# A quasi-Bayesian local likelihood approach to time varying parameter VAR models
# Supplementary Appendix

Katerina Petrova

This Appendix contains supplementary material for the paper. Sections 1.1-1.3 contain the proofs of Propositions 1-3 respectively. Sections 1.4-1.6 include some additional results on the conditional quasi-posterior densities and the corresponding proofs. Section 2 provides a detailed description of the Gibbs algorithms developed in the paper. Finally, Section 3 consists of data and prior description and some additional results.

## 1 Proofs

### 1.1 Proof of Proposition 1

The assumptions for Proposition 1 are the same as in the main paper and provided again for readers' convenience.

**Assumptions.**

For all $j = [\tau T]$ with some fixed $\tau \in (0,1)$ :

1. The parameter space $\Theta$ is a compact subset of $\mathbb{R}^{\dim(\theta_j)}$ and $\theta_j^0 \in int(\Theta)$.

2. For any $\delta > 0$, there exists $\lambda > 0$ such that

$$\liminf_{T\to\infty} P\left\{ \sup_{\|\theta_j - \theta_j^0\| > \delta} \frac{1}{\varkappa_{Tj}} \left( \varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) \right) \leq -\lambda \right\} = 1,$$

   with $\varkappa_{Tj}$ defined in (6) below.

3. For $\theta_j$ in an open neighbourhood of $\theta_j^0$, $\varphi_{Tj}(.)$ admits the following expansion

$$\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) = (\theta_j - \theta_j^0)' \nabla\varphi_{Tj}(\theta_j^0) - \frac{1}{2}(\theta_j - \theta_j^0)' \varkappa_{Tj} J_{Tj}(\theta_j^0)(\theta_j - \theta_j^0) + R_{Tj}(\theta_j) \quad (1)$$

   where the main components

$$\nabla\varphi_{Tj}(\theta_j) = \frac{\partial\varphi_{Tj}(\theta_j)}{\partial\theta_j} \text{ and } J_{Tj}(\theta_j) := -\frac{1}{\varkappa_{Tj}} \mathbb{E}\mathcal{H}\left(\varphi_{Tj}(\theta_j)\right), \quad (2)$$

   and the remainder $R_{Tj}(.)$ satisfy:

   (a) $-J_{Tj}(\theta_j^0)^{1/2} \nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{Tj}} \to_d \mathcal{N}(0, I)$ as $T \to \infty$.

   (b) The matrix $J_{Tj}(\theta_j^0)$ satisfies

$$\lambda^* = \limsup_{T\to\infty} \lambda_{\max}(J_{Tj}(\theta_j^0)) < \infty$$
$$\lambda_* = \liminf_{T\to\infty} \lambda_{\min}(J_{Tj}(\theta_j^0)) > 0, \quad (3)$$

where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximal and minimal eigenvalue of a symmetric matrix $A$.

(c) For every $\varepsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ and $M > 0$ such that

$$\limsup_{T\to\infty} P\left\{ \sup_{M/\sqrt{\varkappa_{Tj}}\leq\left\|\theta-\theta_j^0\right\|\leq\delta} \frac{\|R_{Tj}(\theta)\|}{\varkappa_{Tj}\left\|\theta-\theta_j^0\right\|^2} > \eta \right\} \leq \varepsilon, \tag{4}$$

$$\limsup_{T\to\infty} P\left\{ \sup_{\left\|\theta-\theta_j^0\right\|\leq M/\sqrt{\varkappa_{Tj}}} \|R_{Tj}(\theta)\| > \varepsilon \right\} = 0. \tag{5}$$

4. The prior density $\pi_j(.)$ is strictly positive and Lipschitz continuous function over $\Theta$.

5. The time variation in the true parameters $\theta_j^0$ satisfies one of conditions:
**(i)** $\theta_t$ is a deterministic process $\theta_t = \theta\left(t/T\right)$, where $\theta(.)$ is a piecewise differentiable function.
**(ii)** $\theta_t$ is a stochastic process satisfying: $\sup_{j:|j-t|\leq h}||\theta_t - \theta_j||^2 = O_p\left(h/t\right)$ for $1 \leq h \leq t$ as $t \to \infty$.

We follow the proof of Theorem 1 in Chernozhukov and Hong (2003), which is a generalisation of earlier results by Ibragimov and Has'minskii (1981) and Bickel and Yahav (1969) for a non-likelihood based objective functions. We replace Chernozhukov and Hong (2003)'s asymptotic analysis based on $\sqrt{T}$-neighbourhoods of the true (in their analysis fixed) parameter $\theta^0$ to general $\varkappa_{Tj}^{1/2}$-neighbourhoods of $\theta_j^0$ (now indexed by time) with $\varkappa_{Tj}$ defined as

$$\varkappa_{Tj} := \left(\sum\nolimits_{t=1}^{T} w_{jt}^2\right)^{-1}. \tag{6}$$

This is justified since $\varkappa_{Tj} \sim H\frac{\left(\int K(x)dx\right)^2}{\int K^2(x)dx}$ where $H$, the bandwidth parameter associated with the kernel, satisfies $H \to \infty$ as $T \to \infty$.

In fact, the argument of Chernozhukov and Hong (2003) is valid for an arbitrary $m_T^{1/2}$-neighbourhood as long as $m_T \to \infty$ as $T \to \infty$. The radius of the specified neighbourhood corresponds to the consistency rate of the extremum estimator of $\theta_j^0$.

We start by defining $h_{Tj} = \sqrt{\varkappa_{Tj}}(\theta_j - \xi_{Tj})$ and $\xi_{Tj} := \theta_j^0 + \frac{1}{\varkappa_{Tj}} J_{Tj}(\theta_j^0)^{-1}\nabla\varphi_{Tj}(\theta_j^0)$. Letting

$$\mathbb{H}_{Tj} := \left\{ h_{Tj} \equiv \sqrt{\varkappa_{Tj}}\left(\theta_j - \theta_j^0\right) - J_{Tj}(\theta_j^0)^{-1}\nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{Tj}} : \theta_j \in \Theta \right\} \tag{7}$$

we need to show that

$$\int_{\mathbb{H}_{Tj}} \|h\|^a \left| p_{Tj}^*(h) - p_\infty(h) \right| dh \to_p 0 \tag{8}$$

for all $a \geq 0$. Recall the definition of the quasi-posterior density and using the modified objective function $\varphi_{Tj}(\theta_j)$:

$$p_{Tj}(\theta_j) = \frac{\pi_j(\theta_j)\exp(\varphi_{Tj}(\theta_j))}{\int_\Theta \pi_j(\theta)\exp(\varphi_{Tj}(\theta))d\theta}.$$

We begin by applying the transformation of $\theta_j = h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}$ to obtain

$$
\begin{aligned}
p_{Tj}(h_{Tj}) &= \frac{1}{\sqrt{\varkappa_{Tj}}} p_{Tj}(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) \\
&= \frac{\pi_j(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) \exp(\varphi_{Tj}(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}))}{\int_{\mathbb{H}_{Tj}} \pi_j(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) \exp(\varphi_{Tj}(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) dh} \\
&= \pi_j(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) \exp(\omega_{Tj}(h_{Tj}))/C_{Tj} \qquad (9)
\end{aligned}
$$

where

$$
\omega_{Tj}(h_{Tj}) = \varphi_{Tj}\left(\frac{h_{Tj}}{\sqrt{\varkappa_{Tj}}} + \xi_{Tj}\right) - \varphi_{Tj}(\theta_j^0) - \frac{1}{2\varkappa_{Tj}} \left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1} \left[\nabla\varphi_{Tj}(\theta_j^0)\right] \qquad (10)
$$

and

$$
C_{Tj} = \int_{\mathbb{H}_{Tj}} \pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) \exp\left\{\omega_{Tj}(h)\right\} dh. \qquad (11)
$$

For $h_{Tj}$ belonging to the integration area $\mathbb{H}_{Tj}$ in (7), the following useful identity applies:

$$
\omega_{Tj}(h_{Tj}) = -\frac{1}{2} h'_{Tj} J_{Tj}(\theta_j^0) h_{Tj} + R_{Tj}\left(h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right) \quad \forall h_{Tj} \in \mathbb{H}_{Tj}. \qquad (12)
$$

To prove (12), note that, by definition of $\mathbb{H}_{Tj}$, for any $h_{Tj} \in \mathbb{H}_{Tj}$ there exists a $\theta_j \in \Theta$ satisfying the identities

$$
\begin{aligned}
h_{Tj} &= \sqrt{\varkappa_{Tj}}\left(\theta_j - \theta_j^0\right) - J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{Tj}}, \qquad (13) \\
\theta_j &= \frac{h_{Tj}}{\sqrt{\varkappa_{Tj}}} + \xi_{Tj}. \qquad (14)
\end{aligned}
$$

Applying (13) to (10) and using (1) we obtain,

$$
\begin{aligned}
\omega_{Tj}(h_{Tj}) &= \varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) - \frac{1}{2\varkappa_{Tj}} \left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1} \left[\nabla\varphi_{Tj}(\theta_j^0)\right] \\
&= (\theta_j - \theta_j^0)' \nabla\varphi_{Tj}(\theta_j^0) - \frac{1}{2}(\theta_j - \theta_j^0)' \varkappa_{Tj} J_{Tj}(\theta_j^0)(\theta_j - \theta_j^0) + R_{Tj}(\theta_j) \\
&\quad - \frac{1}{2\varkappa_{Tj}} \left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1} \left[\nabla\varphi_{Tj}(\theta_j^0)\right].
\end{aligned}
$$

Using (14) in each of the terms in of the above expression and collecting terms proves (12). Having established (12), we prove that all subsequential probability limits of $C_{Tj}$ in (11) are strictly positive.

Using (5), dominated convergence theorem and the properties of the Gaussian density we obtain

$$\begin{aligned}C_{Tj} &= \int_{\mathbb{H}_{Tj}} \pi_j\left(\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}\right)\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h+R_{Tj}\left(\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}\right)\right\}dh \\ &= \pi_j\left(\theta_j^0\right)\int_{\mathbb{R}^{\dim(\theta_j^0)}}\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h\right\}dh \\ &\quad +\int_{\mathbb{H}_{Tj}}\left\{\pi_j\left(\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}\right)-\pi_j\left(\theta_j^0\right)\right\}\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h\right\}dh+o_p(1) \\ &= (2\pi)^{\dim(\theta_j^0)/2}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right)+o_p(1),\end{aligned}$$

because Lipschitz continuity of $\pi_j$ in Assumption 4 implies that

$$\left\|\pi_j\left(\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}\right)-\pi_j\left(\theta_j^0\right)\right\|\leq c\left\|\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}-\theta_j^0\right\|\to_p 0 \tag{15}$$

for all $h_{Tj}\in\mathbb{H}_{Tj}$. We conclude that, as $T\to\infty$

$$\left|C_{Tj}-(2\pi)^{\dim(\theta_j^0)/2}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right)\right|\to_p 0, \tag{16}$$

which implies that

$$\liminf_{T\to\infty} P\left(C_{Tj}>\varepsilon\right)\geq 1-\varepsilon \qquad (\forall\varepsilon>0), \tag{17}$$

since Assumption 3b guarantees that $\liminf_{T\to\infty}\det(J_{Tj}(\theta_j^0))^{-1/2}>0$ *a.s.* so that

$$\liminf_{T\to\infty}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right)>0 \quad a.s.$$

by Assumption 4. Using (9), the left side of (8) can be written as

$$\int_{\mathbb{H}_{Tj}}\|h\|^a\left|p_{Tj}^*(h)-p_\infty(h)\right|dh=A_{Tj}C_{Tj}^{-1} \tag{18}$$

where $C_{Tj}$ is defined in (11) and

$$A_{Tj}=\int_{\mathbb{H}_{Tj}}\|h\|^a\left|\pi\left(\frac{h}{\sqrt{\varkappa_{Tj}}}+\xi_{Tj}\right)\exp(\omega(h))-C_{Tj}\phi_{Tj}(h)\right|dh$$

where $\phi_{Tj}(h_{Tj})=(2\pi)^{-\dim(\theta_j)/2}\det(J_{Tj}(\theta_j^0))^{1/2}\exp\left\{-\frac{1}{2}h_{Tj}'J_{Tj}(\theta_j^0)h_{Tj}\right\}$. Adding and subtracting $\exp\left\{-\frac{1}{2}h_{Tj}'J_{Tj}(\theta_j^0)h_{Tj}\right\}\pi_j\left(\theta_j^0\right)$ in the above expression we obtain that $A_{Tj}\leq A_{Tj}^{(1)}+A_{Tj}^{(2)}$ where

$$\begin{aligned}A_{Tj}^{(1)} &= \int_{\mathbb{H}_{Tj}}\|h\|^a\left|\pi_j(h/\sqrt{\varkappa_{Tj}}+\xi_{Tj})e^{\omega(h)}-e^{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h}\pi_j\left(\theta_j^0\right)\right|dh, \\ A_{Tj}^{(2)} &= \int_{\mathbb{H}_{Tj}}\|h\|^a\left|\pi_j\left(\theta_j^0\right)-C_{Tj}(2\pi)^{-\dim(\theta_j)/2}\det(J_{Tj}(\theta_j^0))^{1/2}\right|\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h\right\}dh.\end{aligned} \tag{19}$$

By using (16) and the dominated convergence theorem, $A_{Tj}^{(2)}\to_p 0$ as $T\to\infty$. Also, (17) implies that $\left|C_{Tj}^{-1}\right| = O_p(1)$. Therefore, by (18), the proposition is proven if $A_{Tj}^{(1)} \to_p 0$.

To show this, we partition the area of integration in (7) as follows: $\mathbb{H}_{Tj} = \mathbb{H}_{Tj}^{(1)} \cup \mathbb{H}_{Tj}^{(2)} \cup \mathbb{H}_{Tj}^{(3)}$, where

$$\begin{aligned}
\text{i) } \mathbb{H}_{Tj}^{(1)} &= \{h \in \mathbb{H}_{Tj} : \|h\| \leq M\}, \\
\text{ii) } \mathbb{H}_{Tj}^{(2)} &= \left\{h \in \mathbb{H}_{Tj} : \|h\| > \delta\sqrt{\varkappa_{Tj}}\right\}, \\
\text{iii) } \mathbb{H}_{Tj}^{(3)} &= \left\{h \in \mathbb{H}_{Tj} : M < \|h\| \leq \delta\sqrt{\varkappa_{Tj}}\right\},
\end{aligned}$$

where $\delta$ is some positive number satisfying (4) and $M$ is a fixed number in $(0, \infty)$. It remains to prove $A_{Tj}^{(1)} \to_p 0$ over areas i)-iii). Note that, by (7) and Assumption 3(a),

$$\frac{h_{Tj}}{\sqrt{\varkappa_{Tj}}} = \theta_j - \theta_j^0 + O_p\left(\frac{1}{\sqrt{\varkappa_{Tj}}}\right)$$

so the areas i)-iii) can be expressed in terms of $\theta_j$: for example,

$$\mathbb{H}_{Tj}^{(3)} = \left\{\theta \in \Theta : \frac{M}{\sqrt{\varkappa_{Tj}}} < \|\theta\| \leq \delta\right\}, \tag{20}$$

for all but finitely many $T$.

**(i) Area $\mathbb{H}_{Tj}^{(1)}$.** By finiteness of $M$, it suffices to show that the following quantity is $o_p(1)$:

$$\begin{aligned}
&\sup_{h \in \mathbb{H}_{Tj}^{(1)}} \|h\|^a \left| e^{\omega_{Tj}(h)} \pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) - e^{-\frac{1}{2}h' J_{Tj}(\theta_j^0)h} \pi_j(\theta_j^0) \right| \\
= \;& \max_{\|h\| \leq M} \|h\|^a e^{-\frac{1}{2}h' J_{Tj}(\theta_j^0)h} \left| \exp\left\{R_{Tj}\left(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right)\right\} \pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
\leq \;& M^a \max_{\|h\| \leq M} \left| \exp\left\{R_{Tj}\left(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right)\right\} \pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
\leq \;& M^a \max_{\|h\| \leq M} \exp\left\{R_{Tj}\left(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right)\right\} \max_{\|h\| \leq M} \left| \pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
& + M^a \max_{\|h\| \leq M} \left| \exp\left\{R_{Tj}\left(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right)\right\} - 1 \right| \sup_j \pi_j(\theta_j^0).
\end{aligned} \tag{21}$$

The second term of (21) is $o_p(1)$ by (5) since $\sup_{\|h\| \leq M} \left\|h/\sqrt{\varkappa_{Tj}} + \xi_T - \theta_j^0\right\| = O_p(1/\sqrt{\varkappa_{Tj}})$. Since $\exp\left\{R_{Tj}\left(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj}\right)\right\} = 1 + o_p(1)$ by (5), the first term of (21) is $o_p(1)$ by (15).

**(ii) Area $\mathbb{H}_{Tj}^{(2)}$.** It is sufficient to show that:

a) $\int_{\|h\| > \delta\sqrt{\varkappa_{Tj}}} \|h\|^a \pi(h/\sqrt{\varkappa_{Tj}} + \xi_T) \exp(\omega(h)) dh \to_p 0$ and

b) $\int_{\|h\| > \delta\sqrt{\varkappa_{Tj}}} \|h\|^a \exp\left\{-\frac{1}{2}h' J_{Tj}(\theta_j^0)h\right\} dh \to_p 0$.

To show a), change variables $\theta_j = h_{Tj}/\sqrt{\varkappa_{Tj}} + \xi_{Tj}$, so that the area of integration becomes $\left\|\theta - \xi_{Tj}\right\| > \delta$. The expression in a) is bounded by

$$\varkappa_{Tj}^{(\alpha+1)/2} \int_{\|\theta-\xi_{Tj}\|>\delta} \left\|\theta - \xi_{Tj}\right\|^{\alpha} \pi_j(\theta) \exp\left\{\varphi_{Tj}(\theta) - \varphi_{Tj}(\theta_j^0)\right\}$$
$$\times \exp\left\{-\frac{1}{2\varkappa_{Tj}} \nabla\varphi_{Tj}(\theta_j^0)' J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)\right\} d\theta$$

which in turn, is bounded by

$$\varkappa_{Tj}^{(\alpha+1)/2} C K_n \int_{\|\theta-\xi_{Tj}\|>\delta/2} (1+\|\theta\|^{\alpha}) \pi_j(\theta) \exp\left(\varphi_{Tj}(\theta) - \varphi_{Tj}(\theta_j^0)\right) d\theta \tag{22}$$

with $K_n := \exp(-\frac{1}{2\varkappa_{Tj}} \nabla\varphi_{Tj}(\theta_j^0)' J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)) = O_p(1)$, because $\xi_{Tj} = O_p(1)$. By Assumption 2, $\forall\delta > 0$, $\exists\varepsilon > 0$ such that

$$\lim_{n\to\infty} P\left\{\sup_{\|\theta_j-\theta_j^0\|>\delta/2} \exp\left(\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0)\right) \leq \exp(-\varepsilon\varkappa_{Tj})\right\} = 1$$

so that (22) is bounded by $O_p(1)\varkappa_{Tj}^{(\alpha+1)/2} \exp(-\varepsilon\varkappa_{Tj}) \int_{\Theta} \|\theta\|^{\alpha} \pi_j(\theta) d\theta = o_p(1)$ since $\Theta$ is bounded and $\pi_j(.)$ is continuous by Assumptions 1 and 4 respectively.

b) follows directly since the Gaussian density has infinitely many moments and $\varkappa_{Tj} \to \infty$.

**(iii) Area $\mathbb{H}_{Tj}^{(3)}$.** By integrability of Gaussian density functions, we can choose $M$ to be large enough to make the term in $A_{1n} \exp\left\{-\frac{1}{2}h_{Tj}' J_{Tj}(\theta_j^0) h_{Tj}\right\}$ arbitrarily small. So, it is sufficient to show that for all $\varepsilon > 0$, there exists $M$ such that the remaining term

$$\liminf_{T\to\infty} P\left\{\int_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \|h\|^{a} \left|\pi_j(h/\sqrt{\varkappa_{Tj}} + \xi_T) \exp(\omega_{Tj}(h))\right| dh < \varepsilon\right\} \geq 1 - \varepsilon. \tag{23}$$

By (12),

$$e^{\omega(h_{Tj})} \leq e^{-\frac{1}{2}h_{Tj}' J_{Tj}(\theta_j^0) h_{Tj} + \left|R_{Tj}\left(\xi_T + h_{Tj}/\sqrt{\varkappa_{Tj}}\right)\right|}. \tag{24}$$

Moreover, $\delta$ in the partition of $\mathbb{H}_{Tj}$ is chosen to satisfy (4); by (20) and (4), for all $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ and $M > 0$, such that

$$\begin{aligned}
1 - \varepsilon &\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \frac{\left|R_{Tj}(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj})\right|}{\left\|h + \sqrt{\varkappa_{Tj}}(\xi_{Tj} - \theta_j^0)\right\|^2} \leq \frac{1}{4}\eta\right\} \\
&\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \frac{\left|R_{Tj}(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj})\right|}{2\|h\|^2 + 2\left\|\sqrt{\varkappa_{Tj}}(\xi_{Tj} - \theta_j^0)\right\|^2} \leq \frac{1}{4}\eta\right\} \\
&\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \frac{\left|R_{Tj}(h/\sqrt{\varkappa_{Tj}} + \xi_{Tj})\right|}{\|h\|^2 + C^2} \leq \frac{1}{2}\eta\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \liminf_{T\to\infty} P\left\{ \sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \frac{\left|R_{Tj}(h/\sqrt{\varkappa_{Tj}}+\xi_{Tj})\right|}{\|h\|^2\left(1+\frac{C^2}{M^2}\right)} \leq \frac{1}{2}\eta \right\} \\
&= \liminf_{T\to\infty} P\left\{ \sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{Tj}}} \frac{\left|R_{Tj}(h/\sqrt{\varkappa_{Tj}}+\xi_{Tj})\right|}{\|h\|^2} \leq \frac{1}{2}B\eta \right\} \qquad (25)
\end{aligned}
$$

for all $\varepsilon > 0$ and all $\eta > 0$ and some $C > 0$, with $B = 1 + C^2/M^2$. The equality in (25) follows since

$$
\sqrt{\varkappa_{Tj}}(\xi_{Tj} - \theta_j^0) = \frac{1}{\sqrt{\varkappa_{Tj}}} J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0) = O_p(1)
$$

so $\liminf_{T\to\infty} P\left\{\left\|\sqrt{\varkappa_{Tj}}(\xi_{Tj} - \theta_j^0)\right\| \leq C\right\} = 1$ for some $C > 0$. Choosing $\eta = (2B)^{-1}\lambda_*$ in (25) (with $\lambda_*$ defined in (3)) yields

$$
\liminf_{T\to\infty} P\left\{ \left|R_{Tj}(h_{Tj}/\sqrt{\varkappa_{Tj}}+\xi_{Tj})\right| \leq \frac{1}{4}\|h_{Tj}\|^2\lambda_* \right\} > 1-\varepsilon. \qquad (26)
$$

Going back to (24), we want to show that

$$
\liminf_{T\to\infty} P\left( e^{\omega(h_{Tj})} \leq C e^{-\frac{1}{4}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}} \right) \geq 1-\varepsilon.
$$

Note that, by the identity $\min_{\|x\|=1} x'Ax/\|x\|^2 = \lambda_{\min}(A)$ for any symmetric $A$, we obtain

$$
\begin{aligned}
-\frac{1}{4}h'_{Tj}J_{Tj}\left(\theta_j^0\right)h_{Tj} &\geq -\frac{1}{2}h'_{Tj}J_{Tj}\left(\theta_j^0\right)h_{Tj} + \frac{1}{4}\lambda_{\min}\left(J_{Tj}\left(\theta_j^0\right)\right)\|h_{Tj}\|^2 \\
&\geq -\frac{1}{2}h'_{Tj}J_{Tj}\left(\theta_j^0\right)h_{Tj} + \frac{1}{4}\lambda_*\|h_{Tj}\|^2
\end{aligned}
$$

for all but finitely many $T$. Denoting

$$
\begin{aligned}
G_{1T} &= -\frac{1}{2}h'_{Tj}J_{Tj}\left(\theta_j^0\right)h_{Tj} + \left|R_{Tj}\left(\xi_T + h_{Tj}/\sqrt{\varkappa_{Tj}}\right)\right| \\
G_{2T} &= \frac{1}{4}\lambda_*\|h_{Tj}\|^2 - \left|R_{Tj}\left(\xi_T + h_{Tj}/\sqrt{\varkappa_{Tj}}\right)\right|,
\end{aligned}
$$

the above inequality implies that

$$
\begin{aligned}
\liminf_{T\to\infty} P\left( e^{\omega(h_{Tj})} \leq Ce^{-\frac{1}{4}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}} \right) &\geq \liminf_{T\to\infty} P\left( e^{\omega(h_{Tj})} \leq Ce^{-\frac{1}{2}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}+\frac{1}{4}\lambda_*\|h_{Tj}\|^2} \right) \\
&= \liminf_{T\to\infty} P\left( e^{\omega(h_{Tj})} \leq Ce^{G_{1T}}e^{G_{2T}} \right) \\
&\geq \liminf_{T\to\infty} P\left( e^{\omega(h_{Tj})} \leq Ce^{G_{1T}}e^{G_{2T}} \middle| e^{G_{2T}} \geq 1 \right) P\left(e^{G_{2T}} \geq 1\right) \\
&\geq \liminf_{T\to\infty} P\left(e^{G_{2T}} \geq 1\right) \\
&\geq 1-\varepsilon
\end{aligned}
$$

by (26), where we have used the fact that

$$P\left(e^{\omega(h_{Tj})} \leq Ce^{G_{1T}}e^{G_{2T}}\middle| e^{G_{2T}} \geq 1\right) = 1$$

for any $C \geq 1$ by (24). This completes the proof of Proposition 1.

## 1.2 Proof of Proposition 2

We have that the $M \times 1$ dimensional vector $y_t$ is generated by a time varying parameter VAR model of lag order $k$:

$$y_t = B_{0t} + \sum_{p=1}^{k} B_{pt} y_{t-p} + \varepsilon_t, \tag{27}$$

where $B_{0t}$ is an $M \times 1$ vector of time varying intercepts, and $B_{pt}$ is an $M \times M$ matrix of time varying autoregressive coefficients for lag $p = 1, ..., k$. We restrict our attention only to stable autoregressions and assume that all roots of the polynomial $\psi(z) = \det\left(I_M - \sum_{p=1}^{k} z^p B_{pt}\right)$ lie outside the unit circle. The error term, $\varepsilon_t$, is an $M \times 1$ vector of normally distributed zero mean random variables, with a positive definite symmetric $M \times M$ contemporaneous drifting covariance matrix $R_t^{-1}$, so that $\varepsilon_t = R_t^{-1/2}\eta_t$ where $\eta_t \sim NID(0, I_M)$. In addition, let $x_t := (1, y'_{t-1}, ..., y'_{t-k})$ be a $1 \times (Mk+1)$ vector and $B_t := (B_{0t}, B_{1t}, ..., B_{kt})$ be an $M \times (Mk+1)$ matrix. Then, (27) can be written as $y_t = B_t x'_t + \varepsilon_t$ and after vectorising,

$$y_t = (I_M \otimes x_t)\beta_t + R_t^{-1/2}\eta_t, \tag{28}$$

where $\beta_t := vec(B'_t)$ is an $M(Mk+1) \times 1$ vector.

The weighted likelihood of the sample $(y_1, ..., y_T)$ at each point in time $j$ is given by

$$L_j(y|\beta_j, R_j, X) = (2\pi)^{-M\varkappa_{Tj}/2}\,|R_j|^{\varkappa_{Tj}/2} e^{-\frac{1}{2}\sum_{t=1}^{T}\vartheta_{jt}(y_t-(I_M\otimes x_t)\beta_j)' R_j (y_t-(I_M\otimes x_t)\beta_j)} \tag{29}$$

where $\varkappa_{Tj}$ is defined in (6) and the kernel weights $\vartheta_{jt}$ are defined as $\vartheta_{jt} := \varkappa_{Tj} w_{jt}$ , so that $\varkappa_{Tj} = \sum_{t=1}^{T}\vartheta_{jt}$. Denote by $Y = (y_1, ..., y_T)'$ a $T \times M$ matrix of stacked vectors $y'_1, ..., y'_T$ and define $y = vec(Y)$ as a $TM \times 1$ vector. Define $E = (\varepsilon_1, ..., \varepsilon_T)'$ implying that $\varepsilon := vec(E)$ is a $TM \times 1$ vector. Let $X$ be a $T \times Mk + 1$ matrix defined as $X = (x'_1, ..., x'_T)'$. Then the weighted likelihood can be written in a more compact form as:

$$L_j(y|\beta_j, R_j, X) \propto |R_j|^{tr(D_j)/2} \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta_j)'(R_j \otimes D_j)(y - (I_M \otimes X)\beta_j)\right\}$$

where $D_j := diag(\vartheta_{j1}, ..., \vartheta_{jT})$ for $j \in \{1, ..., T\}$.

Next, specify a prior for $\beta_j$ and $R_j$ that has Normal-Wishart distribution:

$$\begin{aligned} p(\beta_j, R_j) \quad &\propto \quad |R_j|^{(Mk+1)/2} \exp\left\{-\frac{1}{2}(\beta_j - \beta_{0j})'(R_j \otimes \kappa_{0j})(\beta_j - \beta_{0j})\right\} \\ &\quad \times |R_j|^{\frac{\alpha_{0j}-M-1}{2}} \exp\left\{-\frac{1}{2}tr(\gamma_{0j}R_j)\right\} \end{aligned}$$

where $\beta_{0j}$ is a $(MK+1)M \times 1$ vector of prior means, $\kappa_{0j}$ is a $(Mk+1) \times (Mk+1)$ positive definite symmetric precision matrix, $\alpha_{0j}$ is a scalar scale parameter of the Wishart distribution, $\gamma_{0j}$ is a positive definite symmetric matrix, $tr(.)$ denotes the trace operator and $\propto$ denotes proportionality upto a constant. Combining the weighted likelihood with the prior, we obtain the local quasi-posterior density,

$$\begin{aligned} p(\beta_j, R_j|Y,X) \quad \propto \quad & \exp\left\{-\frac{1}{2}(y-(I_M\otimes X)\beta_j)'(R_j\otimes D_j)(y-(I_M\otimes X)\beta_j)\right\}|R_j|^{\varkappa_{Tj}/2}|R_j|^{(Mk+1)/2} \\ & \times|R_j|^{\frac{\alpha_{0j}-M-1}{2}}\exp\left\{-\frac{1}{2}(\beta_j-\beta_{0j})'(R_j\otimes\kappa_{0j})(\beta_j-\beta_{0j})-\frac{1}{2}tr(\gamma_{0j}R_j)\right\}. \qquad (30)\end{aligned}$$

Here, it is useful to employ the following identity:

$$u'Au-2a'u=(u-A^{-1}a)'A(u-A^{-1}a)-a'A^{-1}a. \qquad (31)$$

A direct application of the identity in the expression in the exponential term in (30) yields:

$$\begin{aligned} &(y-(I_M\otimes X)\beta_j)'(R_j\otimes D_j)(y-(I_M\otimes X)\beta_j)+(\beta_j-\beta_{0j})'(R_j\otimes\kappa_{0j})(\beta_j-\beta_{0j}) \\ &=y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}+\beta_j'(R_j\otimes(X'D_jX+\kappa_{0j}))\beta_j-2[R_j\otimes\left(XD_jy+\kappa_{0j}\beta_{0j}\right)]'\beta_j \\ &=y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}+\left(\beta_j-\widetilde{\beta}_j\right)'(R_j\otimes\widetilde{\kappa}_j)\left(\beta_j-\widetilde{\beta}_j\right)-\widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j \qquad (32)\end{aligned}$$

where

$$\begin{aligned} \widetilde{\beta}_j &= \left(I_M\otimes\widetilde{\kappa}_j^{-1}\right)\left[(I_M\otimes X'D_jX)\hat{\beta}_j+(I_M\otimes\kappa_{0j})\beta_{0j}\right], \qquad (33) \\ \widetilde{\kappa}_j &= \kappa_{0j}+X'D_jX. \end{aligned}$$

It remains to study the remaining terms in (32):

$$\exp\left\{-\frac{1}{2}\left[y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}-\widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j+tr(\gamma_{0j}R_j)\right]\right\}. \qquad (34)$$

The first three terms in (34) will be handled in the same way. For example, the first term can be written as:

$$\begin{aligned} y'(R_j\otimes D_j)y &= \left[(I_M\otimes D_j^{1/2})vec(Y)\right]'(R_j\otimes I_T)\left[(I_M\otimes D_j^{1/2})vec(Y)\right] \\ &= vec(D_j^{1/2}Y)'(R_j\otimes I_T)vec(D_j^{1/2}Y) \\ &= tr(D_j^{1/2}YR_jY'D_j^{1/2})=tr(Y'D_jYR_j) \qquad (35)\end{aligned}$$

where the second line is obtained using the equality $vec(ABC)=(C'\otimes A)vec(B)$ and the third line uses the equality $tr(ABC)=vec(A')'(I\otimes B)vec(C)=vec(A)'(B\otimes I)vec(C')$. Similarly,

$$\begin{aligned} \beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j} &= tr(B_{0j}\kappa_{0j}B_{0j}'R_j), \quad \text{and} \qquad (36) \\ \widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j &= tr(\widetilde{B}_j\widetilde{\kappa}_j\widetilde{B}_j'R_j), \qquad (37)\end{aligned}$$

where $\beta_{0j}=vec(B_{0j}')$ and $\widetilde{\beta}_j=vec(\widetilde{B}_j')$. After combining (32), (35),(36) and (37) the quasi-posterior in (30) can be written as: $p(\beta_j, R_j|Y,X) \propto$

$$|R_j|^{(Mk+1)/2}|R_j|^{\frac{\widetilde{\alpha}_j-M-1}{2}} \exp\left\{-\frac{1}{2}\left[(\beta_j-\widetilde{\beta}_j)'(R_j\otimes\widetilde{\kappa}_j)(\beta_j-\widetilde{\beta}_j)+tr\left(\widetilde{\gamma}_jR_j\right)\right]\right\} \tag{38}$$

which is a Normal-Wishart density with parameters $\widetilde{\beta}_j, \widetilde{\kappa}_j, \widetilde{\alpha}_j$ and $\widetilde{\gamma}_j^{-1}$, where $\widetilde{\beta}_j$ and $\widetilde{\kappa}_j$ are defined above in (33) and

$$\widetilde{\alpha}_j = \alpha_{0j} + \varkappa_{Tj}, \quad \widetilde{\gamma}_j = \gamma_{0j} + Y'D_jY + B_{0j}\kappa_{0j}B_{0j}' - \widetilde{B}_j\widetilde{\kappa}_j\widetilde{B}_j',$$

which proves Proposition 2.

### 1.3 Proof of Proposition 3

To obtain the marginal quasi-posterior density of the parameter vector $\beta_j$, we integrate the joint quasi-posterior distribution, $p(\beta_j, R_j|Y,X)dR$ in (38) over the $M(M+1)/2$ distinct variables in $R_j$,

$$\begin{aligned} p(\beta_j|Y,X) &= \textstyle\int_{R>0} p(\beta_j, R|Y,X)dR \\ &= C_j \textstyle\int_{R>0} \exp\left\{-\frac{1}{2}(\beta_j-\widetilde{\beta}_j)'(R_j\otimes\widetilde{\kappa}_j)(\beta_j-\widetilde{\beta}_j)\right\} \\ &\quad \times |R_j|^{(Mk+1)/2}\,|R_j|^{\frac{\widetilde{\alpha}_j-M-1}{2}} \exp\left\{-\frac{1}{2}tr\left(\widetilde{\gamma}_jR_j\right)\right\} dR \end{aligned} \tag{39}$$

where $C_j = \frac{(2\pi)^{-M(Mk+1)/2}|\widetilde{\kappa}_j|^{M/2}}{2^{\frac{1}{2}M\widetilde{\alpha}_j}|\widetilde{\gamma}_j|^{-\widetilde{\alpha}_j/2}\Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}$. Note that the first exponential term in (39) can be written as

$$\begin{aligned} &\exp\left\{-\frac{1}{2}(\beta_j-\widetilde{\beta}_j)'(R_j\otimes\widetilde{\kappa}_j)(\beta_j-\widetilde{\beta}_j)\right\} \\ =\ &\exp\left\{-\frac{1}{2}vec\left[\widetilde{\kappa}_j^{1/2}(B_j-\widetilde{B}_j)'\right]'(R_j\otimes I_{Mk+1})vec\left[\widetilde{\kappa}_j^{1/2}(B_j-\widetilde{B}_j)'\right]\right\} \\ =\ &etr\left\{-\frac{1}{2}(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'R_j\right\}, \end{aligned}$$

where $etr(\cdot)$ is the exponential trace operator, so that

$$p(\beta_j|Y,X) = C_{5j}\textstyle\int_{R>0}|R_j|^{\frac{\widetilde{\alpha}_j-M-1+Mk+1}{2}} etr\left\{-\frac{1}{2}\left[(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+\widetilde{\gamma}_j\right]R_j\right\}dR. \tag{40}$$

By Theorem 2.1.11 in Muirhead (2005), for a $n\times n$ symmetric positive definite matrix $R$, with $\mathrm{Re}(\Psi)>0$ and $\mathrm{Re}(\delta)>0$, the following holds:

$$\textstyle\int_{R>0}|R|^{\delta-\frac{n-1}{2}}\exp\left\{-\frac{1}{2}tr(\Psi^{-1}R)\right\}d(R) = \Gamma_n(\delta)|\Psi|^{\delta}2^{n\delta}.$$

Applying the theorem in (40), we have that

$$
\begin{aligned}
p(\beta_j|Y,X) &= C_{5j}\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)\left|(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+\widetilde{\gamma}_j\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}} 2^{M\frac{\widetilde{\alpha}_j+Mk+1}{2}} \\
&= \frac{(2\pi)^{-M(Mk+1)/2}\left|\widetilde{\kappa}_j\right|^{M/2}}{2^{-\frac{1}{2}M\widetilde{\alpha}_j}\left|\widetilde{\gamma}_j\right|^{-\widetilde{\alpha}_j/2}\Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)\left|\widetilde{\gamma}_j\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}} \\
&\quad\times\left|\widetilde{\gamma}_j^{-1}(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+I_M\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}} 2^{M\frac{\widetilde{\alpha}_j+Mk+1}{2}}
\end{aligned}
$$

$$
= \pi^{-M(Mk+1)/2}\left|\widetilde{\kappa}_j\right|^{M/2}\left|\widetilde{\gamma}_j\right|^{-\frac{Mk+1}{2}}\frac{\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)}{\Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}\left|\widetilde{\gamma}_j^{-1}(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+I_M\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}}.
$$

Recall the definition of a matrix variate $t$-distribution density of a $p\times m$ matrix $B$:

$$
\begin{aligned}
p(B;\nu,M,\Sigma,\Omega) &= \frac{\Gamma_p(\frac{1}{2}(\nu+m+p-1))}{\pi^{\frac{1}{2}mp}\Gamma_p(\frac{1}{2}(\nu+p-1))}\left|\Sigma\right|^{-\frac{1}{2}m}\left|\Omega\right|^{-\frac{1}{2}p} \\
&\quad\times\left|I_p+\Sigma^{-1}(B-M)\Omega^{-1}(B-M)'\right|^{-\frac{1}{2}(\nu+m+p-1)},
\end{aligned}
$$

and note that $p(\beta_j|Y,X)$ is a vectorised counterpart of the above definition with parameters:

$$
m=Mk+1,\quad p=M,\quad \nu=\widetilde{\alpha}_j-Mk,\quad M=\widetilde{B}_j,\quad \Omega^{-1}=\widetilde{\kappa}_j,\quad \Sigma=\widetilde{\gamma}_j.
$$

We therefore have

$$
\beta_j|Y,X=vec(B_j')|Y,X\sim T_{\widetilde{\alpha}_j-Mk}\left(vec(\widetilde{B}_j'),\frac{\widetilde{\gamma}_j\otimes\widetilde{\kappa}_j^{-1}}{\widetilde{\alpha}_j-Mk-2}\right),
$$

which proves Proposition 3.

### 1.4 Conditional quasi-posterior distributions

There are two cases of particular interest to be considered for model (27): first, when the covariance matrix $R_t^{-1/2}$ is known; second, when the parameter vector $\beta_t$ is known. The resulting closed form conditional quasi-posterior densities, derived in Propositions 4 and 5 below, are useful for the design of the Gibbs algorithms for estimating model (27) generated by mixtures of fixed and time varying parameters. These Gibbs algorithms are developed in Section 3.3 of the paper and are described in detail in Section 3 below.

**Case 1.** Conditional on knowing $R_t^{-1}$, the model in (27) can be transformed in the following way:

$$
\begin{aligned}
&\widetilde{y}_t=\widetilde{x}_t\beta_t+\eta_t,\quad \eta_t\sim NID(0_M,I_M),\\
&\widetilde{y}_t:=R_t^{1/2}y_t,\;\; \widetilde{x}_t:=R_t^{1/2}(I_M\otimes x_t).
\end{aligned}
\tag{41}
$$

**Case 2.** For known $\beta_t$, the model in (27) can be written as

$$\varepsilon_t = y_t - (I_M \otimes x_t)\beta_t = R_t^{-1/2}\eta_t, \quad \eta_t \sim NID(0_M, I_M).$$

**Proposition 4.** *If a* $\mathcal{N}\left(\beta_{0j}, V_{0j}^{-1}\right)$ *prior distribution is selected for* $\beta_j$ *in Case 1, the quasi-posterior distribution of* $\beta_j$ is given by $\mathcal{N}\left(\widetilde{\beta}_j, \widetilde{V}_j^{-1}\right)$ *with posterior mean and variance*

$$\widetilde{\beta}_j = \widetilde{V}_j^{-1}\left[\sum\nolimits_{t=1}^T \vartheta_{jt}\widetilde{x}_t'\widetilde{y}_t + V_{0j}\beta_{0j}\right], \quad \widetilde{V}_j = V_{0j} + \sum\nolimits_{t=1}^T \vartheta_{jt}\widetilde{x}_t'\widetilde{x}_t$$

*for each* $j \in \{1, ..., T\}$, *where* $\widetilde{x}_t$ *and* $\widetilde{y}_t$ *are defined in (41).*

**Proposition 5.** *If a Wishart* $\mathcal{W}(\alpha_{0j}, \gamma_{0j}^{-1})$ *prior distribution is selected for* $R_j$ *in Case 2, the quasi-posterior of* $R_j$ *is also Wishart* $\mathcal{W}(\widetilde{\alpha}_j, \widetilde{\gamma}_j^{-1})$ *with posterior parameters for each* $j \in \{1, ..., T\}$ :

$$\widetilde{\alpha}_j = \alpha_{0j} + \sum\nolimits_{t=1}^T \vartheta_{jt}, \quad \widetilde{\gamma}_j = \gamma_{0j} + \sum\nolimits_{t=1}^T \vartheta_{jt}\varepsilon_t'\varepsilon_t.$$

## 1.5 Proof of Proposition 4

Assume that the variance covariance matrix $R_t^{-1/2}$ in equation (27) is known. Then, transform the model in the following way:

$$R_t^{1/2}y_t = R_t^{1/2}(I_M \otimes x_t)\beta_t + \eta_t, \quad \eta_t \sim NID(0_M, I_M).$$

Next, specify a prior for $\beta_t$ of the form:

$$\beta_j \sim \mathcal{N}\left(\beta_{0j}, V_{0j}^{-1}\right) \text{ for } j \in \{1, ..., T\}.$$

Combining the prior with the local likelihood in (29) yields a quasi-posterior density for $\beta_j$ of the form

$$p(\beta_j|Y, X, R_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[\sum\nolimits_{t=1}^T \vartheta_{jt}(\widetilde{y}_t - \widetilde{x}_t\beta_j)'(\widetilde{y}_t - \widetilde{x}_t\beta_j) + (\beta_j - \beta_{0j})'V_{0j}(\beta_j - \beta_{0j})\right]\right\}$$

where $\widetilde{x}_t = R_t^{1/2}(I_M \otimes x_t)$ and $\widetilde{y}_t = R_t^{1/2}y_t$. A direct application of identity (31) implies that the conditional quasi-posterior of $\beta_j$ is $\mathcal{N}\left(\widetilde{\beta}_j, \widetilde{V}_j^{-1}\right)$ :

$$p(\beta_j|Y, X, R_{1:T}) \propto \exp\left[-\frac{1}{2}\left(\beta_j - \widetilde{\beta}_j\right)'\widetilde{V}_j\left(\beta_j - \widetilde{\beta}_j\right)\right]$$

with posterior parameters

$$\widetilde{\beta}_j = \widetilde{V}_j^{-1}\left[\sum\nolimits_{t=1}^T \vartheta_{jt}\widetilde{x}_t'\widetilde{y}_t + V_{0j}\beta_{0j}\right], \quad \widetilde{V}_j = V_{0j} + \sum\nolimits_{t=1}^T \vartheta_{jt}\widetilde{x}_t'\widetilde{x}_t \tag{42}$$

which proves Proposition 4.

### 1.6 Proof of Proposition 5

Assume that $\beta_t$ is known in the model in equation (27). Write the model as:

$$\varepsilon_t := y_t - (I_M \otimes x_t)\beta_t = R_t^{-1/2}\eta_t, \ \eta_t \sim NID(0_M, I_M),$$

and specify a Wishart $\mathcal{W}(\alpha_{0j}, \gamma_{0j}^{-1})$ prior density for $R_t^{-1}$. Combine the prior with the local likelihood (29) to obtain conditional quasi-posterior density for $R_j$ of the form

$$p(R_j|Y, X, \beta_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T} \vartheta_{jt}\varepsilon_t{}' R_j \varepsilon_t + tr(\gamma_{0j}R_j)\right]\right\}|R_j|^{\sum_{t=1}^{T}\vartheta_{jt}/2}|R_j|^{(\alpha_{0j}-M-1)/2} \quad (43)$$

$$= etr\left\{-\frac{1}{2}\left(\sum_{t=1}^{T} \vartheta_{jt}\varepsilon_t\varepsilon_t{}' + \gamma_{0j}\right)R_j)\right\}|R_j|^{(\sum_{t=1}^{T}\vartheta_{jt}+\alpha_{0j}-M-1)/2}, \quad (44)$$

implying that the conditional quasi-posterior density for $R_j$ in (43) is of $\mathcal{W}(\widetilde{\alpha}_j, \widetilde{\gamma}_j^{-1})$ form:

$$p(R_j|Y, X, \beta_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[tr(\widetilde{\gamma}_j R_j)\right]\right\}|R_j|^{\frac{\widetilde{\alpha}_j - M - 1}{2}}$$

with posterior parameters

$$\widetilde{\alpha}_j = \alpha_{0j} + \sum_{t=1}^{T} \vartheta_{jt}, \quad \widetilde{\gamma}_j = \gamma_{0j} + \sum_{t=1}^{T} \vartheta_{jt}\varepsilon_t'\varepsilon_t. \quad (45)$$

## 2 Algorithms

### 2.1 Gibbs Algorithm for homoscedastic BVAR model with time varying parameters

Consider an $M$-dimensional process $y_t$ defined by a VAR model with time-varying autoregressive parameters and time invariant volatility: $y_t = (I_M \otimes x_t)\beta_t + R^{-1/2}\eta_t, \ \eta \sim \mathcal{N}(0_M, I_M)$, where $x_t$ is a $1 \times Mk + 1$ vector process defined in (28) and $\beta_t$ is $M(Mk+1) \times 1$ vector of time varying coefficients.

A Gibbs algorithm that can sample from the joint quasi-posterior of $\beta_t$ and $R^{-1}$ can be constructed in the following way. Conditional on a draw $R$, the model can be re-defined as

$$\widetilde{y}_t = R^{1/2}y_t = \widetilde{x}_t\beta_t + \eta_t \quad (46)$$

where $\widetilde{x}_t := R^{1/2}(I_M \otimes x_t)$. Assuming a normal prior $\mathcal{N}(\beta_{0j}, V_{0j})$ for the parameter process $\beta_j$, the quasi-posterior distribution of $\beta_j$, established in Proposition 4, is also normal $\mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$ at each point in time $j$ with parameters given in (42). On the other hand, conditional on a draw from the history of $\beta_j$, $\beta_{1:T} = (\beta_1, ..., \beta_T)$, the model in (46) can be written as $\widetilde{\varepsilon}_t = y_t - (I_M \otimes X_t)\beta_t = R^{-1/2}\eta_t$. Then, assuming a Wishart prior distribution $\mathcal{W}(\alpha_0, \gamma_0^{-1})$ for $R^{-1}$ and stacking $\widetilde{E} = (\widetilde{\varepsilon}_1, \widetilde{\varepsilon}_2, ..., \widetilde{\varepsilon}_{tT})'$, the conditional posterior of $R^{-1}$ can be written as $R^{-1}|X, Y, \beta_{1:T} \sim \mathcal{W}(\widetilde{\alpha}, \widetilde{\gamma}^{-1})$, where

$$\widetilde{\alpha} = \alpha_0 + T, \quad \widetilde{\gamma} = \gamma_0 + \widetilde{E}'\widetilde{E}. \quad (47)$$

The conditional posterior of $R^{-1}$ is of standard form and the conditional quasi-posterior distribution of $\beta_j$, characterised in Proposition 4, can be easily drawn from; hence, the estimation of the model in (46) permits the use of a Gibbs algorithm with the following steps.

**Algorithm 1.**

**Step 1.** Initialise the algorithm with a guess, $R^{-1,0}$.

Then for $i = 1, ..., N$, iterate between steps 2 and 3.

**Step 2.** Draw $\beta_j^i|Y, X, R^{-1,i-1}$ from $\mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$ with posterior parameters defined in (42) for each point in time $j \in \{1, ..., T\}$.

**Step 3.** Draw $R^{-1,i}|X, Y, \beta_{1:T}^i$ from $\mathcal{W}(\widetilde{\alpha}, \widetilde{\gamma}^{-1})$ with posterior parameters defined in (47).

Standard MCMC results apply as follows. Since the form of the quasi-posterior distributions has been established and since the steps above constitute a Markov chain, iterating between step 2 and 3 results into convergence of the chain to its stationary distribution and hence, draws from the algorithm (after discarding a fair proportion of initial draws) can be used for an approximation of the joint quasi-posterior distribution of $\beta_j$ and $R^{-1}$.

## 2.2 Gibbs Algorithm for heteroscedastic BVAR model with time invariant parameters

Consider an $M$-dimensional process $y_t$ defined by the model

$$y_t = (I_M \otimes x_t)\beta + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R_t^{-1}) \tag{48}$$

where $x_t$ is a $1 \times Mk + 1$ vector process defined in (28) and $\beta$ is $M(Mk+1) \times 1$ vector of time invariant coefficients. The error term $\varepsilon_t$ has a time varying covariance matrix $R_t^{-1}$. Conditional on observing a draw of $R_{1:T}^{-1}$ from the history $(R_1^{-1}, ..., R_T^{-1})$, the model reduces to a GLS problem with known time varying covariance matrix. Pre-multiplying equation (48) with $R_t^{1/2}$, we obtain $R_t^{1/2} y_t = R_t^{1/2}(I_M \otimes x_t)\beta + \eta_t$, $\eta_t \sim \mathcal{N}(0_M, I_M)$, and after defining $\widehat{y}_t := R_t^{1/2} y_t$ and $\widehat{x}_t := R_t^{1/2}(I_M \otimes x_t)$, $\widehat{y}_t$ follows a homoscedastic model $\widehat{y}_t = \widehat{x}_t\beta + \eta_t$. Assuming a normal prior $\mathcal{N}(\beta_0, V_0)$ for $\beta$, the posterior distribution is also normal $\mathcal{N}(\widetilde{\beta}, \widetilde{V})$, with parameters

$$\widetilde{V} = \left(\sum_{t=1t}^{T} \hat{x}_t'\hat{x}_t + V_0^{-1}\right)^{-1}, \quad \widetilde{\beta} = \widetilde{V}\left(\sum_{t=1}^{T} \hat{x}_t'\hat{y}_t + V_0^{-1}\beta_0\right). \tag{49}$$

Conditional on observing the coefficients $\beta$, the model simplifies to $\varepsilon_t = y_t - (I_M \otimes X_t)\beta$ where $\varepsilon_t \sim \mathcal{N}(0_M, R_t^{-1})$ and is also observed. Then, by assuming a Wishart prior $\mathcal{W}(\alpha_{0j}, \gamma_{0j}^{-1})$ for $R_j^{-1}$ at each point in time $j \in \{1, ..., T\}$, Proposition 5 implies that the conditional quasi-posterior of $R_j^{-1}$ at each point in time $j$ is also Wishart $\mathcal{W}(\widetilde{\alpha}_j, \widetilde{\gamma}_j^{-1})$, with parameters given in (45).

Since, the conditional posterior of $\beta$ is of standard form and the quasi-posterior distribution of $R_j^{-1}$ has been characterised in Section 3, a Gibbs algorithm can be constructed in order to recursively draw from the conditional posterior distributions of $\beta$ and $R_{1:T}^{-1}$ to approximate their joint posterior distribution. The algorithm consists of three steps.

**Algorithm 2.**

**Step 1**. Initialise the algorithm with $\beta^0$.

For $i = 1, ..., N$ iterate between steps 2 and 3.

**Step 2.** For each $j \in \{1, ..., T\}$ draw $R_j^{-1,i}|X, Y, \beta^{i-1}$ from $\mathcal{W}(\widetilde{\alpha}_j, \widetilde{\gamma}_j^{-1})$ with posterior parameters defined in (45).

**Step 3.** Draw $\beta^i|Y, X, R_{1:T}^{-1,i}$ from $\mathcal{N}(\widetilde{\beta}, \widetilde{V})$ with posterior parameters defined in (49).

Note that it is easy to extend the analysis and consider models in which only a subset of the regressors enter with time varying coefficients.

### 2.3 Gibbs Algorithm for a TVP structural BVAR model

Algorithms 1 and 2 involve reduced-form VAR models. In macroeconomic applications, structural VARs (SVARs) are often applied to analyse macroeconomic shocks and the transmission of these shocks to key variables. SVAR models orthogonalise the VAR residuals so that the resulting 'structural' shocks have a diagonal covariance matrix. The simplest such orthogonalisation of the $M \times M$ covariance matrix $R_t^{-1}$ considered in the literature involves a Cholesky decomposition. In particular, consider the model:

$$y_t = (I_M \otimes X_t)\beta_t + \varepsilon_t, \quad var(\varepsilon_t) = R_t^{-1} = A^{-1}\Omega_t^{-1}\left(A^{-1}\right)' \tag{50}$$

where $\Omega_t^{-1}$ is diagonal matrix with elements $\omega_{it}$, varying over time, and $A$ is a lower triangular matrix with ones on its main diagonal. Providing a Bayesian treatment for the $M(M-1)/2$ non-zero elements of $A$ is straightforward. As shown in Cogley and Sargent (2002), the VAR model can be written as

$$A(y_t - (I_M \otimes X_t)\beta_t) := y_t^* = \Omega_t^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}(0, I_M). \tag{51}$$

As argued in Primiceri (2005), the researcher might also want to allow the orthogonalisation scheme of the structural shocks to be time-varying. In Primiceri (2005), this involves adding the non zero and one elements of $A_t$ to the state vector and specifying stochastic processes for them (in his application, these follow a random walk process). Below, we present an alternative algorithm that can provide estimates over time of $A_t$, by employing the nonparametric kernel-based QBLL method, which does not require specifying the parameter process.

Conditioning on $\beta_t$ and $\Omega_t$, the model in (50) with drifting lower triangular matrix $A_t$ simplifies to a set of $M-1$ equations in (51), $i = 2, ..., M$, with equation $i$ having $y_{i,t}^*$ as a dependent variable and $-y_{m,t}^*$ as regressor, $m = 1, ..., i-1$:

$$\underbrace{\omega_{it}^{-1}y_{i,t}^*}_{\widehat{Y}_{it}} = \underbrace{-\omega_{it}^{-1}[y_{1,t}^*, ..., y_{i-1,t}^*]}_{\widehat{X}_{it}} \underbrace{\begin{bmatrix} a_{i1,t} \\ \vdots \\ a_{i,i-1,t} \end{bmatrix}}_{a_{it}} + \eta_{it}.$$

Assuming a normal prior for the vector $a_{ij}$, the coefficients of the $i^{th}$ row of the matrix $A_j$, at each point in time $j \in \{1, ..., T\}$ of the form $\mathcal{N}(a_{0,ij}, \lambda_{0,ij})$, the quasi-posterior distribution of $a_{ij}$ is also normal $\mathcal{N}\left(\widetilde{a}_{ij}, \widetilde{V}_{ij}\right)$ ,with parameters

$$\widetilde{V}_{ij} = \left(\lambda_{0,ij}^{-1} + \sum_{t=1}^{T} \vartheta_{jt} \widehat{X}_{it}' \widehat{X}_{it}\right)^{-1}, \quad \widetilde{a}_{ij} = \widetilde{V}_{ij}^{-1} \left(\sum_{t=1}^{T} \vartheta_{jt} \widehat{X}_{it}' \widehat{Y}_{it} + \lambda_{0,ij}^{-1} a_{0,ij}\right). \tag{52}$$

It is straightforward to build this step into one of the previously outlined Gibbs algorithms (with or without time variation in the VARs autoregressive parameters or in the volatility) in order to draw $\beta_t$, $\Omega_t$ and $A_t$ from their conditionally conjugate quasi-(or standard) posterior distributions in order to approximate their joint quasi-posterior distribution. The kernel approach could be extended to the estimation of not necessarily just-identified or recursive time varying identifying restrictions.

# 3 Additional details and results

## 3.1 Details on the Monte Carlo Design

For the Monte Carlo design, we consider sample sizes of 100, 500 and 1000 and, due to computational considerations, 300 replications for each DGP and sample size are simulated. In each case, we compute the bias and the RMSE of the estimators. For simplicity, the average bias and the root of the MSE over time are reported in Tables 1-4 of the paper, summarised by averaging over the autoregressive and covariance parameters respectively. We also report the 95% coverage rates for the parameter estimates, computed as the proportion of time that the true parameter finds itself in the 95% posterior confidence intervals implied by the different models. For the QBLL approach, we use the means of the quasi-posterior density as point estimates, the normal kernel for the kernel weights with bandwidth of $T^{0.5}$ and a Minnesota type prior[1] with loose overall shrinkage $\lambda = 1$. Our choice of bandwidth $H = T^{0.5}$ is motivated by the optimal bandwidth choice used for inference in time varying random coefficient models, see Giraitis, Kapetanios and Yates (2014). For all state space models in the Monte Carlo exercise, we use the algorithm outlined in Cogley and Sargent (2005) with 3,000 Gibbs draws (from which, the first 1,000 have been discarded), modified to allow for time varying Cholesky lower triangular matrix, whose elements are modelled as a random walk, as in Primiceri (2005). The priors and initial values of the state space models are set using the initial 10% of the observations as a presample[2]. We have performed a robustness check with respect to the lag length and the main results of the paper do not change much. In practice, increasing the number of lags creates a number of problems for state space models. First, it increases the state space which makes estimation considerably slower. Second, drawing a stable draw with state space models becomes very hard unless extra shrinkage is added on the random walk state equations. If the code checks stability, then this second point increases computation time for the state space models further.

[1] The priors for the autoregressive parameters are as in Bańbura, Giannone and Reichlin (2010) and the Wishart prior parameters for the volatilities are as in Kadiyala and Karlsson (1997).

[2] In a previous version of the paper, we presented a set of results where the state space model's coverage rates were considerably below the nominal 95% and deteriorated with the sample size, implying the presence of Type I errors. These previous results are available upon request. In the current version of the paper, we removed the stability condition on the autoregressive matrix of the state space models for all DGPs and considerably loosened the prior controlling the shrinkage on the standard deviation parameter in the state equation. This delivered better coverage rates, coupled with a slightly worse RMSE performance and a sizeable proportion of non-stable draws.

## 3.2 Additional Details and Results for the Application

For the empirical application of the paper, we estimate a TVP BVAR(2) model using the macroeconomic series in Primiceri (2005) (annual GDP deflator inflation, civilian unemployment rate and the secondary market rate on the 3-month Treasury bills[3]) with the addition of commodity prices, measured as annual growth of Moody's commodity price index. We use a Normal-inverted-Wishart conjugate prior with overall Minnesota shrinkage coefficient $\lambda = 0.2$ for the autoregressive parameters as in Bańbura et al. (2010) and prior for the Wishart parameters as in Kadiyala and Karlsson (1997). A range of robustness checks were performed. First, our results are robust to different lag orders of the VAR model. In addition, the model was estimated with CPI inflation instead of GDP deflator; Fed Funds rate instead of the 3-month Treasury bill, GDP growth instead of unemployment and oil prices instead of commodity prices. The main results do not change with these different specifications. Our results are also robust to different values of the overall shrinkage coefficient, $\lambda$: we estimated the model with $\lambda = 0.05$, 0.1, 0.5, 1 and 2. The figures below provide supplementary material for the empirical application in the paper. Figure 1 presents the conditional contemporaneous covariances over time.

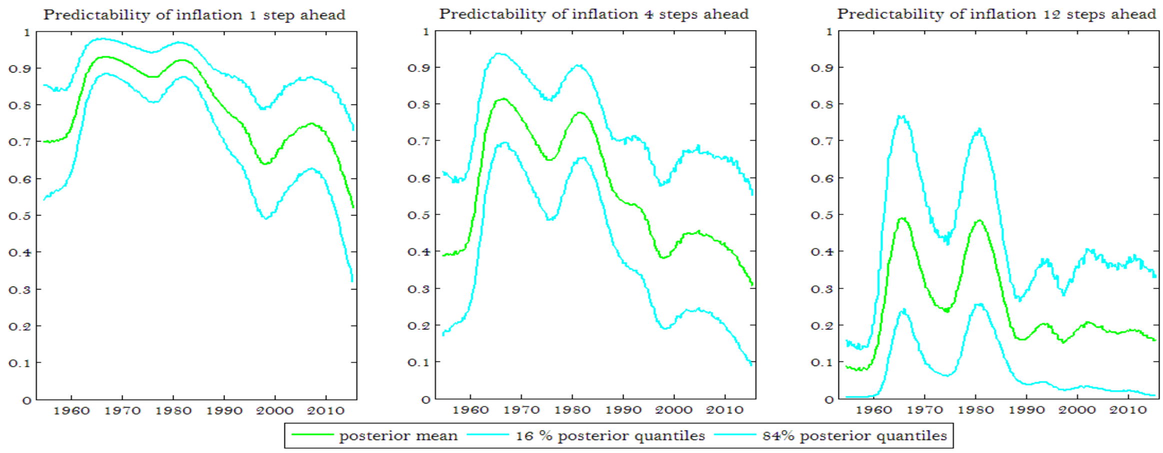
Figure 1: Off-diagonal covariance matrix elements over time


Figure 2 displays the posterior mean of inflation persistence for selected horizons with 16% and 84% confidence bands. The left panel of Figure 2 looks similar to the reported inflation persistence in Cogley, Primiceri and Sargent (2010) for one quarter ahead forecasts. However, their TVP VAR model delivers draws for $R^2_{t,1}$ which are very closely clustered at one around 1980 which they interpret as possible misspecification, while from the left panel of Figure 2, it is clear that

[3]Quarterly series for the unemployment rate and the nominal interest rate are computed as 3-month averages.

the QBLL confidence bands for $R^2_{t,1}$ are bounded away from one. Another difference is that the confidence bands for $R^2_{t,h}$ in Figure 2, especially at medium horizons, are much narrower than the ones presented in Cogley et al. (2010), suggesting more precise estimates.

Figure 2: Inflation Persistence




### 3.3 Additional Details on the Forecasting Exercise

#### 3.3.1 Forecasting with QBLL VAR

We illustrate how a VAR model estimated with the QBLL method can be employed to generate out-of-sample forecasts. Since the method does not require parametric assumptions about the parameter processes, the parameters are kept fixed at the last insample period. In particular, a draw from the predictive density of $y_t$ at $h$ periods in the future is given by:

$$\widehat{y}^i_{T+h} = B^i_{0T} + \Sigma^k_{p=1} B^i_{pT} \widehat{y}^i_{T+h-p} + R_T^{-1/2,i} \eta^i_{T+h},$$

where $B^i_{0T}$, $B^i_{1:k,T}$ and $R_T^{-1/2,i}$ are draws from the quasi-posterior densities, derived in Propositions 2-5, at period $T$ and $\eta^i_{T+1:T+h}$ is a sequence of draws from a multivariate standard Normal.

#### 3.3.2 Additional Estimation Details

The dataset for the forecasting exercise contains 87 quarterly macroeconomic series in first difference[4] and is described in Table 1 below. The forecast origins range from 1970Q2 to 2012Q2 and we compute forecasts for one up to eight quarters ahead. All QBLL BVAR models in the forecasting exercise of the paper are of lag order one and use a Normal-inverted-Wishart conjugate prior with overall shrinkage for the autoregressive parameters as in Bańbura et al. (2010) and prior Wishart parameters as in Kadiyala and Karlsson (1997). We use an overall shrinkage parameter for the small

[4]The choice of using a VAR model in first difference is motivated by results in Carriero, Galvão and Kapetanios (2015), who show in a forecast evaluation containing various models and countries, that their BVAR model in differences performs better on average than the one in levels.

models of 0.3, for the medium models 0.2, and for the large models 0.11, inversely proportional to the model size. A grid for other values of $\lambda$ were considered for each model and size and whereas the absolute forecast performance of the models depends on $\lambda$, the relative models' performance presented in this section is robust to different values of $\lambda$. We also performed a robustness check with a grid for the bandwidth parameter and the main results of the paper do not change. For the three different state space models in the forecasting exercise, we use versions of the algorithms outlined in Cogley and Sargent (2005) with 10,000 Gibbs draws from which the first 5,000 have been discarded. The priors and initial values are set using a presample of 30 observations. Accuracy of point forecasts is measured using the root mean squared forecast error (RMSFE) and forecast bias. Forecast density performance is evaluated using log predictive scores (LPS) and probability integral transformation (PIT). The LPS and PITs are computed with the help of a nonparametric estimator to smooth the draws from the predictive density obtained for each forecast and horizon. We test whether a model is statistically more accurate than the benchmark against the two-sided alternative, with the Diebold and Mariano (1995)'s statistic computed with Newey-West estimator to obtain standard errors. The results of the Diebold-Mariano test are provided for the RMSFEs and LPS. For the bias, we test whether the models' bias is statistically different from zero. In addition, uniformity of the PITs is assessed using the test statistic of Berkowitz (2001), with a null hypothesis of uniform PITs.

## References


Bańbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions, *Journal of Applied Econometrics* **25**(1): 71–92.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics* **19**(4): 465–474.

Bickel, P. J. and Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions, *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**(4): 257–276.

Carriero, A., Galvão, A. and Kapetanios (2015). A comprehensive evaluation of macroeconomic forecast methods, *Working Paper* .

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation, *Journal of Econometrics* **115**: 239–346.

Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010). Inflation-gap persistence in the US, *American Economic Journal: Macroeconomics* **2(1)**: 43–69.

Cogley, T. and Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics, *in* B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 331–88.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post World War II US, *Review of Economic Dynamics* **8**: 262–302.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–263.

Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models, *Journal of Econometrics* **179**(1): 46–65.

Ibragimov, I. and Has'minskii, R. (1981). *Statistical Estimation: Asymptotic Theory*, Springer.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models, *Journal of Applied Econometrics* **12**(2): 99–132.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Mathematical Statistics.

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

| | Series Description | Transformation |
|---|---|---|
| 1 | Real GDP | DLOG(GDP)*100 |
| 2 | GDP Deflator | DLOG(GDPDEF)*100 |
| 3 | Consumer Price Index | DLOG(CPI)*100 |
| 4 | Business Investment | DLOG(INV) |
| 5 | Real Personal Consumption Expenditure | DLOG(CONS)*100 |
| 6 | Civilian Unemployment Rate | D(UNRATE/4) |
| 7 | Industrial Production | DLOG(IP)*100 |
| 8 | 3-Month Treasury Bill: Secondary Market Rate | D(TBILL/4) |
| 9 | ISM Manufacturing: PMI Composite Index | DLOG(PMI) |
| 10 | ISM Manufacturing: New Orders Index | DLOG(NEWORDERS) |
| 11 | Average Hours | DLOG(AVERAGEHOURS)*100 |
| 12 | Nonfarm Business Sector: Real Compensation Per Hour | DLOG(REALWAGE)*100 |
| 13 | Producer Price Index | DLOG(PPI)*100 |
| 14 | Personal Consumption Expenditures: Chain-type Price Index | DLOG(PCE)*100 |
| 15 | Reuters/Jeffries-CRB Total Return Index (w/GFD extension) | DLOG(COMINDEX) |
| 16 | BAA Corporate Spread | D(CORPSPREAD) |
| 17 | NYSE Stock Market Capitalization | DLOG(STOCKCAP) |
| 18 | Industrial Production: Business Equipment | DLOG(IPBUSEQ)*100 |
| 19 | Industrial Production: Consumer Goods | DLOG(IPCONGD)*100 |
| 20 | Industrial Production: Durable Consumer Goods | DLOG(IPDCONGD)*100 |
| 21 | Industrial Production: Durable Materials | DLOG(IPDMAT)*100 |
| 22 | Industrial Production: Final Products (Market Group) | DLOG(IPFINAL)*100 |
| 23 | Industrial Production: Final Products and Nonindustrial Supplies | DLOG(IPFPNSS)*100 |
| 24 | Industrial Production: Manufacturing (SIC) | DLOG(IPMANSICS)*100 |
| 25 | Industrial Production: Materials | DLOG(IPMAT)*100 |
| 26 | Industrial Production: Nondurable Consumer Goods | DLOG(IPNCONGD)*100 |
| 27 | Dow Jones Industrial Total returns index | DLOG(DOW) |
| 28 | ISM Manufacturing: Inventories Index | DLOG(INVENTORIES) |
| 29 | ISM Manufacturing: Supplier Deliveries Index | DLOG(SUPPLIERS) |
| 30 | ISM Manufacturing: PMI Composite Index | DLOG(NAPM) |
| 31 | ISM Manufacturing: PMI Employment index | DLOG(NAPMEI) |
| 32 | ISM Manufacturing: Production index | DLOG(NAPMPI) |
| 33 | ISM Manufacturing: Prices index | DLOG(NAPMPRI) |
| 34 | Civilian Employment | DLOG(EMPLOY)*100 |
| 35 | All Employees: Construction | DLOG(USCONS)*100 |
| 36 | All Employees: Financial Activities | DLOG(USFIRE)*100 |
| 37 | All Employees: Good producing industries | DLOG(USGOOD)*100 |
| 38 | All Employees: Government | DLOG(USGOVT)*100 |
| 39 | All Employees: Trade and Transportation | DLOG(USTPU)*100 |
| 40 | All Employees: retail trade | DLOG(USTRADE)*100 |
| 41 | All Employees: wholesale trade | DLOG(USWTRADE)*100 |
| 42 | All Employees: Durable Goods | DLOG(DMANEMP)*100 |
| 43 | All Employees: Manufacturing | DLOG(MANEMP)*100 |
| 44 | All Employees: Non-Durable Goods | DLOG(NDMANEMP)*100 |
| 45 | All Employees: Service Providing industries | DLOG(SRVPRD)*100 |
| 46 | Total Non-Farm Payrolls | DLOG(PAYEMS)*100 |
| 47 | Real Personal Incomes excluding current transfers | DLOG(W875RX1)*100 |
| 48 | Business Conditions Index | DLOG(BCI) |
| 49 | Real Imports | DLOG(IMPORTS)*100 |
| 50 | Real Exports | DLOG(EXPORTS)*100 |
| 51 | Real Government Spending | DLOG(REALGS)*100 |
| 52 | Real Net Taxes | DLOG(REALTAX) |
| 53 | Number of Civilians unemployed for 15 weeks or over | DLOG(UEMP15OV) |
| 54 | Number of Civilians unemployed for 15 to 26 weeks | DLOG(UEMP15T26) |
| 55 | Number of Civilians unemployed 27 weeks and over | DLOG(UEMP27OV) |
| 56 | Number of Civilians unemployed for 5 to 14 weeks | DLOG(UEMP5TO14) |
| 57 | Number of Civilians unemployed for less than 5 weeks | DLOG(UEMPLT5) |
| 58 | Average Mean Duration of employment | DLOG(UEMPMEAN) |
| 59 | Average Weekly Hours of Production and Nonsupervisory Employees:Goods-Producing | DLOG(CES0600000007)*100 |
| 60 | Average Hourly Earnings of Production and Nonsupervisory Employees:Goods-Producing | DLOG(CES0600000008)*100 |
| 61 | Average Hourly Earnings of Production and Nonsupervisory Employees:Construction | DLOG(CES2000000008)*100 |
| 62 | Average Weekly hours of Production and Nonsupervisory Employees:Manufacturing | DLOG(CES3000000008)*100 |
| 63 | Average Hourly Earnings of Production and Nonsupervisory Employees:Manufacturing | DLOG(AWHMAN)*100 |
| 64 | Civilian Labour Force | DLOG(CLF)*100 |
| 65 | Civilian Participation rate | DLOG(CIVPART)*100 |
| 66 | Nonfarm Business Sector: Unit Labor Cost | DLOG(ULC)*100 |
| 67 | M2 Money Stock | DLOG(M2)*100 |
| 68 | Total Consumer Credit Owned and Securitized, Outstanding | DLOG(CREDIT)*100 |
| 69 | Commercial and Industrial Loans, All Commercial Banks | DLOG(BUSLOANS)*100 |
| 70 | Real Estate Loans, All Commercial Banks | DLOG(REALLN)*100 |
| 71 | Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals | DLOG(PPICMM)*100 |
| 72 | Producer Price Index: Crude Materials for Further Processing | DLOG(PPICRM)*100 |
| 73 | Producer Price Index: Finished Consumer Goods | DLOG(PPIFCG)*100 |
| 74 | Producer Price Index: Finished Goods | DLOG(PPIFGS)*100 |
| 75 | Producer Price Index: Intermediate Materials: Supplies & Components | DLOG(PPIITM)*100 |
| 76 | Consumer Price Index for All Urban Consumers: Apparel | DLOG(CPIAPPSL)*100 |
| 77 | Consumer Price Index for All Urban Consumers: Medical Care | DLOG(CPIMEDSL)*100 |
| 78 | Consumer Price Index for All Urban Consumers: All items less shelter | DLOG(CUUR0000SA0L2_D11)*100 |
| 79 | 10 year Govt Bond Yield minus 3 mth yield | D(GB10-TBILL) |
| 80 | 6-month Treasury bill minus 3 mth yield | D(TBILL6-TBILL) |
| 81 | 1 year Govt Bond Yield minus 3 mth yield | D(GB1-TBILL) |
| 82 | 5 year Govt Bond Yield minus 3 mth yield | D(GB5-TBILL) |
| 83 | AAA Corporate Bond Spread | D(AAA-GB10) |
| 84 | S&P500 Total Return Index | DLOG(STOCK) |
| 85 | S&P500 P/E Ratio | D(PE) |
| 86 | US UK Exchange Rate | DLOG(DOLLARATE) |
| 87 | US Canada exchange rate | DLOG(CNRATE)*100 |

Table 1 Dataset. Large BVARs in the forecasting exercise are estimated with all 87 variables, medium BVARs include variables 1-17.