

***"This is the peer reviewed version of the following article: [Journal of Applied Econometrics, 2020 ] which has been published in final form at [<https://onlinelibrary.wiley.com/doi/10.1002/jae.2793>] purposes in accordance with [Wiley Terms and Conditions for Self-Archiving.](#)"***

# Composite Likelihood Methods for Large Bayesian VARs with Stochastic Volatility\*

Joshua C.C. Chan

Purdue University and UTS

Chenghan Hou

Hunan University

Eric Eisenstat

The University of Queensland

Gary Koop

University of Strathclyde

May 2020

**Abstract:** Adding multivariate stochastic volatility of a flexible form to large Vector Autoregressions (VARs) involving over a hundred variables has proved challenging due to computational considerations and over-parameterization concerns. The existing literature either works with homoskedastic models or smaller models with restrictive forms for the stochastic volatility. In this paper, we develop composite likelihood methods for large VARs with multivariate stochastic volatility. These involve estimating large numbers of parsimonious models and then taking a weighted average across these models. We discuss various schemes for choosing the weights. In our empirical work involving VARs of up to 196 variables, we show that composite likelihood methods forecast much better than the most popular large VAR approach which is computationally practical in very high dimensions: the homoskedastic VAR with Minnesota prior. We also compare our methods to various popular approaches which allow for stochastic volatility using medium and small VARs involving up to 20 variables. We find our methods to forecast appreciably better than these as well.

**Keywords:** Bayesian, large VAR, composite likelihood, prediction pools, stochastic volatility

**JEL Classifications:** C11, C32, C53

---

\*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.

# 1 Introduction

Empirical macroeconomists increasingly use Vector Autoregressions (VARs) with datasets involving a hundred or more variables. However, in many applications, there is strong evidence for stochastic volatility. Adding multivariate stochastic volatility to large VARs has proved challenging. With small data sets, popular specifications for VARs with stochastic volatility (VAR-SV) exist but Bayesian estimation and forecasting with these VAR-SVs is not computationally feasible with large data sets. This paper develops a computationally feasible approach which uses all the data available and allows for stochastic volatility. The idea of our approach is to work with many small VAR-SVs. Each of these contains only a few of the large number of variables available, but every one of the variables appears in one or more of the small models. Forecasts from the many small models are then combined to produce forecasts which reflect all the data available. We use composite likelihood methods to theoretically justify and implement such a strategy.

The fact that large VARs are being found increasingly useful in an era of Big Data needs little justification. The large VAR literature began with the US macroeconomic application of Bańbura, Giannone and Reichlin (2010) but large VARs are now used with a variety of macroeconomic and financial data (see, among others, Bloor and Matheson, 2010, Carriero, Kapetanios and Marcellino, 2010, 2012, Giannone, Lenza, Momferatou and Onorante, 2014, Bańbura, Giannone and Lenza, 2015, Koop and Korobilis, 2016, McCracken, Owyang and Sekhposyan, 2016, Jarociński and Maćkowiak, 2017, and Kastner and Huber, 2017). Similarly, the facts that macroeconomic variables often exhibit structural instabilities and have variances that change over time is increasingly accepted. Papers such as Clark (2011) highlight the particular importance in macroeconomic applications of allowing for time-variation in the error covariance matrix. Hence, this is what we focus on in this paper (although the econometric methods we develop could also be used with the time-varying parameter VAR).

These arguments justify why there is a desire to work with large VAR-SVs. But Bayesian methods, requiring the use of Markov Chain Monte Carlo (MCMC) methods, quickly become computationally infeasible as the number of variables in the VAR increases. Bayesian methods are typically used with large VARs since they allow for prior shrinkage which is of great use with over-parameterized models. For instance,

when working with a large VAR with  $N = 100$  variables and a lag length of  $p = 13$  (as might be required with monthly data), the researcher will have over 100,000 VAR coefficients to estimate and 5,050 free parameters in the error covariance matrix. Bayesian prior shrinkage, often using natural conjugate or Minnesota priors, is used to surmount the problems caused by a shortage of data information relative to the number of coefficients being estimated. Even with these priors, which imply that the posterior and one-step ahead predictive densities have analytical forms, the researcher can face a substantial computational burden (see, e.g., Chan, 2020 or Carriero, Clark and Marcellino, 2016). However, even small departures from the natural conjugate prior VAR leads to posteriors that require the use of MCMC methods which means a huge increase in the computational burden. And it is worth emphasizing that departing from the conjugate prior is often desirable with VARs since its assumptions are very restrictive. These are i) symmetry of the prior across equations, ii) proportionality of the prior covariance to the error variance, and iii) homoskedasticity.

With large VARs this makes many sensible alternative approaches untenable. This holds true for various global-local shrinkage priors (e.g. the variable selection prior of George, Sun and Ni, 2008, Koop, 2013 and Korobilis, 2013, or the Lasso prior of Gefang, 2014). It also holds true for specifications which allow for time-variation in the error covariance matrix. It is the latter which is the focus of the present paper.

As noted, with large VARs standard approaches (e.g. Primiceri, 2005) which allow for multivariate stochastic volatility are not computationally feasible. But there are stochastic volatility specifications that can be used with larger VARs (e.g. Chan, 2020, and Carriero, Clark and Marcellino, 2016, 2019a,b). However, these place restrictions on the form of time variation that is allowed for. And even these have a large computational burden which means they cannot be used for forecasting with the large VARs involving hundreds of dependent variables which are increasingly being used.<sup>1</sup>

These considerations motivate the present paper. Working with many small VAR-SVs is computationally feasible even with very high dimensional data sets and composite likelihood methods can be used to combine forecasts from these many smaller

---

<sup>1</sup>Perhaps the best of the current approaches is developed in Carriero, Clark and Marcellino (2019a). In this paper, impulse responses are presented using a 125 variable VAR, but when forecasting only a 20 variable VAR is used. Repeatedly forecasting with this model on an expanding window of data with the 196 variables used in this paper would take months or more of computing time on a good PC.

models. So the methods we develop are practical and can be used with hundreds of variables. But we also address several other questions to further strengthen the case for our composite likelihood based methods. The first of these is whether there is a theoretically strong justification for use of composite likelihood methods in our context. We discuss relevant econometric theory in the next section of the paper. The second question is: How should the various small models that arise with composite likelihood methods be combined? To answer this question we discuss various methods for doing so, drawing on the literature on opinion pools. The third question is: How well do these methods work in practice? We answer this using a large quarterly US macroeconomic data set involving 196 variables. We find our composite likelihood methods to forecast substantially better than the only computationally practical competitor: a homoskedastic VAR using a natural conjugate prior. We would like to compare our methods to other approaches that involve multivariate stochastic volatility using this large data set, but cannot do it since the computational burden of popular Bayesian alternatives is too large. Instead, we compare our methods to a range of different Bayesian VARs with multivariate stochastic volatility using small data sets involving 7 variables and medium data sets involving 20 variables. We demonstrate the good, computationally efficient, forecasting performance of our composite likelihood methods relative to these alternatives.

## 2 Composite Likelihood Methods for large VARs with Stochastic Volatility

### 2.1 Overview

A traditional likelihood function is based on the p.d.f. of the  $N \times 1$  vector of dependent variables,  $y_t$  for  $t = 1, \dots, T$ . In many empirical cases, particularly if  $N$  is large, computation involving a likelihood function can be difficult or infeasible. In such cases, it may be possible to develop statistical methods for estimation of the parameters or forecasting using the composite likelihood instead of the full likelihood. The composite likelihood is built up as a weighted average of likelihoods for  $y_{i,t}$  for  $i = 1, \dots, M$  which are sub-vectors of  $y_t$ . The likelihoods for these sub-vectors are often called quasi-likelihoods and we will use this terminology. Bayesian methods can then be used by combining a prior with the composite likelihood in the standard way.

Thus, if  $y_{i,t}$  is of much lower dimension than  $y_t$ , a computationally difficult problem of working with a high dimensional likelihood can be turned into a much simpler one of working with many small quasi-likelihoods.

Composite likelihood methods can also have advantages in terms of parsimony. That is, high dimensional models like large VARs are hugely over-parameterized. The correct specification is likely a highly restricted version of the large VAR. The existing Bayesian large VAR literature tries to overcome this problem through the use of prior shrinkage or parameter restrictions. But this prior information often has to be very strong to obtain reasonable forecasts. Using composite likelihood methods, we are only working with small VARs which can yield good forecasts even in the absence of strong prior information. Thus, prior elicitation becomes a less important issue. We do have to make a choice of quasi-likelihoods, but this may be easier to make (and justify) than choosing a particular prior or parametric restriction. For instance, in this paper the desire to forecast a core set of variables of interest in the context of a large data set which includes many other variables motivates our choice of quasi-likelihoods.

The statistical literature on composite likelihood methods (see, e.g., Varin, Reid and Firth, 2011, Ribatet, Cooley and Davison, 2012, and Roche, 2016) provides a theoretical and empirical justification for working with them. They have been exploited in several fields. For instance, Pakel, Shephard, Sheppard and Engle (2014) is a financial application involving a large number of stock returns. These methods have also been used in spatial statistics (e.g. Ribatet, Cooley and Davison, 2012). But they have been rarely used in macroeconomics.<sup>2</sup> To our knowledge, our paper is the first to use them in the large VAR field.

## 2.2 The VAR-SV

We begin by defining the VAR-SVs that our quasi-likelihoods are based on. Specifications identical or similar to this have been used in a huge range of papers, including Primiceri (2005), Koop, Leon-Gonzalez and Strachan (2009), Clark (2011), D'Agostino, Gambetti and Giannone (2013) and Chan and Eisenstat (2018). The

---

<sup>2</sup>Two exceptions to this lie in the field of Dynamic Stochastic General Equilibrium (DSGE) modelling: Canova and Matthes (2018) and Qu (2016).

VAR-SV model can be written as:

$$A_{0t}y_t = c + A_1y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t,$$

where  $c$  is an  $N \times 1$  vector of intercepts,  $A_1, \dots, A_p$  are  $N \times N$  matrices of VAR coefficients,  $\Sigma_t = \text{diag}(e^{h_{1,t}}, \dots, e^{h_{n,t}})$  and  $A_{0t}$  is a time varying  $N \times N$  lower triangular matrix with ones on the diagonal, to be specific,

$$A_{0t} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1,t} & a_{n2,t} & \cdots & 1 \end{pmatrix}.$$

We use notation where  $c_i$  is intercept in the  $i^{\text{th}}$  equation and  $A_{i,j}$  is the  $i^{\text{th}}$  row of  $A_j$  for  $j = 1, \dots, p$ . The VAR coefficients in equation  $i$  are  $\beta_i = (c_i, A_{i,1}, \dots, A_{i,p})$ .

It is convenient to re-write the VAR-SV as

$$y_t = X_t \beta + W_t a_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_t), \quad (1)$$

where  $X_t = I_n \otimes (1, y'_{t-1}, \dots, y'_{t-p})$ ,  $\beta = (\beta_1, \dots, \beta_p)'$ ,  $a_t$  is an  $\frac{N(N-1)}{2} \times 1$  vector consisting of the free elements of  $A_{0t}$  stacked by rows, and  $W_t$  is an  $N \times \frac{N(N-1)}{2}$  matrix,

$$W_t = \begin{pmatrix} 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -y_{1,t} & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -y_{1,t} & -y_{2,t} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & -y_{1,t} & -y_{2,t} & \cdots & -y_{N-1,t} \end{pmatrix}.$$

The log-volatilities  $h_t = (h_{1,t}, \dots, h_{N,t})'$  and the time-varying parameters  $a_t$  are assumed to follow random walk processes:

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

$$a_t = a_{t-1} + \epsilon_t^a, \quad \epsilon_t^a \sim N(0, \Sigma_a), \quad (3)$$

where  $\Sigma_h = \text{diag}(\sigma_{h,1}^2, \dots, \sigma_{h,N}^2)$  and  $\Sigma_a = \text{diag}(\sigma_{a,1}^2, \dots, \sigma_{a, \frac{N(N-1)}{2}}^2)$ .

It can be seen that the VAR-SV can have an enormous number of parameters

when  $N$  is large. This has led large VAR researchers to work with restricted versions of the stochastic volatility process. An influential recent model is the common drifting volatility specification of Carriero, Clark and Marcellino (2016) which we denote by VAR-CCM1 and use in our empirical work. This is the same as the VAR-SV except that  $a_t = 0$  and  $\Sigma_t = e^{h_t} \Sigma$ , where the  $\Sigma$  is an  $N \times N$  positive definite matrix and  $h_t$  is a scalar stochastic volatility process:

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

This much more parsimonious specification has been successfully used with large VARs. But it does severely restrict the form that the time variation in the error covariance matrix can take. In our empirical work, we compare our new approach to the VAR-CCM1. We also use another specification proposed in Carriero, Clark and Marcellino (2019a) which we label VAR-CCM2. This amounts to the VAR-SV with  $a_t$  restricted to be time-invariant.

## 2.3 The Theory of Composite Likelihood Methods

### 2.3.1 Preliminaries

Assuming serially independent errors, the likelihood function (which, in our VAR context, will be a one-step ahead forecast density) for  $y = (y'_1, \dots, y'_T)'$  can be written as:

$$L(y; \theta) = \prod_{t=1}^T L(y_t; \theta), \quad (4)$$

where  $L(y_t; \theta) = p(y_t | \theta)$ . The composite likelihood is defined as

$$L^C(y; \theta) = \prod_{t=1}^T \prod_{i=1}^M L^C(y_{i,t}; \theta)^{w_i}, \quad (5)$$

where  $L^C(y_{i,t}; \theta) = p(y_{i,t} | \theta)$  is the quasi-likelihood and  $w_i$  is the weight attached to each quasi-likelihood with  $\sum_{i=1}^M w_i = 1$ . The weights will be discussed in sub-section 2.3.3.

The maximum composite likelihood estimator (MCLE) involves taking the maximum of  $L^C(y; \theta)$ . Bayesian estimation proceeds using a posterior based on the composite likelihood (i.e. the Bayesian composite posterior is  $p^C(\theta | y) \propto L^C(y; \theta) p(\theta)$ )

where  $p(\theta)$  is the prior).

In theory, the likelihood components used to build a composite likelihood can be anything. That is,  $y_{i,t}$  for  $i = 1, \dots, M$  can be any sub-sets of  $y_t$  and, indeed,  $y_{i,t}$  and  $y_{j,t}$  can overlap. For computational purposes, the key issue is that  $y_{i,t}$  and  $M$  should be small enough to lead to fast estimation. In this paper, all of our likelihood components are four dimensional VAR-SVs involving three core variables of interest that we wish to forecast and one additional variable. We partition  $y_t = \begin{pmatrix} y_t^* \\ z_t \end{pmatrix}$  where  $y_t^*$  is  $N_*$ -dimensional and contains the core variables of interest, while  $z_t$  (with elements denoted by  $z_{i,t}$ ) is the vector of dimension  $N_{other} = N - N_*$  that contains the remaining variables. Then we can let  $y_{i,t} = \begin{pmatrix} y_t^* \\ z_{i,t} \end{pmatrix}$  for  $i = 1, \dots, N_{other}$  and, hence,  $M = N_{other}$  in the number of models. Thus, our composite likelihood VAR-SV (VAR-CL-SV) application will involve quasi-likelihoods which are all  $(N_* + 1)$ -dimensional VAR-SVs.

### 2.3.2 Asymptotic Results

The standard frequentist way of investigating the theoretical properties of composite likelihoods is to assume that  $L(y; \theta)$  is the true data generating process involving a true parameter value  $\theta = \theta^0$  and derive the behavior of the MCLE. Results exist in the literature noting that the MCLE should converge asymptotically to  $\theta^0$  under certain assumptions (see, e.g., Varin, Reid and Firth, 2011, or Ribatet, Cooley and Davison, 2012). But such results are limited and model dependent. In this sub-section we derive some asymptotic results for our choice of quasi-likelihoods.

In (5), we have written the likelihood components as  $L(y_{i,t}; \theta)$  which all depend upon a common parameter vector  $\theta$ . In the VAR-CL-SV this will not be the case. Some parameters will not appear in any of the likelihood components. For instance, consider the equations for  $z_{i,t}$  and  $z_{j,t}$  for  $i \neq j$ . A large VAR-SV will contain a time-varying error covariance between these two equations. However, this error covariance will not appear in the composite likelihood function and so it will be impossible to obtain consistent estimates of it using  $L^C(y; \theta)$ . In other words, our choice of quasi-likelihoods means that we can never aim for asymptotic convergence to an unrestricted large VAR-SV. However, it is interesting to investigate what our methods do converge to. In this sub-section, we prove asymptotic convergence to a particular restricted

VAR-SV. We also highlight the connections between this restricted VAR-SV and the Minnesota prior, but emphasize that our approach allows for stochastic volatility while the conventional Minnesota prior does not.

For the choice of quasi-likelihoods made in the preceding sub-section, we can prove asymptotic convergence of the composite likelihood to that of a restricted VAR-SV of the following form:

$$\begin{pmatrix} A_{y,t} & 0 & 0 & \cdots & 0 & 0 \\ -\alpha'_{z,1,t} & 1 & 0 & 0 & \cdots & 0 \\ -\alpha'_{z,2,t} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \\ -\alpha'_{z,M-1,t} & 0 & 0 & \cdots & 1 & 0 \\ -\alpha'_{z,M,t} & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} y_t^* \\ z_{1,t} \\ z_{2,t} \\ \vdots \\ z_{M-1,t} \\ z_{M,t} \end{pmatrix} = \begin{pmatrix} c_y \\ c_{z,1} \\ c_{z,2} \\ \vdots \\ c_{z,M-1} \\ c_{z,M} \end{pmatrix} + \sum_{j=1}^p \begin{pmatrix} B_{yy,j} & \frac{w_1}{g(M)}\beta_{yz,1,j} & \cdots & \cdots & \frac{w_M}{g(M)}\beta_{yz,M,j} \\ \beta'_{zy,1,j} & \beta_{zz,1,j} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \beta'_{zy,M-1,j} & 0 & \cdots & \beta_{zz,M-1,j} & 0 \\ \beta'_{zy,M,j} & 0 & \cdots & 0 & \beta_{zz,M,j} \end{pmatrix} \begin{pmatrix} y_{t-j}^* \\ z_{1,t-j} \\ \vdots \\ z_{M-1,t-j} \\ z_{M,t-j} \end{pmatrix} + \begin{pmatrix} \epsilon_{y,t} \\ \epsilon_{z,1,t} \\ \vdots \\ \epsilon_{z,M-1,t} \\ \epsilon_{z,M,t} \end{pmatrix},$$

with  $\epsilon_{y,t} \sim N(0, \Sigma_{y,t})$ ,  $\epsilon_{z,i,t} \stackrel{iid}{\sim} N(0, e^{h_{N^*+i,t} - \ln w_i})$  independent of each other and

$$A_{y,t} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\alpha_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_{N^*1,t} & -\alpha_{N^*2,t} & \cdots & 1 \end{pmatrix}, \Sigma_{y,t} = \text{diag}(e^{h_{1,t}}, \dots, e^{h_{N^*,t}}).$$

Observe that this is a VAR-SV of the form

$$\tilde{A}_t y_t = c + \sum_{j=1}^p \tilde{B}_j y_{t-j} + \epsilon_t, \quad (6)$$

with some elements of  $\tilde{A}_t$  and  $\tilde{B}_j$  restricted to zero and some elements of  $\tilde{B}_j$  shrunk towards zero by factors  $\frac{w_1}{g(M)}, \dots, \frac{w_M}{g(M)}$  where  $g(M)$  is a function of  $M$ . We stress that the target model is a restricted VAR-SV with random walk laws of motion as in equations (2) and (3).

To provide some more insight into this restricted VAR-SV, note that the parameters in the VAR can be broken into three groups: those controlling the relationship between: i) the core variables of interest, ii) the core variables and other variables, and iii) the other variables. Note that those in the first group (i.e.  $A_{y,t}$  and  $B_{yy,j}$  for  $j = 1, \dots, p$ ) are left completely unrestricted. These are likely of most importance in ensuring good forecast performance. The parameters in the second group (i.e.  $\alpha_{z,i,t}$  and  $\beta_{yz,i,j}$  and  $\beta_{zy,i,j}$  for  $i = 1, \dots, M$  and  $j = 1, \dots, p$ ) are also likely important, but the restricted VAR only places very mild restrictions on them (i.e. through the presence of the  $\frac{w_i}{g(M)}$  for  $i = 1, \dots, M$  terms). It is only the parameters in the third group that are highly restricted. But since these control the relationships between the variables which are of no direct interest in the forecasting exercise, restricting them is likely to have little impact on forecast performance.

A word of explanation is in order about  $g(M)$ . A sufficient condition for the proof of the following proposition requires  $\frac{\sqrt{M}}{g(M)}$  to be bounded for all  $M$  (e.g. if  $g(M) = \sqrt{M}$  our proof follows standard law of large numbers results). But this condition is exactly what prior shrinkage in VARs usually does. That is, in our approach as  $M$  increases  $N$  also increases and the VAR dimension increases. It is standard for Bayesians working with large VARs to increase prior shrinkage (e.g. using the Minnesota prior) when VAR dimension increases (see, e.g., Table 1 of Banbura Giannone and Reichlin, 2010). Hence, the presence (and interpretation) of  $g(M)$  is justified as being comparable to the types of prior shrinkage commonly used in large Bayesian VARs. Note too that  $g(M)$  only applies to other lags in the equations for the core variables, so the convergence of the composite likelihood to a restricted VAR-SV only depends on the presence of shrinkage on these coefficients.

It is important to emphasize that  $L^C(y; \theta)$  is not a true likelihood in the sense that it is not a density in the data (conditional on parameters) that integrates to one. To compare it to a conventional likelihood for the restricted VAR-SV given in (6),  $L(y; \theta)$ , we consider the normalized composite likelihood

$$\tilde{L}^C(y; \theta) = \frac{L^C(y; \theta)}{\int_y L^C(y; \theta) dy}.$$

A useful measure of the approximation error associated with using  $L^C(y; \theta)$  instead of  $L(y; \theta)$  is the Kulback-Liebler divergence of  $L(y; \theta)$  from  $\tilde{L}^C(y; \theta)$ , denoted  $D_{\text{KL}}(L \parallel \tilde{L}^C)$ , which is summarized in the following proposition.

**Proposition 1** *Assume  $\max\{w_i\}$  is decreasing in  $M$  and  $\frac{\sqrt{M}}{g(M)} < \infty$  for all  $M \geq 1$ . Then*

$$\lim_{M \rightarrow \infty} D_{KL}(L \parallel \tilde{L}^C) = 0.$$

The proof of this proposition is in Section A.1 of the Online Appendix. The assumption that  $\max\{w_i\}$  is decreasing in  $M$  is innocuous as it implies only that when we add a new sub-model it has a non-zero weight which will leave less weight for the other models, including the model with maximum weight. Thus, our composite likelihood using small VAR-SVs as quasi-likelihoods asymptotically converges to the likelihood of a particular large VAR-SV under sensible assumptions.

Of course, given the way we have defined our quasi-likelihoods, it is not possible to asymptotically converge to an unrestricted large VAR-SV since (as noted previously) some of the unrestricted model’s parameters appear in none of the quasi-likelihoods. If interest lies in using composite likelihood methods to provide estimates of all the parameters in a large VAR-SV, then other quasi-likelihoods should be chosen to build a composite likelihood function (e.g. building a set of quasi-likelihoods involving all possible bivariate or tri-variate combinations of the variables). Our choice of quasi-likelihoods is based on our choice of empirical problem. We are interested in forecasting a small number of variables, using the other variables only to improve these forecasts. For this, our choice of quasi-likelihoods is a sensible one.

### 2.3.3 Composite Likelihoods as Opinion Pools

An alternative way of theorizing about composite likelihoods, popular among Bayesians (see, e.g., Roche, 2016) is to begin by assuming there is some feature of interest,  $\theta$  (in our case, the error covariance matrix relating to the core variables). There are many “agents” each of which uses a (possibly agent-specific) information set to produce an “opinion” (i.e. a posterior) about  $\theta$ . The opinions going into the pool can be obtained from any source. The question arises as to how to pool these opinions? There is a literature on such opinion or prediction pools. Hall and Mitchell (2007) and Geweke and Amisano (2011) are influential approaches in econometrics. Genest, Weerahandi, Zidek (1984) and Genest, McConway and Schervish (1986) are influential early references which establish or review many theoretical properties of opinion pools.

If, in our case, we interpret each quasi-likelihood,  $L^C(y_{i,t}; \theta)$ , as arising from an

agent, we can draw on this literature to obtain a theoretical justification for our approach. Papers such Roche (2016) show that Bayesian inference using the composite likelihood can be interpreted as arising from a generalized logarithmic opinion pool. This offers strong theoretical justification for our approach. Genest *et al.* (1984) show that such opinion pools have attractive properties including external Bayesianity. External Bayesianity implies that if all agents agree on the same prior, then it does not matter whether the prior is added before or after the opinions are pooled. Generalized logarithmic opinion pools are the only class of opinion pools that have this property.

An alternative approach is to use linear opinion pools (e.g. Hall and Mitchell, 2007, and Geweke and Amisano, 2011). The use of linear opinion pools means the resulting approach does not satisfy external Bayesianity and does not lead to Bayesian inference based on  $p^C(\theta|y)$  either. However, as discussed in Geweke and Amisano (2011), linear pools sometimes give results that are comparable to logarithmic opinion pools. Hence, even though they are not a composite likelihood approach, they are closely related and we include them in our set of empirical results.

The advantage of drawing on the opinion pool literature is that it offers insights into how the weights,  $w_i$  for  $i = 1, \dots, M$ , can be chosen. In our empirical work, we consider a range of approaches. In the linear opinion pool formulation, Geweke and Amisano (2011) derive a set of weights which are optimal for the linear pool and provide a method for calculating them which we follow in this paper.

In the logarithmic opinion pool formulation, a logical thing to do (see Canova and Matthes, 2018) is to base the weights on some measure of the fit of each quasi-likelihood. In our application, where each quasi-likelihood is a VAR-SV involving a set of core variables ( $y_t^*$ ) and one other variable, it makes sense to use the marginal likelihood or an approximation to it to calculate the weights. Hence, we consider weighting schemes based on the Bayesian information criterion (BIC), the Deviance Information criterion (DIC) and the marginal likelihood (ML). Letting  $\text{BIC}_i$  be the BIC for sub-model  $i$ , we have

$$\text{BIC}_i = -2 \log L(y^*; \hat{\theta}_i) + d \log(T),$$

where  $\hat{\theta}$  is the maximum likelihood estimate using sub-model  $i$ ,  $y^* = (y_1^*, \dots, y_T^*)'$  and  $d$  is the number of free parameters. We stress that, in each quasi-likelihood, we are

only using the core variables (which are common to all quasi-likelihoods) to define the BIC. In other words, the likelihood in the BIC (and the posterior in the formula for the marginal likelihood below) is based on a three-dimensional conditional density (i.e. the density for the core variables conditional on the other variable). The maximum likelihood estimate is computed using the integrated likelihood as in Chan and Eisenstat (2018).

The weight for each sub-model is computed as

$$w_i^{BIC} = \frac{e^{-\frac{1}{2}\text{BIC}_i}}{\sum_{j=1}^M e^{-\frac{1}{2}\text{BIC}_j}}, \quad \text{for } i = 1, \dots, M.$$

Our second set of weights follows the same strategy, but using DIC instead of BIC. DIC is calculated based on the integrated likelihood for the core variables of interest (see Chan and Grant, 2016, for details).

The third weighting scheme is based on the marginal likelihood. We use the following marginal likelihood for sub-model  $i$ :

$$\text{ML}_i = \int p_i(y^*|\theta)p(\theta)d\theta,$$

where  $p_i(y|\theta) = \prod_{t=1}^T L^C(y_{i,t}; \theta)$  and  $p_i(y^*|\theta)$  implies evaluating the marginal likelihood only using the core variables. The weight for each sub-model is computed as

$$w_i^{ML} = \frac{\text{ML}_i}{\sum_{j=1}^M \text{ML}_j}, \quad \text{for } i = 1, \dots, M.$$

We calculate the marginal likelihood using the methods of Chan and Eisenstat (2018). We use the abbreviations, VAR-CL-BIC, VAR-CL-DIC and VAR-CL-ML for composite likelihood methods involving these three different weights.

In the linear opinion pool approach we calculate the optimal weights using the method described in Geweke and Amisano (2011). This involves the following steps. Let  $p_i(y_t^*|y_{1:t-1})$  be the one-step-ahead predictive density for the core variables for the  $i^{\text{th}}$  sub-model and  $w = (w_1, w_2, \dots, w_M)'$ . The predictive log score function is given by

$$f(w) = \sum_{t=1}^T \log \left( \sum_{i=1}^M w_i p_i(y_t^*|y_{1:t-1}) \right).$$

The optimal weight is obtained by solving the optimization problem  $\hat{w} = \text{argmax}_w f(w)$ .

We use VAR-LIN as the abbreviation for this approach. Even though these weights are calculated to be optimal in the linear opinion pool case, we can also use them as weights in the composite likelihood. We refer to such an approach as VAR-CL-LIN.

The main research question of interest in this paper is whether composite likelihood methods involving many small models can forecast well in the presence of large data sets. A subsidiary question though, is whether the general idea of combining many small models for forecasting is a good one. This idea is exploited in many different ways in the econometrics literature (see, e.g., the subset regression approach of Elliott, Gargano and Timmermann, 2013). Here we focus on logarithmic and linear opinion pools. By including linear opinion pooling methods we can address the second question. To preview our empirical findings, we find that all approaches which combine many small models forecast well. That is, it seems that the empirical success of our approach is largely due to the choice of quasi-likelihoods as opposed to the way they are combined. However, it is worth noting that (as detailed in the computational times comparison provided in Section B.3 of the Online Appendix) the linear opinion pool has significant computational drawbacks relative to the composite likelihood approach.

### 3 Bayesian Analysis Using the Composite Posterior

Our goal is to carry out Bayesian analysis on the composite posterior,  $p^C(\theta|y)$ , using MCMC draws from each of the quasi-posterior distributions. This section develops an algorithm for doing so.

#### 3.1 Quasi-Posterior Distributions

We first extend our earlier notation to define the quasi-likelihoods. Remember that each of these is a VAR-SV that combines core variables of interest,  $y_t^*$ , with an additional variable,  $z_{i,t}$ . Thus, quasi-likelihood  $i$  (for  $i = 1, \dots, M$ ) can be expressed

in the form:

$$A_{y,t}y_t^* = X_{y,t}\beta_y + X_{z_i,t}\beta_{y z_i} + \epsilon_{y,t}, \quad \epsilon_{y,t} \sim N(0, \Sigma_{y,t}), \quad (7)$$

$$z_{i,t} = y_t^* \alpha_{z_i,t} + X_t \beta_{z_i} + \epsilon_{z_i,t}, \quad \epsilon_{z_i,t} \sim N(0, e^{h_{N_*+i,t}}), \quad (8)$$

$$A_{y,t} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N_*1,t} & \alpha_{N_*2,t} & \cdots & 1 \end{pmatrix}, \quad \Sigma_{y,t} = \begin{pmatrix} e^{h_{1,t}} & & & \\ & e^{h_{2,t}} & & \\ & & \ddots & \\ & & & e^{h_{N_*t}} \end{pmatrix}.$$

In (7), the matrix  $X_{y,t}$  contains lags of  $y_t^*$ , and the matrix  $X_{z_i,t}$  contains lags of  $z_{i,t}$ . The log-volatilities  $h_{i,t}$  and the time-varying parameters  $\alpha_{z_i,t}$  and  $\alpha_{jk,t}$ ,  $i = 1, \dots, M, j = 2, \dots, N_*, k = 1, \dots, j-1$  are assumed to follow random walk processes:

$$h_{i,t} = h_{i,t-1} + \epsilon_{i,t}^h, \quad \epsilon_{i,t}^h \sim N(0, \sigma_{h,i}^2), \quad (9)$$

$$\alpha_{jk,t} = \alpha_{jk,t-1} + \epsilon_{jk,t}^\alpha, \quad \epsilon_{jk,t}^\alpha \sim N(0, \sigma_{\alpha,jk}^2), \quad (10)$$

$$\alpha_{z_i,t} = \alpha_{z_i,t-1} + \epsilon_{i,t}^\alpha, \quad \epsilon_{i,t}^\alpha \sim N(0, \Sigma_{\alpha,i}), \quad (11)$$

where  $\Sigma_{\alpha,i}$  is a diagonal matrix.

Let  $\theta = \{\beta_y, A_{y,1}, \dots, A_{y,T}, \Sigma_{y,1}, \dots, \Sigma_{y,T}\}$  be the set of parameters that are common in all quasi-likelihoods, and denote by

$$\eta_i = \{\beta_{y z_i}, \beta_{z_i}, \alpha_{z_i,0}, \dots, \alpha_{z_i,T}, h_{N_*+i,0}, \dots, h_{N_*+i,T}, \Sigma_{\alpha,i}\}$$

the parameters that appear only in quasi-likelihood  $i$ . Each quasi-posterior  $i$  is given by

$$p_i(\theta, \eta_i | y^*, z_i) = p(\theta, \eta_i) p(y^*, z_i | \theta, \eta_i) / p(y^*, z_i),$$

where  $z_i = (z_{i,1}, \dots, z_{i,T})'$ .

A key feature of our set-up is that the density that defines each quasi-likelihood

can be conveniently decomposed as:

$$\begin{aligned}
p(y^*, z_i | \theta, \eta_i) &= \prod_{t=1}^T p(y_t^* | y_{t-1}^*, \dots, y_{t-p}^*, z_{i,t-1}, \dots, z_{i,t-p}, \beta_y, \beta_{y_{z_i}}, A_{y,t}, \Sigma_{y,t}) \\
&\quad \times p(z_{i,t} | y_t^*, y_{t-1}^*, \dots, y_{t-p}^*, z_{i,t-1}, \dots, z_{i,t-p}, \beta_{z_i}, \alpha_{z_i,t}, h_{N_*+i,t}), \\
&= \left( \prod_{t=1}^T p(y_t^* | \cdot) \right) \left( \prod_{t=1}^T p(z_{i,t} | y_t^*, \cdot) \right), \\
&= p(y^* | \tilde{z}_i, \theta, \beta_{y_{z_i}}) p(z_i | y^*, \tilde{\eta}_i),
\end{aligned}$$

where  $\tilde{z}_i = \{z_{i,1}, \dots, z_{i,T-1}\}$  and  $\tilde{\eta}_i = \{\beta_{z_i}, \alpha_{z_i,0}, \dots, \alpha_{z_i,T}, h_{N_*+i,0}, \dots, h_{N_*+i,T}, \Sigma_{\alpha,i}\}$ .

In this decomposition,  $p(y^* | \tilde{z}_i, \theta, \beta_{y_{z_i}})$  is the density of a multivariate normal distribution that can be regarded as the likelihood for the model in (7), with  $z_{i,1}, \dots, z_{i,T-1}$  treated as exogenous regressors. Moreover, this density can be integrated analytically with respect to a prior on  $\beta_{y_{z_i}}$  to obtain a density that only contains common parameters  $\theta$ , i.e.,  $p(y^* | \tilde{z}_i, \theta) = \int_{\beta_{y_{z_i}}} p(\beta_{y_{z_i}}) p(y^* | \tilde{z}_i, \theta, \beta_{y_{z_i}}) d\beta_{y_{z_i}}$ . Similarly,  $p(z_i | y^*, \tilde{\eta}_i)$  can be viewed as the multivariate normal likelihood for a time-varying parameter autoregressive distributed lag model (TVP-ARDL) with exogenous  $y_t^*$  defined by (8), with the important feature that it contains only nuisance parameters.

Consequently, if  $\theta$  and  $\tilde{\eta}_i$  are independent in the prior (as we assume in this paper), then they are also independent in the  $i$ -th quasi-posterior. Moreover, this independence carries over to the composite posterior defined as

$$\begin{aligned}
p^C(\theta, \tilde{\eta}_1, \dots, \tilde{\eta}_M | y^*, z_1, \dots, z_M) &\propto p(\theta) \prod_{i=1}^M p(\tilde{\eta}_i) p(y^*, z_i | \theta, \tilde{\eta}_i)^{w_i}, \\
&= p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M) \prod_{i=1}^M p^C(\tilde{\eta}_i | y^*, z_i), \quad (12)
\end{aligned}$$

where

$$\begin{aligned}
p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M) &\propto p(\theta) \prod_{i=1}^M p(y^* | \tilde{z}_i, \theta)^{w_i}, \\
p^C(\tilde{\eta}_i | y^*, z_i) &\propto p(\tilde{\eta}_i) p(z_i | y^*, \tilde{\eta}_i)^{w_i}.
\end{aligned}$$

The decomposition in (12) is crucial as it allows us to sample the common pa-

rameters  $\theta$  and each  $\tilde{\eta}_i, i = 1, \dots, M$  independently. Consequently, we can parallelize the sampling and vastly reduce the computation time. This is taken up in the next sub-section.

### 3.2 Simulation from the Composite Posterior

This section describes our computational algorithm to simulate from the composite posterior  $p^C(\theta, \tilde{\eta}_1, \dots, \tilde{\eta}_M | y^*, z_1, \dots, z_M)$ . Instead of designing an MCMC algorithm to directly sample from this composite posterior, we develop an accept-reject algorithm using MCMC draws from the individual quasi-posterior distributions  $p_i(\theta | y^*, z_i), i = 1, \dots, M$ , as proposals. The key advantage of this approach is that sampling from each of the quasi-posterior distributions can be done in parallel and using standard MCMC methods for small VAR-SV models.

Using the decomposition of the composite posterior in (12), we can generate samples from  $p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M), p^C(\tilde{\eta}_1 | y^*, z_1), \dots, p^C(\tilde{\eta}_M | y^*, z_M)$  independently. We start with simulating the common parameters  $\theta$  from  $p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M)$  by appropriately pooling draws of  $\theta$  from the quasi-posteriors. We develop an accept-reject algorithm for this purpose.

Consider the proposal density  $q(\theta)$  that is a mixture of the  $M$  quasi-posteriors, i.e.,

$$q(\theta) = \sum_{i=1}^M w_i p_i(\theta | y^*, z_i) = p(\theta) \sum_{i=1}^M \frac{w_i p(y^* | \tilde{z}_i, \theta)}{p(y^* | \tilde{z}_i)},$$

where  $p(y^* | \tilde{z}_i) = \int_{\theta} p(\theta) p(y^* | \tilde{z}_i, \theta) d\theta$  can be regarded as the marginal likelihood of the VAR-SV with exogenous variables defined in (7).

Given draws from the  $M$  quasi-posteriors  $p_i(\theta | y^*, z_i)$  for  $i = 1, \dots, M$  and a set of weights  $w_i$  for  $i = 1, \dots, M$ —which can be any of those described in section 2.3.3—it is easy to obtain a set of draws from  $q(\theta)$ . Moreover,  $q(\theta)$  can be readily evaluated:  $p(y^* | \tilde{z}_i)$  can be computed using the algorithm of Chan and Eisenstat (2018) that we use to obtain the marginal likelihood in a VAR-SV (see Section 2.3.3) and  $p(y^* | \tilde{z}_i, \theta)$  is a multivariate normal density.

To show the latter claim, let  $\beta_{yz_i} \sim N(\underline{\beta}_{yz}, \underline{V}_{\beta,z})$  denote the prior for  $\beta_{yz_i}$ . Let  $\alpha_{y,t}$  represent the free elements in  $A_{y,t}$  stacked by row, and let  $W_{y,t}$  denote the associated covariate matrix (see the discussion in Section 2.2). Then,  $p(y^* | \tilde{z}_i, \theta)$  has the following

multivariate normal form:

$$(y^* | \tilde{z}_i, \theta) \sim N \left( W_y \alpha_y + X_y \beta_y + X_{z_i} \beta_{yz}, X_{z_i} V_{\beta,z} X_{z_i}' + \Sigma_y \right),$$

where  $\Sigma_y$  is a block diagonal matrix with diagonal blocks  $\Sigma_{y,t}, t = 1, \dots, T$ . Finally,  $X_y, X_{z_i}, W_y$  and  $\alpha_y$  respectively stack  $X_{y,t}, X_{z_i,t}, W_{y,t}$  and  $\alpha_{y,t}$  over  $t = 1, \dots, T$ .

To be a valid accept-reject algorithm with proposal density  $q(\theta)$ , we need to show that the ratio  $p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M)/q(\theta)$  is bounded for all  $\theta$  in its support. To that end, observe that

$$r(\theta) = \frac{\prod_{i=1}^M [p(y^* | \tilde{z}_i, \theta)/p(y^* | \tilde{z}_i)]^{w_i}}{\sum_{i=1}^M w_i p(y^* | \tilde{z}_i, \theta)/p(y^* | \tilde{z}_i)} \leq 1.$$

This inequality follows from the fact that a geometric average is always less than or equal to the corresponding arithmetic average. Now, write the target density as

$$p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M) = \frac{p(\theta) \prod_{i=1}^M p(y^* | \tilde{z}_i, \theta)^{w_i}}{K_i},$$

where  $K_i = \int p(\theta) \prod_{i=1}^M p(y^* | \tilde{z}_i, \theta)^{w_i} d\theta$  is the normalizing constant. If we let  $K = \prod_{i=1}^M p(y^* | \tilde{z}_i)^{w_i}/K_i$ , then we can show that  $p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M) \leq Kq(\theta)$  for all  $\theta$ :

$$\frac{p^C(\theta | y^*, \tilde{z}_1, \dots, \tilde{z}_M)}{Kq(\theta)} = \frac{p(\theta) \prod_{i=1}^M [p(y^* | \tilde{z}_i, \theta)p(y^* | \tilde{z}_i)]^{w_i}}{p(\theta) \sum_{i=1}^M w_i p(y^* | \tilde{z}_i, \theta)/p(y^* | \tilde{z}_i)} = r(\theta) \leq 1.$$

This suggests an accept-reject sampling approach to pool draws of common parameters obtained from individual quasi-posteriors.<sup>3</sup> We summarize the algorithm as follows:

1. obtain a proposal draw  $\theta^* \sim q(\theta)$  as follows:
  - (a) sample from  $p_i(\theta | y^*, z_i)$  *in parallel* for  $i = 1, \dots, M$ , using standard methods (e.g., the Gibbs sampler of Primiceri, 2005);
  - (b) for each  $p_i(\theta | y^*, z_i)$  compute relevant quantities such as the BIC, DIC

---

<sup>3</sup>For a general discussion of the accept-reject method, see, e.g., Section 3.1.5 in Kroese, Taimre and Botev (2011). Since the proposal draws are obtained from the quasi-posteriors using MCMC, they are correlated by construction. Consequently, the sample obtained from this accept-reject algorithm would also be correlated.

and/or ML (again, using parallel operations and efficient algorithms—e.g., those developed in Chan and Eisenstat, 2018);

- (c) use the BICs, DICs, and/or MLs to compute the composite likelihood weights  $w_1, \dots, w_M$ ;
2. accept  $\theta^*$  with probability  $r(\theta^*)$  using the accept-reject algorithm.

Next, we consider obtaining draws from  $p^C(\tilde{\eta}_i | y^*, z_i)$  for each  $i = 1, \dots, M$ . The draws of  $\tilde{\eta}_i$  are only needed to compute MLs, DICs, and BICs, which are only used to compute the weights  $w_1, \dots, w_M$ . If the weights are known (e.g. as in the equal weights case), then there is no need to draw  $\tilde{\eta}_i$ . Note also that the MLs used to evaluate  $r(\theta)$  in implementing the accept-reject algorithm do not require  $\tilde{\eta}_i$ . Since, in some cases, there is no need to produce draws of  $\tilde{\eta}_i$  and the algorithm is only a slight adaptation of one for the TVP-ARDL model, we relegate technical details to Section A.2 of the Online Appendix.

Once draws of  $\theta$  and  $\tilde{\eta}_1, \dots, \tilde{\eta}_M$  are obtained from the quasi-posterior distribution, it is straightforward to simulate from the joint predictive density of the core variables as well. The exact procedure is described in detail in Section A.3 of the Online Appendix. We use this approach in the following section to conduct a comprehensive pseudo out-of-sample forecasting exercise.

## 4 Forecasting Results

### 4.1 Overview

We carry out an empirical investigation of our composite likelihood methods using a large US quarterly data set involving 196 variables. The data are taken from the Federal Reserve Bank of St. Louis' FRED-QD data set and run from 1959Q1-2015Q3.<sup>4</sup> All data are transformed to stationarity following the recommendations in the FRED-QD data base. We focus on empirical results relating to three variables: CPI inflation, GDP growth and the Federal Funds rate and refer to these as the core variables. We also present results using small and medium data sets of 7 and 20

---

<sup>4</sup>The data are available through <https://research.stlouisfed.org/econ/mccracken/fred-databases/>. See also McCracken and Ng (2016). Complete details of all the variables in the data set are provided there.

variables, respectively. Details of which variables are included in which data set are given in Section B.1 of the Online Appendix. A lag length of four is used for all models.

We compare our composite Bayesian methods to a range of alternative methods. Our choice of these is motivated by (and this section organized around) our wish to answer the following questions.

The first of these is: Do composite likelihood methods forecast well compared to other scalable high dimensional approaches? To answer this question the key comparison is with the homoskedastic large VAR with natural conjugate prior as this is the most popular approach that is computationally feasible in the large data set. This is labelled "Large VAR" in the table below. If we find composite likelihood approaches based on VAR-SVs beat the Large VAR (as we do), then a subsidiary question arises. Are our composite likelihood methods winning due to the fact that they allow for stochastic volatility or some other aspect of the composite likelihood approach itself? To address this question, we also present forecasting results using our composite likelihood methods on homoskedastic VARs. These are exactly like our other composite likelihood approaches, but using homoskedastic VARs to define the quasi-likelihoods. In the interests of brevity, we only present results for the equally-weighted case and label this approach VAR-HM-CL-EQ.

The second question is: Do our large data methods forecast well relative to simpler methods using smaller data sets? Other papers working with similar US quarterly data sets and alternative modelling approaches have tended to find that working with large VARs does improve forecast performance relative to small VARs. However, the evidence is often not that strong. For instance, Koop (2013) finds that, compared to small VARs, moving towards larger VARs does improve forecast performance, but there comes a point where adding extra variables into the VAR offers only modest improvements in forecast performance. To address this question, we present results using the unrestricted VAR-SV using small data sets.

When working with a small data set, the researcher must choose which variables to include. This can potentially have important consequences for forecast performance. It is possible that a bad choice will lead to poor forecast performance. An advantage of our composite Bayesian methods (or any large VAR method) is that such a choice does not have to be made. All the 196 variables are included and the algorithm decides which get more weight in the composite posterior. To illustrate the potential

consequence of making a poor choice in a small data set, we present results from a 7-variate unrestricted VAR-SV involving the 3 core variables and 4 additional variables which have been commonly used for forecasting the core variables in other studies. We refer to these as the four “good” variables. We also present results from this model except that the 4 “good” variables are replaced by 4 “bad” variables. These are chosen in an informal manner as being among the ones which have low correlations with the core variables. Thus we have both a “good” and “bad” VAR-SV with the small data set. Section B.1 of the Online Appendix provides a list of these “good” and “bad” variables.

The third question relates to stochastic volatility and asks: How do our methods compare to other approaches which restrict the form of the multivariate stochastic volatility process? Several approaches have been suggested which are, to different extents, scalable. These include the VAR-CCM1 and VAR-CCM2 models described in Section 2.2. We produce results for these using the “good” small and the medium data sets. In addition, we include the VAR with multivariate stochastic volatility modelled using the factor structure of Kastner (2019), estimated with the medium data set. We present results for two versions of this model, one has a single factor and the other two factors.

Table 1 provides a list of all these models along with their acronyms.

---



---

Table 1: Models used in Forecasting Exercise

---

VAR-CL-BIC	composite Bayesian VAR-SV with BIC based weights
VAR-CL-DIC	composite Bayesian VAR-SV with DIC based weights
VAR-CL-EQ	composite Bayesian VAR-SV with equal weights
VAR-CL-ML	composite Bayesian VAR-SV with marginal likelihood weights
VAR-CL-LIN	composite Bayesian VAR-SV with linear pool weights
VAR-LIN	VAR-SV with linear pool weights as in Geweke and Amisano (2011)
VAR-HM-CL-EQ	composite Bayesian homoscedastic VAR with equal weights
Large VAR	Large Homoskedastic VAR
VAR-SV-3	VAR-SV using core variables only
VAR-SV-g	VAR-SV using small data set (“good” variables)
VAR-SV-b	VAR-SV using small data set (“bad” variables)
VAR-CCM1	Carriero, Clark and Marcellino (2016) using small data set (“good” variables)
VAR-CCM2	Carriero, Clark and Marcellino (2019a) using small data set (“good” variables)
VAR-CCM1-20	Carriero, Clark and Marcellino (2016) using medium data set
VAR-CCM2-20	Carriero, Clark and Marcellino (2019a) using medium data set
VAR-FSV-1f	VAR with factor SV model of Kastner (2019), 1 factor
VAR-FSV-2f	VAR with factor SV model of Kastner (2019), 2 factors

---



---

Further details about the specification of all models, including prior hyperparameter choice, are given in Section A.3 of the Online Appendix. For the VAR coefficients in all models we make standard Minnesota prior choices. Where possible, we make identical specification and prior hyperparameter choices across models. It is worth stressing that, in conventional large VAR approaches where the number of parameters being estimated exceeds the number of observations, prior elicitation is crucial. Priors must be very informative and results can be sensitive to prior choice. An advantage of composite likelihood approaches is that, since all sub-models used are small, prior elicitation is a less important issue. It is possible to use less informative priors and prior sensitivity concerns are mitigated.

To evaluate forecast performance, we use two point forecast metrics and two density forecast metrics for the core variables. Let  $y_t^* = (y_{t,1}^*, y_{t,2}^*, y_{t,3}^*)'$  denote the random variables being forecast and  $y_t^R = (y_{t,1}^R, y_{t,2}^R, y_{t,3}^R)'$  be their realizations. For the point forecast, we report the root mean squared forecast error (RMSFE) and the mean

absolute forecast error (MAFE),

$$\text{RMSFE}_i = \sqrt{\frac{\sum_{t=t_0}^{T-h} (y_{t+h,i}^R - E(y_{t+h,i}^* | y_{1:t}^R))^2}{T-h-t_0+1}}.$$

$$\text{MAFE}_i = \frac{\sum_{t=t_0}^{T-h} |y_{t+h,i}^R - \hat{y}_{t+h,i}|}{T-h-t_0+1},$$

for  $i = 1, 2, 3$  where  $E(y_{t+h} | y_{1:t}^R)$  is the mean of the predictive density and  $\hat{y}_{t+h}^M$  is the median of the predictive density. For the density forecasts, we report the average log-predictive likelihoods (ALPL) and the average continuous rank probability score (ACRPS),

$$\text{ALPL}_i = \frac{\sum_{t=t_0}^{T-h} \log p_{t+h}(y_{t+h,i}^* = y_{t+h,i}^R | y_{1:t}^R)}{T-h-t_0+1},$$

$$\text{ACRPS}_i = \frac{1}{T-h-t_0+1} \sum_{t=t_0}^{T-h} \text{CRPS}_{t,i},$$

for  $i = 1, 2, 3$  where  $\text{CRPS}_{t,i} = \int_{-\infty}^{\infty} (F_{t+h}(z) - 1(y_{t+h}^R < z))^2 dz = E_{p_{t+h}} |y_{t+h,i}^* - y_{t+h,i}^R| - 0.5 E_{p_{t+h}} |y_{t+h,i}^* - y_{t+h,i}^R|$  and  $F_{t+h}(\bullet)$  is the c.d.f. of the predictive density. A small value of the  $\text{ACRPS}_i$  indicates a better forecasting performance.

We also present a joint ALPL for the three core variables of interest:

$$\text{ALPL} = \frac{\sum_{t=t_0}^{T-h} \log p_{t+h}(y_{t+h}^* = y_{t+h}^R | y_{1:t}^R)}{T-h-t_0+1}.$$

We present results for a forecast evaluation period beginning in 1970 and running to the end of the sample.<sup>5</sup>

We provide forecasts of quarterly variables one quarter ( $h = 1$ ) and one year in the future ( $h = 4$ ). We carry out a one-sided sign test of equal predictive accuracy of Diebold and Mariano (1995). All tests compare a specific model to the benchmark. Different tables use different benchmarks depending on which of our three questions is being addressed. In the tables, \*\*\*, \*\* and \* denote findings of statistically significant forecast improvements relative to the benchmark at the 1%, 5% and 10% level of

---

<sup>5</sup>In Section B.4 of the Online Appendix we also present results for a short forecast evaluation period that begins in 2008Q1 so as to take in only the financial crisis and subsequent period. This shorter period yields results which are similar, but slightly more favorable to our methods, than those produced here.

significance, respectively.

The methods developed in this paper are of particular use for forecasting. If estimating large time-varying covariance matrices were the focus, we would choose the sub-models differently. However, it is worth noting that, in Section B.2 of the Online Appendix, we present results from a small Monte Carlo study where we use parameter estimates from the VAR-SV using the small data set to construct a data generating process (DGP). Then we artificially generate 100 artificial data sets from this DGP and produce estimates of the time-varying error variances and covariances for the first three equations (i.e. those for the three core variables). All VAR-CL specifications produce estimates which are close to those of the VAR-SV.

Section B.3 of the Online Appendix also includes evidence on the computational burden of various approaches. As expected, this demonstrates the scalability of the composite likelihood approaches in that the computational burden increases roughly linearly in  $N$ . The same cannot be said, for instance, for VAR-CCM2 where the computational burden increases at a much faster rate. For VAR-CCM2 the computational burden is similar to composite Bayesian approaches when  $N = 20$ , but is almost 10 times more burdensome for  $N = 100$ . It is also worth noting that the linear opinion pool is much more computationally demanding than composite likelihood methods since it involves recursive estimation and numerical optimization (see Section B.3 of the Online Appendix and Geweke and Amisano, 2011).

## 4.2 Do Our Methods Forecast Better Than the Large VAR?

The large homoskedastic VAR using the Minnesota prior (or similar) can be used in large data sets involving hundreds of variables and is the most popular VAR approach at present. The best overall summary of forecast performance involves the entire joint predictive density for the three core variables. These are presented in Table 2 for  $h = 1$  and  $h = 4$ . The evidence in this table is overwhelming. All of our composite Bayesian methods as well as the linear opinion pool forecast much better than the large VAR at both forecast horizons. The Diebold-Mariano tests indicate these improvements are highly statistically significant. Table 3 presents results for the core variables individually and for a wider variety of forecast metrics. This table also indicates strong, statistically significant, improvements in forecasting by our composite likelihood methods for the vast majority of cases. It is instructive to look

at the few cases where the composite Bayesian methods are not significantly better than the large VAR. These occur only for  $h = 1$  and for two variables: inflation and GDP growth. Most noteworthy is the fact that these exceptions are only found when using MSFEs and MAFEs (measures of point forecasts performance) and never found when using ACRPSs and ALPLs (measures based on the predictive density). This highlights the importance of adding stochastic volatility. The homoskedastic large VAR is not able to model time-varying predictive variances. This inability to do so means that, even if its point forecasts are reasonable (at least for inflation and GDP growth for  $h = 1$ ), its predictive densities are not.

When we use our composite likelihood methods with homoskedastic VARs, we also find forecast performance to be worse than approaches which allow for stochastic volatility. This can be seen in the rows of Tables 2 and 3 for VAR-HM-CL-EQ. This provides additional evidence of the presence of stochastic volatility in this data set. In addition, it is interesting to note that VAR-HM-CL-EQ is forecasting much better than the large homoskedastic VAR. This suggests that there are some benefits from using composite Bayesian methods with VARs apart from their ability to easily incorporate stochastic volatility. Working with many smaller parsimonious models, and combining them using composite likelihood methods, is producing better forecasts than those produced by a single large model which induces parsimony through the use of the Minnesota prior.

Tables 2 and 3 also allow us to compare the different ways of weighting used with our composite Bayesian methods. What they show is that the alternative ways of doing the weighting typically do not make a great deal of difference for forecasting. There is slight evidence that BIC and DIC based weights are inferior to the other weights. But marginal likelihood and linear opinion pool weights are roughly the same and forecast very well. There are even cases where the simplest strategy of using equal weights forecasts well. And the VAR-LIN of Geweke and Amisano (2011) also is typically among the top performing approaches, although we remind the reader that this is computationally much more demanding than our composite Bayesian methods (see Section B.3 of the Online Appendix). Accordingly, in the remainder of this paper we will focus on composite likelihood methods using marginal likelihoods as weights. Results with other weighting schemes are similar and are available in Section B.4 of the Online Appendix.

Table 2: Joint average log-predictive likelihoods for 3 core variables under 8 large models and 2 forecast horizons ( $h = 1, 4$ )

Horizon	$h = 1$	$h = 4$
Large VAR	1.040	-0.570
VAR-LIN	8.645***	6.993***
VAR-CL-ML	8.153***	6.455***
VAR-CL-DIC	6.281***	5.076***
VAR-CL-BIC	5.857***	6.190***
VAR-CL-LIN	8.428***	6.416***
VAR-CL-EQ	8.440***	6.855***
VAR-HM-CL-EQ	6.159***	5.170***

Note: DM statistics are based on the large VAR benchmark. \*\*\*, \*\* and \* denote findings of statistically significant forecast improvements relative to the benchmark at the 1%, 5% and 10% level of significance, respectively.

Table 3: Individual forecast performance in terms of RMSFE, MAE, ACRPS and ALPL for the 3 core variables under 8 large models and 2 forecast horizons ( $h = 1, 4$ )

Evaluation of GDP Growth Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
large VAR	0.850	0.660	4.590	1.234	1.037	0.806	5.868	1.135
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-HM-CL-EQ	0.856	0.646	0.982***	3.322***	0.832***	0.618***	1.009***	3.302***
Evaluation of Inflation Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
large VAR	0.580	0.430	4.570	1.237	0.959	0.739	5.852	1.137
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-HM-CL-EQ	0.633	0.473	0.906***	3.377***	0.654***	0.470***	0.955***	3.336***
Evaluation of Interest Rate Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
Large VAR	1.091	0.737	0.655	-1.429	3.020	2.087	1.870	-2.821
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-HM-CL-EQ	1.021	0.651	0.534***	-0.546***	1.513***	1.116***	1.369***	-1.477***

Note: DM statistics are based on the large VAR benchmark, and the RMSFEs, MAEs and ACRPSs are multiplied by 100 for easy comparison. \*\*\*, \*\* and \* denote findings of statistically significant forecast improvements relative to the benchmark at the 1%, 5% and 10% level of significance, respectively.

### 4.3 Do Our Methods Forecast Well Relative to Simpler Methods Using Smaller Data Sets?

It is possible that the good forecast performance of our methods found in the preceding sub-section was solely due to the fact that they allowed for stochastic volatility whereas the large VAR did not and the inclusion of a large data set brought no additional benefits. If this were the case, then it is possible that working with a small

data set suffices, provided stochastic volatility is added to the model. There would be no need to consider composite Bayesian methods since unrestricted VAR-SVs could be estimated. In this sub-section, we investigate this possibility by comparing composite Bayesian methods for unrestricted VAR-SV models using different small data sets. Table 4 presents the results for this case. In it, the Diebold-Mariano test is benchmarked against the 7 dimensional VAR-SV using the “good” variables. A comparison of Table 4 with results in Tables 2 and 3 shows that the small VAR-SVs produce forecasts which are better than the homoskedastic large VAR forecasts (and, as documented in Section B.4 the Online Appendix are typically statistically significant).

In terms of the joint ALPL for the three core variables, we are finding composite Bayesian methods to produce statistically significant forecast improvements relative to the unrestricted VAR-SVs with small data sets. The magnitude of these improvements are not as large as those we found in the comparison with the large homoskedastic VAR, but they are appreciable. If we examine the series individually and consider a broader range of forecast metrics, then we find a similar pattern as we did in the preceding sub-section. That is, composite Bayesian methods are found to perform particularly well when ALPLs and ACRPSs are used as forecast metrics. If we examine metrics based on point forecasts the benefits of our methods are much less. Indeed if we consider only RMSFEs and MAEs for  $h = 1$  then our methods are never better than the 7 dimensional VAR-SV in a statistically significant sense (although they occasionally are for  $h = 4$ ). Thus, we are finding that the extra information in the large data set is helping improve forecasts and that these benefits occur largely through obtaining better estimates of the dispersion of the predictive density.

If we compare the three small unrestricted VAR-SVs (i.e. the two 7 variable models involving “good” and “bad” variables and the 3 variable model involving only the core variables), we find that including the “good” variables improves forecast performance slightly. The deterioration in forecast performance that occurs when we switch from VAR-SV-g to VAR-SV-b may look to be small, but it is often statistically significant. For instance, the Diebold-Mariano test indicates the joint ALPL of VAR-SV-g is better than that for VAR-SV-b for  $h = 1$ , although it is not for  $h = 4$ . This shows that the choice of variables in a small VAR can be important and that, if the researcher gets this wrong, it can have consequences. An advantage of our methods

is that such a choice is not necessary.

Table 4: Individual forecast performance in terms of RMSFE, MAE, ACRPS and ALPL for the 3 core variables under 4 small models and 2 forecast horizons ( $h = 1, 4$ )

Forecasting Evaluation Using Joint ALPL for 3 Core Variables								
Horizon	$h = 1$				$h = 4$			
VAR-SV-3	5.651				4.134			
VAR-SV-g	5.602				3.998			
VAR-SV-b	5.509				3.913			
VAR-CL-ML	8.153***				6.455***			
Evaluation of GDP Growth Forecasts								
Horizon	$h = 1$				$h = 4$			
	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-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-CL-ML	0.837	0.607	0.619	3.938***	0.810	0.589	0.688	3.806***
Evaluation of Inflation Forecasts								
Horizon	$h = 1$				$h = 4$			
	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-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-CL-ML	0.584	0.401	0.463	4.178***	0.624***	0.454***	0.593	3.955***
Evaluation of Interest Rate Forecasts								
Horizon	$h = 1$				$h = 4$			
	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-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-CL-ML	0.998	0.603	0.458**	0.034***	1.314***	0.894***	1.248**	-1.309***

Note: DM statistics are based on the VAR-SV-g benchmark, and the RMSFE, MAE and ACRPS are multiplied by 100 for easy comparison. \*\*\*, \*\* and \* denote findings of statistically significant forecast improvements relative to the benchmark at the 1%, 5% and 10% level of significance, respectively.

#### 4.4 How Do Our Methods Compare to Approaches Which Have Restricted Forms of Stochastic Volatility?

Table 5 has the same format as Table 4, but provides a comparison with various methods which do have stochastic volatility, but of restricted forms. We are again finding that composite likelihood methods are forecasting better than other approaches. VAR-CL-ML has the highest joint ALPL for the three core variables of any of the approaches in Table 5 for both forecast horizons. The differences between it and the VAR-CCM2 benchmark are statistically significant. The fact that VAR-CL-ML

(which allows for time variation in  $a_t$ ) is forecasting better than VAR-CCM2 and VAR-CCM2-20 (which assume  $a_t$  to be constant) suggests that there is some time variation in  $a_t$ . Although there are other potential reasons for this finding. It is possible that the parsimony of the individual sub-models which enter VAR-CL-ML is an advantage relative to VAR-CCM2 and VAR-CCM2-20 which have to estimate high dimensional matrices of VAR coefficients. Or it is possible that the 196 variables used with VAR-CL-ML are providing more information than is available with the 20 or 7 variables data sets used with the VAR-CCM2 models.

Table 5: Joint and individual forecast performance under models with various restricted types of stochastic volatilities.

Forecasting Evaluation Using Joint ALPL for 3 Core Variables								
Horizon	$h = 1$				$h = 4$			
VAR-CCM1	6.079***				4.456***			
VAR-CCM2	5.805				4.089			
VAR-CCM1-20	5.328				3.500			
VAR-CCM2-20	5.804				4.577***			
VAR-FSV-1f	3.106				2.486			
VAR-FSV-2f	2.820				2.510			
VAR-CL-ML	8.153***				6.455***			
Evaluation of GDP Growth Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
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
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-CL-ML	0.837	0.607	0.619	3.938***	0.810	0.589	0.688	3.806***
Evaluation of Inflation Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
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
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-CL-ML	0.584	0.401	0.463	4.178***	0.624***	0.454***	0.593	3.955***
Evaluation of Interest Rate Forecasts								
Horizon	$h = 1$				$h = 4$			
	RMSFE	MAE	ACRPS	ALPL	RMSFE	MAE	ACRPS	ALPL
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
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-CL-ML	0.998	0.603	0.458	0.034***	1.314***	0.894***	1.248	-1.309***

Note: DM statistics are based on the VAR-CCM2 benchmark, and the RMSFEs, MAEs and ACRPSs are multiplied by 100 for easy comparison. \*\*\*, \*\* and \* denote findings of statistically significant forecast improvements relative to the benchmark at the 1%, 5% and 10% level of significance, respectively.

## 5 Summary and Conclusions

Large VARs are emerging as a popular tool in modern macroeconomics. Adding multivariate stochastic volatility to them has emerged as one of the unresolved challenges in the field. It arises since it is not computationally practical to carry out Bayesian estimation in large VARs with multivariate stochastic volatility. Even if computation were possible, conventional approaches can be over-parameterized when working with large data sets leading to problems with over-fitting, imprecise estimation and the need for strong prior information. In this paper, we propose the use of composite likelihood methods for meeting this challenge. These involve averaging over many smaller models. In our context, we use many small VAR-SVs thus enabling computation to be feasible even in data sets involving hundreds of variables. By working with smaller models, concerns with over-parameterization and the need for careful prior elicitation are lessened. We explore these themes in the paper. In addition, we discuss the econometric theory of composite likelihood methods drawing on conventional asymptotic results as well as the literature on prediction pools. All in all, there are strong theoretical reasons for thinking composite likelihood methods may be an attractive way of adding stochastic volatility to large VARs.

The issue of how well composite likelihood methods work in practice is explored in our empirical work. Working with a large US quarterly macroeconomic data set involving 196 variables, we find encouraging results. When we use all 196 variables and compare the forecast performance of our composite likelihood methods against the main practical alternative (a large homoskedastic VAR with natural conjugate prior), we find strong evidence of the superiority of our methods. Clearly, stochastic volatility is an important feature of this data set and our VAR-CL-SV methods allow for this.

We also compare our methods to a range of existing methods which include stochastic volatility in various ways in data sets of various dimensions. In this comparison, our composite likelihood methods also come out well. With few exceptions, they produce the best forecasts and in some cases the improvements over alternative specifications is statistically significant. Overall, we conclude that the strategy of combining forecasts from many small models is computationally feasible even with large VARs and leads to forecast performance that is better than other computationally feasible approaches.

## References

- Banbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 25, 71-92.
- Banbura, M., Giannone, D. and Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting*, 31, pp.739-756.
- Bernanke, B., Boivin, J. and Eliasziw, P. (2005). Measuring monetary policy: A Factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120, 387-422.
- Bloor, C. and Matheson, T. (2010). Analysing shock transmission in a data-rich environment: a large BVAR for New Zealand. *Empirical Economics*, 39, 537-558.
- Canova, F. and Matthes, C. (2018). A composite likelihood approach for dynamic structural models. Manuscript available at: <https://doi.org/10.21144/wp18-12>.
- Carriero, A., Clark, T. and Marcellino, M. (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics*, 34, 375-390.
- Carriero, A., Clark, T. and Marcellino, M. (2019a). Large Vector Autoregressions with stochastic volatility and flexible priors. *Journal of Econometrics*, forthcoming.
- Carriero, A., Clark, T. and Marcellino, M. (2019b). Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, forthcoming.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2010). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400-417.
- Chan, J. (2020). Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics*, 38, 68-79.
- Chan, J. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33, 509-532.
- Chan, J. and Grant, A. (2016). On the observed-data Deviance Information Criterion for volatility modeling. *Journal of Financial Econometrics*, 14, 772-802.
- Clark, T. (2011). Real-time density forecasts from BVARs with stochastic volatility. *Journal of Business and Economic Statistics*, 29, 327-341.
- Creal, D. and Tsay, R. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189, 335-345.
- D'Agostino, A., Gambetti, L. and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28, 82-101.
- Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.
- Durrett, R. (2010). *Probability: Theory and Examples*, Fourth Edition. Cambridge University Press, Cambridge.
- Elliott, G., Gargano, A. and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177, 357-373.

- Elliott, G., Gargano, A. and Timmermann, A. (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, 54, 86-110.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, 30, 1-11.
- Genest, C., Weerahandi, S., Zidek, J. (1984). Aggregating opinions through logarithmic pooling. *Theory and Decision*, 17, 61-70.
- Genest, C., McConway, K. and Schervish, M. (1986). Characterization of externally Bayesian pooling operators. *Annals of Statistics*, 14, 487-501.
- Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164, 130-141.
- Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014.) Short-term inflation projections: a Bayesian vector autoregressive approach, *International Journal of Forecasting*, 30, 635-644.
- Giannone, D., Lenza, M. and Primiceri, G. (2015). Prior selection for Vector Autoregressions, 97, 436-451.
- Hall, S. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23, 1-13.
- Horn, R. and Johnson, C. (1991). *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.
- Jarociński, M. and Maćkowiak, B. (2017). Granger causal priority and choice of variables in vector autoregressions. *Review of Economics and Statistics*, 99, 319-329.
- Kastner, G. (2019). Sparse Bayesian time-varying covariance estimation in many dimensions, *Journal of Econometrics*, 210, 98-115.
- Kastner, G. and Huber, F. (2017). Sparse Bayesian vector autoregressions in huge dimensions. manuscript available at <https://arxiv.org/abs/1704.03239>.
- Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177-203.
- Koop, G. and Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review*, 81, 115-131.
- Koop, G., Leon-Gonzalez, R. and Strachan, R. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33, 997-1017.
- Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28, 204-230.
- Kroese, D., Taimre, T. and Botev, Z. (2011). *Handbook of Monte Carlo Methods*. John Wiley and Sons, New York.
- McCracken, M. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, 34, 574-589.
- McCracken, M., Owyang, M. and Sekhposyan, T. (2016). Real-time forecasting with a large, mixed frequency Bayesian VAR, manuscript available at <http://www.tateviksekhposyan.org/>.

Pakel, C., Shephard, N., Sheppard, K. and Engle, R. (2014). Fitting vast dimensional time-varying covariance models, available at <http://staff.bilkent.edu.tr/cavit/research/>.

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

Qu, Z. (2016). A composite likelihood approach to analyze singular DSGE models, Boston University manuscript.

Ribatet, M., Cooley, D. and Davison, A. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 813-845.

Roche, A. (2016). Composite Bayesian inference, working paper available at <https://arxiv.org/abs/1512.07678>.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21, 5-42.