Online Appendix: Composite Likelihood Methods for Large Bayesian VARs with Stochastic Volatility*

Joshua C.C. Chan
Purdue University and UTS

Eric Eisenstat
The University of Queensland

Chenghan Hou
Hunan University

Gary Koop
University of Strathclyde

May 2020

*We would like to thank participants at Workshops on Forecasting at the Deutsche Bundesbank and the National Bank of Poland as well as seminar participants at Heriot-Watt and Essex Universities for helpful comments. Gary Koop is a Senior Fellow at the Rimini Center for Economic Analysis. Joshua Chan and Eric Eisenstat would like to acknowledge financial support by the Australian Research Council via a Discovery Project (DP180102373). Emails: joshuacc.chan@gmail.com, e.eisenstat@uq.edu.au, chenghan.hou@hotmail.com and gary.koop@strath.ac.uk.

A Technical Appendix

A.1 Proof of Proposition 1

Proof. Defining $\tilde{y}_t^* = A_{y,t}y_t^* - c_y - \sum_{j=1}^p B_{yy,j}y_{t-j}^*$ it is straightforward to show the form of the restricted VAR-SV implies:

$$
\begin{aligned}
p(y_t \mid \ \cdot\ ) \propto \exp&\left\{-\frac{1}{2}\left(\tilde{y}_t^* - \sum_{i=1}^M w_i \sum_{j=1}^p \frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)' \Sigma_{y,t}^{-1}\left(\tilde{y}_t^* - \sum_{i=1}^M w_i \sum_{j=1}^p \frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)\right\} \\
&\times \prod_{i=1}^M \exp\left\{-\frac{1}{2}\left[h_{N^*+i,t} - \ln w_i \right.\right. \\
&\qquad \left.\left. + \ e^{-h_{N^*+i,t}+\ln w_i}\left(z_{i,t} - \alpha_{z,i,t}'y_t^* - c_{z,i} - \sum_{j=1}^p \beta_{zy,i,j}'y_{t-j}^* - \sum_{j=1}^p \beta_{zz,i,j}z_{i,t-j}\right)^2\right]\right\} \\
\propto \exp&\left\{\sum_{i=1}^M -\frac{w_i}{2}\left(\tilde{y}_t^* - \sum_{j=1}^p \frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)' \Sigma_{y,t}^{-1}\left(\tilde{y}_t^* - \sum_{j=1}^p \frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)\right\} \\
&\times \exp\left\{-\frac{1}{2g(M)^2}\left(\sum_{i=1}^M w_i \sum_{j=1}^p \beta_{yz,i,j}z_{i,t-j}\right)' \Sigma_{y,t}^{-1}\left(\sum_{i=1}^M w_i \sum_{j=1}^p \beta_{yz,i,j}z_{i,t-j}\right)\right\} \\
&\times \exp\left\{\sum_{i=1}^M \frac{w_i}{2g(M)^2}\left(\sum_{j=1}^p \beta_{yz,i,j}z_{i,t-j}\right)' \Sigma_{y,t}^{-1}\left(\sum_{j=1}^p \beta_{yz,i,j}z_{i,t-j}\right)\right\} \\
&\times \prod_{i=1}^M \exp\left\{-\frac{1}{2}\left[h_{N^*+i,t} - \ln w_i \right.\right. \\
&\qquad \left.\left. + \ e^{-h_{N^*+i,t}+\ln w_i}\left(z_{i,t} - \alpha_{z,i,t}'y_t^* - c_{z,i} - \sum_{j=1}^p \beta_{zy,i,j}'y_{t-j}^* - \sum_{j=1}^p \beta_{zz,i,j}z_{i,t-j}\right)^2\right]\right\},
\end{aligned}
$$

where we used the fact that $(y_t^*)'\Sigma_{y,t}^{-1}(y_t^*) = \sum_{i=1}^M w_i (y_t^*)'\Sigma_{y,t}^{-1}(y_t^*)$. The likelihood of the restricted VAR-SV is

$$
L(y;\theta) = \prod_{t=1}^T p(y_t \mid \ \cdot\ ). \tag{1}
$$

Now, suppose that our composite likelihood is constructed from sub-models:

$$A_{y,t}y_t = c_y + \sum_{j=1}^{p} B_{yy,j}y^*_{t-j} + \sum_{j=1}^{p} \frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)} + \epsilon_{y,t}, \qquad \epsilon_{y,t} \sim N(0, \Sigma_{y,t}), \tag{2}$$

$$z_{i,t} - \alpha'_{z,i,t}y^*_t = c_{z,i} + \sum_{j=1}^{p} \beta'_{zy,j}y^*_{t-j} + \sum_{j=1}^{p} \beta'_{zz,i,j}z_{i,t-j} + \epsilon_{z,i,t}, \quad \epsilon_{z,i,t} \sim N\left(0, e^{h_{N*+i,t}}\right), \tag{3}$$

which leads to

$$\begin{aligned} p^C(y_t \mid \ \cdot\ ) \propto \exp\left\{\sum_{i=1}^{M} -\frac{w_i}{2}\left(\tilde{\mathbf{y}}^*_t - \sum_{j=1}^{p}\frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)' \Sigma^{-1}_{y,t}\left(\tilde{\mathbf{y}}^*_t - \sum_{j=1}^{p}\frac{\beta_{yz,i,j}z_{i,t-j}}{g(M)}\right)\right\} \\ \times \prod_{i=1}^{M} \exp\left\{-\frac{1}{2}\left[w_i h_{N^*+i,t}\right.\right. \\ \left.\left. +\ e^{-h_{N^*+i,t}+\ln w_i}\left(z_{i,t} - \alpha'_{z,i,t}y^*_t - c_{z,i} - \sum_{j=1}^{p}\beta'_{zy,i,j}y^*_{t-j} - \sum_{j=1}^{p}\beta_{zz,i,j}z_{i,t-j}\right)^2\right]\right\} \end{aligned}$$

and the composite likelihood $L^C(y;\theta) = \prod_{t=1}^{T} p^C(y_t \mid \ \cdot\ )$.

Observe that

$$\begin{aligned} L^C(y;\theta) &\propto L(y;\theta) \\ &\quad \times \exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^{T}\left[\sum_{i=1}^{M} w_i\left(\sum_{j=1}^{p}\beta_{yz,i,j}z_{i,t-j}\right)' \Sigma^{-1}_{y,t}\left(\sum_{j=1}^{p}\beta_{yz,i,j}z_{i,t-j}\right)\right.\right. \\ &\quad \left.\left. -\left(\sum_{i=1}^{M} w_i\sum_{j=1}^{p}\beta_{yz,i,j}z_{i,t-j}\right)' \Sigma^{-1}_{y,t}\left(\sum_{i=1}^{M} w_i\sum_{j=1}^{p}\beta_{yz,i,j}z_{i,t-j}\right)\right]\right\} \\ &\propto L(y;\theta)\exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^{T}\tilde{z}'_t\Xi_t\tilde{z}_t\right\}, \end{aligned}$$

where $\tilde{z}_t = (z_{1,t-1}, \ldots, z_{1,t-p}, \ldots, z_{M,t-1}, \ldots, z_{M,t-p})'$, $B_i = (\beta_{yz,i,1}, \ldots, \beta_{yz,i,p})$, and Ξ_t is a $Mp \times Mp$ positive semi-definite matrix with the (i,k) block given by

$$\Xi_{ik,t} = \begin{cases} w_i(1-w_i)B'_i\Sigma^{-1}_{y,t}B_i & \text{if} \quad i = k, \\ -w_i w_k B'_i\Sigma^{-1}_{y,t}B_k & \text{if} \quad i \neq k. \end{cases}$$

Let $\tilde{z}_i = (z_{i,1}, \ldots, z_{i,T-1})'$, $\tilde{z} = (z_1', \ldots, z_M')'$, $z_T = (z_{i,T}, \ldots, z_{M,T})'$ and $y^* = ((y_1^*)', \ldots, (y_T^*)')'$. Then, we may write the likelihood $L(y;\theta)$ as the density $L(y;\theta) = p(y^*, z_T, \tilde{z} \,|\, \vartheta)$. Consequently,

$$\begin{aligned}\tilde{L}^C(y;\theta) &= \frac{p(y^*, z_T, \tilde{z} \,|\, \theta) \exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right\}}{\int_{\tilde{z}}\int_{\mathbf{y}^*,z_T} p(y^*, z_T, \tilde{z} \,|\, \theta) d(y^*, z_T) \exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right\} d\tilde{z}} \\ &= \frac{p(y^*, z_T, \tilde{z} \,|\, \theta) \exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right\}}{\mathrm{E}_{\tilde{z}}\left(\exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right\}\right)},\end{aligned}$$

and

$$D_{\mathrm{KL}}(L\|\tilde{L}^C) = \ln \mathrm{E}_{\tilde{z}}\left(\exp\left\{-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right\}\right) - \mathrm{E}_{\tilde{\mathbf{z}}}\left(-\frac{1}{2g(M)^2}\sum_{t=1}^T \tilde{z}_t'\Xi_t\tilde{z}_t\right).$$

To prove that $D_{\mathrm{KL}}(L\|\tilde{L}^C) \longrightarrow 0$ as $M \longrightarrow \infty$, note that Ξ_t can be represented by the Hadamard product $\tilde{\Xi}_t \odot (W \otimes \iota_p\iota_p')$, with the $M \times M$ matrix W defined by elements

$$W_{ik} = \begin{cases} w_i(1-w_i) & \text{if} \quad i = k \\ -w_i w_k & \text{if} \quad i \neq k \end{cases},$$

and $\iota_p = (1, \ldots, 1)'$ being the $p \times 1$ vector of ones. In particular, W is positive semi-definite and contains information regarding the weights, while $\tilde{\Xi}_{ik,t} = B_i'\Sigma_{y,t}^{-1}B_k$, for all i and k, depends only on the parameters.

Accordingly,

$$\frac{\tilde{z}_t'\Xi_t\tilde{z}_t}{g(M)^2} = \frac{\tilde{z}_t'\tilde{z}_t}{g(M)^2} \times \frac{\tilde{z}_t'\Xi_t\tilde{z}_t}{\tilde{z}_t'\tilde{z}_t} \leq \frac{\tilde{z}_t'\tilde{z}_t}{g(M)^2}\|\Xi_t\|,$$

where $\|\cdot\|$ denotes the spectral norm. Since $\tilde{\Xi}_t$ and $W \otimes \iota_p\iota_p'$ are positive semi-definite, Schur's inequality (Horn and Johnson, 1991, Theorem 5.5.1) implies $\|\Xi_t\| \leq p\|\tilde{\Xi}_t\|\|W\|$. Moreover, there exists a unit vector u (satisfying $u'u = 1$) such that

$$\|W\| = u'Wu = \sum_{i=1}^M w_i u_i^2 - \left(\sum_{i=1}^M u_i w_i\right)^2.$$

Since $\sum_{i=1}^M w_i u_i^2 \leq \max\{w_i\}\sum_{i=1}^M u_i^2 = \max\{w_i\}$ and $\left(\sum_{i=1}^M u_i w_i\right)^2 \geq 0$, we obtain $\|W\| \leq \max\{w_i\}$. Consequently, $\max\{w_i\} \longrightarrow 0$ implies $\|W\| \longrightarrow 0$ and $\|\Xi_t\| \longrightarrow 0$

follows from the fact that $\|\tilde{\Xi}_t\|$ is constant with respect to M.

It remains to show that $\frac{\tilde{z}_t'\tilde{z}_t}{g(M)^2} = \sum_{j=1}^{p} \frac{\sum_{i=1}^{M} z_{i,t-j}^2}{g(M)^2}$ does not diverge for fixed T and $M \longrightarrow \infty$. Since $z_{i,t-j}$ is normally distributed conditional on y^*, with conditional expectation $\mu_i(y^*) \equiv \mathrm{E}(z_{i,t-j} \,|\, y^*)$ and variance v_i^2, the quantity $\zeta_i = \frac{z_{i,t-j} - \mu_i(y^*)}{g(M)}$ is conditionally independently (though not identically) distributed, and has the following properties:

1. $\mathrm{E}(\zeta_i \,|\, y^*) = 0$,
2. $\mathrm{E}(\zeta_i^2 \,|\, y^*) = \frac{v_i^2}{g(M)^2}$,
3. $\sum_{i=1}^{M} \mathrm{Var}(\zeta_i \,|\, y^*) = \bar{v}\frac{M}{g(M)^2} < \infty$, where $\bar{v} = \frac{1}{M}\sum_{i=1}^{M} v_i^2$,
4. $\sum_{i=1}^{M} \mathrm{Var}(\zeta_i^2 \,|\, y^*) \leq 3\tilde{v}\frac{M}{g(M)^4} < \infty$, where $\tilde{v} = \frac{1}{M}\sum_{i=1}^{M} v_i^4$.

Hence $\sum_{i=1}^{M} \zeta_i$ and $\sum_{i=1}^{M} \zeta_i^2 - \bar{v}\frac{M}{g(M)^2}$ both converge in $\mathbb{R}$ almost surely (Durrett, 2010, Theorem 2.5.3), which implies $\frac{\sum_{i=1}^{M} z_{i,t-j}^2}{g(M)^2}$ converges in $\mathbb{R}$ almost surely. In this case, the product $\frac{\tilde{z}'\tilde{z}_t}{g(M)^2}\|\Xi_t\| \longrightarrow 0$ and $D_{\mathrm{KL}}(L\|\tilde{L}^C)$ vanishes in the limit. ■

A.2 Drawing the Quasi-Likelihood Specific Nuisance Parameters

To draw $\tilde{\eta}_i$ we use methods for the TVP-ARDL model defined by (**??**). Note that each $\tilde{\eta}_i$ is relatively low-dimensional and is independent of θ. Consequently, sampling $\tilde{\eta}_1, \ldots, \tilde{\eta}_M$ can be done in parallel and is fast in practice.

If we did not have to worry about the weights (i.e. $w_i = 1$ for $i = 1, .., M$), then sampling from $p^C(\tilde{\eta}_i \,|\, y^*, z_i)$ is equivalent to sampling from the TVP-ARDL posterior, which is standard. A typical Gibbs sampling algorithm cycles through:

1. $(\beta_{z_i} \mid \alpha_{z_i,1}, \ldots, \alpha_{z_i,T}, h_{N_*+i,1}, \ldots, h_{N_*+i,T})$;
2. $(\alpha_{z_i,1}, \ldots, \alpha_{z_i,T} \mid \beta_{z_i}, h_{N_*+i,1}, \ldots, h_{N_*+i,T}, \alpha_{z_i,0}, \Sigma_{\alpha,i})$;
3. $(\alpha_{z_i,0} \mid \Sigma_{\alpha,i}, \alpha_{z_i,1}, \ldots, \alpha_{z_i,T})$;
4. $(\Sigma_{\alpha,i} \mid \alpha_{z_i,0}, \alpha_{z_i,1}, \ldots, \alpha_{z_i,T})$;
5. $(h_{N_*+i,1}, \ldots, h_{N_*+i,T} \mid \beta_{z_i}, \alpha_{z_i,1}, \ldots, \alpha_{z_i,T}, h_{N_*+i,0}, \sigma_{h,i}^2)$;

6. $(h_{N_*+i,0} \mid \sigma^2_{h,i}, h_{N_*+i,1}, \ldots, h_{N_*+i,T})$;

7. $(\sigma^2_{h,i} \mid h_{N_*+i,0}, h_{N_*+i,1}, \ldots, h_{N_*+i,T})$.

With (independent) Normal priors on β_{z_i}, $\alpha_{z_i,0}$, $h_{N_*+i,0}$, and with (independent) Inverse-Gamma priors on $\Sigma_{\alpha,i}$, $\sigma^2_{h,i}$, each step leads to a conditionally conjugate distribution. Specifically, the distributions in Steps 1, 2, 3, 5, and 6 are Normal while the distributions in Steps 4 and 7 are Inverse-Gamma.

For the general case, where it is not the case that $w_i = 1$ for $i = 1, .., M$, the standard Gibbs sampler outlined above needs only minor modifications. In particular, the conditional distributions for $\alpha_{z_i,0}$, $\Sigma_{\alpha,i}$, $h_{N_*+i,0}$ and $\sigma^2_{h,i}$ (conditional on draws of $\alpha_{z_i,1}, \ldots, \alpha_{z_i,T}$ and $h_{N_*+i,1}, \ldots, h_{N_*+i,T}$) are identical to those in Steps 3, 4, 6, and 7. The conditional distributions for β_{z_i} and $\alpha_{z_i,1}, \ldots, \alpha_{z_i,T}$ are also very similar to those in Steps 1 and 2—the only modification needed is to replace $h_{N_*+i,t}$ by $\bar{h}_{N_*+i,t} = h_{N_*+i,t} - \ln w_i$ for all $t = 1, \ldots, T$ and $i = 1, \ldots, M$ in the conditional densities. Finally, we adjust Step 5 to sample $h_{N_*+i,1}, \ldots, h_{N_*+i,T}$ from its conditional distribution

$$\begin{aligned} p(h_{N_*+i,1}, \ldots, h_{N_*+i,T} \,|\, h_{N_*+i,0}, \sigma^2_{h,N_*+1}, \alpha_{z_i,0}, \alpha_{z_i,1}, \ldots, \alpha_{z_i,T}, \beta_{z_i}, z_i, y^*) \\ \propto p(h_{N_*+i,1}, \ldots, h_{N_*+i,T} \,|\, h_{N_*+i,0}, \sigma^2_{h,N_*+1}) \\ p(z_i \,|\, y^*, h_{N_*+i,1}, \ldots, h_{N_*+i,T}, \alpha_{z_i,1}, \ldots, \alpha_{z_i,T}, \beta_{z_i})^{w_i}. \quad (4) \end{aligned}$$

This is done by considering the following auxiliary state space model

$$\begin{aligned} z_{i,t} &= y_t^{*\prime}\alpha_{z_i,t} + x_t'\beta_{z_i} + \epsilon_{z_i,t}, & \epsilon_{z_i,t} &\sim N\left(0, e^{\tilde{h}_{N_*+i,t}}\right), \\ \tilde{h}_{N_*+i,t} &= \frac{T-t+1}{2}(1-w_i)\sigma^2_{h,N_*+i} + \tilde{h}_{N_*+i,t-1} + \epsilon^h_{N_*+i,t}, & \epsilon^h_{N_*+i,t} &\sim N\left(0, \sigma^2_{h,N_*+i}\right). \end{aligned}$$

Clearly, we can sample $\tilde{h}_{N_*+i,1}, \ldots, \tilde{h}_{N_*+i,T}$ using standard methods for stochastic volatility models. Given these draws, we set $h_{N_*+i,t} = \tilde{h}_{N_*+i,t} + \ln w_i$. It can be shown that the draws thus obtained follow the same conditional distribution given in (**??**).

A.3 Forecasting

In this section we describe how one can compute the joint predictive density of the core variables using simulation. To start we first introduce some notation. For a time series

$x_1, \ldots, x_T$, we use $x_{s:t}$ to denote the observations from time s to time t, i.e., $x_{s:t} = \{x_s, x_{s+1}, \ldots, x_{t-1}, x_t\}$. For example, $\theta_{1:t}$ represents the set of common parameters from time 1 to time t, i.e., $\theta_{1:t} = \{\beta_y, A_{y,1}, \ldots, A_{y,t}, \Sigma_{y,1}, \ldots, \Sigma_{y,t}\}$. Furthermore, let $z_{t-p:t-1}$ denote the set of non-core variables: $\{z_{1,t-p:t-1}, \ldots, z_{M,t-p:t-1}\}$.

The one-step-ahead composite predictive density, conditional on the parameters up to time t, is given by:

$$p^C(y_t^*, z_{1,t}, \ldots, z_{M,t} \,|\, y_{t-p:t-1}, z_{t-p:t-1}, \theta_{1:t}, \tilde{\eta}_{1,1:t}, \ldots, \tilde{\eta}_{M,1:t}) = \\ p^C(y_t^* \,|\, y_{t-p:t-1}^*, z_{t-p:t-1}, \theta_{1:t}) \prod_{i=1}^{M} p^C(z_{i,t} \,|\, y_{t-p:t}^*, z_{i,t-p:t-1}, \tilde{\eta}_{i,1:t}),$$

where

$$p^C(y_t^* \,|\, y_{t-p:t-1}^*, z_{t-p:t-1}, \theta_{1:t}) \propto \prod_{i=1}^{M} p(y_t^* \,|\, y_{t-p:t-1}^*, z_{i,t-p:t-1}, \theta_{1:t})^{w_i},$$
$$p^C(z_{i,t} \,|\, y_{t-p:t}^*, z_{i,t-p:t-1}, \tilde{\eta}_{i,1:t}) \propto p(z_{i,t} \,|\, y_{t-p:t}^*, z_{i,t-p:t-1}, \tilde{\eta}_{i,1:t})^{w_i}.$$

The density $p^C(y_t^* \,|\, y_{t-p:t-1}^*, z_{t-p:t-1}, \theta_{1:t})$ is multivariate normal and has the form

$$(y_t^* \,|\, y_{t-p:t-1}^*, z_{t-p:t-1}, \theta_{1:t}) \sim N(\hat{y}_t, V_{y,t}),$$
$$\hat{y}_t = W_{y,t}\alpha_{y,t} + X_{y,t}\beta_y + V_{y,t}\left(\sum_{i=1}^{M} w_i V_{y,i,t}^{-1} X_{z_i} \underline{\beta}_{yz}\right),$$
$$V_{y,t} = \left(\sum_{i=1}^{M} w_i V_{y,i,t}^{-1}\right)^{-1},$$
$$V_{y,i,t} = X_{z_i,t} \underline{V}_{\beta,z} X_{z_i,t}' + \Sigma_{y,t}.$$

The density $p^C(z_{i,t} \,|\, y_{t-p:t}^*, z_{i,t-p:t-1}, \tilde{\eta}_{i,1:t})$ is also normal and has the form

$$(z_{i,t} \,|\, y_{t-p:t}^*, z_{i,t-p:t-1}, \tilde{\eta}_{i,1:t}) \sim N\left(y_t^{*\prime}\alpha_{z_i,t} + X_t\beta_{z_i}, e^{h_{N_*+i,t} - \ln w_i}\right).$$

Accordingly, the one-step ahead predictive density is given by

$$p^C(y^*_{t+1} \,|\, y^*_{1:t}, z_{1,1:t}, \ldots, z_{M,1:t}) = \int_{\theta_{1:t+1}} p^C(y^*_{t+1} \,|\, y^*_{t-p+1:t}, z_{t-p+1:t}, \theta_{1:t+1}) \\ p^C(\theta_{1:t} \,|\, y^*_{1:t}, \tilde{z}_{1,1:t}, \ldots, \tilde{z}_{M,1:t}) p(\theta_{t+1}|\theta_t) d\theta_{1:t+1},$$

where $p(\theta_{t+1}|\theta_t)$ is a product of normal densities implied by the state equations (9)–(10) in the paper. Hence, we can obtain the one-step ahead predictive density as follows: given a posterior draw of $\theta_{1:t}$, we use the state equations (9)–(10) in the paper to obtain θ_{t+1}. Conditional on these draws, $p^C(y^*_{t+1} \,|\, y^*_{t-p+1:t}, z_{t-p+1:t}, \theta_{1:t+1})$ is a normal density given above. Finally, we average these densities over the posterior simulator output.

This predictive simulation method can be applied to generate forecasts for longer horizons. Specifically, the same procedure can be applied, once we generate future core and auxiliary variables using the model. Furthermore, observe that sampling from the one-step-ahead predictive density $p^C(y^*_{t+1} \,|\, y^*_{1:t}, z_{1,1:t}, \ldots, z_{M,1:t})$ does not require draws of $\tilde{\eta}_{i,1:t}$, and therefore, the extra steps involved in sampling $\tilde{\eta}_{i,t}$ can be omitted if the researcher is interested only in one-step ahead forecasting or uses the direct method of forecasting. The empirical results in Section 4 use the direct method of forecasting.

A.4 Priors and Specification Choices

For the unrestricted VAR-SV models, we assume normal priors for the initial condition $a_0 \sim N(0, V_a)$ and $h_0 \sim N(0, V_h)$. Moreover, we assume an independent prior for parameters in Σ_h and Σ_a which are distributed as

$$\sigma^2_{h,i} \sim IG(\nu_{h,i}, S_{h,i}), \quad \sigma^2_{a,j} \sim IG(\nu_{a,j}, S_{a,2}),$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, \frac{N(N-1)}{2}$. We set $\nu_{h,i} = 10$, $S_{h,i} = 0.1^2(\nu_{h,i} - 1)$, $\nu_{a,j} = 10$ and $S_{h,j} = 0.01^2(\nu_{h,j} - 1)$. For the initial states, we set $V_h = 10 \times I_N$ and $V_a = 10 \times I_{\frac{N(N-1)}{2}}$.

For the VAR coefficients $\beta = \text{vec}\,((c, A_1, , \ldots, A_p)')$, we use a Minnesota prior and assume $\beta \sim N(\beta_0, V_\beta)$. For the prior mean, we set $\beta_0 = 0$. The prior covariance

matrix V_β is set to be diagonal and its corresponding values are set as follows:

$$\begin{aligned}
&\text{Var}(c) = 10 \times I_N, \\
&\text{Var}(A_l^{ij}) = \begin{cases} \frac{\lambda_1^2 \lambda_2}{l^{\lambda_3}} \frac{\sigma_i}{\sigma_j} & \text{for } l = 1, \ldots, p \text{ and } i \neq j, \\ \frac{\lambda_1^2}{l^{\lambda_3}} & \text{for } l = 1, \ldots, p \text{ and } i = j. \end{cases}
\end{aligned}$$

where A_l^{ij} denotes the (i, j) th element of the matrix A_l and σ_r is set equal to the standard deviation of the residual from an AR(p) model for the variable r. For the hyperparameters, we set $\lambda_1 = 0.2$, $\lambda_2 = 0.5$, $\lambda_3 = 2$, $p = 4$.

The VAR-CCM2 is the same as the VAR-SV except that the a_t is restricted to be time-invariant, i.e. $a_t = a$. We assume a normal prior $a \sim N(0, \Omega_a)$ with $\Omega_a = 10 \times I_{\frac{N(N-1)}{2}}$. The priors for other parameters are set the same as those in the VAR-SV.

For the Large Homoskedastic VAR

$$y + X\beta + \epsilon, \quad \epsilon_t \sim N(0, I_N \otimes \Sigma),$$

we assume an independent prior for the model parameters. The prior for the VAR coefficients is set equal to that in the VAR-SV. For the covariance matrix, we set $\Sigma \sim IW(\Sigma_0, \nu_0)$ with $\nu_0 = N + 2$ and $\Sigma_0 = (\nu_0 - N - 1)I_N$, where $IW(\cdot, \cdot)$ denotes the inverse Wishart distribution. This implies that the prior mean $E(\Sigma) = I_N$. We also include a natural conjugate prior version of the homoskedastic VAR for use with the large data set. For this we choose the same prior with the exception that the prior covariance matrix for β is the same as for VAR-CCM1 (see below).
For the VAR-CCM1, we first let $x_t' = (1, y_{t-1}', \ldots, y_{t-p}')$. It is convenient to specify the model as

$$Y = XA + U, \quad \text{vec}(U) \sim N(0, \Sigma \otimes \Omega)$$

where $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, $A = (c, A_1, \ldots, A_p)'$ and $\Omega = \text{diag}(e^{h_1}, \ldots, e^{h_T})$. Recall that the log volatility follow an AR(1) process

$$h_t = \rho h_{t-1} + \epsilon_t^h, \quad \rho \sim N(0, \sigma_h^2),$$

with $|\rho| < 1$. A standard normal-inverse-Wishart prior are set for model parameters (A, Σ) as

$$\Sigma \sim IW(\Sigma_0, \nu_0), \quad \text{vec}(A)|\Sigma \sim N(\text{vec}(A_0), \Sigma \otimes V_{\mathbf{A}}).$$

The hyperparameters Σ_0 and ν_0 are set equal to those in Large Homoskedastic VAR. We set $A_0 = 0$ for the prior mean of the VAR coefficients. For the covariance matrix, we assume it to be $V_A = \text{diag}(v_1, \ldots, v_k)$ and set $v_i = \frac{\lambda_1^2 \sigma_r}{l^{\lambda_3}}$ for coefficients associated to lag l of variable r for $i = 2, \ldots, k$ and $v_1 = 10$. The other hyperparameters are set equal to those in VAR-SV. For the AR coefficient and the variance of the log volatility process, we assume

$$\rho \sim N(\rho_0, V_\rho) \text{ for } |\rho| < 1, \quad \sigma_h^2 \sim IG(\nu_h, S_h)$$

with $\rho_0 = 0.9$, $V_\rho = 0.2^2$, $\nu_h = 10$ and $S_h = 0.1^2(\nu_h - 1)$.

For the VAR-FSV, we use the approach proposed by Kastner (2019) to model the time-varying error covariance of the VAR model. To be specific, the VAR-FSV can be written as:

$$\begin{aligned} y_t &= X_t\beta + \Lambda f_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_t), \\ f_t &\sim N(0, V_t), \end{aligned}$$

where $\Sigma_t = \text{diag}(e^{h_{1,t}}, \ldots, e^{h_{n,t}})$, Λ is a $n \times n_f$ factor loading, and f_t is a $n_f \times 1$ vector of factors with time-varying error covariance $V_t = \text{diag}(e^{h^f_{1,t}}, \ldots, e^{h^f_{n_f,t}})$. The log-volatilities are assumed to follow AR(1) processes:

$$\begin{aligned} h_{i,t} &= \mu_i + \phi_i(h_{i,t} - \mu_i) + \epsilon^h_{i,t}, \quad \epsilon^h_{i,t} \sim N(0, \sigma^2_{h,i}), \text{ for } i = 1, \ldots, n, \\ h^f_{j,t} &= \phi^f_j h^f_{j,t-1} + \epsilon^f_{j,t}, \quad \epsilon^f_{j,t} \sim N(0, \sigma^2_{f,j}), \text{ for } j = 1, \ldots, n_f. \end{aligned}$$

The priors for the VAR coefficients β and the variances of the log-volatility processes, $\sigma^2_{h,1}, \ldots, \sigma^2_{h,n}$, are set equal to those in the VAR-SV. For the other parameters in the

log-volatility processes, we assume

$$\begin{aligned}
\phi_i &\sim N(\phi_{i,0}, V_{\phi_i}) \text{ for } |\phi_i| < 1, \text{ and } i = 1, \ldots, n, \\
\mu_i &\sim N(\mu_{i,0}, V_{\mu_i}) \text{ for } i = 1, \ldots, n, \\
\phi_j^f &\sim N(\phi_{j,0}^f, V_{\phi_j^f}) \text{ for } |\phi_j^f| < 1, \text{ and } j = 1, \ldots, n_f, \\
\sigma_{f,j}^2 &\sim IG(\nu_{f,j}, S_{f,j}), \text{ for } j = 1, \ldots, n_f,
\end{aligned}$$

with $\phi_{i,0} = 0.9$, $V_{\phi_i} = 0.1^2$, $\mu_{i,0} = 0$, $V_{\mu_i} = 1$, $\phi_{j,0}^f = 0.9$, $V_{\phi_j^f} = 0.1^2$, $\nu_{f,j} = 5$, and $S_{f,j} = 0.1^2(\nu_{f,j} - 1)$, for $i = 1, \ldots, n$ and $j = 1, \ldots, n_f$. Following Kastner (2019), a Normal-Gamma prior is imposed on the factor loading matrix Λ. More specifically, we let

$$\Lambda_{i,j} \sim N(0, \tau_{i,j}^2), \quad \tau_{i,j}^2 \sim G(\psi, \frac{\psi\delta^2}{2}), \quad \delta^2 \sim G(c_1, c_2),$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n_f$. We set the hyperparameters $c_1 = c_2 = 0.001$ and impose an exponential distributed prior centered on unity for the hyperparameter ψ, i.e. $\psi \sim Exp(1)$. For the MCMC sampler, we refer reader to Kastner (2019) and Huber and Feldkircher (2019) for more details. In our forecasting exercise, we only consider the cases for $n_f = 1, 2$.

B Empirical Appendix

B.1 Data

We use the quaterly data set from the Federal Reserve Bank of St. Louis 'FRD-QD' data set of Mccracken and Ng (2016) covers period from 1960Q1 - 2015Q3. All data are transformed as in Mccracken and Ng (2016) to achieve stationarity.[1] The three core variables are listed in the paper itself. The"good" variables, "bad" variables and variables used in the 7-variable VAR-SVs are listed in Table A1 and Table A2, respectively. The 20 variables models, VAR-CCM1-20, VAR-CCM2-20, VAR-FSV-1f and VAR-FSV-2f, use the core variables, the "good" variables and the variables listed in Table A3.

[1]The one exception to this is that we do not transform the Effective Federal Funds Rate.

Table A1: Four "good" variables.

fred	description
UNRATE	Civilian Unemployment Rate
INPRO	Industral Production Index
M2REALx	Real Money Base
S&P 500	S&P's common Stoch Price Index: Composite

Table A2: Four "bad" variables.

fred	description
HOUSTNE	Housing Starts in Northeast Census Region
DIFSRG3Q086SBEA	Personal Consumption Expenditures: Financial Services and Insurance
NONBORRES	Reserves of Depository Institutions, Nonborrowed
CUUR0000SAD	Consumer Price Index for All Urban Consumers: Durables

Table A3: Description of other 13 variables.

fred	description
	Industral Production
CUMFNS	Capacity Utilization: Manufacturing
	Emplyment and Unemployment
CLAIMSx	Initial Claims
PAYEMS	All Employees: Total nonfarm
AWHMAN	Avg Weekly Hours: Manufacturing
	Housing
HOUST	Housing Starts: Total New Privately Owned
	Money and Credit
TOTRESNS	Total Business Inventories
	Inventories, Orders and Sales
AMDMNOx	New Orders for Durable Goods
NAPMNOI	ISM Manufacturing: New Orders Index
NAPMII	ISM Manufacturing: Inventories Index
	Interest Rate
TB3MS	3-Month Treasury Bill
GS10	10-Year Treasury Rate
	Price
OILPRICEx	Crude Oil, spliced WTI and Cushing
CPIULFSL	CPI: All Items Less Food

B.2 Monte Carlo Study

The DGP is obtained by first estimating the VAR-SV (equations (1)-(3) in the paper) using the small data set so as to obtain estimates (posterior means) of a_t, h_t and β. We then generate 100 artificial datasets (with same sample size as the actual data) from

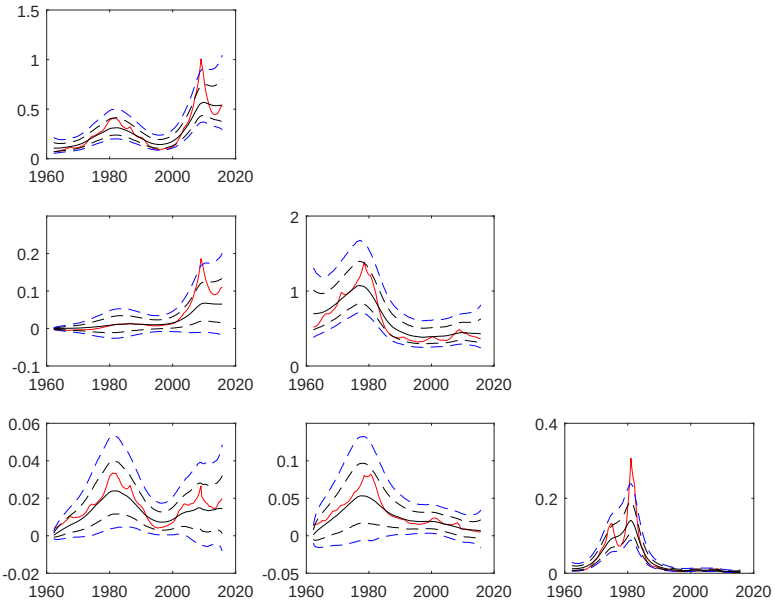

Figure 1: Monte Carlo Results for VAR-CL-ML. Solid red line: true parameter path. Black line: posterior median. Dotted lines: 16th/84th and 5th/95th percentiles.

the VAR-SV with parameters and states set to these estimates. For each dataset, we use various VAR-CL approaches to estimate σ_{ijt} for $i, j = 1, 2, 3$ where σ_{ijt} denotes the $(i, j)^{th}$ element of the error covariance matrix at time t. The results are in Figures 1 through 4. All lines in these figures (except the one for the true parameter path) represents an average over the 100 datasets. It can be seen that the average of the point estimates for all approaches tracks the true parameter path fairly well and that the coverage of the intervals is excellent for all the choices of weights used with the composite likelihood approaches. Even the use of equal weights leads to good coverage properties.

B.3 The Computational Advantages of Composite Likelihood Methods

We have argued in this paper that the main advantage of our composite likelihood approach is computational. It is computationally feasible in a Big Data context where other approaches which incorporate stochastic volatility are not. To reinforce this

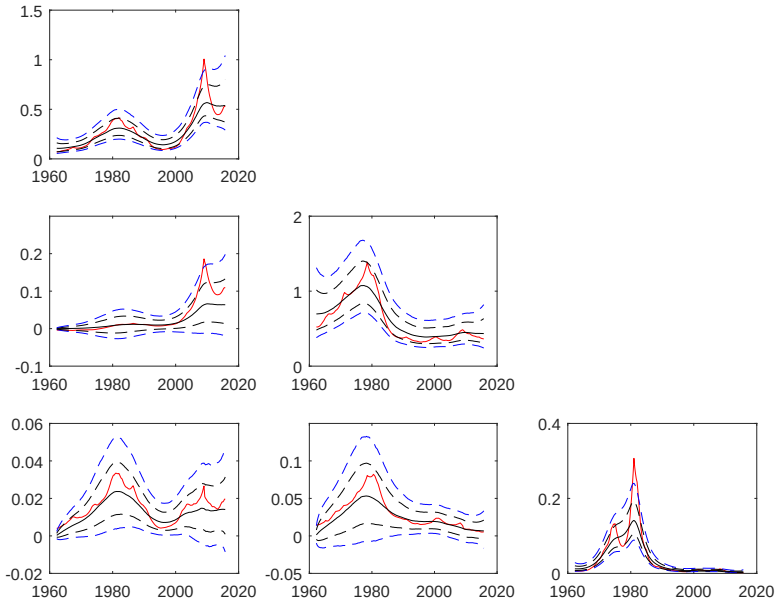

Figure 2: Monte Carlo Results for VAR-CL-BIC. Solid red line: true parameter path. Black line: posterior median. Dotted lines: 16th/84th and 5th/95th percentiles.

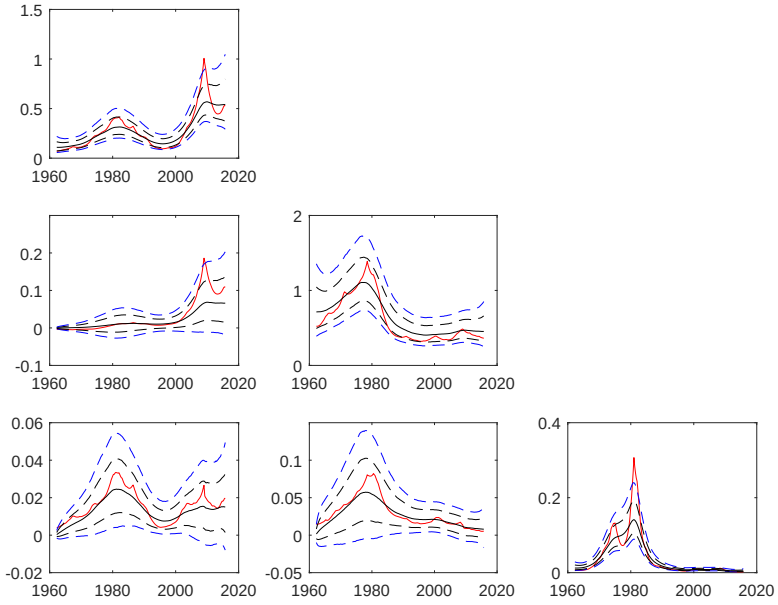

Figure 3: Monte Carlo Results for VAR-CL-DIC. Solid red line: true parameter path. Black line: posterior median. Dotted lines: 16th/84th and 5th/95th percentiles.

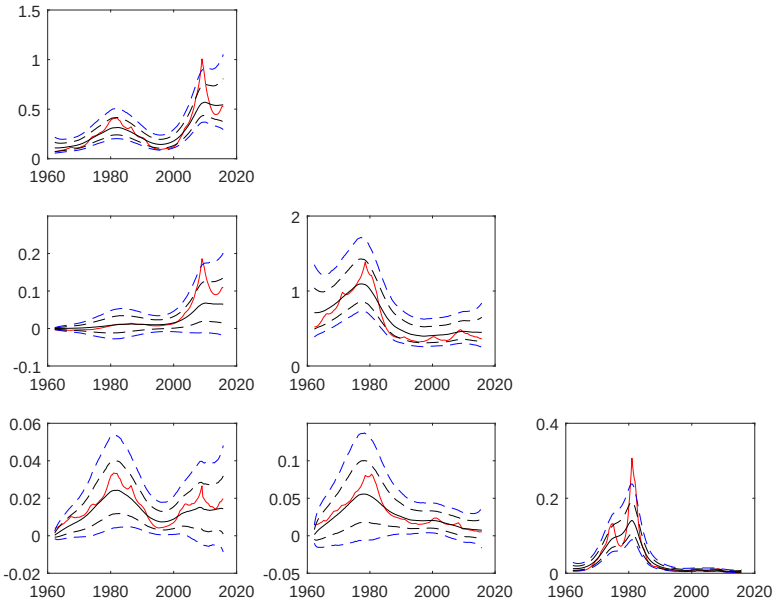

Figure 4: Monte Carlo Results for VAR-CL-EQ. Solid red line: true parameter path. Black line: posterior median. Dotted lines: 16th/84th and 5th/95th percentiles.

point, in this section we present some results showing the computational properties of the composite likelihood approaches relative to others.

Table B1 presents results relating to the composite likelihood approach when using the large data set and doing one run of our simulation algorithm using the full sample.[2] Note that our algorithm involves two steps: (i) estimating all the quasi-posteriors using MCMC (labelled "Estimation" in Table 1) and (ii) using an accept-reject algorithm to pool draws (labelled "Pooling" in Table 1). Note also that, within step (i), we run things in parallel across different quasi-posteriors and some variants require marginal likelihood or information criteria estimation. The results in the table are based on taking 22,000 draws from each quasi-posterior which, we have found, is the minimum necessary to obtain reasonable effective sample sizes for all our MCMC algorithms. The first 2,000 draws from each quasi-posterior are burn-in draws which are dropped. The remaining 20,000 are then thinned to 1,000 to reduce correlation between draws. This leaves us with 1,000×193=193,000 draws which are used in

[2]All computation was done on a Dell Precision Tower 7910 with 2 Intel Xeon 3.10Ghz processors (total of 20 cores) and 256GB of memory.

step (ii). The time to do step (ii) is calculated as the time taken to obtain 1,000 retained draws from these 193,000 draws. Note that these final 1,000 may contain some repeating draws and, if such repeats are too high, this will make the effective sample size of the algorithm low. To show that this is not a substantive problem with our composite likelihood approaches, the table also contains a column labelled "Unique" which is the percentage of draws which are unique and do not repeat.

It can be seen that, even with our very large data set, computation can easily be done by a good PC with running time being roughly an hour. The equally-weighted composite likelihood approach is faster due to the fact that it does not require the calculation of marginal likelihoods or an information criterion.

Table B1: Computational Time in Minutes of Composite Likelihood Approaches

	Estimation	Pooling	Total	Unique
VAR-CL-ML	62.4	5.5	67.9	46.7%
VAR-CL-DIC	60.6	0.4	61.0	65.0%
VAR-CL-BIC	60.6	2.7	63.3	53.2%
VAR-CL-EQ	34.3	11.2	45.5	99.9%

It is worth noting that the linear opinion pool (VAR-LIN) is much more computationally demanding since it involves recursive estimation and numerical optimization (see Geweke and Amisano, 2011). The computational time comparable to those reported in Table B1 is 62.4 hours.

Table B2 presents computational time for VARs of different dimensions for the alternative approaches which allow for stochastic volatility and for one of our composite likelihood approaches. For the composite likelihood approach, the time reported is to carry out the same exercise as was used to produce the numbers in Table 1. For the other approaches, it is the time to produce 22,000 MCMC draws.

Table B2 shows that the composite likelihood approach and VAR-CCM1 are the only approaches likely to be computationally feasible in truly large VARs. But, as we have seen, VAR-CCM1 is likely to be too restrictive in many empirical contexts. Computational times, of course, increase with VAR dimension. But with the composite likelihood approach this increase is approximately linear in N. With VAR-SV the increase is much more rapid (not quadratic, but close to it). Even for $N = 100$, computation time with the VAR-SV is more than a week on a good PC for a single

run of the algorithm. VAR-CCM2 is not this bad (for $N = 100$ its running time is a few hours), but running time is much more than for our composite likelihood approaches and it is increasing at a more than linear rate in N. The latter fact is likely to preclude its use in very large VARs.

Table B2: Computation Time in Minutes for Different VAR dimensions				
N	VAR-CL-EQ	VAR-SV	VAR-CCM1	VAR-CCM2
3	n.a.	0.74	0.27	0.44
7	2.67	6.34	0.40	1.34
20	4.01	112.67	0.82	5.42
50	10.95	1602.61	1.79	37.89
100	21.14	13071.09	7.72	160.24

B.4 Additional Empirical Results

In Tables 2 through 5 in the paper, we presented a selection of the results of our forecasting exercise organized around addressing three questions. In this section, we provide the complete set of forecasting results which underlies these tables. The DM statistics use the the Large VAR benchmark. The RMSFEs, MAEs and ACRPs are multiplied by 100 to allow for easy comparison. This section also provides a complete set of forecasting results for a shorter forecast evaluation period which begins in 2008Q1.

Table B3: Forecasting Evaluation Using Joint ALPL for 3 Core Variables				
Horizon	$h = 1$		$h = 4$	
Evaluation begins:	1970Q1	2008Q1	1970Q1	2008Q1
VAR-SV-3	5.650	6.780	4.133	4.431
VAR-CCM1	6.079***	6.050***	4.456***	3.909***
VAR-CCM2	5.805***	6.553***	4.089***	3.759***
Large VAR	1.040	2.089	−0.570	0.987
VAR-SV-g	5.602***	6.404***	3.998***	3.838***
VAR-SV-b	5.509***	6.371***	3.913***	3.463***
VAR-LIN	8.645***	9.760***	6.993***	7.269***
VAR-CL-ML	8.153***	9.010***	6.455***	7.148***
VAR-CL-DIC	6.281***	8.182***	5.076***	5.067***
VAR-CL-BIC	5.857***	9.227***	6.190***	7.228***
VAR-CL-LIN	8.428***	9.391***	6.416***	6.655***
VAR-CL-EQ	8.440***	9.573***	6.855***	7.075***
VAR-CCM1-20	5.328***	6.367***	3.500***	4.878***
VAR-CCM2-20	5.804***	6.019***	4.577***	4.185***
VAR-FSV-1f	3.106***	3.424***	2.486***	2.984***
VAR-FSV-2f	2.820***	3.205***	2.510***	3.274***
VAR-HM-CL-EQ	6.159***	7.106***	5.170***	6.060***

Table B4: Evaluation of Inflation Forecasts Beginning in 1970								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	0.570**	0.390**	0.390***	3.467***	0.760***	0.550***	0.480***	3.327***
VAR-CCM1	0.720	0.480	0.380***	3.711***	0.900***	0.614***	0.489***	3.441***
VAR-CCM2	0.570**	0.380**	0.420***	3.411***	0.810***	0.588***	0.529***	3.274***
large VAR	0.580	0.430	4.570	1.237	0.959	0.739	5.852	1.137
VAR-SV-g	0.560***	0.380***	0.410***	3.431***	0.811***	0.592***	0.525***	3.282***
VAR-SV-b	0.630**	0.420***	0.460***	3.416***	0.819***	0.585***	0.591***	3.252***
VAR-LIN	0.576*	0.394**	0.336***	4.362***	0.591***	0.427***	0.458***	4.215***
VAR-CL-ML	0.584	0.401	0.463***	4.178***	0.624***	0.454***	0.593***	3.955***
VAR-CL-DIC	0.745	0.493	1.743***	3.392***	0.747***	0.518***	1.561***	3.334***
VAR-CL-BIC	1.746	0.905	3.102***	3.351***	0.587***	0.439***	0.846***	3.822***
VAR-CL-LIN	0.581	0.400	0.418***	4.273***	0.614***	0.442***	0.790***	3.974***
VAR-CL-EQ	0.577*	0.395**	0.415***	4.270***	0.581***	0.421***	0.498***	4.152***
VAR-CCM1-20	0.710	0.490	0.700***	3.246***	1.000	0.710*	1.010***	2.963***
VAR-CCM2-20	0.530	0.380	0.410***	3.434***	0.760***	0.530***	0.450***	3.352***
VAR-FSV-1f	1.260	1.160	0.950***	2.819***	1.470	1.330	1.100***	2.717***
VAR-FSV-2f	1.400	1.260	1.010***	2.776***	1.440	1.300	1.080***	2.733***
VAR-HM-CL-EQ	0.633	0.473***	0.906***	3.377***	0.654***	0.470***	0.955***	3.336***

Table B5: Evaluation of Inflation Forecasts Beginning in 2008								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	0.740	0.450	0.430***	3.316***	0.820*	0.500**	0.480***	3.239***
VAR-CCM1	0.730	0.490	0.390***	3.482***	0.741**	0.434**	0.401***	3.258***
VAR-CCM2	0.730	0.450	0.460***	3.300***	0.845*	0.567**	0.527***	3.190***
large VAR	0.730	0.450	2.780	1.563	0.885	0.656	3.154	1.504
VAR-SV-g	0.730	0.450	0.460***	3.307***	0.856*	0.572**	0.528***	3.192***
VAR-SV-b	0.960	0.620	0.720***	3.238***	1.047	0.658	1.045***	2.946***
VAR-LIN	0.740	0.448	0.416***	4.231***	0.757**	0.460**	0.455***	4.138***
VAR-CL-ML	0.746	0.465	0.583***	3.984***	0.728***	0.460***	0.551***	3.993***
VAR-CL-DIC	0.864	0.510	1.231***	3.745***	0.632***	0.538**	2.000***	3.192***
VAR-CL-BIC	0.772	0.458	0.624***	3.977***	0.711***	0.451**	0.517***	4.054***
VAR-CL-LIN	0.766	0.490	0.491***	4.091***	0.727***	0.447**	0.594***	3.896***
VAR-CL-EQ	0.742	0.456	0.462***	4.143***	0.727**	0.459**	0.502***	4.075***
VAR-CCM1-20	0.690	0.500	0.450***	3.545***	0.850*	0.590	0.540***	3.272***
VAR-CCM2-20	0.760	0.500	0.500***	3.264***	0.880*	0.570*	0.530***	3.210***
VAR-FSV-1f	1.290	1.110	0.930***	2.870***	1.360	1.110	1.010***	2.843***
VAR-FSV-2f	1.490	1.240	1.000***	2.820***	1.240	1.030	0.980***	2.864***
VAR-HM-CL-EQ	0.706	0.457	0.675***	3.630***	0.717***	0.405***	0.720***	3.595***

Table B6: Evaluation of Interest Rate Forecasts Beginning in 1970								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	1.002***	0.602***	0.473***	−1.062***	2.368	1.774	1.331***	−2.368
VAR-CCM1	1.000***	0.577***	0.465***	−1.100***	2.347	1.776	1.361***	−2.350
VAR-CCM2	0.946***	0.559***	0.435***	−0.876***	2.240**	1.674**	1.266***	−2.260**
Large VAR	1.091	0.737	0.655	−1.429	3.020	2.087	1.870	−2.821
VAR-SV-g	0.971***	0.583***	0.462***	−1.079***	2.314	1.723	1.302***	−2.355*
VAR-SV-b	1.006***	0.611***	0.483***	−1.087***	2.364	1.773	1.338***	−2.375
VAR-LIN	0.971***	0.567***	0.427***	0.117***	1.337***	0.952***	1.294***	−1.315***
VAR-CL-ML	0.998***	0.603***	0.458***	0.034***	1.314***	0.894***	1.248***	−1.309***
VAR-CL-DIC	0.983***	0.578***	0.634***	−0.184***	1.447***	1.011***	1.377***	−1.394***
VAR-CL-BIC	1.330*	0.877*	1.569**	−0.623***	1.321***	0.917***	1.321***	−1.315***
VAR-CL-LIN	0.980***	0.572***	0.444***	0.074***	1.382***	0.939***	1.558***	−1.396***
VAR-CL-EQ	0.973***	0.577***	0.440***	0.098***	1.340***	0.950***	1.283***	−1.323***
VAR-CCM1-20	0.967***	0.581***	0.495***	−1.057***	2.509***	1.685**	1.442***	−2.384***
VAR-CCM2-20	0.965***	0.551***	0.484***	−0.888***	2.733	2.061	1.161***	−1.950***
VAR-FSV-1f	4.942	4.743	3.028	−3.056	5.070	4.605	3.300	−3.425
VAR-FSV-2f	5.965	5.586	3.615	−3.226	5.321	4.752	3.358	−3.328
VAR-HM-CL-EQ	1.021	0.651	0.534***	−0.546***	1.513***	1.116***	1.369***	−1.477***

Table B7: Evaluation of Interest Rate Forecasts Beginning in 2008								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	0.379	0.209	0.170	0.215	1.688	1.480	0.988	−1.944
VAR-CCM1	0.470	0.389	0.281***	−0.714***	1.696	1.566	1.110	−2.222
VAR-CCM2	0.417***	0.288***	0.216***	−0.138***	1.645	1.455	1.029	−2.123**
Large VAR	0.660	0.492	0.456	−1.040	1.618	1.114	1.113	−1.917
VAR-SV-g	0.419***	0.300***	0.222***	−0.156***	1.626	1.430	0.985	−2.032
VAR-SV-b	0.415***	0.271***	0.227***	0.060***	1.659	1.456	1.012	−1.974
VAR-LIN	0.370***	0.198***	0.160***	1.412***	1.130	1.079	0.926**	−0.910***
VAR-CL-ML	0.481***	0.245***	0.196***	1.172***	0.711	0.618*	0.683***	−0.739***
VAR-CL-DIC	0.364***	0.192***	0.249***	0.877***	1.390	1.336	1.156	−1.195***
VAR-CL-BIC	0.532***	0.256***	0.194***	1.396***	1.067	0.908	0.867**	−0.756***
VAR-CL-LIN	0.365***	0.199***	0.155***	1.310***	1.151	1.093	0.934	−1.027***
VAR-CL-EQ	0.364***	0.189***	0.162***	1.383***	1.212	1.151	0.966	−0.974***
VAR-CCM1-20	0.582***	0.303***	0.305***	−0.622***	1.440	0.784	0.836***	−1.692***
VAR-CCM2-20	0.883**	0.439**	0.416***	−0.560***	2.648	1.871	1.349	−2.121
VAR-FSV-1f	3.770	3.752	2.305	−2.775	3.705	3.565	2.408	−2.956
VAR-FSV-2f	3.746	3.669	2.288	−2.800	2.418	2.231	1.605	−2.489
VAR-HM-CL-EQ	1.013	0.717	0.779***	3.560***	1.069	0.761	0.772***	3.550***

Table B8: Evaluation of GDP Growth Forecasts Beginning in 1970								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	0.800**	0.600**	0.520***	3.220***	0.850***	0.610***	0.560***	3.162***
VAR-CCM1	0.790	0.610	0.480***	3.405***	0.924**	0.664***	0.551***	3.242***
VAR-CCM2	0.750*	0.580**	0.550***	3.213***	0.870***	0.622***	0.608***	3.122***
large VAR	0.850	0.660	4.590	1.234	1.037	0.806	5.868	1.135
VAR-SV-g	0.750*	0.580**	0.540***	3.220***	0.870***	0.626	0.607***	3.120***
VAR-SV-b	0.990	0.680	0.640***	3.152***	0.921***	0.652***	0.687***	3.099***
VAR-LIN	0.791**	0.596***	0.462***	4.161***	0.820**	0.589**	0.512***	4.093***
VAR-CL-ML	0.837**	0.607**	0.619***	3.938***	0.810***	0.589***	0.688***	3.806***
VAR-CL-DIC	1.054	0.807	2.621***	3.070***	0.951***	0.659**	1.961***	3.131***
VAR-CL-BIC	2.549	1.024	3.721***	3.126***	0.830***	0.618***	0.992***	3.683***
VAR-CL-LIN	0.821**	0.607**	0.548***	4.077***	0.885**	0.606**	0.884***	3.837***
VAR-CL-EQ	0.805*	0.599*	0.540***	4.070***	0.830***	0.599**	0.573***	4.023***
VAR-CCM1-20	0.830	0.620	0.780***	3.113***	0.950*	0.710	1.060***	2.847***
VAR-CCM2-20	0.700***	0.530***	0.530***	3.230***	0.870***	0.650***	0.580***	3.148***
VAR-FSV-1f	1.180	0.920	0.940***	2.859***	1.240	1.000	1.010***	2.818***
VAR-FSV-2f	1.180	0.970	0.960***	2.829***	1.240	1.030	1.010***	2.808***
VAR-HM-CL-EQ	0.856	0.646	0.982***	3.322***	0.832***	0.618***	1.009***	3.302***

Table B9: Evaluation of GDP Growth Forecasts Beginning in 2008								
	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
VAR-SV-3	0.880	0.620	0.520^{***}	3.206^{***}	0.890	0.590	0.530^{***}	3.152^{***}
VAR-CCM1	0.900	0.670	0.530^{***}	3.144^{***}	1.019	0.716	0.597^{***}	2.873^{***}
VAR-CCM2	0.790	0.580	0.530^{***}	3.223^{***}	0.910	0.631	0.587^{***}	3.097^{***}
large VAR	0.790	0.520	2.800	1.560	0.966	0.687	3.176	1.500
VAR-SV-g	0.790	0.580	0.520^{***}	3.235^{***}	0.904	0.627	0.579^{***}	3.106^{***}
VAR-SV-b	1.630	1.020	1.020^{***}	3.006^{***}	1.205	0.788	1.182^{***}	2.893^{***}
VAR-LIN	0.902	0.614	0.525^{***}	4.115^{***}	0.978	0.661	0.503^{***}	4.054^{***}
VAR-CL-ML	0.929	0.609	0.709^{***}	3.849^{***}	0.971^{*}	0.641^{***}	0.632^{***}	3.893^{***}
VAR-CL-DIC	0.684	0.526	1.523^{***}	3.554^{***}	1.325	0.782	2.314^{***}	3.065^{***}
VAR-CL-BIC	0.961	0.662	0.772^{***}	3.848^{***}	0.994	0.671	0.584^{***}	3.947^{***}
VAR-CL-LIN	0.963	0.635	0.584^{***}	3.988^{***}	0.972	0.674	0.673^{***}	3.795^{***}
VAR-CL-EQ	0.880	0.618	0.540^{***}	4.053^{***}	1.008	0.689	0.571^{***}	3.982^{***}
VAR-CCM1-20	0.720	0.530	0.510^{***}	3.394^{***}	1.000	0.740	0.670^{***}	3.030^{***}
VAR-CCM2-20	0.600	0.450	0.500^{***}	3.300^{***}	0.860	0.600	0.620^{***}	3.109^{***}
VAR-FSV-1f	0.850	0.690	0.810^{***}	3.018^{***}	1.160	1.010	1.010^{***}	2.856^{***}
VAR-FSV-2f	1.060	0.940	0.910^{***}	2.900^{***}	1.270	1.120	1.050^{***}	2.810^{***}
VAR-HM-CL-EQ	0.508^{*}	0.365^{**}	0.324^{***}	-0.104^{***}	1.382	1.344	1.026	1.115^{***}

References

Kastner, G. (2019). Sparse Bayesian time-varying covariance estimation in many dimensions, Journal of Econometrics, 210, 98-115.
Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. Journal of Business & Economic Statistics 37, 27-39.