WORKING PAPER 3/2024 (ECONOMICS AND STATISTICS)

US Interest Rates: Are Relations Stable?

Sune Karlsson, Tamás Kiss, Hoang Nguyen and Pär Österholm

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

US Interest Rates: Are Relations Stable?

Sune Karlsson[(a)]*, Tamás Kiss[(a)], Hoang Nguyen[(b)] and Pär Österholm [(a,c)]

[(a)] School of Business, Örebro University

[(b)] Department of Management and Engineering, Linköping University

[(c)] National Institute of Economic Research

March 8, 2024

Abstract

In this paper, we assess whether key relations between US interest rates have been stable over time. This is done by estimating trivariate hybrid time-varying parameter Bayesian VAR models with stochastic volatility for the three-month Treasury bill rate, the slope of the Treasury yield curve and the corporate bond-yield spread. As a methodological contribution, we also allow for disturbances with heavy tails. We analyse monthly data from April 1953 to February 2023 both within- and out-of-sample. Our results indicate that the relations have not been stable; more specifically, there is evidence that the equation of the corporate bond-yield spread is subject to time variation in its parameters. We also find that an increase in the corporate bond-yield spread decreases the risk free rate. Finally, we note that while allowing for heavy tails receives a fair amount of support within sample, it appears to be of more limited importance from a forecasting perspective.

JEL Classification: C11, C32, C52, E44, E47, G17

Keywords: Bayesian inference; Stochastic volatility; Orthogonal Student's t distribution; Time-varying parameter VAR

*Corresponding author.School of Business, Örebro University. Address: Örebro University, Fakultetsgatan 1, 702 81, Örebro, Sweden. E-mail: sune.karlsson@oru.se

1 Introduction

During the decade following the Great Recession, policy interest rates in many countries reached historically low levels. However, several central banks – including the Federal Reserve, the Bank of England, the ECB, the Swiss National Bank and Sveriges Riksbank – judged that a very low policy interest rate did not make monetary policy sufficiently expansive and in order to further ease the stance of monetary policy, the central banks also employed unconventional monetary-policy tools.[1] Interest rates were widely affected. Effects in the United States have been established on, for example, Treasury yields, swap rates, yields on mortgage-backed securities and corporate bond yields; see, for example, Hancock and Passmore (2011), Gagnon et al. (2018), Ihrig et al. (2018).[2] Despite these monetary-policy measures, inflation was typically low during this period, leaving policy rates at low levels. As the corona crisis hit in 2020, there was accordingly little scope for central banks to conduct conventional monetary policy – that is, cutting policy rates – and further unconventional measures were a common response. However, in the recovery after the corona pandemic, inflation has surged in many countries and central banks have hiked policy rates at an unusually high pace.

In addition to these more recent developments, it is also the case that interest rates have varied substantially in a somewhat longer perspective. For example, interest rates (and inflation) were notably higher in the late 1970s and early 1980s than during the 1960s or the Great Moderation. Parts of this evolution can be explained by monetary policy being conducted differently over time; see, for example, Romer and Romer (1989), Sargent (1999), Clarida et al. (2000), Boivin (2006), Beechey and Österholm (2012) and Belongia and Ireland (2016). An important question that arises in light of this is whether the empirical relations between variables capturing various aspects of the interest-rate markets are stable or whether they should be modelled as time-varying.

In this paper, we assess the stability of the relations between three key interest-rate variables in the United States: the three-month Treasury bill yield, the slope of the Treasury yield curve and the corporate bond-yield spread. In line with Duffee (1998), our empirical analysis takes its starting point in a vector autoregressive (VAR) model. In order to study

[1] One important part of these unconventional means has been large-scale asset purchases, where focus typically was on government bonds (Cecioni et al., 2011; IMF, 2013). Other types of assets have also been purchased though. In the United States, the Federal Reserve has, for example, purchased mortgage backed securities. Bernanke (2020) and Bhattarai and Neely (2022) provide a recent summary and review of the literature on unconventional monetary policy and asset purchase programs.

[2] Further studies establishing effects on various markets include d'Amico et al. (2012), Rosa (2012) Wu (2014), Neely (2015), Inoue and Rossi (2019), Dedola et al. (2021), Rossi (2021) and Miranda-Agrippino and Nenova (2022).

the issue of stability, we cast the model in a recently developed econometric framework, namely the hybrid time-varying parameter Bayesian VAR with stochastic volatility of Chan and Eisenstat (2018b). However, we extend Chan and Eisenstat's methodology by allowing for disturbances with heavy tails. This is done in order to address the empirical finding that large shocks – often associated with increasing levels of risk and risk aversion, leading to abruptly changing term and credit spreads during crisis times such as the oil crises of the 1970s, the Great Recession and the covid crisis – appear to arrive more frequently than if shocks were drawn from a normal distribution.[3] Relying on this extended framework, we can conduct formal Bayesian model selection between VARs with time variation in none, one, two or three of the equations.[4]

Our paper makes two contributions to the literature. First, we assess the stability of key relations between US interest rates. In doing this, we provide further evidence on the interaction between corporate credit spreads and financial and macroeconomic variables. From a more macroeconomic perspective, our study is hence part of the research related to the fact that an increase in the corporate credit spread tends to be associated with a decrease in real economic activity and a lower short risk-free rate; for studies addressing the effects of the corporate credit spread on various variables – such as Treasury yields, inflation, GDP, employment, housing starts and industrial production – see, for example, Stock and Watson (1989), Friedman and Kuttner (1998), Mody and Taylor (2004), Gilchrist et al. (2009), Gilchrist and Zakrajšek (2012), Boivin et al. (2013) and Prieto et al. (2016). However, our paper is also related to a strand of research in financial economics whose emphasis lies on the effect that the risk-free rate should have on corporate bond spreads. A negative relationship between corporate bond-yield spreads and the risk free interest rate is typically found in this literature; see, for example, Nielsen and Ronn (1996), Duffee (1998), Collin-Dufresne et al. (2001), Campbell and Taksler (2003), Batten and Hogan (2003), Van Landschoot (2008), Giesecke et al. (2011), Österholm (2018) and Karlsson and Österholm (2019). Such a result is in line with structural theoretical pricing models of risky corporate debt (Merton, 1974; Longstaff and Schwartz, 1995; Collin-Dufresne and Goldstein, 2001) in which a negative relationship

[3] In line with this observation, there is a growing literature pointing to the potential importance of allowing for heavy tails when modelling macroeconomic time series; see, for example, Fagiolo et al. (2008), Ascari et al. (2015), Cross and Poon (2016), Chiu et al. (2017), Liu (2019), Kiss and Österholm (2020), and Brunnermeier et al. (2021).

[4] It can be noted that there is a large literature pointing to the benefits of using models that explicitly address the issue of heteroskedasticity in macroeconomics; see, for example, Fountas and Karanasos (2007), Clark (2011), Carriero et al. (2015), Chan (2017), Trypsteen (2017), Koop and Korobilis (2019), Clark et al. (2020) and Karlsson and Osterholm (2020, 2023). In line with this literature – and Chan and Eisenstat (2018*b*) – we take stochastic volatility as a given feature of our model. This modelling choice also has empirical support. In Table 4 in Appendix C, we provide marginal likelihoods which compare all models in Table 1 to the corresponding homoskedastic models. The support for stochastic volatility is very strong.

between the risk-free interest rate and the corporate bond yield spread stems from the fact that a higher risk-free interest rate is associated with higher expected cash-flow – hence decreasing future default probability. While there is little ambiguity regarding the sign of the relationship, its strength is more debated. [5] Our analysis adds to this literature by allowing for time variation in the relationship between the risk-free rate and corporate bond yields.

Second, by extending the hybrid time-varying parameter Bayesian VAR with stochastic volatility framework of Chan and Eisenstat (2018*b*), we provide a methodological setting that could be useful for researchers, analysts and policy makers in both macroeconomics and finance. Bayesian VARs with time-varying (drifting) parameters and stochastic volatility appear to be an increasingly popular tool in macroeconomic modelling; a few recent examples include Bianchi and Civelli (2015), Akram and Mumtaz (2019), Karlsson and Österholm (2020), Mumtaz and Petrova (2022) and Karlsson and Österholm (2023).[6] Chan and Eisenstat (2018*a*) made an important contribution to this field by providing tools for model comparison based on marginal likelihoods. In a related paper – which is the methodological foundation for our paper – they then also showed how inference can be drawn regarding time variation in the parameters of a subset of the equations of the VAR (Chan and Eisenstat, 2018*b*). Such analysis can be useful when analysing low-dimensional models; see, for example, Karlsson and Österholm (2020, 2023). In this paper, we extend Chan and Eisenstat (2018*b*)'s framework so that it also allows for disturbances with heavy tails in a similar vein to Karlsson et al. (2023). As pointed out above, this is a feature that has been shown to have relevance in macroeconomic and financial modelling, even if its importance seems to differ somewhat depending on the application; see, for example, Chan (2020) and Kiss et al. (2023). In summary, our framework allows for formal model selection based on marginal likelihoods for hybrid time-varying parameter Bayesian VAR with stochastic volatility and heavy tails, thereby providing a flexible and convenient tool for assessing various empirically relevant aspects of a VAR model.

We apply our modelling framework to US data ranging from April 1953 to February 2023. In addition to conducting within-sample analysis – where we both perform model selection and describe model dynamics – we also assess the models' performance out-of-sample; this is done by evaluating the forecasting performance with respect to both point and density

[5] In a seminal paper, Duffee (1998) finds that the callability of the corporate bonds is crucial, and the empirical relationship between the risk free rate and the yield of non-callable corporate bonds is much weaker; this result is also supported by Batten et al. (2005) who study the Australian bond market. Further empirical analysis suggests that the relationship is weaker once the dynamics of the interest rate series is taken into account (Neal et al., 2015) and for investment-grade corporate bonds (Dupoyet et al., 2023).

[6] Seminal contributions to this type of model were made by Cogley and Sargent (2005) and Primiceri (2005).

forecasts.

Briefly mentioning our results, we find that the relations between the variables in the model have not been stable. Both the within-sample evidence based on marginal likelihoods and our out-of-sample forecast evaluation suggest that improvements are to be had when parameters in the corporate bond-yield spread equation are allowed to vary. In addition, we find that an increase in the corporate bond-yield spread decreases the risk-free rate, and that an increase in the three-month Treasury bill rate tends to decrease the corporate bond-yield spread. Finally, we note that even though heavy-tailed innovations receive a fair amount of support within sample, it appears to be of more limited importance from a forecasting perspective.

The rest of this paper is organised as follows: In Section 2, we describe our methodological framework. Data are presented in Section 3. In Section 4, we present the results from our empirical analysis. Finally, Section 5 concludes.

2 Econometric framework

In this section, we first present the original framework of time-varying parameter VAR with stochastic volatility (TVP-VAR-SV) developed by Primiceri (2005) and Cogley and Sargent (2005). We then show how it can be extended within the hybrid TVP-VAR-SV framework of Chan and Eisenstat (2018*b*) – in which some equations in the VAR can have constant parameters while other equations have time-varying parameters – to allow for disturbances with heavy tails. Finally, we show how Bayesian inference can be conducted in this extended setting.

2.1 A TVP-VAR model with Gaussian SV innovations

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})'$ be a k-dimensional vector of endogenous variables. A time-varying parameter VAR model with Gaussian stochastic volatility innovations is then given by

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{b}_t + \mathbf{B}_{1t}\mathbf{y}_{t-1} + \ldots + \mathbf{B}_{pt}\mathbf{y}_{t-p} + \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{b}_t$ is a k-dimensional vector of time-varying intercept and $\mathbf{B}_{jt}$ is a $k \times k$ dimensional matrix of regression coefficients for $j = 1, \ldots, p$; $\mathbf{A}_t$ is a $k \times k$ lower triangular matrix with ones on the diagonal that describes the time-varying contemporaneous interaction of the endogenous variables; $\mathbf{H}_t = \text{diag}(\exp(h_{1,t}), \ldots, \exp(h_{k,t}))$ is a $k \times k$ diagonal matrix and

$\boldsymbol{\epsilon}_t$ is a k-dimensional vector of standardized Gaussian error terms. We assume that the log-volatilities and the time-varying parameters follow random walk processes,

$$\begin{aligned} \mathbf{h}_t &= \mathbf{h}_{t-1} + \boldsymbol{\zeta}_t, \\ \boldsymbol{\theta}_{i,t} &= \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\eta}_{i,t}, \end{aligned} \tag{2}$$

where $\mathbf{h}_t = (h_{1t}, \ldots, h_{kt})'$, $\boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_h)$ for $\boldsymbol{\Sigma}_h = \text{diag}(\sigma^2_{h,1}, \ldots, \sigma^2_{h,k})$; $\boldsymbol{\theta}_{it}$ stacks the vector of the time-varying regression coefficients in the ith equation and the free element in the ith row of the lower triangular matrix $\mathbf{A}_t$ and $\boldsymbol{\eta}_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\theta,i})$ for $\boldsymbol{\Sigma}_{\theta,i} = \text{diag}(\boldsymbol{\sigma}^2_{\theta,i}) = \text{diag}(\sigma^2_{\theta,i,1}, \ldots, \sigma^2_{\theta,i,k_i})$ where k_i is the number of time-varying parameters in equation i.

2.2 A hybrid TVP-VAR model with orthogonal Student's t SV innovations

In order to account for heavy tails, we extend Equation 1 by including a mixing variable, $\mathbf{W}_t$, in the innovations such that the model is given as

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{b}_t + \mathbf{B}_{1t}\mathbf{y}_{t-1} + \ldots + \mathbf{B}_{pt}\mathbf{y}_{t-p} + \mathbf{e}_t, \quad t = 1, \ldots, T, \tag{3}$$

where $\mathbf{e}_t = \mathbf{W}_t^{1/2}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$ and $\mathbf{W}_t = \text{diag}(w_{1,t}, \ldots, w_{k,t})$ with $w_{i,t} \sim \mathcal{IG}(\frac{\nu_i}{2}, \frac{\nu_i}{2})$; $\mathbf{H}_t$ and $\boldsymbol{\epsilon}_t$ are defined as above. The structural form error terms $\mathbf{e}_t$ are thus t-distributed with ν_i degrees of freedom and the marginal distribution of $y_{i,t}$ is a linear combination of orthogonal Student's t (OT) distributions; see the discussion in Karlsson et al. (2023).

The hybrid TVP-VAR of Chan and Eisenstat (2018*b*) works by leveraging the triangular structure of $\mathbf{A}_t$. Let $\mathbf{x}_{i,t} = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p}, -y_{1,t}, \ldots, -y_{i-1,t})$. We can then write equation i of the VAR in equation (3) as

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\theta}_{i,t} + \sqrt{w_{i,t}}\exp(h_{i,t}/2)\epsilon_{i,t}. \tag{4}$$

The recursive structure together with the independent evolution of $\boldsymbol{\theta}_{i,t}$ makes equation by equation estimation straightforward if there also is prior independence between the equations. Time-varying parameters is then turned on or off for an equation by allowing the diagonal elements of $\boldsymbol{\Sigma}_{\theta,i}$ to be positive or restricting them to zero.

In general, the hybrid TVP-VAR model with orthogonal Student's t stochastic volatility (OT-SV) innovations provides a flexible framework for modelling the joint evolution of endogenous variables. The magnitude and the frequency of the shocks are driven by the heavy

tails and stochastic volatility components, while the interaction among endogenous variables is captured through the time-varing parameter setting. The hybrid TVP-VAR model with OT innovations reduces to the Gaussian case when the degrees of freedom $\nu \to \infty$.

2.3 Inference and model selection

We assume vague but proper prior distributions for the parameters of the TVP-VAR model. For example, $\boldsymbol{\theta}_{i,0}$ follows a Minnesota prior with the overall shrinkage $l_1 = 0.2$ and the cross-variable shrinkage $l_2 = 0.5$; see Koop and Korobilis (2010). The prior for the degrees of freedom is a Gamma distribution. Following Kastner and Frühwirth-Schnatter (2014), the prior for the variance of the of shocks to the volatility and time-varying parameters are Gamma distributions; this should be less influential compared to the conjugate inverse Gamma prior when true variance values are small. Details on the prior distributions and and overview of the MCMC scheme used for posterior inference is provided in Appendix A.

We compare the hybrid TVP-VAR models with different assumptions of innovations using the marginal likelihood computed via the multi-stage importance sampling as in Chan and Eisenstat (2018*b*). In order to select a good proposal distribution to approximate the high-dimensional integration of the marginal likelihood, we divide the model parameters into fix parameters and TVP latent variables. First, we use the cross-entropy method (Chan and Eisenstat, 2018*a*) to learn the proposal distribution of the fix parameters. Then, given a sample of the fix parameters, we calculate the integrated likelihood of the TVP equation using another level of importance sampling; see details in Appendix B. When comparing models, the model with the highest marginal likelihood is preferred by the data.

3 Data

For our empirical analysis, we use monthly interest-rate data from the United States. The sample ranges from April 1953 to February 2023. The three interest-rate variables we use are *i)* the three-month Treasury bill rate, *ii)* the difference between the yield on the ten-year Treasury bond and the three-month Treasury bill rate, and *iii)* the corporate yield spread. The corporate yield spread is defined as the yield on BAA corporate bonds minus the yield on a ten-year Treasury bond. These variables are henceforth referred to as the short rate, the slope and the BAA spread, respectively.

Figure 1 shows the time series of the three variables. The shaded areas mark NBER recession periods. These periods are usually associated with a rise in the BAA spread,

signalling tighter financial conditions in association with economic downturns.

A visual inspection of the short rate in the first panel supports the claim that our sample period contains some different interest-rate environments. The sample starts with moderate interest rates in the 1950s and 1960s, which is followed by high nominal rates (and high inflation) around the oil crisis and the collapse of the gold standard. Since the early 1980s – following the disinflationary efforts of the Federal Reserve ("the Volcker disinflation") – interest rates (and inflation) have largely been on a declining trend. However, the era of falling and low interest rates – which included the Great Moderation and the period after the Great Recession – came to a halt in the recovery after the corona pandemic. Relatively strong aggregate demand – in parts induced by fiscal policy – combined with Russia's attack war on Ukraine and other supply distortions, made inflation surge and interest rates rise. We next assess whether these changes to the interest-rate environment are associated with changing empirical relationships over time in our chosen reduced-form setting.

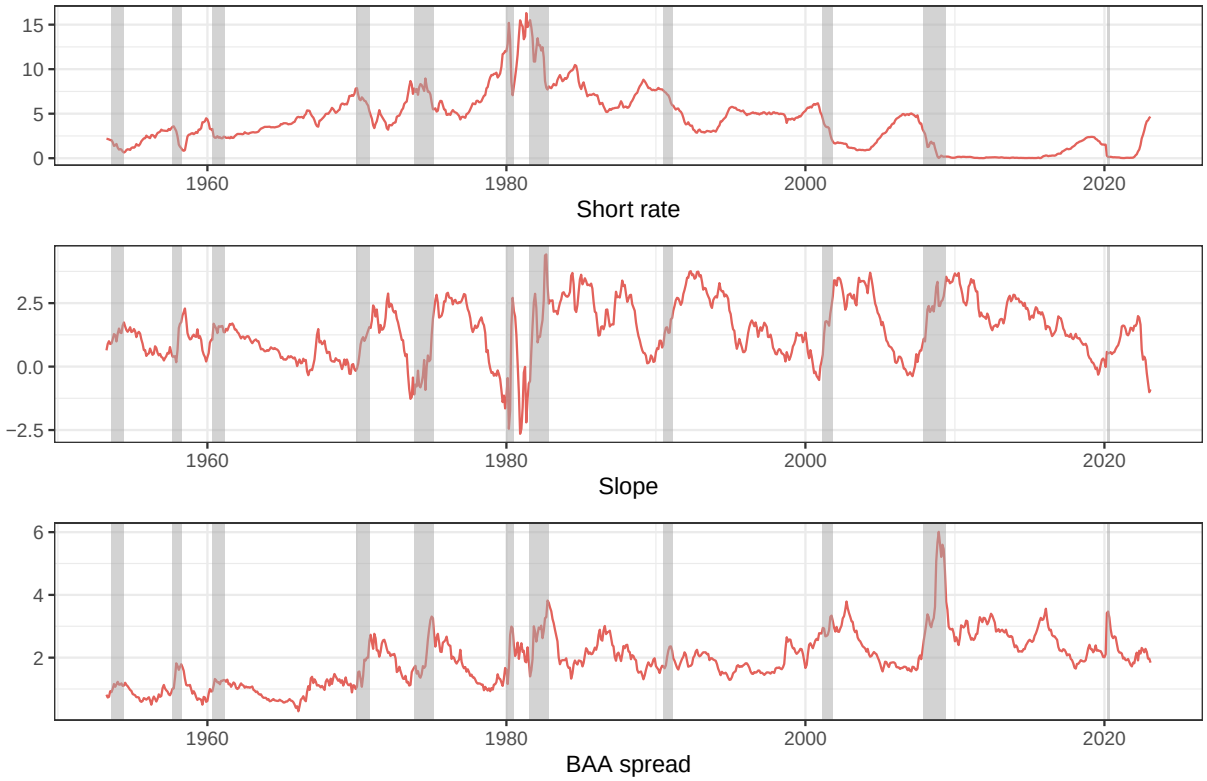

Figure 1: Data

Percent on the vertical axis for the three-month Treasury bill rate. Percentage points on the vertical axis for the slope and the BAA spread.

4 Empirical analysis

In accordance with Duffee (1998), we estimate trivariate VAR models. We define the vector of dependent variables in equation 1 as $\mathbf{y}_t = (i_t, s_t, c_t)'$, where i_t is the short rate, s_t is the slope and c_t is the BAA spread.[7] Lag length in all models is set to $p = 3$.

The within-sample analysis focuses on whether time variation in parameters – in combination with heavy-tailed innovation distributions – can improve the model, while the out-of-sample analysis investigates how these features impact forecasting performance. Hence, our analysis sheds light on whether relations have changed over our sample period, and whether these changes impact the way we should conduct modelling and forecasting.

4.1 Within-sample results

We first analyse the importance of time-varying parameters and heavy tails in our proposed empirical setup. We do that by estimating VAR models with different specifications for time variation in the coefficients and different innovation distributions. Since we deal with a low dimensional model, we consider all permutations – that is, all possibilities with zero, one, two or three equations with time-varying coefficients. Moreover, we estimate all models with Gaussian as well as heavy-tailed (orthogonal Student's t) innovation distributions. Log marginal likelihoods are presented in Table 1, together with the estimated degrees of freedom in the models allowing for heavy tails.

If we first compare the models with Gaussian innovations, it can be seen that the model with time variation only in the equation for the BAA spread has the highest log marginal likelihood, and hence it is the preferred specification. It is also interesting to note that the second best model is the specification with constant parameters; this suggests that drifting parameters is an important feature in the equation for the BAA spread, but not in the equations for the two other variables. This interpretation is further supported by the rest of the log marginal likelihoods. As can be seen from the table, the model which has time variation in the equation for the BAA spread and the equation for the slope is ranked third, and the model which has time variation in the equation for the BAA spread and the equation for the short rate is ranked fourth. The model ranked last – by a wide margin – is that with time variation in the equations for the short rate and the slope, but not in the equation for BAA spread.

[7] A similar trivariate specification was also used by Österholm (2018) and Karlsson and Österholm (2019) on Australian data.

Table 1: Log marginal likelihood for VAR models and estimated degrees of freedom in t-distribution (where applicable)

Model	Gaussian	Orthogonal Student's t	
	Log marginal likelihood	Log marginal likelihood	Degrees of freedom
000	1125.39	1133.78	(13.89; 24.48; 12.50)
100	1077.77	1085.11	(10.38; 25.62; 13.40)
010	1079.97	1087.25	(13.48; 25.19; 12.85)
001	1158.11	1165.70	(14.10; 27.05; 8.80)
110	1030.99	1038.52	(10.05; 27.26; 13.13)
101	1108.88	1115.44	(11.84; 26.13; 9.49)
011	1111.56	1119.50	(13.05; 25.73; 8.31)
111	1061.56	1067.67	(12.46; 26.13; 9.40)

The cross entropy methods of Chan and Eisenstat (2018*b*) are used to calculate the log marginal likelihoods. The model number indicates the combination of constant and time-varying parameters, for example 001 indicates that only the third equation in the TVP VAR model has time-varying parameters. The order of the variables are 1) the short rate, 2) the slope and 3) the BAA spread. We first sample 120,000 total draws from the conditional posterior distributions with 20,000 discarded as burn-in. Then, all log marginal likelihoods are estimated using 20,000 draws from the proposal distributions; for details, see Appendix B. Estimated degrees of freedom is given by the mean from the posterior distribution.

Turning to t-distributed innovations, heavy tails improve the model. Comparing models pairwise to the case with Gaussian innovations, the difference is substantial in all eight cases; using the scale of two times the difference in log marginal likelihood and the terminology of Kass and Raftery (1995), the evidence in favour of heavy tails is in all cases "very strong".[8] As can be seen from the rightmost column in Table 1, the innovations of the equations of the short rate and the BAA spread are associated with heavier tails than the innovations of the equation for the slope. The estimated degrees of freedom for the former two range between approximately eight and 14; for the slope, the corresponding range is 24 to 27.

The ranking of the models with t-distributed innovations is identical to that we found with Gaussian errors. Accordingly, the model with t-distributed innovations and time variation only in the equation for the BAA spread has the highest log marginal likelihood of all models. It can be noted that the evidence in favour of time variation in the equation for the BAA spread (but no other equation) also is "very strong"; comparing it to the best model with a different specification concerning the dynamics – that is, the model with constant parameters in all equations and t-distributed innovations – we find that two times the difference in log marginal likelihood is 63.84.

Some insights as to why the model with time variation in the equation for the BAA

[8]The cutoff point for "very strong" – which is the highest category – is 10.

spread was selected can be provided through a visual inspection of the estimated parameters of the model in which we allow for time variation in all the equations of the VAR (that is, the model denoted 111 in Table 1). We use the version of the fully time-varying model which allows for fat-tailed innovation distribution. The evolution of the parameters – given by the posterior mean and the 80 percent credible band – is shown in Figures 3 to 5 in Appendix C. For the short rate (in Figure 3), the coefficients – with the possible exception of the first lag of the slope – move very modestly over time. Turning to slope (in Figure 4), we again see a limited amount of movement over time; in this case the coefficients on the first own lag and the contemporaneous short rate stand out as showing more movement. However, looking at the coefficients for the BAA spread in Figure 5, we see quite a few coefficients that show what appears to be a reasonably large amount of movement over time. For this model there are substantial movements in the coefficients for the first lag of all variables, for all the lagged values of the BAA spread itself, as well as the coefficients governing contemporaneous relations.

It was established above that the model with t-distributed innovations and time variation in the equation for the BAA spread was the preferred one. We next look further into the dynamics of this model. Seeing that we are working with VAR models, our focus will be on the impulse-response functions. These are given in Figure 2.

It should first be noted that the size of the impulses is one standard deviation. Since we have a model with stochastic volatility, the standard deviation is time-varying. For each variable in the model, this is given by the horizon-zero response of the variable to a shock to itself; for clarity, it is also provided in Figure 6 in Appendix C. As can be seen, there is non-negligible variation in shock volatility over time for all three variables. Substantial peaks for the short rate (first row, first column) and slope (second row, second column) can be found in the early 1980s when there were large movements in the short rate – movements which partly were due to the policy decisions associated with the Volcker disinflation. For the BAA spread (third row, third column), we see the most pronounced peaks around the Great Recession and the corona crisis.

Turning to the effect that a shock to the short rate has, it can be seen that it – not surprisingly – compresses the slope (second row, first column). It can also be noted that the effect of the shock dies out quite slowly; this feature applies to several of the impulse-response functions of the model and reflects the fact that the variables of the system have a fairly high persistence – in particular the short rate. In line with the majority of the literature studying the effect of the risk-free interest rate on the corporate bond spread, we also find that this shock tends to decrease the BAA spread (third row, first column). While the point

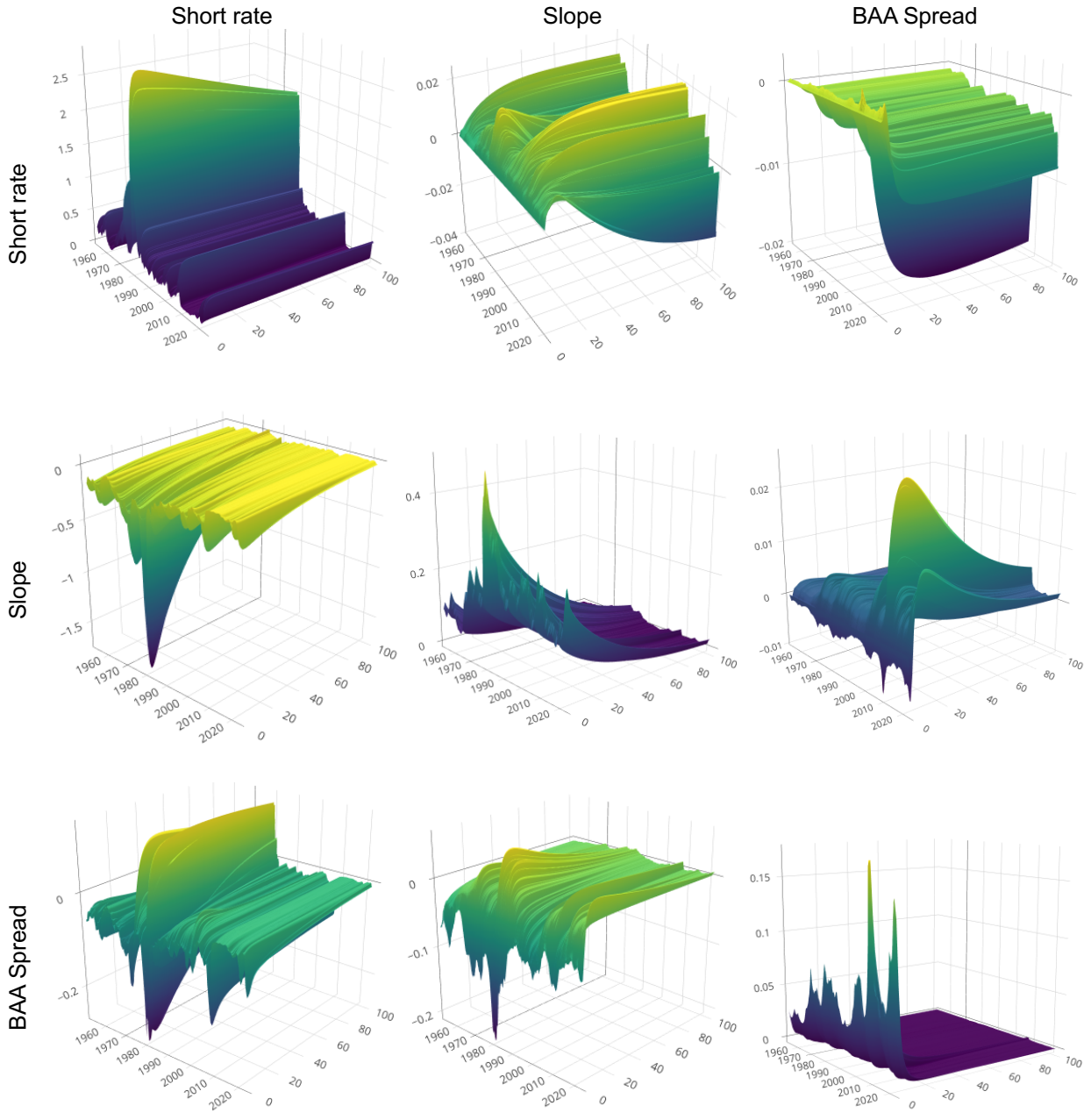

Figure 2: The impulse response functions of the model with time varying parameters in the equation for the BAA spread and innovations with heavy tails.

Impulses are shown in the column headers, response variables are given by the rows. The size of the impulse is one standard deviation. The size of the effect in percentage points is represented on the vertical axis. The two horizontal axes represent the date of the shock (left horizontal axis) and the horizon in months (right horizontal axis), respectively.

estimates indicate that the dynamic effect that this shock has on the BAA spread differs a fair bit over the sample, it is only the initial decline in the BAA spread that is a robust finding in terms of being significantly different from zero; this is illustrated in Figures 7 and 8 which give the impulse-response functions of the model at December 2008 (that is, during the Great Recession) and February 2023 (that is, at the end of our sample). The significance of the initial decline – but not of effects at longer horizons – also applies to the period around the Volcker disinflation where there are some fairly large positive point estimates at longer horizons.

The point estimate of a shock to the slope is generally associated with an increase in the short rate (first row, second column), even if a decrease is sometimes found at longer horizons. However, the effect is small and not significant; see Figures 7 and 8 for illustrations. A shock to the slope also tends to decrease the BAA spread – at least at short horizons (third row, second column); this is in line with such shocks indicating a stronger business cycle.

Finally, a shock to the BAA spread decreases the short rate (first row, third column); see also Figures 7 and 8. This is an effect consistent with what studies relating the corporate bond-yield spread to the macroeconomy have previously found. In line with this, we also find an increase in the slope (second row, third column). It is worth noting that the large differences in the effect on these variables that can be seen over time largely is due to the size of the shock, where the volatility of the shock peaked around the Great Recession (and is at similar levels at the end of the sample).

4.2 Forecasting performance

The within-sample evidence presented above favours both time-varying parameters in the BAA equation and innovations with heavy tails. We next conduct an out-of-sample forecast exercise to see if these features also help improve forecasting performance.

We consider both point and density forecasts, and analyse six forecast horizons: one month, two months, three months, six months, one year and two years. Our benchmark model is the constant parameter trivariate VAR model with Gaussian innovations. This is compared to both the constant parameter trivariate VAR model with heavy-tailed innovations and three models with time-varying parameters (discussed in Section 4.1) under both Gaussian and heavy-tailed innovation distributions. All models are estimated allowing for stochastic volatility, seeing the overwhelming evidence for that in macroeconomic data (see footnote 4). We established above that the evidence for time-varying parameters is strongest for the equation of the BAA spread and weakest for the equation of the short rate. We ac-

cordingly first allow the equation of the BAA spread to have time-varying parameters. Then, we also allow the slope equation to have time-varying dynamics, and, lastly, we consider the fully time-varying model.

The forecast evaluation is performed on an expanding sample; the parameters are re-estimated each time the sample is expanded. We use the mean squared forecast error (MSFE) to evaluate point forecasts. At forecast horizon h, the MSFE is calculated as

$$\text{MSFE}_{i,h} = \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \left[\left(\bar{y}_{i,t+h|t} - y^o_{i,t+h} \right)^2 \right],$$

where $\bar{y}_{i,t+h|t}$ is the point forecast – the posterior mean – using all the information up to time t and $y^o_{i,t+h}$ is the realization of the variable i, h-steps ahead. A lower MSFE indicates better model performance.

For density forecasts, we use the log predictive score (LPS). The LPS of the posterior predictive distribution is evaluated as

$$\begin{aligned}
\text{LPS}_{i,h} &= \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[\log \int p(y^o_{i,t+h}|\mathbf{\Theta}, \mathbf{y}_{1:t}) p(\mathbf{\Theta}|\mathbf{y}_{1:t}) d\mathbf{\Theta} \right] \\
&= \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[\log \int \int p(y^o_{i,t+h}|\mathbf{\Theta}, \mathbf{y}_{1:t}, \mathbf{y}_{t+}) p(\mathbf{y}_{t+}|\mathbf{\Theta}, \mathbf{y}_{1:t}) p(\mathbf{\Theta}|\mathbf{y}_{1:t}) d\mathbf{y}_{t+} d\mathbf{\Theta} \right] \\
&\approx \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \left[\log \frac{1}{R} \sum_{r=1}^{R} p(y^o_{i,t+h}|\mathbf{\Theta}^{(r)}, \mathbf{y}_{1:t}, \mathbf{y}^{(r)}_{t+}) \right],
\end{aligned}$$

with $\mathbf{\Theta}^{(1)}, \ldots, \mathbf{\Theta}^{(R)}$ being the posterior samples of the corresponding VAR model, $\mathbf{y}^{(r)}_{t+}$ represents a simulated path of intermediate observations from $t+1$ to $t+h-1$ using $\mathbf{\Theta}^{(r)}$ and $p(y^o_{i,t+h}|\mathbf{\Theta}^{(r)}, \mathbf{y}_{1:t}, \mathbf{y}^{(r)}_{t+})$ is the 1-step ahead posterior predictive density function (the score) evaluated at the realization of the variable, conditional on the posterior sample of parameters $\mathbf{\Theta}^{(r)}$ and the simulated path. The LPS is defined such that a larger value indicates a better forecast performance of the model.

To compare both point and density forecasting performance, we apply the one-sided test of Diebold and Mariano (1995), where the standard errors are calculated using the Newey-West estimator (Newey and West, 1987). This is in line with previous work on density forecast comparisons, for example, Clark (2011) and Clark and Ravazzolo (2015). The first estimation sample ranges from April 1953 to December 1999, and the last forecast is made based on the sample April 1953 to February 2021. This allows us to evaluate 255 forecasts

for all horizons.

4.2.1 Point forecasts

We initially focus on the point forecasting results. Table 2 presents MSFEs for the short rate, the slope and the BAA spread, respectively. Absolute numbers are shown for the benchmark model with constant parameters and Gaussian innovations, and relative MSFEs are given for the other specifications. The relative MSFE is defined as the MSFE for the model of interest divided by the MSFE of the benchmark model. Hence, a relative MSFE smaller than one indicates that the model of interest is better than the benchmark. Models that outperform the benchmark using the Diebold and Mariano (1995) test are marked with asterisks.

Looking at the MSFE results for the short rate we can note that time variation in the parameters typically does not materially affect point forecasting results. Relative MSFEs of the Gaussian models with time-varying parameters tend to vary closely around one across horizons and specifications; in a couple of cases, MSFEs are around ten percent lower than that of the benchmark model but no significant improvements are found. In contrast, allowing for heavy-tailed innovations helps in the model with constant parameters, where improvements in MSFE are significant for all horizons; the improvements are, however, quantitatively negligible and not economically significant. Time variation together with heavy tails does not improve forecasts much either; relative MFSEs are generally close to one and the two cases that are statistically significant are not economically significant.

The results for the slope, presented in panel (b) of Table 2, are also meagre in terms of improvements relative to the baseline model. Significantly lower MSFEs only appear for the constant-parameter model with heavy-tailed error distributions at the two longest horizons – improvements which are, again, quantitatively negligible. It can also be noted that at the two longest horizons, MSFEs are substantially higher for the models with time-varying parameters. To some extent, this appears to be driven by the fact that the benchmark Gaussian model with constant parameters tends to have dynamics that are associated by less persistence; this is illustrated in Figures 9 and 10 in Appendix C. This lower persistence – which generates more rapid reversion of the forecasts of the slope to the unconditional mean – pays off on average and very clearly at certain points in time, for example for the forecasts generated between early 2002 and mid-2004. During this period short term movements in the slope were substantial, which the persistent dynamics of the time-varying model could not capture.

Results for the BAA spread are markedly different. Time variation in the parameters

Table 2: MSFEs and relative MSFEs

	1M	2M	3M	6M	12M	24M
(a) Short rate						
G000	0.026	0.079	0.147	0.438	1.332	3.909
G001	0.995	0.995	0.996	1.008	0.994	0.985
G011	0.996	0.998	1.004	1.021	0.947	0.895
G111	0.989	0.978	0.952	0.905	0.942	1.076
T000	0.995*	0.992*	0.990*	0.986*	0.985**	0.990**
T001	0.991*	0.989	0.991	1.000	0.979	0.972
T011	0.993*	0.996	1.001	1.016	0.933	0.882
T111	0.988	0.981	0.960	0.911	0.943	1.071
(b) Slope						
G000	0.051	0.135	0.210	0.406	0.678	1.113
G001	0.998	0.999	1.011	1.070	1.212	1.450
G011	1.005	0.988	0.944	0.926	1.225	1.550
G111	1.004	0.994	0.952	0.928	1.372	1.921
T000	1.002	1.006	1.004	0.996	0.989**	0.989**
T001	1.004	1.009	1.022	1.074	1.205	1.434
T011	1.008	0.989	0.943	0.924	1.213	1.529
T111	1.005	0.993	0.948	0.925	1.366	1.920
(c) BAA spread						
G000	0.042	0.127	0.217	0.470	0.898	1.434
G001	0.911*	0.902*	0.876*	0.836*	0.769**	0.738**
G011	0.920*	0.910	0.884	0.834*	0.745**	0.712**
G111	0.925	0.926	0.910	0.878	0.809	0.820
T000	1.001	1.002	0.999	0.998	0.998	1.001
T001	0.913*	0.916	0.880*	0.831*	0.760**	0.734**
T011	0.922	0.923	0.887	0.839*	0.742**	0.711**
T111	0.928	0.934	0.905	0.861	0.783	0.806*

The first line of each panel reports the MSFE of the benchmark Gaussian VAR model without time-varying parameters during January 2000 to March 2021 (255 estimations). The relative performance is computed as the ratio of the MSFE of alternative specifications over the benchmark; entries less than 1 indicate that the given model is better. ***,**,* denote that the corresponding model significantly outperforms the benchmark at 1%, 5%, 10% level based on the one-sided Diebold and Mariano (1995) test where the standard errors of the test statistics are computed with the Newey–West estimator (Clark, 2011).

brings large improvements relative to the constant-parameter model when Gaussian innovations are assumed. Most of this effect can be attributed to allowing for time variation in the BAA-spread equation, while additional improvements can be had at the longer horizons if we allow for time-variation also in the slope equation. The fully time-varying model always has a higher MSFE than the other specifications with time-varying parameters. Flexibly modelling the innovations does not benefit point forecasts of the BAA spread. Comparing the models with heavy-tailed innovations to their Gaussian counterparts, we see very small differences in relative MSFEs.

Overall, the point forecast performance partly supports the in-sample results with respect to the issue of time variation in coefficients. For example, both the Gaussian and heavy-tailed models with time variation in the equation for the BAA spread generate the lowest MSFEs at horizons up to six months for the BAA spread when compared to their counterparts with other dynamic specifications. However, benefits in terms of forecasting performance from using these models do not carry over to all variables and horizons. We also note that with respect to the question of heavy-tailed innovations, substantial support for heavy tails was found within-sample; however, we see virtually no improvements in forecast precision from allowing for this instead of Gaussian innovations.

4.2.2 Density forecasts

In this section, we consider how time variation in the parameters and heavy-tailed innovations impact density forecasting performance. Table 3 shows results for all variables in a similar structure as for point forecasts – that is, the benchmark model is the constant parameter VAR with Gaussian innovations, for which the values of the LPS are presented. For the other specifications, the improvement over the benchmark is given in units of LPS.

Starting with the short rate, results in panel (a) of Table 3 show that neither time variation in the parameters nor heavy-tailed innovations do much to improve density forecast performance; there is only one significant entry (namely for the model with constant parameters and heavy-tailed innovations). While the magnitude of the difference in LPS is difficult to interpret economically, it can, however, be noted that differences tend to be very small.

For the slope, there are some improvements to be had when one departs from the baseline specification. According to panel (b) of Table 3, some models with time variation in the parameters significantly improve density forecasts at some of the shorter horizons. In line with what we found for point forecasts, it seems that at longer horizons, time-varying parameters can actually deteriorate forecast performance. Having taken time variation into account,

Table 3: LPS and difference in LPS relative to the Gaussian VAR model without time-varying parameter

	1M	2M	3M	6M	12M	24M
(a) Short rate						
G000	0.998	0.374	0.022	-0.718	-1.500	-2.356
G001	-0.001	-0.002	-0.013	-0.022	-0.033	-0.024
G011	-0.004	-0.012	-0.023	-0.044	-0.040	0.022
G111	-0.046	-0.034	-0.026	0.038	0.090	0.122
T000	0.015**	0.008	0.003	0.005	-0.006	-0.009
T001	0.005	0.000	-0.010	-0.018	-0.031	-0.043
T011	0.006	-0.001	-0.019	-0.037	-0.034	0.002
T111	-0.029	-0.013	-0.019	0.043	0.099	0.100
(b) Slope						
G000	0.106	-0.385	-0.639	-1.045	-1.340	-1.607
G001	-0.000	-0.000	-0.005	-0.035	-0.101	-0.200
G011	0.024	0.053*	0.087**	0.105	-0.040	-0.217
G111	0.030	0.059*	0.101**	0.147*	0.011	-0.183
T000	0.006**	0.006**	0.005	0.010***	0.014**	0.007*
T001	0.005*	0.004	-0.004	-0.031	-0.092	-0.197
T011	0.027	0.056*	0.092**	0.110*	-0.026	-0.205
T111	0.031	0.060*	0.102**	0.149*	0.015	-0.184
(c) BAA spread						
G000	0.406	-0.191	-0.510	-1.032	-1.569	-1.962
G001	0.072**	0.065*	0.087	0.227**	0.439***	0.606***
G011	0.051	0.006	0.016	0.233**	0.449***	0.593***
G111	0.059*	0.008	0.035	0.233**	0.435**	0.576***
T000	0.015**	0.010*	0.001	-0.004	-0.001	-0.035
T001	0.098***	0.051	0.092*	0.230**	0.451***	0.626***
T011	0.067**	0.057	0.095*	0.236**	0.454***	0.595***
T111	0.097***	0.058	0.057	0.233**	0.445**	0.579***

The first line of each panel reports the LPS of the benchmark Gaussian VAR model without time-varying parameters during January 2000 to March 2021 (255 estimations). The difference in LPS is computed as the LPS of the alternative specifications minus the LPS of the benchmark model; entries greater than 0 indicate that the given model is better. ***,**,* denote that the corresponding model significantly outperforms the benchmark at 1%, 5%, 10% level based on the one-sided Diebold and Mariano (1995) test where the standard errors of the test statistics are computed with the Newey–West estimator (Clark, 2011).

there appears to be limited benefits from also allowing for heavy tails; the differences in LPS between models with Gaussian and heavy-tailed disturbances are very small.

Lastly, there is strong evidence in favour of time-varying parameter specifications when it comes to predicting the BAA spread. Most specifications at most horizons show significant improvements in density forecasting performance. However, we once again find that the benefits of using models with heavy-tailed disturbances are quite limited once time-varying parameters have been taken into account. This holds for both models with constant parameters and models where time variation in the dynamic parameters are taken into account.

Summing up the density forecast results, we note that they echo the findings from the point forecast evaluation in that allowing for time-varying parameters is beneficial when forecasting the BAA spread, but not necessarily for the other variables of the model. And we again note that allowing for heavy-tailed instead of Gaussian innovations does little to improve forecast performance.[9]

5 Conclusions

In this paper, we have investigated whether the relations between three key interest-rate variables in the United States have been time invariant. This was done using hybrid time-varying parameter Bayesian VARs with stochastic volatility. Extending existing methodology, we allow for the models to have heavy-tailed innovations.

Our results indicate that the relations have not been stable. Model comparison based on marginal likelihoods strongly suggests that the equation of the corporate bond-yield spread is subject to time variation in its parameters but that the other two equations have been stable. Out-of-sample forecasts show that both point and density forecasts of the corporate bond-yield spread tend to improve by taking this time variation into account. The forecasting performance with respect to the three-month Treasury bill rate and the slope, on the other hand, lends less support to the models with time-varying parameters.

Strong support is also found for heavy tails based on the within-sample analysis (marginal likelihood calculations). However, our out-of-sample analysis suggests that it appears to be of more limited importance. This is in line with the somewhat mixed evidence in the related literature; see, for example, Chiu et al. (2017), Liu (2019), Kiss and Österholm (2020), Brunnermeier et al. (2021) and Kiss et al. (2023). It is unclear what this discrepancy between the within- and out-of-sample analysis is due to but it does not seem unlikely that

[9] This is also illustrated by cumulative differences in LPS; see Figures 11 and 12 in Appendix C.

the Bayesian methods that we employ are subject to the traditional problem of within-sample overfitting. However, it can also be the case that it is due to the fact that out-of-sample analysis uses the data inefficiently; see, for example, the discussion in Diebold (2015).

With respect to the effects that the shocks of the preferred model have, these are qualitatively largely in line with what one would expect ex ante. We specifically note that an increase in the corporate bond yield spread decreases the three-month Treasury bill rate, and that an increase in the three-month Treasury bill rate tends to decrease the corporate bond yield spread. However, analysts, forecasters and policymakers – and others concerned with the co-evolution of these variables – should consider the instability established in this paper when conducting analysis or making decisions.

Acknowledgments

We are grateful to seminar participants at Stockholm University, the 12th European Seminar on Bayesian Econometrics in Salzburg and the 5th University of Sydney Time Series and Forecasting Symposium for valuable commments. Financial support from Jan Wallanders och Tom Hedelius stiftelse (grants no. Bv18-0018, P18-0201 and W19-0021) is gratefully acknowledged. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at NSC funded by the Swedish Research Council through grant agreement no. 2022-06725.

Appendix

A Bayesian inference

This section discusses the prior distributions for the parameters in the hybrid TVP-VAR models and the general MCMC inference scheme. The Bayesian inference using a Gibbs sampler is extended to make inference on model parameters.

Following Chan (2023), we obtain the hybrid TVP-VAR by writing the structural form VAR in equation (3) into a recursive system equation by equation using the non-centered parameterization (NCP),

$$\begin{aligned} y_{i,t} &= \mathbf{x}_{i,t}\boldsymbol{\theta}_{i,0} + M_i\mathbf{x}_{i,t}\boldsymbol{\Sigma}_{\theta,i}^{1/2}\widetilde{\boldsymbol{\theta}}_{i,t} + w_{i,t}^{0.5}\exp(0.5h_{i,t})\epsilon_{i,t}, \\ \widetilde{\boldsymbol{\theta}}_{i,t} &= \widetilde{\boldsymbol{\theta}}_{i,t-1} + \widetilde{\boldsymbol{\eta}}_{i,t}, \quad \widetilde{\boldsymbol{\eta}}_{i,t} \sim \mathcal{N}(0, \boldsymbol{I}), \end{aligned} \tag{5}$$

where $\mathbf{x}_{i,t} = (1, \mathbf{y}_{t-1}^{'}, \ldots, \mathbf{y}_{t-p}^{'}, -y_{1,t}, \ldots, -y_{i-1,t})$; $w_{i,t}$ and $\epsilon_{i,t}$ are independently distributed and $\mathbf{w}_t = (w_{1,t}, \ldots w_{k,t})$ is the vector of the mixing variables; $M_i \in \{0, 1\}$ is a binary variable indicating that the ith equation of the VAR system has time-varying coefficients if $M_i = 1$, and constant coefficients if $M_i = 0$ in which case $\boldsymbol{\theta}_{i,0}$ is the vector of fixed regression coefficients. With time-varying coefficents the centered parameterization (CP) is obtained as $\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,0} + M_i\boldsymbol{\Sigma}_{\theta,i}^{1/2}\widetilde{\boldsymbol{\theta}}_{i,t}$. It is also straightforward to see that $\boldsymbol{\theta}_{i,t} - \boldsymbol{\theta}_{i,t-1} = M_i\boldsymbol{\Sigma}_{\theta,i}^{1/2}(\widetilde{\boldsymbol{\theta}}_{i,t} - \widetilde{\boldsymbol{\theta}}_{i,t-1})$. Here, we fix M_i and obtain different model specifications.

The observations in equation (5) can be written as a seemingly unrelated regression,

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\theta}_{i,0} + \mathbf{Z}_i\widetilde{\boldsymbol{\theta}}_i + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i, \\ \mathbf{T}_i\widetilde{\boldsymbol{\theta}}_i &= \widetilde{\boldsymbol{\eta}}_i, \end{aligned} \tag{6}$$

where $\mathbf{Y}_i = (y_{i,1}, \ldots, y_{i,T})^{'}$, $\mathbf{X}_i = (\mathbf{x}_{i,1}^{'}, \ldots, \mathbf{x}_{i,T}^{'})^{'}$, $\mathbf{Z}_i = M_i\,\mathrm{diag}(\mathbf{x}_{i,1}\boldsymbol{\Sigma}_{\theta,i}^{1/2}, \ldots, \mathbf{x}_{i,T}\boldsymbol{\Sigma}_{\theta,i}^{1/2})$, $\widetilde{\boldsymbol{\theta}}_i = (\widetilde{\boldsymbol{\theta}}_{i,1}^{'}, \ldots \widetilde{\boldsymbol{\theta}}_{i,T}^{'})^{'}$, $\mathbf{W}_i = \mathrm{diag}(\mathbf{w}_i) = \mathrm{diag}(w_{i,1}, \ldots, w_{i,T})$, $\mathbf{H}_i = \mathrm{diag}(\exp(\mathbf{h}_i)) = \mathrm{diag}(\exp(h_{i,1}), \ldots, \exp(h_{i,T}))$, $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots \epsilon_{i,T})^{'}$, $\widetilde{\boldsymbol{\eta}}_i = (\widetilde{\boldsymbol{\eta}}_{i,1}^{'}, \ldots \widetilde{\boldsymbol{\eta}}_{i,T}^{'})^{'}$ and $\mathbf{T}_i = \begin{pmatrix} \mathbf{I}_{k_i} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_{k_i} & \mathbf{I}_{k_i} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & -\mathbf{I}_{k_i} & \mathbf{I}_{k_i} \end{pmatrix}$.

A.1 Prior distribution

We denote $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{1:k,0}, \boldsymbol{\Sigma}_{\theta,1:k}, \boldsymbol{\Sigma}_{h,1:k}, \boldsymbol{\nu}, \widetilde{\boldsymbol{\theta}}_{1:k,1:T}, \mathbf{h}_{1:k,0:T}, \mathbf{w}_{1:k,1:T}\}$ as the parameter set of the hybrid TVP-VAR model with OT-SV innovations. Following the prior assumption in Karlsson et al. (2023), we assume vague but proper prior distributions for the model parameters. For example, the prior for the degree of freedom is $\nu_i \sim \mathcal{G}(2, 0.1)$ such that the prior mean of the degrees of freedom is 20. The priors for the variances of the shocks to the volatility and time-varying parameters are $\sigma^2_{h,i} \sim \mathcal{G}(\frac{1}{2}, \frac{1}{2V_h})$ for each i=1,...,k, and $\sigma^2_{\theta,1:k_i} \sim \mathcal{G}(\frac{1}{2}, \frac{1}{2V_\theta})$ where k_i is the number of time-varying parameters in equation i for each i=1,...,k. The hyper parameters are set $V_h = V_\theta = 1$, which lead to $\pm\sqrt{\sigma^2_{h,1:k}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ and $\pm\sqrt{\sigma^2_{\theta,1:k_i}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_i})$; see Kastner and Frühwirth-Schnatter (2014). This prior is less influential in comparison to the conjugated inverse gamma prior especially when the true value is small. The prior of the initial volatility $\mathbf{h}_0$ is $\mathbf{h}_0 \sim N(\mathbf{a}_{h_0}, \mathbf{V}_{h_0})$ where $\mathbf{a}_{h_0} = \log \hat{\Sigma}_{OLS}$ is the estimated residual variance of a univariate AR(p) model using the ordinary least square method, and $\mathbf{V}_{h_0} = 4\mathbf{I}_k$; see Clark (2011). The prior on the initial regression coefficients $\boldsymbol{\theta}_{i,0}$ follows a Minnesota prior with the overall shrinkage $l_1 = 0.2$ and the cross-variable shrinkage $l_2 = 0.5$; see Koop and Korobilis (2010). Finally, the prior for the mixing variables is based on the model assumption $w_{i,t}|\nu_i \sim \mathcal{IG}(\frac{\nu_i}{2}, \frac{\nu_i}{2})$.

A.2 Estimation procedure

We outline the Gibbs sampler for the conditional posteriors. The implementation is described in detail in the Online Appendix.

Let $\boldsymbol{\Psi}$ be a set of conditional parameters except the one that we sample from.

1. Sample $(\widetilde{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i})$ parameters for $i = 1, \ldots, k$, using the ancillarity-sufficiency interweaving strategy (ASIS) by proposed Yu and Meng (2011),
   (a) Draw $\pi(\widetilde{\boldsymbol{\theta}}_i|\boldsymbol{\Psi})$ in the non-centered parameterization (NCP) setting.
   (b) Draw $\pi(\boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i}|\boldsymbol{\Psi})$ in the NCP setting.
   (c) Move $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i,0} + M_i\boldsymbol{\Sigma}^{1/2}_{\theta,i}\widetilde{\boldsymbol{\theta}}_i$ in the centered parameterization (CP) setting.
   (d) Draw $\pi(\boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i}|\boldsymbol{\Psi})$ in the CP setting.
   (e) Move $\widetilde{\boldsymbol{\theta}}_i = (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})/M_i\boldsymbol{\Sigma}^{1/2}_{\theta,i}$ in the NCP setting.
2. Sample $\pi(\mathbf{h}_i, \boldsymbol{\Sigma}_{h,i}|\boldsymbol{\Psi})$ using the ASIS proposed by Kastner and Frühwirth-Schnatter (2014), for $i = 1, \ldots, k$.

3. Sample $w_{i,t}$ for $i = 1, \ldots, k$, and $t = 1, \ldots, T$ from Inverse Gamma distributions following Geweke (1993).

4. Sample $\pi(\nu_i|\boldsymbol{\Psi})$ for $i = 1, \ldots, k$, using an adaptive random walk Metropolis-Hastings algorithm; see Roberts and Rosenthal (2009).

B Model selection

The marginal likelihoods of the hybrid TVP-VAR model with OT-SV innovations requires integrating over a high-dimensional parameter space,

$$p(\mathbf{y}_{1:T}) = \int p(\mathbf{y}_{1:T}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}.$$

We divide the model parameters into two groups with the static parameters $\boldsymbol{\Theta}_1 = \{\boldsymbol{\theta}_{1:k,0}, \boldsymbol{\Sigma}_{\theta,1:k}, \boldsymbol{\Sigma}_{h,1:k}, \boldsymbol{\nu}\}$ and the latent states $\boldsymbol{\Theta}_2 = \{\mathbf{h}_{1:k}, \mathbf{w}_{1:k}\}$. Note that the conditional likelihood $p(\mathbf{Y}_i|\boldsymbol{\Theta}_2, \boldsymbol{\Theta}_1)$ can be derived analytically by integrating over $\widetilde{\boldsymbol{\theta}}_{1:k}$ following Chan and Eisenstat (2018*a*). The marginal likelihood can be obtained via importance sampling

$$\begin{aligned} p(\mathbf{y}_{1:T}) &= \int p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1)\pi(\boldsymbol{\Theta}_1)d\boldsymbol{\Theta}_1, \\ &= \int \frac{p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1)\pi(\boldsymbol{\Theta}_1)}{f(\boldsymbol{\Theta}_1|\lambda)} f(\boldsymbol{\Theta}_1|\lambda)d\boldsymbol{\Theta}_1. \end{aligned} \tag{7}$$

The accuracy of the marginal likelihood depends on the approximation of the integrated likelihood, $p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1)$, and the proposal distribution, $f(\boldsymbol{\Theta}_1|\lambda)$. We apply the cross-entropy methods to learn the proposal distribution $f(\boldsymbol{\Theta}_1; \lambda)$ as an approximation of the posterior samples. We specify that the parametric families of $f(\boldsymbol{\Theta}_1; \lambda)$ are the multivariate Gaussian distribution for $(\boldsymbol{\theta}_{1:k,0})$, the independent Gamma distribution for $\boldsymbol{\nu}$ and independent Gamma distribution for $(\boldsymbol{\Sigma}_{\theta,1:k}, \boldsymbol{\Sigma}_{h,1:k})$. Given the posterior samples $\boldsymbol{\Theta}_1^{(1)}, \ldots, \boldsymbol{\Theta}_1^{(R)}$ from the posterior density $\pi(\boldsymbol{\Theta}|\mathbf{y}_{1:T})$, we find λ^* that maximizes the likelihood estimate,

$$\lambda^* = \arg\max_{\lambda} \frac{1}{R} \sum_{r=1}^{R} \log f(\boldsymbol{\theta}_1^{(r)}; \lambda).$$

The marginal likelihood can be approximated via importance sampling,

$$p(\mathbf{y}_{1:T}) \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1^{(n)})p(\boldsymbol{\Theta}_1^{(n)})}{f(\boldsymbol{\Theta}_1^{(n)}|\lambda^*)},$$

where $\boldsymbol{\Theta}_1^{(1)}, \ldots, \boldsymbol{\Theta}_1^{(N)} \sim f(\boldsymbol{\Theta}_1; \lambda^*)$ are the samples from the proposal distribution of the static parameters. The number of samples N is chosen such that the variance of the estimated quantity using importance sampling is less than one. Here, the integrated likelihood $p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1)$ is calculated using an inner importance sampling loop based on a sparse matrix representation; see the discussion in Chan and Eisenstat (2018*a*),

$$\begin{aligned} p(\mathbf{y}_{1:T}|\boldsymbol{\Theta}_1) &= \prod_{i=1}^{k} p(\mathbf{Y}_i|\boldsymbol{\Theta}_1), \\ &= \prod_{i=1}^{k} \left[\int p(\mathbf{Y}_i|\boldsymbol{\Theta}_2, \boldsymbol{\Theta}_1) p(\boldsymbol{\Theta}_2|\boldsymbol{\Theta}_1) d\boldsymbol{\Theta}_2 \right], \\ &= \prod_{i=1}^{k} \left[\int \frac{p(\mathbf{Y}_i|\mathbf{h}_i, \mathbf{w}_i, \boldsymbol{\Theta}_1) p(\mathbf{h}_i, \mathbf{w}_i|\boldsymbol{\Theta}_1)}{f(\mathbf{h}_i, \mathbf{w}_i)} f(\mathbf{h}_i, \mathbf{w}_i) d\mathbf{h}_i d\mathbf{w}_i \right]. \end{aligned} \tag{8}$$

with importance function $f(\mathbf{h}_i, \mathbf{w}_i) = f(\mathbf{w}_i|\mathbf{h}_i) f(\mathbf{h}_i)$ with a Gaussian distribution for the log volatilities, $\mathbf{h}_i \sim \mathcal{N}(\widehat{\mathbf{h}}_i, \mathbf{K}_{hi}^{-1})$, and inverse Gamma for the mixing variables, $w_{it}|h_{it} \sim \mathcal{IG}(a_{it}, b_{it})$. The Online Appendix describes how the parameters of the importance functions are obtained.

C Further results

Table 4: Log marginal likelihood for VAR models with and without stochastic volatility

Model	Gaussian-nonSV	OT-nonSV	Gaussian-SV	OT-SV
000	185.25	609.24	1125.39	1133.78
100	452.03	632.06	1077.77	1085.11
010	193.39	585.97	1079.97	1087.25
001	202.16	576.10	1158.11	1165.70
110	459.72	604.56	1030.99	1038.52
101	468.76	593.39	1108.88	1115.44
011	210.24	547.85	1111.56	1119.50
111	475.90	561.71	1061.56	1067.67

The cross entropy methods of Chan and Eisenstat (2018*b*) are used to calculate the log marginal likelihoods. The model number indicates the combination of constant and time-varying parameters, for example 001 indicates that only the third equation in the TVP VAR model has time-varying parameters. The order of the variables are 1) the short rate, 2) the slope and 3) the BAA spread. We first sample 120,000 total draws from the conditional posterior distributions with 20,000 discarded as burn-in. Then, all log marginal likelihoods are estimated using 20,000 draws from the proposal distributions; for details, see Appendix B.

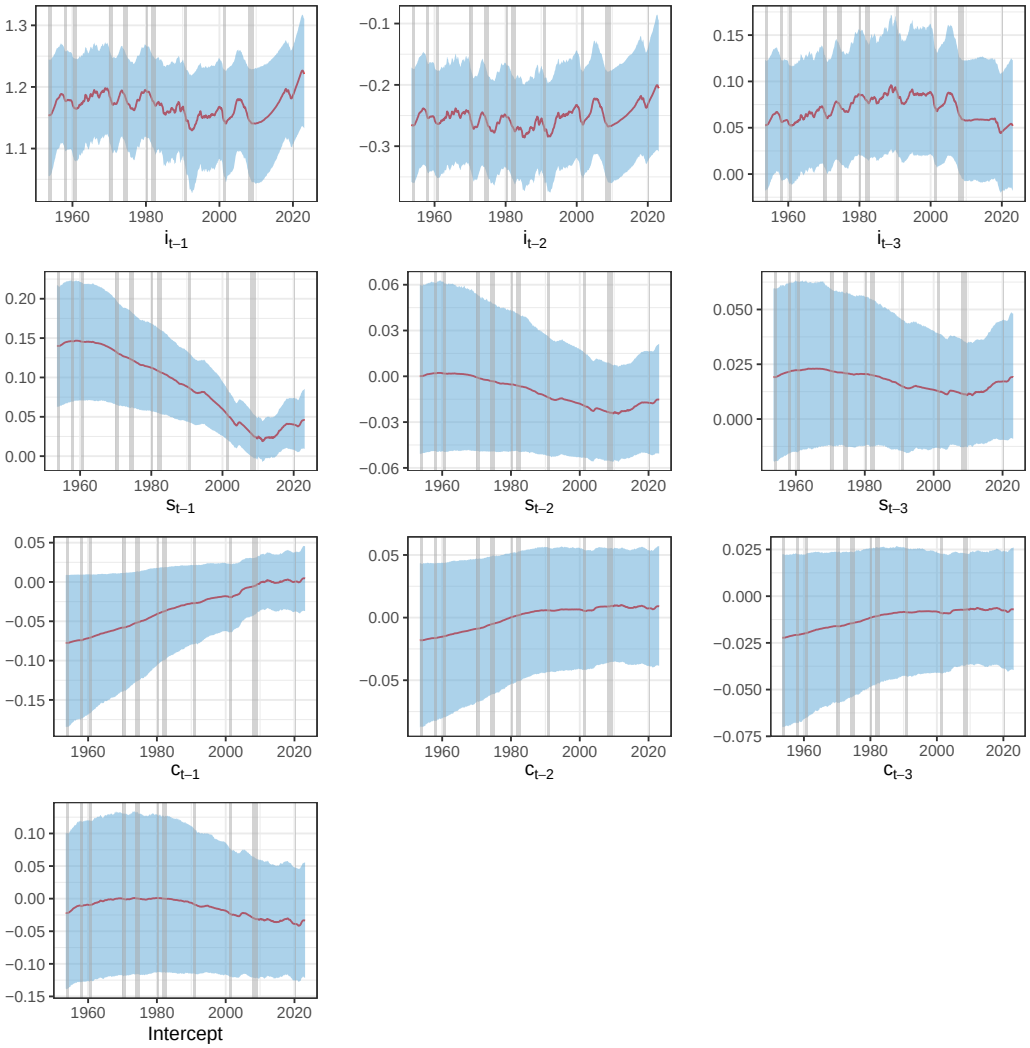

Figure 3: The time-varying parameters $\boldsymbol{\theta}$ of the equation for the short rate of the model with time-varying parameters in all three equations and heavy-tailed innovations.

The figure plots the evolution of the posterior distribution for the time-varying parameters $\boldsymbol{\theta}$ in the short rate equation for the model where dynamic parameters are time-varying in all three equations and innovations are modelled as heavy-tailed. The solid (red) line shows the posterior mean while the blue band indicates the corresponding 80% credible interval. The first, second and the third rows show the coefficients on the short rate, the slope and the BAA-spread, respectively, while different columns representing different lags of the same variables. The last row shows the intercept of the equation.

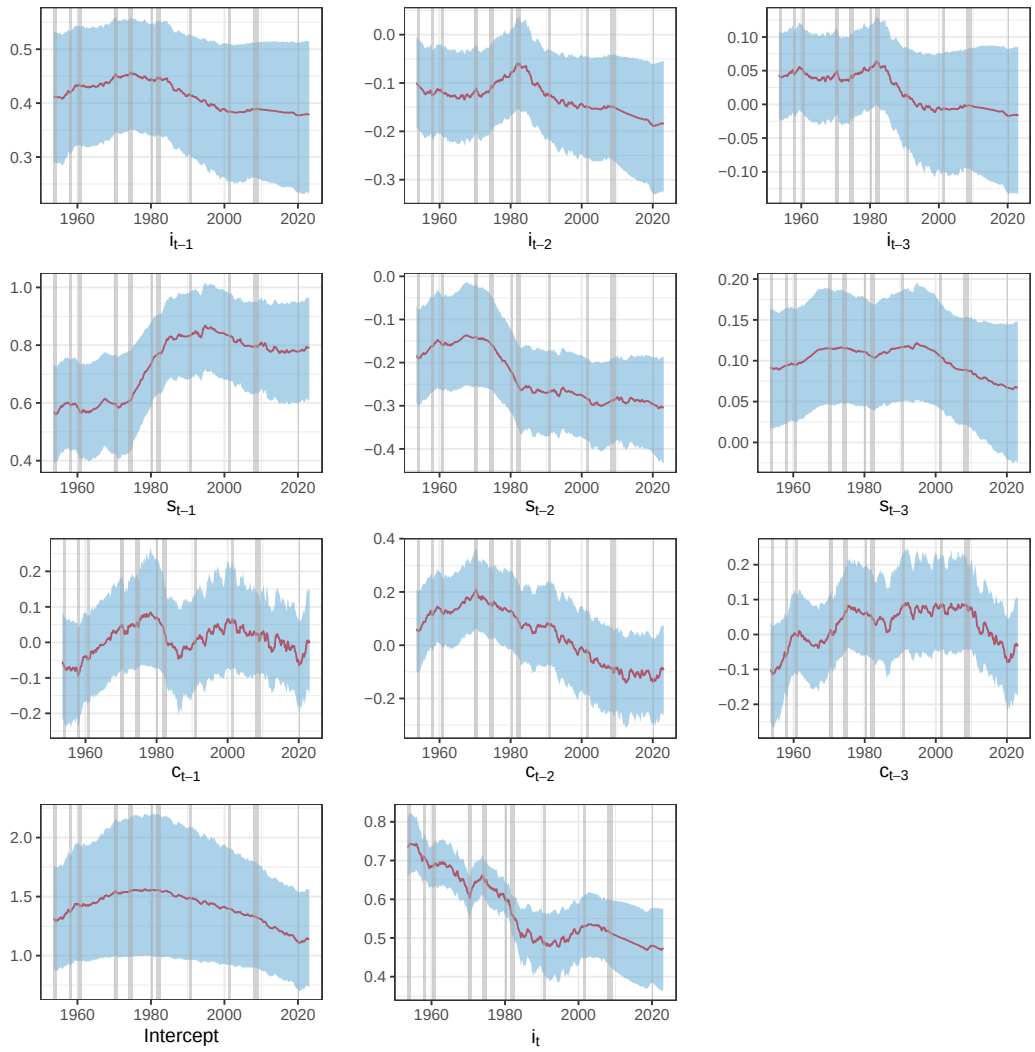

Figure 4: The time-varying parameters $\boldsymbol{\theta}$ of the equation of the slope of the model with time-varying parameters in all three equations and heavy-tailed innovations.

The figure plots the evolution of the posterior distribution for the time-varying parameters $\boldsymbol{\theta}$ in the slope equation for the model where dynamic parameters are time-varying in all three equations and innovations are modelled as heavy-tailed. The solid (red) line shows the posterior mean while the blue band indicates the corresponding 80% credible interval. The first, second and the third rows show the coefficients on the short rate, the slope and the BAA-spread, respectively, while different columns representing different lags of the same variables. The last row shows the intercept of the equation and the parameter capturing the contemporaneous relationship between the short rate and the slope.

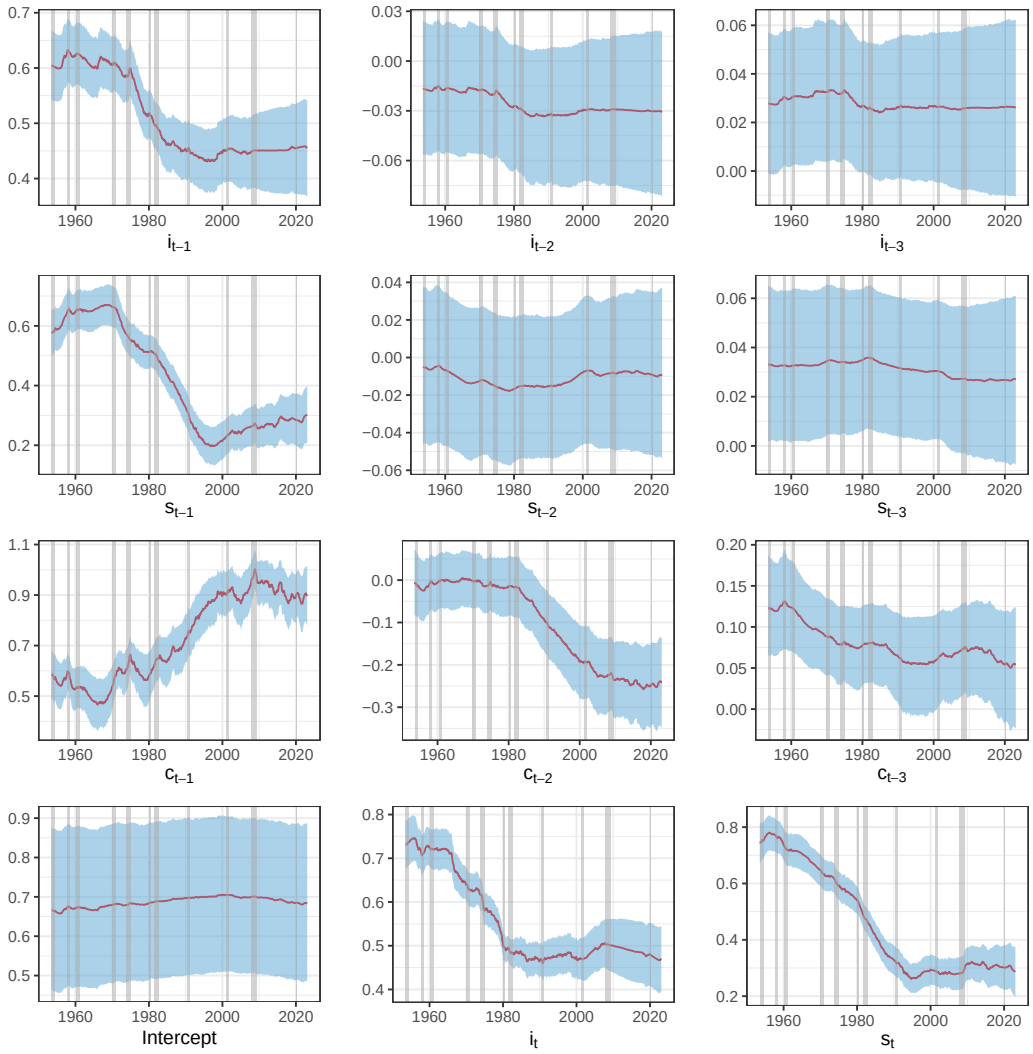

Figure 5: The time-varying parameters $\boldsymbol{\theta}$ of the equation of the BAA spread of the model with time-varying parameters in all three equations and heavy-tailed innovations.

The figure plots the evolution of the posterior distribution for the time-varying parameters $\boldsymbol{\theta}$ in the BAA-spread equation for the model where dynamic parameters are time-varying in all three equations and innovations are modelled as heavy-tailed. The solid (red) line shows the posterior mean while the blue band indicates the corresponding 80% credible interval. The first, second and the third rows show the coefficients on the short rate, the slope and the BAA-spread, respectively, while different columns representing different lags of the same variables. The last row shows the intercept of the equation and the parameters capturing the contemporaneous relationships between the BAA spread and the other variables.

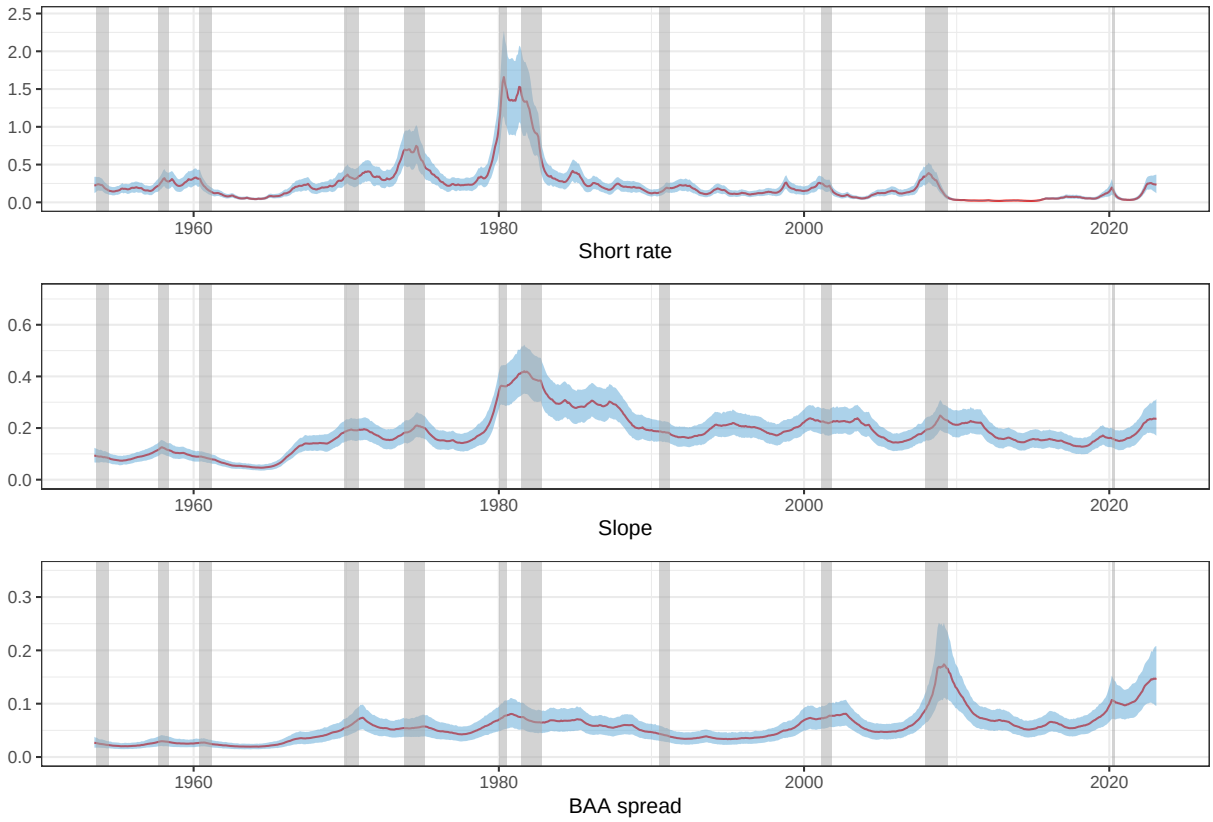

Figure 6: Stochastic volatility estimates $(\exp(\widehat{\mathbf{h}_t}))^{0.5}$ from the model with time-varying parameters in the equation for the BAA spread and innovations with heavy tails.

The red solid line gives the point estimate under the assumption of an Orthogonal Student's t-distribution. The coloured bands bands show the 80% credible interval.

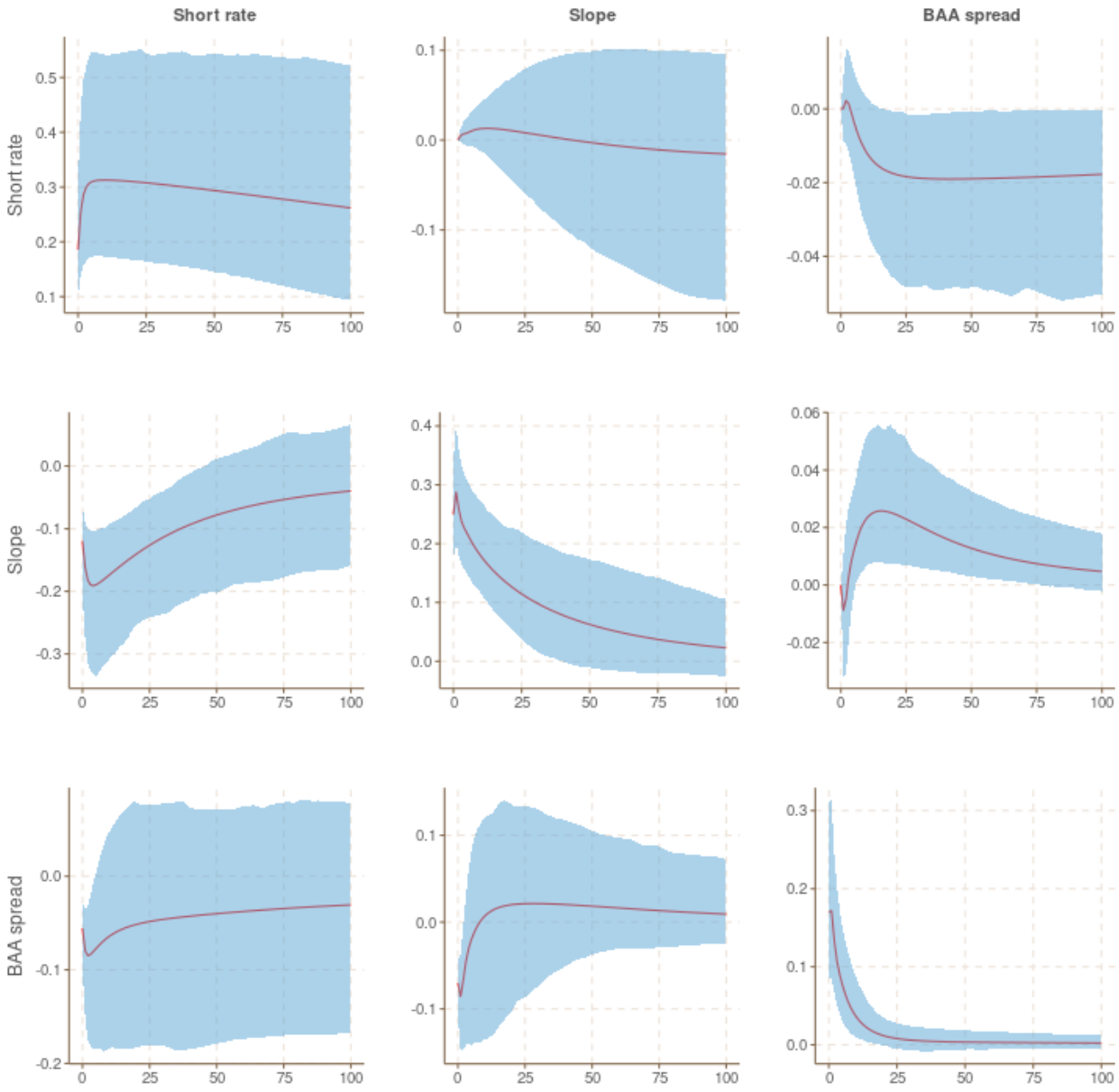

Figure 7: The impulse response functions of the model with time-varying parameters in the equation for the BAA spread and innovations with heavy tails in December 2008.

Impulses are shown in the column headers, response variables are given by the rows. The size of the impulse is one standard deviation. The red solid line gives the point estimate and the coloured bands show the 80% credible interval. The size of the effect in percentage points is represented on the vertical axis. The horizontal axis shows the horizon in months.

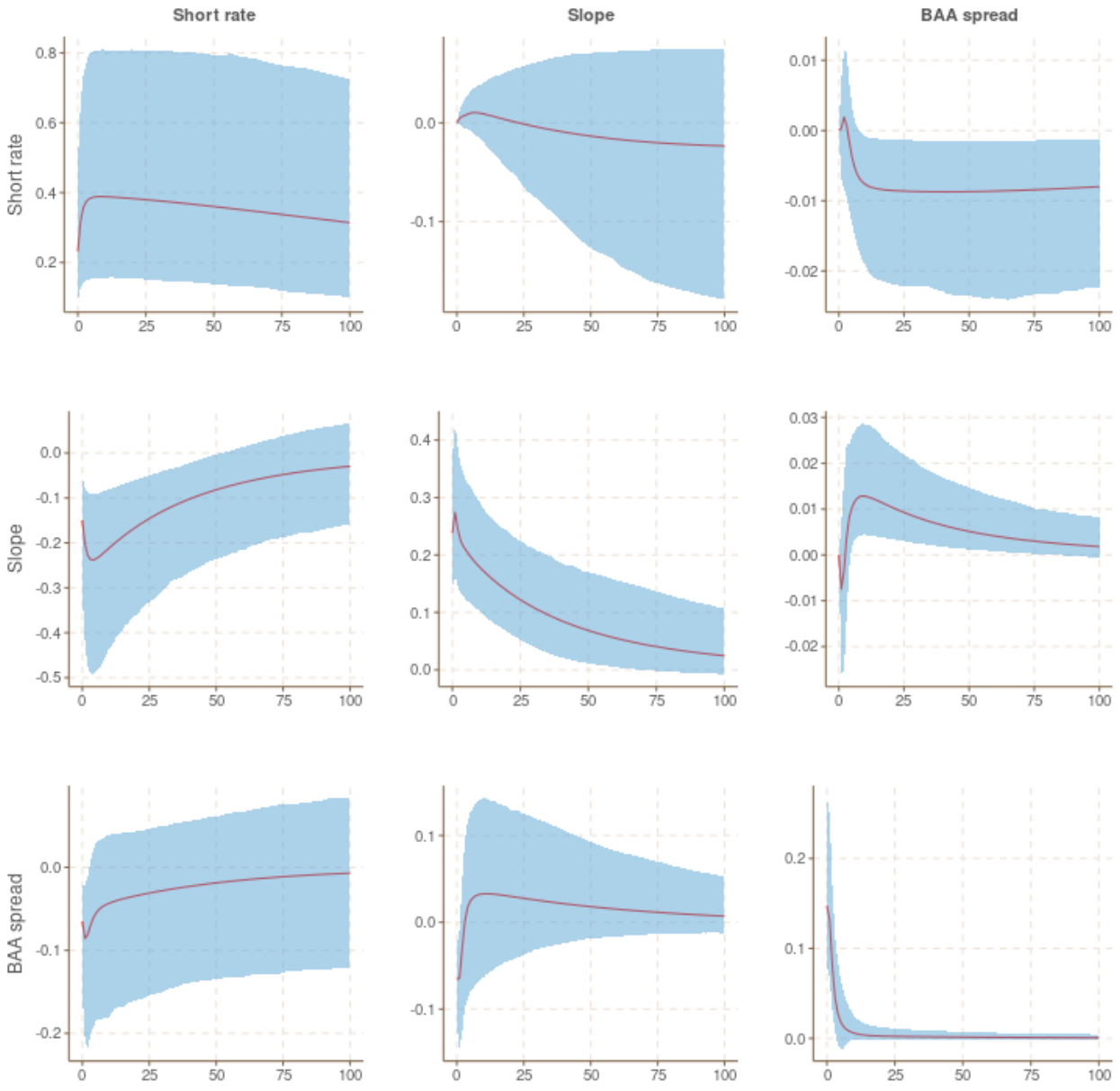

Figure 8: The impulse response functions of the model with time-varying parameters in the equation for the BAA spread and innovations with heavy tails in February 2023.

Impulses are shown in the column headers, response variables are given by the rows. The size of the impulse is one standard deviation. The red solid line gives the point estimate and the coloured bands show the 80% credible interval. The size of the effect in percentage points is represented on the vertical axis. The horizontal axis shows the horizon in months.

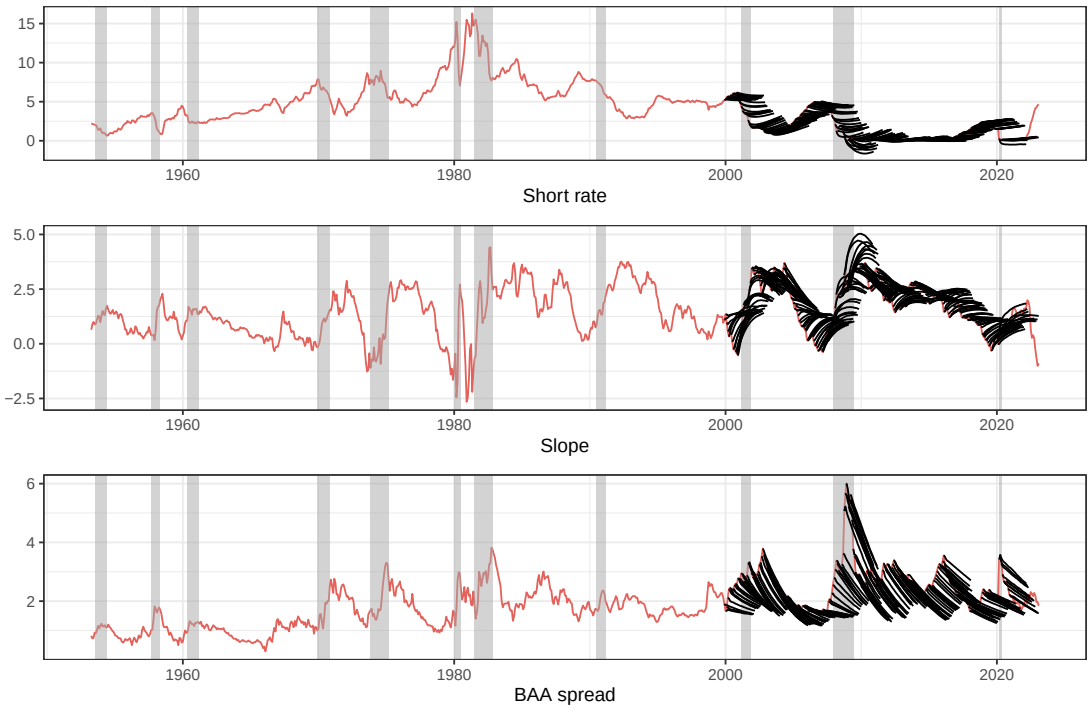

Figure 9: Point forecasts of the model with constant parameters and Gaussian innovations.

The red line represents the realisations of the respective variable and the black lines are its forecasts (up to a two-years horizon), made at different points in time. The model used is the one with constant parameters and Gaussian innovations.

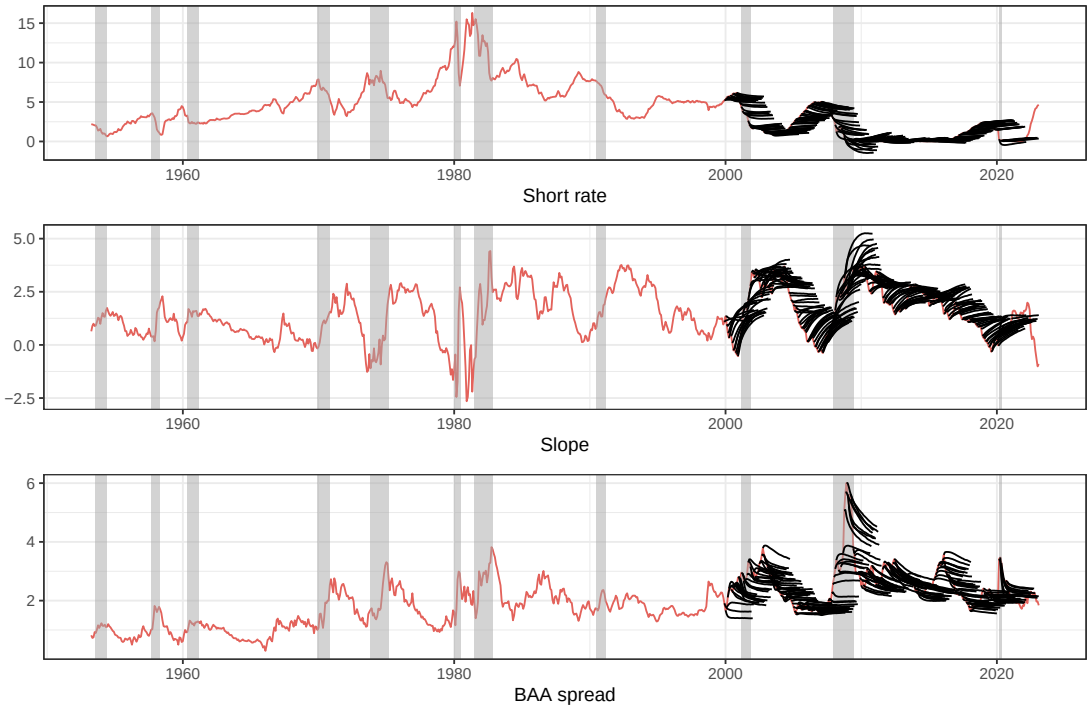

Figure 10: Point forecasts of the model with time-varying parameters in the equation for the BAA-spread and heavy-tailed innovations.

The red line represents the realisations of the respective variable and the black lines are its forecasts (up to a two-years horizon), made at different points in time. The model used is the one with time-varying parameters in the equation for the BAA-spread and heavy-tailed innovations.

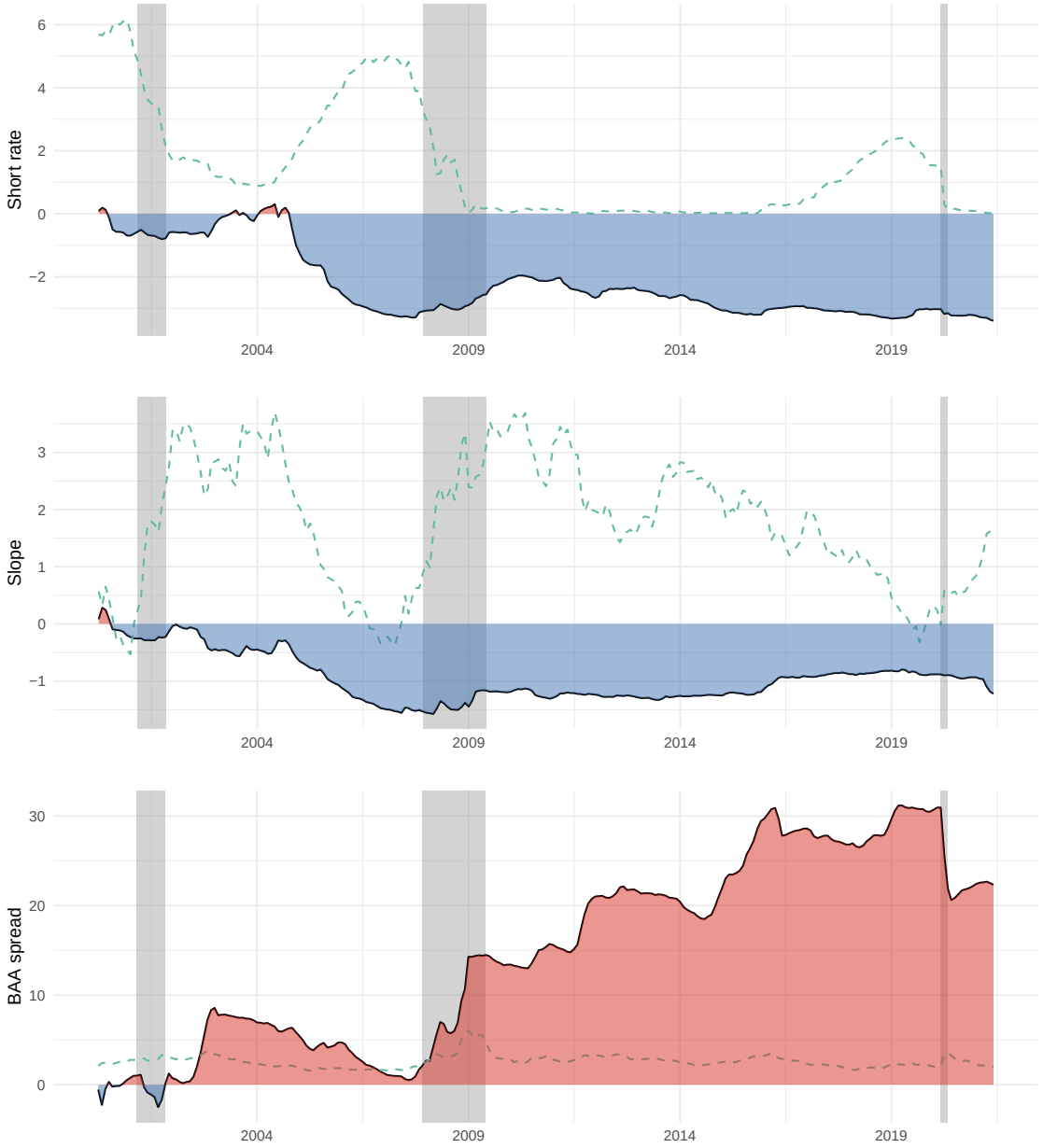

Figure 11: Cumulative difference in log predictive score between the model with time-varying parameters in the equation for the BAA-spread (and innovations with heavy tails) and the constant parameter model (with Gaussian innovations).

Positive values (red) means that the model with time-varying parameters in the equation for the BAA spread predicts better and negative values (blue) means that the constant parameter model does better. The forecast horizon is three months. In both cases, Gaussian innovations are used. The dashed lines illustrate the scale values of the original variables. See Geweke and Amisano (2010) for details.

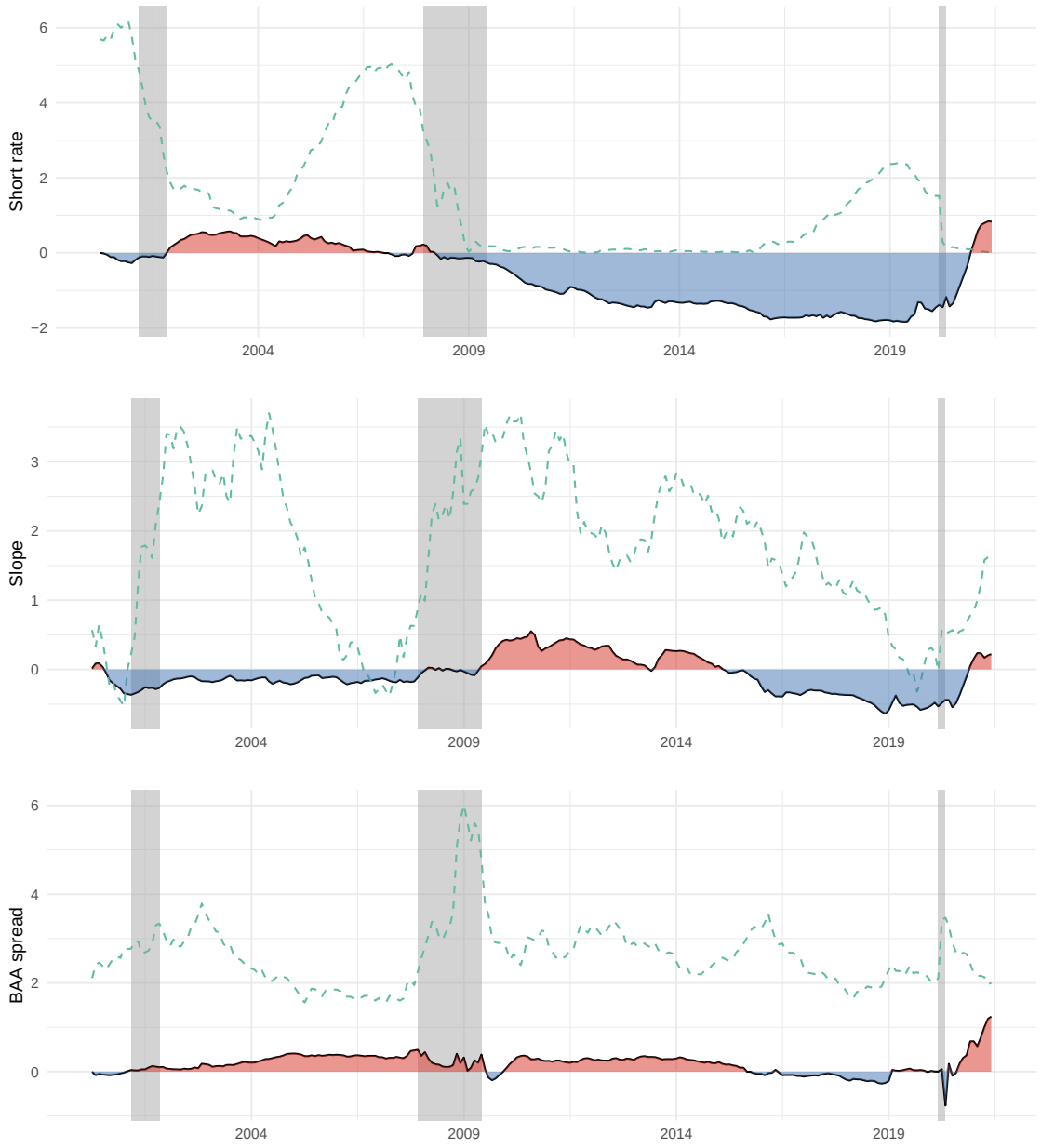

Figure 12: Cumulative difference in log predictive score between the models with time-varying parameter in the equation for the BAA-spread, using heavy-tailed and Gaussian innovations.

Positive values (red) means that the model with heavy-tailed innovations predicts better and negative values (blue) means that the Gaussian model does better. The forecast horizon is three months. The dashed lines illustrate the scale values of the original variables. See Geweke and Amisano (2010) for details.

References

Akram, F. Q. and Mumtaz, H. (2019), 'Time-varying dynamics of the norwegian economy', *The Scandinavian Journal of Economics* **121**(1), 407–434.

Ascari, G., Fagiolo, G. and Roventini, A. (2015), 'Fat-tail distributions and business-cycle models', *Macroeconomic Dynamics* **19**(2), 465–476.

Batten, J. A. and Hogan, W. P. (2003), 'Time variation in the credit spreads on Australian Eurobonds', *Pacific-Basin Finance Journal* **11**(1), 81–99.

Batten, J. A., Hogan, W. P. and Jacoby, G. (2005), 'Measuring credit spreads: evidence from Australian Eurobonds', *Applied Financial Economics* **15**(9), 651–666.

Beechey, M. and Österholm, P. (2012), 'The rise and fall of u.s. inflation persistence', *International Journal of Central Banking* **8**(3), 55–86.

Belongia, M. T. and Ireland, P. N. (2016), 'The evolution of u.s. monetary policy: 2000-2007', *Journal of Economic Dynamics and Control* **73**, 78–93.

Bernanke, B. S. (2020), 'The new tools of monetary policy', *American Economic Review* **110**(4), 943–83.

Bhattarai, S. and Neely, C. J. (2022), 'An analysis of the literature on international unconventional monetary policy', *Journal of Economic Literature* **60**(2), 527–97.

Bianchi, F. and Civelli, A. (2015), 'Globalization and inflation: Evidence from a time-varying var', *Review of Economic Dynamics* **18**(2), 406–433.

Boivin, J. (2006), 'Has u.s. monetary policy changed? evidence from drifting coefficients and real-time data', *Journal of Money, Credit and Banking* **38**(5), 1149–1173.

Boivin, J., Giannoni, M. P. and Stevanovi, D. (2013), 'Dynamic effects of credit shocks in a data-rich environment', *FRB of New York Staff Report 615* .

Brunnermeier, M., Palia, D., Sastry, K. A. and Sims, C. A. (2021), 'Feedbacks: financial markets and economic activity', *American Economic Review* **111**(6), 1845–1879.

Campbell, J. Y. and Taksler, G. B. (2003), 'Equity volatility and corporate bond yields', *The Journal of Finance* **58**(6), 2321–2350.

Carriero, A., Clark, T. E. and Marcellino, M. (2015), 'Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility', *Journal of the Royal Statistical Society. Series A,(Statistics in Society)* **178**(4), 837.

Cecioni, M., Ferrero, G. and Secchi, A. (2011), 'Unconventional monetary policy in theory and in practice', *Banca D'Italia Occasional Papers, No. 102* .

Chan, J. C. (2017), 'The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling', *Journal of Business and Economic Statistics* **35**(1), 17–28.

Chan, J. C. (2020), 'Large Bayesian VARs: A flexible Kronecker error covariance structure', *Journal of Business and Economic Statistics* **38**(1), 68–79.

Chan, J. C. (2023), 'Large hybrid time-varying parameter VARs', *Journal of Business and Economic Statistics* **41**(3), 890–905.

Chan, J. C. and Eisenstat, E. (2018*a*), 'Bayesian model comparison for time-varying parameter VARs with stochastic volatility', *Journal of Applied Econometrics* **33**(4), 509–532.

Chan, J. C. and Eisenstat, E. (2018*b*), 'Comparing hybrid time-varying parameter VARs', *Economics Letters* **171**, 1–5.

Chiu, C.-W. J., Mumtaz, H. and Pinter, G. (2017), 'Forecasting with VAR models: Fat tails and stochastic volatility', *International Journal of Forecasting* **33**(4), 1124–1143.

Clarida, R., Galí, J. and Gertler, M. (2000), 'Monetary policy rules and macroeconomic stability: Evidence and some theory', *Quarterly Journal of Economics* **115**, 147–180.

Clark, T. E. (2011), 'Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility', *Journal of Business and Economic Statistics* **29**(3), 327–341.

Clark, T. E., McCracken, M. W. and Mertens, E. (2020), 'Modeling time-varying uncertainty of multiple-horizon forecast errors', *Review of Economics and Statistics* **102**(1), 17–33.

Clark, T. E. and Ravazzolo, F. (2015), 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics* **30**(4), 551–575.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: Monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics* **8**(2), 262–302.

Collin-Dufresne, P. and Goldstein, R. S. (2001), 'Do credit spreads reflect stationary leverage ratios?', *The Journal of Finance* **56**(5), 1929–1957.

Collin-Dufresne, P., Goldstein, R. S. and Martin, J. S. (2001), 'The determinants of credit spread changes', *The Journal of Finance* **56**(6), 2177–2207.

Cross, J. and Poon, A. (2016), 'Forecasting structural change and fat-tailed events in Australian macroeconomic variables', *Economic Modelling* **58**, 34–51.

Dedola, L., Georgiadis, G., Gräb, J. and Mehl, A. (2021), 'Does a big bazooka matter? Quantitative easing policies and exchange rates', *Journal of Monetary Economics* **117**, 489–506.

Diebold, F. X. (2015), 'Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests', *Journal of Business and Economic Statistics* **33**(1), 1–1.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**(3), 134–144.

Duffee, G. R. (1998), 'The relation between treasury yields and corporate bond yield spreads', *The Journal of Finance* **53**(6), 2225–2241.

Dupoyet, B., Jiang, X. and Zhang, Q. (2023), 'A new take on the relationship between interest rates and credit spreads', *Applied Economics* pp. 1–17.

d'Amico, S., English, W., López-Salido, D. and Nelson, E. (2012), 'The federal reserve's large-scale asset purchase programmes: rationale and effects', *The Economic Journal* **122**(564), F415–F446.

Fagiolo, G., Napoletano, M. and Roventini, A. (2008), 'Are output growth-rate distributions fat-tailed? Some evidence from OECD countries', *Journal of Applied Econometrics* **23**(5), 639–669.

Fountas, S. and Karanasos, M. (2007), 'Inflation, output growth, and nominal and real uncertainty: empirical evidence for the G7', *Journal of International Money and Finance* **26**(2), 229–250.

Friedman, B. M. and Kuttner, K. N. (1998), 'Indicator properties of the paper—bill spread: Lessons from recent experience', *Review of Economics and Statistics* **80**(1), 34–44.

Gagnon, J., Raskin, M., Remache, J. and Sack, B. (2018), 'The financial market effects of the federal reserve's large-scale asset purchases', *24th issue (Mar 2011) of the International Journal of Central Banking* .

Geweke, J. (1993), 'Bayesian treatment of the independent Student-t linear model', *Journal of Applied Econometrics* **8**(S1), S19–S40.

Geweke, J. and Amisano, G. (2010), 'Comparing and evaluating Bayesian predictive distributions of asset returns', *International Journal of Forecasting* **26**(2), 216–230.

Giesecke, K., Longstaff, F. A., Schaefer, S. and Strebulaev, I. (2011), 'Corporate bond default risk: A 150-year perspective', *Journal of Financial Economics* **102**(2), 233–250.

Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009), 'Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets', *Journal of Monetary Economics* **56**(4), 471–493.

Gilchrist, S. and Zakrajšek, E. (2012), 'Credit spreads and business cycle fluctuations', *American Economic Review* **102**(4), 1692–1720.

Hancock, D. and Passmore, W. (2011), 'Did the federal reserve's mbs purchase program lower mortgage rates?', *Journal of Monetary Economics* **58**(5), 498–514.

Ihrig, J., Klee, E., Li, C., Wei, M. and Kachovec, J. (2018), 'Expectations about the federal reserve's balance sheet and the term structure of interest rates', *International Journal of Central Banking* **14**(2), 341–390.

IMF (2013), 'Unconventional monetary policies—recent experience and prospects.'.

Inoue, A. and Rossi, B. (2019), 'The effects of conventional and unconventional monetary policy on exchange rates', *Journal of International Economics* **118**, 419–447.

Karlsson, S., Mazur, S. and Nguyen, H. (2023), 'Vector autoregression models with skewness

and heavy tails', *Journal of Economic Dynamics and Control* **146**, 104580.

Karlsson, S. and Österholm, P. (2019), 'Volatilities, drifts and the relation between treasury yields and corporate bond yield spreads in Australia', *Finance Research Letters* **30**, 378–384.

Karlsson, S. and Österholm, P. (2020), 'The relation between the corporate bond-yield spread and the real economy: Stable or time-varying?', *Economics Letters* **186**, 108883.

Karlsson, S. and Österholm, P. (2023), 'Is the US Phillips curve stable? Evidence from Bayesian vector autoregressions', *Scandinavian Journal of Economics* **125**(1), 287–314.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kastner, G. and Frühwirth-Schnatter, S. (2014), 'Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models', *Computational Statistics and Data Analysis* **76**, 408–423.

Kiss, T., Mazur, S., Nguyen, H. and Österholm, P. (2023), 'Modeling the relation between the us real economy and the corporate bond-yield spread in bayesian vars with non-gaussian innovations', *Journal of Forecasting* **42**(2), 347–368.

Kiss, T. and Österholm, P. (2020), 'Fat tails in leading indicators', *Economics Letters* **193**, 109317.

Koop, G. and Korobilis, D. (2010), *Bayesian multivariate time series methods for empirical macroeconomics*, Now Publishers Inc.

Koop, G. and Korobilis, D. (2019), 'Forecasting with high-dimensional panel VARs', *Oxford Bulletin of Economics and Statistics* **81**(5), 937–959.

Liu, X. (2019), 'On tail fatness of macroeconomic dynamics', *Journal of Macroeconomics* **62**, 103154.

Longstaff, F. A. and Schwartz, E. S. (1995), 'A simple approach to valuing risky fixed and floating rate debt', *The Journal of Finance* **50**(3), 789–819.

Merton, R. C. (1974), 'On the pricing of corporate debt: The risk structure of interest rates', *The Journal of Finance* **29**(2), 449–470.

Miranda-Agrippino, S. and Nenova, T. (2022), 'A tale of two global monetary policies', *Journal of International Economics* **136**, 103606.

Mody, A. and Taylor, M. P. (2004), 'Financial predictors of real activity and the financial accelerator', *Economics Letters* **82**(2), 167–172.

Mumtaz, H. and Petrova, K. (2022), 'Changing impact of shocks: A time-varying proxy SVAR approach', *Journal of Money, Credit and Banking* .

Neal, R., Rolph, D., Dupoyet, B. and Jiang, X. (2015), 'Interest rates and credit spread dynamics', *Journal of Derivatives* **23**(1), 25.

Neely, C. J. (2015), 'Unconventional monetary policy had large international effects', *Journal of Banking and Finance* **52**, 101–111.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Nielsen, S. S. and Ronn, E. I. (1996), *The valuation of default risk in corporate bonds and interest rate swaps*, Citeseer.

Österholm, P. (2018), 'The relation between treasury yields and corporate bond yield spreads in Australia: Evidence from VARs', *Finance Research Letters* **24**, 186–192.

Prieto, E., Eickmeier, S. and Marcellino, M. (2016), 'Time variation in macro-financial linkages', *Journal of Applied Econometrics* **31**(7), 1215–1233.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Roberts, G. O. and Rosenthal, J. S. (2009), 'Examples of adaptive MCMC', *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Romer, C. and Romer, D. (1989), Does monetary policy matter? a new test in the spirit of friedman and schwartz, *in* O. Blanchard and S. Fischer, eds, 'NBER Macroeconomics Annual', Vol. 4, MIT Press, pp. 121–170.

Rosa, C. (2012), 'How 'unconventional' are large-scale asset purchases? the impact of monetary policy on asset prices', *FRB of New York Staff Report 560* .

Rossi, B. (2021), 'Identifying and estimating the effects of unconventional monetary policy: How to do it and what have we learned?', *The Econometrics Journal* **24**(1), C1–C32.

Sargent, T. J. (1999), *The conquest of American inflation*, Princeton University Press.

Stock, J. H. and Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *NBER Macroeconomics Annual* **4**, 351–394.

Trypsteen, S. (2017), 'The importance of time-varying volatility and country interactions in forecasting economic activity', *Journal of Forecasting* **36**(6), 615–628.

Van Landschoot, A. (2008), 'Determinants of yield spread dynamics: Euro versus us dollar corporate bonds', *Journal of Banking and Finance* **32**(12), 2597–2605.

Wu, M. T. (2014), *Unconventional monetary policy and long-term interest rates*, International Monetary Fund.

Yu, Y. and Meng, X.-L. (2011), 'To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency', *Journal of Computational and Graphical Statistics* **20**(3), 531–570.

US Interest Rates: Are Relations Stable? Online appendix

Sune Karlsson[(a)], Tamás Kiss[(a)], Hoang Nguyen[(b)] and Pär Österholm [(a,c)]

[(a)] School of Business, Örebro University

[(b)] Department of Management and Engineering, Linköping University

[(c)] National Institute of Economic Research

March 8, 2024

A Bayesian inference

This section describes the MCMC inference scheme for the parameters in the hybrid TVP-VAR models in detail.

We begin with the hybrid TVP-VAR model as a seemingly unrelated regression,

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\theta}_{i,0} + \mathbf{Z}_i\widetilde{\boldsymbol{\theta}}_i + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i, \\ \mathbf{T}_i\widetilde{\boldsymbol{\theta}}_i &= \widetilde{\boldsymbol{\eta}}_i, \end{aligned} \tag{1}$$

where $\mathbf{Y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\mathbf{X}_i = (\mathbf{x}_{i,1}', \ldots, \mathbf{x}_{i,T}')'$, $\mathbf{Z}_i = M_i \operatorname{diag}(\mathbf{x}_{i,1}\boldsymbol{\Sigma}_{\theta,i}^{1/2}, \ldots, \mathbf{x}_{i,T}\boldsymbol{\Sigma}_{\theta,i}^{1/2})$, $\widetilde{\boldsymbol{\theta}}_i = (\widetilde{\boldsymbol{\theta}}_{i,1}', \ldots \widetilde{\boldsymbol{\theta}}_{i,T}')'$, $\mathbf{W}_i = \operatorname{diag}(\mathbf{w}_i) = \operatorname{diag}(w_{i,1}, \ldots, w_{i,T})$, $\mathbf{H}_i = \operatorname{diag}(\exp(\mathbf{h}_i)) = \operatorname{diag}(\exp(h_{i,1}), \ldots, \exp(h_{i,T}))$, $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots \epsilon_{i,T})'$, $\widetilde{\boldsymbol{\eta}}_i = (\widetilde{\boldsymbol{\eta}}_{i,1}', \ldots \widetilde{\boldsymbol{\eta}}_{i,T}')'$, $\mathbf{T}_i = \begin{pmatrix} \mathbf{I}_{k_i} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_{k_i} & \mathbf{I}_{k_i} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & -\mathbf{I}_{k_i} & \mathbf{I}_{k_i} \end{pmatrix}$.

Given the latent variables $\mathbf{w}_{1:k,1:T}$, the conditional posterior distributions of the remaining parameters in the TVP VAR model with OT-SV innovations are similar to those in the VAR model with Gaussian-SV innovations. The Bayesian inference using a Gibbs sampler is

extended to make inferences on model parameters. To simplify the notation, let $\boldsymbol{\Psi}$ be a set of conditional parameters except the one that we sample from.

1. We sample $(\widetilde{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i})$ for $i = 1, \ldots, k$ using the ASIS by proposed Yu and Meng (2011),

   (a) Draw $\pi(\widetilde{\boldsymbol{\theta}}_i|\boldsymbol{\Psi})$ in the NCP setting by writing Equation 1 as a multivariate linear regression,

   $$\begin{aligned}\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0} &= \mathbf{Z}_i\widetilde{\boldsymbol{\theta}}_i + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i,\\ \mathbf{T}_i\widetilde{\boldsymbol{\theta}}_i &= \widetilde{\boldsymbol{\eta}}_i.\end{aligned}$$

   As the prior distribution is $\widetilde{\boldsymbol{\theta}}_i \sim \mathcal{N}(\mathbf{0}, (\mathbf{T}_i'\mathbf{T}_i)^{-1})$, the conditional posterior distribution of $\widetilde{\boldsymbol{\theta}}_i$ is a conjugate Gaussian distribution $\widetilde{\boldsymbol{\theta}}_i \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_i, \mathbf{K}_{\theta,i}^{-1})$ for $i = 1, \ldots k$, where

   $$\begin{aligned}\hat{\boldsymbol{\theta}}_i &= \mathbf{K}_{\theta,i}^{-1}\mathbf{d}_{\theta,i},\\ \mathbf{K}_{\theta,i} &= \mathbf{T}_i'\mathbf{T}_i + \mathbf{Z}_i'\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}\mathbf{Z}_i,\\ \mathbf{d}_{\theta,i} &= \mathbf{Z}_i'\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0}).\end{aligned} \tag{2}$$

   (b) Draw $\pi(\boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i}|\boldsymbol{\Psi})$ in the NCP setting by writing Equation 1 as a multivariate linear regression,

   $$\begin{aligned}\mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\theta}_{i,0} + M_i\left(\mathbf{X}_i \odot \widetilde{\boldsymbol{\theta}}_i\right)\boldsymbol{\sigma}_{\theta,i} + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i\\ &= \begin{pmatrix}\mathbf{X}_i & M_i\left(\mathbf{X}_i \odot \widetilde{\boldsymbol{\theta}}_i\right)\end{pmatrix}\begin{pmatrix}\boldsymbol{\theta}_{i,0}\\ \boldsymbol{\sigma}_{\theta,i}\end{pmatrix} + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i,\\ \mathbf{Y}_i &= \widetilde{\mathbf{X}}_i\boldsymbol{b}_i + \mathbf{W}_i^{1/2}\mathbf{H}_i^{1/2}\boldsymbol{\epsilon}_i,\end{aligned}$$

   where $\widetilde{\mathbf{X}}_i = \begin{pmatrix}\mathbf{X}_i & M_i\left(\mathbf{X}_i \odot \widetilde{\boldsymbol{\theta}}_i\right)\end{pmatrix}$ and $\boldsymbol{b}_i = \begin{pmatrix}\boldsymbol{\theta}_{i,0}\\ \boldsymbol{\sigma}_{\theta,i}\end{pmatrix}$. As the $\boldsymbol{\theta}_{i,0}$ follows a Minnesota prior and $\pm\sqrt{\sigma^2_{\theta,1:k}} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\theta)$, $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{b}_{i0}, \mathbf{V}_{\mathbf{b}_{i,0}})$. The conditional posterior distribution of $\boldsymbol{b}_i$ is a conjugate Gaussian distribution

   $$\pi(\boldsymbol{b}_i|\boldsymbol{\Psi}) \sim \mathbf{N}(\boldsymbol{b}_i^*, \mathbf{V}_{\boldsymbol{b}_i}^*),$$

   where

   $$\begin{aligned}\mathbf{V}_{\mathbf{b}_i}^{*-1} &= \mathbf{V}_{\mathbf{b}_{i,0}}^{-1} + \sum_{t=1}^{T}\widetilde{\mathbf{X}}_i'\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}\widetilde{\mathbf{X}}_i,\\ \mathbf{b}_i^* &= \mathbf{V}_{\mathbf{b}_i}^*\left[\mathbf{V}_{\mathbf{b}_{i,0}}^{-1}\mathbf{b}_{i0} + \sum_{t=1}^{T}\widetilde{\mathbf{X}}_i'\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}\mathbf{Y}_i\right].\end{aligned}$$

(c) Move $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i,0} + M_i \boldsymbol{\Sigma}_{\theta,i}^{1/2} \widetilde{\boldsymbol{\theta}}_i$ in the CP setting.

(d) Draw $\pi(\boldsymbol{\theta}_{i,0}, \boldsymbol{\Sigma}_{\theta,i} | \boldsymbol{\Psi})$ in the CP setting.

The conditional posterior $\pi(\sigma_{\theta,i}^2 | \boldsymbol{\Psi})$ is generalized inverse Gaussian distribution (GIG) as,

$$\pi(\sigma_{\theta,i}^2 | \boldsymbol{\Psi}) \propto \left(\sigma_{\theta,i}^2\right)^{-\frac{T}{2}} \exp\left(- \frac{\sum_{t=1}^{T} (\theta_{i,t} - \theta_{i,t-1})^2}{2\sigma_{\theta,i}^2} \right) \left(\sigma_{\theta,i}^2\right)^{-\frac{1}{2}} \exp\left(-\frac{\sigma_{\theta,i}^2}{2V_\theta} \right).$$

We sample $\sigma_{\theta,i}^2 \sim GIG(\lambda, \psi, \chi)$ where $\lambda = -0.5(T-1)$, $\chi = \sum_{t=1}^{T} (\theta_{i,t} - \theta_{i,t-1})^2$ and $\psi = 1/V_\theta$, see Hörmann and Leydold (2014) for more details. And we sample $\pi(\boldsymbol{\theta}_{i,0} | \boldsymbol{\Psi}) \sim \mathbf{N}(\boldsymbol{\theta}_{i,0}^*, \mathbf{V}_{\boldsymbol{\theta}_{i,0}}^*)$ where

$$\begin{aligned} \mathbf{V}_{\boldsymbol{\theta}_{i,0}}^{*-1} &= \mathbf{V}_{\boldsymbol{\theta}_{i,0}}^{-1} + \boldsymbol{\Sigma}_{\theta,i}^{-1}, \\ \boldsymbol{\theta}_{i,0}^* &= \mathbf{V}_{\boldsymbol{\theta}_{i,0}}^{*} \left[\mathbf{V}_{\boldsymbol{\theta}_{i,0}}^{-1} \boldsymbol{\theta}_{i,0} + \boldsymbol{\Sigma}_{\theta,i}^{-1} \boldsymbol{\theta}_{i,1} \right]. \end{aligned}$$

(e) Move $\widetilde{\boldsymbol{\theta}}_i = \left(M_i \boldsymbol{\Sigma}_{\theta,i}^{1/2} \right)^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}) /$ in the NCP setting.

2. We sample $\pi(\mathbf{h}_{1:k}, \boldsymbol{\Sigma}_{h,1:k} | \boldsymbol{\Psi})$ using the ASIS proposed by Kastner and Frühwirth-Schnatter (2014). Let $\tilde{u}_{i,t} = w_{i,t}^{-1/2} (y_{i,t} - \mathbf{x}_{i,t} \boldsymbol{\theta}_{i,0} + M_i \mathbf{x}_{i,t} \boldsymbol{\Sigma}_{\theta,i}^{1/2} \widetilde{\boldsymbol{\theta}}_{i,t})$, for each series $i = 1, \ldots, k$, we have that $\log \tilde{u}_{it}^2 = h_{it} + \log \epsilon_t^2$. We sample $\pi(\mathbf{h}_i | \boldsymbol{\Psi})$ in the CP-NCP setting.

(a) Draw $\pi(\mathbf{h}_{1:k} | \boldsymbol{\Psi})$ in the CP setting (Kastner and Frühwirth-Schnatter, 2014) and draw $\mathbf{h}_0 \sim \mathcal{N}(\widehat{\mathbf{h}}_0, \mathbf{K}_{h_0}^{-1})$ in CP setting where

$$\begin{aligned} \mathbf{K}_{h_0} &= \mathbf{V}_{h_0}^{-1} + \boldsymbol{\Sigma}_h^{-1}, \\ \widehat{\mathbf{h}}_0 &= \mathbf{K}_{h_0}^{-1} (\mathbf{V}_{h_0}^{-1} \mathbf{a}_{h_0} + \boldsymbol{\Sigma}_h^{-1} \mathbf{h}_1). \end{aligned}$$

(b) Draw $\pi(\boldsymbol{\Sigma}_{h,i} | \boldsymbol{\Psi})$ for $i = 1, \ldots, k$ using a $GIG(\lambda, \psi, \chi)$ where $\lambda = -0.5(T-1)$, $\chi = \sum_{t=1}^{T} (h_{i,t} - h_{i,t-1})^2$ and $\psi = 1/V_h$, see Hörmann and Leydold (2014), as the

conditional posterior $\pi(\sigma^2_{h,i}|\boldsymbol{\Psi})$ is

$$\pi(\sigma^2_{h,i}|\boldsymbol{\Psi}) \propto \left(\sigma^2_{h,i}\right)^{-\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^{T}(h_{i,t}-h_{i,t-1})^2}{2\sigma^2_{h,i}}\right)\left(\sigma^2_{h,i}\right)^{-\frac{1}{2}} \exp\left(-\frac{\sigma^2_{h,i}}{2V_h}\right).$$

(c) Move $\widetilde{\mathbf{h}}_i = \mathbf{h}_i/\sigma_{h,i}$ for $i = 1, \ldots, k$.

(d) Draw $\pi(\boldsymbol{\Sigma}_{h,1:k}|\boldsymbol{\Psi})$ in the NCP setting (Kastner and Frühwirth-Schnatter, 2014).

(e) Move $\mathbf{h}_i = \widetilde{\mathbf{h}}_i\sigma_{h,i}$ for $i = 1, \ldots, k$.

3. We sample $w_{i,t} \sim \mathcal{IG}\left(\frac{\nu_i+1}{2}, \frac{\nu_i}{2} + \frac{(y_{i,t}-\mathbf{x}_{i,t}\boldsymbol{\theta}_{i,0}-M_i\mathbf{x}_{i,t}\widetilde{\boldsymbol{\theta}}_{i,t})^2}{2\exp(h_{i,t})}\right)$ for $i = 1, \ldots, k$, $t = 1, \ldots, T$ following Geweke (1993).

4. We sample $\pi(\nu_i|\boldsymbol{\Psi}) \propto \mathcal{G}(\nu_i; 2, 0.1)\prod_{t=1}^{T}\mathcal{IG}\left(w_t; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right)$ for $i = 1, \ldots, k$, using an adaptive random walk Metropolis-Hastings algorithm to accept/reject the draw $\nu_i^{(*)} = \nu_i + \eta_i \exp(c_i)$, where $\eta_i \sim \mathcal{N}(0, 1)$ and the adaptive variance c_i is adjusted automatically such that the acceptance rate is around 0.25 (Roberts and Rosenthal, 2009).

B Marginal likelihood: The inner importance sampling

Recall that the objective is to estimate the integrated likelihood $p(\mathbf{y}_{1:T}|\mathbf{\Theta}_1)$, that is, we need to integrate out $\mathbf{\Theta}_2 = \{\mathbf{h}_{1:k}, \mathbf{w}_{1:k}\}$ from $p(\mathbf{y}_{1:T}|\mathbf{\Theta}_1, \mathbf{\Theta}_2)$. This is done using an inner importance sampling loop based on a sparse matrix representation as in Chan and Eisenstat (2018). Write the integrated likelihood as

$$\begin{aligned} p(\mathbf{y}_{1:T}|\mathbf{\Theta}_1) &= \prod_{i=1}^{k} p(\mathbf{Y}_i|\mathbf{\Theta}_1), \\ &= \prod_{i=1}^{k} \left[\int p(\mathbf{Y}_i|\mathbf{\Theta}_2, \mathbf{\Theta}_1) p(\mathbf{\Theta}_2|\mathbf{\Theta}_1) d\mathbf{\Theta}_2 \right], \\ &= \prod_{i=1}^{k} \left[\int \frac{p(\mathbf{Y}_i|\mathbf{h}_i, \mathbf{w}_i, \mathbf{\Theta}_1) p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{\Theta}_1)}{f(\mathbf{h}_i, \mathbf{w}_i)} f(\mathbf{h}_i, \mathbf{w}_i) d\mathbf{h}_i d\mathbf{w}_i \right]. \end{aligned} \tag{3}$$

The ideal zero-variance importance sampling density is

$$\begin{aligned} f(\mathbf{h}_i, \mathbf{w}_i) &= p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{y}_{1:T}, \mathbf{\Theta}_1), \text{or} \\ f(\mathbf{w}_i|\mathbf{h}_i) f(\mathbf{h}_i) &= \frac{p(\mathbf{h}_i, \mathbf{w}_i, \widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \mathbf{\Theta}_1)}{p(\widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \mathbf{h}_i, \mathbf{w}_i, \mathbf{\Theta}_1)}. \end{aligned} \tag{4}$$

However, the zero-variance importance density is not available due to the unknown normalizing constant. We approximate $p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{Y}_i, \mathbf{\Theta}_1)$ using a Gaussian density for the log volatility and Inverse Gamma density for the mixing variables, $\mathbf{h}_i \sim \mathcal{N}(\widehat{\mathbf{h}}_i, \mathbf{K}_{hi}^{-1})$ and $w_{it}|h_{it} \sim \mathcal{IG}(a_{it}, b_{it})$. For the proposal density $f(\mathbf{h}_i)$, we find $\widehat{\mathbf{h}}_i$ as the mode of $\log p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{Y}_i, \mathbf{\Theta}_1)$ using an expectation maximization (EM) algorithm, and $\mathbf{K}_{hi}$ as the negative Hessian of this log density evaluated at the mode. In the E-step, let $\mathbf{lw}_i = \log(\mathbf{w}_i)$, we compute

$$\begin{aligned} &\mathbf{Q}(\mathbf{h}_i, \mathbf{lw}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i) = \mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i} \Big(\log p(\mathbf{h}_i, \mathbf{lw}_i, \widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \mathbf{\Theta}_1) \Big) \\ &= const - \mathbf{1}_T^{'} \frac{\nu_i}{2} \mathbf{lw}_i - \mathbf{1}_T^{'} \frac{\nu_i}{2} e^{-\mathbf{lw}_i} \\ &\quad - \frac{1}{2} (\mathbf{h}_i - \boldsymbol{\alpha}_{hi})^{'} \mathbf{T}_{hi}^{'} \mathbf{\Sigma}_{hi}^{-1} \mathbf{T}_{hi} (\mathbf{h}_i - \boldsymbol{\alpha}_{hi}) - \frac{1}{2} \mathbf{1}_T^{'} \mathbf{h}_i - \frac{1}{2} \mathbf{1}_T^{'} \mathbf{lw}_i \\ &\quad - \frac{1}{2} \text{tr} \Big(diag(e^{-\mathbf{h}_i} e^{-\mathbf{lw}_i}) (\mathbf{Z}_i \mathbf{K}_{\theta,i}^{-1} \mathbf{Z}_i^{'} + (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}_{0,i} - \mathbf{Z}_i \widehat{\boldsymbol{\theta}}_i)(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}_{0,i} - \mathbf{Z}_i \widehat{\boldsymbol{\theta}}_i)^{'}) \Big) \end{aligned}$$

where $\widehat{\boldsymbol{\theta}}_i = (\mathbf{T}_i^{'}\mathbf{T}_i + \mathbf{Z}_i^{'}\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}\mathbf{Z}_i)^{-1}(\mathbf{Z}_i^{'}\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0}))$ and $\mathbf{K}_{\theta,i}^{-1} = (\mathbf{T}_i^{'}\mathbf{T}_i + \mathbf{Z}_i^{'}\mathbf{W}_i^{-1}\mathbf{H}_i^{-1}\mathbf{Z}_i)^{-1}$ are the mean and variance of the conditional posterior distribution of $\widetilde{\boldsymbol{\theta}}_i$ given $(\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i)$;

for details, see the Online Appendix; $\mathbf{T}_{hi} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{pmatrix}$, $\boldsymbol{\alpha}_{hi} = \mathbf{T}_{hi}^{-1}\widetilde{\boldsymbol{\alpha}}_{hi}$, $\widetilde{\boldsymbol{\alpha}}_{hi} = (a_{h_0,i}, 0, \ldots, 0)'$ and $\boldsymbol{\Sigma}_{hi} = diag(\sigma_{h,i}^2 + V_{h_0,i}, \sigma_{h,i}^2, \ldots, \sigma_{h,i}^2)$.

In the M-step, we maximize the function $\mathbf{Q}(\mathbf{h}_i, \mathbf{lw}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i)$ with respect to $(\mathbf{h}_i, \mathbf{lw}_i)$ using the Newton–Raphson method. Chan and Eisenstat (2018) derive the gradient and Hessian for $\mathbf{h}_i$; the gradient of $\mathbf{lw}_i$ can also be obtained similarly,

$$\begin{aligned}
\mathbf{g}_{\mathbf{h}_i} &= \frac{\partial \mathbf{Q}(\mathbf{h}_i, \mathbf{lw}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i)}{\partial \mathbf{h}_i} = -\mathbf{T}_{hi}'\boldsymbol{\Sigma}_{hi}^{-1}\mathbf{T}_{hi}(\mathbf{h}_i - \boldsymbol{\alpha}_{hi}) - \frac{1}{2}(\mathbf{1}_T - e^{-\mathbf{h}_i}e^{-\mathbf{lw}_i} \odot \widehat{\mathbf{z}}_i), \\
\mathbf{g}_{\mathbf{w}_i} &= \frac{\partial \mathbf{Q}(\mathbf{h}_i, \mathbf{lw}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i)}{\partial \mathbf{lw}_i} = -\frac{\nu_i}{2}\mathbf{1}_T + \frac{\nu_i}{2}e^{-\mathbf{lw}_i} - \frac{1}{2}(\mathbf{1}_T - e^{-\mathbf{h}_i}e^{-\mathbf{lw}_i} \odot \widehat{\mathbf{z}}_i),
\end{aligned}$$

where $\widehat{\mathbf{z}}_i$ is the diagonal of $\mathbf{Z}_i\mathbf{K}_{\theta,i}^{-1}\mathbf{Z}_i' + (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)'$ and $\odot$ is the component-wise vector multiplication. After obtaining the mode $(\widehat{\mathbf{h}}_i, \widehat{\mathbf{lw}}_i)$ of $p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{Y}_i, \boldsymbol{\Theta}_1)$, the negative Hessian at the mode is calculated as,

$$\begin{aligned}
\mathbf{K}_{hi} &= -\frac{\partial^2 \log p(\mathbf{h}_i, \mathbf{w}_i|\mathbf{Y}_i, \boldsymbol{\Theta}_1)}{\partial \mathbf{h}_i^2}\bigg|_{(\widehat{\mathbf{h}}_i, \widehat{\mathbf{lw}}_i)} \\
&= -\left[\frac{\partial^2 \mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i}\left(\log p(\mathbf{h}_i, \mathbf{w}_i, \widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \boldsymbol{\Theta}_1)\right)}{\partial \mathbf{h}_i^2} - \frac{\partial^2 \mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\widetilde{\mathbf{h}}_i, \widetilde{\mathbf{lw}}_i}\left(\log p(\widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \boldsymbol{\Theta}_1, \mathbf{h}_i, \mathbf{w}_i)\right)}{\partial \mathbf{h}_i^2}\right]\bigg|_{(\widehat{\mathbf{h}}_i, \widehat{\mathbf{lw}}_i)} \\
&= \left(\mathbf{T}_{hi}'\boldsymbol{\Sigma}_{hi}^{-1}\mathbf{T}_{hi} + \frac{1}{2}diag(e^{-\widehat{\mathbf{h}}_i}e^{-\widehat{\mathbf{lw}}_i} \odot \widehat{\mathbf{z}}_i)\right) + \frac{1}{2}\mathbf{V}_i' \odot (\mathbf{I}_T - \mathbf{V}_i),
\end{aligned}$$

where $\mathbf{V}_i = diag(e^{-\widehat{\mathbf{h}}_i}e^{-\widehat{\mathbf{lw}}_i})\mathbf{Z}_i\mathbf{K}_{\theta,i}^{-1}\mathbf{Z}_i'$; see the derivation in Chan and Eisenstat (2018). Conditional from a sample $\mathbf{h}_i \sim \mathcal{N}(\widehat{\mathbf{h}}_i, \mathbf{K}_{hi}^{-1})$, the proposal distribution $\mathbf{w}_i$ would be

$$\begin{aligned}
\log f(\mathbf{w}_i|\mathbf{h}_i) \propto& \mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\mathbf{h}_i, \widetilde{\mathbf{lw}}_i}\left(\log p(\mathbf{w}_i, \widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i, \boldsymbol{\Theta}_1, \mathbf{h}_i)\right), \\
\propto& -\mathbf{1}_T'\left(\frac{\nu_i + 1}{2} + 1\right)\mathbf{lw}_i - \mathbf{1}_T'\frac{\nu_i}{2}e^{-\mathbf{lw}_i} \\
&- \frac{1}{2}\mathrm{tr}\left(diag(e^{-\mathbf{h}_i}e^{-\mathbf{lw}_i})\mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\mathbf{h}_i, \widetilde{\mathbf{lw}}_i}\left(\mathbf{Z}_i\mathbf{K}_{\theta,i}^{-1}\mathbf{Z}_i' + (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)'\right)\right).
\end{aligned}$$

The calculation of $\log f(\mathbf{w}_i|\mathbf{h}_i)$ requires to carry out another run of the EM algorithm for the expectation $\mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\mathbf{h}_i, \widetilde{\mathbf{lw}}_i}\left(\mathbf{Z}_i\mathbf{K}_{\theta,i}^{-1}\mathbf{Z}_i' + (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_{0,i} - \mathbf{Z}_i\widehat{\boldsymbol{\theta}}_i)'\right)$ and it is quite

computationally expensive. Therefore, we propose a simplification of the calculation that

$$\begin{aligned}\log f(\mathbf{w}_i|\mathbf{h}_i) \propto & \mathbf{E}_{\widetilde{\boldsymbol{\theta}}_i|\mathbf{h}_i,\widehat{\mathbf{lw}}_i}\left(\log p(\mathbf{w}_i,\widetilde{\boldsymbol{\theta}}_i|\mathbf{Y}_i,\boldsymbol{\Theta}_1,\mathbf{h}_i)\right) \\ \propto & -\mathbf{1}_T^{'}\left(\frac{\nu_i+1}{2}+1\right)\mathbf{lw}_i-\mathbf{1}_T^{'}\left(\frac{\nu_i}{2}+\frac{1}{2}e^{-\mathbf{h}_i}\odot\widetilde{\mathbf{z}}_i\right)e^{-\mathbf{lw}_i},\end{aligned}$$

with $\widetilde{\mathbf{z}}_i$ is $\widehat{\mathbf{z}}_i$ evaluated at $\mathbf{h}_i$, hence $w_{it}|h_{it} \sim \mathcal{IG}(a_{it},b_{it})$ where $a_{it} = \frac{\nu_i+1}{2}$ and $b_{it} = \left(\frac{\nu_i}{2}+\frac{1}{2}e^{-h_{it}}\widetilde{z}_{it}\right)$.

References

Chan, J. C. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, 8(S1):S19–S40.

Hörmann, W. and Leydold, J. (2014). Generating generalized inverse Gaussian random variates. *Statistics and Computing*, 24(4):547–557.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis*, 76:408–423.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Yu, Y. and Meng, X.-L. (2011). To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570.