

Elsevier required licence: © 2018

This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The definitive publisher version is available online at

[10.1016/j.econlet.2018.06.031](https://doi.org/10.1016/j.econlet.2018.06.031)

Comparing Hybrid Time-Varying Parameter VARs

Joshua C.C. Chan

Economics Discipline Group,
University of Technology Sydney

Eric Eisenstat

School of Economics,
University of Queensland

June 2018

Abstract

Empirical questions such as whether the Phillips curve or the Okun's law is stable can often be framed as a model comparison—e.g., comparing a vector autoregression (VAR) in which the coefficients in one equation are constant versus one that has time-varying parameters. We develop Bayesian model comparison methods to compare a class of time-varying parameter VARs we call hybrid TVP-VARs—VARs with time-varying parameters in some equations but constant coefficients in others. Using US data, we find evidence that the VAR coefficients in some, but not all, equations are time varying. Our finding highlights the empirical relevance of these hybrid TVP-VARs.

Keywords: state space, marginal likelihood, Bayesian model comparison

JEL classifications: C11, C52, E32, E52

1 Introduction

Time-varying parameter vector autoregressions (TVP-VARs) developed by Cogley and Sargent (2001, 2005) and Primiceri (2005) have become workhorse models for analyzing the evolving inter-relationships between multiple macroeconomic variables. Models with time-varying parameters and stochastic volatility are also found to forecast better than homoskedastic, constant-coefficient models, as demonstrated in many recent papers.¹ However, there is an emerging literature that has expressed concerns about the potential over-parameterization of these TVP-VARs,² and it remains unclear if all forms of time variation are needed.

In Chan and Eisenstat (2018) we take a first step in addressing this model specification issue by developing model comparison techniques for these TVP-VARs. Using a US dataset, we find overwhelming evidence in favor of the model of Primiceri (2005) compared to a conventional constant-coefficient VAR. But most of the gains appear to have come from allowing for stochastic volatility rather than time variation in the VAR coefficients. However, our earlier work does not provide a more nuanced understanding of the role of time-varying coefficients. For instance, empirical researchers are often only interested in assessing the stability of one or a few equations in a VAR system. For example, Karlsson and Österholm (2018) address the question of whether the US Phillips curve is stable, which amounts to testing if the inflation equation in a VAR has time-varying coefficients—and the hypothesis is silent about the coefficients in other equations.

In view of this, we consider a class of models we call hybrid TVP-VARs—i.e., VARs that allow some equations to have time-varying parameters while coefficients in other equations remain constant. In terms of methodological contribution, we develop an improved version of the importance sampling estimator for marginal likelihood in Chan and Eisenstat (2018). The key improvement is dimension reduction: instead of a very high-dimensional importance sampling estimator for the joint VAR system, we use an equivalent, recursive representation to obtain a sequence of lower-dimensional importance sampling estimators, the product of which gives the same marginal likelihood of the model. By vastly reducing the dimension of the importance sampling, the improved version can substantially reduce the computational time and allow us to potentially scale up to larger systems.

¹Recent examples include Clark (2011), D’Agostino, Gambetti, and Giannone (2013), Koop and Korobilis (2013), Clark and Ravazzolo (2014) and Cross and Poon (2016).

²See, e.g., Chan, Koop, Leon-Gonzalez, and Strachan (2012), Nakajima and West (2013) and Belmonte, Koop, and Korobilis (2014).

On the empirical side, we confirm our earlier finding that the data favor a constant-coefficient VAR with stochastic volatility relative to the more flexible model of Primiceri (2005) that allows for time-varying coefficients. However, we find strong evidence that the VAR coefficients in some, but not all, equations are time varying, highlighting the empirical relevance of the hybrid TVP-VARs.

2 Hybrid TVP-VARs with Stochastic Volatility

We outline in this section the class of time-varying parameter models we wish to compare. Let $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})'$ be an $n \times 1$ vector of endogenous variables at time t . First consider the following standard TVP-VAR with stochastic volatility:

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{b}_t + \mathbf{B}_{1,t} \mathbf{y}_{t-1} + \dots + \mathbf{B}_{p,t} \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (1)$$

where \mathbf{b}_t is an $n \times 1$ vector of time-varying intercepts, $\mathbf{B}_{1,t}, \dots, \mathbf{B}_{p,t}$ are $n \times n$ VAR coefficient matrices, \mathbf{A}_t is an $n \times n$ lower triangular matrix with ones on the diagonal and $\boldsymbol{\Sigma}_t = \text{diag}(\exp(h_{1,t}), \dots, \exp(h_{n,t}))$. The law of motion of each of the log-volatilities $\mathbf{h}_t = (h_{1,t}, \dots, h_{n,t})'$ is specified as an independent random walk:

$$h_{i,t} = h_{i,t-1} + \zeta_{i,t}, \quad \zeta_{i,t} \sim \mathcal{N}(0, \sigma_{i,h}^2), \quad (2)$$

where the initial conditions $h_{i,0}$ are treated as parameters to be estimated.

The system in (1) is written in the structural form and the covariance matrix $\boldsymbol{\Sigma}_t$ by construction is diagonal. Hence, we can estimate this recursive system equation by equation. A similar representation is used in Carriero, Clark, and Marcellino (2016) to estimate a large VAR with a standard stochastic volatility specification. To estimate the model equation by equation, we first introduce some notations. Let $b_{i,t}$ denote the i -th element of \mathbf{b}_t and $\mathbf{B}_{i,j,t}$ the i -th row of $\mathbf{B}_{j,t}$. Then, $\boldsymbol{\beta}_{i,t} = (b_{i,t}, \mathbf{B}_{i,1,t}, \dots, \mathbf{B}_{i,p,t})'$ is the intercept and VAR coefficients for the i -th equation. Further, let $\boldsymbol{\alpha}_{i,t}$ denote the free elements in the i -th row of the impact matrix \mathbf{A}_t , i.e., $\boldsymbol{\alpha}_{i,t} = (A_{i1,t}, \dots, A_{i(i-1),t})'$. Then, the i -th equation of the system in (1) can be rewritten as:

$$y_{i,t} = \mathbf{w}_{i,t} \boldsymbol{\alpha}_{i,t} + \tilde{\mathbf{x}}_t \boldsymbol{\beta}_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, e^{h_{i,t}}),$$

where $\mathbf{w}_{i,t} = (-y_{1,t}, \dots, -y_{i-1,t})$ and $\tilde{\mathbf{x}}_t = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$.³

Letting $\mathbf{x}_{i,t} = (\mathbf{w}_{i,t}, \tilde{\mathbf{x}}_t)$, we can further simplify the i -th equation as:

$$y_{i,t} = \mathbf{x}_{i,t} \boldsymbol{\theta}_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, e^{h_{i,t}}), \quad (3)$$

where $\boldsymbol{\theta}_{i,t} = (\boldsymbol{\alpha}'_{i,t}, \boldsymbol{\beta}'_{i,t})'$ is of dimension $k_{\theta_i} = (i-1) + (np+1) = np+i$. Finally, the coefficients evolve as a random walk:

$$\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\eta}_{i,t}, \quad \boldsymbol{\eta}_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\theta_i}), \quad (4)$$

where the initial conditions $\boldsymbol{\theta}_{i,0}$ are treated as unknown parameters.

The standard TVP-VAR specifies each equation to be time varying as in (3). In contrast, a hybrid TVP-VAR allows some equations to have time-varying coefficients while others have time-invariant coefficients.⁴ More specifically, each hybrid TVP-VAR of n variables is defined by a vector of binary variables $\mathbf{M} = (M_1, \dots, M_n) \in \{0, 1\}^n$. For $i = 1, \dots, n$, if $M_i = 1$, then the i -th equation of the system is specified by (3)–(4); otherwise the i -th equation has constant coefficients

$$y_{i,t} = \mathbf{x}_{i,t} \boldsymbol{\gamma}_i + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, e^{h_{i,t}}).$$

3 Importance Sampling for Marginal Likelihood

This section presents an improved version of the importance sampling estimator for marginal likelihood proposed in Chan and Eisenstat (2018). The key improvement is dimension reduction: instead of a very high-dimensional importance sampling estimator for the joint system (1), we use the equivalent representation in (3) to develop a sequence of lower-dimensional importance sampling estimators that gives the same marginal likelihood estimator. By reducing the dimension of the importance sampling, the improved approach version can substantially reduce the computational time and allow us to potentially scale up to larger systems.

³Note that $y_{i,t}$ depends not only on lagged values of all variables, but it also depends on contemporaneous variables $y_{1,t}, \dots, y_{i-1,t}$. However, the system is recursive and the density function remains the same form when we perform the change of variables from ε_t to \mathbf{y}_t —in fact, it can be shown that determinant of the Jacobian is one.

⁴In all models we include stochastic volatility as it is found to be empirically important; see, e.g., Clark (2011) and Chan and Eisenstat (2018).

Below we outline the new importance sampling estimator. To fix ideas, we focus on the TVP-VAR case; hybrid TVP-VARs can be handled similarly. For notational convenience, stack $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,T})'$, $\boldsymbol{\theta}_i = (\boldsymbol{\theta}'_{i,1}, \dots, \boldsymbol{\theta}'_{i,T})'$ and $\mathbf{h}_i = (h_{i,1}, \dots, h_{i,T})'$ over $t = 1, \dots, T$, and define $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^n$, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}_{i=1}^n$ and $\mathbf{h} = \{\mathbf{h}_i\}_{i=1}^n$. Moreover, the time-invariant parameters of the model are $\boldsymbol{\Sigma}_\theta = \{\boldsymbol{\Sigma}_{\theta_i}\}_{i=1}^n$, $\boldsymbol{\Sigma}_h = \{\sigma_{i,h}^2\}_{i=1}^n$, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{1,0}, \dots, \boldsymbol{\theta}'_{n,0})'$ and $\mathbf{h}_0 = (h_{1,0}, \dots, h_{n,0})'$. Then, the *marginal likelihood* of the model is defined as the integral

$$p(\mathbf{y}) = \int p(\mathbf{y} | \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) p(\boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) d(\boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0),$$

where $p(\boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0)$ is the prior density and

$$p(\mathbf{y} | \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) = \int p(\mathbf{y} | \boldsymbol{\theta}, \mathbf{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) p(\boldsymbol{\theta}, \mathbf{h} | \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) d(\boldsymbol{\theta}, \mathbf{h})$$

is the *integrated likelihood*, i.e., the marginal density of the data unconditional of the latent states $\boldsymbol{\theta}$ and \mathbf{h} . The major challenge is to evaluate this integrated likelihood as the integration is very high dimensional. Chan and Eisenstat (2018) propose an importance sampling estimator for that purpose. Here we improve upon the earlier algorithm by dividing this high dimensional integration into n lower dimensional integrations.

To that end, first note that the *conditional likelihood* given the latent states $\boldsymbol{\theta}$ and \mathbf{h} can be decomposed as

$$p(\mathbf{y} | \boldsymbol{\theta}, \mathbf{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) = p(\mathbf{y} | \boldsymbol{\theta}, \mathbf{h}) = \prod_{i=1}^n p(\mathbf{y}_i | \boldsymbol{\theta}_i, \mathbf{h}_i),$$

where $p(\mathbf{y}_i | \boldsymbol{\theta}_i, \mathbf{h}_i)$ is the product of T univariate Gaussian densities implied by (3). Furthermore, recall that $h_{i,t}$ and $\boldsymbol{\theta}_{i,t}$ evolve according to independent random walks given in (2) and (4), respectively. Therefore, their joint density can be decomposed as:

$$p(\boldsymbol{\theta}, \mathbf{h} | \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) = \prod_{i=1}^n p(\boldsymbol{\theta}_i | \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}).$$

Consequently, the integrated likelihood can be written as

$$\begin{aligned} p(\mathbf{y} | \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \mathbf{h}_0) &= \prod_{i=1}^n \int p(\mathbf{y}_i | \boldsymbol{\theta}_i, \mathbf{h}_i) p(\boldsymbol{\theta}_i | \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}) d(\boldsymbol{\theta}_i, \mathbf{h}_i) \\ &= \prod_{i=1}^n \int p(\mathbf{y}_i | \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}) d\mathbf{h}_i. \end{aligned} \quad (5)$$

Both terms in the integrand in (5) have an analytical expression. The first term is the density of the data marginal of $\boldsymbol{\theta}_i$, and the second term is the prior density of \mathbf{h}_i implied by (2); their closed-form expressions are given in Appendix A. Next, we adapt the approach in Chan and Eisenstat (2018) to estimate the integrated likelihood using (5).

More specifically, each of the n integrations in (5) can be estimated by importance sampling. In this case, the ideal zero-variance importance sampling density is given by

$$p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\theta_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0}) \propto p(\mathbf{y}_i | \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0})p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}).$$

However, this density cannot be used directly as the importance sampling density because its normalization constant is unknown. To proceed, we approximate this ideal zero-variance importance sampling density using a Gaussian approximation. In particular, its mean vector is set to be the mode of $p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\theta_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0})$ and its precision matrix—i.e., the inverse of the covariance matrix—is set to be the negative Hessian of $\log p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\theta_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0})$ evaluated at the mode. Independent samples from this Gaussian importance sampling density can be obtained efficiently using the algorithm in Chan and Jeliazkov (2009). Finally, the integrated likelihood is simply the product of the estimates of these n integrals.

Once we can evaluate the integrated likelihood, we use the cross-entropy method of Chan and Eisenstat (2015) to further integrate out the time-invariant parameters $\boldsymbol{\Sigma}_{\theta}$, $\boldsymbol{\Sigma}_h$, $\boldsymbol{\theta}_0$ and \mathbf{h}_0 . The cross-entropy method is an adaptive importance sampling algorithm that uses the posterior draws of the time-invariant parameters to construct the optimal importance sampling density. This step is relatively easier as the dimension of the time-invariant parameters is much lower. We refer the readers to Chan and Eisenstat (2015) for a more detailed discussion.

4 Application: Are VAR Coefficients Time Varying?

In this empirical application we compare a number of hybrid TVP-VARs with different combinations of time-varying parameters and constant coefficients. We use a dataset that consists of US quarterly observations on the GDP deflator, real GDP, and short-term interest rate from 1954Q3 to 2017Q4, sourced from the Federal Reserve Bank of St. Louis economic database. The GDP deflator and real GDP are transformed to annualized growth rates. The short-term interest rate is the effective Federal Funds rate

and is not transformed. Following standard practice, we set the lag length to be $p = 2$. We compute the log marginal likelihoods of the hybrid TVP-VARs using the importance sampling algorithm described in Section 3. Each log marginal likelihood estimate is based on 20000 evaluations of the integrated likelihood. The importance sampling density for the time-invariant parameters is constructed using the cross-entropy method with 20000 posterior draws after a burn-in period of 5000.

Before presenting the marginal likelihood estimates, we compare the speed of the new algorithm with that of the original one in Chan and Eisenstat (2018) for computing the marginal likelihoods of three models: TVP-VARs with and without stochastic volatility (respectively TVP-SV and TVP) and a constant-coefficient VAR with stochastic volatility (VAR-SV). The algorithms are implemented using MATLAB on a standard desktop with an Intel Core i5-4590S @3.0 GHz processor and 8 GB of RAM. The computation times are reported in Table 1. As shown in the table, the improved algorithm can reduce the computation times by about 35%-65%.

Table 1: Computation times of marginal likelihoods of TVP-SV, TVP and VAR-SV (in minutes).

	TVP-SV	TVP	VAR-SV
Chan and Eisenstat (2018)	568	5.9	1.4
new approach	206	2.9	0.9

There are three variables in the VAR and in each equation the coefficients can either be constant or time varying. Hence, there are in total eight hybrid TVP-VARs, including the full-fledged TVP-VAR of Primiceri (2005). The marginal likelihood estimates are reported in Table 2. Consistent with the model comparison results in Chan and Eisenstat (2018), the model with constant VAR coefficients and stochastic volatility is slightly preferred by the data relative to the more flexible TVP-VAR of Primiceri (2005). Since the marginal likelihood has a built-in penalty for model complexity, a more flexible model will have a higher marginal likelihood value only if the increase in model fit is substantial relative to the increase in model complexity. Here we have an example where a more flexible model has a lower marginal likelihood value.

In addition, we find evidence that some of the VAR coefficients are time varying. In particular, the best model is the one that specifies constant coefficients in the interest rate equation while allowing time variation in both the GDP deflator and real GDP equations. The Bayes factor in favor of this model relative to the constant coefficient

VAR with stochastic volatility is 9.8×10^6 , showing overwhelming evidence in favor of the former. Moreover, models with time-varying parameters in the interest rate equation have consistently lower marginal likelihood values compared to their counterparts with constant coefficients in interest rate equation.⁵ This finding might be surprising given that several authors have found that the late 1970s and early 1980s were characterized by a clear change in the monetary policy regime (Clarida, Gal’i, and Gertler, 2000; Lubik and Schorfheide, 2004; Benati and Surico, 2009; Castelnuovo and Surico, 2010; Castelnuovo and Fanelli, 2015). One possible reason could be that a one-time change in the monetary policy regime over a long sample period is not well approximated by drifting parameters. To investigate this possibility, we redid the model comparison using a shortened sample from 1970Q1 to 1999Q4. With this shorter sample, we find evidence that the coefficients in the interest rate equation are time varying (detailed results are reported in Appendix B).

Table 2: Log marginal likelihood estimates of the hybrid TVP-VARs and the corresponding numerical standard errors.

GDP deflator eq.	real GDP eq.	interest rate eq.	log ML	NSE
constant	constant	constant	-1227.1	0.21
constant	constant	time varying	-1243.7	0.32
constant	time varying	constant	-1218.9	0.11
constant	time varying	time varying	-1235.9	0.10
time varying	constant	constant	-1219.2	0.15
time varying	constant	time varying	-1236.2	0.18
time varying	time varying	constant	-1211.0	0.08
time varying	time varying	time varying	-1228.0	0.13

Overall, our model comparison exercise suggests that if we only compare the constant coefficient VAR with the TVP-VAR, we might conclude that the VAR coefficients have not changed over the sample period. This conclusion, however, overlooks the possibility that only a subset of the VAR coefficients have changed over time. Our results therefore highlight the empirical relevance of the hybrid TVP-VARs.

⁵Our sample covers the periods when the Fed Funds rate effectively hit the zero lower bound. For robustness check, we redid the model comparison exercise by replacing the Fed Funds rate with the shadow rate constructed by Wu and Xia (2016). The results are similar and are reported in Appendix B. In particular, the data favor models where the interest rate equation has constant coefficients. We thank an anonymous referee for this suggestion.

5 Concluding Remarks and Future Research

With the aim of expanding the toolkit of empirical researchers, we considered a class of TVP-VARs in which the VAR coefficients are time varying in some equations but constant in others. We then adapted the importance sampling estimator for marginal likelihood in Chan and Eisenstat (2018) to this class of hybrid TVP-VARs. Using a US dataset, we confirmed our earlier finding that the data prefer a constant-coefficient VAR with stochastic volatility compared to the full-fledged TVP-VAR of Primiceri (2005). However, we found evidence that only some VAR coefficients are time varying, while others are constant.

In future work, it would be interesting to extend the methods to compare large TVP-VARs. In particular, it would be helpful to compare different shrinkage priors and different ways to model stochastic volatility in high-dimensional settings. Such a model comparison exercise would provide useful guidelines for empirical researchers.

Appendix A: Computational Details

In this appendix we provide the technical details for evaluating the integrated likelihood of the TVP-VAR with stochastic volatility. Using the decomposition in (5), it suffices to consider the evaluation of

$$\int p(\mathbf{y}_i | \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}) d\mathbf{h}_i. \quad (6)$$

Below we give explicit expressions for the two terms in the integral and the exposition follows closely Chan and Eisenstat (2018). Define $\boldsymbol{\mu}_{\theta_i} = \mathbf{1}_T \otimes \boldsymbol{\theta}_{i,0}$, $\mathbf{S}_{\mathbf{y}_i} = \text{diag}(\mathbf{h}_i)$ and $\mathbf{S}_{\theta_i} = \mathbf{I}_T \otimes \boldsymbol{\Sigma}_{\theta_i}$, where $\mathbf{1}_T$ is a $T \times 1$ column of ones. Furthermore, let \mathbf{H}_{θ_i} denote the first difference matrix of dimension k_{θ_i} , i.e.,

$$\mathbf{H}_{\theta_i} = \begin{pmatrix} \mathbf{I}_{k_{\theta_i}} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_{k_{\theta_i}} & \mathbf{I}_{k_{\theta_i}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & -\mathbf{I}_{k_{\theta_i}} & \mathbf{I}_{k_{\theta_i}} \end{pmatrix}.$$

Then, using a similar derivation in Chan and Grant (2016), one can obtain the first term in the integral (in log) as follows:

$$\begin{aligned} \log p(\mathbf{y}_i | \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) &= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \mathbf{1}'_T \mathbf{h}_i - \frac{T}{2} \log |\boldsymbol{\Sigma}_{\theta_i}| - \frac{1}{2} \log |\mathbf{K}_{\theta_i}| \\ &\quad - \frac{1}{2} \left(\mathbf{y}'_i \mathbf{S}_{\mathbf{y}_i}^{-1} \mathbf{y}_i + \boldsymbol{\mu}'_{\theta_i} \mathbf{H}'_{\theta_i} \mathbf{S}_{\theta_i}^{-1} \mathbf{H}_{\theta_i} \boldsymbol{\mu}_{\theta_i} - \widehat{\boldsymbol{\theta}}'_i \mathbf{K}_{\theta_i} \widehat{\boldsymbol{\theta}}_i \right), \end{aligned} \quad (7)$$

where

$$\mathbf{K}_{\theta_i} = \mathbf{H}'_{\theta_i} \mathbf{S}_{\theta_i}^{-1} \mathbf{H}_{\theta_i} + \mathbf{X}'_i \mathbf{S}_{\mathbf{y}_i}^{-1} \mathbf{X}_i, \quad \widehat{\boldsymbol{\theta}}_i = \mathbf{K}_{\theta_i}^{-1} (\mathbf{H}'_{\theta_i} \mathbf{S}_{\theta_i}^{-1} \mathbf{H}_{\theta_i} \boldsymbol{\mu}_{\theta_i} + \mathbf{X}'_i \mathbf{S}_{\mathbf{y}_i}^{-1} \mathbf{y}_i). \quad (8)$$

The second term in the integral in (6) is in fact a Gaussian density:

$$\log p(\mathbf{h}_i | \sigma_{i,h}^2, h_{i,0}) = -\frac{T}{2} \log(2\pi\sigma_{i,h}^2) - \frac{1}{2\sigma_{i,h}^2} (\mathbf{h}_i - \boldsymbol{\mu}_{h_i})' \mathbf{H}'_h \mathbf{H}_h (\mathbf{h}_i - \boldsymbol{\mu}_{h_i}),$$

where $\boldsymbol{\mu}_{h_i} = \mathbf{1}_T \otimes h_{i,0}$ and \mathbf{H}_h is the first difference matrix of dimension T .

Given the analytical expressions of the integrand in (6), we next integrate out \mathbf{h}_i using importance sampling. To obtain a good importance sampling density, we use the EM algorithm to find the maximum of the marginal density $p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\theta_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0})$, which

is the ideal zero-variance importance sampling density (but it cannot be used directly because its normalization constant is unknown).

To implement the E-step, we compute the following conditional expectation

$$\mathcal{Q}(\mathbf{h}_i | \tilde{\mathbf{h}}_i) = \mathbb{E}_{\boldsymbol{\theta}_i | \tilde{\mathbf{h}}_i} [\log p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0})],$$

where the expectation is taken with respect to $p(\boldsymbol{\theta}_i | \mathbf{y}_i, \tilde{\mathbf{h}}_i, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_i}, \boldsymbol{\theta}_{0,i})$ for an arbitrary vector $\tilde{\mathbf{h}}_i$. Moreover, this conditional distribution is Gaussian:

$$(\boldsymbol{\theta}_i | \mathbf{y}_i, \tilde{\mathbf{h}}_i, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_i}, \boldsymbol{\theta}_{0,i}) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_i, \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}),$$

where $\hat{\boldsymbol{\theta}}_i$ and $\mathbf{K}_{\boldsymbol{\theta}_i}$ are given in (8). We emphasize that both the mean vector $\hat{\boldsymbol{\theta}}_i$ and precision matrix $\mathbf{K}_{\boldsymbol{\theta}_i}$ are functions of $\tilde{\mathbf{h}}$, and they are computed using $\tilde{\mathbf{h}}_i$. It can be shown that $\mathcal{Q}(\mathbf{h} | \tilde{\mathbf{h}})$ has the following explicit expression:

$$\begin{aligned} \mathcal{Q}(\mathbf{h} | \tilde{\mathbf{h}}) = & -\frac{1}{2\sigma_{i,h}^2}(\mathbf{h}_i - \boldsymbol{\mu}_{h_i})' \mathbf{H}'_h \mathbf{H}_h (\mathbf{h}_i - \boldsymbol{\mu}_{h_i}) - \frac{1}{2} \mathbf{1}'_T \mathbf{h}_i \\ & - \frac{1}{2} \text{tr} \left(\text{diag}(e^{-\mathbf{h}_i}) \left(\mathbf{X}_i \mathbf{K}_{\boldsymbol{\theta}_i}^{-1} \mathbf{X}'_i + (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i)(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i)' \right) \right) + c_1, \end{aligned} \quad (9)$$

where $\text{tr}(\cdot)$ is the trace operator and c_1 is a constant not dependent on \mathbf{h}_i .

In the M-step, we maximize the function $\mathcal{Q}(\mathbf{h}_i | \tilde{\mathbf{h}}_i)$ with respect to \mathbf{h}_i . This is done using the Newton-Raphson method with gradient

$$\mathbf{g}_{\mathcal{Q}} = -\frac{1}{\sigma_{i,h}^2} \mathbf{H}'_h \mathbf{H}_h (\mathbf{h}_i - \boldsymbol{\mu}_{h_i}) - \frac{1}{2} (\mathbf{1}_T - e^{-\mathbf{h}_i} \odot \hat{\mathbf{z}}_i),$$

and Hessian

$$\mathbf{H}_{\mathcal{Q}} = -\frac{1}{\sigma_{i,h}^2} \mathbf{H}'_h \mathbf{H}_h - \frac{1}{2} \text{diag} (e^{-\mathbf{h}_i} \odot \hat{\mathbf{z}}_i), \quad (10)$$

where \odot denotes the entry-wise product, $\hat{\mathbf{z}}_i = (s_{i,1}^2 + \hat{\varepsilon}_{i,1}^2, \dots, s_{i,T}^2 + \hat{\varepsilon}_{i,T}^2)'$, $s_{i,t}^2$ is the t -th diagonal element of $\mathbf{X}_i \mathbf{K}_{\boldsymbol{\theta}_i}^{-1} \mathbf{X}'_i$ and $\hat{\varepsilon}_t$ is the t -th element of $\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i$. Since both $\mathbf{g}_{\mathcal{Q}}$ and $\mathbf{H}_{\mathcal{Q}}$ can be computed efficiently using sparse and band matrix algorithms and the Hessian $\mathbf{H}_{\mathcal{Q}}$ is negative definite for all \mathbf{h}_i , the convergence of the Newton-Raphson method is fast.

After obtaining the mode $\hat{\mathbf{h}}_i$ of the ideal importance sampling density, we next compute the Hessian of the log density evaluated at $\hat{\mathbf{h}}_i$. For that purpose we use the identity

$$\log p(\mathbf{h}_i | \mathbf{y}_i, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_i}, \sigma_{i,h}^2, \boldsymbol{\theta}_{i,0}, h_{i,0}) = \mathcal{Q}(\mathbf{h}_i | \mathbf{h}_i) + \mathcal{H}(\mathbf{h}_i | \mathbf{h}_i), \quad (11)$$

where $\mathcal{H}(\mathbf{h}_i | \mathbf{h}_i) = -\mathbb{E}_{\boldsymbol{\theta}_i | \mathbf{h}_i} [\log(\boldsymbol{\theta}_i | \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_i}, \boldsymbol{\theta}_{0,i})]$. Hence, the Hessian of the log marginal density evaluated at $\widehat{\mathbf{h}}_i$ is simply the sum of the Hessians of \mathcal{Q} and \mathcal{H} with $\mathbf{h}_i = \widehat{\mathbf{h}}_i$. The former comes out as a by-product of the EM algorithm; an analytical expression is given in (10). The latter has the following explicit expression:

$$\mathbf{H}_{\mathcal{H}} = -\frac{1}{2} \mathbf{Z}' \odot (\mathbf{I}_T - \mathbf{Z}_i),$$

where $\mathbf{Z}_i = \text{diag}(e^{-\mathbf{h}_i}) \mathbf{X}_i \mathbf{K}_{\boldsymbol{\theta}_i}^{-1} \mathbf{X}_i'$. Finally, let $\mathbf{H}_{\mathcal{Q}}$ denote the Hessian of $\mathcal{Q}(\mathbf{h} | \mathbf{h})$ evaluated at $\mathbf{h}_i = \widehat{\mathbf{h}}_i$. Then, the negative Hessian of the log marginal density of \mathbf{h}_i evaluated at $\mathbf{h}_i = \widehat{\mathbf{h}}_i$ is simply $\mathbf{K}_h = -(\mathbf{H}_{\mathcal{Q}} + \mathbf{H}_{\mathcal{H}})$, which is used as the precision matrix of the Gaussian approximation.

Appendix B: Additional Results

In this appendix we report additional model comparison results. Table 3 presents the log marginal likelihood estimates of hybrid TVP-VARs using a dataset in which we replace the Fed Funds rate with the shadow rate constructed by Wu and Xia (2016).

Table 3: Log marginal likelihood estimates of the hybrid TVP-VARs and the corresponding numerical standard errors; replacing Fed Funds Rate with the shadow rate in Wu and Xia (2016).

GDP deflator eq.	real GDP eq.	interest rate eq.	log ML	NSE
constant	constant	constant	-1258.5	0.91
constant	constant	time varying	-1271.7	0.39
constant	time varying	constant	-1250.3	0.54
constant	time varying	time varying	-1266.1	0.40
time varying	constant	constant	-1249.6	0.70
time varying	constant	time varying	-1264.6	0.47
time varying	time varying	constant	-1244.4	0.59
time varying	time varying	time varying	-1259.0	0.47

Table 4 reports the log marginal likelihood estimates of hybrid TVP-VARs using the original dataset (with Fed Funds rate) but the sample period is shortened to 1970Q1-1999Q4.

Table 4: Log marginal likelihood estimates of the hybrid TVP-VARs and the corresponding numerical standard errors; Fed Funds Rate; sample period 1970Q1-1999Q4.

GDP deflator eq.	real GDP eq.	interest rate eq.	log ML	NSE
constant	constant	constant	-688.6	0.24
constant	constant	time varying	-686.3	0.24
constant	time varying	constant	-680.0	0.17
constant	time varying	time varying	-678.0	0.15
time varying	constant	constant	-683.7	0.25
time varying	constant	time varying	-681.2	0.33
time varying	time varying	constant	-675.5	0.13
time varying	time varying	time varying	-673.5	0.06

References

- M. Belmonte, G. Koop, and D. Korobilis. Hierarchical shrinkage in time-varying coefficients models. *Journal of Forecasting*, 33(1):80–94, 2014.
- L. Benati and P. Surico. VAR analysis and the Great Moderation. *American Economic Review*, 99(4):1636–1652, 2009.
- A. Carriero, T. E. Clark, and M. Marcellino. Large vector autoregressions with stochastic volatility and flexible priors. *Working Paper 1617, Federal Reserve Bank of Cleveland*, 2016.
- E. Castelnuovo and L. Fanelli. Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test. *Journal of Applied Econometrics*, 30(6):924–947, 2015.
- E. Castelnuovo and P. Surico. Monetary policy, inflation expectations and the price puzzle. *Economic Journal*, 120(549):1262–1283, 2010.
- J. C. C. Chan and E. Eisenstat. Marginal likelihood estimation with the cross-entropy method. *Econometric Reviews*, 34(3):256–285, 2015.
- J. C. C. Chan and E. Eisenstat. Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of Applied Econometrics*, 2018. Forthcoming.
- J. C. C. Chan and A. L. Grant. Fast computation of the deviance information criterion for latent variable models. *Computational Statistics and Data Analysis*, 100:847–859, 2016.
- J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1):101–120, 2009.
- J. C. C. Chan, G. Koop, R. Leon-Gonzalez, and R. Strachan. Time varying dimension models. *Journal of Business and Economic Statistics*, 30(3):358–367, 2012.
- R. Clarida, J. Gal’i, and M. Gertler. Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, 115(1):147–180, 2000.
- T. E. Clark. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, 29(3):327–341, 2011.
- T. E. Clark and F. Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 2014. Forthcoming.
- T. Cogley and T. J. Sargent. Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual*, 16:331–388, 2001.

- T. Cogley and T. J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005.
- J. Cross and A. Poon. Forecasting structural change and fat-tailed events in Australian macroeconomic variables. *Economic Modelling*, 58:34–51, 2016.
- A. D’Agostino, L. Gambetti, and D. Giannone. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28:82–101, 2013.
- S. Karlsson and P. Österholm. Is the US Phillips curve stable? evidence from Bayesian VARs. Technical report, 2018.
- G. Koop and D. Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013.
- T. Lubik and F. Schorfheide. Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review*, 94(1):190–217, 2004.
- J. Nakajima and M. West. Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics*, 31(2):151–164, 2013.
- G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852, 2005.
- J. C. Wu and F. D. Xia. Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit, and Banking*, 48(2-3):253–291, 2016.