MPRA

Time-Varying Vector Autoregressions: Efficient Estimation, Random Inertia and Random Mean

Romain Legrand

10 September 2018

Time-Varying Vector Autoregressions: Efficient Estimation, Random Inertia and Random Mean

Romain Legrand *

This version: September 2018

Abstract

Time-varying VAR models have become increasingly popular and are now widely used for policy analysis and forecast purposes. They constitute fundamental tools for the anticipation and analysis of economic crises, which represent rapid shifts in dynamic responses and shock volatility. Yet, despite their flexibility, time-varying VARs remain subject to a number of limitations. On the theoretical side, the conventional random walk assumption used for the dynamic parameters appears excessively restrictive. It also conceals the potential heterogeneities existing between the dynamic processes of different variables. On the application side, the standard two-pass procedure building on the Kalman filter proves excessively complicated and suffers from low efficiency.

Based on these considerations, this paper contributes to the literature in four directions:
i) it introduces a general time-varying VAR model which relaxes the standard random walk assumption and defines the dynamic parameters as general auto-regressive processes with variable-specific mean values and autoregressive coefficients.
ii) it develops an estimation procedure for the model which is simple, transparent and efficient. The procedure requires no sophisticated Kalman filtering methods and reduces to a standard Gibbs sampling algorithm.
iii) as an extension, it develops efficient procedures to estimate endogenously the mean values and autoregressive coefficients associated with each variable-specific autoregressive process.
iv) through a case study of the Great Recession for four major economies (Canada, the Euro Area, Japan and the United States), it establishes that forecast accuracy can be significantly improved by using the proposed general time-varying model and its extensions in place of the traditional random walk specification.

---

*ESSEC Business School (visiting researcher). Email: romain.legrand@essec.edu

1   Introduction

VAR models have become the cornerstone of applied macroeconomics. Since the seminal work of Sims (1980), they have been used extensively by financial and economic institutions to perform routine policy analysis and forecasts. While convenient, VAR models with static coefficients and residual variance often turn out to be excessively restrictive in capturing the dynamics of time-series, which often exhibit some form of non-linearity in their behaviours. This motivated the introduction of time-varying coefficients in VAR models (Doan et al. (1983), Canova (1993), Stock and Watson (1996), Cogley (2001), Ciccarelli and Rebucci (2003)), along with stochastic volatility (Harvey et al. (1994), Jacquier et al. (1995), Uhlig (1997), Chib et al. (2006)), and more recently both features (Cogley and Sargent (2005), Primiceri (2005)), in order to account for potential shifts in the transmission mechanism and volatility of the underlying structural shocks.

Since then, time-varying VAR models have become increasingly popular. They have been used for a wide range of applications related to policy analysis, including the evolution of monetary policy over the last decades (Primiceri (2005), Mumtaz and Zanetti (2013), Korobilis (2013)), the response to exchange rate movements (Mumtaz and Sunder-Plassmann (2010), Kavtaradze and Mokhtari (2018), Abbate and Marcellino (2018b)), the impact of fiscal policy (Gerba and Hauzenberger (2013), Eisenstat et al. (2016)), or the general analysis of macroeconomic fluctuations (Gali and Gambetti (2008), Koop and Korobilis (2012)). Time-varying vector autoregressions have also become a benchmark for forecasting as a well-established literature demonstrates that they generally perform better than their static counterparts (Clark (2011), D'Agostino et al. (2011), Aastveit et al. (2017), Abbate and Marcellino (2018a)).

Lately, time-varying VAR models have received much attention regarding the analysis and anticipation of economic crises, in particular the events of the Great Recession. The literature has considered two main classes of explanations for this episode of severe economic decline. The first view focuses on the heteroskedasticity of the exogenous shocks (Stock and Watson (2012), Doh and Connolly (2013), Bijsterbosch and Falagiarda (2014), Gambetti and Musso (2017)). It interprets the Great Recession primarily as an episode of sharp volatility of the structural disturbances affecting the economy. The second view emphasizes the changes in the transmission mechanism (Baumeister and Benati (2010), Benati and Lubik (2014), Ellington et al. (2017) among many others). It considers the Great Recession essentially as a period of altered response of macroeconomic variables to economic policy. In either case, there is strong evidence that modelling time variation adequately is important to the accuracy of both policy analysis and forecasts in a context of crisis. In this respect, it seems plausible that the Great Recession could have been better apprehended with a proper use of time-varying VAR models. Time-varying VARs may also constitute a benchmark tool in the future to predict economic downturns and accurately forecast their evolutions.

Despite the versatility of time-varying VAR models, there remain concerns about their performances. The most widely employed methodology, in line with the seminal contributions of Primiceri (2005) and Del Negro and Primiceri (2015), implies a number of limitations of both theoretical and applied order.

On the theoretical side, the first limitation consists in the choice of a random walk specification for the laws of motion of the different dynamic parameters. This formulation has been widely adopted by the literature, both for the VAR coefficients and the log volatilities of the structural shocks. Though convenient for its simplicity and parsimony, it may be inconsistent with the behaviour of the data. A random walk implies that the range of values taken by the dynamic parameters of the model increases over time and becomes eventually unbounded, resulting in an explosive behaviour in the limit. This is at odd with both empirical observations and economic theory such as the notion of balanced growth path. Most importantly, it is unlikely that such a formulation proves appropriate to describe the short-term fluctuations of economic data. For instance, a well-known feature of the random walk is that it grants equal weight to all past shocks. But if an economy experiences rapid shifts in its dynamics due to a series of large disturbances, as would be the case in a context of crisis, it becomes crucial to capture correctly the effect of the most recent shocks while granting less weight to past shocks. This supposes the use of more general formulations, possibly stationary and mean-reverting, in place of the standard random walk.

The random walk specification is further criticisable as it results de facto in a homogeneity assumption. It indeed implies that all the dynamic parameters follow a similar unit-root process. There is yet no legitimate reason to assume that the dynamic parameters of different variables evolve homogenously. In fact, it is quite likely that different economic variables are characterised by different behaviours of their dynamic coefficients and residual volatilities. Following, the state equations of the parameters should be formulated on a variable-specific basis to account for the potential heterogeneities prevailing from one variable to another.

While switching from a homogeneous random walk specification to a set of variable-specific stationary processes is conceptually trivial, it complicates the estimation procedure. Perhaps for this reason, and even though such alternative formulations have attracted considerable attention in the univariate ARCH literature (Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Jacquier et al. (2004), Eisenstat and Strachan (2016), among others), the contributions on the multivariate side are considerably more limited. Doan et al. (1983) consider a general stationary formulation for the VAR coefficients of their model, but set the autoregressive coefficient to 0.999, which effectively turns the formulation into a random walk. Ciccarelli and Rebucci (2003) and Lubik and Matthes (2015) also propose a general stationary formulation for the law of motion of their time-varying VAR models, but retain the random walk for empirical applications. Clark and Ravazzolo (2015) test for different specifications of stochastic volatility in VAR models, including a stationary autoregressive specification. Their results with the competing random walk formulation are overall inconclusive.

The main challenge while adopting a general stationary process for the dynamic parameters is that of a choice for the mean values and autoregressive coefficients. As economic theory provides little guidance, the literature has mostly avoided the issue by adopting the random walk specification. Yet another option consists in estimating those parameters endogenously. While there has been again much attention to this question in the univariate ARCH literature (see references hereabove), effort has been scarce on the multivariate side. In a first attempt to determine the mean of the structural shock volatility, Uhlig (1997) relies on a set of Beta prior distributions. Primiceri (2005) questions the random walk assumption and tests for exogenous estimation of the autoregressive coefficients on the dynamic processes. He obtains near random walk estimates and concludes that no relevant differences exist compared to the homogeneous

random walk specification. Nevertheless, he does not include a mean term in the autoregressive processes and does not adopt an variable-specific formulation, which may significantly affect the results. Mumtaz and Zanetti (2013) endogenously estimate a single autoregressive coefficient on stochastic volatility, assumed to be common to all the structural shocks.

Aside from theoretical considerations, estimation remains the main challenge for time-varying VAR models. Except for a marginal number of contributions building on frequentist methods (Gorgi et al. (2017)), the Bayesian approach has been unanimously adopted by the literature for its flexibility. So far the benchmark methodology relies on the state-space formulation proposed by Primiceri (2005), building on the algorithm developed by Carter and Kohn (1994). The technique involves a two-pass procedure. It starts with an initial forward filtering pass making use of the Kalman filter to produce recursively the predictive mean for each period, followed by a subsequent backward filtering pass drawing the states in reverse order. A first limit of the procedure resides in its complexity. The use of multiple filtering stages combined with the general Kalman filter approach makes the technique complicated to understand and implement. It also limits the transparency and intuitiveness of the procedure.

A second concern comes for the efficiency of the process. The multiple loops through time and the building of the states in a recursive fashion considerably slow down the procedure. It is not uncommon for time-varying Bayesian VARs to be estimated in hours or even days. This significantly reduces the attractiveness of the model, feasibility remaining a key factor in empirical applications. In this respect, an important contribution was made by Chan and Jeliazkov (2009). These authors propose to replace the usual state-space resolution method with a precision sampler based on a full sample formulation. Despite its efficiency, the procedure has remained confidential and applications have been limited. Poon (2018) expands the approach to time-varying panel VAR models, using a structural factor approach, while Chan and Eisenstat (2018) expand the use of the precision sampler to the estimation of the structural identification matrix. These preliminary expansions remain nevertheless limited. First, they only extend the precision sampler to a fraction of the parameters involved in the model. Second, the dynamic parameters remain defined by the standard random walk approach. Third, the formulation of the precision sampler is not optimised, resulting in reduced efficiency benefits.

Based on these considerations, this paper contributes to the literature in four directions. First, it introduces a general time-varying VAR model which adopts a fully variable-specific approach. For each dynamic parameter, the homogeneous random assumption is relaxed and replaced with a general autoregressive processes with variable-specific mean values and autoregressive coefficients. Second, it proposes an estimation procedure for the model which replaces the usual state-space approach by the precision sampler of Chan and Jeliazkov (2009), generalised to all the dynamic parameters of the model. The procedure is simple, intuitive, and reduces to a standard Gibbs sampling algorithm. It it also optimised in its formulation and provides considerable efficiency gains. Third, it proposes extension procedures to endogenously estimate the mean terms and autoregressive coefficients associated with the laws of motion of each dynamic parameters. The employed priors are informative and contribute to improve the performances of the model. Finally, it conducts a case study on the Great Recession in four major economies (Canada, the European Union, Japan and the United States) which shows that the forecasts produced by the general time-varying model and its extensions outperform the competing homogeneous random walk specifications. Following, it suggests that the crisis could have been better

predicted with a proper use of time-varying VAR models. In this respect, this paper adds to a growing literature discussing the optimal specification of time-varying VAR models regarding forecast accuracy (Clark and Ravazzolo (2015), Aastveit et al. (2017), Kalli and Griffin (2018)).

The remaining of the paper is organised as follows: section 2 introduces the general time-varying model and provides the details of the estimation procedure; section 3 discusses the efficiency of the procedure compared to the usual state-space formulation, along with alternative estimation strategies for the stochastic volatility components of the model; section 4 and 5 respectively develop the extensions allowing for endogenous estimation of the autoregressive coefficients (random inertia) and mean terms (random mean) of the dynamic parameters; section 6 presents the results of the case study on the Great Recession and discusses the benefits of the general time-varying model and its extensions in terms of forecast accuracy; section 7 concludes.

2 A general time-varying model

2.1 The model

Consider the general time-varying model:

$$y_t = C_t z_t + A_{1,t} y_{t-1} + \cdots + A_{p,t} y_{t-p} + \varepsilon_t \qquad t = 1, \cdots, T \qquad , \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma_t) \qquad (1)$$

y_t is a $n \times 1$ vector of observed endogenous variables, z_t is a $m \times 1$ vector of observed exogenous variables such as constant or trends, and ε_t is a $n \times 1$ vector of reduced-form residuals.[1] The residuals are heteroskedastic disturbances following a normal distribution with variance-covariance matrix Σ_t. $C_t, A_{1,t}, \cdots, A_{p,t}$ are matrices of time-varying VAR coefficients comfortable with z_t and the lagged values of y_t. Stacking in a vector β_t the set of VAR coefficients, (1) rewrites:

$$y_t = X_t \beta_t + \varepsilon_t \qquad (2)$$

with:

$$X_t = I_n \otimes x_t \ , \ \ x_t = \begin{pmatrix} z_t' & y_{t-1}' & \cdots & y_{t-p}' \end{pmatrix} \ , \ \ \beta_t = vec(B_t) \ , \ \ B_t = \begin{pmatrix} C_t & A_{1,t} & \cdots & A_{p,t} \end{pmatrix}' \quad (3)$$

Considering specifically row i of (2), the equation for variable i of the model rewrites:

$$y_{i,t} = x_t \beta_{i,t} + \varepsilon_{i,t} \qquad (4)$$

where $\beta_{i,t}$ is the $k \times 1$ vector obtained from column i of B_t. Stacking (4) over the T sample periods yields a full sample formulation for variable i:

$$Y_i = X \beta_i + \mathcal{E}_i \qquad (5)$$

with:

$$Y_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{pmatrix} \quad , \quad X = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_T \end{pmatrix} \quad , \quad \beta_i = \begin{pmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \vdots \\ \beta_{i,T} \end{pmatrix} \quad , \quad \mathcal{E}_i = \begin{pmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,T} \end{pmatrix} \qquad (6)$$

---

[1]Unlike Primiceri (2005) and part of the literature, the model is introduced in reduced form rather than as a structural VAR. There are a number of reasons for doing so, including intuitiveness and flexibility in the implementation of a potential structural decomposition. The correspondence between the two formulations is nevertheless straightforward, as the matrix Δ^{-1} developed in this model is readily identified as the structural matrix in Primiceri (2005).

The variance-covariance matrix Σ_t for the reduced form residuals is decomposed into:

$$\Sigma_t = \Delta_t \Lambda_t \Delta_t' \tag{7}$$

Δ_t is a unit lower triangular matrix, and Λ_t is a diagonal matrix with positive diagonal entries, taking the form:

$$\Delta_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \delta_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \delta_{n1,t} & \cdots & \delta_{n(n-1),t} & 1 \end{pmatrix} , \quad \Lambda_t = \begin{pmatrix} s_1\ exp(\lambda_{1,t}) & 0 & \cdots & 0 \\ 0 & s_2\ exp(\lambda_{2,t}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & s_n\ exp(\lambda_{n,t}) \end{pmatrix} \tag{8}$$

The decomposition of the variance-covariance matrix Σ_t implemented in (9) is common in time-series models (see for instance Hamilton (1994)). The coefficients in Δ_t and Λ_t can be respectively interpreted as the covariance and volatility components of Σ_t. The s_i terms are positive scaling hyperparameters which represent the equilibrium value of the diagonal entries of Λ_t. For technical reasons which will become clear later, it is more convenient to work with Δ_t^{-1} than with Δ_t. The transformation is harmless since there is a one-to-one correspondence between the two terms. As Δ_t is unit lower triangular, so is Δ_t^{-1}:

$$\Delta_t^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \delta_{21,t}^{-1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \delta_{n1,t}^{-1} & \cdots & \delta_{n(n-1),t}^{-1} & 1 \end{pmatrix} \tag{9}$$

Denoting by $\delta_{i,t}^{-1}$ the vector of non-zero and non-one terms in row i of Δ_t^{-1} so that $\delta_{i,t}^{-1} = (\delta_{i1,t}^{-1} \cdots \delta_{i(i-1),t}^{-1})'$, $\delta_{i,t}^{-1}$ represents the (inverse) residual covariance terms of variable i with the other variables of the model.

The dynamics of the model's time varying parameters is specified as follows:

$$
\begin{aligned}
\beta_{i,t} &= (1 - \rho_i)b_i + \rho_i\beta_{i,t-1} + \xi_{i,t} & t &= 2,3,\ldots,T & \xi_{i,t} &\sim \mathcal{N}(0,\Omega_i) \\
\beta_{i,1} &= b_i + \xi_{i,1} & t &= 1 & \xi_{i,1} &\sim \mathcal{N}(0,\tau\Omega_i) \\
\lambda_{i,t} &= \gamma_i\lambda_{i,t-1} + \nu_{i,t} & t &= 2,3,\ldots,T & \nu_{i,t} &\sim \mathcal{N}(0,\phi_i) \\
\lambda_{i,1} &= \nu_{i,1} & t &= 1 & \nu_{i,1} &\sim \mathcal{N}(0,\mu\phi_i) \\
\delta_{i,t}^{-1} &= (1-\alpha_i)d_i^{-1} + \alpha_i\delta_{i,t-1}^{-1} + \eta_{i,t} & t &= 2,3,\ldots,T & \eta_{i,t} &\sim \mathcal{N}(0,\Psi_i) \\
\delta_{i,1}^{-1} &= d_i^{-1} + \eta_{i,1} & t &= 1 & \eta_{i,1} &\sim \mathcal{N}(0,\epsilon\ \Psi_i)
\end{aligned} \tag{10}
$$

ρ_i, γ_i and α_i represent variable-specific autoregressive coefficients while b_i, s_i and d_i^{-1} represent the mean values of the processes. In the base version of the model, these are treated as exogenously set hyperparameters, an assumption which will be relaxed in sections 4 and 5. Clearly, each law of motion nests the usual random walk specification as a special case setting the autoregressive coefficient to 1. For each process, the initial period is formulated to be consistent with the overall dynamics of the parameters. The mean corresponds to the unconditional expectation of the process, while the variance is scaled by the hyperparameters $\tau, \mu, \epsilon > 1$ in order to account

for the greater uncertainty associated with the initial period.

All the innovations in the model are assumed to be jointly normally distributed with the following assumptions on the variance covariance matrix:

$$Var\begin{pmatrix}\varepsilon_t \\ \xi_{i,t} \\ \nu_{i,t} \\ \eta_{i,t}\end{pmatrix} = \begin{pmatrix}\Sigma_t & 0 & 0 & 0 \\ 0 & \Omega_i & 0 & 0 \\ 0 & 0 & \phi_i & 0 \\ 0 & 0 & 0 & \Psi_i\end{pmatrix} \tag{11}$$

This concludes the description of the model. The parameters of interest to be estimated are: the dynamic VAR coefficients $\beta_i = \{\beta_{i,t} : i = 1, \ldots, n; t = 1, \ldots, T\}$, the dynamic volatility terms $\lambda_i = \{\lambda_{i,t} : i = 1, \ldots, n; t = 1, \ldots, T\}$, the dynamic inverse covariance terms $\delta_i^{-1} = \{\delta_{i,t}^{-1} : i = 2, \ldots, n; t = 1, \ldots, T\}$, and the associated variance-covariance parameters Ω_i, ϕ_i and Ψ_i. To these six base parameters must be added an additional parameter $r_i = \{r_{i,t} : i = 1, \ldots, n; t = 1, \ldots, T\}$, whose role will be clarified shortly.

2.2 Bayes rule

Following most of the litterature on time-varying VAR models, Bayesian methods are used to evaluate the posterior distributions of the parameters of interest. Given the model, Bayes rule is given by:

$$\pi(\beta, \Omega, \lambda, \phi, \delta^{-1}, \Psi, r | y) \propto f(y | \beta, \lambda, \delta^{-1}, r)$$
$$\times \left(\prod_{i=1}^{n} \pi(\beta_i | \Omega_i)\pi(\Omega_i)\right)\left(\prod_{i=1}^{n} \pi(\lambda_i | \phi_i)\pi(\phi_i)\right)\left(\prod_{i=2}^{n} \pi(\delta_i^{-1} | \Psi_i)\pi(\Psi_i)\right)\left(\prod_{i=1}^{n}\prod_{t=1}^{T} \pi(r_{i,t})\right) \tag{12}$$

2.3 Likelihood function

A standard formulation of the likelihood function can be obtained from (2):

$$f(y | \beta, \lambda, \delta^{-1}, r) = \prod_{t=1}^{T}(2\pi)^{-n/2}|\Sigma_t|^{-1/2}exp\left(-\frac{1}{2}(y_t - X_t\beta_t)'\Sigma_t^{-1}(y_t - X_t\beta_t)\right) \tag{13}$$

(13) does not permit to estimate the different parameters of the model since it is not expressed in variable-specific terms. After some manipulations, it reformulates as:

$$f(y | \beta, \lambda, \delta^{-1}, r) = (2\pi)^{-nT/2}\left(\prod_{i=1}^{n} s_i^{-T/2}\right)exp\left(-\frac{1}{2}\sum_{i=1}^{n}\lambda_i'1_T\right)$$
$$\times\; exp\left(-\frac{1}{2}\sum_{i=1}^{n} s_i^{-1}\left\{(Y_i - X\beta_i)'\tilde{\Lambda}_i(Y_i - X\beta_i) + (\mathcal{E}_{-i}\delta_i^{-1})'\tilde{\Lambda}_i(\mathcal{E}_{-i}\delta_i^{-1}) + 2(Y_i - X\beta_i)'\tilde{\Lambda}_i(\mathcal{E}_{-i}\delta_i^{-1})\right\}\right) \tag{14}$$

with:

$$
\lambda_i = \begin{pmatrix} \lambda_{i,1} \\ \lambda_{i,2} \\ \vdots \\ \lambda_{i,T} \end{pmatrix} \quad , \quad 1_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad , \quad \tilde{\Lambda}_i = \begin{pmatrix} exp(-\lambda_{i,1}) & 0 & \cdots & 0 \\ 0 & exp(-\lambda_{i,2}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & exp(-\lambda_{i,T}) \end{pmatrix}
$$

$$
\mathcal{E}_{-i} = \begin{pmatrix} \varepsilon'_{-i,1} & 0 & \cdots & 0 \\ 0 & \varepsilon'_{-i,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \varepsilon'_{-i,T} \end{pmatrix} \quad , \quad \varepsilon_{-i,t} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{i-1,t} \end{pmatrix} \quad , \quad \delta_i^{-1} = \begin{pmatrix} \delta_{i,1}^{-1} \\ \delta_{i,2}^{-1} \\ \vdots \\ \delta_{i,T}^{-1} \end{pmatrix} \tag{15}
$$

(14) proves convenient for the estimation of β_i and δ_i^{-1}, but does not provide any conjugacy for λ_i due to the presence of the exponential term $\tilde{\Lambda}_i$. This is a well-known issue of models with stochastic volatility, and the most efficient solution is the so-called normal offset mixture representation proposed by Kim et al. (1998).[2] The procedure consists in reformulating the likelihood function in terms of the transformed shock $e_t = (\Delta_t \Lambda_t^{1/2})^{-1} \varepsilon_t$. It is trivially shown that e_t is a vector of structural shock with $e_t \sim \mathcal{N}(0, I_n)$. Considering specifically the shock $e_{i,t}$ in the vector, squaring, taking logs and rearranging eventually yields:

$$
\hat{e}_{i,t} = \log(e_{i,t}^2) = \hat{y}_{i,t} - \lambda_{i,t} \qquad \hat{y}_{i,t} = \log\left(s_i^{-1}(\varepsilon_{i,t} + \delta_{i,t}^{-1\prime}\varepsilon_{-i,t})^2 \right) \tag{16}
$$

$\hat{e}_{i,t}$ follows a log chi-squared distribution which does not grant any conjugacy. Kim et al. (1998) thus propose to approximate the shock as an offset mixture of normal distributions. The approximation is given by:

$$
\hat{e}_{i,t} \approx \sum_{j=1}^{7} \mathbb{1}(r_{i,t} = j) \, z_j \qquad , \qquad z_j \sim \mathcal{N}(m_j, v_j) \qquad , \qquad Pr(r_{i,t} = j) = q_j \tag{17}
$$

The values for m_j, v_j and q_j can be found in Table 4 of Kim et al. (1998). The constants m_j and v_j respectively represent the mean and variance components of the normally distributed random variable z_j. $r_{i,t}$ is a categorical random variable taking discrete values $j = 1, \ldots, 7$, the probability of obtaining each value being equal to q_j. Finally, $\mathbb{1}(r_{i,t} = j)$ is an indicator function taking a value of 1 if $r_{i,t} = j$, and a value of 0 otherwise. To draw from the log chi-squared distribution, the mixture first randomly draws a value for $r_{i,t}$ from its categorical distribution; once $r_{i,t}$ is known, its value determines which component z_j of the mixture is selected. $\hat{e}_{i,t}$ then turns into a regular normal random variable with mean m_j and variance v_j. Given (16) and the offset mixture (17), an approximation of the likelihood function obtains as:

$$
f(y|\beta, \lambda, \delta^{-1}, r) = \prod_{i=1}^{n} \prod_{t=1}^{T} \sum_{j=1}^{7} \mathbb{1}(r_{i,t} = j) \left\{ (2\pi v_j)^{-1/2} \exp\left(-\frac{1}{2} \frac{(\hat{y}_{i,t} - \lambda_{i,t} - m_j)^2}{v_j} \right) \right\} \tag{18}
$$

For the estimation of λ_i, a more convenient joint formulation can be adopted. Defining $r_i = (r_{i,1} \quad \ldots \quad r_{i,T})'$, denoting by J any possible value for r_i, by m_J and v_J the resulting mean and variance vectors, and defining $V_J = diag(v_J)$, the likelihood function rewrites as a mixture of multivariate normal distributions:

---

[2]Section 3.3 discusses a number of alternative strategies to overcome this issue.

$$f(y|\beta, \lambda, \delta^{-1}, r)$$

$$= \prod_{i=1}^{n} \sum^{J} \mathbb{1}(r_i = J) \left\{ (2\pi)^{-T/2} |V_J|^{-1/2} exp\left(-\frac{1}{2}(\hat{Y}_i - \lambda_i - m_J)' V_J^{-1}(\hat{Y}_i - \lambda_i - m_J)\right) \right\} \quad (19)$$

with:

$$\hat{Y}_i = \begin{pmatrix} \hat{y}_{i,1} & \hat{y}_{i,2} & \dots & \hat{y}_{i,T} \end{pmatrix}' = log(s_i^{-1} \mathcal{Q}_i) \qquad \mathcal{Q}_i = (\mathcal{E}_i + \mathcal{E}_{-i} \delta_i^{-1})^2 \quad (20)$$

2.4 Priors

The formulation of the priors for the dynamic parameters obtains from a generalisation of the procedure by Chan and Jeliazkov (2009). Consider first the VAR coefficients β_i. Starting from (10), the law of motion can be expressed in compact form as:

$$\begin{pmatrix} I_k & 0 & \cdots & 0 \\ -\rho_i I_k & I_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\rho_i I_k & I_k \end{pmatrix} \begin{pmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \vdots \\ \beta_{i,T} \end{pmatrix} = \begin{pmatrix} b_i \\ (1-\rho_i)b_i \\ \vdots \\ (1-\rho_i)b_i \end{pmatrix} + \begin{pmatrix} \xi_{i,1} \\ \xi_{i,2} \\ \vdots \\ \xi_{i,T} \end{pmatrix} \quad (21)$$

or:

$$(F_i \otimes I_k)\, \beta_i = \bar{b}_i + \xi_i \qquad F_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\rho_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\rho_i & 1 \end{pmatrix} \qquad \bar{b}_i = \begin{pmatrix} b_i \\ (1-\rho_i)b_i \\ \vdots \\ (1-\rho_i)b_i \end{pmatrix} \qquad \xi_i = \begin{pmatrix} \xi_{i,1} \\ \xi_{i,2} \\ \vdots \\ \xi_{i,T} \end{pmatrix} \quad (22)$$

Also:

$$Var(\xi_i) = \begin{pmatrix} \tau\Omega_i & 0 & \cdots & 0 \\ 0 & \Omega_i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Omega_i \end{pmatrix} = I_\tau \otimes \Omega_i \qquad I_\tau = \begin{pmatrix} \tau & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad (23)$$

(22) and (23) respectively imply $\beta_i = (F_i \otimes I_k)^{-1}\bar{b}_i + (F_i \otimes I_k)^{-1}\xi_i$ and $\xi_i \sim \mathcal{N}(0, I_\tau \otimes \Omega_i)$. From this and rearranging, the prior distribution eventually obtains as:

$$\pi(\beta_i|\Omega_i) \sim \mathcal{N}(\beta_{i0}, \Omega_{i0}) \qquad \beta_{i0} = 1_T \otimes b_i \qquad \Omega_{i0} = (F_i' I_\tau^{-1} F_i \otimes \Omega_i^{-1})^{-1} \quad (24)$$

Using for λ_i and δ_i^{-1} equivalent procedures and notations, it is straightforward to obtain:

$$\pi(\lambda_i|\phi_i) \sim \mathcal{N}(0, \Phi_{i0}) \qquad \Phi_{i0} = \phi_i (G_i' I_\mu^{-1} G_i)^{-1}$$
$$\pi(\delta_i^{-1}|\Psi_i) \sim \mathcal{N}(\delta_{i0}^{-1}, \Psi_{i0}) \qquad \delta_{i0}^{-1} = 1_T \otimes d_i^{-1} \qquad \Psi_{i0} = (H_i' I_\epsilon^{-1} H_i \otimes \Psi_i^{-1})^{-1} \quad (25)$$

Once the prior distributions for the dynamic parameters are determined, it remains to set the priors for their associated variance-covariance parameters. The choice is that of standard inverse Wishart and inverse Gamma distributions. Precisely:

$$\pi(\Omega_i) \sim IW(\zeta_0, \Upsilon_0) \qquad \pi(\phi_i) \sim IG\left(\frac{\kappa_0}{2}, \frac{\omega_0}{2}\right) \qquad \pi(\Psi_i) \sim IW(\varphi_0, \Theta_0) \quad (26)$$

Finally, from (17), it is immediate that the prior distribution for $r_{i,t}$ is categorical:

$$\pi(r_{i,t}) \sim Cat(q_1, \dots, q_7) \quad (27)$$

2.5 Posteriors

The joint posterior obtained from (12) is analytically intractable. Following standard practices, the marginal posteriors are then estimated from a Gibbs sampling algorithm relying on conditional distributions.

For β_i, Bayes rule (12) implies $\pi(\beta_i|y,\beta_{-i}) \propto f(y|\beta,\lambda,\delta^{-1},r)\pi(\beta_i|\Omega_i).$[3] From the likelihood (14), the prior (24) and rearranging, it follows that:

$$\pi(\beta_i|y,\beta_{-i}) \sim \mathcal{N}(\bar{\beta}_i,\bar{\Omega}_i) \qquad \text{with:}$$
$$\bar{\Omega}_i = \left(s_i^{-1}X'\tilde{\Lambda}_iX + F_i'\ I_\tau^{-1}\ F_i \otimes \Omega_i^{-1}\right)^{-1}$$
$$\bar{\beta}_i = \bar{\Omega}_i\left(s_i^{-1}X'\tilde{\Lambda}_i\left[Y_i + \mathcal{E}_{-i}\delta_i^{-1}\right] + F_i'I_\tau^{-1}F_i1_T \otimes \Omega_i^{-1}b_i\right) \tag{28}$$

For λ_i, Bayes rule (12) implies $\pi(\lambda_i|y,\lambda_{-i}) \propto f(y|\beta,\lambda,\delta^{-1},r)\pi(\lambda_i|\phi_i)$. From the approximate likelihood (19), the prior (25) and rearranging, it follows that:

$$\pi(\lambda_i|y,\lambda_{-i}) \sim \mathcal{N}(\bar{\lambda}_i,\bar{\Phi}_i) \qquad \text{with:}$$
$$\bar{\Phi}_i = (V_J^{-1} + \phi_i^{-1}G_i'I_{-\mu}G_i)^{-1} \qquad \bar{\lambda}_i = \bar{\Phi}_i(V_J^{-1}[\hat{Y}_i - m_J]) \tag{29}$$

For δ_i^{-1}, Bayes rule (12) implies $\pi(\delta_i^{-1}|y,\delta_{-i}^{-1}) \propto f(y|\beta,\lambda,\delta^{-1},r)\pi(\delta_i^{-1}|\Psi_i)$. From the likelihood (14), the prior and rearranging, it follows that:

$$\pi(\delta_i^{-1}|y,\delta_{-i}^{-1}) \sim \mathcal{N}(\bar{\delta}_i^{-1},\bar{\Psi}_i) \qquad \text{with:}$$
$$\bar{\Psi}_i = (s_i^{-1}\mathcal{E}_{-i}'\tilde{\Lambda}_i\mathcal{E}_{-i} + H_i'\ I_{-\epsilon}\ H_i \otimes \Psi_i^{-1})^{-1}$$
$$\bar{\delta}_i^{-1} = \bar{\Psi}_i(-s_i^{-1}\mathcal{E}_{-i}'\tilde{\Lambda}_i\mathcal{E}_i + H_i'I_{-\epsilon}H_i1_T \otimes \Psi_i^{-1}d_i^{-1}) \tag{30}$$

Consider now the associated variance-covariance parameters. For Ω_i, Bayes rule (12) implies $\pi(\Omega_i|y,\Omega_{-i}) \propto \pi(\beta_i|\Omega_i)\pi(\Omega_i)$. From the priors (24) and (26) then rearranging, it follows that:

$$\pi(\Omega_i|y,\Omega_{-i}) \sim IW(\bar{\zeta},\bar{\Upsilon}_i) \qquad \text{with:} \qquad \bar{\zeta} = T + \zeta_0 \qquad \bar{\Upsilon}_i = \tilde{B}_i + \Upsilon_0$$
$$\tilde{B}_i = (B_i - 1_T' \otimes b_i)\ (F_i'\ I_{-\tau}\ F_i)(B_i - 1_T' \otimes b_i)' \qquad B_i = \begin{pmatrix} \beta_{i,1} & \beta_{i,2} & \cdots & \beta_{i,T} \end{pmatrix} \tag{31}$$

For ϕ_i, Bayes rule (12) implies $\pi(\phi_i|y,\phi_{-i}) \propto \pi(\lambda_i|\phi_i)\pi(\phi_i)$. From the priors (25) and (26) then rearranging, it follows that:

$$\pi(\phi_i|y,\phi_{-i}) \sim IG(\bar{\kappa},\bar{\omega}_i) \qquad \text{with:} \qquad \bar{\kappa} = \frac{T + \kappa_0}{2} \qquad \bar{\omega}_i = \frac{\lambda_i'(G_i'I_{-\mu}G_i)\lambda_i + \omega_0}{2} \tag{32}$$

For Ψ_i, Bayes rule (12) implies $\pi(\Psi_i|y,\Psi_{-i}) \propto \pi(\delta_i^{-1}|\Psi_i)\pi(\Psi_i)$. From the priors (25) and (26) then rearranging, it follows that:

$$\pi(\Psi_i|y,\Psi_{-i}) \sim IW(\bar{\varphi},\bar{\Theta}_i) \qquad \text{with:} \qquad \bar{\varphi} = T + \varphi_0 \qquad \bar{\Theta}_i = \tilde{D}_i + \Theta_0$$
$$\tilde{D}_i = (D_i - 1_T' \otimes d_i^{-1})\ (H_i'\ I_{-\epsilon}\ H_i)(D_i - 1_T' \otimes d_i^{-1})' \qquad D_i = \begin{pmatrix} \delta_{i,1}^{-1} & \delta_{i,2}^{-1} & \cdots & \delta_{i,T}^{-1} \end{pmatrix} \tag{33}$$

Finally, for $r_{i,t}$, Bayes rule (12) implies $\pi(r_{i,t}|y,r_{-i,t}) \propto f(y|\beta,\lambda,\delta^{-1},r)\pi(r_{i,t})$. From the approximate likelihood (18) and the prior (27), it follows immediately that:

$$\pi(r_{i,t}|y,r_{-i,t}) \sim Cat(\bar{q}_1,\ldots,\bar{q}_7) \quad \text{with:} \quad \bar{q}_j = (2\pi v_j)^{-1/2}\exp\left(-\frac{1}{2}\frac{(\hat{y}_{i,t}-\lambda_{i,t}-m_j)^2}{v_j}\right)q_j \tag{34}$$

---

[3]For θ_i any parameter, $\pi(\theta_i|\theta_{-i})$ is used to denote the density of θ_i conditional on all the model parameters except θ_i.

2.6 MCMC algorithm

Once the conditional posteriors are obtained, it is possible to introduce the MCMC algorithm for the model. The latter reduces to a simple 7-step procedure, as follows:

Algorithm 1: MCMC algorithm for the general time-varying model:

1. Sample λ_i from $\pi(\lambda_i|y, \lambda_{-i}) \sim \mathcal{N}(\bar{\lambda}_i, \bar{\Phi}_i)$.

2. Sample β_i from $\pi(\beta_i|y, \beta_{-i}) \sim \mathcal{N}(\bar{\beta}_i, \bar{\Omega}_i)$.

3. Sample δ_i^{-1} from $\pi(\delta_i^{-1}|y, \delta_{-i}^{-1}) \sim \mathcal{N}(\bar{\delta}_i^{-1}, \bar{\Psi}_i)$.

4. Sample Ω_i from $\pi(\Omega_i|y, \Omega_{-i}) \sim IW(\bar{\zeta}, \bar{\Upsilon}_i)$.

5. Sample ϕ_i from $\pi(\phi_i|y, \phi_{-i}) \sim IG(\bar{\kappa}, \bar{\omega}_i)$.

6. Sample Ψ_i from $\pi(\Psi_i|y, \Psi_{-i}) \sim IW(\bar{\varphi}, \bar{\Theta}_i)$.

7. Sample $r_{i,t}$ from $\pi(r_{i,t}|y, r_{-i,t}) \sim Cat(\bar{q}_1, \ldots, \bar{q}_7)$.

Two remarks can be made about the algorithm. First, observe that the ordering of the steps in the algorithm differs from the one used for the presentation of the model. It introduces λ_i first, then the other model parameters, and eventually the offset mixture parameter $r_{i,t}$. This specific ordering is necessary to recover the correct posterior distribution if the normal offset mixture is used to provide an approximation of the likelihood function. See Del Negro and Primiceri (2015) for details. Second, due to the large dimension of β_i and δ_i^{-1}, it is not advisable nor efficient to compute explicitly the parameters $\bar{\beta}_i, \bar{\Omega}_i, \bar{\delta}_i^{-1}$ and $\bar{\Psi}_i$ defining the normal distributions. A better option consists in taking advantage of the sparse and banded nature of $\bar{\Omega}_i^{-1}$ and $\bar{\Psi}_i^{-1}$ to proceed efficiently by backward and forward substitution. See Chan and Jeliazkov (2009), Algorithm 1 for details.

3 Efficiency analysis

3.1 Estimation

As a preliminary exercise and for the sake of comparison, the methodology introduced in section 2 is used to estimate the small U.S. economy model of Primiceri (2005). The data set includes 3 variables: a series of inflation rate and unemployment rate representing the non-policy block, and a short-term nominal interest rate representing the policy block. Estimation is conducted with two lags and one constant on quarterly data running from 1963q1 to 2001q3, resulting in a sample of size $T = 153$ quarters.

For the priors, one possibility consists in calibrating the hyperparameters with a training sample, as done by Primiceri (2005). Since there is no evidence that such a strategy improves on the estimates, simple values are used instead. For the inverse Wishart priors on the variance-covariance hyperparameters Ω_i and Ψ_i, the degrees of freedom are set to a small value of 5 additional to the parameter dimension, namely $\zeta_0 = k + 5$ and $\varphi_0 = (i - 1) + 5$. The scale parameters are

set to $\Upsilon_0 = 0.01 I_k$ and $\Theta_0 = 0.01 I_{i-1}$. Combined with the degrees of freedom, this implies an average 0.05 standard deviation for the shocks on the dynamic processes, or in other words a 5% difference between consecutive values of β_i and δ_i^{-1}. For the stochastic volatility part of the model, the prior is slightly looser. The shape and scale parameters of the inverse Gamma prior distribution on ϕ_i are set to $\kappa_0 = 1$ and $\omega_0 = 0.01$ to generate a weakly informative prior. Finally, the initial period variance scaling terms are set to $\tau = \mu = \epsilon = 5$ in order to obtain a variance over the initial periods which is roughly equivalent to that prevailing for the rest of the sample. For the dynamic processes, the autoregressive coefficients are set to $\rho_i = \gamma_i = \alpha_i = 0.9$, inducing stationarity but a susbtantial degree of inertia. For the mean of the dynamic processes, static OLS estimates are used. b_i is set to its OLS counterpart $\hat{\beta}_i$. Similarly, the static OLS estimate $\hat{\Sigma}$ is decomposed into $\hat{\Sigma} = \hat{\Delta} \hat{\Lambda} \hat{\Delta}'$. s_i is then set as the ith diagonal entry of $\hat{\Lambda}$, and d_i^{-1} is determined as the free elements of the ith row of $\hat{\Delta}^{-1}$. The model is run from 10000 iterations of the MCMC algorithm, discarding the initial 2000 iterations as burn-in sample. As shown in Appendix A, The convergence diagnostics are satisfactory, indicating proper convergence to the posterior distribution.

To highlight the main characteristics of the model, Figure 1 compares the results obtained with the methodologies of Primiceri (2005) (left panels, without corrigendum), Del Negro and Primiceri (2015) (middle panels, integrating the corrigendum), and the general time-varying model (right panels).[4] The three top panels report the historical evolution of the volatility of the structural shocks while the three bottom panels focus on the developments of the dynamic coefficients (own first lag of each variable).

On the qualitative side, the different models produce comparable outcomes in terms of shock volatility. For inflation and unemployment, the three models detect an initial fuelling in volatility somewhere around 1975 followed by a moderate recurrence in the course of the 1980's. For the interest rate, the three models adequately identify the high volatility peak occurring in 1982. Interestingly enough, the results obtained with the general time-varying model are qualitatively closer to those obtained from the original Primiceri (2005) model than to those provided by the corrigendum model by Del Negro and Primiceri (2015) which are considerably smoother. The conclusion concerning the VAR coefficients are quite different: while the estimates from the general time-varying model display a significant amount of variation accross the sample, those produced by the two comparative models are virtually flat.

---

[4]The results for the models by Primiceri (2005) and Del Negro and Primiceri (2015) are obtained from the Matlab code written by Marco Del Negro and Giorgio Primiceri, and cross-checked with alternative versions of the code provided by Haroon Mumtaz, Dimitris Korobilis and Benedikt Kolb. I am grateful and indebted to these authors for providing their material.
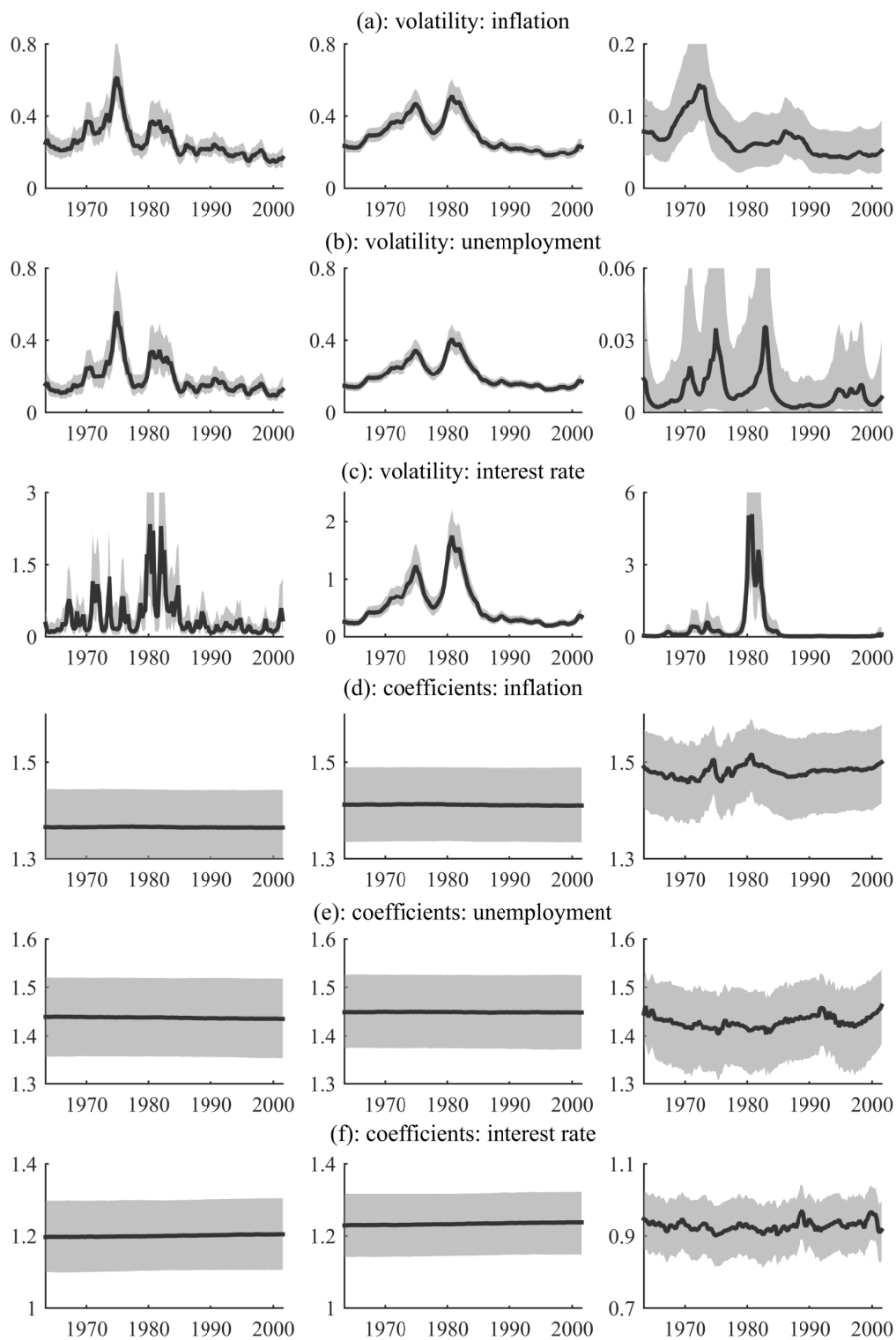
Figure 1: Median and 70% credibility interval for the volatility and VAR coefficients (own first lag) of inflation, unemployment and the interest rate
left panels: Primiceri (2005) (no corrigendum), middle panels: Del Negro and Primiceri (2015) (with corrigendum), right panels: general time-varying model

The explanation comes from the quantitative side of the models. The three top panels reveal that the overall range of volatility induced by the general time-varying model is considerably smaller than with the other models. This is true for all the variables, despite the higher peak in volatility for the interest rate in 1982. This is a consequence of the inclusion of the variable-specific mean terms s_i in the dynamic processes of stochastic volatility, as stated in (8). This sets s_i as the prior equilibrium value of the process, which drives by construction the posterior towards it. By contrast, the log-normal random walk formulations of Primiceri (2005) and Del Negro and Primiceri (2015) effectively amount to scaling the means to $s_i = 1$ for all the variables. This normalisation of the prior equilibrium is not innocuous as it pushes the posterior estimates upward, resulting in higher levels of fluctuation. As a consequence, most of the variation observed in the data is attributed to stochastic volatility. This leaves only a marginal extent of variation to be explained by the dynamic coefficients, hence the remarkably flat estimates. On the other hand, because the stochastic volatility contributions of the general time-varying model remain modest, a larger share of the observed data variability is left to be explained by the dynamic VAR coefficients, hence the wider range of fluctuation. Overall, these conclusions question the common belief that time-varying models attribute the bulk of observed fluctuations to stochastic volatility, while the dynamic responses play a marginal role. This feature may in fact be a technical artefact produced by the random walk assumption, which disappears once a more general formulation is adopted.

3.2 Efficiency

This section discusses the computational efficiency of the general time-varying model compared to the standard Primiceri (2005) methodology, integrating the corrigendum of Del Negro and Primiceri (2015). To do so, three models are considered. The first is the model developed in the previous section, labeled as the "small" model. As a reminder, the model includes three variables (inflation, unemployment and interest rate), two lags and runs from 1963q1 to 2001q3, which represents an estimation sample of 153 quarters. The second is an abridged version of the small model which contains only two variables (inflation and unemployment), one lag, and runs for a smaller period ranging from 1980q1 to 2001q3, resulting in an estimation sample of 86 quarters. This is labelled as the "tiny" model. The final "medium" model to be estimated is an expanded version of the initial model. It comprises four variables (inflation, unemployment and interest rate supplemented with a series of real GDP growth), three lags and covers a longer period ranging from 1953q1 to 2018q1, for a total of 218 quarters. Table 1 reports the approximate estimation time to achieve 10000 iterations with the Primiceri (2005) methodology and the general time-varying model [5]. The table reveals significant efficiency gains from using the general time-varying model methodology. The computational benefits range from about 55% for the medium model to more than 90% for the tiny model. For a typical small-sized time-varying model like the small US economy model of Primiceri (2005), the computational gain is greater than 80%. Clearly, the returns are diminishing with the number of parameters to be estimated. The benefit remains however considerable even when the number of parameters is quite large, as in the case of the medium model. In fact, for any reasonably sized time-varying VAR model, the benefit will remain sizable.

---

[5] All the estimations were conducted on a computer equipped with a 2 GHz Intel Core processor and 4 Go of RAM, for a Windows performance rating of 5.1/10, i.e., a fairly average computer. While the absolute numerical performances depend on the technical capacities of every machine, the ratio of the relative performance on estimating different models remains invariant to the computer used.

	Number of estimated parameters	Methodology of Primiceri (2005)	General time-varying model	Efficiency gain
Tiny model	774	558 s (9 m 18 s)	51 s (0 m 51 s)	90.9 %
Small model	4131	1186 s (19 m 46 s)	218 s (3 m 38 s)	81.6 %
Medium model	16132	2650 s (44 m 10 s)	1195 s (19 m 55 s)	54.9 %

Table 1: Summary of estimation performances for the different methodologies (for 10000 iterations; time in seconds)

There are two main sources for the observed efficiency gains. The first consists in the adoption of the precision sampler of Chan and Jeliazkov (2009) in place of the usual Kalman filter procedure of Carter and Kohn (1994) for drawing the dynamic parameters β_i, λ_i and δ_i^{-1}. The benefit from the procedure is double: while the standard approach proceeds period by periods and requires a two-pass filtering process, the precision sampler draws for all the periods at once from the highly multivariate posterior distribution of each parameter. The efficiency of the method depends on the size of the matrices involved (see $F_i, I_\tau^{-1}, \Omega_i, G_i, I_{-\mu}, H_i, I_{-\epsilon}$ and Ψ_i in (28), (29) and (30)) which themselves depend on the dimension of the model n, the number of coefficients per equation k, as well as the number of sample periods T. Larger matrices involve a more than linear increase in the number of computations and result in a relative loss of efficiency, though the number of calculations evolve at less than a square rate due to the sparse and banded nature of $\bar{\Omega}_i^{-1}, \bar{\Phi}_i^{-1}$ and $\bar{\Psi}_i^{-1}$. It is then not surprising that the benefit from the precision sampler gets smaller as the overall number of parameters increases, even though it remains substantial for any reasonable model.

The second source of gains lies in the optimised formulation of the precision sampler. While Chan and Jeliazkov (2009) and Chan and Eisenstat (2018) realise the computations at the largest scale, the present model is formulated to take advantage of the Kronecker structure of the formulas, allowing to work at a smaller scale. For instance, the computation of the posterior parameter $\bar{\Omega}_i^{-1}$ in (28) only involves the update of the $k \times k$ matrix Ω_i^{-1} at each iteration of the MCMC algorithm, which is then enlarged through the Kronecker product $F_i' I_\tau^{-1} F_i \otimes \Omega_i^{-1}$, where $F_i' I_\tau^{-1} F_i$ is a constant term which needs only to be computed once, prior to the initiation of the algorithm. By contrast, Chan and Jeliazkov (2009) explicitly re-compute the whole term $(F_i \otimes I_k)(I_\tau \otimes \Omega_i)^{-1}(F_i \otimes I_k)$ at every iteration,[6] which involves products and inversions of matrices of size $Tk \times Tk$. Also, rather than relying on the simple parameter τ to determine the distribution of the first period, these authors create an additional step which endogenously estimate an initial condition for period 0. While the gains from these formulations may sound modest, they eventually add up to generate substantial benefits once applied to all the dynamic parameters and repeated over thousands of iterations. [7]

---

[6]In fact, the term computed by Chan and Jeliazkov (2009) is only an equivalent of $(F_i \otimes I_k)(I_\tau \otimes \Omega_i)^{-1}(F_i \otimes I_k)$, the formulation of their model being slightly different from the present general time-varying model.

[7]Chan and Jeliazkov (2009) report an approximate 60% efficiency gain from their precision sampler procedure applied to a small-sized model which only includes time-varying VAR coefficients. The present time-varying model, which includes both time-varying VAR coefficients and stochastic volatility results in more than 80% efficiency gains, for a model of similar size.

3.3 Alternative estimation strategies

The main difficulty in the estimation of time-varying models comes from the standard log-normal formulation of the stochastic volatility processes. This assumption results in a likelihood function containing double exponential stochastic volatility terms such as the $\tilde{\Lambda}_i$ matrices in (14). These terms are challenging and prevent any conjugacy with a normal prior distribution for λ_i. The solution adopted for the general time-varying model is the normal offset mixture strategy of Kim et al. (1998). While it yields a convenient reformulation of the likelihood function, it also involves the estimation of the extra set of parameters r_i which is undesirable as it generates additional computations which contribute to reduce efficiency. It is thus important to consider alternative estimation strategies which may prove more efficient. To introduce the alternative solutions, observe first that the likelihood function (14) can rewrite as:

$$
f(y|\beta, \lambda, \delta^{-1}, r) = (2\pi)^{-nT/2} \left(\prod_{i=1}^{n} s_i^{-T/2} \right) exp \left(-\frac{1}{2} \sum_{i=1}^{n} \left\{ \lambda_i' 1_T + s_i^{-1} \tilde{\lambda}_i' \; \mathcal{Q}_i \right\} \right) \tag{35}
$$

where $\tilde{\lambda}_i = (exp(-\lambda_{i,1}) \quad exp(-\lambda_{i,2}) \quad \cdots \quad exp(-\lambda_{i,T}))'$. Clearly, $\tilde{\lambda}_i$ is the equivalent of $\tilde{\Lambda}_i$ in (14). It constitutes the log normal term which generates the difficulties in obtaining analytical forms for the posterior distribution of λ_i. Indeed, Bayes rule (12) implies that $\pi(\lambda_i|y, \lambda_{-i}) \propto f(y|\beta, \lambda, \delta^{-1}, r)\pi(\lambda_i|\phi_i)$; substituting then for (35) and (25) and rearranging eventually yields:

$$
\pi(\lambda_i|y, \lambda_{-i}) \propto exp \left(-\frac{1}{2} \left\{ \lambda_i' 1_T + s_i^{-1} \tilde{\lambda}_i' \; \mathcal{Q}_i + \lambda_i' \Phi_{i0}^{-1} \lambda_i \right\} \right) \tag{36}
$$

This cannot be reformulated as a multivariate normal density due to the presence of $\tilde{\lambda}_i$. As such, this posterior density is not workable. Besides the normal offset mixture approach of Kim et al. (1998), the literature has provided two classes of solutions for this issue. The first consists in the adoption of an accept-reject algorithm approach, while the second relies on the Metropolis-Hastings methodology. Both strategies can be applied either for all the sample periods simultaneously, or on a period-by-period basis.

Consider first the accept-reject approach. This strategy was advocated for model with stochastic volatility by Kim et al. (1998). Noting that the problematic term $\tilde{\lambda}_i$ in (36) can be approximated by a first-order Taylor series around 0 as $\tilde{\lambda}_i = exp(-\lambda_i) \geq 1_T - \lambda_i$, where the inequality follows from the convexity of $\tilde{\lambda}_i$, one obtains:

$$
\begin{aligned}
\pi(\lambda_i|y, \lambda_{-i}) \quad &\propto exp \left(-\frac{1}{2} \left\{ \lambda_i' 1_T + s_i^{-1} \tilde{\lambda}_i' \; \mathcal{Q}_i + \lambda_i' \Phi_{i0}^{-1} \lambda_i \right\} \right) \\
&\leq exp \left(-\frac{1}{2} \left\{ \lambda_i' 1_T + s_i^{-1} (1_T - \lambda_i)' \; \mathcal{Q}_i + \lambda_i' \Phi_{i0}^{-1} \lambda_i \right\} \right) \\
&\propto exp \left(-\frac{1}{2} (\lambda_i - \bar{\lambda}_i)' \Phi_{i0}^{-1} (\lambda_i - \bar{\lambda}_i) \right)
\end{aligned} \tag{37}
$$

with:

$$
\bar{\lambda}_i = \frac{1}{2} \Phi_{i0}(s_i^{-1} \mathcal{Q}_i - 1_T) \tag{38}
$$

The final row of (37) is immediately recognisable as the kernel of a multivariate normal density. In other words, the actual posterior density (36) is dominated by a simple multivariate normal density. A natural accept-reject approach thus consists in drawing a candidate value from $\lambda_i \sim \mathcal{N}(\bar{\lambda}_i, \Phi_{i0})$, and then accept it with a probability equal to the ratio of the actual density to the candidate density, given by $p(a) = exp\left(-\frac{1}{2}s_i^{-1}(\tilde{\lambda}_i + \lambda_i - 1_T)\mathcal{Q}_i\right)$. If the candidate is rejected, a new candidate is considered until acceptance is obtained. In theory, the algorithm can be very efficient, especially if the acceptance rate is high. The additional benefits are obvious: unlike the offset mixture approach, the accept-reject approach does not involve the estimation of any additional random variables; the formulas involved in the algorithm are trivial to compute; also, efficiency is maximised from the fact that in case of rejection, a new candidate is immediately drawn until acceptance is achieved.

In practice however, the algorithm works poorly and often ends up being trapped repeatedly at the rejection stage. There are two main explanations for this situation. The first is the quality of the approximation provided by the dominating function: a poor approximation results in low acceptance rates, except if the candidate is drawn very close to the approximation point. For the considered application, it is clear that the first-order Taylor series $\tilde{\lambda}_i \approx 1_T - \lambda_i$ constitutes only a crude approximation. The second explanation lies in the dimensionality of the target distribution: a high dimension contributes to reduce the acceptance rate of the algorithm as the distances between the actual and candidate distributions add up over the dimensions. It is in fact a well-known property that the acceptance probability of any accept-reject algorithm falls to zero as the dimensionality of the candidate distribution approaches infinity. Due to its high dimensionality of T, the joint accept-reject approach for λ_i becomes quickly inefficient.

In an attempt to suppress the dimensionality issue, one may opt for a period-by-period approach. From Bayes rule (12), the likelihood function (35), the prior (25), a Taylor approximation and some rearrangement, one obtains:[8]

$$\pi(\lambda_{i,t}|y, \lambda_{-i-t}) \propto exp\left(-\frac{1}{2}\left\{\lambda_{i,t} + s_i^{-1}\tilde{\lambda}_{i,t}\ \mathcal{Q}_{i,t} + \frac{(\lambda_{i,t} - \hat{\lambda}_{i,t})}{\hat{\phi}_i}\right\}\right) \leq exp\left(-\frac{1}{2}\frac{(\lambda_{i,t} - \bar{\lambda}_{i,t})}{\bar{\phi}_i}\right) \quad (39)$$

with:

$$\bar{\phi}_i = \hat{\phi}_i = \frac{\phi_i}{1 + \gamma_i^2} \qquad \bar{\lambda}_{i,t} = \hat{\lambda}_{i,t} + \frac{\bar{\phi}_i}{2}(s_i^{-1}\mathcal{Q}_{i,t} - 1) \qquad \hat{\lambda}_{i,t} = \frac{\gamma_i}{1 + \gamma_i^2}(\lambda_{i,t-1} + \lambda_{i,t+1}) \quad (40)$$

Following, a period-by-period accept-reject procedure consists in drawing a candidate value from $\lambda_{i,t} \sim \mathcal{N}(\bar{\lambda}_{i,t}, \bar{\phi}_i)$ and accept it with probability $p(a) = exp\left(-\frac{1}{2}s_i^{-1}(\tilde{\lambda}_{i,t} + \lambda_{i,t} - 1)\mathcal{Q}_{i,t}\right)$. While this procedure yields slightly better results than the simultaneous period approach, its performance remains poor. Even though the dimensionality issue is settled, the approximation remains too rough and the acceptance probability stays close to zero.[9] In the end, the accept-reject approach proves unsuccessful, except in the case of extremely tight priors for λ_i which generate candidates close to zero, the point of approximation. Consequently, it does not represent a viable alternative to the offset mixture approach in general.

---

[8] The values are slightly different for the initial and final periods and are omitted to save space.

[9] It may be tempting to use higher order Taylor series to refine the approximation. However, such approximations do not dominate the target distribution anymore, or yield formulas which are not conjugate. Attempts with alternative dominating functions were not successful either.

The second class of methodologies relies on the Metropolis-Hastings approach. This strategy was proposed by Cogley and Sargent (2005) for multivariate models with stochastic volatility. A first possibility consists in adopting a joint period approach. Notice that (36) can rewrite as:

$$\pi(\lambda_i|y, \lambda_{-i}) \propto exp\left(-\frac{1}{2}\left\{\lambda_i'1_T + s_i^{-1}\tilde{\lambda}_i' \ \mathcal{Q}_i\right\}\right) \ exp\left(-\frac{1}{2}\lambda_i'\Phi_{i0}^{-1}\lambda_i\right) \tag{41}$$

The first term on the right-hand side represents the contribution of the data, while the second represents the prior distribution. A simple strategy consists in drawing a candidate from the prior (the second term) and then confront it to the data (the likelihood contribution) to decide on acceptance or rejection.[10] Specifically, this defines an independent transition kernel which yields an acceptance probability of:

$$p(a) = exp\left(-\frac{1}{2}\{(\lambda_i^{(n)} - \lambda_i^{(n-1)})'1_T + s_i^{-1}(\tilde{\lambda}_i^{(n)} - \tilde{\lambda}_i^{(n-1)})' \ \mathcal{Q}_i\}\right) \tag{42}$$

where $\lambda_i^{(n-1)}$ denotes the value inherited from the previous iteration, and $\lambda_i^{(n)}$ denotes the candidate at iteration n. If the candidate is rejected, the value from the previous iteration is retained. The advantage of this Metropolis-Hastings algorithm lies in the simplicity of the approach and the formulas involved in the calculations. On the other hand, a well-known drawback of such algorithm is that it produces repeated values at rejection, which often leads to increase the number of iterations or implement thinning of the draws to reduce the autocorrelation of the chain. In the case of the multivariate approach developed here, however, the prime concern is the failure to reach the posterior distribution. Because the candidates are drawn from the prior distribution, they may only partially cover the posterior distribution. Following, even though the accepted values belong to the posterior, the full distribution will not be recovered except if the prior fully overlaps with the posterior, which is not generally true.

For this reason, it is preferable to follow a period-by-period approach. Observe that (39) may rewrite:

$$\pi(\lambda_{i,t}|y, \lambda_{-i-t}) \propto exp\left(-\frac{1}{2}\left\{\lambda_{i,t} + s_i^{-1}\tilde{\lambda}_{i,t} \ \mathcal{Q}_{i,t}\right\}\right) exp\left(-\frac{1}{2}\frac{(\lambda_{i,t} - \hat{\lambda}_{i,t})}{\hat{\phi}_i}\right) \tag{43}$$

The procedure then consists in drawing a candidate from the the second term and use the first term to decide on acceptance or rejection. This again defines an independent transition kernel which yields an acceptance probability of:

$$p(a) = exp\left(-\frac{1}{2}\{(\lambda_{i,t}^{(n)} - \lambda_{i,t}^{(n-1)}) + s_i^{-1}(\tilde{\lambda}_{i,t}^{(n)} - \tilde{\lambda}_{i,t}^{(n-1)}) \ \mathcal{Q}_{i,t}\}\right) \tag{44}$$

The main difference between the multivariate approach and the period-by-period approach is that in the latter the candidates are not drawn from the prior distribution. As can be seen from (40), the mean value $\hat{\lambda}_{i,t}$ of the candidate distribution involves $\lambda_{i,t-1}$ and $\lambda_{i,t+1}$ which come directly from the posterior distribution. For this reason, a period-by-period algorithm will properly converge to the posterior distribution and cover its whole support.

---

[10]Alternatively, one may draw the candidates from $exp\left(-\frac{1}{2}\left\{\lambda_i'1_T + \lambda_i'\Phi_{i0}^{-1}\lambda_i\right\}\right)$ rearranged as a multivariate normal density, and use the remaining term $exp\left(-\frac{1}{2}\left\{s_i^{-1}\tilde{\lambda}_i' \ \mathcal{Q}_i\right\}\right)$ to decide on acceptance or rejection. This produces similar results.

In terms of efficiency, the performances of the algorithm are satisfactory. On the one hand, the iterative scheme of the procedure makes it intrinsically slower than the offset mixture methodology, but on the other hand the omission of the additional parameter r_i permits to spare some computations. In the end, the two algorithms are roughly equivalent on a per iteration basis. Also, the concern that Metropolis-Hastings produces repeated values is not to bee taken too seriously in this case: empirical applications suggest that the acceptance rate of the procedure typically exceeds 80%, which produces only few repetitions. The only significant difference between the two approaches resides in the number of iterations required for convergence. While the offset mixture typically obtains convergence in 2000 iterations or so, the Metropolis-Hastings algorithm is more sluggish and requires between 5000 and 10000 iterations to achieve convergence. This is due to the recursive nature of the algorithm which only allows for small moves of the chain for each sample period, the amplitude being restricted by the neighbouring values. In the end, the offset mixture approach remains the most efficient method, though the period-by-period Metropolis-Hastings algorithm constitutes an acceptable alternative.

4 Random inertia

In the basic version of the general time-varying model, the autoregressive coefficients ρ_i, γ_i and α_i associated with the dynamic processes in (10) are treated as exogenous hyperparmeters. The traditional choice in the literature consist in setting $\rho_i = \gamma_i = \alpha_i = 1$ which corresponds to the homogeneous random walk assumption. Rather, the calibration proposed for the general time-varying model uses $\rho_i = \gamma_i = \alpha_i = 0.90$, which reduces the inertia of the process and attributes some weight to the mean component. While this choice is reasonable and preferable to the usual random walk specification, it is in not necessarily optimal. For this reason, this section proposes a simple procedure to estimate endogenously the autoregressive coefficients ρ_i, γ_i and α_i.

4.1 Priors

The literature has produced a number of options to define the prior distributions of autoregressive coefficients. On the univariate side, the Beta distribution has sometimes been favoured for its support producing values between zero and one (Kim et al. (1998)). The Beta is however not conjugate with the normal distribution, which leads to an inefficient Metropolis-Hastings step in the estimation. On the multivariate side, a simpler alternative has consisted in using normal distributions (Primiceri (2005), Mumtaz and Zanetti (2013)). The prior is diffuse to let the data speak and produce posteriors centered on OLS estimates. While simple, this strategy is unadvisable for two reasons. First, as the support of the normal distribution is unrestricted, part of the posterior distribution may lie outside of the zero-one interval, which is not meaningful from an economic point of view. Second, the use of a diffuse prior is suboptimal as relevant information can be introduced at the prior stage. For these reasons, the prior is chosen here to be a truncated normal distributions with informative hyperparameters. Considering for instance ρ_i in (10), the prior distribution is a normal distribution with mean ρ_{i0} and variance π_{i0}, truncated over the [0 1] interval:

$$\pi(\rho_i) \sim \mathcal{N}_{[0\ 1]}(\rho_{i0}, \pi_{i0}) \tag{45}$$

An informative and reasonable prior belief consists in assuming that with 95% probability an autoregressive coefficient value should be comprised between 0.6 and 1. This is obtained by setting a mean value of $\rho_{i0} = 0.8$ and a standard deviation of 0.1, yielding a variance of $\pi_{i0} =$

18

0.01. This way, the prior is sufficiently loose to allow for significant differences in the posterior distributions of the different ρ_i's, but also sufficiently restrictive to avoid posteriors that would be too far away from the prior and implausible. Finally, the truncation operated at the prior stage ensures that the posterior distribution is restricted over the same range $[0\ 1]$, effectively ruling out irrelevant parts of the support. A similar strategy is applied to the other autoregressive coefficients in (10):

$$\pi(\gamma_i) \sim \mathcal{N}_{[0\ 1]}(\gamma_{i0}, \varsigma_{i0}) \qquad\qquad \pi(\alpha_i) \sim \mathcal{N}_{[0\ 1]}(\alpha_{i0}, \iota_{i0}) \qquad\qquad (46)$$

The mean and variance parameters are set to $\gamma_{i0} = \alpha_{i0} = 0.8$ and $\varsigma_{i0} = \iota_{i0} = 0.01$.

4.2 Bayes rule

Bayes rule must be slightly amended due to the inclusion of ρ_i, γ_i and α_i as random variables. The updated version obtains as:

$$\pi(\beta, \Omega, \rho, \lambda, \phi, \gamma, \delta^{-1}, \Psi, \alpha, r|y) \propto f(y|\beta, \lambda, \delta^{-1}, r) \left(\prod_{i=1}^{n} \pi(\beta_i|\Omega_i, \rho_i)\pi(\Omega_i)\pi(\rho_i) \right)$$

$$\times \left(\prod_{i=1}^{n} \pi(\lambda_i|\phi_i, \gamma_i)\pi(\phi_i)\pi(\gamma_i) \right) \left(\prod_{i=2}^{n} \pi(\delta_i^{-1}|\Psi_i, \alpha_i)\pi(\Psi_i)\pi(\alpha_i) \right) \left(\prod_{i=1}^{n}\prod_{t=1}^{T} \pi(r_{i,t}) \right) \qquad (47)$$

4.3 Posteriors

As for the basic model, the marginal posteriors are estimated from a Gibbs sampling algorithm relying on conditional distributions. For ρ_i, Bayes rule (47) implies $\pi(\rho_i|y, \rho_{-i}) \propto \pi(\beta_i|\Omega_i, \rho_i)\pi(\rho_i)$. From the priors (24) and (45) and some rearrangement, it follows that:

$$\pi(\rho_i|y, \rho_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\rho}_i, \bar{\pi}_i) \qquad\qquad \text{with:}$$
$$\bar{\pi}_i = (\ddot{\beta}'_{i,t-1}\ddot{\beta}_{i,t-1} + \pi_{i0}^{-1})^{-1} \qquad\qquad \bar{\rho}_i = \bar{\pi}_i(\ddot{\beta}'_{i,t-1}\ddot{\beta}_{i,t} + \pi_{i0}^{-1}\rho_{i0})$$
$$\ddot{\beta}_{i,t} = vec\left(\Omega_i^{-1/2\prime}(\beta_{i,2} - b_i \quad \cdots \quad \beta_{i,T} - b_i)\right) \qquad\qquad (48)$$

For γ_i, Bayes rule (47) implies $\pi(\gamma_i|y, \gamma_{-i}) \propto \pi(\lambda_i|\phi_i, \gamma_i)\pi(\gamma_i)$. From the priors (25) and (46) and some rearrangement, it follows that:

$$\pi(\gamma_i|y, \gamma_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\gamma}_i, \bar{\varsigma}_i) \qquad\qquad \text{with:}$$
$$\bar{\varsigma}_i = (\ddot{\lambda}'_{i,t-1}\ddot{\lambda}_{i,t-1} + \varsigma_{i0}^{-1})^{-1} \qquad\qquad \bar{\gamma}_i = \bar{\varsigma}_i(\ddot{\lambda}'_{i,t-1}\ddot{\lambda}_{i,t} + \varsigma_{i0}^{-1}\gamma_{i0})$$
$$\ddot{\lambda}_{i,t} = \phi_i^{-1/2}(\lambda_{i,2} \quad \ldots \quad \lambda_{i,T})' \qquad\qquad (49)$$

Finally for α_i, Bayes rule (47) implies $\pi(\alpha_i|y, \alpha_{-i}) \propto \pi(\delta_i^{-1}|\Psi_i, \alpha_i)\pi(\alpha_i)$. From the priors (25) and (46) and some rearrangement, it follows that:

$$\pi(\alpha_i|y, \alpha_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\alpha}_i, \bar{\iota}_i) \qquad\qquad \text{with:}$$
$$\bar{\iota}_i = (\ddot{\delta}_{i,t-1}^{-1}{}'\ddot{\delta}_{i,t-1}^{-1} + \iota_{i0}^{-1})^{-1} \qquad\qquad \bar{\alpha}_i = \bar{\iota}_i(\ddot{\delta}_{i,t-1}^{-1}{}'\ddot{\delta}_{i,t}^{-1} + \iota_{i0}^{-1}\alpha_{i0})$$
$$\ddot{\delta}_{i,t}^{-1} = vec(\Psi_i^{-1/2\prime}(\delta_{i,2}^{-1} - d_i^{-1} \quad \cdots \quad \delta_{i,T}^{-1} - d_i^{-1})) \qquad\qquad (50)$$

The other posteriors are unchanged.

4.4 MCMC algorithm

The Markov Chain Monte Carlo algorithm for the model with random inertia is fundamentally similar to the one of the general time-varying model. The 7 steps of Algorithm 1 are thus unchanged, but 3 additional steps specific to the model must be inserted between steps 6 and 7:

Algorithm 2: additional steps of the MCMC algorithm for the model with random inertia:

1. Sample ρ_i from $\pi(\rho_i|y, \rho_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\rho}_i, \bar{\pi}_i)$.

2. Sample γ_i from $\pi(\gamma_i|y, \gamma_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\gamma}_i, \bar{\varsigma}_i)$.

3. Sample α_i from $\pi(\alpha_i|y, \alpha_{-i}) \sim \mathcal{N}_{[0\ 1]}(\bar{\alpha}_i, \bar{\iota}_i)$.

5 Random mean

The standard version of the general time-varying model treats the mean parameters b_i, s_i and d_i^{-1} in (8) and (10) as exogenously supplied hyperparameters. While simple, this assumption may be overly restrictive. In particular, the preliminary conclusions obtained from Figure 1 suggest that the mean terms may be of considerable importance as they determine the share of data variation endorsed by each component of the model (dynamic coefficients, stochastic volatility and residual covariance). Endogenous estimation then constitutes a natural extension. While the univariate ARCH literature has paid some attention to this question in the context of stochastic volatility processes (Jacquier et al. (1994), Kim et al. (1998)), the subject has been almost completely neglected in multivariate models. One notable exception is the contribution of Chiu et al. (2015) who integrate a (period-specific) mean component to the dynamic variance of the residuals. This section fills the gap by proposing simple estimation procedures for the mean components of the dynamic processes.

5.1 Priors

For b_i, the choice for the prior is that of a simple multivariate normal distribution with mean b_{i0} and variance-covariance matrix Ξ_{i0}:

$$\pi(b_i) \sim \mathcal{N}(b_{i0}, \Xi_{i0}) \tag{51}$$

In the basic version of the general time-varying model, b_i is set according to its static OLS estimate $\hat{\beta}_i$, which constitutes a reasonable starting point. For this reason, the prior mean b_{i0} is set to $\hat{\beta}_i$ while the prior standard deviation is set to a fraction ϖ_i of this value, resulting in $\Xi_{i0} = diag((\varpi_i\hat{\beta}_i)^2)$. Small values of ϖ_i generate a tight and hence informative prior around $\hat{\beta}_i$ while larger values can be used to achieve diffuse and uninformative priors. Given the lack of economic theory concerning the equilibrium value of the time-varying coefficients, the prior is set to be informative but somewhat looser than usual in order to leave sufficient weight to the data. This is achieved by setting $\varpi_i = 0.25$, implying that b_i lies within 50% of $\hat{\beta}_i$ with 95% confidence.

Similar strategies are applied for s_i and d_i^{-1}. For the s_i which are positive scaling terms, the inverse Gamma represents a natural candidate. Specifically, the prior for each s_i is inverse Gamma with shape χ_{i0} and scale ϑ_{i0}:

$$\pi(s_i) \sim IG\left(\frac{\chi_{i0}}{2}, \frac{\vartheta_{i0}}{2}\right) \tag{52}$$

The hyperparameter values χ_{i0} and ϑ_{i0} are then chosen to imply a prior mean of $\hat{s}_i$, the OLS estimate used for the general time-varying model, and a prior standard deviation equal to a fraction ψ_i of this value.[11] . As a base case, ψ_i is set to 0.25 in order to generate, again, an informative but sufficiently loose prior.

Finally, the prior for each d_i^{-1} is multivariate normal with mean d_{i0}^{-1} and variance-covariance matrix Z_{i0}:

$$\pi(d_i^{-1}) \sim \mathcal{N}(d_{i0}^{-1}, Z_{i0}) \tag{53}$$

The prior mean is set as $d_{i0}^{-1} = \hat{d}_i^{-1}$, the OLS estimate used for the general time-varying model, and the the prior standard deviation is set to a fraction ϱ_i of this value, resulting in $Z_{i0} = diag((\varrho_i\hat{d}_i^{-1})^2)$. An informative but loose prior is achieved by setting $\varrho_i = 0.25$.

5.2   Bayes rule

Given the model, the updated version of Bayes rule is given by:

$$\pi(\beta, \Omega, b, \lambda, \phi, s, \delta^{-1}, \Psi, d^{-1}, r | y) \propto f(y|\beta, \lambda, s, \delta^{-1}, r) \left(\prod_{i=1}^{n} \pi(\beta_i|\Omega_i, b_i)\pi(\Omega_i)\pi(b_i)\right)$$

$$\times \quad \left(\prod_{i=1}^{n} \pi(\lambda_i|\phi_i)\pi(\phi_i)\right)\left(\prod_{i=1}^{n} \pi(s_i)\right)\left(\prod_{i=2}^{n} \pi(\delta_i^{-1}|\Psi_i, d_i^{-1})\pi(\Psi_i)\pi(d_i^{-1})\right)\left(\prod_{i=1}^{n}\prod_{t=1}^{T} \pi(r_{i,t})\right) \tag{54}$$

5.3   Posteriors

For b_i, Bayes rule (54) implies $\pi(b_i|y, b_{-i}) \propto \pi(\beta_i|\Omega_i, \rho_i)\pi(b_i)$. From the priors (24) and (51) and some rearrangement, it follows that:

$\pi(b_i|y, b_{-i}) \sim \mathcal{N}(\bar{b}_i, \bar{\Xi}_i)$ with:

$\bar{\Xi}_i = \left(\tilde{\tau}_i\Omega_i^{-1} + \Xi_{i0}^{-1}\right)^{-1}$ $\bar{b}_i = \bar{\Xi}_i\left(\Omega_i^{-1}(\tilde{\rho}_i \otimes I_k)\beta_i + \Xi_{i0}^{-1}b_{i0}\right)$

$\tilde{\tau}_i = \tau^{-1} + (1-\rho_i)^2(T-1)$ $\tilde{\rho}_i = \left(\tau^{-1} - (1-\rho_i)\rho_i \quad (1-\rho_i)^2 \quad \cdots \quad (1-\rho_i)^2 \quad (1-\rho_i)\right)$

$$\tag{55}$$

For s_i, Bayes rule (54) implies $\pi(s_i|y, s_{-i}) \propto f(y|\beta, \lambda, s, \delta^{-1}, r)\pi(s_i)$. From the likelihood function (35), the prior (52) and some rearrangement, it follows that:

$\pi(s_i|y, s_{-i}) \sim IG(\bar{\chi}_i, \bar{\vartheta}_i)$ with:

$\bar{\chi}_i = \dfrac{T + \chi_{i0}}{2}$ $\bar{\vartheta}_i = \dfrac{\tilde{\lambda}_i' \ Q_i + \vartheta_{i0}}{2}$ $\tag{56}$

---

[11]This is conveniently achieved by exploiting the fact that the inverse Gamma distribution defines a unique correspondence between any pair of mean/variance values and shape/scale parameters.

Finally for d_i^{-1}, Bayes rule (54) implies $\pi(d_i^{-1}|y, d_{-i}^{-1}) \propto \pi(\delta_i^{-1}|\Psi_i, d_i^{-1})\pi(d_i^{-1})$. From the priors (25) and (53) and some rearrangement, it follows that:

$$\pi(d_i^{-1}|y, d_{-i}^{-1}) \sim \mathcal{N}(\bar{d}_i^{-1}, \bar{Z}_i) \quad \text{with:}$$
$$\bar{Z}_i = \left(\tilde{\epsilon}_i \Psi_i^{-1} + Z_{i0}^{-1}\right)^{-1} \qquad \bar{d}_i^{-1} = \bar{Z}_i \left(\Psi_i^{-1}(\tilde{\alpha}_i \otimes I_{i-1})\delta_i^{-1} + Z_{i0}^{-1}d_{i0}^{-1}\right)$$
$$\tilde{\epsilon}_i = \epsilon^{-1} + (1-\alpha_i)^2(T-1) \quad \tilde{\alpha}_i = \left(\epsilon^{-1} - (1-\alpha_i)\alpha_i \quad (1-\alpha_i)^2 \quad \cdots \quad (1-\alpha_i)^2 \quad (1-\alpha_i)\right) \tag{57}$$

The other posteriors are unchanged.

5.4 MCMC algorithm

The Markov Chain Monte Carlo algorithm for the model with random mean is fundamentally similar to the one of the general time-varying model. The 7 steps of Algorithm 1 are thus unchanged, but 3 additional steps specific to the model must be inserted between steps 6 and 7:

Algorithm 3: additional steps of the MCMC algorithm for the model with random mean:

1. Sample b_i from $\pi(b_i|y, b_{-i}) \sim \mathcal{N}(\bar{b}_i, \bar{\Xi}_i)$.

2. Sample s_i from $\pi(s_i|y, s_{-i}) \sim IG(\bar{\chi}_i, \bar{\vartheta}_i)$.

3. Sample d_i^{-1} from $\pi(d_i^{-1}|y, d_{-i}^{-1}) \sim \mathcal{N}(\bar{d}_i^{-1}, \bar{Z}_i)$.

6 A case study on the Great Recession

6.1 Setup

To conclude this work, a short case study on the Great Recession is proposed. The study focuses on four major economies which have been severely impacted by the crisis: Canada, the Euro area, Japan, and the United States. The model considered is, again, the small economy model of Primiceri (2005). It comprises two lags, a constant, and 3 variables: unemployment and inflation, which represent the non policy block of the model, and a series of interest rate which represents the policy part of the setup. The data is quarterly. For each country, it starts in 1971q1 and runs up to 2010q1. Unemployment series are harmonised rates. Inflation is computed as a hundred times the difference of the log CPI with the log CPI of previous year at the same period. For Canada, the Euro area and the United States the interest rate is chosen as the 3-month interbank rate. For Japan the discount rate is retained, due to both unavailability of long series of interbank rates and the fact that the discount rate was used until recently as the main instrument of monetary policy. For Canada, Japan and the US the data comes from the OECD, while for the Euro area it is obtained from the Area Wide Model Database of Fagan et al. (2001) which has become the standard for academic research on the Euro Area.

22

The aim of the exercise consists in assessing the forecast performances of different models for key periods of the crisis. Figure 2 displays the growth rate of GDP [12] for the four considered economies over the Great Recession periods. For each country, three critical periods of the crisis are considered. The first period is the recession period, the period at which the country enters into negative growth. For Canada, the Euro area, Japan and the United States, this respectively occurs in 2009q1, 2008q4, 2008q2 and 2008q3. The second period considered is the reversion period, the period at which GDP growth reaches its minimum before it starts increasing again. This respectively happens in 2009q3, 2009q1, 2009q1 and 2009q2. The final period considered is the recovery period, the period at which the economy hits positive growth again. For the four countries considered, this happens in 2010q1. These periods are of special importance for policy makers as they corresponds to the points where the crisis respectively initiates, reverts and ends. It is crucial to anticipate them correctly in order to provide an adequate answer to the rapidly changing economic conditions.
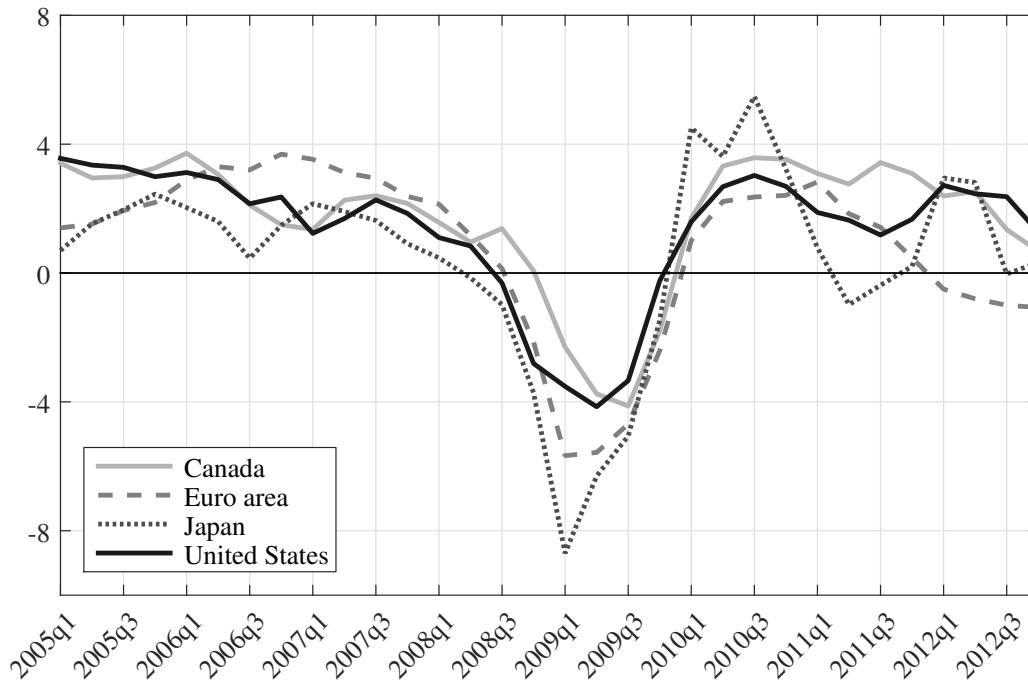
Figure 2: Year-on-year GDP growth for the four major economies

The forecasting exercise is performed in pseudo real time, that is, it does not use information which is not available at the time the forecast is made. For this reason, for each country and each considered period of the crisis the model is estimated from 1971q1 up to the period preceding the forecast period. The forecast is then obtained as the one period-ahead out-of-sample prediction. For each forecast, two criteria are considered. The first criterion is the classical Root Mean Squared Error (RMSE) which considers the accuracy of point forecasts. Denoting by $\tilde{y}_{t+h}$ the h-step ahead prediction and by y_{t+h} the realised value, it is defined as:

---

[12]Estimates are year-on-year growth rates from chained volume series, obtained by taking a hundred times the difference of the log value with the log value of previous year at the same period.

$$RMSE_{t+h} = \sqrt{\frac{1}{h}\sum_{i=1}^{h}(\tilde{y}_{t+h} - y_{t+h})^2} \tag{58}$$

The second criterion is the Continuous Ranked Probability Score (CRPS) of Gneiting and Raftery (2007) which evaluates density forecasts. As pointed by those authors, this criterion presents advantages over alternative density scores such as the log score as it rewards more density points close to the realised value and is less sensitive to outliers. Denoting by F the cumulative distribution function of the h-step ahead forecast density and by $\hat{y}_{t+h}$ and $\hat{y}'_{t+h}$ independent random draws from this density, the CRPS is defined as:

$$CRPS_{t+h} = \int_{-\infty}^{\infty}(F(x) - \mathbb{1}(x \geq y_{t+h}))^2 dx = \mathbb{E}\,|\hat{y}_{t+h} - y_{t+h}| - \frac{1}{2}\,\mathbb{E}\,|\hat{y}_{t+h} - \hat{y}'_{t+h}| \tag{59}$$

Empirical estimations of the CRPS use the second term on the right-hand side of (59), approximated from the Gibbs sampler draws. For both criteria, a lower score indicates a better performance.

Finally, the exercise considers five competing models. The first model is the homogenous random walk (Hrw) specification of Primiceri (2005), which obtains as a special case of the general time-varying model by setting the autoregressive coefficients of the dynamic processes to $\rho_i = \gamma_i = \alpha_i = 1$ for all $i = 1, \cdots, n$.[13] The second model is the general time-varying (Gtv) model developed in section 2, following the calibration proposed in section 3.1. The third and fourth models respectively consist in the general time-varying model augmented by the random inertia (Ri) extension developed in section 4 and the random mean (Rm) extension developed in section 5. The last model combines the two extensions, thus adding both random inertia and random mean (Rim) to the general time-varying model.

6.2 Results

Table 2 reports the results of the experiment for unemployment. The results for inflation and the interest rate are fairly similar and relegated to Appendix B in order to save space. Table 2 comprises three parts: the top, middle and bottom panels which respectively correspond to forecast evaluations for the recession, reversion and recovery periods. In each panel, the left part summarizes the results for the RMSE while the right part reports the results for the CRPS. For each country and each criterion, the bold entry corresponds to the model achieving the best forecast performance among the five competitors. Also, a shaded entry for the general time-varying model (second entry) indicates that the model performs better than the homogenous random walk formulation (first entry). In total, the experiment considers three periods, three variables, four countries and two evaluation criteria for a total of 72 forecast evaluations, each carried on a set of five competing models.

---

[13]Beyond the random walk assumption on the dynamic parameters, there exist a number of additional differences in the specification of the model of Primiceri (2005) and the present general time-varying model. As the objective of the exercise consists primarily in assessing the impact of the random walk assumption, a meaningful comparison necessitates that the competing models are similar in every other respect. For this reason, the exercise is conducted with the general time-varying model specified a random walk rather than with the actual model of Primiceri (2005). For the sake of robustness, the exercise was replicated with the code of Primiceri (2005). The results are similar to those obtained in section 6.2, the main difference being found in the recovery phase for which the Primiceri (2005) model performs slightly better than the general time-varying formulated as a random walk.

| Unemployment: recession period | | | | | | | | | |
| | RMSE | | | | | CRPS | | | | |
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	1.120	**0.808**	0.870	0.924	0.965	0.944	**0.594**	0.705	0.732	0.806
Euro area	0.299	**0.291**	0.331	0.316	0.339	0.198	**0.185**	0.231	0.212	0.235
Japan	0.066	**0.005**	0.026	0.013	0.023	0.047	0.041	0.038	0.038	**0.037**
United States	0.456	**0.406**	0.445	0.461	0.491	0.290	**0.247**	0.316	0.309	0.363

| Unemployment: reversion period | | | | | | | | | |
| | RMSE | | | | | CRPS | | | | |
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.853	0.470	0.296	0.512	**0.290**	0.606	0.281	0.176	0.319	**0.173**
Euro area	0.632	**0.376**	0.470	0.474	0.516	0.450	**0.252**	0.353	0.341	0.400
Japan	**0.416**	0.438	0.443	0.461	0.445	**0.299**	0.329	0.347	0.352	0.349
United States	0.089	0.284	**0.018**	0.044	0.166	0.171	0.179	**0.078**	0.110	0.116

| Unemployment: recovery period | | | | | | | | | |
| | RMSE | | | | | CRPS | | | | |
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	**0.095**	0.186	0.171	0.119	0.135	**0.103**	0.140	0.118	0.111	0.110
Euro area	0.154	0.050	**0.028**	0.034	0.037	0.098	0.071	**0.056**	0.065	0.060
Japan	0.060	0.136	0.028	0.108	**0.000**	0.067	0.086	0.051	0.075	**0.050**
United States	0.303	0.222	**0.191**	0.267	0.214	0.192	0.158	**0.132**	0.169	0.138

Table 2: Forecast evaluation criteria for unemployment

Overall, four main conclusions can be drawn from these results. First and foremost, the results unambiguously disqualify the homogenous random walk as the best formulation regarding forecast accuracy. When considering the homogenous random walk against the general time-varying model only (the shaded entries in the tables), the latter performs better in 57 out of 72 evaluations, that is, in 79% of the cases. When all the five competing models are considered (the bold entry providing the best performance), the homogenous random walk is over-performed in 64 out of 72 evaluations, or 89% of the cases. This questions the common belief that a random walk formulation may represent an optimal, or even reasonably efficient representation for the dynamics of time-varying coefficient models. Looking more closely at the results, the relative inefficiency of the homogenous random walk assumption appears to be consistent across variables, countries and periods. The only significant difference appears in the case of inflation for which the random walk performs especially poorly and never represents the optimal formulation. Also, the RMSE and the CRPS select almost systematically the same candidate, which demonstrates that the results are not affected by the evaluation being conducted on the point estimate only or on the whole density. Overall, this further supports the robustness of the results.

Second, the gains in forecast accuracy resulting from relaxing the random walk assumption are substantial. Table 3 reports the average forecast accuracy gain from adopting the general time-varying model compared to the homogenous random walk assumption, for the three variables of the model and both in terms of RMSE and CRPS. For each variable and each criterion, the average accuracy gain is defined as the ratio of the sums of the criteria under the general time-varying model over the sum of the criteria under the random walk specification. For unemployment and the interest rate, the gains are significant. For unemployment, the general time-varying model results in an average 19% cut in the RMSE and 26% decline in the CRPS. The gain is slightly more modest for the interest rate, with average drops of 15% in RMSE ad 16% in CRPS. Overall, those gains are already considerable. The gains however become remarkably high in the case inflation where they attain 44% for the CRPS and exceed 50% for the RMSE. It is not clear what the reasons are for these significant differences in gain across variables, but this demonstrates that any variable may benefit from adopting the general time-varying model, though in general the gain may vary considerably from one variable to another.

	RMSE	CRPS
Unemployment	19.17 %	26.03 %
Inflation	50.09 %	44.06 %
Interest rate	15.90 %	16.13 %

**Table 3: Average forecast accuracy gain
compared to random walk formulation**

Third, the random inertia and random mean extensions significantly contribute to improve forecast accuracy. Over the 64 evaluations where the homogenous random walk specification is dominated, 61% of the cases (39 evaluations) retain either random inertia, random mean or both as the optimal model. Stated differently, in almost two third of the cases forecasts can be improved by adopting one extension or both. This is not surprising as the mean and autoregressive coefficient values used by the general time-varying model constitute reasonable starting points, but in no way optimal candidates. Leaving some freedom to the data to endogenously determine these parameters may logically contribute to produce estimates that are closer to the actual data generating process, resulting in better forecasts.

Fourth and last, there does not seem to exist a hierarchical pattern between random inertia, random mean or their combination. Over all forecast evaluations, random inertia constitutes the best formulation 13 times, random mean 11 times, and the combination of both 15 times. These results are not sufficiently different one from each other to assert that one extension proves better than the other. This suggests that which extension is the best may be a dataset-specific matter, and perhaps even a variable-specific one. For a given model, it might therefore be advisable to conduct testing before an extension is retained in order to assess the suitability of this extension for the considered setting.

## 6.3 Crisis analysis with random inertia and random mean

This brief section proposes an economic analysis of the Great Recession from the perspective of random mean and random inertia. It outlines the differences - both in terms of results and interpretations - of the Primiceri (2005) approach compared to the general time-varying model augmented by the random mean and random inertia extensions.

The exercise focuses on the United States, using again the small economy model of Primiceri (2005). The sample is extended to 2017q4 in order to provide a broader view of the Great Recession by including its aftermath. The first 40 periods of the dataset (1971q1-1980q4) are used to generate the training sample required for the Primiceri (2005) methodology, so the estimation is conducted from 1981q1 onwards for both models. The number of lags is reduced to one in order to provide direct interpretation to the estimated VAR coefficients. In particular, the shocks associated to the three variables of the model (unemployment, inflation and the interest rate) can be readily interpreted as real, nominal and monetary policy shocks. The dynamic coefficients can then be interpreted as the impulse response at impact of each variable to these shocks. Finally, the random inertia and random mean hyperparameters are increased to $\pi_{i0} = \varsigma_{i0} = \iota_{i0} = \varpi_i = \psi_i = \varrho_i = 1$ in order to implement a slightly loser prior and attribute increased weight to data information.

Figure 3 reports the volatility estimates for the unemployment and interest rate shocks. The left panels provide the estimates obtained from the Primiceri (2005) model. In these panels the dashed lines represent the empirical sample mean, which provides an approximation of the long-run volatility level. The right panels provide the volatility estimates obtained from the random mean extension. The dashed line in the plots represents the theoretical equilibrium level s_i, endogenously estimated by random mean. These equilibrium values are also reported in Table 4.

The left panels reveal that the stochastic volatility components of the Primiceri (2005) model are characterised by large shocks and wide amplitudes of fluctuations. The figure suggests indeed that the long-run levels of volatility are quite high: around 0.24 for real activity shocks, and 0.38 for monetary policy shocks. More importantly, the deviations from the equilibrium levels become considerable in a context of crisis. For the Great Recession, the peak in volatility for unemployment culminates at about four times the long-run level, while monetary policy shocks attain roughly twice their equilibrium value.

By contrast, the estimates obtained from random mean on the right panels suggest that the volatility levels endogenously estimated with random mean are considerably smaller than the one produced with the Primiceri (2005) approach. The equilibrium level for unemployment shocks is at about 0.017, more than 14 times smaller than the value of 0.241 obtained with the Primiceri (2005) methodology. The same holds for the long-run levels of inflation and interest rate shocks which take values of 0.161 and 0.067, respectively more than 3 and 5 times smaller than their Primiceri (2005) counterparts. Table 4 shows that in general, the data does not endorse the Primiceri (2005) estimates: the posterior estimates obtained from random mean prove considerably smaller than the prior means, rather than tending to the larger Primiceri (2005) values, if the latter were correct.
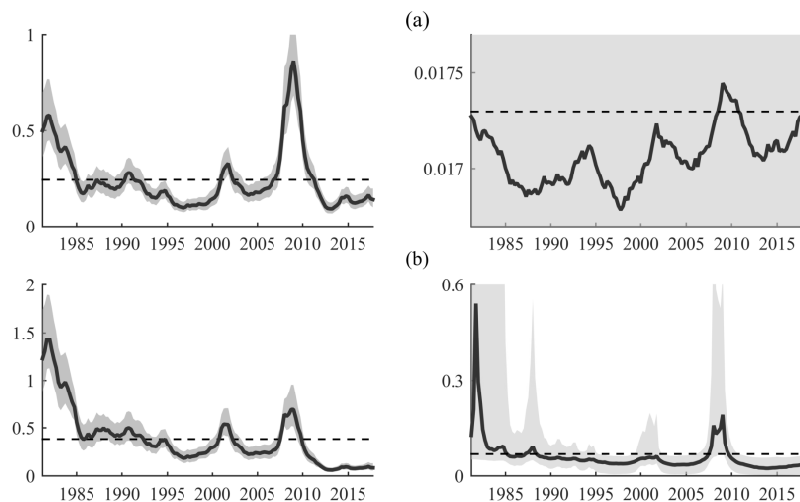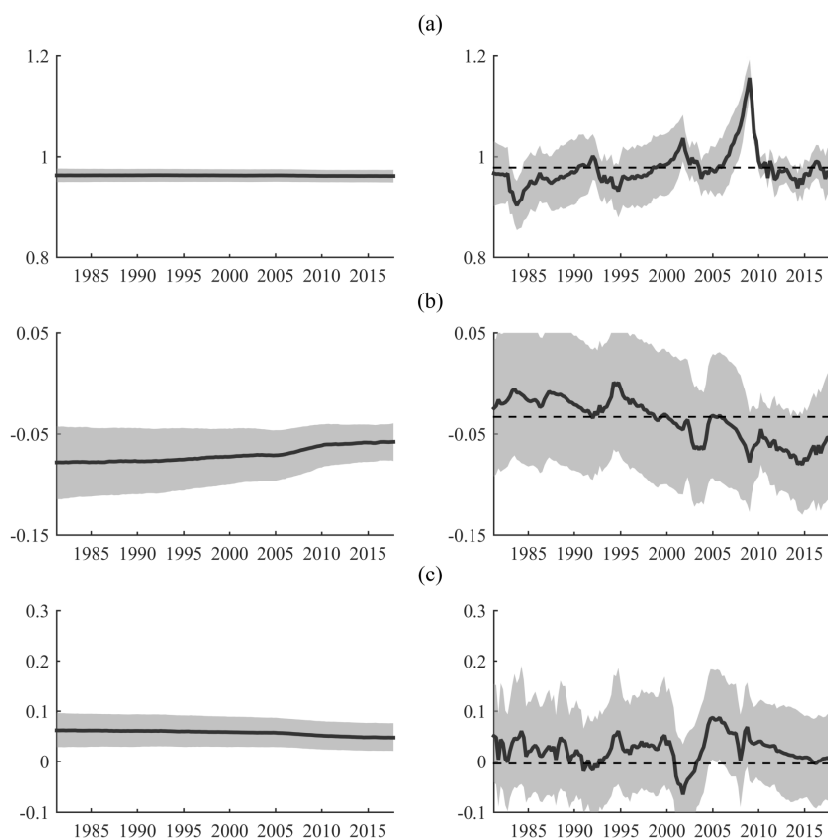
Figure 3: Volatility of (a) unemployment shocks, (b) interest rate shocks
left panels: Primiceri (2005), right panels: random mean

	Unemployment	Inflation	Interest rate
Primiceri (2005): empirical posterior mean	0.241	0.568	0.381
Random mean: prior	0.092	0.406	0.307
Random mean: posterior	0.017	0.161	0.067

Table 4: Summary of random mean estimates for s_i (United States)

The amplitude of the fluctuations in volatility obtained with random mean is also substantially smaller than those suggested by the Primiceri (2005) methodology. While interest rate shocks did double in size with both methodologies during the Great Recession, the absolute magnitude of increase with random mean is limited to a mere 0.1, against 0.4 with the Primiceri (2005) approach. As for unemployment shocks, random mean indicates that they increased only marginally during the Great Recession, rising by a mere 3% (from 0.017 to 0.0175) and in fact hardly exceeding the long-run level, far from the fourfold rise observed with the Primiceri (2005) model.

Consider then the dynamic coefficients of the model. Figure 4 display the dynamic responses of unemployment (to unemployment shocks) and the interest rate (to unemployment and interest rate shocks). The left panels display the estimates obtained from the Primiceri (2005) model and the right panels the estimates from the general time-varying model augmented with random mean. On these panels, the dashed line represents the theoretical equilibrium value b_i endogenously estimated by random mean.

28

Figure 4: Dynamic response of (a) unemployment to unemployment shocks, (b) interest rate to unemployment shocks, (c) interest rate to inflation shocks
left panels: Primiceri (2005), right panels: random mean

The Primiceri (2005) estimates displayed on the left panels suggest that the changes in the dynamic responses of the economy have been negligible. Panel (a) reveals a virtually flat response of unemployment to real shocks over the whole sample. In other words, the propagation mechanism of the real side of the economy has remained remarkably stable, even during the periods of recession. The same conclusion holds for Monetary policy. Panels (b) and (c) show that monetary policy hardly experienced any adjustment over the sample. The only noticeable evolution consists in a moderate increases in the reaction of the interest to real shocks from 2007 on and a simultaneous marginal cut in the reaction to inflation.

By contrast, a substantial level of fluctuation obtains under the random mean model. Concerning the response of unemployment to real shocks, panel (a) reveals a sudden departure from the long-run level during the Great Recession to reach a peak at more than 20% over the equilibrium level in 2009, at the worse of the crisis. The response then returned to its equilibrium value in 2010, after the end of the crisis. According to panels (b) and (c), monetary policy also adjusted significantly during the crisis. Considering the response of the interest rate to real shocks, the right panel in (b) shows that from 2007 on the FED had been placing its response below its long-run level, continuously loosening its stance regarding real activity over the Great Recession. It stiffened its response again from 2009 on as the economy initiated its recovery, before initiating a new episode of accommodating policy from 2010 on.

29

The results for the response of the interest rate to inflation shocks are quite singular. They reveal that even during the Great Recession, the FED maintained a primarily anti-inflationary stance. From 2004 on, its started stiffening its response to inflation, pushing it above its long-run level. It then steadily loosened it until late 2008, to the point where it went back to its long-run level. But then, rather than accommodating inflation further as the crisis was reaching its worse, the FED pushed again its reaction to inflation upward. In fact it maintained it durably above the equilibrium level, reducing it gradually from 2009 onward to return to its equilibrium level only in 2015.

Critical to the analysis is also the degree of inertia characterising the dynamic responses of the economy. A possible explanation for the lack of variability of the dynamic coefficients estimated with the Primiceri (2005) approach is the sluggishness involved by the random walk formulation. This formulation seems at odd with the high frequency movements observed for the coefficients of the general time-varying model. To answer this question, Table 5 reports the estimates of the autoregressive coefficients obtained with the random inertia extension. The results confirm that the data does not support the random walk formulation of Primiceri (2005). While the volatility processes are unambiguously characterised by a near random walk process, the dynamic coefficients experience a much lower degree of inertia, with autoregressive coefficients comprised between 0.47 and 0.71. This suggests that the changes in the dynamic responses of the economy may in fact be very swift as little weight is attributed to past disturbances.

In general, the results of Table 4 and Table 5 also confirm the validity of the variable-specific approach. Be it for the mean level of the dynamic processes or their autoregressive coefficients, the random mean and random inertia extensions unambiguously demonstrate that different variables experience significantly different dynamics. They also reveal that these singularities affect profoundly the results and the ensuing analysis.

	Unemployment	Inflation	Interest rate
VAR coefficients	0.71	0.47	0.53
Shock volatility	0.99	0.97	0.99
Shock covariance		0.54	0.38

Table 5: Summary of random inertia estimates (United States)

Gathering all this information, two different narratives of the Great Recession emerge, depending on the model considered. The Primiceri (2005) approach yields the first narrative, which is consistent with the mainstream interpretation of the Great Recession. This narrative considers the crisis primarily from the perspective of structural shocks. In this scenario, the Great Recession was initiated by a series of large adverse financial disturbances which severely affected the real side of the economy (Figure 3, left panel (a)). As the Housing bubble burst in late 2007, banks and financial institutions incurred heavy losses in asset value leading them to either bankruptcy or being bailed out by the government. Coupled with the strong decline in the housing construction sector, economic activity was dragged down and entered into recession. In this narrative, the dynamic response of the corporate and private side of the economy was left essentially unchanged (Figure 4, left panel (a)), the shocks constituting the principal drivers of the crisis. Similarly, the monetary authorities maintained a rather conservative stance and did

not alter regular monetary policy profoundly (Figure 4, left panels (b) and (c)). Rather, they focused on protecting the banks against the risks of exposure to bad mortgages, intervening by the way of unconventional policies to restore the flow of liquidities. The policies consisted in expended access to the FED discount window, and the settlement of temporary programmes like the Term Auction Facility (TAF) and the Troubled Asset Relief Program (TARP). Such policies can then be interpreted as a series of expansionary monetary policy shocks in an attempt to offset the impact of the negative financial shocks (Figure 3, left panel b)).

With this interpretation the Great Recession becomes essentially exogenous, conditioned by large-sized adverse shocks triggering and prolonging the crisis. Recovery is then understood as an automatic process resulting from the return of the shocks to their equilibrium levels, possibly with the support of positive shocks from the monetary authorities.

A different narrative obtains however from the random mean and random inertia extensions. In this alternative scenario, three actors become involved in the crisis: the shocks, the response of the monetary authorities, and the response of the private sector. As the Housing bubble burst in late 2007, a series of adverse real disturbances impacted negatively the banking and housing sectors of the economy, initiating the financial crisis. These shocks however remained of moderate size and in fact barely exceeded their long-run levels (Table 4 and Figure 3, right panel (a)). As a consequence the installing crisis may have remained mild, had the shocks been properly accommodated.

The response of the FED however proved inadequately restrictive: the monetary authorities maintained a very strong stance against inflation, perceived as the major risk (Figure 4, right panels (b) and (c)). For this reason the nominal interest rate experienced a fairly slow decline, the FED taking no less than eight quarters to eventually cut it below 0.5% (from 5.4% at the beginning of the financial crisis in 2007q3 to 0.3% in 2009q3, as the crisis ended). The effects of this restrictive monetary policy ended up impacting the real side of the economy as well. As investment banks fell short of liquidities, they disturbed the flow of credit to consumers and corporations, reducing aggregate demand and nominal GDP. By preventing the normal level of transactions to be conducted, the FED policy contributed to turn a financial crisis initially triggered by the housing slump into a general macroeconomic crisis affecting the response of the entire corporate sector (Figure 4, right panels (a)). The propagation of the crisis to the entire private sector proved fast, due to the low persistence of past disturbances in the dynamics of the real side of the economy (Table 5).

In this narrative, the adverse financial shocks merely acted as the trigger for the recession, far from representing the main culprit as suggested by the Primiceri (2005) approach. Also, while the unconventional monetary policy conducted by the FED might have had some supportive effects, the latter remained most likely modest due to the small size of the implied shocks (Figure 3, right panel (b)). In this scenario, the main drivers of the Great Recession consist in fact in the lack of reactivity of the Federal Reserve, and the overall fall in aggregate demand and nominal GDP faced by the private sector. Such a narrative is not mainstream and remains controversial. It is nevertheless supported by the data and consistent with the findings of certain economists who emphasize the role and responsibility of the FED in the events of the Great Recession (see for instance Beckworth (2012) and Sumner (2017)).

7 Conclusion

This paper introduces a general methodology for time-varying Bayesian VAR models. The model takes a variable-specific approach and provides general autoregressive formulations for the dynamic parameters, in contrast to the random walk assumption used by the canonical approach of Primiceri (2005). The developed estimation methodology is simple and yields significant efficiency gains compared to the standard Carter and Kohn (1994) procedure. From a case study on four countries during the Great Recession, overwhelming evidence is found against the homogenous random walk formulation of Primiceri (2005). In general, it is shown that forecast accuracy can be significantly improved by adopting the general time-varying model and its extensions, random mean and random inertia. The results obtained from these extensions also suggest that contrary to the common belief, stochastic volatility plays only a marginal role in time-varying VAR models while the dynamic coefficients represent the main source of time variation.

References

Aastveit, K. A., Carriero, A., Clark, T. E., and Marcellino, M. (2017). Have standard VARs remained stable since the crisis? *Journal of applied econometrics*, 32(5):931–951.

Abbate, A. and Marcellino, M. (2018a). Modelling and forecasting exchange rates with time-varying parameter models. *Journal of the Royal Statistical Society*, forthcoming.

Abbate, A. and Marcellino, M. (2018b). Point, interval and density forecasts of exchange rates with time varying parameter models. *Journal of the Royal Statistical Society*, 181(1).

Baumeister, C. and Benati, L. (2010). Unconventional monetary policy and the great recession: estimating the impact of a compression in the yield spread at the zero lower bound. Working Paper Series 1258, European Central Bank.

Beckworth, D. (2012). *Boom and Bust Banking: The Causes and Cures of the Great Recession*. Independent Institute.

Benati, L. and Lubik, T. (2014). The time-varying beveridge curve. In van Gellecom, F. S., editor, *Advances in Non-linear Economic Modeling. Dynamic Modeling and Econometrics in Economics and Finance*, volume 17. Springer.

Bijsterbosch, M. and Falagiarda, M. (2014). Credit supply dynamics and economic activity in euro area countries: a time-varying parameter var analysis. Working Paper Series 1714, European Central Bank.

Canova, F. (1993). Modelling and forecasting exchange rates with a bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control*, 17:233–261.

Carter, C. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3).

Chan, J. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of applied econometrics*.

Chan, J. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chib, S. and Griffiths, W. (2008). *Bayesian Econometrics*. Emerald Group Publishing.

Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371.

Chib, S., Nardarib, F., and Shephard, N. (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108.

Chiu, C.-W., Mumtaz, H., and Pinter, G. (2015). Forecasting with VAR models: fat tails and stochastic volatility. Bank of England working papers 528, Bank of England.

Ciccarelli, M. and Rebucci, A. (2003). Measuring contagion with a bayesian, time-varying coefficient model. Working Paper Series 263, European Central Bank.

Clark, T. (2011). Real-time density forecasts from BVARs with stochastic volatility. *Journal of Business and Economic Statistics*, 29(3):327–341.

Clark, T. and Ravazzolo, F. (2015). The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility. *Journal of applied econometrics*, 30(4):551–575.

Cogley, T. (2001). How fast can the new economy grow? a bayesian analysis of the evolution of trend growth. ASU Economics Working Paper 16/2001, ASU.

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.

D'Agostino, A., Gambetti, L., and Giannone, D. (2011). Macroeconomic forecasting and structural change. *Journal of applied econometrics*, 28(1):82–101.

Del Negro, M. and Primiceri, G. E. (2015). Time-varying structural vector autoregressions and monetary policy: a corrigendum. *Review of Economic Studies*, 82(4):1342–1345.

Doan, T., Litterman, R., and Sims, C. (1983). Forecasting and conditional projection using realistic prior distributions. Working Paper Series 1202, NBER.

Doh, T. and Connolly, M. (2013). The state space representation and estimation of a time-varying parameter var with stochastic volatility. In Zeng, Y. and Wu, S., editors, *State-Space Models. Statistics and Econometrics for Finance*, volume 1. Springer.

Eisenstat, E., Chan, J., and Strachan, R. (2016). Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews*, 35(8-10).

Eisenstat, E. and Strachan, R. (2016). Modelling inflation volatility. *Journal of applied econometrics*, 31(5).

Ellington, M., Florackis, C., and Milas, C. (2017). Liquidity shocks and real GDP growth: Evidence from a bayesian time-varying parameter VAR. *Journal of International Money and Finance*, 72:93–117.

Fagan, G., Henry, J., and Mestre, R. (2001). An area-wide model (AWM) for the euro area. Working Paper Series 42, European Central Bank.

Gali, J. and Gambetti, L. (2008). On the sources of the great moderation. NBER Working Papers 14171, National Bureau of Economic Research, Inc.

Gambetti, L. and Musso, A. (2017). Loan supply shocks and the business cycle. *Journal of Applied Econometrics*, 32:764–782.

Gerba, E. and Hauzenberger, K. (2013). Estimating US fiscal and monetary interactions in a time varying var. Studies in economics, School of Economics, University of Kent.

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gorgi, P., Koopman, S. J., and Schaumburg, J. (2017). Time-varying vector autoregressive models with structural dynamic factors. Working paper.

Hamilton, J. (1994). *Time-series analysis*. Princeton University Press.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–264.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–417.

Jacquier, E., Polson, N., and Rossi, P. (1995). Models and priors for multivariate stochastic volatility. CIRANO Working Papers 95s-18, CIRANO.

Jacquier, E., Polson, N., and Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212.

Kalli, M. and Griffin, J. (2018). Bayesian nonparametric vector autoregressive models. *Journal of econometrics*, 203:267–282.

Kavtaradze, L. and Mokhtari, M. (2018). Factor models and time-varying parameter framework for forecasting exchange rates and inflation: a survey. *Journal of Economic Surveys*, 32(2).

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393.

Koop, G. and Korobilis, D. (2012). Large time-varying parameter VARs. Working Paper 2012-04, Business School - Economics, University of Glasgow.

Korobilis, D. (2013). Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics*, 75(2):157–179.

Lubik, T. A. and Matthes, C. (2015). Time-varying parameter vector autoregressions: Specification, estimation, and an application. *Economic Quarterly*, 101(4):323–352.

Mumtaz, H. and Sunder-Plassmann, L. (2010). Time-varying dynamics of the real exchange rate. a structural VAR analysis. Working Paper 382, Bank of England.

Mumtaz, H. and Zanetti, F. (2013). The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking*, 45(4):535–558.

Poon, A. (2018). Assessing the synchronicity and nature of australian state business cycles. Working paper, University of Strathclyde.

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

Raftery, A. and Lewis, S. (1992). How many iterations in the gibbs sampler? In Bernardo, J., Berger, J., Dawid, A., and Smith, F., editors, *Bayesian Statistics*, pages 763–773. (Oxford: Oxford University Press.

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Stock, J. and Watson, M. (1996). Evidence on structural instability in macroeconomic time-series relations. *Journal of Business and Economic Statistics*, 14(1):11–30.

Stock, J. and Watson, M. (2012). Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity*, 1:81–135.

Sumner, S. (2017). Monetary policy rules in light of the great recession. *Journal of Macroeconomics*, 54(PA):90–99.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, 65(1):59–74.

Appendix A Convergence of the MCMC algorithm

Because the estimation methodology of this paper differs from the standard methodology developed by Carter and Kohn (1994), it is important to check that the MCMC algorithm converges properly to the posterior distribution. Kim et al. (1998) suggest that 2000 burn-in iterations are sufficient when using the normal offset mixture approximation of the likelihood function. This also corresponds to the number of initial iterations discarded by Primiceri (2005) for his small model of the U.S economy. As a base case, the model is thus estimated on the basis of 10000 total simulations, including 2000 initial iterations trimmed as burn-in sample.

To check for convergence, the chain is first started from different starting points (randomly selected), yielding similar results. The estimation results are also not affected by increasing the number of burn-in and post burn-in iterations (attempts with 20000 burn-in iterations and 50000 post burn-in iterations, retaining all the draws or only one over ten, produce results with no visible differences). As a second set of checks, and following Primiceri (2005), the model is then submitted to 3 different types of tests. The first and second tests respectively focus on the autocorrelation functions and inefficiency factors of the draws, while the third test consists in the diagnostic test of Raftery and Lewis (1992) for the total number of runs of the algorithm. Given the 153 sample points, 3 variables and 2 lags, the total number of dynamic parameters β_i, λ_i and δ_i^{-1} to estimate is 4131. Therefore, for the sake of saving space the results are only presented for these parameters, leaving aside the static variance-covariance hyperparameters Ω_i, ϕ_i and Ψ_i.

Figure 5 presents the results for the different diagnostics. For the three panels, the leftmost part of the plot (3213 points) represents the results for the dynamic VAR coefficients β, while the right and rightmost parts (459 points each) respectively provide the results for the volatility and inverse covariance terms λ and δ^{-1}. The 20th-order autocorrelation functions are summarised in panel (a). Low autocorrelation functions are usually interpreted as evidence that the draws are almost independent. All the values remain below 0.25, and, in fact most values don't even exceed 0.10, particularly for the VAR coefficients. The second diagnostics consists in the calculation of the inefficiency factors, presented in panel (b). Following Chib and Griffiths (2008), the inefficiency factor is computed as $1 + 2\sum_{k=1}^{K} \rho_k$, where ρ_k represents the kth-order autocorrelation function and K is chosen as the value at which the autocorrelation functions tapper off, from a 4% tapered window. Inefficiency factor values around 20 or below are considered satisfactory. Again, a vast majority of the values are around or below 20. Further information about the inefficiency factors is provided in Table 6. The most inefficient coefficients are the volatility terms λ for which the inefficiency factors commonly reach a value of 30, a value which remains nevertheless reasonably close to 20. For the other two sets of coefficients, 90% of the estimated values are below 25, with remarkably low inefficiency factors for the VAR coefficients β.

	Median	Mean	Min	Max	10th percentile	90th percentile
β	1.13	1.55	0.95	8.21	1.03	2.82
λ	9.45	13.37	2.19	64.59	4.69	30.4551
δ^{-1}	7.06	9.59	1.70	52.02	2.43	22.09

Table 6: Summary of the distribution of the inefficiency factors for the dynamic parameters
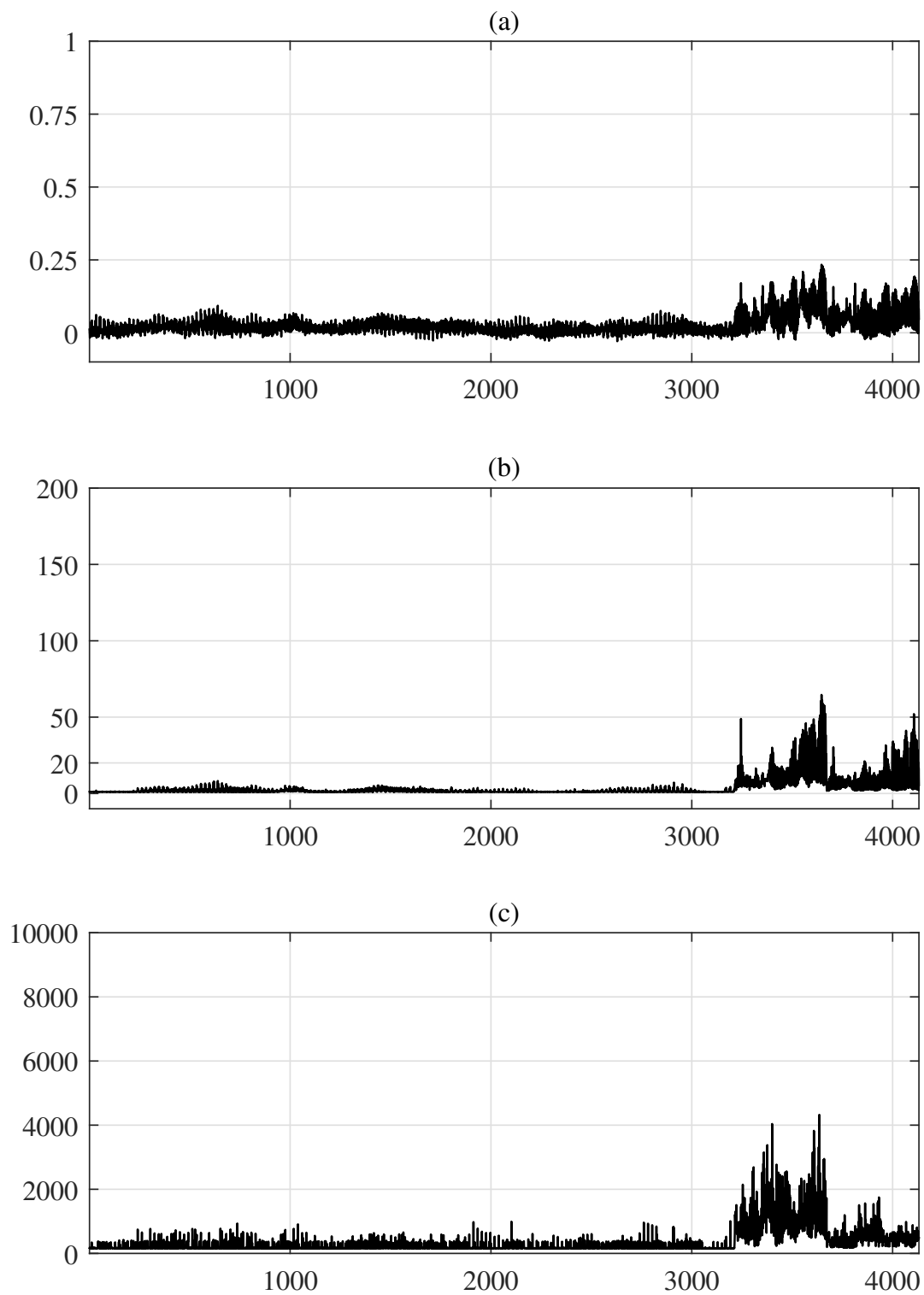
Figure 5: Convergence diagnostics for the time-varying parameters
(a) 20-th-order sample autocorrelation, (b) inefficiency factors, (c) Raftery and Lewis's total number of runs

As a final diagnostics, panel (c) plots the Raftery and Lewis (1992) total number of runs required to achieve a given level of accuracy. The parameters are set as follows: quantile = 0.025; required precision = 0.025; probability of attaining the required precision = 0.95. The total number of runs is always far below the number of 10000 iterations used to conduct the estimation, and suggests that 4000 iterations might in fact be sufficient. Again, the number of draws required to achieve convergence for the VAR coefficients β appears to be particularly low.

Overall, given the complexity and high dimensionality of the model, the convergence checks are satisfactory and provide evidence that convergence to the posterior distribution is efficient and properly achieved.

Appendix B   Results for the case study on the Great Recession

Inflation: recession period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.468	**0.073**	0.237	0.197	0.252	0.290	0.186	0.203	**0.183**	0.190
Euro area	1.722	1.677	1.637	**1.594**	1.610	1.530	1.379	1.390	**1.340**	1.394
Japan	0.427	**0.341**	0.402	0.541	0.508	0.254	**0.232**	0.243	0.329	0.299
United States	1.138	0.869	**0.862**	1.027	0.953	0.742	**0.537**	0.552	0.691	0.647

Inflation: reversion period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.751	**0.030**	0.279	0.488	0.417	0.448	**0.200**	0.221	0.300	0.263
Euro area	1.850	**0.365**	0.561	0.525	0.581	1.372	**0.225**	0.346	0.316	0.364
Japan	0.979	0.570	0.532	**0.514**	0.531	0.657	0.352	0.319	0.315	**0.312**
United States	2.181	0.663	0.686	0.430	**0.003**	1.342	0.446	0.448	0.376	**0.312**

Inflation: recovery period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.648	0.035	0.195	**0.035**	0.110	0.417	0.205	0.210	**0.184**	0.187
Euro area	0.582	**0.127**	0.327	0.208	0.363	0.370	0.158	0.207	**0.160**	0.220
Japan	1.090	0.801	0.921	**0.542**	0.626	0.750	0.492	0.595	**0.336**	0.375
United States	0.433	0.572	**0.107**	0.409	0.134	0.541	0.462	0.390	0.402	**0.367**

Table 7: Forecast evaluation criteria for inflation

Interest rate: recession period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.932	**0.689**	0.792	0.911	0.901	0.645	**0.426**	0.548	0.619	0.655
Euro area	0.903	**0.749**	0.758	0.781	0.765	0.685	**0.487**	0.552	0.521	0.571
Japan	0.076	0.019	**0.001**	0.003	0.003	0.054	0.057	0.043	0.050	**0.042**
United States	0.866	0.689	0.567	0.570	**0.540**	0.542	0.417	0.343	0.344	**0.329**

Interest rate: reversion period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.539	0.341	0.440	0.247	**0.068**	0.322	0.286	0.293	0.232	**0.194**
Euro area	2.178	**1.184**	1.246	1.312	1.257	1.800	**0.870**	1.031	1.002	1.041
Japan	0.255	0.226	0.205	**0.136**	0.143	0.154	0.134	0.122	0.089	**0.087**
United States	**0.964**	1.764	1.786	1.085	1.359	**0.580**	1.165	1.313	0.629	0.888

Interest rate: recovery period										
	RMSE					CRPS				
	Hrw	Gtv	Ri	Rm	Rim	Hrw	Gtv	Ri	Rm	Rim
Canada	0.577	0.576	**0.369**	0.672	0.484	0.345	0.352	**0.258**	0.415	0.294
Euro area	0.358	0.256	**0.143**	0.491	0.260	0.226	0.211	**0.144**	0.296	0.174
Japan	**0.010**	0.107	0.083	0.128	0.112	**0.054**	0.085	0.064	0.087	0.074
United States	0.484	0.247	0.547	**0.064**	0.318	0.309	0.304	0.360	**0.259**	0.275

Table 8: Forecast evaluation criteria for the interest rate