	q25%	mean	q75%	q25%	mean	q75%
φ	0.732	0.759	0.785	0.637	0.669	0.702	0.612	0.65	0.689	0.515	0.556	0.597
π^P	0.484	0.531	0.576	0.43	0.454	0.48	-	-	-	-	-	-
σ_ν	0.28	0.288	0.295	-	-	-	0.271	0.279	0.287	-	-	-
ρ	-	-	-	0.715	0.775	0.846	-	-	-	0.673	0.738	0.81
p	-	-	-	0.016	0.031	0.041	-	-	-	0.013	0.028	0.038
σ_w	-	-	-	0.606	0.967	1.213	-	-	-	0.548	0.969	1.227
σ_ϵ	-	-	-	0.255	0.343	0.428	-	-	-	0.329	0.418	0.507
σ_η	-	-	-	-	-	-	0.048	0.069	0.085	0.045	0.061	0.075
δ	-	-	-	-	-	-	0.116	0.141	0.163	0.112	0.135	0.153

The main result is that the intrinsic inflation persistence reduces when we include the TVIT (Model 2) and SVWJ (Model 3). Considering both TVIT and SVWJ (Model 4), the intrinsic inflation persistence falls drastically from a mean of 0.76 to 0.56. Indeed, the entire posterior distribution of φ shifts to the left when we consider these two characteristics, as illustrated by figure 2.2. This result indicates that both the TVIT and SVWJ affect intrinsic inflation persistence. Thus, neglecting both the time-varying mean and variance can bias the intrinsic inflation persistence. Note that the forward-looking inflation persistence $(1-\delta)$ is almost unaffected by including stochastic variance (see table 2.1).

There are results about inflation persistence in the literature that disregards the time-varying effects of inflation target and volatility. Dossche and Everaert (2005) consider the effect of a TVIT, but their models are homoskedastic. Antonakakis et al. (2016)

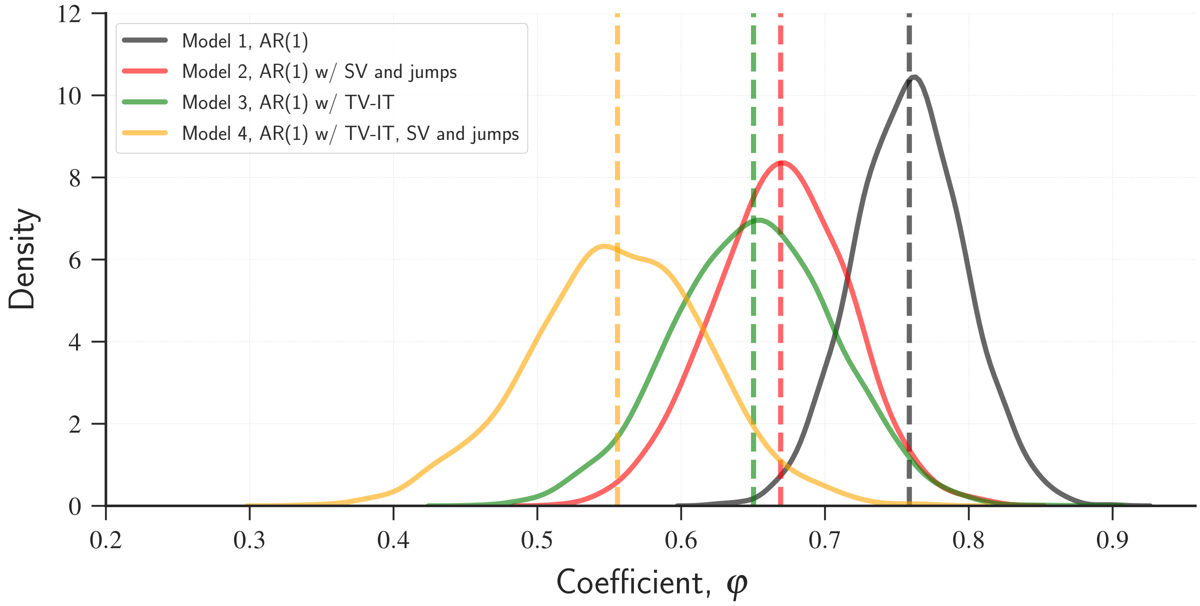

Figure 2.2 – Posterior distribution of autoregressive coefficient for all models.

shows that there is a difference between inflation persistence when measured by online and official price indexes but also ignores the effects of heteroskedasticity. Luengo-Prado, Rao and Sheremirov (2018), which considers sectoral inflation data to estimate inflation persistence, considers only structural breaks in the mean process and also considers homoskedastic errors for inflation. As shown by our results, this homoskedastic assumption for the inflation process may bias the estimation of inflation persistence.

For the Brazilian case, the inflation target perceived by the private agents moves smoothly, as the unobservable component extracted from Model 4 indicates (see figure 2.3). At the beginning of the sample, the perceived inflation target is higher, which is an expected result since the economy was in a hyperinflation process before 1995. After 1999, with the implementation of the inflation target system, the perceived inflation target shows some picks like in mid-2003 and mid-2015, both periods, marked by conturbation in political issues.

The agents' perception of macroeconomic variables has been a relevant topic in the literature. This is the case, for instance, of the work of Krane (2011), Jain (2019) and, more recently, Clements (2021). This line of inquiry uses professional forecasters' data and the revision of their forecasts to identify the agents' perception of the shocks affecting the economy. While Krane (2011) and Clements (2021) concentrate on the agents' view of GDP shocks, Jain (2019) specifically considers the perception of forecasters to build a measure of perceived inflation persistence. Her results indicate that the proposed perceived inflation persistence is well below the inflation persistence of the actual data. The author attributes this difference to the informational rigidity faced by the forecasters.

Our measure of expectation-based inflation persistence is related to the literature on agents' perception of macroeconomic variables since it is related to the perceived

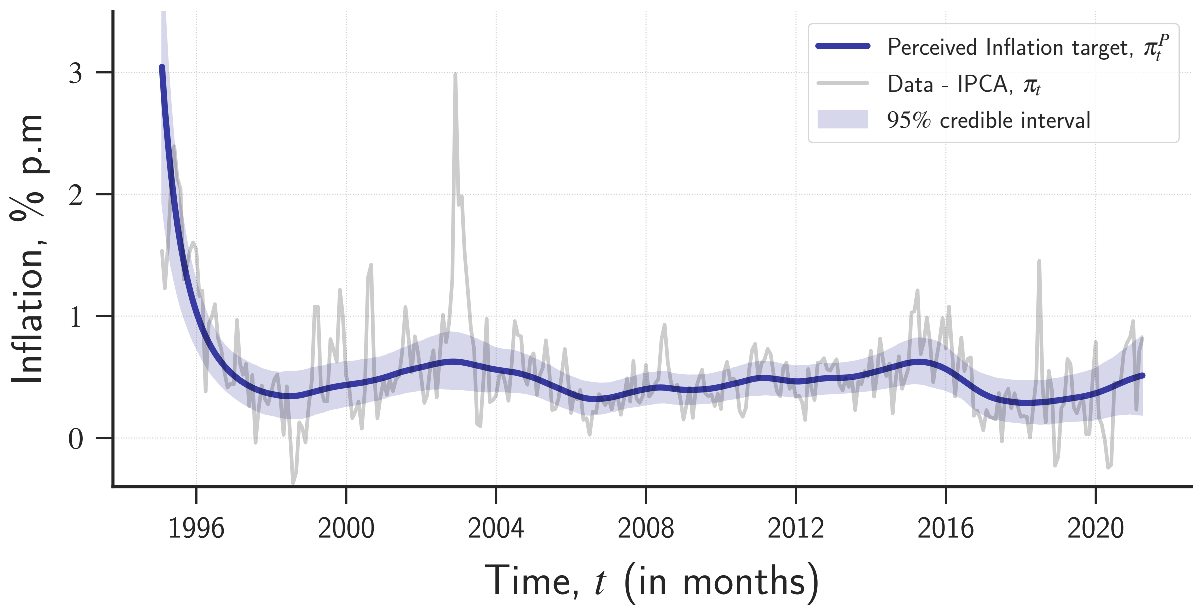

Figure 2.3 – Observed inflation and the unobserved perceived inflation target.

inflation target. The expectation-based inflation persistence considers the presence of information rigidity, as the agents do not have full information about the actual inflation target that the central bank is pursuing. As argued by Dossche and Everaert (2005), this rigidity is similar to those by Mankiw and Reis (2002). The expectation-based inflation persistence, however, cannot be directly compared to the perceived inflation persistence proposed by Jain (2019) because it depends on the perceived inflation target, which is a lower frequency time series than inflation forecasts. Finally, while our approach considers the general perceptions of agents in the economy, Jain (2019) considers only the perception of professional forecasts. Thus, a natural extension for future research is to apply our decomposition to professional forecasters' data.

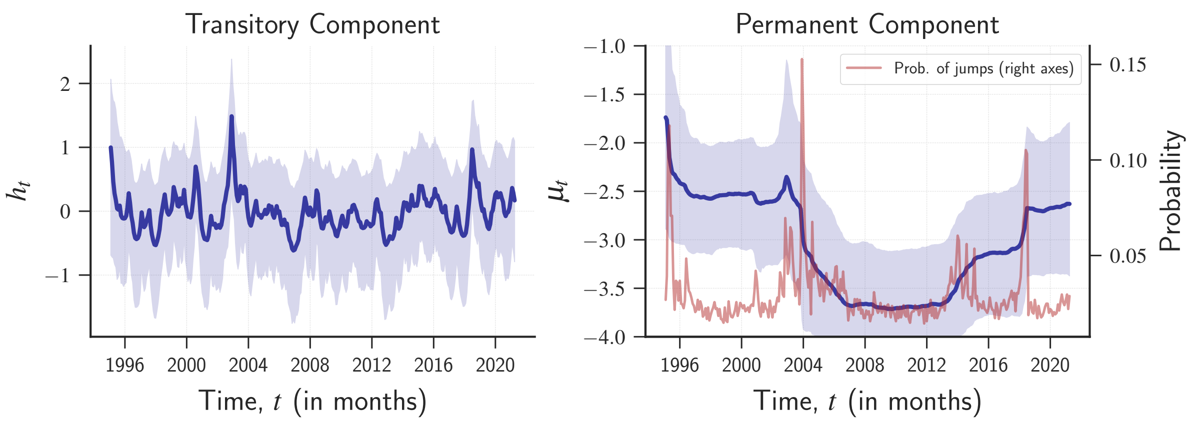

Figure 2.4 – Transitory and permanent component of stochastic volatility.

The stochastic volatility extracted from Model 4 confirms the effects of time-varying variance for both transitory (left panel) and permanent components (right panel), see figure 2.4. The right axis of the left panel of 2.4 measures the probability of the jumps that occur at each period.

There are three probability peaks. The first occurs in April 1995, which bring the volatility to a lower level, and is possibly associated with the stabilization plan adopted in the previous year. The second jump was in November 2003, and the permanent component of volatility also decreases. In mid-2002 and mid-2003, Brazil experienced a confidence crisis triggered by the presidential election. Since the victorious candidate continued to follow the policies initiated by the previous government, the permanent component volatility has decreased and remained at a low level until the next jump in May 2018.

2.5 Conclusion

We have assessed if TVIT and SVWJ affect inflation persistence. We find that including these characteristics reduces intrinsic inflation persistence. Moreover, results indicate the importance of including jumps in stochastic volatility.

Neglecting the time variation in inflation target and inflation volatility with jumps results in an upward-biased estimation of persistence. Thus, including these characteristics to model inflation persistence results in a better measure of inflation persistence. These results are potentially relevant to the inflation literature since measuring inflation persistence and measuring the magnitude of shocks are fundamental to guiding policymakers' decisions.

Future research could apply the decomposition proposed in this paper to other economies. Since our model does not make any specific assumption about the Brazilian economy, it can be used without modifications for other countries, especially for those whose current inflation is accelerating and with higher volatility.

2.6 Appendix

2.6.1 Bayesian diagnostics

All parameters of the model were estimated using MCMC methods and we have to assess the convergence. As latent processes such as h_t and π_t^P increase with the sample size, the number of parameters to be assessed is large for some of the models. Thus, we present here the potential scale reduction factor (PSRF)proposed by (GELMAN; RUBIN, 1992) in a graphical approach (except for Model 1, in which the number of parameters is small). Figure 2.5 presents the frequency of the PSRF for two chains for Models 2, 3, and 4.

In the figure above we can note that the maximum PSRF was smaller than 1.05 and for all models the greater frequency is around 1.00 and 1.001, indicating convergence

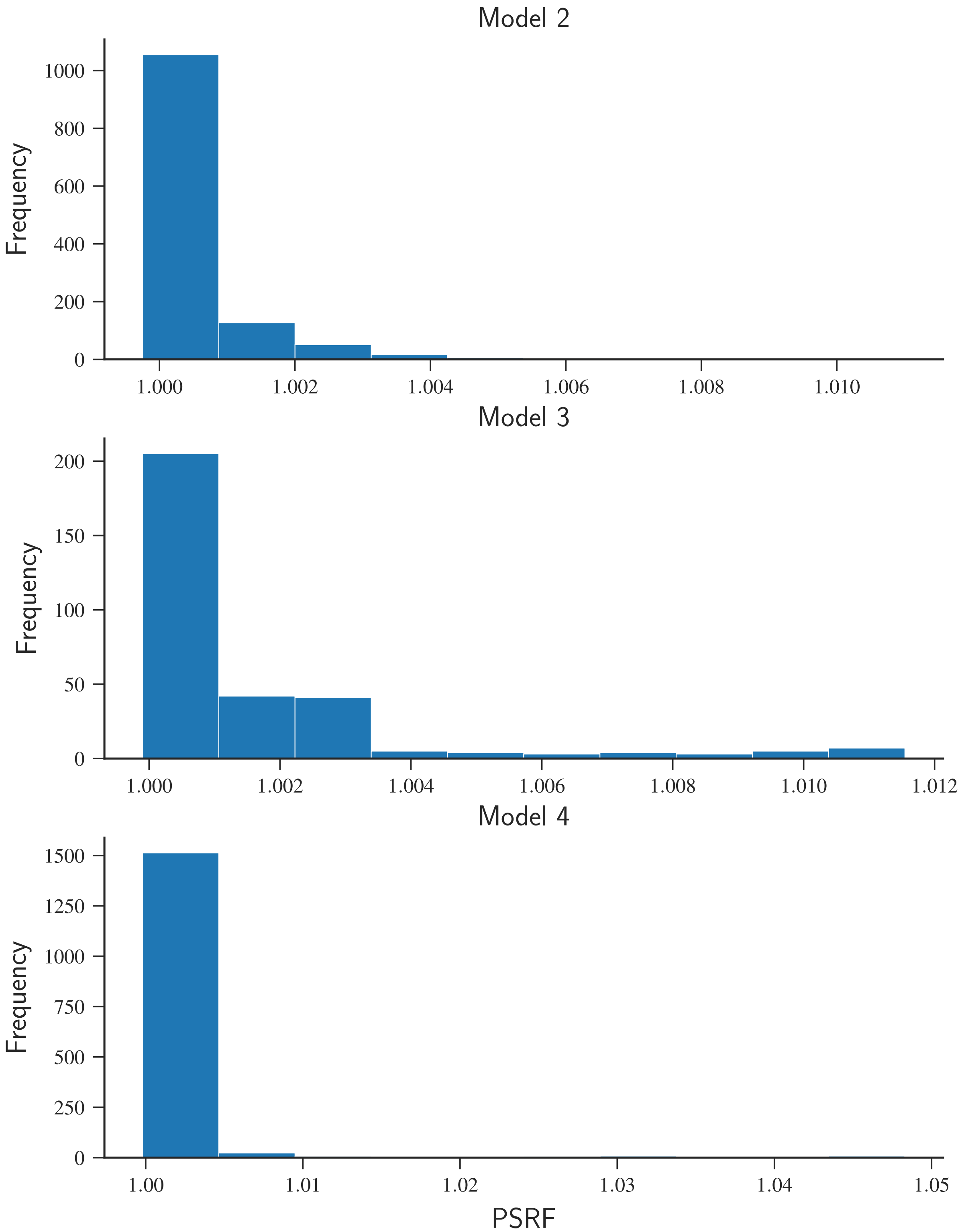

Figure 2.5 – Potential Scale Reduction Factor for Model 2, 3, and 4

of the chains.

3 Can Brazilian Central Bank Communication help to predict the yield curve?

Abstract

This paper investigates whether Brazilian Central Bank communication helps to forecast the yield curve. Our forecast strategy involves two steps: First, we analyze textual Central Bank documents to extract sentiment variables that describe its communication, and then, we include those sentiment variables as additional factors into the dynamic Nelson-Siegel term structure model. We found that sentiment variables contain predictive information for yield curve forecasting. Specifically, when combined with macroeconomic variables, the sentiment variables improve the accuracy of the forecast for short maturities and forecast horizons. In addition, sentiment variables are useful in forecasting for medium and long forecast horizons for all maturities. Besides finding a new source of information to forecast the yield curve, the results indicate that the information provided by Central Bank affects market participants, proving to be a useful tool for monetary policy.

Keywords: Yield Curve, Sentiment Analysis, Bayesian Estimation, Central Bank Communication.

3.1 Introduction

Over the last decades, Central Bank communication has been a relevant theme in the interest rate literature. Central bankers use textual reports to explain their decisions and expectations to the private sector. These communications represent a tool for managing the expectations that link the short-term interest rate, which the Central Bank controls, to the entire term structure of interest rates (i. e., the yield curve), which guides economic decisions (WOODFORD, 2001). For this reason, the yield curve plays a crucial role in the economy, requiring accurate forecasts by financial and macroeconomists. There is in-sample evidence for the relation between the Central Bank communication and the yield curve (BOUKUS; ROSENBERG, 2006; LUCCA; TREBBI, 2009; CHAGUE et al., 2015). This relation is still little explored for out-of-sample forecast analysis, although the Central Bank's words may contain new information about the future path of the yield curve.

This paper proposes to use Central Bank communication to enhance the information set when forecasting the yield curve. We provide out-of-sample evidence that the Central Bank communication contains predictive information for yield curve forecasting. The innovation of our approach is to process the textual content of official documents published by the Central Bank and use it as a new factor in the widely used Dynamic Nelson-Siegel (DNS) model of Diebold, Rudebusch and Aruoba (2006a). Specifically, we use sentiment analysis to classify the content of those documents that the Central Bank uses to communicate with the private sector and construct sentiment variables using this classification to extend the Dynamic Nelson-Siegel model. Then, we use the extended model to forecast the yield curve, providing out-of-sample evidence for the connection between the Central Bank communication and the yield curve.

Our forecast approach produces accuracy gains for all forecast horizons. In particular, for long forecast horizons, the model with sentiment variables improves the accuracy of the traditional DNS model. It also beats the random walk benchmark, a competitive yield curve forecaster. At the twelve-month horizon, the forecasts of this model are $10\% - 18\%$ more accurate than forecasts of the random walk benchmark. For other forecast horizons, including Central Bank communication in the information set improves the accuracy for every maturity relative to the traditional DNS. Moreover, when combined with macroeconomic factors, the model with sentiment variables can beat the benchmark random walk for the one-step horizon and short maturities.

The information incorporated in the Central Bank communication is crucial for understanding the improvement in forecast accuracy. This communication with the private sector presents a wide range of information. It informs, for instance, the short-term interest rate decisions, describes the Central Bank vision of the macroeconomic outlook, and possibly announce some guidance for future decisions on the short-term interest rate. Market participants consume this type of information when allocating their assets, which affects the prices of bonds and, consequently, the term structure of interest rate. The use of the content published by the Central Bank is especially relevant when markets are being affected by an exogenous shock, such as the covid-19 pandemic. Given the exogenous shocks, private agents pay attention to the Central Bank's assessment of the economic situation.

Our research is related and contributes to two strands of the literature. First, we contribute to the literature that attempts to identify good predictors for the yield curve. The literature has incorporate both unobservable (LITTERMAN; SCHEINKMAN, 1991; DIEBOLD; RUDEBUSCH; ARUOBA, 2006a) and observable factors (ANG; PIAZZESI, 2003; HÖRDAHL; TRISTANI; VESTIN, 2006; VIEIRA; FERNANDES; CHAGUE, 2017) to forecast the yield curve. The former uses the level, slope, and curvature factors to form the information set, while the latter uses or combines these latent factors with macroeconomic and financial factors. Besides employing these factors, our approach extends the information set to include Central Bank communication and access its power to predict the yield curve. Thus, we contribute to this literature by verifying if Central Bank communication is a good predictor of the yield curve.

The second branch of literature connected to this work is the one that advocates that textual analysis of documents published by the Central Bank is beneficial to understanding the term structure of interest rate (BOUKUS; ROSENBERG, 2006; LUCCA; TREBBI, 2009; CHAGUE et al., 2015; MÁTÉ; SEBŐK; BARCZIKAY, 2021). The explanation for using textual analysis in this context is that agents extract information from the Central Bank communications and use it when pricing assets. In this regard, the sentiment analysis of Central Bank documents may explain the behavior of the yield curve. Some researchers argue that the Central Bank communication affects short maturities of the yield curve (BOUKUS; ROSENBERG, 2006; CHAGUE et al., 2015), while some authors claim that long maturities (LAMLA; LEIN, 2011) of the yield curve are affected. These studies present only in-sample evidence, while our investigation presents an out-of-sample analysis of the effects of Central Bank communication over each maturity of the yield curve.

In short, our evidence shows that Brazilian Central Bank communication helps to predict the yield curve for every maturity, at least concerning the traditional DNS model. We claim that this improvement is due to the information content about the fundamentals of the economy.

The rest of the paper is organized as follows. Section 3.2 presents how we analyze Brazilian Central Bank communication through sentiment analysis and how we extend the DNS approach to include Central Bank communication in the forecast analysis. In Section 3.3, we present the results and discuss it in Section 3.4. Finally, in Section 3.5, we provide the conclusion.

3.2 Sentiment analysis and yield curve modelling

Our strategy to forecast the yield curve has two steps. First, we quantify the Brazilian Central Bank communication by extracting the sentiment of the documents published after the monetary policy committee's meeting. Second, we use the sentiment and macroeconomic variables in a Dynamic Nelson-Siegel model to make direct forecasts for the term structure of interest rate. In the following subsections, we explain these two steps and how we evaluate the performance of forecasts.

3.2.1 Quantifying the Central Bank Communication

The Monetary Policy Committee (COPOM - *Comitê de Política Monetária* - Monetary Policy Committee in Brazilian Portuguese) meetings of Central Bank of Brazil represent a relevant source of information about monetary policy decisions. The content of these meetings is summarized in text format in two documents: the COPOM statements and the COPOM minutes. The former is short and only informs the decisions of the meetings, while the latest describes the economic environment in detail. Both documents reflect the viewpoints of the Central Bank about the economic outlook, and market participants closely monitor these views. Thus, to capture the Central Bank's perspective, we examine the textual content of these documents.

We use sentiment analysis to extract information from the two documents published by the Central Bank. First, we collect raw text data and treat it appropriately. After collecting the raw data, for each meeting, we remove punctuation, blank lines, stop words, and the name of the participants of the monetary policy committee, to exclude words that should not affect the sentiment of the text. It is worth noting that the Brazilian Central Bank started to publicize the COPOM statements and minutes on regular dates[1] only from 2006, which defines the starting year of our sample in this paper. Second, we measure the sentiment of the Central Bank of Brazil from this treated text data by classifying the content of these documents according to their semantic categories. We point out that these preprocessing steps are not exclusive to financial studies but are universal in textual analysis.

We base our sentiment analysis of the COPOM statements and minutes on a dictionary-oriented approach. The vocabulary of these documents is very specialized, and the words used in these documents have specific meanings, usually related to financial meaning. For this reason, we use the financial dictionary proposed by Loughran and McDonald (2011) to categorize the words of the documents. This dictionary allows us to identify the semantic content of each word in the text, separating them into positive, negative, and uncertain.

[1] Since 2006, the Central Bank publishes the COPOM statements and minutes every 45 days.

We build the sentiment variables representing Central Bank communication using these three semantic categorizations. For both COPOM statements and COPOM minutes, we consider as sentiment variables the proportion of each category to the total number of words of each document. Formally, for each period t that the COPOM meeting takes place, we define the sentiment variable as follows:

$$s_{\ell,t} \equiv \frac{\#\text{ of words of category } \ell}{\#\text{ of total words}}, \tag{3.1}$$

where $\ell \in \{\text{positive}, \text{negative}, \text{uncertain}\}$. Additionally, for the COPOM minutes, which have a large number of words, we follow Cannon (2015) and define the tone of the document as

$$\tau_t \equiv \frac{\#\text{ of positive words} - \#\text{ of negative words}}{\#\text{ of positive words} + \#\text{ of negative words}}, \tag{3.2}$$

summarizing the sentiment of the Central Bank of Brazil in only one variable.

The interpretation of the tone variable is straightforward: If $\tau_t > 0$, then the document has a overall positive tone; if $\tau_t < 0$, then it has a negative tone; otherwise, the discourse is neutral. To capture the change in the tone of Central Bank, we use the first difference of the tone, that is, $\Delta\tau_t \equiv \tau_t - \tau_{t-1}$. The tone variable was not computed for the COPOM statements because the shortness of these documents implies that for some periods, the sum in the denominator of Equation (3.1) is equal to zero, which is not well defined.

3.2.2 Forecasting the yield curve

3.2.2.1 The dynamic Nelson-Siegel model

The seminal paper of Nelson and Siegel (1987) introduces a preeminent class of models for the yield curve. These authors establish a parsimonious model for the term structure of interest rate that relies on three latent factors, which are interpreted as level, slope, and curvature of the yield curve and are allowed to vary over time in its dynamic version (DIEBOLD; RUDEBUSCH; ARUOBA, 2006a). Besides the latent factor, the literature has identified observable factors that may help to predict the yield curve, such as macroeconomic variables (see Ang and Piazzesi (2003) and Hördahl, Tristani and Vestin (2006), for instance). Given the ability to deal with unobservable and observable factors, we build our model based on the Dynamic Nelson-Siegel (DNS) approach, augmenting it with macroeconomic and sentiment variables.

The DNS decomposition of the yield curve allows us to make a prediction of the entire yield curve based on the level, slope, and curvature latent factors, denoted by $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$, respectively. The yield at maturity m and period t may be approximated by

$$y_t(m) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda m}}{\lambda m}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right) + \eta_t, \tag{3.3}$$

where η_t is normally distributed with zero mean and variance given by σ_η^2.

Since $\beta_{i,t}$, for $i \in \{1, 2, 3\}$, are unobservable, the challenge is to find the best possible estimate of these latent factors to predict the yield curve accurately. In addition to the autoregressive structure proposed by Diebold, Rudebusch and Aruoba (2006a), we include both sentiment and macroeconomic variables as explanatory variables for the latent factors. Following Diebold, Rudebusch and Aruoba (2006b), we use three macro variables: inflation, π_t, the capacity of utilization CU_t and Central Bank interest rate target (Selic Rate), r_t, that are collected in the (3×1) vector X_t. Denoting the vector of sentiment variables by S_t, which may include $s_{\ell,t}$ and/or $\Delta\tau_t$, the general structure of β_i is specified as follow:

$$\beta_{i,t} = c_i + \phi_i\beta_{i,t-h} + \sum_{j=1}^{J}\gamma_j X_{j,t-h} + \sum_{k=1}^{K}\alpha_k S_{k,t-h} + \varepsilon_{i,t}, \tag{3.4}$$

where $\varepsilon_{i,t}$ is normally distributed with zero mean and variance given by σ_i^2, and h is the horizon forecast.

Based on the general model in Equation (3.4), we consider eight competing specifications, which differ depending on the variables included in S_t, and whether macro and sentiment variables are considered. Table 3.1 summarizes these models. The first column presents the abbreviated name of each model, and the other two columns show the associated variable for each specification.

Table 3.1 – Competing specifications used to forecast yield curve

	Models	
Abbreviated model name	S_t	X_t
DNS-only	Not included	Not included
DNS-macro	Not included	π_t, CU_t, r_t
DNS-macro-state	s_ℓ from statements	π_t, CU_t, r_t
DNS-macro-min	s_ℓ from minutes	π_t, CU_t, r_t
DNS-macro-tone	$\Delta\tau_t$ from minutes	π_t, CU_t, r_t
DNS-tone	$\Delta\tau_t$ from minutes	Not included
DNS-min	s_ℓ from minutes	Not included
DNS-state	s_ℓ from state	Not included

We can estimate the model exposed in Equations (3.3) and (3.4) using both the Classical and Bayesian techniques. In the Classical approach, we can apply the maximum likelihood estimator by using the prediction error decomposition obtained from Kalman

Filter, as proposed by Diebold, Rudebusch and Aruoba (2006b). A drawback of this approach is that the numerical maximization of the likelihood function may be problematic. To overcome this limitation, some authors have advocated the Bayesian estimation for this class of models (LAURINI; HOTTA, 2010; HAUTSCH; YANG, 2012). More specifically, Equations (3.3) and (3.4) together with prior distributions[2] for the parameters form a hierarchical model that can be estimated using Bayesian techniques. Therefore, we use the Bayesian approach to simulate the posterior distribution of each parameter of this hierarchical model.

The Bayesian estimation of the DNS requires Markov Chain Monte Carlo (MCMC) methods since the posterior distributions of parameters do not have a closed formula. Following Batista and Laurini (2016), we use the Hamiltonian Monte Carlo (HMC) method to simulate the posterior distribution of each parameter. Batista and Laurini (2016) showed that the HMC is a viable method to estimate models of DNS class presenting faster convergence than the standard Metropolis-Hastings algorithm. Specifically, we use the HMC and the No-U-Turn sampling, proposed by Hoffman and Gelman (2014) and implemented in 'Rstan' package. In the appendix, we briefly describe how HMC works, but more details can be found in Neal et al. (2011) and Hoffman and Gelman (2014).

3.2.2.2 Forecasting and Performance evaluation

The yield curve forecast is built using the draws from the posterior of the parameters of the model. Let τ be the last period used in the estimation, and $h \in \{1, 6, 12\}$ be the forecast horizon. The predictive density for a yield of maturity m uses the $\mathcal{S}$-th draw from the posterior of $\lambda, \beta_{i,\tau}$, for $i \in \{1, 2, 3\}$, to build:

$$\hat{y}^{(\mathcal{S})}_{\tau+h|\tau}(m) = \beta^{(\mathcal{S})}_{1,\tau} + \beta^{(\mathcal{S})}_{2,\tau}\left(\frac{1-e^{-\lambda^{(\mathcal{S})}m}}{\lambda^{(\mathcal{S})}m}\right) + \beta^{(\mathcal{S})}_{3,\tau}\left(\frac{1-e^{-\lambda^{(\mathcal{S})}m}}{\lambda^{(\mathcal{S})}m} - e^{-\lambda^{(\mathcal{S})}m}\right), \tag{3.5}$$

where $\hat{y}^{(\mathcal{S})}_{t+h|t}(m)$ denotes predictive value of the h-steps-ahead forecast of the yield of maturity m, using the $\mathcal{S}$-th draw from the latent factors, $\beta^{(\mathcal{S})}_{i,\tau}$, and from the decay parameter, $\lambda^{(\mathcal{S})}$. We use the mean of $\{\hat{y}^{(\mathcal{S})}_{\tau+h|\tau}(m)\}_{\mathcal{S}=1}^{N_{\text{iter}}}$ as a point forecast for the yields, where N_{iter} is the number of iterations of the Hamiltonian Monte Carlo.

We perform out-of-sample forecasts for the last 24 periods of our sample. We use a rolling window scheme as follows. We begin by estimating a specification considering the first window with size $K = 156$. Then, we sequentially re-estimate the specification, including the next month and excluding the first month of the previous estimation window. We repeat this process until we exhaust the entire sample. Finally, we compare the forecast of each specification to the actual data to evaluate the performance.

[2] The prior distributions used in this paper are briefly described in appendix.

To analyze the performance of forecasts, we use the root of mean squared forecast error (RMSE) as a loss function. Given the maturity and forecast horizon, this loss function was computed for each model l as follows:

$$\text{RMSE}_l(m,h) = \sqrt{\frac{1}{24}\sum_{t=1}^{24}\left[\hat{y}_{t+h|t,l}(m) - y_{t+h}(m)\right]^2}. \tag{3.6}$$

The model that possesses the minimum RMSE is the one with the best performance.

Although extensively used in the forecasting literature, the RMSE only offers a measure of accuracy for the entire sample. To overcome this drawback, we also analyze the cumulative squared forecast error (CSFE), proposed by Welch and Goyal (2008). The CSFE requires the choice of a benchmark model. We choose the Random Walk (RW) model as a benchmark since it usually produces good forecast for the yield curve. Thus, the CSFE is defined as:

$$\text{CSFE}_{l,T}(m,h) = \sum_{t=1}^{T}\left\{\left[\hat{y}_{t+h|t,\text{RW}}(m) - y_{t+h}(m)\right]^2 - \left[\hat{y}_{t+h|t,l}(m) - y_{t+h}(m)\right]^2\right\}, \tag{3.7}$$

for each model l, maturity m and forecast horizon h. Equation (3.7) allows us to evaluate the performance of a model for all periods in the forecast window. When CSFE is increasing, it indicates outperformance of the benchmark, and a decrease in CSFE indicates that the model underperforms the benchmark.

Additionally, we apply the Model Confidence Set (MCS) procedure, proposed by Hansen, Lunde and Nason (2011), to verify whether the differences encountered in the loss functions among models were indeed significant. This procedure allows us to construct a set of superior models from the set of all alternative models, given a level of significance. Formally, let $d_{ij,t}$ denotes the difference between the loss function of two models i and j and let $\mathcal{M}_0$ denotes the set of all models. Starting from the set of all models, that is, $\mathcal{M} = \mathcal{M}_0$, the MCS procedure consists in sequentially testing the null hypothesis:

$$H_0 : \mu_{ij} = 0, \forall i, j \in \mathcal{M}, \tag{3.8}$$

where $\mu_{ij} = \mathbb{E}[d_{ij,t}]$. If the null hypothesis is rejected, the model is excluded from $\mathcal{M}$; otherwise, the test continues until we cannot reject the null hypothesis, given a level of significance α. In this manner, the procedure returns the set of superior models $\mathcal{M}^*_{1-\alpha}$.

The test statistic associated with (3.8) is constructed as follows. For each model $i, j \in \mathcal{M}$, compute

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\hat{var}(\bar{d}_{ij})}}, \tag{3.9}$$

where $\bar{d}_{ij} = (1/M)\sum_{j=1}^{M} d_{ij}$. We compute the variance that appears in the denominator in (3.9) through block bootstrap. Then, the test statistic can be calculated using $T_R = \max_{i,j\in\mathcal{M}} |t_{ij}|$. The asymptotic distribution of this statistic is nonstandard, and we also estimate it via bootstrap.

3.3 Empirical analysis

This section empirically assesses the accuracy of the alternative specifications discussed above. We begin by describing the data on bond yields used to estimate and forecast the yield curve. Then, we present the extracted Central Bank sentiment variables and explain their relationship with the term structure of the interest rate. We then give the forecast accuracy findings for the various models discussed in the section before (see Table 3.1).

3.3.1 Data description

To investigate the out-of-sample performance of the alternative specifications, we use monthly Brazilian Treasury yields data from January 2006 to December 2020, totaling $T = 180$ observations. The data were taken from the B3, former BM&F-Bovespa, website[3], and interpolated using a smoothing spline. The maturities of the yields are 1, 2, 3, 6, 9, 12, 24, and 36 months. The forecast window starts in January 2019 and ends in December 2020. Figure 3.1 illustrates the monthly yield data used to produce the forecasts.

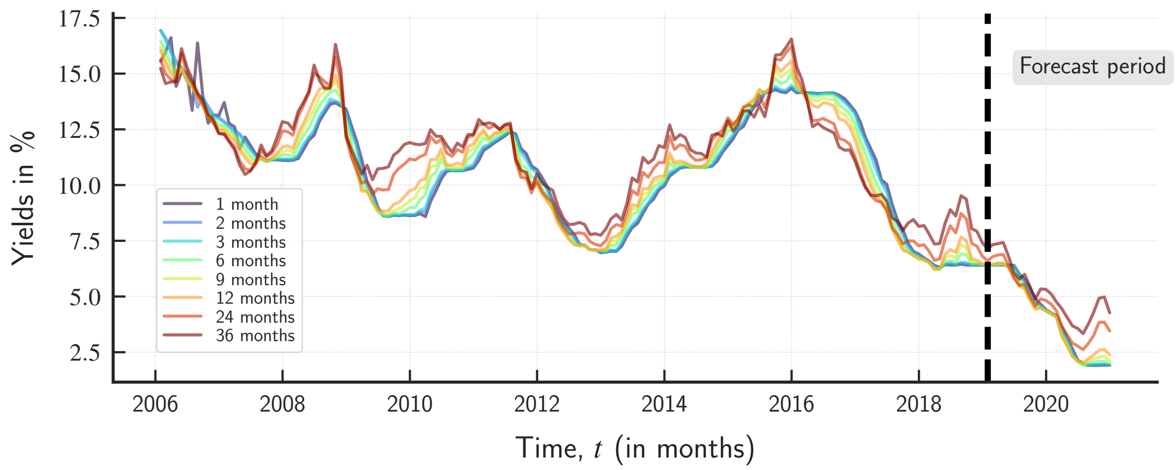

Figure 3.1 – Brazilian Treasury yields, by maturity, from January 2006 to December 2020. The dashed vertical line indicates the first period of the forecast window.

[3] http://www2.bmf.com.br/pages/portal/bmfbovespa/lumis/lum-taxas-referenciais-bmf-ptBR.asp

Maturity (in months)	Mean	Std	Min	25%	50%	75%	Max	$\rho(1)$	$\rho(6)$	$\rho(12)$
1	9.97	3.36	1.91	7.28	10.70	12.51	16.61	0.97	0.73	0.43
2	9.98	3.38	1.92	7.30	10.70	12.48	16.94	0.97	0.73	0.43
3	9.97	3.39	1.93	7.26	10.66	12.41	16.93	0.97	0.73	0.43
6	9.98	3.39	1.92	7.28	10.73	12.46	16.46	0.97	0.74	0.43
9	10.01	3.38	1.94	7.32	10.79	12.53	16.19	0.97	0.74	0.44
12	10.06	3.36	2.03	7.45	10.82	12.50	16.03	0.97	0.74	0.44
24	10.36	3.21	2.61	7.86	11.08	12.67	16.22	0.97	0.73	0.44
36	10.68	3.00	3.33	8.45	11.38	12.64	16.56	0.96	0.72	0.43

Table 3.2 – Descriptive statistics for yield curves. The last three columns represents the sample autocorrelations of order 1, 6 and 12.

The term structure of interest rate presents periods of decreasing and increasing along with the data range for every maturity. However, in general, the average behavior of the interest rates shows a decreasing tendency. Specifically, the forecast period presents the lowest interest rates in the sample, as illustrated in Figure 3.1. In Table 3.2, we present descriptive statistics for the yields for the entire sample. Three points are worth noting about these statistics: First, the typical yield curve is upward sloping; second, in general, longer maturities have lower volatilities; and third, the persistence is similar across the maturities.

We also use macroeconomic and sentiment variables in the estimation. Regarding macro variables, we use the IPCA (*Índice Nacional de Preços ao Consumidor Amplo* - Broad national consumer price index) as the inflation rate, π_t, the SELIC (*Sistema Especial de Liquidação e de Custódia* - Special System for Settlement and Custody) interest rate as the monetary policy interest rate, r_t, and the capacity utilization of the industry denoted by CU_t. Except for SELIC, the macroeconomic variables are available only in monthly data, which defines the frequency of our model. These three macroeconomic variables are available on the Ipeadata website[4]. Regarding the sentiment variable that reflects Central Bank communication, we describe it separately in the next section.

3.3.2 The Central Bank communication

To measure the Central Bank communication, we extract the sentiment of the Brazilian Central Bank using textual analysis tools, as described in Section 2.1. We construct the sentiment variables for the days when COPOM statements or minutes were published. For both documents, the first characteristic identified by the analysis of textual data was a change in the behavior of Central Bank communication in the last decade, as shown in Figure 3.2.

[4] www.ipeadata.gov.br/

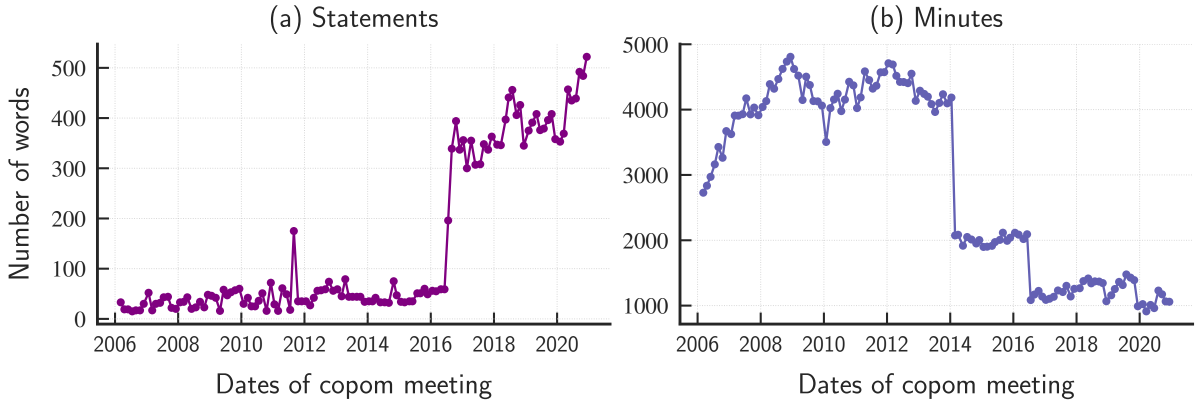

Figure 3.2 – Sentiment of Central Bank of Brazil. Left panel shows the sentiment extracted from COPOM statement and right panel shows the sentiment extracted from COPOM minutes.

The main change in Central Bank communication is related to a Central Bank's chair change in June 2016, which began to publish more detailed COPOM statements. In effect, the average number of words before June 2016 was 42 word per meeting; after that, the average increased to 382. This result indicates an effort by the Central Bank to increase the transparency of its actions. While the COPOM statements present an increase in the number of words, this number decreases for the COPOM minutes along the sample period (see Panel (b) of Figure 3.2). In this case, there are two breaking points in the number of words. The first breaking point happened in January 2014, and the second in June 2016. The latter occurs for the same reason as the COPOM statement, but, in this case, the number of words decreases. This evidence shows a tendency of the Central Bank to prefer the COPOM statements instead of the COPOM minutes to communicate with the private sector.

We now present the time series of the sentiment variables used in the estimation. Panel (a) of Figure 3.3 shows the behavior of these sentiment variables for each day that the COPOM meeting takes place. It displays the proportion of negative, positive, and uncertainty words extracted from COPOM statements. Until 2016, the proportions of words had values equal to zero for several periods. From this year onward, the variables were always positive. This behavior occurs because the small number of words in COPOM statements before June 2016, as shown before, implies that any of the words match the Loughran and McDonald (2011) dictionary.

The sentiment variables extracted from COPOM minutes, in turn, always presented values greater than zero for the proportion of words, as shown by panel (b) of Figure 3.3. In this case, the change previously observed in 2016 is not so evident in the proportion of words, except that the fraction of uncertain words grows from that date. Note, however, that after June 2016, the behavior of the positive, negative, and uncertain words are very similar when compared to the result of the proportions found using COPOM statement

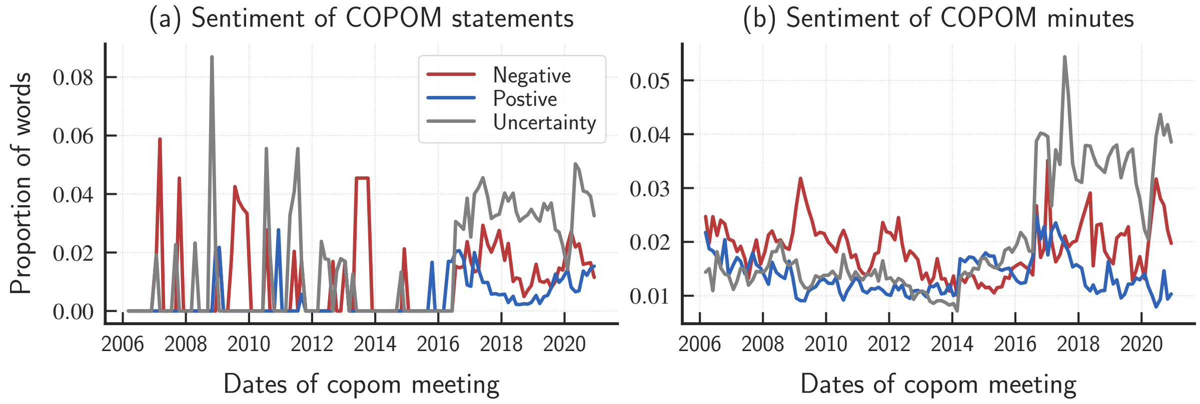

Figure 3.3 – Sentiment of Central Bank of Brazil. The left panel shows the sentiment extracted from COPOM statement, and the right panel shows the sentiment extracted from COPOM minutes.

(see Panel (a) of Figure 3.3).

An advantage of the sentiment extracted from COPOM minutes is that we can compute the tone of the Central Bank (as defined in Equation (3.2)) since the sum of positive and negative words is always strictly positive. Figure 3.4 plots the tone of the Central Bank together with short-, medium-, and long-term interest rates (1, 9, and 36 months, respectively). First, we note that for almost all periods, the Central Bank's tone was negative. Second, the tone seems to move in the same direction as yields, independent of maturity. Its correlation with the various maturity interest rate confirms this statement, as shown in Figure 3.4.

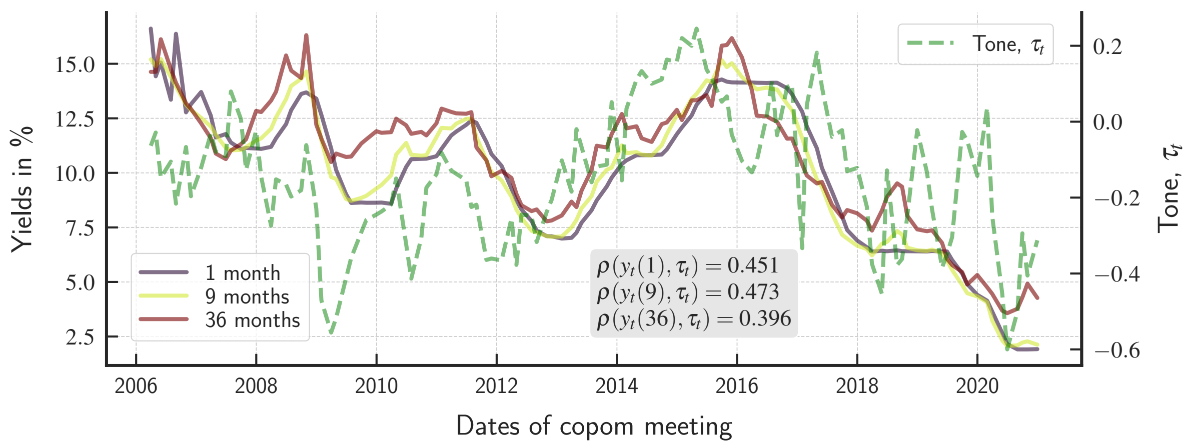

Figure 3.4 – Tone of Brazilian Central Bank and yields. We select short, medium, and long maturities of yields to compare with the tone. $\rho(y_t(m), \tau_t)$ indicates the correlation.

For the days that there was no COPOM meeting, and consequently no Central Bank document publication, we set all sentiment variables to zero and build a series in daily frequency[5]. Then, we aggregate these series to obtain the monthly frequency by

[5] We also build the sentiment variable repeating the value of the variable in the last meeting for dates

the maximum value of each month. This step was necessary since the macroeconomic variables are available only in monthly frequency.

3.3.3 Out-of-sample forecasts

Table 3.3 presents the accuracy of the forecasts for all models and all forecast horizons. Besides the models explained in Section 3.2, we also compare our forecasts with the Random Walk (RW) model since there is evidence that the RW model is competitive, especially for short forecasting horizons (JOSLIN; SINGLETON; ZHU, 2011). As we use the RW as a benchmark, only the column 'RW' of Table 3.3 presents the RMSE (in percentage points), while the remaining column displays the root of the mean squared error of each model relative to the RW benchmark. Then, values less than one indicate that the model outperforms the benchmark, and values greater than one indicate underperformance. Additionally, Table 3.3 also highlights in bold-type the outperformance model in maturities.

We first note that the RW confirm to be a challenging competitor. In contrast to the findings of Diebold, Rudebusch and Aruoba (2006a) and, Vicente and Tabak (2008), which find that the DNS-only produces a better forecast at longer horizons, the RW outperforms the DNS-only for every maturity and forecast horizons in our dataset. We are not the first to find such a result. Caldeira and Torrent (2017), for example, found that the DNS-only outperformed the RW only for maturities longer than 36 months for the US yield curve, which is the longer maturity considered here. However, the performance concerning the benchmark RW varies with maturities and forecast horizons when we add macro or sentiment variables in the DNS model.

Including macroeconomic variables substantially improves the predictive performance for short forecast horizons and maturities. Considering maturities of up to three months and the one-month horizon forecast, the DNS-macro outperforms the benchmark. Including the sentiment of the Central Bank together with macroeconomic variables makes the accuracy even better. In particular, the DNS-macro-state is the best forecaster for maturities up to three months. It is worth noting that for six months of maturity, the DNS-macro-tone specification still outperforms the benchmark.

At the six-month horizon, the RW benchmark is the best model for all maturities. However, the macroeconomic and sentiment variables also help the traditional DNS model to make better forecasts. Comparing the DNS-only with the DNS-macro, we observe that adding macro factors increases accuracy for maturities up to six months. Moreover, except for the DNS-macro-tone specification, using the sentiment and macro variables jointly improves the accuracy even when compared to the DNS-macro for maturities up to six months. The two last columns of Table 3.3 also report that, for the six-step-ahead,

that the meeting does not take place. The out-of-sample result is very similar, as shown in the appendix.

Maturity (in months)	RW	DNS-only	DNS-macro	DNS-macro-state	DNS-macro-min	DNS-macro-tone	DNS-tone	DNS-min	DNS-state
				1-month forecast horizon					
1	0.272*	2.508	0.705*	**0.684***	0.825*	0.699*	2.531	1.967	2.260
2	0.274*	2.387	0.771*	**0.752***	0.903*	0.764*	2.409	1.879	2.129
3	0.280*	2.312	0.832*	**0.821***	0.972*	0.825*	2.332	1.847	2.056
6	0.288*	2.251	1.000*	1.006*	1.152	**0.996***	2.269	1.845	1.994
9	**0.296***	2.211	1.084*	1.105*	1.228	1.082*	2.227	1.827	1.959
12	**0.306***	2.211	1.118*	1.164*	1.259	1.117*	2.225	1.853	1.973
24	**0.353***	1.977	1.185	1.254	1.303	1.184	1.988	1.73	1.815
36	**0.409***	1.588	1.095	1.170	1.210	1.095	1.598	1.483	1.520
				6-month forecast horizon					
1	**1.401***	1.329	1.153*	1.126*	1.155*	1.559	1.542	1.180*	1.143*
2	**1.398***	1.313	1.175*	1.145*	1.175*	1.602	1.526	1.166*	1.128*
3	**1.397***	1.308	1.205*	1.172*	1.202*	1.650	1.521	1.161*	1.124*
6	**1.381***	1.317*	1.306	1.263*	1.290*	1.790	1.533	1.158*	1.129*
9	**1.374***	1.323	1.395	1.343	1.364	1.890	1.540	1.155*	1.135*
12	**1.375***	1.334	1.478	1.419	1.432	1.960	1.549	1.160*	1.149*
24	**1.391***	1.318	1.594	1.519	1.506	1.984	1.514	1.151	1.161
36	**1.405***	1.236	1.531	1.450	1.424	1.831	1.408	1.098*	1.109*
				12-month forecast horizon					
1	2.389*	1.094*	0.984*	1.467	1.501	2.085	1.281	1.067*	**0.892***
2	2.391*	1.083*	1.006*	1.523	1.549	2.109	1.273	1.058*	**0.881***
3	2.396*	1.078*	1.031*	1.577	1.595	2.133	1.270	1.049*	**0.873***
6	2.385*	1.080*	1.115	1.729	1.727	2.211	1.278	1.024*	**0.862***
9	2.373*	1.085*	1.188	1.836	1.820	2.262	1.286	0.996*	**0.856***
12	2.362*	1.095*	1.248	1.91	1.886	2.294	1.297	0.970*	**0.859***
24	2.317*	1.111	1.353	1.955	1.924	2.265	1.299	0.883*	**0.875***
36	2.311*	1.079	1.316	1.797	1.780	2.099	1.239	**0.823***	0.872*

Table 3.3 – RMSE for 1-, 6- and 12-step-ahead prediction. The column RW reports the root of mean squared error (in percentage points) of the random walk benchmark, while the other columns display the root of mean squared error of each model relative to the benchmark. Values in bold indicate outperformance in the maturity. We indicate by * the models in the set of superior models for each maturity at 10% of significance.

the DNS specification augmented by the sentiment of the Central Bank improves the accuracy of the DNS-macro model. In particular, the DNS-state is better than DNS-macro-state for all maturities, except for one-month maturity. These results indicate that Central Bank communication can help to forecast the yield curve for the medium-horizon forecast.

The best performance of the specifications with only sentiment variables (DNS-min and DNS-state specification, for example) occurs in the longer forecast horizon. At the twelve-month horizon, the forecasts of this model are 10% − 18% more accurate than forecasts of the Random Walk benchmark. In this case, although the model with macro variable outperformed the benchmark RW for the shortest maturity, the DNS-state was the best model for all maturities, except the longer one, in which the DNS-min presented the most accurate forecast, outperforming even the RW benchmark. This evidence indicates that Central Bank communication also helps to forecast the yield curve at the long horizon.

Table 3.3 also presents the models included in the Model Confidence Set. First, note that the RW is in the superior set models for all maturities and forecast horizons. Specifications with macro and sentiment variables are also in the set of superior models for short and medium maturities at the one-month horizon. At the six-month horizon, the specification with only sentiment variables is among the best models for almost every maturity. The specifications with sentiment and macro variables are in the set of best models only for maturities shorter than six months. The DNS-min and DNS-state are among the best at the twelve-month horizon, independent of maturity.

Finally, to further investigate the accuracy along the forecast period, we analyze the cumulative squared forecast error. We construct this measure of accuracy using the RW as a benchmark. Here, we present the result only for the six-months horizon, and the appendix shows the other forecast horizons. Figure 3.5 presents the CSFE for $h = 6$.

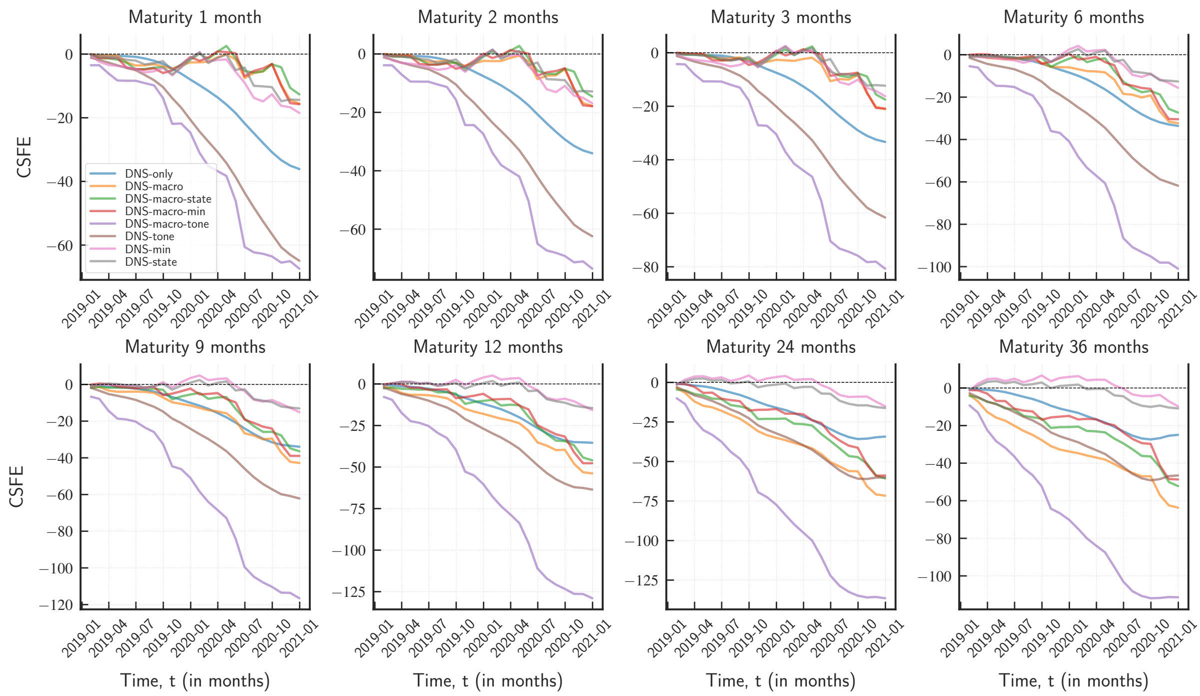

Figure 3.5 – Cumulative squared forecast errors (CSFE) six-step-ahead: An increase in a line indicates better performance of the named variable; a decrease in a line indicates better performance relative to the random walk.

Figure 3.5 shows that the CSFE alternate between periods of positive and negative slope. We observe that between the last months of 2019 and the first months of 2020, there is an increasing trend in the CSFE for various models. However, at the end of the sample, the lines are downward slope, implying a worse performance than the benchmark. This result indicates that the accuracy of each model compared to the RW may vary in time. Also, note that for the DNS-macro-tone and DNS-tone, the lines of the CSFE are decreasing for every period. Thus, these models underperform the benchmark in the entire forecast window.

3.4 Discussion

The main question of this paper, which is whether the Brazilian Central Bank communication can help to predict the yield curve, is answered by our results. The Brazilian Central Bank communication does contain predictive information about the yield curve. The out-of-sample analysis pointed to three main results. First, the DNS specifications augmented by sentiment variables (but without macro) are a good competitor for long horizons. Second, including macroeconomic variables improves the forecast accuracy for the short horizon, and this result agrees with the literature (ANG; PIAZZESI, 2003; HÖRDAHL; TRISTANI; VESTIN, 2006). However, while our results showed an improvement for short forecast horizons, Vereda, Lopes and Fukuda (2008) argues there is no significant improvement for short-term forecasts. Third, besides the macro variables, incorporating the sentiment of the Central Bank of Brazil in the DNS specifications makes the forecasts even better for short horizons.

The finding that Central Bank communication helps to forecast the yield curve corroborates the argument that agents use the information published by the Central Bank to form their expectations and pricing assets. In this regard, our results are consistent with earlier studies linking Central Bank communication with the yield curve, such as Boukus and Rosenberg (2006), Lucca and Trebbi (2009), Chague et al. (2015), Máté, Sebők and Barczikay (2021). While these researches focus on in-sample analysis, our results contribute to the literature by providing out-of-sample evidence of the relation between the yield curve and Central Bank communication.

The reason why Central Bank communication is a good predictor for medium and long horizons is related to the *information effect*, discussed in Romer and Romer (2000) and explored in Nakamura and Steinsson (2018). In the COPOM minutes and statements, the Central Bank informs its vision about the economic fundamentals. The Central Bank announcements lead the private agents to update their beliefs about the monetary policy path and the macroeconomic outlook. The change in agents' beliefs may alter the composition of fixed income portfolios, moving from long to short-term bonds or vice-versa, depending on Central Bank assessment of economic fundamentals. This change will impact the price and yield of the bonds and thus make the Central Bank communication a crucial ingredient to the information set in the yield curve forecast. In addition, Hansen, McMahon and Tong (2019) argues that the news on economic uncertainty affects the yield curve. In our approach, we also consider the proportion of uncertain words in Central Bank communication, which may be information that impacts agents' behavior.

Although our results are promising, they have some limitations which may be addressed in future publications. First, we use a dictionary approach to classify the words of Central Bank documents. This approach has the disadvantage that the chosen dictionary (LOUGHRAN; MCDONALD, 2011) is a financial dictionary rather than a dictionary for macro or monetary policy context. Among the available dictionaries, however, a financial dictionary is the one that is close to the Central Bank documents. Future research can measure Central Bank communication using a dictionary-free approach, using tools such as the Latent Dirichlet Allocation for topic classification. Our strategy, however, is independent of the parametric assumptions implicit in Latent Dirichlet Allocation (BLEI; NG; JORDAN, 2003), which may be an advantage. A further possible advantage to our approach compared with Latent Dirichlet Allocation is that Latent Dirichlet Allocation may not perform well when the texts are short, as is the case with tweets and microblogs (QUAN et al., 2015; MEHROTRA et al., 2013). Many of the texts used in this paper are rather short, and despite their short length our analysis shows that their content is relevant in yield curve prediction. A second limitation is related to external validity since our results only reflect the Brazilian Central Bank communication. Future research may use Central Bank communication to forecast the yield curve for economies that explicitly uses forward guidance as a monetary policy tool.

In conclusion, our result shows that Central Bank communication does contain predictive information about the yield curve, especially for long movements. We argue that the improvement in accuracy is due to the information effect. Since it affects long horizons and maturities of the term structure of interest rate, Central Bank communication reveals to be a relevant tool for monetary policy.

3.5 Conclusion

In this paper, we augment the Dynamic Nelson-Siegel model to include Central Bank communication and use this model to forecast the Brazilian yield curve. We conclude that Central Bank communication, measured by the sentiment of the Central Bank, helps to predict the term structure of interest rate. The biggest accuracy gain is for long forecast horizons and long maturity yields. This improvement happens because when the Central Bank publicizes its assessment of the macroeconomic outlook, market participants change the composition of their fixed income portfolio, affecting the entire term structure of interest rate. Thus, the use of a measure of Central Bank communication improves yield curve predictability.

3.6 Appendix

3.6.1 Bayesian estimation of the DNS model by the Hamiltonian Monte Carlo (HMC) method

Suppose we collect all parameters of interest in a vector q. The idea underlying the Hamiltonian Monte Carlo method is to use the Hamiltonian dynamics from physics to sample from a (posterior) distribution $\pi(q)$. In this context, we view the vector q as a position variable, and it is necessary to introduce a momentum variable, which we denote by p. Intuitively, the HMC consists of building a Markov Chain as follows: Generate proposals of the position by exploring the evolution of the Hamiltonian system, and use a Metropolis acceptance step to decide whether the position variable moves or returns to its previous values.

Formally, it is necessary to relate the posterior distribution and the Hamiltonian dynamics. Consider that the joint probability of the momentum and position variables is given by $\pi(q,p) = \pi(p|q)\pi(q)$, and define the Hamiltonian $H(p,q) \equiv -\log(\pi(q,p))$, such that:

$$H(p,q) = K(p,q) + V(q),$$

where $K(p,q) \equiv -\log(\pi(p|q))$ is the Kinetic energy and $V(q) \equiv -\log(\pi(q))$ is the potential energy, so that $H(p,q)$ is a energy function. Using the concept of canonical distribution from statistical mechanics, the joint distribution of (q,p) is given by $P(q,p) \propto \exp(-H(q,p))$. Thus, we can analyze the dynamic of (q,p) by using Hamilton equations:

$$\begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p} = \frac{\partial K}{\partial p} \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial q} = -\frac{\partial K}{\partial q} - \frac{\partial V}{\partial q}. \end{aligned}$$

Since the moment p is independent of q, then $\pi(p|q) = \pi(p)$, which implies that $\partial K/\partial q = 0$. Consequently the system of differential equation is given by $dp/dt = -\partial V/\partial q$ and $dq/dt = \partial K/\partial p$. This system can be solved using the leapfrog integrator, an algorithm that operates through discrete time steps in a small interval ϵ. Then, it toggle between an update at $t + \epsilon/2$ to q and an update at $t + \epsilon$ to p and q, as follow:

$$p^{(t+\epsilon/2)} \longleftarrow p^{(t)} - (\epsilon/2)\frac{\partial V^{(t)}}{\partial q} \tag{3.10}$$

$$q^{(t+\epsilon)} \longleftarrow q^{(t)} + \epsilon\Sigma p(t + \epsilon/2) \tag{3.11}$$

$$p^{(t+\epsilon)} \longleftarrow p^{(t+\epsilon/2)} - (\epsilon/2)\frac{\partial V^{(t+\epsilon/2)}}{\partial q}. \tag{3.12}$$

After L discrete iterations, the algorithm returns the final state, denoted by (q^*, p^*). This states is accepted or rejected according to a Metropolis condition, that is, is accepted with probability $\min\{1, \exp[H(q,p) - H(q,p)]\}$. More details about the use of HMC in the Bayesian estimation of DNS models can be found in (BATISTA; LAURINI, 2016).

We run a different number of iterations for each model described in Table 3.1. As a general rule, the simpler the model, the smaller the number of iterations required to achieve convergence. Specifically, for the DNS-macro, we run 2×10^4 iterations, discarding the first 50% of the chain as burn-in. For the DNS-macro and DNS-macro-state, we run 3.5×10^4, discarding the first 1.5×10^4 draws as burn-in. For the DNS-min, we run 4.5×10^4, discarding the first 2×10^4 as burn-in. For DNS-macro-tone, we run a total of 8×10^4 iterations, discarding the first 50% of the chain as burn-in. For the DNS-tone, we run 1.6×10^4 and discard the first 6000, and for DNS-min, we run 2.5×10^4, excluding the first 10^4. Finally, the DNS-state required 5.5×10^4 with the first 2.5×10^4 used as burn-in.

The prior distribution can be briefly summarized as follow. For the variance parameters, σ_i^2 and σ_η, we specify gamma distributions to ensure the positiveness of these parameters. The decay parameter λ also needs to be positive. Therefore, we use a lognormal distribution for this parameter. For the persistence parameters, ϕ_i, we use a normal distribution truncated in the range $(0, 1)$ to guarantee the stationary of the model. The prior distribution for the rest of the parameters, c_i, γ_i, and α_i, are standard normal distributions. More details about the prior distributions are available upon request.

3.6.2 Cumulative Squared Forecasting Error

Besides the cumulative squared forecasting error (CSFE) presented in the main text, here we show the CSFE for $h = 1$ and $h = 12$. Figures 3.6 and 3.7 depict the result for these forecast horizons. Again, each line represents a specific model. A positive slope indicates outperformance, and a negative slope represents underperformance in a given period.

For the one-step-ahead forecast, the CSFE presents a positive slope for all models that include macroeconomic variables (DNS-macro, DNS-macro-state, and DNS-macro-tone) for maturities up to three months, outperforming the benchmark for every period. Regarding six-month maturity, we observe some periods of positive slope for these models. However, in the last months of the forecast period, the CSFE is decreasing, implying a worst overall performance than the benchmark RW. For maturities greater than six months, all models underperform the RW. In particular, DNS-only and the models that include only sentiment variables dramatically decrease over the entire period, independent of the maturity.

Finally, Figure 3.7 displays the CSFE for $h = 12$. In this case, only the DNS-min and DNS-state specifications present a positive slope, while the remaining competitors

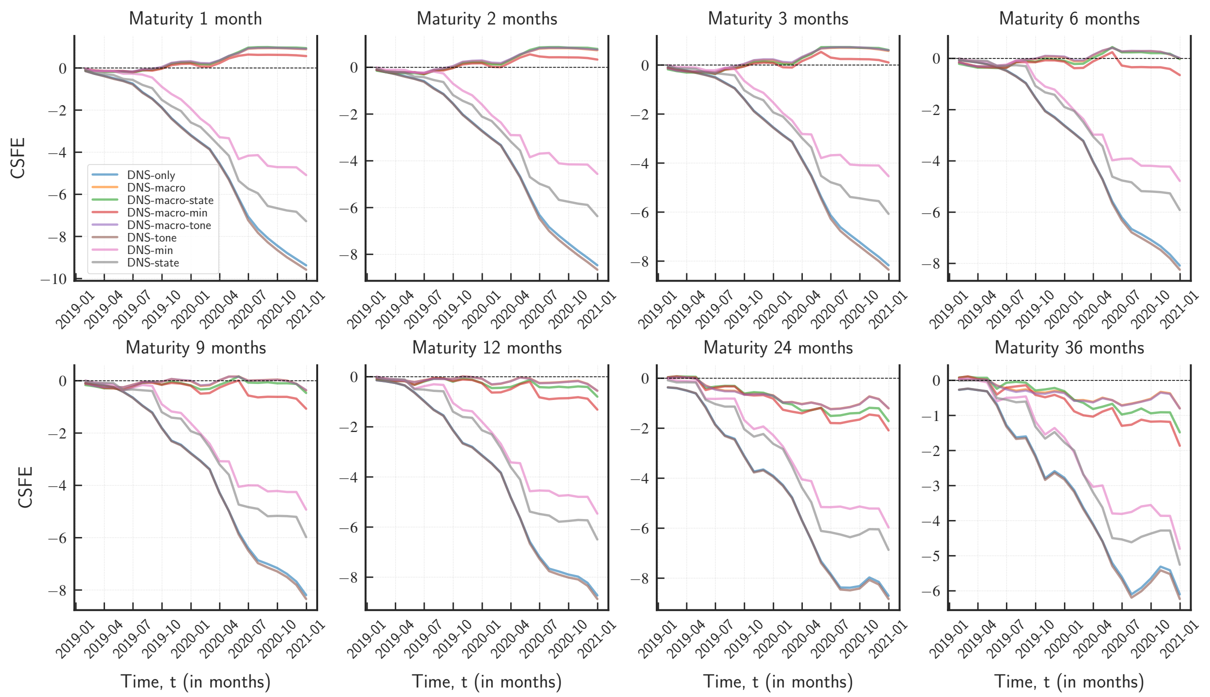

Figure 3.6 – Cumulative squared forecast errors (CSFE) one-step-ahead: An increase in a line indicates better performance of the named variable; a decrease in a line indicates better performance relative to the random walk.

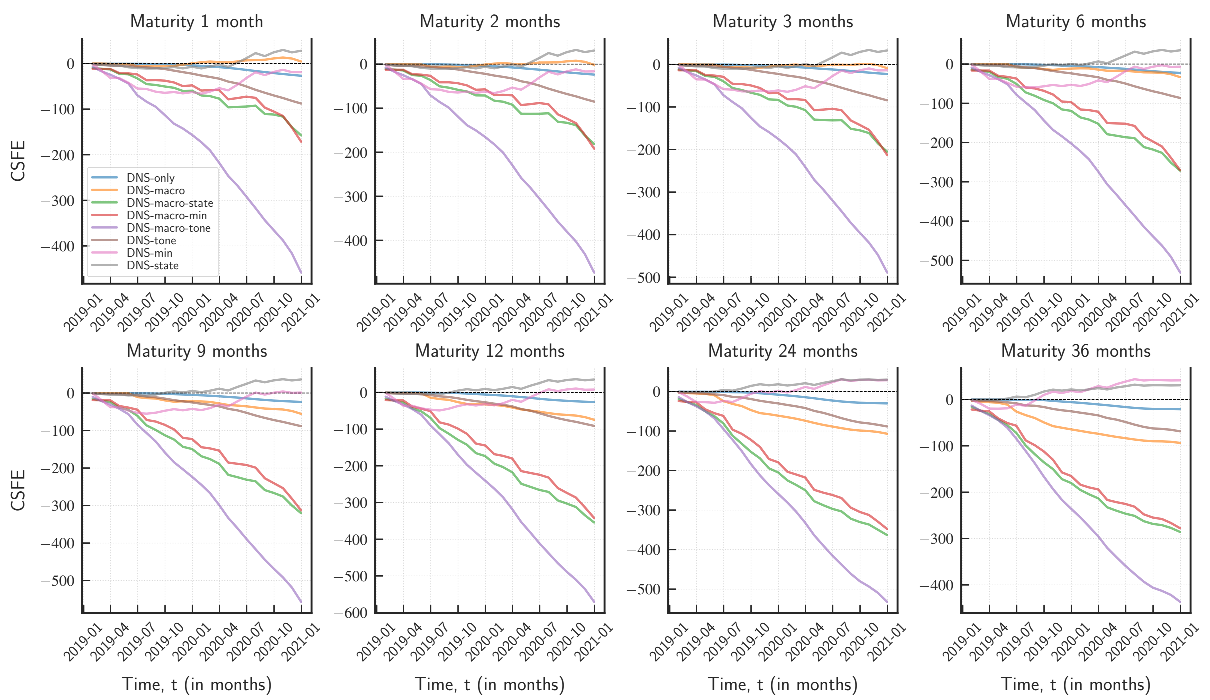

Figure 3.7 – Cumulative squared forecast errors (CSFE) twelve-step-ahead: An increase in a line indicates better performance of the named variable; a decrease in a line indicates better performance relative to the Random Walk.

present decreasing line, indicating underperformance relative to the RW model.

3.6.3 Robustness analysis

We forecast the yield curve again, considering some modifications. The first one concerns the construction of the sentiment variable. Instead of filling in the gaps created by the absence of Central Bank meetings with zeros, we use the sentiment variable's value from the most recent Central Bank meeting. The second one concerns the forecast period. We changed the forecast window to exclude the covid-19 pandemic period. The last modification is related to the change in the number of words in Central Bank documents, shown in Figure 3.2. We redo the forecasts considering a dummy variable to separate the "few words" period from the "many words" period. None of these modifications changed the conclusion of the results presented in the main text. In what follows, we show the results of these modifications.

Table 3.4 presents the forecast accuracy considering a different forecast period. We forecast the period between 2018 and 2019 rather than the range between 2019 and 2020. Also, the specifications that consider Central Bank communication use a different the construction of the sentiment variables, as explained before. The result is very similar to those presented in Table 3.3 in the text, as shown in Table 3.4.

Additionally, we run specifications that include a dummy variable to control for the different partners in the number of words in the period after June 2016. Thus, for every model that includes sentiment variables, we run the following specification:

$$\beta_{i,t} = c_i + \phi_i \beta_{i,t-h} + \sum_{j=1}^{J} \gamma_{ij} X_{j,t-h} + \sum_{k=1}^{K} (\alpha_{ik} + \delta_{ik} d_{t-h}) S_{k,t-h} + \varepsilon_{i,t}, \tag{3.13}$$

where $\varepsilon_{i,t}$ is normally distributed with zero mean and variance given by σ_i^2 and d_t is equal to 1 for every period after June 2016 and 0 otherwise.

In general, the result of the forecast produced by Equation (3.13) and those produced by Equation (3.4) are very similar, and the dummy variable d_t does not affect the accuracy of the forecast. However, for some specifications, specifically for 12−steps ahead forecasts, the MCMC diverges from the stationary distribution, even when we use a large number of MCMC iterations. For those specifications that the MCMC achieves convergence, the forecasts are very similar. Figure 3.8 illustrates the similarity between DNS-state with and without the dummy variable.

Maturity (in months)	RW	DNS-only	DNS-macro	DNS-macro-state	DNS-macro-min	DNS-macro-tone	DNS-tone	DNS-min	DNS-state
				1-month forecast horizon					
1	**0.185***	2.738	1.389	1.288*	1.478	1.411	2.771	1.896	2.162
2	**0.188***	2.548	1.452	1.362*	1.570	1.467	2.574	1.758	1.898
3	**0.194***	2.419	1.503	1.421*	1.651	1.516	2.439	1.704	1.729
6	**0.234***	2.072	1.471*	1.405*	1.665*	1.484*	2.088	1.612	1.428
9	**0.282***	1.827	1.335*	1.282*	1.514*	1.344	1.839	1.485	1.279
12	**0.332***	1.696	1.210*	1.153*	1.346*	1.208*	1.702	1.371	1.213
24	**0.461***	1.467	1.113*	1.044*	1.143	1.098*	1.465	1.181	1.129*
36	**0.521***	1.261*	1.043*	1.005*	1.073*	1.034*	1.263*	1.084*	1.040*
				6-months forecast horizon					
1	1.013*	1.271	1.074	0.922	0.926	1.709	1.548	1.171	**0.746***
2	1.000*	1.242	1.085	0.894	0.896	1.760	1.519	1.175*	**0.733***
3	0.983*	1.240	1.127	0.896	0.896	1.832	1.516	1.204*	**0.745***
6	0.963*	1.254	1.291	0.961*	0.952*	2.015	1.522	1.307*	**0.807***
9	0.981*	1.260	1.431	1.045*	1.031*	2.106	1.510	1.357*	**0.854***
12	1.041*	1.251	1.510	1.104*	1.087*	2.104	1.475	1.329*	**0.869***
24	1.240*	1.198*	1.539	1.147*	1.125*	1.923	1.345	1.179*	**0.881***
36	1.339*	1.138*	1.415	1.061*	1.041*	1.720	1.241	1.121*	**0.881***
				12-months forecast horizon					
1	2.681	1.026	0.801	1.135	1.769	1.377	1.200	1.177	**0.751***
2	2.616	1.011	0.809	1.202	1.753	1.409	1.198	1.235	**0.740***
3	2.532	1.010	0.831	1.282	1.759	1.458	1.210	1.313	**0.739***
6	2.309	1.022	0.924	1.531	1.809	1.623	1.267	1.567	**0.751***
9	2.152	1.036	1.027	1.743	1.861	1.768	1.317	1.800	**0.765***
12	2.068	1.046	1.118	1.893	1.893	1.870	1.351	1.961	**0.778***
24	1.964	1.042	1.293	2.053	1.876	1.990	1.368	2.213	**0.811***
36	1.968	1.006	1.284	1.910	1.757	1.906	1.304	2.244	**0.829***

Table 3.4 – RMSE for 1-, 6- and 12-step-ahead prediction. The column RW reports the root of mean squared error (in percentage points) of the random walk benchmark, while the other columns display the root of mean squared error of each model relative to the benchmark. Except for the first column, values less than 1 indicates outperformance relative to the benchmark. Values in bold indicate outperformance in the maturity. We indicate by * the models in the set of superior models for each maturity at 10% of significance.

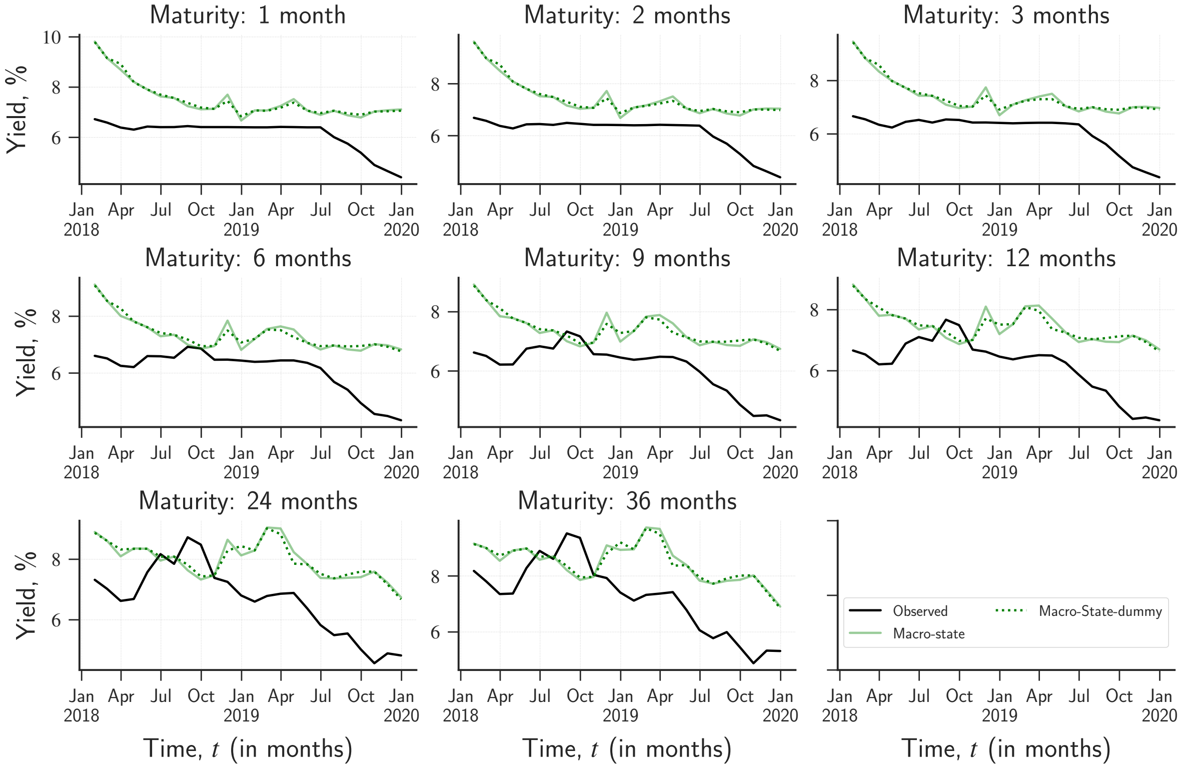

Figure 3.8 – Forecasts using Equation 3.4 versus forecasts using Equation 3.13. The dotted line represents the forecast with dummy variables.

4 The effects of Brazilian Central Bank Communication on the yield curve

Abstract

This paper investigates the bidirectional relation between the Brazilian Central Bank Communication and the yield curve. Using latent factors, observable macroeconomic variables, and observable variables representing central bank communication, we estimate a model that summarizes the yield curve. We find evidence of the effects of Brazilian Central Bank Communication on the movements of the yield curve and the impact of the yield curve components in Brazilian central bank communication. In particular, Central Bank Communication can shape yield curve curvature and slope. Additionally, we find a strong relation between Central Bank Communication and the curvature of the yield curve. These results show that Central Bank Communication impacts market players, making it a valuable instrument for monetary policy.

Keywords: Yield Curve, Sentiment Analysis, Bayesian Estimation, Central Bank Communication.

4.1 Introduction

Researchers and practitioners have increasingly recognized the importance of central bank communication on macroeconomic variables, such as inflation and interest rates. It is still unclear, however, how the Central Bank Communication impacts the term structure of interest rates and how it is affected by changes in those interest rates. Some authors argue that central bank communication affects the long maturity of the yield curve (LEOMBRONI et al., 2021, for instance), and others advocate that it affects only the short maturity of the yield curve (MÁTÉ; SEBŐK; BARCZIKAY, 2021, for instance). Also, the effects of yield curve shape on Central Bank Communication, if any, are unclear. This paper investigates the bi-direction effects between the Central Bank Communication and the term structure of the interest rate.

The most common monetary policy instrument is the short-term interest rate, which policymakers can control directly. Nevertheless, policymakers cannot directly affect longer maturities of interest rates. Since the entire term structure of interest rates affects investment and consumption decisions, all maturities are relevant to policymakers. The change in the short-term interest rate by the central bank may cause an indirect impact on the yield curve because investors will respond to the new short-term interest rate by trading their assets. How the investor will relocate their portfolio, however, will depend on their expectation about the future path of monetary policy and other economic fundamentals, and it is out of the control of the interest rate policy instrument.

The recent literature on central banking recognized that Central Bank Communication may also serve as a monetary policy instrument. This sort of policy instrument, in turn, may affect medium- and long-term interest by shaping the agent's expectations. By telling their vision about future interest rate decisions, inflation, and economic activity, the central bank may update the agent's beliefs about the economic fundamentals, affecting their trading and, ultimately, the price and the assets' returns. This mechanism is known as the *information effect*: Central Bank announcements lead the private sector to update its belief about the monetary policy path and future time path economic fundamentals (ROMER; ROMER, 2000; NAKAMURA; STEINSSON, 2018).

The empirical evidence of how central bank communication affects the term structure of interest rates is conflicting. Part of the literature argues that Central Bank communication impacts only short-term yield curve maturities (LUCCA; TREBBI, 2009; MÁTÉ; SEBŐK; BARCZIKAY, 2021) and the other part says that, instead, it affects long-term yield curve maturities (LAMLA; LEIN, 2011; CHAGUE et al., 2015; LEOMBRONI et al., 2021). This conflicting evidence may be related to different data and hypotheses used to model and estimate yield curves and central bank communication. For instance, an assumption in the literature is that the equation drove the economy and the central bank communication have no contemporaneous or lagged interest rate for any maturity[1]. Also, in some works, the effect of central bank communication on the interest rate is modeled for a given maturity (LUCCA; TREBBI, 2009; LAMLA; LEIN, 2011; MÁTÉ; SEBŐK; BARCZIKAY, 2021), requiring estimating a different model for each maturity. Likewise, a branch of the literature assumes that Central Bank Communication is determined independently of the shape of the yield curve. This hypothesis implies a unidirectional effect from CBC to yield curve (CBC-yield assumption). The Central bank may, however, react to changes in the shape of the yield curve since it is related to several macroeconomic variables.

This paper aims to investigate the bidirectional relationship between central bank communication and the yield curve. In doing so, we allow for both the yield curve affecting the CBC and CBC affecting the yield curve. Specifically, we extend the DNS of Diebold and Li (2006) to also include the central bank communication in addition to the latent factors (level, slope, and curvature) and macroeconomic factors (inflation, capacity of utilization, and interest rate). This model allows us to study the bidirectional effect between CBC and the yield curve and also to consider the lagged effects of the yield curve due to the autoregressive structure of latent factors.

Our results showed that central bank communication indeed impacts the shape of the yield curve. Specifically, if the central bank talks in a more positive way the curvature of the yield curve response is positive. If instead, the central bank has a negative discourse, then the slope and curvature response is negative. Since changes in curvature are related to intermediate maturity changes, our results indicate that the central bank can affect more than just the short-term interest rate by using central bank communication as a policy instrument. We also find that, conversely, changes in the shape of the yield curve affect central bank communication. Finally, we find a relation between the central bank communication to the curvature factor of the yield curve.

Our findings complement and connect to the following works of literature. While Diebold, Rudebusch and Aruoba (2006a) relates the level and slope factors to macroeconomic variables, we relate the curvature factor with central bank communication. Chun (2011) relates the fluctuations in bonds with the expected path of monetary policy and macroeconomy using analyst forecasts. Our approach can be seen as an alternative that uses Central Bank Communication as an agent expectation. Han, Jiao and Ma (2021) relate the shape of the yield curve, associated with a time-varying factor loading, with real macroeconomic variables and argue that this time-varying factor loading contains information about the market perception of the economic risk and uncertainty. In our approach, we use central bank communication to include the perception of economic risk in the model. We believe that our results shed new light on the effects of central bank communication on the yield curve.

[1] See, for example, Lucca and Trebbi (2009), p. 30.

4.2 Quantifying the Central Bank Communication

The most usual form Central Bank Communicates with the private sector is by issuing press releases. This sort of communication often occurs after a council meeting, which defines the short-term interest rate considering the council members' vision of the economic outlook. The content is documented in the textual press release and made available for private agents in the economy. In the case of the Brazilian Central Bank, this communication takes place every forty-five days using two press releases: the COPOM Minutes and the COPOM Statements. While the Statement is a short document that briefly explains the decision on a short-term interest rate, the Minutes are a longer document detailing the macroeconomic outlook and discussing the decision made by the committee.

A natural way to quantify the Central Bank Communication is to analyze the press release issued by the Central Bank, and researchers widely use this practice (BOUKUS; ROSENBERG, 2006; LUCCA; TREBBI, 2009; ROSA, 2011; CHAGUE et al., 2015). To transform the textual information into numerical variables, we used sentiment analysis to extract information from the COPOM minutes made available for the Central Bank. We start by acquiring unprocessed text data and appropriately handling it. After collecting the raw data at each meeting, we remove punctuation, blank lines, stop words, and the names of the members of the monetary policy committee in order to remove words that should not impact the tone of the text. It is important to note that the COPOM announcements and minutes were only regularly published by the Brazilian central bank starting in 2006, which serves as the starting year for our sample in this article. By categorizing the content of these papers into semantic categories, we utilize this processed text data to measure the Central Bank of Brazil's attitude.

We use a dictionary-oriented technique to analyze the sentiment of the COPOM minutes. The vocabulary of these documents is highly specialized, and the phrases employed have specific connotations, usually with financial meanings. As a result, we classified the words in the papers using the financial dictionary offered by Loughran and McDonald (2011). This dictionary helps us to identify each word's semantic content, categorizing it as positive, negative, or uncertain. We create the sentiment variables representing central bank communication using these three semantic categorizations. We regard the proportion of each category to the total amount of words in each document as sentiment variables for COPOM minutes. For every period t that the COPOM meetings take place, we denote the proportion of positive, negative, and uncertainty words by $s_{p,t}$, $s_{n,t}$ and $s_{u,t}$. We also follow Cannon (2015) to define the tone of the central bank by:

$$\tau_t \equiv \frac{s_{p,t} - s_{n,t}}{s_{p,t} + s_{n,t}}. \tag{4.1}$$

The tone summarizes the sentiment of the central bank in one variable, and it is helpful to analyze the central bank communication and yield curve relationship. We used these four variables to analyze how central bank communication affects the term structure of interest rates.

4.3 The Yield Curve and the Central Bank Communication

To analyze the interconnection between the yield curve and central bank communication, we augmented the Dynamic Nelson and Siegel (DNS) proposed by Diebold and Li (2006) to include Central Bank Communication as a new factor. Considering observables and latent factors, the DNS has proved to perform well in fitting and forecasting the yield curve. The original DNS of Diebold and Li (2006) decomposes the yield curve using only the unobserved factors known as level (L_t), slope (S_t), and curvature (C_t). Then, Diebold, Rudebusch and Aruoba (2006a) extend the original model to include a bi-direction relation between the unobserved factors and observed macroeconomic factors, specifically inflation, the capacity of utilization, and interest rate. Our model extends this macro-yield model to include central bank communication.

To include the central bank communication in the DNS model, let us define the vector of factors as $X_t \equiv (L_t, S_t, C_t, s_{p,t}, s_{n,t}, s_{u,t}, CU_t, R_t, \pi_t)'$. We assume that the factors follow a Vector Auto-Regression (VAR) representation:

$$X_t = c + \Phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(\mathbf{0}, Q), \tag{4.2}$$

where c is a vector of constant, Φ is a 9×9 matrix of VAR coefficients and Q is assumed to be diagonal. In line with the DNS model, we use the unobserved factors to build the entire term structure of interest rate:

$$y_t(m) = L_t + S_t \left(\frac{1 - e^{-\lambda m}}{\lambda m} \right) + C_t \left(\frac{1 - e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \right) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \tag{4.3}$$

where $y_t(m)$ is the interest rate in period t of a bond with maturity m and λ is a factor decay.

The model formed by Equations (4.2) and (4.3) describes the dynamics of the yield curve considering macroeconomic variables and central bank communication. The parameters of the model can be summarized in a vector $\theta = (c, \phi_{ij}, q_i, \lambda, \sigma_\eta^2)$, where ϕ_{ij} is the elements of the matrix Φ, q_i is the diagonal elements of the matrix Q and $i, j \in \{1, \cdots, 9\}$. For a given set of parameters θ, we can assess how the central bank communication (and macroeconomic variables) affects the latent factor L_t, S_t, and C_t. Depending on the size of the factor decay λ, the latent factors will build the entire term structure of interest rate, so we can indirectly assess the effect of communication and macroeconomic shocks in the yield curve. The level, slope, and curvature are unobservable, and we need to estimate them.

We can estimate the vector of parameter θ and latent components in a Frequentist or Bayesian approach. In the Frequentist case, we can construct the likelihood function by the prediction error decomposition produced from the Kalman Filter and then maximize this function (DIEBOLD; RUDEBUSCH; ARUOBA, 2006a). The maximum likelihood estimation, however, may present numerical problems, particularly when the number of factors increases. We use a Bayesian technique estimation (LAURINI; HOTTA, 2010) to prevent this kind of issue. In the Bayesian case, we elicit prior distribution to mix with likelihood function information. Then, we use Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution. In Appendix 4.6.1, we present prior distribution and MCMC details.

4.4 Results

4.4.1 Yield curve fit

We estimate the model for yields of eight different maturities, presented in the first column of Table 4.1. The DNS model augmented with central bank communication fits the yield curve well since the difference between the fitted yield curve and the observed yield curve, using the estimated latent factors, is negligible. Table 4.1 shows some statistics of this difference.

Table 4.1 – Summary statistics for the difference between the fitted yield curve and the observed yield curve.

Maturities (in months)	Mean	Std. dev.	Min	25%	50%	75%	Max
1	-0.0002	0.0001	-0.0004	-0.0002	-0.0002	-0.0001	0.0001
2	0.0000	0.0000	-0.0001	0.0000	0.0000	0.0001	0.0002
3	0.0001	0.0000	-0.0001	0.0000	0.0001	0.0001	0.0002
6	0.0001	0.0001	-0.0001	0.0001	0.0001	0.0002	0.0003
9	0.0001	0.0001	-0.0003	0.0000	0.0001	0.0002	0.0004
12	-0.0000	0.0001	-0.0005	-0.0001	-0.0000	0.0001	0.0004
24	0.0000	0.0003	-0.0011	-0.0002	0.0000	0.0002	0.0009
36	0.0003	0.0004	-0.0015	-0.0000	0.0003	0.0006	0.0015

Figure 4.1 shows the estimated latent factors used to build the fitted yield curve. The level factor presents high persistence behavior and displays a decreasing tendency overall. In contrast, the estimated slope and curvature are less persistent and assume both positive and negative values. Also, there is a correlation of 0.5 between the slope and curvature movements.

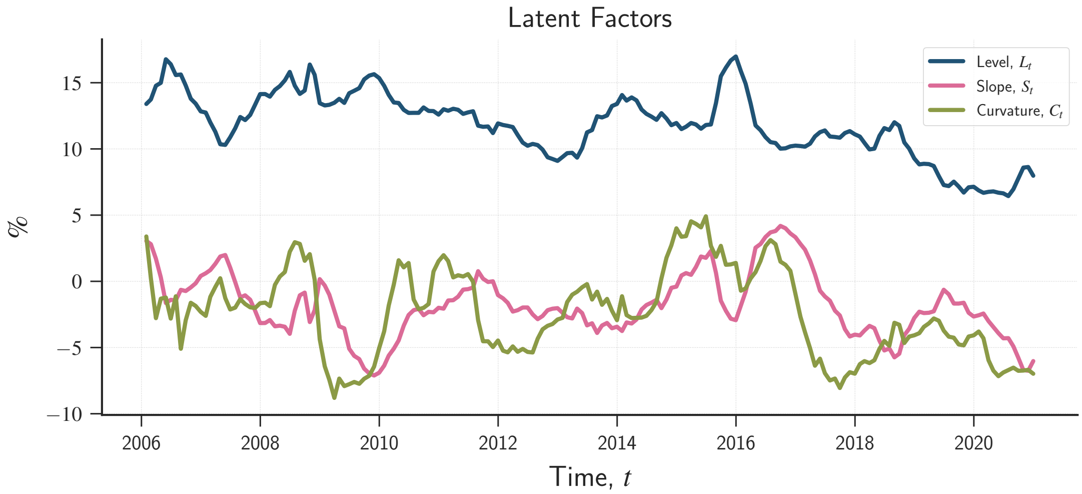

Figure 4.1 – Latent factors: level, slope and curvature

Since these three unobservable factors form the entire yield curve in the DNS model, analyzing their dynamic behavior and relation to other observed factors helps us to understand the yield curve behavior.

4.4.2 Unobservable factors and related variables

We begin by analyzing the correlation between the latent factors and the extracted sentiment of central bank communication, described in the proportion of positive, negative, and uncertain words. Figure 4.2 displays the correlation matrix of latent factors and sentiment variables. In general, the correlation between latent factors and the sentiment of the central bank is high. Specifically, the curvature of the yield curve is related to all sentiment variables. The slope, in turn, is only weakly correlated to the proportion of positive and uncertain words, but it is highly associated with proportion of negative words. The level is also weak correlated to the positive words. This fact leads us to investigate further the relationship between the curvature factor and central bank communication.

A large body of literature links the level factor with inflation and the slope factor with the economic activity cycle. Links between observable variables and the curvature of the yield curve are less frequent. Here, we summarize the central bank communication in the tone variable, as defined in Equation (4.1), and link the curvature of the yield curve with this variable. Figure 4.3 displays the estimated latent factors and the linked comparison series.

In panel (a) of Figure 4.3, we can observe that the level of the yield curve is correlated with inflation, as highlighted in the literature (see, e. g., Diebold, Rudebusch and Aruoba (2006a)). The literature also usually connects the slope factor with the economic cycle. We use the year-on-year economic growth to represent the cycle. In contrast to the finds in the literature, the correlation between the cycle and the slope factor is low (see panel (b) of Figure 4.3). Surprisingly, the curvature factor and the

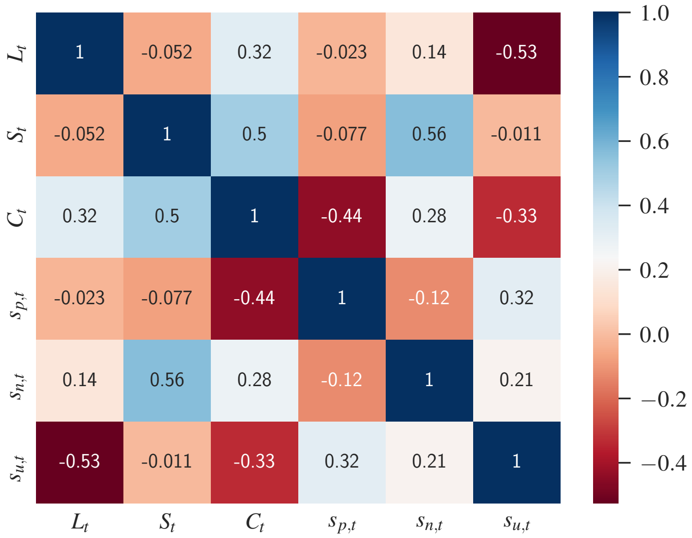

Figure 4.2 – Heat-map of correlation matrix: Yield curve latent factors and central bank communication

central bank communication (represented by the tone variable) move in the same direction overall, presenting a Pearson correlation of 0.50, as shown by panel (c) of Figure 4.3.

The linkage between the curvature factor and central bank communication sheds new light on how Central bank communication can serve as a tool to affect the term structure of interest rates. Increasing the curvature of the yield curve means that medium-term maturities will have higher interest rates (LITTERMAN; SCHEINKMAN, 1991; DIEBOLD; LI, 2006). Accordingly, our find implies that central bank communication is related to the movements in medium-term maturities, which is in line with the literature that argues that the impact of the CBC is beyond the short-term but can affect other maturities too (LAMLA; LEIN, 2011; CHAGUE et al., 2015; LEOMBRONI et al., 2021).

For the Brazilian yield curve, the slope and curvature are positively correlated. Thus, the tone of central bank communication is also associated with the yield curve slope, presenting a correlation of 0.42. The literature argues that the yield curve slope can predict recessions (FAMA, 1986; ESTRELLA; MISHKIN, 1998; RUDEBUSCH; WILLIAMS, 2009; BENZONI; CHYRUK; KELLEY, 2018). Financial economists claim that the yield curve slope "contains information about current and expected future monetary actions" (BENZONI; CHYRUK; KELLEY, 2018, p. 1).

We claim that the central bank informs the private sector about the current and expected future monetary policy in their communications and also their point of view about the economic scenario. Therefore, CBC also contains information about the cycle, which is in line with previous results of Gardner, Scotti and Vega (2022), who argue that their sentiment index of FOMC describes good and bad times. Indeed, for the

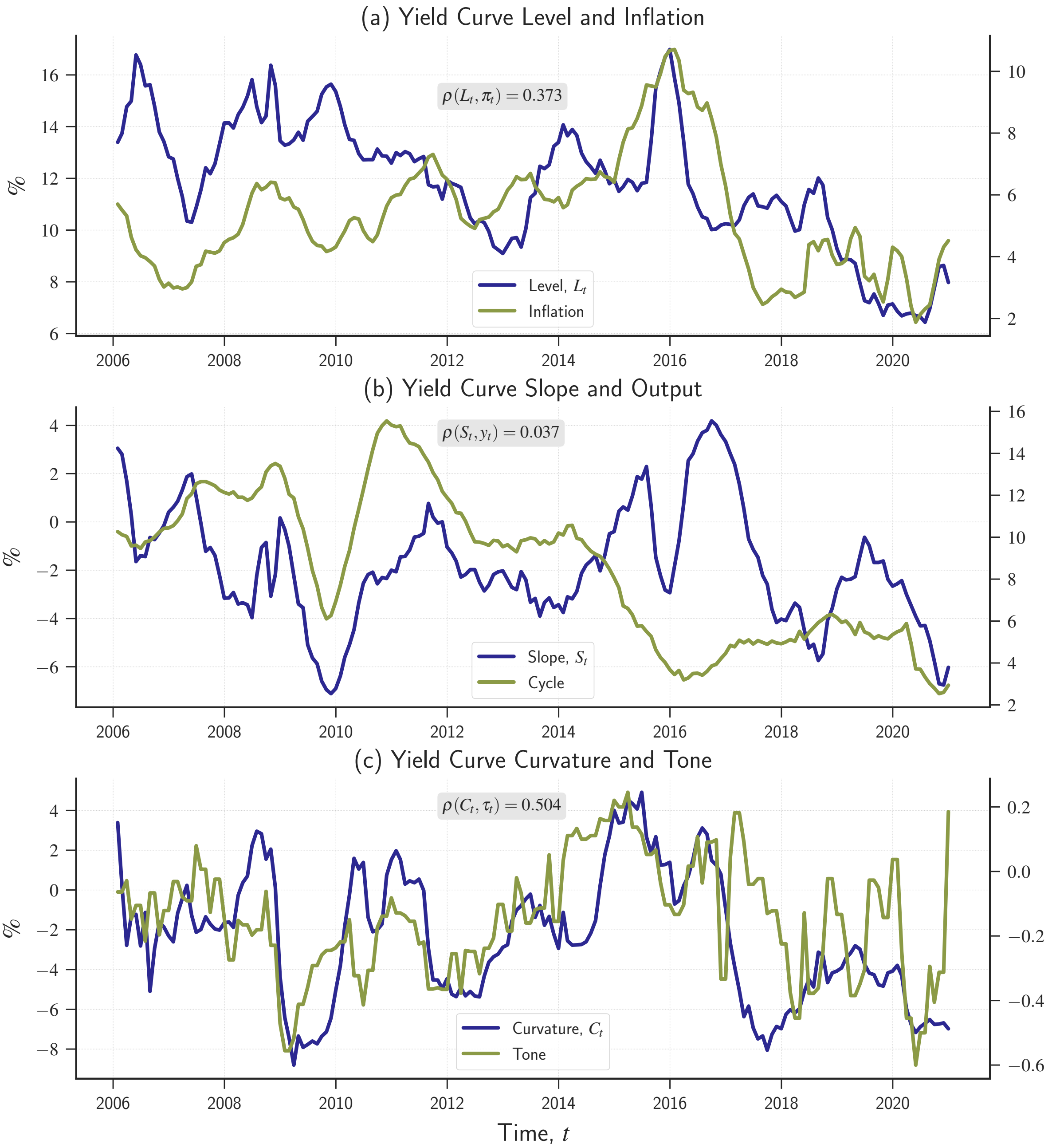

Figure 4.3 – Level, Slope and curvature factors and its empirical counterparts.

Brazilian case, the tone of the Brazilian central bank presents a decreasing tendency during recessions, as shown in Figure 4.6, in Appendix 4.6.2. Han, Jiao and Ma (2021) also argues that, before recessions, the role of curvature and slope is smaller, and during recessions, it tends to play a more relevant role. The tone of the Brazilian central bank, which is correlated with the slope and curvature in the data, captures this movement and informs the market. This knowledge changes the investors' beliefs, possibly implying a change in their portfolios, affecting prices and interest rates of different maturities.

4.4.3 The yield curve and central bank communication dynamics

We now carry out an impulse response analysis to verify how the effects of the different types of discourse (more positive, negative, or uncertain) impact the yield curve. We divide the impulse response function analysis into two groups. The first analyzes how central bank communication shocks affect the latent factors of the yield curve and how long these shocks last to dissipate. The second group examines how the latent factors shocks affect central bank communication. In Appendix 4.6.2, we present the complete impulse response function in Figure 4.7[2].

Let us now describe the first group of impulse responses. Figure 4.4 displays the response of level, slope, and curvature to shocks in negative, positive, and uncertainty words, all in the proportion of total words. The response of the level factor to shocks in negative words is negligible, but it is non-negligible for shocks in the proportion of positive and negative words. In reaction to the proportion of positive word shocks, the level rises and then dissipates slowly. The level response to uncertain word shocks is negative and more persistent.

The intriguing reactions to analyze are the response of slope and curvature since the results of section 4.4.2 indicate a relation of these two factors with central bank communication. An increase in the proportion of negative words (which decreases the tone) reduces both the slope and the curvature factor. Around five months ahead, the curvature achieves a negative one-to-one response. Similarly, the response of the curvature to an increase in the proportion of positive words (which increases the tone) reaches a positive one-to-one around five months. The reaction of slope to shocks in the proportion of positive words is almost negligible, considering the 50% credible interval.

We can interpret these responses of curvature and slope to the shocks in positive and negative words as an effect of the central bank communication in the short and medium term of interest rate. When the Central bank surprisingly informs the market of its vision about the economic outlook, increasing the proportion of positive (negative) words in its communications, private agents will change their portfolios in such a way that the curvature of the yield curve will increase (decrease). This increase (decreasing) means that medium-term yield bonds are also increasing (decreasing). Since the slope also declines with a positive shock in the proportion of negative words, the short-term yield bonds will also reduce. Thus, central bank communication can serve as a tool to affect not only short-term interest rates but also medium-term yields. Again, this result is in line with the literature that argues that the impact of the CBC is beyond the short-term but can affect other maturities too (LAMLA; LEIN, 2011; CHAGUE et al., 2015; LEOMBRONI et al., 2021).

[2] The impulse response function in Figure 4.7 of appendix 4.6.2 also shows the 90% credible interval. In the main text, however, we only present the 50% credible interval to take a closer look at the behavior of point estimation of the responses, represented by the median of the posterior.

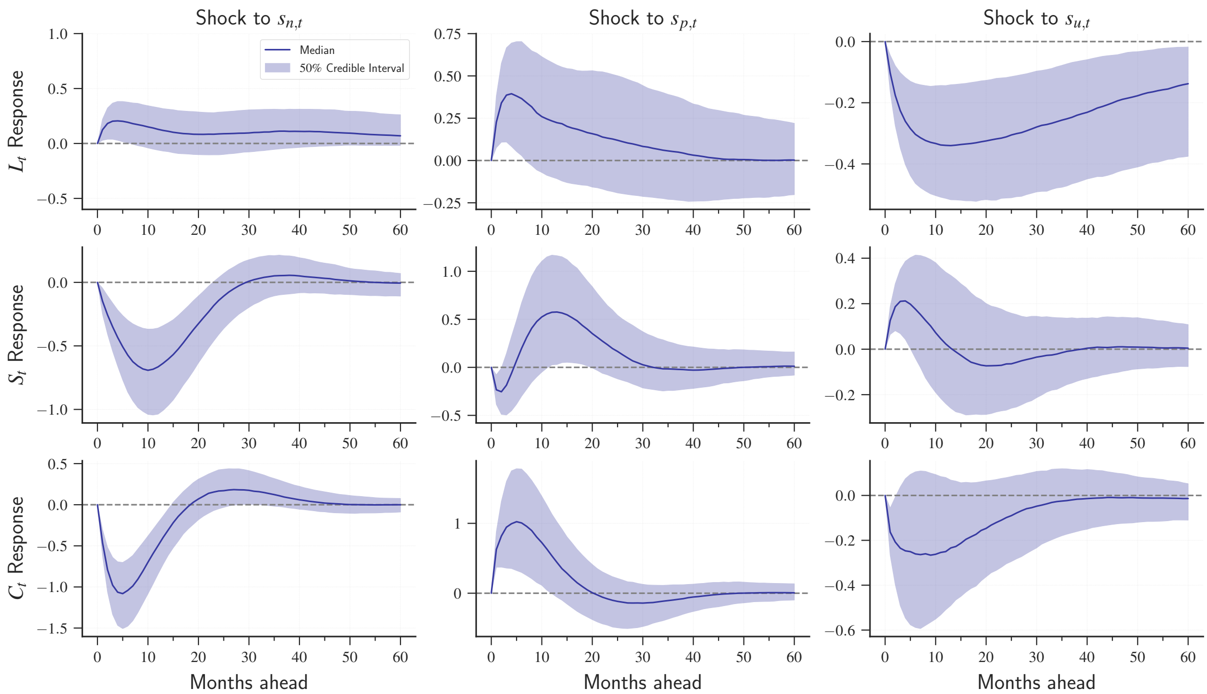

Figure 4.4 – Impulse response function: level, slope, and curvature response to shocks in sentiment variables. The solid blue line represents the median and the shaded blue area represents the 50% credible interval.

Finally, let us describe the second group of impulse responses, which analyzes how central bank communication reacts in response to changes in the shape of the yield curve. To do so, we analyze the impulse response function presented in Figure 4.5. In general, the proportion of negative, positive, and uncertain words increases – some months ahead – considering the credible interval 50%. An increase in the level factor does not affect the proportion of negative words in the first months, considering the credible interval. Around ten months ahead, however, the increase in the level also raises the proportion of the negative words publicized by the central bank. We find a similar pattern for slope and curvature shocks. This result means that the central bank does not react immediately to the proportion of negative words, although it will increase at some point.

The proportion of positive and uncertain words increases immediately for shocks in any of the latent factors[3] and the response dissipates around thirty months ahead. The reaction of the proportion of uncertain words is also positive and immediate for all three shocks. Thus, we can conclude that the central bank communication also reacts to the changes in the yield curve shape, represented by the latent factors.

[3] An exception is the reaction of the proportion of positive words to the curvature shock, which decreases in the first month but then starts to increase. Considering the credible interval, however, we can view the pattern as equal to the other shocks

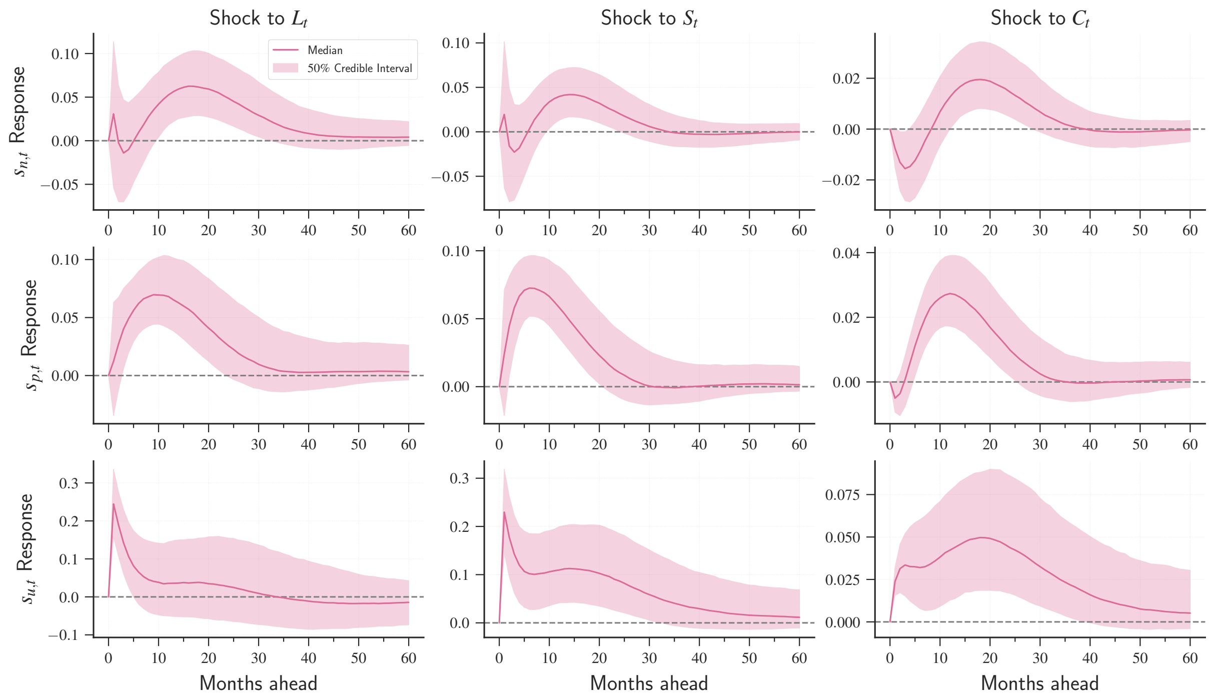

Figure 4.5 – Impulse response function: Proportion of negative, positive, and uncertainty response to shocks in latent factors of the yield curve. The solid red line represents the median and the shaded red area represents the 50% credible interval.

4.5 Conclusion

In this paper, we study the effect of central bank communication on the yield curve. To do so, we used an augmented dynamic Nelson and Siegel model that makes the shape of the yield curve depend on central bank communication. We find that the yield curve, represented by its latent factors, affects and is affected by the Central bank communication. Specifically, we find that the curvature of the yield curve is closely related to Central bank communication. A central bank with a more positive discourse is associated with a greater curvature of the yield curve. Central bank communication is also related to the slope factor, showing that short and medium maturities of interest rates are affected by Central bank communication. Therefore, Central bank communication may serve as a monetary policy tool cable to impact medium-term interest rates in addition to the traditional short-term interest rate instrument.

4.6 Appendix

4.6.1 Bayesian estimation details

The prior distribution can be briefly summarized as follows. For the variance parameters, σ_i^2 and σ_η, we specify gamma distributions to ensure the positiveness of these parameters. The decay parameter λ also needs to be positive. Therefore, we use a log-normal distribution for this parameter. For the persistence parameters, ϕ_{ii}, we use a normal distribution truncated in the range $(0, 1)$, since we are interested in a stationary system. For the parameters ϕ_{ij}, with $j \neq i$ we assume a prior normal distribution. The prior distribution for the rest of the parameters, c_i, γ_i, and α_i, are standard normal distributions. More details about the prior distributions are available upon request.

We run three chains with a total of $20,000$ iterations to estimate the posterior distribution of the model and exclude the first $10,000$ as a burn-in period. For this number of iterations, the MCMC algorithms converge, following usual diagnostics (Gelman and Rubin (1992), Geweke (1992), and visual inspection of trace plots.).

The method used to sample from the posterior of the augmented DNS model was the Hamiltonian Monte Carlo. The Hamiltonian Monte Carlo (HMC) is a Markov Chain Monte Carlo (MCMC) algorithm that uses gradient information to sample from the posterior distribution efficiently. The samples are generated by simulating Hamiltonian dynamics in the posterior distribution. By using gradient information, the HMC can explore the parameter space more efficiently than traditional methods, such as Random Walk Metropolis-Hastings. In short, the algorithm consists in proposing a new draw by combining gradient information and the simulated Hamiltonian dynamics and, then, accepts this proposal using a metropolis step. See Laurini and Hotta (2010) for a detailed explanation of the DNS model Bayesian estimation using HMC, and see Hoffman and Gelman (2014) and Betancourt (2017) for a detailed explanation of the HMC algorithm.

4.6.2 Additional Figures

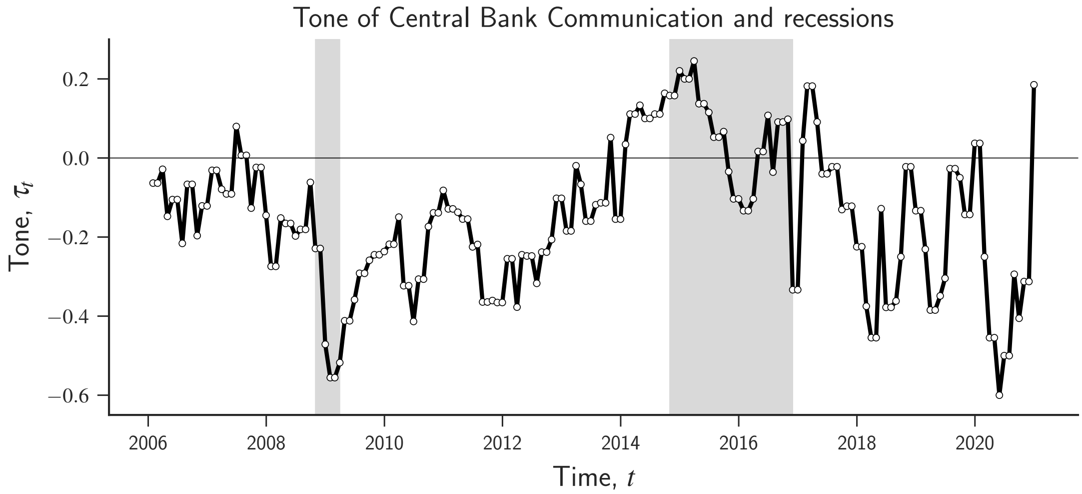

Figure 4.6 – Tone and Recessions

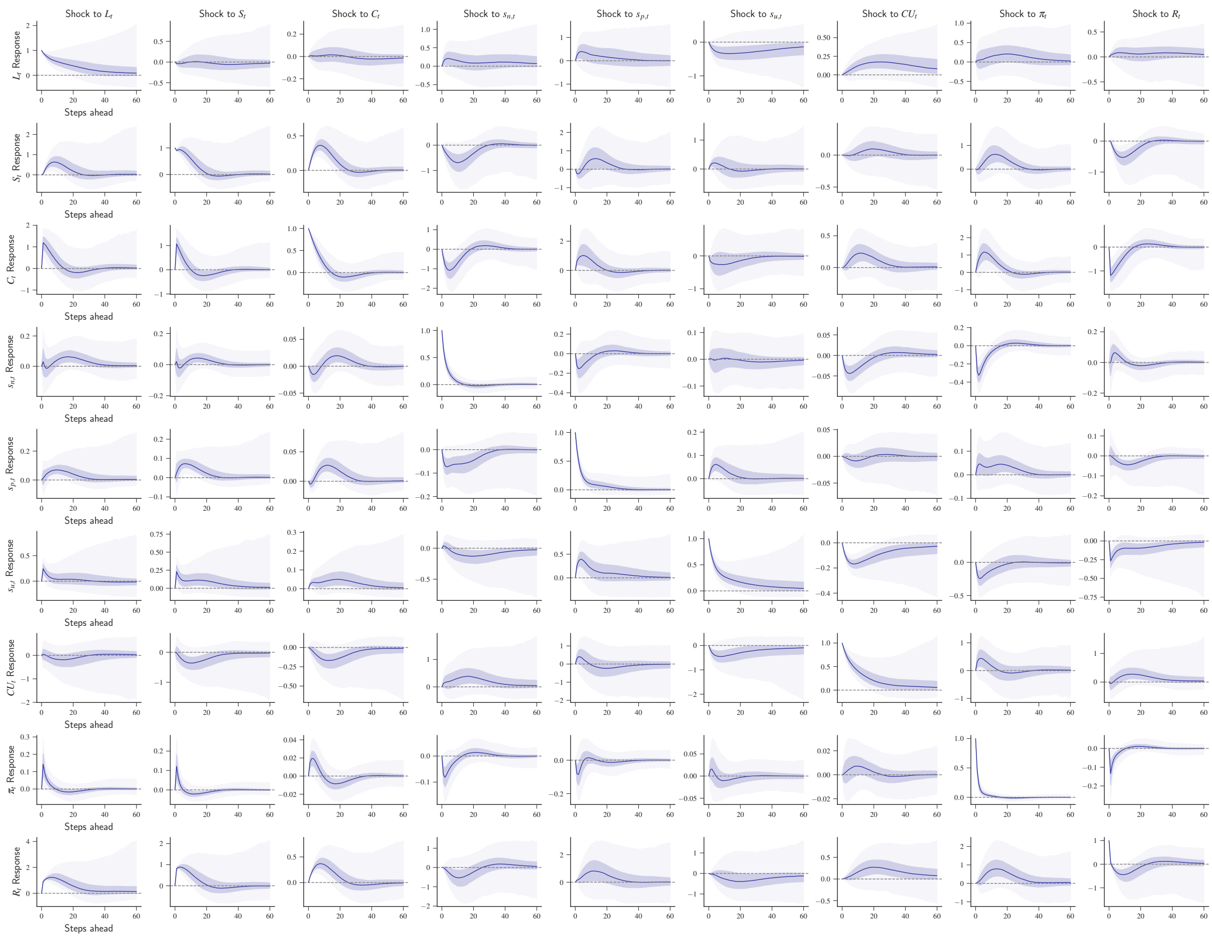

Figure 4.7 – Complete impulse response functions

5 Estimating the Capital Asset Pricing Model with many instruments: A Bayesian Approach

Abstract

This paper proposes an instrumental variable Bayesian shrinkage approach to estimate the Capital Asset Pricing Model (CAPM) using a large set of instruments. Using simulated data, we find that our approach may reduce the size of mean bias caused by error-in-variables. While the traditional two-stage least squares estimation for the CAPM beta becomes biased as the number of instruments increases, our approach corrects the bias for the case of many instruments. However, this bias reduction is attenuated as the number of instruments increases. In an empirical application, the estimated CAPM beta using our approach is subtly different from the ordinary least squares estimator and the traditional two-stage least squares, and this difference is economically relevant. Moreover, when using our approach to explain the variation in average cross-section asset returns the explanatory power of CAPM increases in relation to those found by the two-stage least squares estimator.

Keywords: Bayesian Estimation, Shrinkage, Instrumental Variables, CAPM

5.1 Introduction

Many asset pricing models use the return of the market portfolio as an independent variable in the estimation process. The Sharpe (1964) and Lintner (1965) Capital Asset Pricing Model (CAPM) is perhaps the most famous example of these models, due to its theoretical simplicity and ease of interpretation. Estimating the CAPM requires a surrogate for the return of the market portfolio given the market portfolio's unobservability (ROLL, 1977; STAMBAUGH, 1982; PRONO, 2015). Using a substitute for the market portfolio return introduces an error-in-variables (EIV) problem, which biases the estimates and makes it difficult to interpret the results. This is known as Rolls' critique (ROLL, 1977). The error-in-variables (EIV) problem makes it challenging to estimate the CAPM and evaluate its empirical performance.

The usual econometric solution to the error-in-variables (EIV) problem is to use instrumental variables. However, it can be challenging to find 'strong' instruments for the market return. (MENG; HU; BAI, 2011; SIMMET; POHLMEIER, 2020). The data-rich environment of financial data sets offers many candidates for instrumental variables, although they are usually only weakly correlated to the return of the market portfolio. Alternatively, in contrast to low-dimensional model settings, in which we select a small set of instruments – imposing an *ad hoc* sparsity –, all these many candidates for instrumental variables may be incorporated into the model, leading to high-dimensional model setting. Unfortunately, conventional econometric techniques cannot deal with high-dimensional asset pricing model settings (NAGEL, 2021).

This paper proposes an instrumental variable Bayesian shrinkage approach to estimate the capital asset pricing model using a large set of instruments. Bayesian shrinkage techniques can deal with high-dimensional models by using regularization priors. This approach has been increasingly adopted in financial econometrics (HOTZ-BEHOFSITS; HUBER; ZÖRNER, 2018; KOWAL; MATTESON; RUPPERT, 2019; KOZAK; NAGEL; SANTOSH, 2020; NARD, 2022, for example). Regularization priors are particularly helpful when there are several potential instruments, as in the case of CAPM estimation. Without these regularization priors, using many instruments generates biased estimates (BEKKER, 1994; NEWEY; SMITH, 2004; NG; BAI, 2009). In terms of the two-stage least squares (2SLS), for example, a large set of instruments will imply overfitting in the first-stage, because of the tendency of the ordinary least squares (OLS) to fit too well. Then, the second-stage will be closer to a simple OLS, which is biased in the presence of EIV. The regularization priors avoid this overfitting in the first-stage and, consequently, also avoid the bias in the many instruments setting. Thus, the high-dimensional model setting combined with prior regularization may offer a new approach to deal with Rolls' critique.

Although high-dimensional models have increasingly become popular in financial econometric literature, high-dimensional models for instrumental variables combined to regularization techniques are still a little-explored field. In this paper, we estimate the capital asset pricing model using a large set of instruments and shrinkage priors over the parameters associated with the instruments. To do so, we use the Bayesian approach proposed by Hahn, He and Lopes (2018) to shrink unimportant instruments and compare the size of the estimated bias with that produced by the traditional estimation methods (ordinary least squares and two-stage least squares).

We compare our approach both in simulated data and in observed data. In the simulation exercises, we analyze whether the shrinkage method can help to improve the inference on the CAPM beta. In the empirical application, we use the shrinkage approach to verify if it delivers better estimates for the beta CAPM by comparing the beta estimates between methods. We also verify if our proposal can help explain the cross-section of returns by running the two-step procedure of (FAMA; MACBETH, 1973).

The results indicate that the regularization over the instruments coefficient improves the estimates of CAPM beta. In the Monte Carlo simulation analysis, we find that the regularized Bayesian instrumental variable dramatically reduces the mean bias concerning the traditional 2SLS. Moreover, the Root of Mean Squared Error and Mean Absolute Bias are smaller when using the regularization technique. This evidence shows that high-dimensional settings offer a better way to deal with the CAPM error-in-variable problem. Using many instruments, we can find more precise and unbiased measures of CAPM stock's beta, which helps us accurately evaluate the systematic risk. In addition, more accurate estimates for betas in the first-pass time-series regression offer an adequate input for the second-pass cross-section regression in the Fama and MacBeth (1973) two-step procedure.

In the empirical application, the beta estimates using our approach present a subtle difference to OLS and 2SLS estimates that vary across the assets. This difference in estimated betas is economically relevant since many financial models are sensitive to beta. To further investigate whether the difference among estimated betas is indeed relevant, we run the second-step of Fama-Macbeth procedure for both individual stocks and portfolios sorted by size and book-to-market. For the portfolio data, the results show that our approach can explain around 15% of the cross-section of portfolio return, while the standard 2SLS explains only 5%. For the stocks individually, the betas from our approach can explain 3% of the cross-section of the stock's return, compared to nearly 0% of the standard 2SLS.

These results shed new light on the estimation of asset pricing models with measurement errors. By using high-dimensional data and proper techniques, we can attenuate the error-in-variables problem in the estimation of systematic risk, which in turn will do a better job in explaining the cross-section of return. We note that, even if the power of explanation of the cross-section of returns found in our empirical application is small, compared to multi-factor models, the use of many instruments and regularization priors can improve this power concerning the unregularized approaches. Thus, the use of many instruments with regularization priors can be useful in other contexts too, especially in asset pricing models that include variables with large measurement errors, such as macroeconomic variables.

This paper is organized as follows. Section 5.2 reviews the Capital Asset Pricing Model and introduces the notation used in the paper. Section 5.3 explains the Bayesian estimation and prior regularization. Section 5.4 presents the Monte Carlo analysis and the empirical results of the paper. Finally, Section 5.5 concludes the paper.

5.2 The CAPM and measurement errors

The seminal paper of Markovitz (1959) prepared the framework for the *Capital Asset Pricing Model* (CAPM). The author established the investor problem in terms of a trade-off between risk and return and defined the mean-variance efficiency concept of portfolio allocation. This definition states that, for a given level of return, the portfolio is mean-variance efficient if it minimizes the variance. Sharpe (1964) and Lintner (1965) worked on Markovitz (1959) results to analyze the implication for the asset pricing and developed what is called the Sharpe-Lintner CAPM, or just CAPM.

Assuming that investors possess homogeneous expectations Sharpe (1964) and Lintner (1965) showed that, in the absence of market frictions, if all investors choose an efficient portfolio, then the market portfolio is also mean-variance efficient. In this context, the market portfolio includes all assets in the economy, for instance, stocks, real estate, and commodities, which makes it an unobserved variable. In practice, usual surrogates for the market portfolio are market indexes, such as S&P500, but these indexes do not contain all assets and, consequently, the market portfolio is observed only with errors. Despite this practical difficulty, the efficiency of the market portfolio will imply a relation between assets risk-premium and the market risk premium:

$$\mathbb{E}\left[R_i\right] - R_f = \beta_i \left(\mathbb{E}\left[R_m\right] - R_f\right), \tag{5.1}$$

where $\beta_i \equiv \sigma_{im}/\sigma_m^2$. Therefore, the CAPM summarized in Equation (5.1) is an equilibrium result that holds for a single period.

The relation established in Equation (5.1) for one period is not enough to empirically assess the CAPM. To proceed with econometric analysis, an additional assumption is required: the returns are independent and identically distributed along time and multivariate Gaussian. Although this hypothesis is a strong one, it possesses some benefits. First, it is consistent with the CAPM holding for each period in time. Moreover, it is a good approximation for monthly returns (CAMPBELL et al., 1997). Under this assumption, the CAPM may be represented by the single index model, which is described by

$$R_{it} - R_{ft} = \gamma_i + \beta_i \left(R_{mt} - R_{ft}\right) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2). \tag{5.2}$$

In Equation (5.2), if γ_i is equal to zero, then the CAPM holds for each period in time.

The representation of the CAPM model given by Equation (5.2) started a tradition of testing the model that became known as the time series approach. To empirically test the CAPM model, Jensen, Black and Scholes (1972) proposed to use time series for the return of assets, for return risk-free assets, and a proxy to the return of the market portfolio to estimate Equation (5.2). The usual choice for the risk-free asset is the US Treasure Bill and S&P5500 for the return of the market portfolio. Then, their approach suggests testing whether the estimated intercept is equal to zero, which may be done by using a Wald test or the test proposed by (GIBBONS; ROSS; SHANKEN, 1989).

Testing the CAPM using the Jensen, Black and Scholes (1972) approach is problematic once the return of the market portfolio, R_{mt}, is a variable contaminated by measurement errors. The source of measurement errors appears because the market indexes used to estimate the model contain only a subset of assets. Moreover, even if all universe of assets were observed, the measurement error could appear due to misspecification in the weights of assets. This problem is known as Roll's critique, due to Roll (1977) who argued that, once the market portfolio is not observed, the CAPM cannot be tested. According to this author, a rejection of the CAPM could be due to measurement errors in the return of the market portfolio. In an econometric sense, the present problem is a case of classical measurement errors and should be treated as such.

To put the problem in terms of the classical measurement errors, let $\tilde{R}_{mt}$ denote the observed return of the market portfolio. Also, denote by $x_t^* \equiv R_{mt} - R_{ft}$ the excess of the return on the true market portfolio and by $x_t \equiv \tilde{R}_{mt} - R_{ft}$ the excess of the return on the observed market portfolio. The excess of return on the asset i is denoted by y_{it}, and there is no error in variable in this case. Instead of Equation (5.2), the model to be estimated to test the CAPM should be

$$y_{it} = \gamma_i + \beta_i x_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{5.3}$$

$$x_t = x_t^* + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma_u^2). \tag{5.4}$$

Equation (5.4) assumes that the measurement error is additive. If one ignores this additive measurement error and estimates Equation (5.3) using least squares, then the estimates of betas will suffer from attenuation bias and the intercept will be upward biased, implying positive alphas, even if CAPM holds. Thus, to appropriately deal with error-in-variable problems, the equations (5.3) and (5.4) must be considered to estimate and test CAPM model.

5.3 Methods and Data

The data-rich environment available in the financial data set allows us to use many instruments to correct the bias caused by measurement errors, even though these instruments are possibly weak. The many instruments setting needs to be used carefully, as long as it can itself be a source of bias. To overcome this inconvenience, we need a regularization step, such as variable selection or shrink of less important parameters. Examples of regularization methods are LASSO, ridge, Elastic Net, or via Bayesian shrinkage priors, which penalize the number of covariates in some form.

In instrumental variable regression, it is interesting to use a method that jointly estimates “two stages" and the Bayesian approach has this advantage. The regression of the treatment variable on the instruments and the estimation of the target variable on the treated variable can be estimated in a single step. In this sense, the Bayesian shrinkage priors are preferred rather than other regularization methods. In particular, the factor-based prior proposed by Hahn, He and Lopes (2018) has the advantage of linearly combining the information in all possibly weak instruments in such a way that, taken together, makes them stronger. In the next subsection, we present this structure of shrinkage prior to the IV regression context.

5.3.1 Bayesian regularization methods in IV regression

When dealing with measurement error, instrumental variable regression may be used. Consider the model:

$$x_t = z_t'\delta + \varepsilon_{x_t}, \tag{5.5}$$

$$y_t = \gamma + x_t\beta + \varepsilon_{y_t}, \quad t \in \{1, \cdots, n\} \tag{5.6}$$

where x_t are the endogenous or treatment variable, z_t is a $(p \times 1)$ vector of instruments, y_t is the response variable, and it is supposed that

$$\begin{bmatrix} \varepsilon_{x_t} \\ \varepsilon_{y_t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_x^2 \end{bmatrix}\right)$$

Since p may be large, some regularization on Equation (5.5) is necessary. The Bayesian solution to this problem is to impose shrinkage priors on δ to shrink those parameters which have little power to explain x_t. By imposing such a prior, the usual Gibbs sampler scheme (LOPES; POLSON, 2014) used to estimate model (5.5)-(5.6) cannot be employed. Hahn, He and Lopes (2018) developed an elliptical slice sampler that can deal with arbitrary priors on δ, allowing us to use shrinkage prior, such as Laplace distribution, as well as the factor-based-prior also developed by the same authors. Then, it is instructive to describe the estimation for an arbitrary prior distribution on δ. To understand the

Bayesian estimation of IV regression, consider the reduced form of equations (5.5) and (5.6):

$$x_t = z_t'\delta + \nu_{x_t} \tag{5.7}$$
$$y_t = \gamma + z_t'\delta\beta + \nu_{y_t} \tag{5.8}$$

where $\nu_{x_t} \equiv \varepsilon_{x_t}$ and $\nu_{y_t} \equiv \beta\varepsilon_{x_t} + \varepsilon_y$. Defining $T = \begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix}$ implies that:

$$\Omega \equiv \text{Cov}\left(\begin{bmatrix} \nu_{x_t} \\ \nu_{y_t} \end{bmatrix}\right) = TST' = \begin{bmatrix} \sigma_x^2 & (\alpha+\beta)\sigma_x^2 \\ (\alpha+\beta)\sigma_x^2 & (\alpha+\beta)^2\sigma_x^2 + \xi^2 \end{bmatrix},$$

with $\alpha \equiv \frac{\sigma_y}{\sigma_x}\rho$ and $\xi^2 \equiv (1-\rho^2)\sigma_y$. Note that, the parameters to be estimated are $\Theta = (\sigma_x^2, \delta, \xi^2, \gamma, \beta, \alpha)$. Then, conditional on the set of instruments, the likelihood function may be written as:

$$f(x, y|z, \Theta) = f(y|x, \delta, \alpha, \beta, \xi^2) \times f(x|z, \sigma_x^2\delta) \tag{5.9}$$
$$= N\left(\gamma + x_t\beta + \alpha(x_t - z_t'\delta), \xi^2\right) \times \mathcal{N}(z_t'\delta, \sigma_x^2). \tag{5.10}$$

This decomposition of the likelihood function allows us to form a Gibbs sampler scheme by choosing the following prior distributions:

$$\delta \sim \text{arbitrary}, \tag{5.11}$$
$$\sigma_x^2 \sim \mathcal{IG}(\text{shape} = k_x, \text{scale} = s_x), \tag{5.12}$$
$$(\xi^2, \gamma, \beta, \alpha)' \sim \mathcal{NIG}\left(0, \xi^2\Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2}\right). \tag{5.13}$$

Combining these priors with the likelihood function in Equation (5.10), gives us the posterior distribution. To sample from this posterior distribution, it is possible to break it into three full conditional posteriors to form a Gibbs sampler scheme. To explain these three blocks, it is useful to introduce some definitions. Define[1] $\tilde{x} \equiv (1, x, z'\delta)$, $M \equiv \Sigma_0 + \tilde{x}'\tilde{x}$, $a \equiv k + n$ and $b \equiv s + y'y - y'\tilde{x}M^{-1}\tilde{x}'y$. It is possible to show that $f(y|x, z, \delta) \propto |M|^{-\frac{a}{2}} b^{-\frac{1}{2}}$. With these definitions, we can describe each of these blocks:

Full conditional posterior for $\delta|\Theta$, data: given Θ, from Equations (5.10) and (5.11) the conditional posterior is proportional to $f(x|\Theta)f(y,|x,\Theta)\pi(\delta)$. Since we are considering an arbitrary prior for δ, this full conditional posterior may not have a closed form, requiring alternative methods to sample it. Although traditional Metropolis-Hastings can

[1] In general, the notation with no subscript t means that the variable contains all observations. For example, $x \equiv (x_1, x_2, ..., x_n)$. A particular case is the vector of instruments, z_t, for which all observations are denoted by Z since it becomes a matrix of dimension $(p \times n)$.

be used in this case, it scales poorly due to possibly high dimension and multimodality of the full conditional posterior. Instead, Hahn, He and Lopes (2018) proposed to sample it using an elliptical slice sampler, which only requires the ability to evaluate $\pi(\delta)$. This algorithm is described below:

Algorithm 1 Elliptical Slice Sampler

procedure Slice Sampler$(\delta, \sigma_x^2, x, Z, y)$
  Define $\hat{\delta} = (ZZ')^{-1}Zx$ and $\Delta = \delta - \hat{\delta}$
  Draw $\zeta \sim \mathcal{N}(0, \sigma_x(ZZ')^{-1})$ and $v \sim U(0,1)$
  Compute $\ell \equiv \log(f(y|x, Z, \delta) + \log(\pi(\delta)) + \log(v)$
  Draw an angle $\varphi \sim U(0, 2\pi)$, and do $\mathsf{Lower} \leftarrow \varphi - 2\pi$ and $\mathsf{Upper} \leftarrow \varphi$
  Update Δ and δ: $\bar{\Delta} = \Delta\cos(\varphi) + \zeta\sin(\varphi)$ and $\bar{\delta} = \hat{\delta} + \Delta$
  while $\log(f(y, x, Z, \bar{\delta})) + \log(\pi(\bar{\delta})) < \ell$ **do**
    If $\varphi < 0$, then $\mathsf{Lower} \leftarrow \varphi$. Else $\mathsf{Upper} \leftarrow \varphi$
    Draw a new angle $\varphi \sim U(\mathsf{Lower}, \mathsf{Upper})$
    Update Δ and δ: $\bar{\Delta} = \Delta\cos(\varphi) + \zeta\sin(\varphi)$ and $\bar{\delta} = \hat{\delta} + \Delta$
  $\delta \leftarrow \hat{\delta} + \bar{\Delta}$.
  return δ

Note that the only requirement of the Algorithm (1) is the ability to evaluate the prior density $\pi(\delta)$.

Full conditional posterior for $\sigma_x^2|\Theta,$ data: Fortunately, for an inverse-gamma prior on σ_x^2 the full conditional posterior for σ_x^2 have a closed form. Combining the likelihood $f(x|Z, \delta, \sigma_x^2)$ with the prior given in (5.12), it is possible to show that the full conditional posterior is an Inverse-Gamma with shape parameter $k_x + n$ and scale $s_x + \sum_{i=1}^n (x_t - z_t'\delta)^2$ (see proof in Appendix).

Full conditional posterior for $(\xi^2, \gamma, \beta, \alpha)'|\Theta,$ data: This block also have a closed form. By using the bivariate normal properties, we can write the likelihood in terms of the transformed variable $\tilde{x}$, which follows $y_t|\tilde{x}_t \sim N(\tilde{x}_t\theta, \xi^2)$. Combining this likelihood with the prior give in (5.13), it can be shown that the full conditional posterior $(\xi^2, \gamma, \beta, \alpha)'|\Theta,$ data follows a Normal-Inverse-Gamma distribution. Specifically:

$$(\xi^2, \gamma, \beta, \alpha)'|\Theta, \text{data} \sim \mathcal{NIG}\left(M^{-1}\tilde{x}'y, \xi^2 M^{-1}, \text{shape} = \frac{a}{2}, \text{scale} = \frac{b}{2}\right).$$

(See Appendix for proof).

We can use these full conditional posteriors to form a three-block Gibbs sampler, by iterative sampling over the blocks. This methodology is interesting because we can choose arbitrary priors for δ, and it still works well. In particular, we can elicit several shrinkage priors over δ, since the many instruments setting requires regularization.

5.3.2 Shrinkage priors for instruments coefficients

There is a large range of shrinkage priors in the literature (ERP; OBERSKI; MULDER, 2019). The underlying idea of these priors is to give a higher prior probability around zero, such that if the parameter is not too important, it shrinks to zero. In what follows, we present some of these priors that can be directly applied to $\delta = (\delta_1, \cdots, \delta_p)$. Then, we proceed with the factor-based prior distribution.

5.3.2.1 Heavy-tailed priors

Popular choices of shrinkage priors are the Cauchy, Laplace, and Horseshoe densities (CARVALHO; POLSON; SCOTT, 2010). The horseshoe density is the stronger one, in the sense that it concentrates high probability density around zero. In the same sense, the Laplace prior is strong, while the Cauchy density is relatively weaker, although it also concentrates much density around zero, and so it also works as a shrinkage prior. The left panel of Figure 5.1 depicts these three priors.

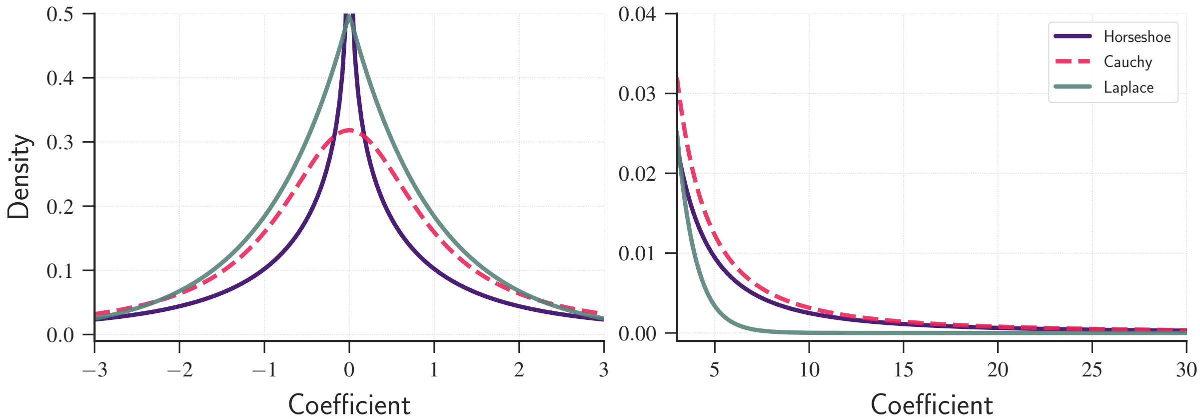

Figure 5.1 – Three examples of shrinkage priors: Horseshoe, Double-exponential, and Cauchy, all of them centered at zero.

An advantage of the Horseshoe prior is that at the same time, it concentrates high probability density around zero, shrinking unimportant coefficients to this point, it also has heavy tails. The right panel of Figure 5.1 compares the right tails of the three priors considered here. Notice that the tail of the Horseshoe prior is above the Laplace and closer to the Cauchy tail. This heavy tail allows the identification of parameters that are different from zero.

In the IV regression case, we can choose one of these priors for each δ_j and assume that they are independent for all $i \neq j$, with $i, j \in \{1, \cdots, p\}$. Although it may work, it neglects the covariance between the instruments. To consider the covariance between the instruments, a more sophisticated prior is required, and this subject is discussed in the next subsection.

5.3.2.2 Factor-based shrinkage prior

The idea underlying the factor-based prior, proposed by Hahn, He and Lopes (2018), is to explore the covariance of instruments to extract factors that represent 'strong' instruments. To formalize this intuition, consider the following decomposition of the covariance matrix of instruments:

$$\text{Cov}(z_t) = BB' + \Psi^2, \tag{5.14}$$

where, B is a $(p \times k)$ matrix and Ψ^2 is a diagonal $(p \times p)$. Despite the fact that every covariance matrix admits this decomposition, the interest here is on the case where $k << p$, where k represents the number of factors to be extracted, denoted by f_t. Suppose that the instruments z_t and the factors f_t are jointly Normally distributed as follows:

$$\begin{bmatrix} z_t \\ f_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} BB' + \Psi^2 & B \\ B' & I_k \end{bmatrix} \right).$$

This assumption implies that $\mathbb{E}[f_t|z_t] = Az_t =: \hat{f}_t$, with $A \equiv B'(BB' + \Psi^2)^{-1}$.

Now, consider the factor regression model:

$$x_t = \theta \hat{f}_t + \varepsilon_t, \tag{5.15}$$

where θ is a $(1 \times k)$. From Equation (5.15) and the definition of $\hat{f}_t$ it is possible to show that $\delta' = \theta A$. However, this specification is only correct if δ lies in the row space of A; otherwise, the model is misspecified. Then, it is necessary to extend the model to include the possibility that δ lies in the row space of A. For that end, the specification in Equation (5.15) needs to be modified to

$$x_t = \theta \hat{f}_t + \eta \hat{r}_t, + \varepsilon_t \tag{5.16}$$

where η is a $(1 \times p)$ vector of parameter and $\hat{r}_t \equiv (I_p + A^+ A)z_t$ and A^+ denote the Moore-Penrose pseudo-inverse of A. In this case, it can be shown that $\delta' = \theta A + \eta(I_p + A^+ A)$.

Defining $\tilde{\delta}' = (\theta, \eta)$, we note that $\delta = H\tilde{\delta}$, where

$$H' = \begin{bmatrix} A \\ I_p + A^+ A \end{bmatrix}.$$

Consequently, we can rewrite (5.16) as $x_t = H\tilde{\delta} z_t + \varepsilon_t$. Assuming we know A (and then H), this specification allows us to attribute a prior over δ by imposing strong shrinkage prior over $\tilde{\delta}$. If we solve the system $\delta = H\tilde{\delta}$, using the theory of pseudo-inverses, we have $\tilde{\delta} = H^+\delta + (I_{p+k} + H^+ H)\omega$, for an arbitrary vector ω. With this identity, conditional on

ω, we can impose a Horseshoe prior, for instance, on $\tilde{\delta}$ and it induces a prior on δ. That is:

$$\pi(\delta|\omega) = \prod_{j=1}^{k+p} \left\{ \left(2\pi^3\right)^{-\frac{1}{2}} \log\left(1 + \frac{4}{\tilde{\delta}_j^2}\right) \right\}$$
$$\pi(\delta|\omega) = \prod_{j=1}^{k+p} \left\{ \left(2\pi^3\right)^{-\frac{1}{2}} \log\left(1 + \frac{4}{\left\{\left[H^{+}\delta + \left(I_{k+p} - H^{+}H\right)\omega\right]_j\right\}^2}\right) \right\}.$$

Following Hahn, He and Lopes (2018), we assume that $\omega \sim \mathcal{N}(0, I_{k+p})$. Once we know ω, we can evaluate the prior $\pi(\delta|\omega)$, which is the only requirement of the slice sampler presented in Algorithm 1. Then we can sample δ by inducing a prior on δ via horseshoe prior over $\tilde{\delta}$. Note that, under this specification, the factor structure derived in this section is taken into account in the prior over δ. In practice, however, the matrices B, Ψ are unknown and, consequently, A and H is also unknown. Instead of estimating it in a Bayesian fashion, we use point estimates of these matrices, which is found by minimizing the trace of $\text{Cov}(z_t) - D$, by choosing D, subject to D be diagonal and positive-definite.

Finally, for all shrinkage priors, we incorporate a global shrinkage parameter λ. Introducing this as a parameter in the model is a key feature in Bayesian regularization because it avoids procedures like cross-validation or setting it as a fixed parameter. We sample the global shrinkage parameter via a Metropolis-Hastings step.

5.3.3 Data

To analyze whether our empirical method performs well, we start by using it in simulated data by means of a Monte Carlo analysis. When we know the true generating process, we can calculate the error of the estimate and compare it with alternative methodology (for instance, OLS, 2SLS, etc.). Besides, in the simulation exercise, we also apply the empirical method in real financial data. To estimate the CAPM, we need asset return data, a surrogate for the market return, and a risk-free asset. As the risk-free asset, we consider the one-month Treasury Bill rate and take the surrogate of the market return data from Kenneth French's web site[2]. Finally, we consider the return of 277 stocks listed in the S&P500 with data availability in the last five years. All data is daily and ranges from 2017-01-01 to 2021-12-31, resulting in 1,260 observations.

[2] The return on market return and the risk-free asset is taken form https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. The return of the stocks is based on the close price of the stocks and is taken from yahoo finance.

5.4 Results

In this section, we describe and discuss the results of our paper. We begin by describing the outcome of a simulation exercise, in which we compare the Bayesian regularization discussed in the previous section with traditional ordinary least squares and two-stage least squares estimation. Then, we present the result of CAPM estimation for observed data using the proposed Bayesian shrinkage approach.

5.4.1 Monte Carlo analysis: Simulation procedures

To simulate the CAPM we consider a classical additive measurement error model, as follows:

$$x_t = x_t^* + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2) \text{ and } x_t^* \sim \mathcal{N}(0, \sigma_x^2), \tag{5.17}$$

$$y_{it} = \beta_i x_t^* + \varepsilon_{it}, \tag{5.18}$$

for $i \in \{1, ..., p+1\}$ and $t \in \{1, ..., n\}$. In Equations (5.17) and (5.18) x_t^* is the true market return, and it is assumed to be Gaussian with mean zero and variance σ_x^2, u_t is Gaussian measurement error, with mean zero and variance σ_u^2, x_t is the observed market return. The sensitivity to the market return, which is measured by β_i, is assumed to be known, and based on these values and given the error term ε_t, we construct the assets return, as described in Equation (5.18). The error term ε_t is also assumed to be Gaussian, with mean zero and variance σ_ε^2.

To simulate the model, we need to calibrate the parameters $(\sigma_u^2, \sigma_x^2, \sigma_\varepsilon^2, \beta_i)$. For the β_i, we consider a linear grid between 0.3 and 1.5. Based on data of a proxy of market return, we calibrated $\sigma_x = 0.01$. We calibrated one of the assets with $\sigma_\varepsilon = 0.001$ and the other $p-1$ with $\sigma_\varepsilon = 0.9$. We use these different values for σ_ε to create one strong instrument and $p-1$ weak instruments. A lower standard deviation creates assets that will be stronger instruments than those with a higher standard deviation. Besides these p assets, we consider an additional one with $\beta = 1$ and $\sigma_\varepsilon = 0.04$, which is used as a target variable in the CAPM estimation. Finally, we calibrate $\sigma_u = 2\sigma_x$ to create a situation with a high measurement error. These parameters calibration allow us to simulate all variables of interest in CAPM. We set the hyper-parameters of the remaining prior distribution parameters to reflect a diffuse prior.

To evaluate the accuracy of each method, we simulate the model $N_{\text{sim}} = 1,000$ times. At each iteration, we estimated the parameters using six methods. The first one was the traditional Ordinary Least Squares (OLS), regressing y_{it} on x_t, which is known to be inconsistent in the presence of measurement errors. Second, we consider the two-stage least squares (2SLS), using all asset returns but the regressand in the CAPM Equation as instruments (see (MENG; HU; BAI, 2011), for a similar approach). We believe that these variables satisfy the requirement of an instrument: they are correlated with the market return but uncorrelated with the error term in the CAPM equation. Third, we use the Limited Information Maximum Likelihood (LIML) (FULLER, 1977) estimator with the same set of instruments. Although our main interest is to verify whether the Bayesian regularization of the two-stage can improve the inference about beta, we include the results of the LIML estimator since it is known to be unbiased in the presence of many instruments (HANSEN; HAUSMAN; NEWEY, 2008) [3]. In the last three methods, we consider the same set of instruments to estimate the model using the Bayesian method described in Section 5.3. We use the Horseshoe, Laplace, and Factor-Based Shrinkage prior distributions over δ. This estimation is referred to as BHS, BLA, and BFB.

Table 5.1 – Measures for the beta estimation error, $n = 1000$

		OLS	2SLS	LIML	BHS	BLA	BFB
$p = 2$	Mean bias	-0.798	**-0.001**	0.003	0.007	0.012	0.013
	Mean abs. bias	0.798	0.633	**0.117**	0.117	0.119	0.119
	RMSE	0.800	4.970	**0.143**	0.144	0.146	0.146
$p = 10$	Mean bias	-0.799	-0.029	0.007	0.017	0.018	**-0.005**
	Mean abs. bias	0.799	0.123	0.121	0.120	0.123	**0.119**
	RMSE	0.801	0.154	0.151	0.151	0.155	**0.149**
$p = 20$	Mean bias	-0.799	-0.066	**0.012**	0.021	0.023	-0.055
	Mean abs. bias	0.799	0.126	0.122	0.121	0.125	**0.120**
	RMSE	0.801	0.158	0.155	0.155	0.160	**0.152**
$p = 40$	Mean bias	-0.798	-0.146	**0.003**	0.010	0.006	-0.166
	Mean abs. bias	0.798	0.167	0.126	**0.122**	0.159	0.181
	RMSE	0.800	0.198	0.161	**0.158**	0.308	0.211
$p = 80$	Mean bias	-0.803	-0.252	0.017	**0.008**	-1.133	-0.320
	Mean abs. bias	0.803	0.253	0.135	**0.129**	1.234	0.320
	RMSE	0.805	0.279	0.172	**0.163**	1.444	0.339
$p = 160$	Mean bias	-0.799	-0.407	**0.010**	-0.079	-1.149	-0.508
	Mean abs. bias	0.799	0.407	0.158	**0.138**	1.153	0.508
	RMSE	0.801	0.421	0.199	**0.168**	1.249	0.517

Note: We report mean bias, mean absolute bias, and the root mean of squared error (RMSE) of each estimator. We highlight in bold type the best estimator for each criterion and number of instruments.

We simulate and estimate the model for different numbers of assets and hence different numbers of instruments. Specifically, we start with $p = 2$, and then increase it to 10, 20, 40, 80, and 160. In the estimation process, we consider the asset with $\beta = 1$. Because of measurement error, the OLS estimator is downward biased. To assess the ability of each estimation method to correct this bias, we use three criteria: mean bias,

[3] The LIML estimator, however, is also known to have no moments. The modification version of LIML, due to Fuller (1977), solves this drawback. Still, the modification introduces an additional parameter that must be chosen by the econometrician.

mean absolute bias, and the root mean of squared error (RMSE). According to these criteria, the lower its absolute value, the better the estimator.

Table 5.1 summarizes the Monte Carlo results. In this table, we highlight in bold the best estimator for the criterion and the number of instruments. The Bayesian "two-stage" procedure with Horseshoe regularization prior (BHS) is better than traditional 2SLS for all criteria and choices of p, except for the mean bias criteria and $p = 2$. This result shows that the regularization over the instrumental variables indeed avoids bias in the presence of many weak instruments. The mean bias of BHS is closer to the mean bias of LIML. For $p = 80$, it is even smaller than those of LIML, which corrects the many instruments bias (HANSEN; HAUSMAN; NEWEY, 2008). We note that, for the other values of $p \neq 80$, the LIML has a smaller mean bias. However, considering the mean bias and the root of mean squared error, which also penalizes for variance, the Bayesian approach always dominates the LIML, except for $p = 2$. Thus, we conclude that the Bayesian regularization in the estimation of the CAPM model improves the inference about the betas in two ways: decreasing the bias and reducing the estimates' variance.

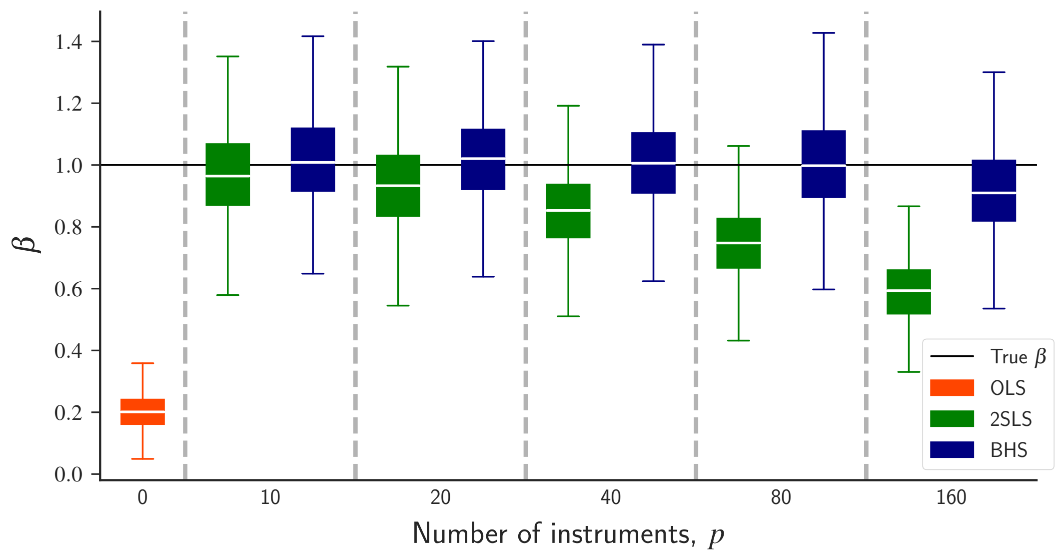

Figure 5.2 – Boxplots of the estimated CAPM betas for different numbers of instruments: Comparison between Bayesian Horseshoe (BHS) and Two-Stage Least Squares (2SLS). The horizontal black line represents the true CAPM beta value, which is one

Among the three types of regularization priors, the horseshoe outperforms the other at least in terms of mean bias, except for $p = 10$. For this reason, we now focus on the horseshoe prior in comparison with the traditional 2SLS. In Figure 5.2 we present the distribution of the BHS and 2SLS for several number of instruments, as well as the OLS estimates. For small sets of instruments, say up to 20, the 2SLS bias is small and can be entirely corrected by Bayesian regularization. When the number of instruments increases to 40 and 80, the bias becomes greater, and the Bayesian regularization still entirely corrects the bias, maintaining the distribution of the estimates around the true CAPM beta, which is one. The bias of the traditional 2SLS can be explained by the tendency of the OLS, in the first-stage, to fit too well, as the number of instruments increases, see (DAVIDSON; MACKINNON et al., 1993, pp. 222). The BHS penalizes the "first-stage" estimation, avoiding this tendency to overfitting, and hence reducing the bias (see Figure 5.2). The size of the correction diminishes when we increase the number of instruments to 160, but it is still better than 2SLS, as shown both by Figure 5.2 and Table 5.1.

In the appendix, we present a robustness check, in which we change some parameters of the simulated CAPM.

5.4.2 Empirical Application

This section uses the Bayesian regularization procedure to estimate the CAPM using observed data and compare it with the traditional 2SLS, LIML, and OLS estimators of CAPM beta. We estimate the model for 278 stocks listed in the S&P500 index. The instruments for each model consist of all other stock returns listed in the S&P500, but the one used in the CAPM target equation. The original set contains 366 stocks for the period between January 2017 to December 2021, but we exclude 88 stocks with a correlation above 0.7. This exclusion is needed to avoid numerical approximation errors in the inversion of ZZ', a requirement of Algorithm 1. All data is daily and ranges from January 2017 to December 2021, totaling $1,260$ observations.

The posterior distribution for CAPM beta was similar for the factor-based prior and Laplace prior (BFB and BLA, see Figure 5.6 in Appendix). This result indicates that, for these stocks, these two priors perform equally. The reason why the Factor-Based shrinkage prior cannot do a better job than the straight horseshoe prior may be related to the covariance structure of the instruments. Indeed, the eigenvalues of the covariance matrix decay drastically from the first to the second eigenvalues and then decline slowly. The covariance of the minimized-trace (the $\text{Cov}(z_t) - D$ discussed in Section 5.3.2.2) presents a similar behavior, see Figure 5.7 in Appendix. Thus, we cannot isolate "commonalities" and, hence, the prior information is unable to help in the shrinkage of parameters. The Horseshoe prior presents a subtly different result, compared to BLA and BFB (see Figure 5.6 in Appendix). We focus on the Horseshoe prior because of their better performance in the simulation exercise.

When comparing the Bayesian IV estimates with OLS and 2SLS estimates, we find different results across the stocks. While for some stocks the Bayesian IV delivers greater betas than the other methods for some stocks, there is also a group of stocks whose BIV estimates are less than (or very similar to) the unregularized estimates. On one hand, a beta greater than the OLS is a better estimate, since the OLS is downward biased in the presence of EIV. On the other hand, the appearance of betas smaller than the OLS may reflect another hypothesis that may be failing in the model, such as an omitted variable.

Figure 5.3 brings two examples of stocks where BIV delivers a greater beta than the other estimators. The stocks are Nvidia (NVDAO) and AMD, which are enterprises from the technology sector and present the highest return in our data set. Since they possess the high returns we expected, according to the CAPM theory, they hold higher systematic risk and, consequently, a higher beta.

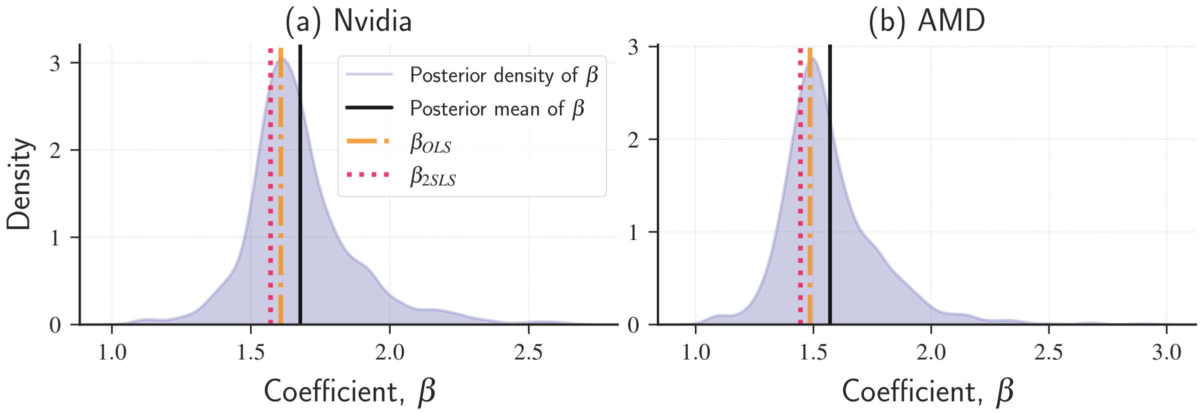

Figure 5.3 – Posterior distribution of CAPM beta estimated by Bayesian Instrumental Variable with Horseshoe prior for three assets: (a) Nvidia and (b) AMD. The figure also presents the mean of the posterior distribution and the OLS and 2SLS estimates.

For both the Nvidia and AMD stock, the 2SLS and OLS estimates are similar, though the 2SLS is slightly smaller. It is an expected result since when the number of the instruments is large (as in this case, $p = 277$) relative to the number of observations ($n = 1,260$), the 2SLS estimates tend to the OLS estimates[4]. In addition, since the market return contains measurement error, we know that the OLS estimates are downward biased and, consequently, so are the 2SLS estimates.

The Bayesian approach, in turn, delivers a greater CAPM beta for the two stocks, considering the posterior mean as a point estimation (see black line in Figure 5.3). For Nvidia stock, while $\hat{\beta}_{OLS} = 1.61$ and $\hat{\beta}_{2SLS} = 1.57$ the posterior mean of beta is 1.68, using the Horseshoe prior. For AMD stock, while $\hat{\beta}_{OLS} = 1.49$ and $\hat{\beta}_{2SLS} = 1.44$ the posterior mean of beta is 1.68. Although we do not know the true beta in this case, we know that OLS is downward biased, which puts our approach in a better position. Even though the difference between the BIV and 2SLS beta estimates is not too stark, these discrepancies in estimated beta may have drastic implications for finance practitioners. As noted by Malloch, Philip and Satchell (2021), some analysis in finance, such as valuation, is very sensitive to the estimated beta. Thus, the many instruments approach with shrinkage priors can offer a better way to make such financial analysis.

[4] In the extreme case where $p = n$ the two estimators are equivalent.

The above discussion only considers two stocks with the highest return of the analyzed data set. We extend this analysis to include all 277 stocks and, hence, consider all the differences among the different estimators. To do so, we analyze how the estimated betas of each estimator can explain the heterogeneity in the cross-section return by running the second-pass regression of the two-pass procedure proposed by Fama and MacBeth (1973). The idea of the second pass regression is to regress the cross-section of the returns against the betas. Thus, we can verify how much of the cross-sectional variation in the returns is explained by the betas. Specifically, we follow an approach similar to Lettau and Ludvigson (2001), regressing:

$$\bar{R}_i = \lambda_0 + \lambda_1 \hat{\beta}_i + u_i \tag{5.19}$$

where $\bar{R}_i \equiv \sum_{t=0}^{n} R_{it}$.

The second-pass regression also suffers from the EIV problem because $\hat{\beta}_i$ carries error estimation from the first-pass (the time-series regression). As argued in the simulation exercise, our approach can improve the estimation of the betas both in terms of bias and variance. Thus, a more efficient estimate and possibly bias-corrected can improve the estimation of the second-pass, alleviating its EIV problem. Traditionally, the two-pass procedure is done by grouping the stock in portfolios to avoid the EIV in the second pass [5]. Since our approach improves the estimation in the first step, we consider the two-pass procedure for stocks instead of for portfolios. We also consider, however, the analysis for the portfolio to allow some comparison to previous literature.

Figure 5.4 plots the results of the second-pass regression using the betas from 2SLS and from BHS. In this figure, the horizontal axis measures the realized return, and the vertical axis the fitted return obtained from the regression (5.19). If the CAPM fitted the data perfectly all points should lie exactly in the 45-degree line. The CAPM, however, has been challenging in explaining the cross-section return. In part (a) of Figure 5.4, we observe that the return of the market portfolio can explain little of the cross-section variation of stock returns when we use the betas from 2SLS ($R^2 = 0\%$). Using betas from BHS, in part (b) of Figure 5.4, we can observe an increase in the explanation of cross-section return, with a $R^2 = 3\%$. This power of explanation is still low compared to alternative models in the literature, such as multi-factor models, but it represents some improvement in the explanation of heterogeneity of cross-section return by the market factor.

[5] Grouping stocks in portfolios have some shortcomings. See Jegadeesh et al. (2019) for a discussion.

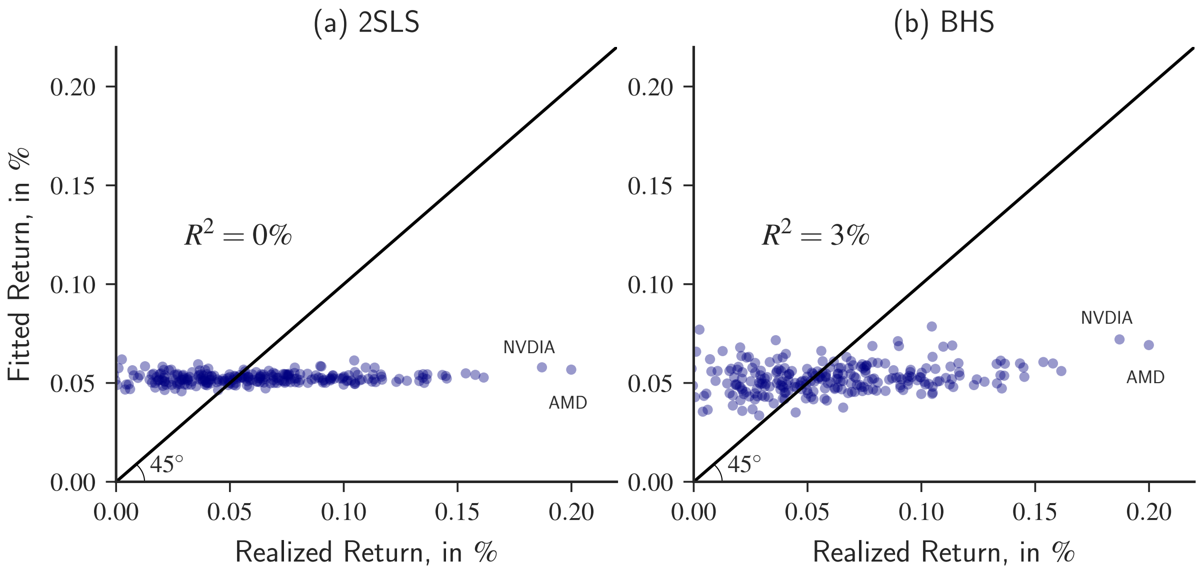

Figure 5.4 – Realized versus fitted *stocks* return: (a) fitted return obtained from 2SLS betas estimates; (b) fitted return obtained from BHS.

We can interpret this result as a "possible improvement in the measure". Using many instruments and regularization techniques, we can directly deal with the EIV of the return on the market portfolio, alleviating the problems caused by the EIV. In a certain sense, our approach deals with Roll (1977) critique: even though we cannot measure the return on the market portfolio exactly, we can use a data-rich environment to treat the mismeasured variable. The shrinkage approach makes it possible to downplay the significance of unimportant instruments and use all the relevant information in the data. This better use of the data allows for a better estimate of the betas in the first step; hence, the second step can explain a higher variation of the cross-section return.

To illustrate the argument above, we highlight in Figure 5.4 the points representing the AMD and Nvidia stocks. Notice that these two stocks present the highest realized return among all stocks. They are, however, distant from the 45-degree line. For a given estimate of λ_0 and $\lambda_1 > 0$ [6], a higher Nvidia or AMD beta would increase its fitted return, bringing the points of these stocks closer to the 45-degree line. As shown in Figure 5.3, the point BHS estimate of the CAPM beta is greater than the 2SLS, which brings the points closer to the 45-degree line and increases the R^2.

We also analyze the two-pass procedure using portfolios instead of stocks. Specifically, we consider the 25-Fama-French portfolios sorted by size and book-to-market to run the Fama and MacBeth (1973) two-step procedure, as in the case of the stocks presented above. The portfolio data ranges from 1963-Q3 to 2019-Q4. Following Lettau and Ludvigson (2001) and Meng, Hu and Bai (2011), we use the data in quarterly frequency.

[6] We note, however, that different values of β_i would affect the estimates of λ_j, $j = 0, 1$. We also note that, in our case, the estimated λ_0 and λ_1 are positive both for betas from 2SLS and betas from BHS, with the last being greater than the former.

Figure 5.5 presents the fitted return versus the realized return of the 25 portfolios. Comparing the results using the betas from 2SLS in part (a) with the results using BHS in part (b) of Figure 5.5, we again observe an increase in the R^2. That is, using the betas from the BHS increases the percentage of variation in the portfolio returns explained by the market beta.

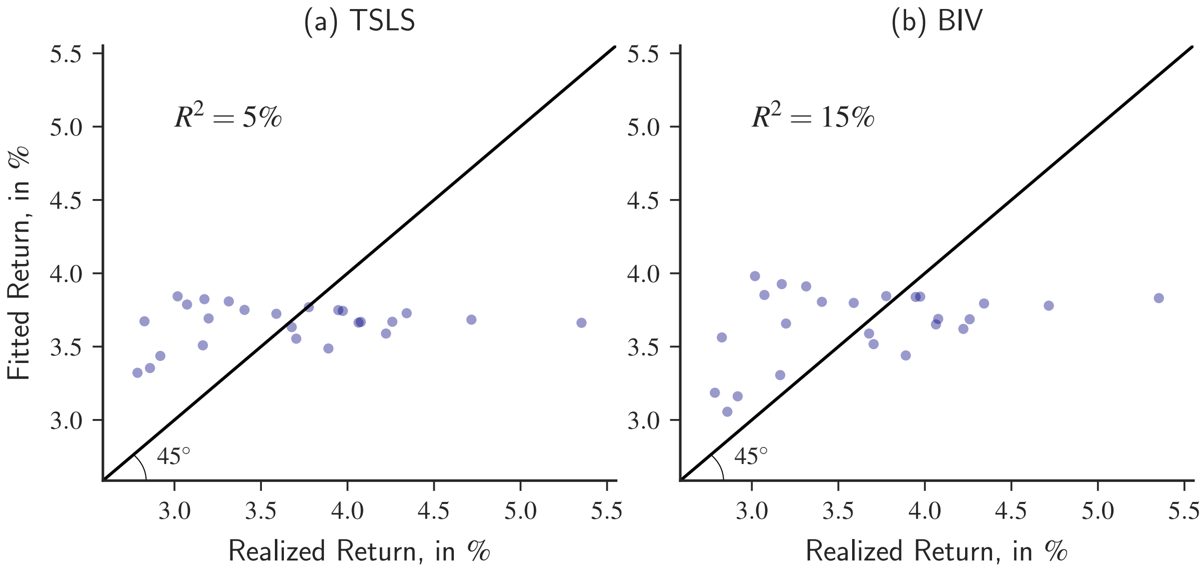

Figure 5.5 – Realized versus fitted *portfolio* return: (a) fitted return obtained from 2SLS betas estimates; (b) fitted return obtained from BHS.

A large body of the literature documented the CAPM's incapability to explain the heterogeneity of cross-section average return (FAMA; FRENCH, 1992; FAMA; FRENCH, 1993). By using methods that ignore the EIV problems, such as OLS, or instrumental variables without regularization techniques and high-dimensional settings, previous literature has reported that the CAPM can explain around only 1% of the variation in cross-section return of 25 Fama-French portfolios (LETTAU; LUDVIGSON, 2001; MENG; HU; BAI, 2011). In contrast, by exploring the high-dimensional settings in instrumental variables, our approach shows that CAPM betas can explain 16% of the variation in the return of 25 Fama-French portfolios.

These findings provide fresh insight into the estimate of asset pricing models with measurement errors. We can lessen the error-in-variables problem in the assessment of systematic risk, which in turn will better explain the cross-section of return, by employing high-dimensional data and appropriate approaches. Although the power of explanation of the cross-section of returns reported in our empirical application is modest compared to multi-factor models, we emphasize that the inclusion of several instruments and regularization priors might increase this power relative to unregularized techniques. Therefore, using a variety of instruments with regularization priors can be beneficial in other contexts as well, particularly in asset pricing models that incorporate variables with significant measurement errors, such as macroeconomic variables.

5.5 Conclusion

This paper contributes to the literature that works on the solution of Roll (1977) critique, in particular the strand of literature that uses the instrumental variable. The data-rich environment available in finance offers a wide range of instruments to solve the error-in-variable present in the capital asset pricing model. These instruments, however, are usually weakly correlated with the endogenous market return, and using too many instruments may induce bias. This paper proposes to estimate the capital asset pricing model using a large set of instruments and shrinkage priors over the parameters associated with the instruments.

In a simulation exercise, the results show that the proposed approach reduces the bias of the CAPM beta estimates. In the empirical application, we argue that the high-dimensional settings can also improve the estimation of CAPM. We verify empirically that for some stocks our proposal produces greater betas than the OLS and 2SLS methods. Since OLS is downward biased, the regularized Bayesian approach delivers better results than the traditional 2SLS for this application. This find is supported by a large literature that shows beta's incapacity to appropriately measure the magnitude of risk. This difference in estimated betas is economically important since many financial models are sensitive to beta. Moreover, for both stocks individually and for portfolio analysis, our approach helps to explain the heterogeneity of the cross-section of return.

Future research can extend the results of this work in at least two ways. The first is to generalize the model presented here to consider the joint estimation of all beta assets. This generalization can be a way to increase the efficiency of the estimates since it will consider the covariance of all stock returns in the estimation. Second, a limited information Bayesian approach may be designed to also include prior regularization, in an analogous manner as presented here.

An important comment about the contribution of our work is its general application in more general contexts than CAPM estimation. Note that the same estimation methodology can be applied in the context of multifactor risk models, especially in the construction of risk factors based on asset sorting using characteristics such as the so-called Fama-French factors (FAMA; FRENCH, 1993). These factors are built by ordering the assets based on observable characteristics, and then building the risk factors based on quantile separation. Note that in this procedure there are several possible sources of measurement risk, such as the frequency of reordering based on the observed characteristic, the choice of quantile level for separating the portfolios, etc. These ad hoc choices can introduce a component of measurement error in the construction of these risk factors, as discussed for example in Jegadeesh et al. (2019) and Giglio and Xiu (2021).

The primary focus of our work was to analyze the performance and properties of Bayesian estimation methods when dealing with measurement errors and utilizing shrinkage within the CAPM framework. However, the natural extension of our findings lies in the realm of risk pricing for multifactor models, where the presence of contaminations by measurement errors can lead to biased estimates of risk premia.

In conclusion, our research not only addresses CAPM estimation but also opens up avenues for improving risk pricing methodologies in more complex multifactorial contexts by considering and mitigating the impact of measurement errors.

5.6 Appendix

5.6.1 Full conditional posterior

First full conditional posterior $\sigma_x^2|\delta, \beta, \alpha, \xi^2$

From Equation (5.5) we know that

$$f(x_t|z_t, \delta, \sigma_x^2) = \mathcal{N}(z_t'\delta, \sigma_x^2),$$

which represents the likelihood. It implies that the density of $x \equiv (x_1, ..., x_n)'$ given $z \equiv (z_1, ..., z_n)'$ is:

$$f(x|z, \delta, \sigma_x^2) \propto (\sigma_x^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_x^2}(x - z'\delta)'(x - z'\delta)\right\}.$$

Consider the inverse-gamma prior with $s_x/2$ scale parameter and $k_x/2$, it follows that the conditional posterior is:

$$\pi(\sigma_x|x, y, Z, \delta) \propto (\sigma_x^2)^{\frac{-k_x-n-2}{2}} \exp\left\{-\frac{1}{2\sigma_x^2}\left[s_x + \sum_{i=1}^{n}(x_t - z_t'\delta)^2\right]\right\},$$

which is the kernel of an inverse-gamma.

Second full conditional posterior $\delta|\sigma_x^2, \alpha, \beta, \xi^2$

Use the elliptical slice sampler for this parameter, described in Algorithm 1.

Third full conditional posterior $(\gamma, \beta, \alpha, \xi^2)|\delta, \sigma_x^2, \alpha, \beta, \xi^2$

To simplify the notation, define $\theta \equiv (\gamma, \beta, \alpha)'$, which is a (3×1) vector. From eq. (5.5) and (5.6) and using the bivariate normal properties, we can find the conditional distribution $\varepsilon_{y_t}|\varepsilon_{x_t} \sim N\left(\alpha(x_t - z_t'\delta), \xi^2\right)$, where $\alpha \equiv \rho\frac{\sigma_y}{\sigma_x}$, $\xi^2 \equiv (1-\rho^2)\sigma_y^2$ and note that $\varepsilon_{x_t} = x_t - z_t'\delta$. Then, from eq. (5.6), we can conclude that: $y_t|x_t \sim \mathcal{N}\left(\gamma + x_t\beta + \alpha(x_t - z_t'\delta), \xi^2\right)$.

Define $\tilde{x}_t \equiv (1, x_t, x_t - z_t'\delta)$,which is a (1×3). It will be useful to consider $n \times 3$ matrix of observation: $\tilde{x} = (\tilde{x}_1', ..., \tilde{x}_n')'$. Thus, we can write $y_t|x_t \sim N\left(\tilde{x}_t\theta, \xi^2\right)$. The conditional likelihood will be:

$$f(y|x, Z, \cdots) \propto (\xi^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\xi^2}(y'y - y'\tilde{x}\theta - \theta'\tilde{x}'y - \theta'\tilde{x}'\tilde{x}\theta)\right\}$$

Combining this likelihood with the Normal-inverse-gamma prior:

$$\pi(\theta, \xi^2|\xi^2) \sim \mathcal{NIG}\left(0, \xi^2\Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2}\right)$$

and defining $a \equiv k + n$ and $M = \Sigma_0 + \tilde{x}'\tilde{x}$, allow us to find the third full conditional posterior:

$$\pi(\theta, \xi^2 | y, x, Z, \sigma_x^2, \delta) = \mathcal{NIG}\left(M^{-1}\tilde{x}'y, \xi^2 M^{-1}, \text{shape} = \frac{a}{2}, \text{scale} = \frac{b}{2}\right)$$

where $b \equiv s + y'y - y'\tilde{x}M^{-1}\tilde{x}'y$.

5.6.2 Robustness check

We also consider alternative calibration in the simulation exercise. Besides the results presents in the main text, we also consider a context in which the number of weak instruments and strong instruments are always equal, that is, half of instruments are strong, and half are weak. Specifically, we set the standard deviation of half of the assets to $\sigma_i = 0.01$ and the other half to $\sigma_i = 0.04$. We also consider a less intense measurement error, setting the standard deviation of measurement error to $\sigma_u = 0.9\sigma_x$. The calibration of the other parameters is the same as described in the main text. Table 5.2 presents the results for this setting.

Table 5.2 – Measures for the beta estimation error, $n = 1000$, equal number of weak and strong instruments

		OLS	2SLS	LIML	BHS	BLASSO	BFBS
$p = 2$	Mean bias	-0.449	1.009	**0.008**	0.022	0.018	0.019
	Mean abs. bias	0.449	2.000	0.230	0.236	**0.235**	0.236
	RMSE	0.458	28.312	0.288	0.297	**0.295**	0.296
$p = 10$	Mean bias	-0.452	-0.024	**-0.008**	-0.011	-0.017	-0.019
	Mean abs. bias	0.452	0.154	0.150	0.149	**0.148**	0.148
	RMSE	0.462	0.196	0.188	0.187	**0.185**	0.185
$p = 20$	Mean bias	-0.445	-0.013	0.009	**0.001**	-0.011	-0.014
	Mean abs. bias	0.445	0.124	0.125	0.123	**0.120**	0.120
	RMSE	0.455	0.158	0.158	0.156	0.154	**0.154**
$p = 40$	Mean bias	-0.449	-0.044	**-0.006**	-0.022	-0.044	-0.053
	Mean abs. bias	0.449	0.121	0.118	**0.117**	0.118	0.120
	RMSE	0.458	0.152	0.151	**0.147**	0.149	0.150
$p = 80$	Mean bias	-0.445	-0.070	**-0.004**	-0.042	-0.084	-0.105
	Mean abs. bias	0.445	0.115	0.109	**0.106**	0.117	0.128
	RMSE	0.454	0.145	0.137	**0.134**	0.148	0.159
$p = 160$	Mean bias	-0.450	-0.120	**0.000**	-0.114	-0.179	-0.221
	Mean abs. bias	0.450	0.143	**0.113**	0.136	0.184	0.222
	RMSE	0.460	0.171	**0.142**	0.163	0.209	0.244

Note: We report mean bias, mean absolute bias and the root mean of squared error (RMSE) of each estimator. We highlight in bold-type the best estimator for each criterion and number of instruments.

From Table 5.2 we can notice that the Bayesian horseshoe approach dominates the 2SLS for all criteria and choices of p, in line with the results in the main text. Also, except for $p = 2$ the horseshoe prior performs better than the other regularization priors in terms of mean bias. In this setting, however, the improvement of the Bayesian horseshoe in relation to the unregularized 2SLS is attenuated as the number of instruments increases.

5.6.3 Additional Bayesian estimation results

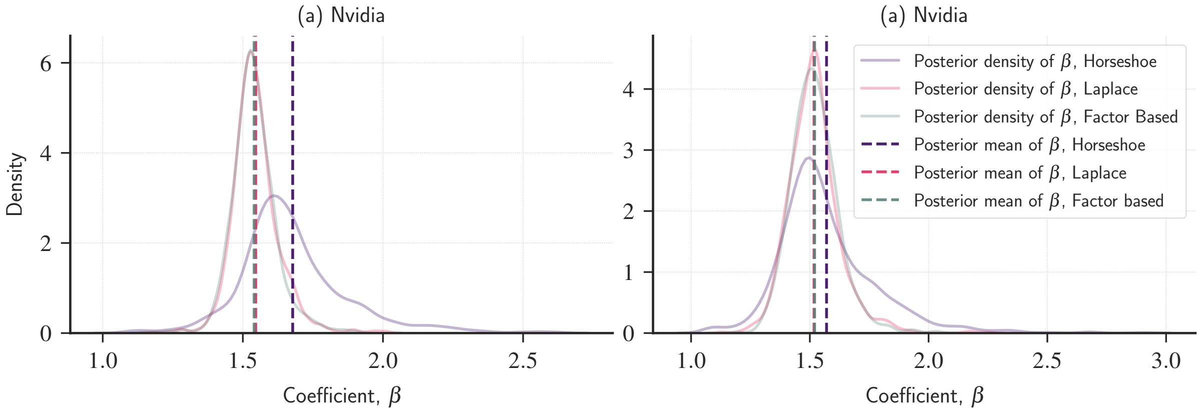

Figure 5.6 – Posterior distributions of beta for different the three types of priors.

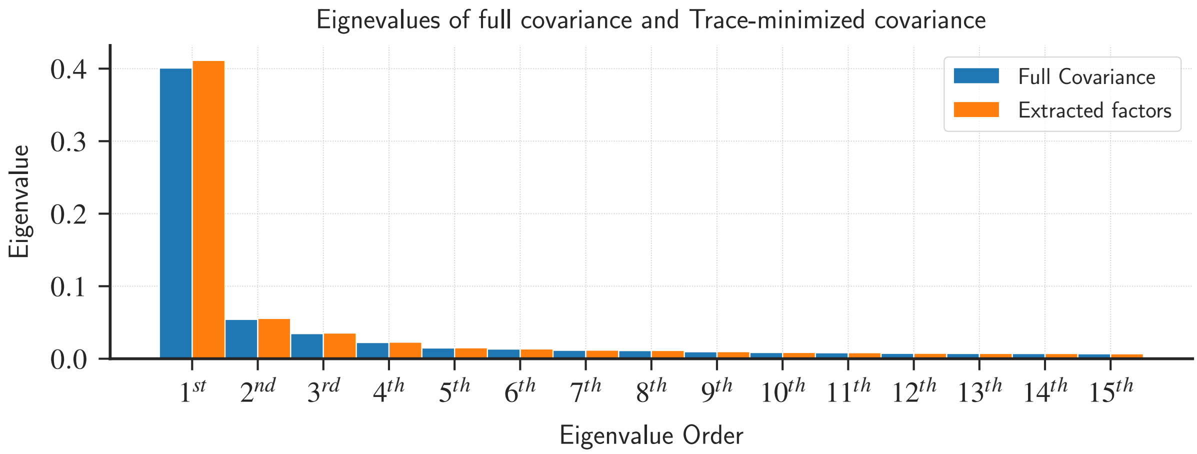

Figure 5.7 – The eigenvalues of a full covariance matrix and the eigenvalues of the trace-minimized covariance matrix

Bibliography

Bibliography

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.

ANTONAKAKIS, N. et al. Is inflation persistence different in reality? **Economics Letters**, v. 148, p. 55–58, 2016. ISSN 0165-1765.

BATISTA, R. L.; LAURINI, M. Bayesian estimation of term structure models: An application of the Hamiltonian Monte Carlo method. **Communications in Statistics: Case Studies, Data Analysis and Applications**, Taylor & Francis, v. 2, n. 3-4, p. 79–91, 2016.

BEKKER, P. A. Alternative approximations to the distributions of instrumental variable estimators. **Econometrica: Journal of the Econometric Society**, JSTOR, v. 62, n. 30, p. 657–681, 1994.

BENZONI, L.; CHYRUK, O.; KELLEY, D. Why does the yield-curve slope predict recessions? **Available at SSRN 3271363**, 2018.

BETANCOURT, M. A conceptual introduction to hamiltonian monte carlo. **arXiv preprint arXiv:1701.02434**, 2017.

BILICI, B.; ÇEKIN, S. E. Inflation persistence in Turkey: A TVP-estimation approach. **The Quarterly Review of Economics and Finance**, Elsevier, v. 78, p. 64–69, 2020.

BLEI, D. M.; NG, A.; JORDAN, M. Latent dirichlet allocation. **Journal of Machine Learning Research**, v. 3, p. 993–1022, 2003.

BOUKUS, E.; ROSENBERG, J. V. The information content of FOMC minutes. **Available at SSRN 922312**, 2006.

CALDEIRA, J.; TORRENT, H. Forecasting the us term structure of interest rates using nonparametric functional data analysis. **Journal of Forecasting**, Wiley Online Library, v. 36, n. 1, p. 56–73, 2017.

CAMPBELL, J. Y. et al. **The Econometrics of Financial Markets**. [S.l.]: Princeton University Press, 1997.

CANNON, S. Sentiment of the FOMC: Unscripted. **Economic Review-Federal Reserve Bank of Kansas City**, Citeseer, v. 5, 2015.

CARVALHO, C. M.; POLSON, N. G.; SCOTT, J. G. The horseshoe estimator for sparse signals. **Biometrika**, Oxford University Press, v. 97, n. 2, p. 465–480, 2010.

CHAGUE, F. et al. Central bank communication affects the term-structure of interest rates. **Revista Brasileira de Economia**, SciELO Brasil, v. 69, n. 2, p. 147–162, 2015.

CHAIM, P.; LAURINI, M. P. Volatility and return jumps in bitcoin. **Economics Letters**, Elsevier, v. 173, p. 158–163, 2018.

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The american statistician**, Taylor & Francis, v. 49, n. 4, p. 327–335, 1995.

CHUN, A. L. Expectations, bond yields, and monetary policy. **The Review of Financial Studies**, Society for Financial Studies, v. 24, n. 1, p. 208–247, 2011.

CLEMENTS, M. P. Individual forecaster perceptions of the persistence of shocks to GDP. **Journal of Applied Econometrics**, Wiley Online Library, forthcoming, 2021.

COGLEY, T.; SARGENT, T. J. Evolving post-world War II US inflation dynamics. **NBER Macroeconomics Annual**, MIT press, v. 16, p. 331–373, 2001.

_____. Drifts and volatilities: monetary policies and outcomes in the post WWII US. **Review of Economic Dynamics**, Elsevier, v. 8, n. 2, p. 262–302, 2005.

DAVIDSON, R.; MACKINNON, J. G. et al. **Estimation and Inference in Econometrics**. [S.l.]: Oxford: New York, 1993.

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.

DIEBOLD, F. X.; RUDEBUSCH, G. D.; ARUOBA, S. B. The macroeconomy and the yield curve: A dynamic latent factor approach. **Journal of Econometrics**, Elsevier, v. 131, n. 1-2, p. 309–338, 2006.

_____. The macroeconomy and the yield curve: A dynamic latent factor approach. **Journal of Econometrics**, Elsevier, v. 131, n. 1-2, p. 309–338, 2006.

DOSSCHE, M.; EVERAERT, G. Measuring inflation persistence: a structural time series approach. **National Bank of Belgium Working Paper**, n. 70, 2005.

DUANE, S. et al. Hybrid monte carlo. **Physics letters B**, Elsevier, v. 195, n. 2, p. 216–222, 1987.

ERP, S. V.; OBERSKI, D. L.; MULDER, J. Shrinkage priors for Bayesian penalized regression. **Journal of Mathematical Psychology**, Elsevier, v. 89, p. 31–50, 2019.

ESTRELLA, A.; MISHKIN, F. S. Predicting us recessions: Financial variables as leading indicators. **Review of Economics and Statistics**, MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046, USA journals . . . , v. 80, n. 1, p. 45–61, 1998.

FAMA, E. F. Term premiums and default premiums in money markets. **Journal of Financial Economics**, Elsevier, v. 17, n. 1, p. 175–196, 1986.

FAMA, E. F.; FRENCH, K. R. The cross-section of expected stock returns. **the Journal of Finance**, Wiley Online Library, v. 47, n. 2, p. 427–465, 1992.

_____. Common risk factors in the returns on stocks and bonds. **Journal of Financial Economics**, Elsevier, v. 33, n. 1, p. 3–56, 1993.

FAMA, E. F.; MACBETH, J. D. Risk, return, and equilibrium: Empirical tests. **Journal of Political Economy**, University of Chicago Press, v. 81, n. 3, p. 607–636, 1973. ISSN 00223808, 1537534X.

FUHRER, J. C. Inflation persistence. In: **Handbook of Monetary Economics**. [S.l.]: Elsevier, 2010. v. 3, p. 423–486.

FULLER, W. A. Some properties of a modification of the limited information estimator. **Econometrica**, JSTOR, p. 939–953, 1977.

GARDNER, B.; SCOTTI, C.; VEGA, C. Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements. **Journal of Econometrics**, Elsevier, v. 231, n. 2, p. 387–409, 2022.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, JSTOR, v. 7, n. 4, p. 457–472, 1992.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. **Bayesian statistics**, Clarendon Press, v. 4, p. 641–649, 1992.

GIBBONS, M. R.; ROSS, S. A.; SHANKEN, J. A test of the efficiency of a given portfolio. **Econometrica**, [Wiley, Econometric Society], v. 57, n. 5, p. 1121–1152, 1989. ISSN 00129682, 14680262.

GIGLIO, S.; XIU, D. Asset pricing with omitted factors. **Journal of Political Economy**, v. 129, n. 7, p. 1947–1990, 2021.

HAHN, P. R.; HE, J.; LOPES, H. Bayesian factor model shrinkage for linear IV regression with many instruments. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 36, n. 2, p. 278–287, 2018.

HAN, Y.; JIAO, A.; MA, J. The predictive power of Nelson–Siegel factor loadings for the real economy. **Journal of Empirical Finance**, Elsevier, v. 64, p. 95–127, 2021.

HANSEN, C.; HAUSMAN, J.; NEWEY, W. Estimation with many instrumental variables. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 26, n. 4, p. 398–422, 2008. Disponível em: <https://doi.org/10.1198/073500108000000024>.

HANSEN, P. R.; LUNDE, A.; NASON, J. M. The model confidence set. **Econometrica**, Wiley Online Library, v. 79, n. 2, p. 453–497, 2011.

HANSEN, S.; MCMAHON, M.; TONG, M. The long-run information effect of central bank communication. **Journal of Monetary Economics**, Elsevier, v. 108, p. 185–202, 2019.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Oxford University Press, 1970.

HAUTSCH, N.; YANG, F. Bayesian inference in a stochastic volatility Nelson–Siegel model. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 11, p. 3774–3792, 2012.

HITCHCOCK, D. B. A history of the metropolis–hastings algorithm. **The American Statistician**, Taylor & Francis, v. 57, n. 4, p. 254–257, 2003.

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. 1, p. 1593–1623, 2014.

HÖRDAHL, P.; TRISTANI, O.; VESTIN, D. A joint econometric model of macroeconomic and term-structure dynamics. **Journal of Econometrics**, Elsevier, v. 131, n. 1-2, p. 405–444, 2006.

HOTZ-BEHOFSITS, C.; HUBER, F.; ZÖRNER, T. O. Predicting crypto-currencies using sparse non-gaussian state space models. **Journal of Forecasting**, Wiley Online Library, v. 37, n. 6, p. 627–640, 2018.

JACQUIER, E.; POLSON, N. G.; ROSSI, P. E. Bayesian analysis of stochastic volatility models. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 20, n. 1, p. 69–87, 2002.

JAIN, M. Perceived inflation persistence. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 37, n. 1, p. 110–120, 2019.

JEGADEESH, N. et al. Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. **Journal of Financial Economics**, v. 133, n. 2, p. 273 – 298, 2019. ISSN 0304-405X.

JENSEN, M. C.; BLACK, F.; SCHOLES, M. S. The capital asset pricing model: Some empirical tests. **Studies in the Theory of Capital Markets**, Praeger Publishers Inc, 1972. Disponível em: <https://ssrn.com/abstract=908569>.

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.

KOWAL, D. R.; MATTESON, D. S.; RUPPERT, D. Dynamic shrinkage processes. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 81, n. 4, p. 781–804, 2019.

KOZAK, S.; NAGEL, S.; SANTOSH, S. Shrinking the cross-section. **Journal of Financial Economics**, Elsevier, v. 135, n. 2, p. 271–292, 2020.

KOZICKI, S.; TINSLEY, P. A. Permanent and transitory policy shocks in an empirical macro model with asymmetric information. **Journal of Economic Dynamics and Control**, Elsevier, v. 29, n. 11, p. 1985–2015, 2005.

KRANE, S. D. Professional forecasters' view of permanent and transitory shocks to GDP. **American Economic Journal: Macroeconomics**, v. 3, n. 1, p. 184–211, 2011.

LAMLA, M. J.; LEIN, S. M. What matters when? The impact of ECB communication on financial market expectations. **Applied Economics**, Taylor & Francis, v. 43, n. 28, p. 4289–4309, 2011.

LAURINI, M.; MAUAD, R.; AIUBE, F. Multivariate stochastic volatility-double jump model: An application for oil assets [working paper n. 415]. **Banco Central do Brasil**, 2016.

LAURINI, M.; VIEIRA, H. A dynamic econometric model for inflationary inertia in Brazil. **Journal of Statistical and Econometric Methods**, v. 2, n. 2, p. 51–83, 2013.

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.

LEOMBRONI, M. et al. Central bank communication and the yield curve. **Journal of Financial Economics**, Elsevier, v. 141, n. 3, p. 860–880, 2021.

LETTAU, M.; LUDVIGSON, S. Resurrecting the (c) CAPM: A cross-sectional test when risk premia are time-varying. **Journal of Political Economy**, The University of Chicago Press, v. 109, n. 6, p. 1238–1287, 2001.

LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. **Review of Economics and Statistics**, v. 47, n. 1, p. 13–37, 1965.

LITTERMAN, R.; SCHEINKMAN, J. Common factors affecting bond returns. **Journal of Fixed Income**, v. 1, n. 1, p. 54–61, 1991.

LOPES, H. F.; POLSON, N. G. Bayesian instrumental variables: Priors and likelihoods. **Econometric Reviews**, Taylor & Francis, v. 33, n. 1-4, p. 100–121, 2014.

LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. **The Journal of Finance**, Wiley Online Library, v. 66, n. 1, p. 35–65, 2011.

LUCCA, D. O.; TREBBI, F. **Measuring central bank communication: An automated approach with application to FOMC statements**. [S.l.], 2009.

LUENGO-PRADO, M. J.; RAO, N.; SHEREMIROV, V. Sectoral inflation and the Phillips curve: What has changed since the Great Recession? **Economics Letters**, Elsevier, v. 172, p. 63–68, 2018.

MALLOCH, H.; PHILIP, R.; SATCHELL, S. Estimation with Errors in Variables via the Characteristic Function. **Journal of Financial Econometrics**, n. 1-35, 2021. ISSN 1479-8409. Nbab011. Disponível em: <https://doi.org/10.1093/jjfinec/nbab011>.

MANKIW, N. G.; REIS, R. Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. **The Quarterly Journal of Economics**, MIT Press, v. 117, n. 4, p. 1295–1328, 2002.

MARKOVITZ, H. M. **Portfolio selection:Efficient diversification of investments**. [S.l.]: John Wiley, 1959.

MÁTÉ, Á.; SEBŐK, M.; BARCZIKAY, T. The effect of central bank communication on sovereign bond yields: The case of Hungary. **Plos One**, Public Library of Science San Francisco, CA USA, v. 16, n. 2, p. e0245515, 2021.

MEHROTRA, R. et al. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In: **SIGIR '13: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval**. [S.l.: s.n.], 2013. p. 889–892.

MENG, J. G.; HU, G.; BAI, J. OLIVE: a simple method for estimating betas when factors are measured with error. **Journal of Financial Research**, Wiley Online Library, v. 34, n. 1, p. 27–60, 2011.

METROPOLIS, N. et al. Equation of state calculations by fast computing machines. **The journal of chemical physics**, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953.

METROPOLIS, N.; ULAM, S. The monte carlo method. **Journal of the American statistical association**, Taylor & Francis, v. 44, n. 247, p. 335–341, 1949.

MURRAY, I.; ADAMS, R.; MACKAY, D. Elliptical slice sampling. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. **Proceedings of the thirteenth international conference on artificial intelligence and statistics**. [S.l.], 2010. p. 541–548.

NAGEL, S. Machine learning in asset pricing. In: **Machine Learning in Asset Pricing**. [S.l.]: Princeton University Press, 2021.

NAKAMURA, E.; STEINSSON, J. High-frequency identification of monetary non-neutrality: the information effect. **The Quarterly Journal of Economics**, Oxford University Press, v. 133, n. 3, p. 1283–1330, 2018.

NARD, G. D. Oops! i shrunk the sample covariance matrix again: Blockbuster meets shrinkage. **Journal of Financial Econometrics**, Oxford University Press, v. 20, n. 4, p. 569–611, 2022.

NEAL, R. M. et al. Mcmc using Hamiltonian dynamics. **Handbook of markov chain monte carlo**, v. 2, n. 11, p. 2, 2011.

NELSON, C. R.; PLOSSER, C. R. Trends and random walks in macroeconomic time series: Some evidence and implications. **Journal of Monetary Economics**, Elsevier, v. 10, n. 2, p. 139–162, 1982.

NELSON, C. R.; SIEGEL, A. F. Parsimonious modeling of yield curves. **Journal of Business**, JSTOR, p. 473–489, 1987.

NEWEY, W. K.; SMITH, R. J. Higher order properties of gmm and generalized empirical likelihood estimators. **Econometrica**, [Wiley, Econometric Society], v. 72, n. 1, p. 219–255, 2004. ISSN 00129682, 14680262.

NG, S.; BAI, J. Selecting instrumental variables in a data rich environment. **Journal of Time Series Econometrics**, v. 1, n. 1, p. 1–32, 2009. Disponível em: <https://doi.org/10.2202/1941-1928.1014>.

PERRON, P. The great crash, the oil price shock, and the unit root hypothesis. **Econometrica**, JSTOR, p. 1361–1401, 1989.

PRONO, T. Market proxies as factors in linear asset pricing models: Still living with the roll critique. **Journal of Empirical Finance**, v. 31, p. 36 – 53, 2015. ISSN 0927-5398.

QU, Z.; PERRON, P. A stochastic volatility model with random level shifts and its applications to S&P 500 and NASDAQ return indices. **The Econometrics Journal**, Oxford University Press Oxford, UK, v. 16, n. 3, p. 309–339, 2013.

QUAN, X. et al. Short and sparse text topic modeling via self-aggregation. In: **Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)**. [S.l.: s.n.], 2015. p. 2270–2278.

ROLL, R. A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. **Journal of Financial Economics**, v. 4, n. 2, p. 129 – 176, 1977. ISSN 0304-405X.

ROMER, C. D.; ROMER, D. H. Federal reserve information and the behavior of interest rates. **American Economic Review**, v. 90, n. 3, p. 429–457, 2000.

ROSA, C. Talking less and moving the market more: Evidence from the ECB and the FED. **Scottish Journal of Political Economy**, Wiley Online Library, v. 58, n. 1, p. 51–81, 2011.

RUDEBUSCH, G. D.; WILLIAMS, J. C. Forecasting recessions: the puzzle of the enduring power of the yield curve. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 27, n. 4, p. 492–503, 2009.

RUPPERT, D.; MATTESON, D. S. **Statistics and data analysis for financial engineering**. [S.l.]: Springer, 2011.

SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. **The journal of Finance**, Wiley Online Library, v. 19, n. 3, p. 425–442, 1964.

SIMMET, A.; POHLMEIER, W. The CAPM with measurement error: 'there's life in the old dog yet!'. **Jahrbücher für Nationalökonomie und Statistik**, De Gruyter Oldenbourg, Berlin, Boston, v. 240, n. 4, p. 417 – 453, 2020.

STAMBAUGH, R. F. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. **Journal of Financial Economics**, v. 10, n. 3, p. 237 – 268, 1982. ISSN 0304-405X.

STOCK, J. H.; WATSON, M. W. Why has US inflation become harder to forecast? **Journal of Money, Credit and Banking**, Wiley Online Library, v. 39, p. 3–33, 2007.

VEREDA, L.; LOPES, H.; FUKUDA, R. Estimating VAR models for the term structure of interest rates. **Insurance: Mathematics and Economics**, Elsevier, v. 42, n. 2, p. 548–559, 2008.

VICENTE, J.; TABAK, B. M. Forecasting bond yields in the Brazilian fixed income market. **International Journal of Forecasting**, Elsevier, v. 24, n. 3, p. 490–497, 2008.

VIEIRA, F.; FERNANDES, M.; CHAGUE, F. Forecasting the Brazilian yield curve using forward-looking variables. **International Journal of Forecasting**, Elsevier, v. 33, n. 1, p. 121–131, 2017.

WELCH, I.; GOYAL, A. A comprehensive look at the empirical performance of equity premium prediction. **The Review of Financial Studies**, Society for Financial Studies, v. 21, n. 4, p. 1455–1508, 2008.

WOODFORD, M. **Monetary policy in the information economy**. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 2001.