Estimation of large dimensional time varying VARs using copulas

Mike G. Tsionas[†] **Marwan Izzeldin**[†*] **Lorenzo Trapani**[*]

Abstract:

This paper provides a simple, yet reliable, alternative to the (Bayesian) estimation of large multivariate VARs with time variation in the conditional mean equations and/or in the covariance structure. The original multivariate, n-dimensional model is treated as a set of n univariate estimation problems, and cross-dependence is handled through the use of a copula. This makes it possible to run the estimation of each univariate equation in parallel. Thus, only univariate distribution functions are needed when estimating the individual equations, which are often available in closed form, and easy to handle with MCMC (or other techniques). Thereafter, the individual posteriors are combined with the copula, so obtaining a joint posterior which can be easily resampled. We illustrate our approach using various examples of large time-varying parameter VARs with 129 and even 215 macroeconomic variables.

Keywords and phrases: Vector AutoRegression; Time-Varying parameters; Heteroskedasticity; Copulas.

JEL codes: C11, C13.

1. Introduction

Following the seminal contributions by Sims (1980) and Litterman (1986), Vector AutoRegression (VAR) models and their variants are now widely applied to multivariate time series, as a flexible alternative to structural models. There is now a huge body of literature on both theory and applications: useful surveys are provided *inter alia* by Watson (1994) and Lütkepohl (2005). Although in their standard form VARs already offer a relatively flexible modelling approach, extensions have been considered to accommodate time variation. This may occur in the coefficients of the conditional mean equations (see e.g. Doan et al., 1984; Canova, 1993; Sims, 1993; Stock and Watson, 1996; and Cogley and Sargent, 2001), as a flexible alternative to models with abrupt breaks. These are termed Time-Varying Parameter VAR (TVP-VAR) models. Time variation is also considered in the covariance matrix of the error term, thereby allowing for time-varying heteroskedasticity. Following seminal papers by Uhlig (1997), Cogley and Sargent (2001) and Primiceri (2005), recent examples where the assumption of homoskedasticity is relaxed include Koop and Korobilis (2013) and Koop et al. (2019), who attempt to reduce the dimensionality issue by imposing a factor structure onto the volatilities; see also Clark (2011), Carriero et al. (2015), Clark and Ravazzolo (2015) and Carriero et al. (2016). In a recent landmark paper, Carriero et al. (2019) propose a less restrictive set-up, which allows for inference in the presence of heteroskedasticity.

Across the extensive literature on multivariate models, virtually all studies have two issues in common: the dimensionality of the model and its computational burden. On the one hand, unless the number of variables involved in the model is relatively large, omitted variable bias may impair the forecasting ability of the model (see Giannone and Reichlin, 2006). Carriero et al. (2019) make a compelling case for the superior performance of large dimensional VARs. On the other hand, computational difficulties may arise when there are a large number of variables and, more substantially, over-parameterisation can occur. Thus, the literature has focused on finding techniques that allow for the estimation of large VARs: see Bańbura et al. (2010) for an excellent review of a variety of approaches. In the case of homoskedastic VARs, the dimensionality issue can be handled by the choice of appropriate (conjugate) prior distributions, as shown by Bańbura et al. (2010) who apply their technique to the estimation of a VAR with 130 variables. However, in the case of a heteroskedastic VAR, the computational burden cannot be resolved through the choice of an appropriate prior. As explained in Carriero et al. (2019), heteroskedasticity invalidates the so-called "symmetry"across

equations that characterises homoskedastic VARs. By that property, a homoskedastic VAR can be thought of as a SUR model with the same the regressors across all equations; this implies that the covariance matrix of the VAR coefficients has a Kronecker structure, which makes estimation simpler than having to deal with large matrices with no simplifying structure. Few contributions consider estimation of a large VAR with time variation in the coefficients of both the conditional mean equations and of the covariance matrix of the error term. Koop and Korobilis (2013) and Koop et al. (2019) propose an estimation technique for large, possibly heteroskedastic TVP-VARs, which essentially relies on the Kalman filter. However, their approach is not fully Bayesian, and it is liable to mis-specification issues if the assumed model for coefficient variation is not correct (see also recent contributions by Kapetanios et al., 2019 and Chan, 2020 which, through a nonparametric approach and a reduced-rank regression approach respectively, provide a solution). To address this issue, Carriero et al. (2019) propose a new, equation-by-equation estimation algorithm which performs very well in out-of-sample forecasting and also in structural analysis. However, their paper does not consider the presence of time varying coefficients in the VAR specification.

Proposed methodology and main contribution of this paper

This paper proposes a copula-based Bayesian estimation methodology for large TVP-VARs with heteroskedasticity, which is suitable for forecasting even with very large VARs (see Section 4). Full details are in Section 2. Here, we give a heuristic preview of our methodology.
Given a multivariate model of (possibly) very large dimension n, we reduce it into n univariate models, which are then separately estimated. In order to recover the cross-dependence among equations, we use a (mixture of normals) copula-like term. In consequence, the likelihood function in our system factors into the likelihoods of the individual autoregressive models, plus the likelihood of the copula term. By this design, there is a balance between two competing needs: (i) greater computational speed (by treating each univariate estimation problem separately); and (ii) greater forecasting accuracy (by exploiting information contained in the cross-equation dependence, through the copula).

In allowing for time variation in the VAR equations, and in breaking down the multivariate estimation problem into separate univariate problems, our approach builds on earlier contributions by: Koop et al. (2009), Feldkircher et al. (2017) and Bitto and Frühwirth-Schnatter (2019) (especially for the time-varying, non-standard VAR part); Chan (2020), Lopes et al. (2016) and Huber et al. (2019)

(especially for the "divide and conquer"approach), *inter alia*; and Korobilis and Pettenuzzo (2019) and Carriero et al. (2019) (who also exploit the equation-by-equation estimation approach). However, in contrast to the last two papers, in our approach we allow for full time variation, in the conditional mean and in the variance equations. Using copulas to model dependence in a Bayesian context has been shown to be effective in several contributions (see especially Gruber and Czado, 2015, Gruber and Czado, 2018, Kreuzer and Czado, 2019 and Kreuzer et al., 2019), including non-parametric Bayesian analysis (see *inter alia* the contributions of Rodriguez and ter Horst (2008), Taddy (2010), Nieto-Barajas et al. (2012), Di Lucca et al. (2013), Bassetti et al. (2014) and Nieto-Barajas and Quintana (2016)).

Our approach also allows for greater flexibility in the specification of the univariate models, without dramatically affecting the forecasting accuracy. For example, in the simplest application of our methodology, each series is modelled as an AR(1) model. However, to allow more sophisticated model selection, and inspired by the results in Koop et al. (2019), we also exploit Bayesian compression methods in order to induce sparsity in the univariate equations (see Guhaniyogi and Dunson, 2015). Other methods for dimensionality reduction could also be applied to the univariate equations - we refer to Huber et al. (2020) for a thorough treatment of sparsity-inducing methods in the context of TVP-VARs.

Summary of main findings

Our methodology delivers superior predictive ability and reduced computational costs compared to existing approaches, and: (i) the information loss in using the copula is lower than that of equation-by-equation computations; (ii) our dimensionality reduction approach addresses the increase in the size of the parameter space due to the copula (Section 2.2.1); and (iii) the use of Bayesian compression reduces the dimensionality of the univariate equations.

Our empirical results indicate that forecasting gains are attributable to use of the copula. Compared to the use of an AR specification in the univariate equations, Bayesian compression (see Section A.2) may yield better results, but those improvements are of second order when set against forecasting gains with the copula. When compared to equation-by-equation estimation methods, such as Carriero et al. (2019), the improved computational speed is jointly the result of our copula dimension reduction strategies, and of the use of Bayesian compression in the univariate equations. Thus, dimension reduction techniques seem a valid alternative to shrinkage priors.

The remainder of the paper is organised as follows. Our methodology is spelt out in Section 2. The main empirical applications are in Section 3 (further applications and analysis are also in Section 4 and in the appendix). Section 5 concludes. Further results and technical details are reported in Appendix. [1]

2. Methodology

We consider the TVP-VAR(p)

$$\mathbf{y}_t = \sum_{j=1}^{p} A_{t,j}\mathbf{y}_{t-j} + \mathbf{u}_t, \quad p+1 \leq t \leq T, \tag{2.1}$$

where $\mathbf{y}_t$ is an $n \times 1$ vector and $\mathbf{u}_t$ is a zero mean, Gaussian process with possibly time varying variance - we discuss the specification of the second moment later on. Model (2.1) could be extended to consider e.g. exogenous regressors, latent regressors such as common factors, or deterministics such as a constant, (linear or nonlinear) trends and seasonal dummies. Moreover, (2.1) could also have an MA(q) structure, in the spirit of Chan and Eisenstat (2013); or it could have no autoregressive structure at all, and only time varying heteroskedasticity as in the case of Creal and Tsay (2015). We prefer to focus on a simpler specification, so that the discussion is not overshadowed by the algebra. Similarly, the assumption that $\mathbf{u}_t$ is Gaussian is made only for simplicity. Nevertheless, even with this simple set-up, the number of parameters grows rapidly with p and n.

2.1. Theory: the univariate equations, the copula and the likelihood function

Univariate equations

In the context of (2.1), we consider two possible models.

As a benchmark, it is possible to employ the following univariate AR(p) specifications

$$y_{i,t} = \sum_{j=1}^{p} \beta_{i,t,j} y_{i,t-j} + u_{i,t} = \beta_{i,t}^{(1)\prime} z_{i,t}^{(1)} + u_{i,t}, \tag{2.2}$$

[1] All code and datasets are available from the authors.

for $1 \leq i \leq n$ with $u_{i,t} = e^{h_{i,t}/2} u^*_{i,t}$, where $u^*_{i,t} \sim i.i.d.N(0,1)$ and

$$h_{i,t} = \alpha_i + \gamma_i h_{i,t-1} + e_{i,t}, \tag{2.3}$$

with $e_{i,t} \sim i.i.d.N(0, \delta_i)$. As noted above, (2.2) can be extended and/or modified to incorporate e.g. a different number of lags p_i for each unit, an MA(q_i) component, exogenous regressors, deterministics, (conditional or unconditional) heteroskedasticity, etc.. Similarly, (2.3) could be replaced e.g. by a GARCH specification to allow for conditional heteroskedasticity (see also the discussion in Uhlig, 1997 on the relative merits of possible specifications for time heteroskedasticity).

In (2.2) $y_{i,t}$ is predicted using only its own past vales. As previously mentioned, this is meant to be a benchmark, although results in Section A.2 show that predictive ability is preserved when using simple AR(p_i) specifications. An alternative approach is based on the Bayesian compression algorithm developed in Guhaniyogi and Dunson (2015). We consider the specification

$$y_{i,t} = \beta^{(2)\prime}_{i,t} z^{(2)}_{i,t} + u_{i,t}, \tag{2.4}$$

with $u_{i,t}$ still satisfying (2.3). As in (2.2), $z^{(2)}_{i,t}$ is a subset of the regressors in each equation of the unrestricted VAR (say $\widetilde{z}_{i,t}$). However, in the case of (2.4), the vector $z^{(2)}_{i,t}$ can include lags of $y_{i,t}$ and also lags of $y_{j,t}$ for $j \neq i$. In order to select the components of $z^{(2)}_{i,t}$, Guhaniyogi and Dunson (2015) suggest the following technique. Let $z^{(2)}_{i,t} = \Phi \widetilde{z}_{i,t}$ with Φ a $p \times np$ matrix whose entries are defined as

$$\Phi_{ij} = \begin{cases} -\phi^{-1/2} & \phi^2 \\ 0 \quad \text{with probability} & 2\phi(1-\phi) \\ \phi^{-1/2} & (1-\phi)^2 \end{cases},$$

and ϕ and p are drawn uniformly from $(0.1, 1)$ and $\{1, ..., p^{\max}\}$, with $p^{\max}$ chosen so that the marginal likelihood has a global peak. The matrix Φ is then normalised via the Gram-Schmidt orthonormalisation. [2] [3]

[2] We refer to Guhaniyogi and Dunson (2015) for details; in all our applications, the number of variables which is selected in each equation, with this method, is around 10-20%. Heuristically, the low percentage of variables not set to zero corresponds to making the dimensionality of each equation linear, as opposed to quadratic, in n.

[3] As we make clear in the following, the use of copulas is computationally practicable with the small numbers of RHS variables produced by random compression. Conversely, using traditional Bayesian shrinkage priors as a substitute for sparsification may, in large dimensions, result in loss of computational speed.

Copula

We now introduce the copula term to model dependence among the univariate equations. Let X denote a continuous k-dimensional random variable whose density is given by $f(x)$. Define $I = [0,1]$, and consider the function $c^* : I^k \to I$, with the following properties:

- $c^*(v_1, ..., v_k) = 0$ if $v_j = 0$ for at least one $1 \leq j \leq k$, and $c^*(v_1, ..., v_k) = v_i$ if $v_j = 1$ for all $j \neq i$;
- $c^*(v_1, ..., v_k)$ is non-decreasing.

Then it holds that

$$f(x) = \left(\prod_{j=1}^{k} f_j(x_j) \right) c^*(v_1, ..., v_k), \tag{2.5}$$

where $f_j(x_j)$ is the density of the j-th coordinate of X, $v_j = F_j(x_j) = \int_{-\infty}^{x_j} f_j(u)\, du$, and $c^*(v_1, ..., v_k)$ is the copula density. The result implied by (2.5) can be equivalently re-written as

$$c^*(v_1, ..., v_k) = P\left(X_1 \leq F_1^{-1}(v_1), ..., X_k \leq F_k^{-1}(v_k)\right), \tag{2.6}$$

This result is known as Sklar's theorem (see Sklar, 1996). A comprehensive summary of the properties of copulas goes beyond the scope of this paper, but Nelsen (2007) offers a comprehensive introduction. According to Sklar's theorem, every valid distribution can be factorised as the product of the individual, univariate marginals, and a copula - the copula itself is, mathematically, a density. Further, Sklar's theorem states that, whenever X is continuous, the copula function is unique. [4] Equation (2.5) can equivalently be written as

$$\ln f(x) = \sum_{j=1}^{k} \ln f_j(x_j) + \ln c^*(v_1, ..., v_k). \tag{2.7}$$

Equation (2.7) illustrates the source of the computational gains: the fact that the joint distribution can be expressed as the product of the marginals and the copula entails that the log-likelihood takes an additive form, where each set of parameters is separate from the others. This allows for parallelisation, which in turn is the reason underpinning the improved computational speed.

Likelihood function

[4] We note that the uniqueness property is in general not true in the case of discrete random variables.

Henceforth, we use $z_{i,t}$ as short-hand for both $z_{i,t}^{(1)}$ and $z_{i,t}^{(2)}$; $\beta_{i,t}$ for both $\beta_{i,t}^{(1)}$ and $\beta_{i,t}^{(2)}$ in (2.2) and (2.4) respectively. We assume the following law of motion

$$\beta_{i,t} = A_{\beta,i}\beta_{i,t-1} + \epsilon_{i,t}, \tag{2.8}$$

with $\epsilon_{i,t} \sim i.i.d.N(0, \Sigma_i)$, independent across i. In (2.8), we do not impose the typical random walk model for the time-varying parameters (see e.g. Koop and Korobilis, 2013), which makes our set-up more general. For simplicity, we do not allow for time variation in any other parameter (i.e., we do not allow for time variation in the copula parameters, or the coefficients in (2.3)).

Let $b_i = (\alpha_i, \gamma_i, \delta_i)$. Then, the marginal density of $y_{i,t}$ conditional on $z_{i,t}$ can be denoted as $f_i(y_{i,t}|z_{i,t}; \beta_{i,t}, b_i)$ (we omit dependence on $A_{\beta,i}$, Σ_i and $\beta_{i,0}$ for short). Then, by (2.5), it holds that

$$f_i(\mathbf{y}_t|z_t; \beta_{i,t}, b_i) = \left(\prod_{i=1}^{n} f_i(y_{i,t}|z_{i,t}; \beta_{i,t}, b_i)\right) c^*(v_{1,t}, ..., v_{n,t}), \tag{2.9}$$

having defined $v_{i,t} = \int_{-\infty}^{y_{i,t}} f_i(u|z_{i,t}; \beta_{i,t}, b_i)\, du$, with $f_i(u|z_{i,t}; \beta_{i,t}, b_i)$ denoting the density of $y_{i,t}$ conditional on $z_{i,t}$.

The choice of the copula

Sklar's theorem ensures that the copula density $c^*(v_{1,t}, ..., v_{n,t})$ is unique, at least for continuous random variables. Although (2.6) offers an implicit method to construct the copula function c^*, this requires knowledge of the distributions of the marginals. From a practical point of view, this entails that no explicit expression for the copula function is provided.

One possibility would be to consider a non-parametric copula, and we refer to Scaillet and Fermanian (2003), Ibragimov (2005) and Chen and Huang (2007), and the references therein, for the relevant theory in a time series context. In this paper, we consider a Gaussian mixture copula model (GMCM henceforth; see Tewari et al., 2011) model, viz.

$$c^*(v_{1,t}, ..., v_{n,t}) = \frac{\sum_{g=1}^{G} p_g f_N\left(\Psi_1^{-1}(v_{1,t}), ..., \Psi_n^{-1}(v_{n,t}) | \mu_g, \Omega_g\right)}{\prod_{i=1}^{n} \phi_i\left(\Psi_i^{-1}(v_{i,t})\right)} \equiv c(\mathbf{y}_t|\alpha), \tag{2.10}$$

where

$$\phi_i = \text{ marginal pdf of } \sum_{g=1}^{G} p_g f_N\left(\mathbf{y}_t|\mu_g,\Omega_g\right),\ 1\leq i\leq n,$$
$$\Psi_i = \text{ marginal cdf of } \sum_{g=1}^{G} p_g f_N\left(\mathbf{y}_t|\mu_g,\Omega_g\right),\ 1\leq i\leq n,$$

and $\{p_g\}_{g=1}^{G}$ (such that $\sum_{g=1}^{G} p_g = 1$ and $p_1 < ... < p_G$) is a set of weights and $f_N\left(\cdot|\mu_g,\Omega_g\right)$ is the density of an n-variate Gaussian random variable with mean μ_g and covariance matrix Ω_g. In (2.10), we use the short-hand notation $\alpha = \left(\left(p_1,...,p_G\right)',\mu_1',...,\mu_G',\left(vech\left(\Omega_1\right)\right)',...,\left(vech\left(\Omega_G\right)\right)'\right)'$.

Finally, let $\beta_t = \left(\beta_{1,t}',...,\beta_{n,t}'\right)'$, $\omega = \left(b_1',...,b_n'\right)'$, $A_\beta = \{A_{\beta,1},...,A_{\beta,n}\}$, $\Sigma = \{\Sigma_1,...,\Sigma_n\}$, and, for short, $\theta = \left(vec\left(\alpha\right)',\omega',vec\left(A_\beta\right)',vech\left(\Sigma^{1/2}\right)',vec\left(\beta_0\right)'\right)'$.
Putting everything together, the resulting likelihood function (conditional on the initial observations $\{\mathbf{y}_t\}_{t=1}^{p}$) is given by

$$\begin{aligned} L\left(\{\mathbf{y}_t\}_{t=p+1}^{T}|\alpha,\omega,A_\beta,\Sigma,\beta_0,\{\beta_t\}_{t=1}^{T}\right) &= L\left(\{\mathbf{y}_t\}_{t=p+1}^{T}|\theta,\{\beta_t\}_{t=1}^{T}\right) \\ &= \prod_{t=p+1}^{T}\left[\prod_{i=1}^{n} f_i\left(y_{i,t}|z_{i,t};\beta_{i,t},A_{\beta,i},\Sigma_i,\beta_{i,0},b_i\right)\right] c\left(\mathbf{y}_t|\alpha\right), \end{aligned} \tag{2.11}$$

where we emphasize the dependence of the marginal densities on $A_{\beta,i}$, Σ_i and $\beta_{i,0}$. It follows that

$$\ln L\left(\{\mathbf{y}_t\}_{t=p+1}^{T}|\theta,\{\beta_t\}_{t=1}^{T}\right) = \sum_{t=p+1}^{T}\ln c\left(\mathbf{y}_t|\alpha\right) + \sum_{i=1}^{n}\sum_{t=p+1}^{T}\ln f_i\left(y_{i,t}|z_{i,t};\beta_{i,t},A_{\beta,i},\Sigma_i,\beta_{i,0},b_i\right). \tag{2.12}$$

2.2. Dimension reduction and estimation

We choose the prior

$$p\left(\theta\right) = p\left(\alpha\right)\prod_{i=1}^{n} p\left(\omega_{i,0}\right)p\left(A_{\beta,i}\right)p\left(\Sigma_i\right)p\left(\beta_{i,0}\right), \tag{2.13}$$

where: $p\left(\alpha\right)$ and $p\left(\omega_{i,0}\right)$ are flat priors (in the latter, coefficients are restricted to be non-negative); $p\left(\Sigma_i\right) \propto \left|\Sigma_i\right|^{-(n+1)/2}$ is a standard non-informative prior; $p\left(A_{\beta,i}\right)$ and $p\left(\beta_{i,0}\right)$ are Gaussian priors.

Hence, the posterior is given by

$$p\left(\theta,\{\beta_t\}_{t=1}^T \mid \{\mathbf{y}_t\}_{t=1}^T\right) \propto L\left(\{\mathbf{y}_t\}_{t=p+1}^T |\theta,\{\beta_t\}_{t=1}^T\right) p\left(\{\beta_t\}_{t=1}^T |\theta\right) p\left(\theta\right). \tag{2.14}$$

Despite the presence of the copula term, the number of parameters in θ is proportional to n^2, which does not fully resolve the challenge represented by dimensionality in a large VAR. More specifically, from (2.10), it is apparent that, when estimating μ_g, the number of parameters to be estimated is Gn; conversely, the covariance matrices Ω_g contain each $\frac{n(n+1)}{2}$ elements and this is where the dimensionality issue arises from. To address this problem, in Section 2.2.1 we consider two ways of restricting the Ω_gs: both reduce the number of free parameters in the copula, making this proportional to n instead of n^2.

Univariate regressions estimation

Both (2.2) and (2.4) are regressions with time varying parameters and stochastic volatility. Thus, we use the approach by Kim et al. (1998) to estimate $\beta_{i,t}$ and b_i (and the other hyperparameters). [5]

Copula density estimation

As is typical with copula models, we first obtain an estimate of the univariate densities

$$f_i\left(y_{i,t}|z_{i,t};A_{\beta,i},\Sigma_i,\beta_{i,0},b_i,\beta_{i,t}\right).$$

We then obtain the probability integral transforms, $v_{i,t}$, and use these as data to estimate α. [6]

2.2.1. Copula dimension reduction strategies - α

The dimensionality issue can be further addressed by imposing some restrictions on α. We discuss two possible approaches (denoted as *S1* and *S2*), where the priors employed are flat.

[5] Further details are in Section A.1.2

[6] This procedure can be viewed as "two-stage" Bayesian, as opposed to a "full-information" Bayesian estimator (see also Creal and Tsay, 2015). This could be carried out by modifying the MCMC algorithm; however, we have tried to use it in some of our empirical applications, but results were - if anything - marginally worse than with the two-step procedure proposed here.

Dimension reduction: strategy S1

The first dimension reduction strategy is based on a recursive model for the Ω_gs:

$$\Omega_g = h_g \Omega_{g-1} + V_g, \quad 2 \leq g \leq G, \tag{2.15}$$

having initialised (2.15) by leaving Ω_1 unrestricted (and thus setting $V_1 = 0$). In (2.15), h_g is a scalar to be estimated, and the idiosyncratic shock V_g is restricted to $V_g = diag\{v_{g,1}, ..., v_{g,n}\}$.
Consequently, the number of parameters associated to the copula is $(G-1)(n+1)$, which grows linearly (instead of quadratically) with n.

Dimension reduction: strategy S2

The second dimension reduction approach is closely related to Principal Components (see Humphreys et al., 2015), and to the Bayesian compression literature (Guhaniyogi and Dunson, 2015). We again leave Ω_1 unrestricted, and model the Ω_gs, for $2 \leq g \leq G$, as

$$\Omega_g = Q_{0,g} Q_{0,g}' + D_g. \tag{2.16}$$

In (2.16), $D_g = diag\{d_{g,1}, ..., d_{g,n}\}$ and $Q_{0,g}$ is an $n \times k$ matrix.
We make no attempt to estimate $Q_{0,g}$. Instead, we randomly generate the elements of $Q_{0,g}$, say $\{Q_{0,g}\}_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq k$, as independent of each other with $\{Q_{0,g}\}_{i,j} \sim N\left(0, q_g^2\right)$ for a total of $1,000,000$ iterations, choosing the specification which maximises the log marginal likelihood. Thus, the only parameters that need to be estimated are q_g and $\{d_{g,1}, ..., d_{g,n}\}$. Under these restrictions, the number of parameters is $(G-1)(n+1)$, which is the same as in *S1*.

3. Empirical application: forecasting large VARs

We conduct a forecasting horserace to compare the performance of our methodology against several alternative methodologies (Koop et al., 2019, and Carriero et al., 2019). [7]

[7] While we were revising our paper, it was brought to our attention by two anonymous Referees that the algorithm in Carriero et al. (2019) contains a (minor) error. Carriero et al. (2021) propose a (computationally marginally more intense) correction; however, when applying the correct algorithm to their original data, the authors show that results do not change much, indicating that the algorithm in Carriero et al. (2019) still provides - at least with our dataset and model - a sufficiently good approximation.

3.1. Forecasting performances

In order to ensure a fair comparison, we use the same datasets as in those studies: a set of macroeconomic variables from the FRED-MD dataset from January 1960 to December 2014. We use the same $n = 129$ and $n = 125$ variables as Koop et al. (2019) and Carriero et al. (2019), respectively. A full description of the data is provided in McCracken and Ng (2016). We carry out the same forecasting exercise as the authors, predicting seven variables of interest: total nonfarm employment (PAYEMS), consumer price inflation (CPIAUCSL), the change in the Fed funds rate (FEDFUNDS), industrial production growth (INDPRO), the unemployment rate (UNRATE), the finished good producer price inflation (PPIFGS), and the change in the 10 year T-bill rate (GS10). We also undertake the same exercise for a smaller VAR, using the same $n = 19$ variables as in the original study. To replicate the exercises in Koop et al. (2019) and Carriero et al. (2019) as closely as possible, all variables are transformed as recommended by the authors using their code (see the original paper for details). All models considered have the same specification: we use 13 lags and no deterministics, except for a constant. We consider three scenarios: the baseline case of a simple, homoskedastic and non time-varying VAR; a TVP-VAR with heteroskedasticity; and, to allow for a direct comparison with Carriero et al. (2019), a fixed coefficients, heteroskedastic VAR.

3.1.1. Forecasting with homoskedastic, fixed coefficient VARs

In Tables A and B, we evaluate the relative forecasting ability of a small ($n = 19$) and a medium-sized ($n = 129$) VAR. As far as Koop et al. (2019) is concerned, we consider the best prediction for each variable and for each forecasting horizon, so that the model employed may differ from variable to variable and across different values of h. The MSFEs reported in both tables are the result of using our methodology with Bayesian compression applied to the univariate equations, and strategy *S2* for dimension reduction in the GMCM copula. This model has been used for each variable and each forecasting horizon h.

[Insert Tables A and B somewhere here]

The results show that our methodology always improves on the MSFE, for all variables and at all forecasting horizons. All of our MSFE ratios are below 1, so indicating that the benchmark AR(1)

model is always beaten, even for the finished good producer price inflation (PPIFGS), and the change in the 10 year T-bill rate (GS10), for which the predictions obtained by Koop et al. (2019) almost always underperformed the benchmark. We have also run the forecasting exercise using different combinations of dimension reduction strategy for the univariate models (i.e., replacing the Bayesian compression algorithm with univariate AR(p_i) models), and of copula dimension reduction (i.e. applying strategy *S1* to the GMCM copula). Our core finding is that the MSFEs are very similar to the ones in Tables A and B. Thus, the general conclusion is that our new approach, based on approximating the VAR with n univariate models and linking these together through a copula, delivers superior predictive ability. Further improvements may stem from alternative specifications within our methodology.

In Table C, we report the average log predictive likelihood (ALPL) as a measure of out-of-sample density forecast performance, as suggested by Geweke and Amisano (2010) - as far as Koop et al. (2019) is concerned, once again we report the best results from their paper. [8] Finally, in Table D we report results for multivariate forecasts for the cases $n = 19$ and $n = 129$. Results reinforce the findings above: in all cases, our methodology offers superior forecasting ability with respect to the best results in Koop et al. (2019) and Carriero et al. (2019). As with the findings in Tables A and B, in Table C all ALPL differentials are always positive, again indicating that the benchmark is always beaten even in terms of this measure of forecasting ability.

[Insert Tables C and D somewhere here]

3.1.2. TVP-VARs with heteroskedasticity

In another set of exercises, we assess the (relative) forecasting performance - again in terms of MSFE and ALPL - when using a TVP-VAR with heteroskedasticity. The other specifications remain the same: we use 13 lags and no deterministics except for a constant; we employ a Bayesian compression algorithm applied to both conditional mean and variance, and strategy *S2* to reduce dimensionality in the GMCM copula. To gauge the extent of sparsity induced by this approach, in each univariate equation, around $10 - 20\%$ of the variables are kept. As far as Koop et al. (2019) is concerned, we draw a comparison with the forecasting performance of the Bayesian compressed VAR (denoted as

[8] These correspond to Table 4 in their paper

BCVAR_{tvp-sv}). [9]

[Insert Table E somewhere here]

Results in Table E contain the same message as the previous tables. *Ceteris paribus* (i.e., given time variation, Bayesian compression, and the same specification), results are always better (indeed, with one exception, the ALPL for CPIAUCSL at $h = 3$) when using our copula-based approach, both at a univariate, single-variable prediction level and at a multivariate level. As in the previous exercises, the relative MSFEs are always below 1, and the ALPL differentials always positive. Note also that adding time variation may improve forecasting ability: when considering point forecasts, results are qualitatively similar to those in Koop et al. (2019) and Carriero et al. (2019).

3.1.3. Heteroskedastic, fixed coefficients VARs

In order to elicit a direct comparison with Carriero et al. (2019), we consider time variation only in the error covariance matrix. In Table F, we report the MSFE and the ALPL for the multivariate forecasts when considering a VAR with heteroskedasticity and no other time variation otherwise - the other specifications are the same as before (namely, we have used 13 lags and constant only). As before, we compare against the best available results in the papers by Koop et al. (2019) and Carriero et al. (2019), with such results being always improved upon.

[Insert Table F somewhere here]

3.2. Prior sensitivity and MCMC technical details

In this section, we discuss the performance of our methodology by considering the sensitivity to the prior specification, and the convergence properties of the MCMC algorithm.

3.2.1. Prior sensitivity

We explore the sensitivity of our methodology to the choice of the (main) priors on $A_{\beta,i}$ and $\beta_{i,0}$. The main focus is the copula-based dimensionality reduction; thus, we propose flat priors in gen-

[9] These are taken from Table 6 in the original paper.

eral, although some parameters undergo nonlinear transformations (see Jeffreys, 1998 for an early treatment of the issue).

For each element in the vector $\left(vec\left(A_{\beta,i}\right)', \beta_{i,0}'\right)'$, we use the prior $N(\overline{b}, \overline{s}_b^2)$, independent across elements. In the context of (2.10), we use both dimension reduction strategies *S1* and *S2*, with:

$$p_g = \frac{e^{-r_g^2}}{\sum_{g=1}^{G} e^{-r_g^2}},$$

where

$$p\left(r_g\right) = N(\overline{r}, \overline{s}_r^2), \tag{3.1}$$

and

$$p\left(\mu_g\right) = N(\overline{m}_g, \overline{s}_m^2), \tag{3.2}$$

again independent for $1 \leq g \leq G$. Finally, in (2.16), we have used $\Omega_g = C_g C_g'$, with :

$$p\left(c_g\right) = N(\overline{c}, \overline{s}_c^2), \tag{3.3}$$

for $2 \leq g \leq G$, where we have defined $c_g = vech\left(C_g\right)$. We have set the priors parameters as follows:

$$\begin{aligned} \overline{b} = 0, \overline{s}_b^2 = 10, \\ \overline{r} = 0, \overline{s}_r^2 = 100, \\ \overline{m}_g = 0, \overline{s}_m^2 = 10, \\ \overline{c} = 0, \overline{s}_c^2 = 100. \end{aligned} \tag{3.4}$$

In Figures 1-3, we assess the sensitivity to our choice of priors by comparing our methodology against the best results in Koop et al. (2019) and Carriero et al. (2019). Our results have been obtained using strategy *S2*; unreported results using *S1* are very similar.

We have used the normal prior described above as a benchmark, and a prior based on using Cauchy distributions, i.e. with tails considerably fatter than the normal distribution. We employ Bayes factors (BF henceforth; see Kass and Raftery, 1995) as an indicator. We consider three scenarios: a baseline, fixed coefficient, homoskedastic VAR (see Figure 1); a homoskedastic TVP-VAR (see Figure 2); and a heteroskedastic TVP-VAR (see Figure 3). We have used $1,000$ different priors by sampling randomly from (3.1)-(3.3), given the parameters defined in (3.4), and doing the same with

Cauchy distributions. For each prior, we have used MCMC sampling, employing $10,000$ iterations starting from the posterior moments delivered by the benchmark prior.

[Insert Figures 1-3 somewhere here]

Results indicate (in all cases) minimal discrepancies between the relative forecasting performance obtained using either prior distribution. All BFs are very high (i.e. higher than the typically employed threshold of 100), which reinforces the conclusion that our methodology, irrespective of the prior, offers superior forecasting ability. We have also carried out the same sensitivity analysis for the empirical exercises discussed in Section A.3 - the results, available upon request, confirm the robustness to the choice of the prior distribution.

3.2.2. MCMC convergence diagnostics and computational efficiency

We evaluate the numerical efficiency of our algorithm, described in Section A.1.2, in Figure 4. We note that, in Figure 4, we have used only the $n = 125$ variables used in Carriero et al. (2019); using the full $n = 129$ variables employed in Koop et al. (2019) does not result, as can be expected, in virtually any changes.

[Insert Figure 4 somewhere here]

Results show that our MCMC algorithm is quite efficient. Specifically, in sub-panel *(a)-(d)* of each panel of Figure 4 we use Relative Numerical Efficiency (RNE) to monitor convergence. The RNE is clustered around 0.75, and never below 0.6 - this corresponds to an Inefficiency Factor which, in the worst case scenario, is below 2, which indicates that the MCMC algorithm requires less than twice the number of *i.i.d.* draws to produce the same information content. This indicates that the chain mixes (very) well; sub-panels *(b)* and *(e)* show that autocorrelations between MCMC samples decline quickly, and the convergence diagnostics reported in sub-panels *(c)* and *(f)* reinforce this finding.

In order to gauge the computational efficiency, we compare the CPU timings of our methodology with the one developed in Carriero et al. (2019) as a function of the VAR dimension n. [10]

[Insert Figure 11 somewhere here]

Figure 11 indicates that our procedure is faster for all values of n, with relative computational speed growing with n. This result can be explained by the sparsification of the univariate equations (achieved e.g. through Bayesian compression), and the copula dimension reduction strategies developed in Section 2.2.1.

4. Further evidence and applications

4.1. Extension to VARMA specifications

Chan et al. (2016) make a compelling case for the use of VARMAs, in light of their superior predictive ability (see also the theory in Lütkepohl and Poskitt, 1996). Yet, VARMAs, as well as suffering from well-known identification issues (see e.g. the recent contribution by Gouriéroux et al., 2019) are liable to overparameterisation, and therefore dimensionality, in this context, is a very important issue, which may explain their relative lack of popularity. [11]

Prior to presenting the application, we note that, as far as the methodology is concerned, we use exactly the same approach as described above; specifically, we use Bayesian compression and dimension reduction strategy *S2*. The only (minor) variation lies in the resampling scheme, described in Section A.1.2.

We compare the forecasting ability of a heteroskedastic TVP-VARMA against competing approaches, reporting BFs, and using a normal and a Cauchy prior as in Section 3.2.1.

[Insert Figure 12 somewhere here]

[10] In order to make the comparison fairer, in this case we used a single processor. Also, although our procedure is a two-stage one, we tried using a full-information one, but it only marginally increased computational times, also with marginally worse forecasts.

[11] In the Supplement, we consider another application, based on the paper byChan et al. (2016) applied to US macro data.

The TVP-VARMA yields superior forecasting ability, with all Bayes factors well above 100. This result holds for both small and large n. The choice of the prior is essentially uninfluential on this, similarly to what found for the case of a TVP-VAR in the Section 3.2.1.

We have also run a small scale exercise to explicitly compare the forecasting ability of a TVP-VARMA against a TVP-VAR, both with heteroskedasticity, and both estimated using our methodology based on Bayesian compression and strategy *S2* to reduce the copula dimensionality. Results are in Figures 13-14.

[Insert Figures 13-14 somewhere here]

Finally, we consider a forecasting exercise under "unusual circumstances", predicting 7 series during the Covid-19 outbreak (see also a recent paper by Lenza and Primiceri, 2020) towards the end of our sample. [12]

[Insert Table G and Figure 15 somewhere here]

4.1.1. Impulse response analysis

As a complement to our forecasting exercise, we study the response of our system to exogenous (or "non-systematic") monetary policy shocks. We base our exercise on the seminal paper by Primiceri (2005). Similarly to that paper, we investigate the responses of inflation and unemployment to a 1% shock to the monetary policy interest rate (the FEDFUNDS series). We consider both a medium and a large VAR ($n = 19$ and $n = 125$ respectively), and we use the dataset in Koop et al. (2019) used above, until September 2001. [13] We use a TVP-VAR with heteroskedasticity, carrying out estimation via Bayesian compression in the univariate equations and strategy *S2* to reduce the copula dimensionality. We generate the impulse using a simple (lower-triangular) Choleski factorisation,

[12] The are the same as in Section 3.1, namely: total nonfarm employment (PAYEMS), consumer price inflation (CPIAUCSL), the change in the Fed funds rate (FEDFUNDS), industrial production growth (INDPRO), the unemployment rate (UNRATE), the finished good producer price inflation (PPIFGS), and the change in the 10 year T-bill rate (GS10).

[13] Primiceri (2005) uses quarterly series, with sample starting in 1953Q1 till 2001Q3.

leaving the monetary policy variable last in our ordering; as in Primiceri (2005), we order inflation before unemployment, although we note that changing the ordering leaves the responses virtually unaffected. Similarly, we evaluate the impact of non-systematic monetary policy actions in three different dates: January 1975, July 1981 and January 1996. Results are in Figures 16-19.

[Insert Figures 16-19 somewhere here]

With some exceptions, results are qualitatively similar to those found in Primiceri (2005), with responses having the same shape (as expected on account of economic theory). With inflation, the same ordering of responses holds in all three periods (namely, the inflation response is lower in more recent periods, so reflecting faster response to monetary policy). In the case of the medium VAR, responses are very close to each other (discrepancies are statistically insignificant) and similar to the ones documented in Primiceri (2005); inflation is reined in, after a period of price puzzle, at essentially the same time as found in Primiceri (2005). Conversely, when using a large VAR, discrepancies become more evident, making the case for time variation. In particular, inflation in the Greenspan era exhibits virtually no price puzzle, whereas the presence of this is more apparent in the other two periods considered, confirming the findings in (*inter alia*) Balke et al. (1994), where it is argued that the price puzzle became less evident over time. Considering unemployment, even in this case the response when using a medium VAR looks similar to the one derived by Primiceri (2005), with greater discrepancies arising when using a large VAR.

4.2. Application to large dimensional VARs

We complement the findings in the previous section by investigating the performance of our estimation technique when applied to heteroskedastic TVP-VARs of very large dimension. We use the dataset in Kastner and Huber (2020) (Section 4.2.1), and we complement the empirical analysis by reporting some evidence from synthetic data and CPU times for various values of n (Section 4.2.2).

4.2.1. Forecasting large dimensional VARs

We follow Kastner and Huber (2020), and use the $n = 215$ quarterly series from the McCracken and Ng (2016) dataset, with a sample spanning the period between 1959:Q1 and 2015:Q4. To ensure a fair

comparison, data are transformed to be approximately stationary as suggested in McCracken and Ng (2016); each series is standardised to have zero mean and unit variance. We focus on forecasting the same eleven series as Kastner and Huber (2020): gross domestic product (GDPC96), industrial production (INPRO), total non-farm payroll (PAYEMS), civilian unemployment rate (UNRATE), new privately owned housing units started (HOUST), consumer price index inflation (CPIAUCSL), producer price index for finished goods inflation (PPIFGS), effective federal funds rate (FEDFUNDS), 10-year treasury constant maturity rate (GS10), U.S./U.K. exchange rate (EXUSUK), and the S&P 500 (S.P.500).

We begin by reporting an overall evaluation of forecasting ability, based on the cumulative log-predictive scores as in the original paper. Again in order to ensure a fair comparison, we have used the same approach as Kastner and Huber (2020) - namely, we have carried out the initial estimation using a sample ranging from 1959:Q3 to 1990:Q2; and we have then computed one-quarter-ahead predictive densities for the first period in the remainder of the sample (i.e. 1990:Q3), expanding the estimation sample after predicting each quarter ahead. The relative cumulated quarterly scores obtained this way are reported in Figure 20.

[Insert Figure 20 somewhere here]

In our comparison, we have used both a VAR(1) with constant only, fixed coefficients and heteroskedasticity, and a VARMA(1,1) with the same specifications for time variation. In both cases, we do not impose a factor structure on the covariance matrix, so that no choice/estimation of the number of common factors is required.

Both models yield better results than the methodology proposed in Kastner and Huber (2020), although the VARMA seems to perform better. The impact of the prior specification is, as found in Section 3.2.1, minimal. Further, more specific comparisons (both in terms of overall performance, and for individual series of interest), are in Tables H and I.

Insert Tables H and I somewhere here.

4.2.2. Evidence from synthetic data

We report the outcome of a small set of Monte Carlo experiments, based on various combinations of (n, T), using a VAR(1) specification with no time variation in the conditional mean equation. [14] We allow for different levels of sparsity in the partial autocorrelation matrix A (we refer to Kastner and Huber, 2020 for details). In the vast majority of the cases, forecasts are superior when using our approach; this is especially true in settings where sparsity is not an adequate assumption.

[Insert Table J somewhere here]

Finally, in Figure 21, we report CPU time for a single draw from the posterior as a function of the dimension of the model n. These results complement the ones in Figure 11 - as can be noted, these increase with n, but in a sublinear way, staying below 1' even when $n = 300$.

[Insert Figure 21 somewhere here]

5. Discussion and conclusions

Our paper develops an alternative methodology for the estimation of large TVP-VARs with possible heteroskedasticity. The original multivariate model is decomposed into n simpler models, whose interactions are modelled separately through a copula. Our applications show that our approach is computationally more efficient than directly estimating multivariate models, yielding excellent goodness of fit and predictive ability. [15]

Our extensive empirical analysis leads to the following conclusions. The main gains in forecasting ability come from the use of a copula to take into account the cross dependence among the univariate equations. Our first conclusion is that the copula manages to capture features of the data that the original, standard multivariate models are likely to miss. We use the GMCM copula, whose good performance in our context is in line with the conclusions of other studies (see e.g. Geweke and Keane,

[14] We use the specification

$$\mathbf{y}_t = A\mathbf{y}_{t-1} + \mathbf{u}_t;$$

in our simulations, allowing for heteroskedasticity, and using the dimension reduction strategy *S2*.

[15] Further output is reported in the Supplement, reinforcing our conclusions.

2007 and Villani et al., 2009); we also note, based also on further results in appendix, that using a data-driven dimensionality reduction for the GMCM copula works particularly well. Although it would be feasible, in principle, to use other copula specifications, this remains an area for future research.

Second order improvements in forecasting ability also come from the methodology employed to reduce the dimensionality of the individual equations. As our applications show (see especially Section A.3), it is possible to use a pure AR(1) structure in which each series is predicted using solely its own lags, with good forecasting ability: this indicates that the copula is very successful in exploiting the information coming from cross-equation dependence. However, superior results are obtained using a different model reduction strategy, based on Bayesian compression, which in this paper is proposed as the main strategy. After using Bayesian compression, our results are shown to be robust to the specification of the prior. This presents our contribution as complementing recent advances in the Bayesian analysis of large VARs, such as those developed in Bańbura et al. (2010) and Giannone et al. (2015), where - instead of sparsifying the univariate equations and subsequently using copulas - new, more sophisticated priors are proposed as a way to deal with large VARs (see also the results on prior sensitivity analysis in the appendix).

In addition to improvements in forecasting, our methodology has faster computational speed. These gains arise from sparsification (using both Bayesian compression and our proposed copula dimension reduction strategies) and from parallelisation.

Finally, our applications mainly focus on "reduced form" examples, as can be seen by the emphasis on forecasting ability. We conjecture however that, in light of its superior performance, our technique could also be employed in the context of more structural applications. This issue is currently under examination by the authors.

Acknowledgment

We are grateful to the Editor (Florin Bilbiie), an anonymous Associate Editor, and three anonymous Referees, whose comments and careful reading have helped us substantially improve this paper. We are also grateful to Luca Fanelli, to Todd Clark for bringing the corrected version of the algorithm in Carriero et al. (2019) to our attention, and to Gerry Steele for editorial suggestions and support.

References

Balke, N. S., K. M. Emery, et al. (1994). Understanding the price puzzle. *Federal Reserve Bank of Dallas Economic Review, Fourth Quarter*, 15–26.

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

Bassetti, F., R. Casarin, and F. Leisen (2014). Beta-product dependent Pitman–Yor processes for Bayesian inference. *Journal of Econometrics 180*(1), 49–72.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*(1), 387–422.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Canova, F. (1993). Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control 17*(1-2), 233–261.

Cappé, O., E. Moulines, and T. Rydén (2009). Inference in hidden Markov models. In *Proceedings of EUSFLAT conference*, pp. 14–16.

Carriero, A., J. Chan, T. Clark, and M. Marcellino (2021). Corrigendum to: Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. Technical report.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Carriero, A., T. E. Clark, and M. Marcellino (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics 34*(3), 375–390.

Carriero, A., T. E. Clark, and M. Marcellino (2019). Large Bayesian Vector Autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics 212*, 137–154.

Chan, J. C. (2020). Large bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics 38*(1), 68–79.

Chan, J. C. and E. Eisenstat (2013). Gibbs samplers for VARMA and its extensions. Technical report, Australian National University, College of Business and Economics.

Chan, J. C., E. Eisenstat, and G. Koop (2016). Large Bayesian VARMAs. *Journal of Econometrics 192*(2), 374–390.

Chen, S. X. and T.-M. Huang (2007). Nonparametric estimation of copula functions for dependence

modelling. *Canadian Journal of Statistics 35*(2), 265–282.

Christiano, L. J., M. Eichenbaum, and C. Evans (1996). The effects of monetary policy shocks: Evidence from the flow of funds. *The Review of Economics and Statistics 78*(1), 16–34.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Cogley, T. and T. J. Sargent (2001). Evolving post-World War II US inflation dynamics. *NBER macroeconomics annual 16*, 331–373.

Creal, D. D. and R. S. Tsay (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics 189*(2), 335–345.

Di Lucca, M. A., A. Guglielmi, P. Müller, and F. A. Quintana (2013). A simple class of Bayesian nonparametric autoregression models. *Bayesian Analysis (Online) 8*(1), 63.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*(1), 1–100.

Feldkircher, M., F. Huber, and G. Kastner (2017). Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs? *arXiv preprint arXiv:1711.00564*.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 1317–1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics 4*, 641–649.

Geweke, J. and G. Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230.

Geweke, J. and G. Amisano (2014). Analysis of variance for Bayesian inference. *Econometric Reviews 33*(1-4), 270–288.

Geweke, J. and M. Keane (2007). Smoothly mixing regressions. *Journal of Econometrics 138*(1), 252–290.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Giannone, D. and L. Reichlin (2006). Does information help recovering structural shocks from past

observations? *Journal of the European Economic Association 4*(2-3), 455–465.

Girolami, M. and B. Calderhead (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(2), 123–214.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2019). Identification and estimation in non-fundamental structural VARMA models. *The Review of Economic Studies*, to appear.

Gruber, L. and C. Czado (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis 10*(4), 937–963.

Gruber, L. F. and C. Czado (2018). Bayesian model selection of regular vine copulas. *Bayesian Analysis 13*(4), 1107–1131.

Guhaniyogi, R. and D. B. Dunson (2015). Bayesian compressed regression. *Journal of the American Statistical Association 110*(512), 1500–1514.

Huber, F., G. Kastner, and M. Feldkircher (2019). Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models. *Journal of Applied Econometrics 34*(5), 621–640.

Huber, F., G. Koop, and L. Onorante (2020). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics* (just-accepted), 1–48.

Humphreys, D. A., P. M. Harris, M. Rodríguez-Higuero, F. A. Mubarak, D. Zhao, and K. Ojasalo (2015). Principal component compression method for covariance matrices used for uncertainty propagation. *IEEE Transactions on Instrumentation and Measurement 64*(2), 356–365.

Ibragimov, R. (2005). Copula-based dependence characteriztions and modeling for time series. *Harvard Institute of Economic Research Discussion Paper* (2094).

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2010). Investment shocks and business cycles. *Journal of Monetary Economics 57*(2), 132–145.

Kapetanios, G., M. Marcellino, and F. Venditti (2019). Large time-varying parameter VARs: A non-parametric approach. *Journal of Applied Econometrics*, to appear.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kastner, G. and F. Huber (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting 39*(7), 1142–1165.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Koop, G. and D. Korobilis (2013). Large time-varying parameter VARs. *Journal of Econometrics 177*(2), 185–198.

Koop, G., D. Korobilis, and D. Pettenuzzo (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics 210*(1), 135–154.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control 33*(4), 997–1017.

Korobilis, D. and D. Pettenuzzo (2019). Adaptive hierarchical priors for high-dimensional vector autoregressions. *Journal of Econometrics 212*, 241–271.

Kreuzer, A. and C. Czado (2019). Bayesian inference for dynamic vine copulas in higher dimensions. *arXiv preprint arXiv:1911.00702*.

Kreuzer, A., L. D. Valle, and C. Czado (2019). Bayesian multivariate nonlinear state space copula models. *arXiv preprint arXiv:1911.00448*.

Lenza, M. and G. E. Primiceri (2020). How to estimate a VAR after March 2020. Technical report, National Bureau of Economic Research.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Lopes, H. F., R. E. McCulloch, and R. S. Tsay (2016). Parsimony inducing priors for large scale state-space models. *Bayesian Anal.*

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lütkepohl, H. and D. S. Poskitt (1996). Specification of echelon-form VARMA models. *Journal of Business & Economic Statistics 14*(1), 69–79.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nemeth, C., C. Sherlock, and P. Fearnhead (2016). Particle Metropolis-adjusted Langevin algorithms. *Biometrika 103*(3), 701–717.

Nieto-Barajas, L. E., P. Müller, Y. Ji, Y. Lu, and G. B. Mills (2012). A time-series DDP for functional proteomics profiles. *Biometrics 68*(3), 859–868.

Nieto-Barajas, L. E. and F. A. Quintana (2016). A Bayesian non-parametric dynamic AR model for multiple time series analysis. *Journal of Time Series Analysis 37*(5), 675–689.

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

Roberts, G. O. and J. S. Rosenthal (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(1), 255–268.

Rodriguez, A. and E. ter Horst (2008). Bayesian dynamic density estimation. *Bayesian Analysis 3*(2), 339–365.

Scaillet, O. and J.-D. Fermanian (2003). Nonparametric estimation of copulas for time series. *Journal of Risk* (5), 25–54.

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

Sims, C. A. (1993). A nine-variable probabilistic macroeconomic forecasting model. In *Business Cycles, Indicators and Forecasting*, pp. 179–212. University of Chicago Press.

Sklar, A. (1996). Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*, 1–14.

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

Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association 105*(492), 1403–1417.

Tewari, A., M. J. Giering, and A. Raghunathan (2011). Parametric characterization of multimodal distributions with non-gaussian modes. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 286–292. IEEE.

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

Villani, M., R. Kohn, and P. Giordani (2009). Regression density estimation using smooth adaptive Gaussian mixtures. *Journal of Econometrics 153*(2), 155–173.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of Econometrics 4*, 2843–2915.

TABLE A
Relative MSFE at various horizons h - VAR with n = 19.

Variable	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
	0.774	0.428	0.420	0.515	0.613	0.682
PAYEMS	(0.830)	(0.554)	(0.522)	(0.686)	(0.824)	(0.931)
	[0.810]	[0.950]	[0.471]	[0.612]	[0.803]	[0.782]
	0.912	0.893	0.911	0.922	0.924	0.930
CPIAUCSL	(0.949)	(0.936)	(0.978)	(0.979)	(0.960)	(0.969)
	[0.920]	[0.910]	[0.956]	[0.981]	[0.953]	[0.957]
	0.955	0.919	0.922	0.933	0.885	0.913
FEDFUNDS	(0.962)	(0.945)	(1.001)	(0.986)	(0.921)	(0.975)
	[0.960]	[0.932]	[0.986]	[0.975]	[0.915]	[0.945]
	0.776	0.783	0.789	0.813	0.822	0.877
INDPRO	(0.810)	(0.825)	(0.928)	(0.957)	(0.958)	(0.974)
	[0.801]	[0.802]	[0.902]	[0.903]	[0.913]	[0.925]
	0.644	0.712	0.763	0.847	0.902	0.918
UNRATE	(0.783)	(0.805)	(0.850)	(0.939)	(0.951)	(0.968)
	[0.714]	[0.785]	[0.804]	[0.910]	[0.923]	[0.957]
	0.886	0.916	0.925	0.944	0.962	0.977
PPIFGS	(0.970)	(1.012)	(1.016)	(1.026)	(1.004)	(1.000)
	[0.902]	[1.001]	[1.000]	[1.000]	[1.000]	[0.989]
	0.876	0.944	0.952	0.961	0.973	0.980
GS10	(0.988)	(1.003)	(1.032)	(1.006)	(0.997)	(1.000)
	[0.974]	[0.966]	[0.982]	[0.980]	[0.999]	[1.000]

All figures reported in the table are *relative* MSFEs, where the benchmark model is a univariate AR(1) specification for each variable - see equation (19) in Koop et al. (2019) for the formula.
Similarly to Koop et al. (2019), we have used the first half of the sample (January 1960 till June 1987) to obtain initial parameter estimates (and the initial predictions). All subsequent predictions are then computed using recursive estimates of the model.
In each cell, results for this paper are the first number. Numbers in round brackets are taken from Table 1 in Koop et al. (2019), and are the best MSFEs for each variable and each h. We point out that the models used in Koop et al. (2019) are the DFM model (see equation (17) in the original paper for details), the FAVAR model of Bernanke et al. (2005), the BVAR with the Minnesota prior suggested by Bańbura et al. (2010), and the two VARs with bayesian compression in Koop et al. (2019). Numbers in square brackets are the MSFEs obtained using the methodology proposed by Carriero et al. (2019).

TABLE B
Relative MSFE at various horizons h - VAR with n = 129.

Variable	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
	0.688	0.317	0.302	0.513	0.584	0.655
PAYEMS	(0.748)	(0.481)	(0.474)	(0.620)	(0.743)	(0.848)
	[0.712]	[0.401]	[0.415]	[0.495]	[0.615]	[0.773]
	0.788	0.793	0.815	0.867	0.832	0.799
CPIAUCSL	(0.860)	(0.887)	(0.904)	(0.916)	(0.885)	(0.872)
	[0.812]	[0.871]	[0.876]	[0.910]	[0.865]	[0.855]
	0.882	0.830	0.887	0.866	0.873	0.910
FEDFUNDS	(0.965)	(0.892)	(0.967)	(0.959)	(0.969)	(0.976)
	[0.913]	[0.875]	[0.944]	[0.912]	[0.810]	[0.954]
	0.655	0.714	0.748	0.840	0.872	0.903
INDPRO	(0.778)	(0.801)	(0.893)	(0.967)	(0.975)	(0.989)
	[0.762]	[0.785]	[0.853]	[0.913]	[0.916]	[0.925]
	0.688	0.673	0.710	0.747	0.855	0.899
UNRATE	(0.750)	(0.769)	(0.836)	(0.886)	(0.938)	(0.979)
	[0.722]	[0.713]	[0.874]	[0.866]	[0.910]	[0.944]
	0.910	0.923	0.948	0.955	0.962	0.977
PPIFGS	(0.938)	(1.013)	(1.034)	(1.041)	(1.011)	(1.032)
	[0.915]	[1.000]	[1.020]	[1.030]	[0.981]	[0.980]
	0.917	0.903	0.940	0.932	0.935	0.935
GS10	(1.009)	(1.005)	(1.049)	(1.022)	(1.000)	(1.003)
	[0.985]	[1.000]	[0.980]	[0.981]	[0.977]	[0.969]

The figures in the table are the same as in Table A. The numbers in round brackets are taken from Table 3 in Koop et al. (2019).

TABLE C
Out-of-sample density forecast performance at various horizons h - VAR with $n = 129$.

Variable	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
	0.448	0.523	0.551	0.510	0.413	0.389
PAYEMS	(*0.302*)	(*0.471*)	(*0.447*)	(*0.296*)	(*0.129*)	(*0.110*)
	[*0.352*]	[*0.489*]	[*0.461*]	[*0.315*]	[*0.256*]	[*0.210*]
	0.232	0.244	0.257	0.288	0.271	0.244
CPIAUCSL	(*0.052*)	(*0.098*)	(*0.121*)	(*0.227*)	(*0.212*)	(*0.060*)
	[*0.133*]	[*0.142*]	[*0.133*]	[*0.285*]	[*0.262*]	[*0.210*]
	0.318	0.322	0.327	0.310	0.288	0.256
FEDFUNDS	(*0.291*)	(*0.247*)	(*0.228*)	(*0.186*)	(*0.275*)	(*0.211*)
	[*0.301*]	[*0.296*]	[*0.282*]	[*0.197*]	[*0.236*]	[*0.224*]
	0.244	0.271	0.282	0.270	0.255	0.241
INDPRO	(*0.092*)	(*0.238*)	(*0.065*)	(*0.082*)	(*0.110*)	(*0.062*)
	[*0.107*]	[*0.241*]	[*0.187*]	[*0.190*]	[*0.186*]	[*0.181*]
	0.233	0.244	0.255	0.262	0.271	0.280
UNRATE	(*0.157*)	(*0.163*)	(*0.106*)	(*0.084*)	(*0.045*)	(*0.034*)
	[*0.173*]	[*0.182*]	[*0.180*]	[*0.175*]	[*0.179*]	[*0.162*]
	0.122	0.137	0.144	0.119	0.112	0.110
PPIFGS	(*0.059*)	(*−0.015*)	(*0.086*)	(*0.003*)	(*0.099*)	(*−0.130*)
	[*0.101*]	[*0.122*]	[*0.132*]	[*0.100*]	[*0.105*]	[*−0.001*]
	0.044	0.049	0.052	0.047	0.043	0.032
GS10	(*0.006*)	(*0.012*)	(*0.032*)	(*0.012*)	(*0.039*)	(*0.034*)
	[*0.100*]	[*0.013*]	[*0.017*]	[*0.022*]	[*0.030*]	[*0.036*]

The figures in the table are the differentials between the Average Predictive Likelihood (ALPL) and the benchmark, univariate AR(1) model. We refer to equation (21) in Koop et al. (2019) for details.
All density forecasts are generated as out-of-sample, using recursive estimates starting from July 1987 - see also the comments below Table A.
In each cell, results for this paper are the first number. Numbers in round brackets are taken from Table 4 in Koop et al. (2019), and are the best ALPLs for each variable and each h. The models considered are the same as in Table A. Numbers in square brackets have been obtained applying the methodology developed in Carriero et al. (2019).

TABLE D
Out-of-sample forecasting performance at various horizons h - multivariate results for VARs with $n = 19$ *and* $n = 129$.

	$n = 19$		$n = 129$	
	WMSFE	MVALPL	WMSFE	MVALPL
	0.832	1.042	0.892	1.055
$h = 1$	(*0.916*)	(*0.979*)	(*0.907*)	(*0.996*)
	[*0.910*]	[*0.982*]	[*0.903*]	[*1.032*]
	0.830	1.125	0.890	1.177
$h = 2$	(*0.926*)	(*1.068*)	(*0.908*)	(*1.139*)
	[*0.903*]	[*1.074*]	[*0.901*]	[*1.125*]
	0.932	1.133	0.898	1.213
$h = 3$	(*0.940*)	(*1.097*)	(*0.916*)	(*1.179*)
	[*0.939*]	[*1.117*]	[*0.912*]	[*1.201*]
	0.913	1.225	0.882	1.252
$h = 6$	(*0.954*)	(*1.030*)	(*0.933*)	(*1.131*)
	[*0.942*]	[*1.010*]	[*0.944*]	[*1.102*]
	0.914	1.221	0.910	1.246
$h = 9$	(*0.957*)	(*1.021*)	(*0.938*)	(*1.076*)
	[*0.937*]	[*1.015*]	[*0.927*]	[*1.131*]
	0.910	1.189	0.908	1.235
$h = 12$	(*0.988*)	(*0.927*)	(*0.968*)	(*1.009*)
	[*0.975*]	[*0.963*]	[*0.975*]	[*1.104*]

The figures in the table are the ratio between the weighted MSFEs (WMSFE) and the benchmark univariate AR(1) specifications, and the differential between the multivariate average predictive likelihood and the corresponding indicator for the benchmark model (MVALPL) - we refer to equations (20) and (22) in Koop et al. (2019) for details; in those formulas, similarly to the authors, we have used $n = 7$ since the predictive exercise focuses on 7 series only.
All density forecasts are generated as out-of-sample, using recursive estimates starting from July 1987 - see also the comments below Table A.
In each cell, results for this paper are the first number. Numbers in round brackets are taken from Table 5 in Koop et al. (2019), and are the best results for each variable and each h. The models considered are the same as in Table A. As in the previous table, numbers in square brackets refer to the methodology developed by Carriero et al. (2019).

TABLE E

Out-of-sample forecasting performance at various horizons h - VARs with n = 19 and n = 129.

	MSFE						*ALPL*					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
	$n=19$						$n=19$					
	0.614	0.413	0.410	0.429	0.452	0.510	0.449	0.416	0.422	0.420	0.386	0.377
PAYEMS	*(0.700)*	*(0.565)*	*(0.565)*	*(0.651)*	*(0.769)*	*(0.872)*	*(0.338)*	*(0.391)*	*(0.352)*	*(0.078)*	*(−0.422)*	*(−0.533)*
	[0.685]	*[0.520]*	*[0.537]*	*[0.581]*	*[0.703]*	*[0.860]*	*[0.342]*	*[0.395]*	*[0.366]*	*[0.105]*	*[0.202]*	*[−0.100]*
	0.878	0.865	0.867	0.832	0.820	0.822	0.416	0.445	0.440	0.432	0.417	0.403
CPIAUCSL	*(0.924)*	*(0.872)*	*(0.884)*	*(0.869)*	*(0.841)*	*(0.845)*	*(0.284)*	*(0.211)*	*(0.461)*	*(0.191)*	*(0.280)*	*(0.292)*
	[0.913]	*[0.865]*	*[0.872]*	*[0.855]*	*[0.830]*	*[0.810]*	*[0.287]*	*[0.225]*	*[0.460]*	*[0.203]*	*[0.207]*	*[0.301]*
	0.810	0.803	0.800	0.894	0.912	0.933	0.893	0.890	0.785	0.723	0.615	0.514
FEDFUNDS	*(0.879)*	*(0.892)*	*(0.924)*	*(0.995)*	*(0.967)*	*(1.061)*	*(0.760)*	*(0.594)*	*(0.423)*	*(0.382)*	*(0.303)*	*(0.365)*
	[0.861]	*[0.887]*	*[0.910]*	*[0.976]*	*[0.953]*	*[1.012]*	*[0.792]*	*[0.610]*	*[0.515]*	*[0.402]*	*[0.212]*	*[0.297]*
	0.872	0.870	0.865	0.910	0.915	0.922	0.317	0.320	0.320	0.315	0.287	0.280
INDPRO	*(0.899)*	*(0.925)*	*(0.940)*	*(0.978)*	*(0.980)*	*(0.989)*	*(−0.030)*	*(−0.224)*	*(−0.128)*	*(−0.509)*	*(−0.414)*	*(−0.225)*
	[0.885]	*[0.914]*	*[0.935]*	*[0.966]*	*[0.974]*	*[0.982]*	*[0.103]*	*[0.117]*	*[0.204]*	*[0.221]*	*[0.227]*	*[0.229]*
	0.811	0.803	0.800	0.855	0.872	0.913	0.310	0.321	0.287	0.217	0.188	0.180
UNRATE	*(0.846)*	*(0.847)*	*(0.876)*	*(0.939)*	*(0.971)*	*(1.011)*	*(0.123)*	*(0.104)*	*(0.095)*	*(0.059)*	*(0.036)*	*(−0.009)*
	[0.837]	*[0.833]*	*[0.857]*	*[0.921]*	*[0.954]*	*[1.000]*	*[0.175]*	*[0.192]*	*[0.103]*	*[0.116]*	*[0.072]*	*[0.089]*
	0.911	0.913	0.922	0.923	0.930	0.944	0.322	0.445	0.313	0.310	0.312	0.317
PPIFGS	*(0.968)*	*(0.991)*	*(1.001)*	*(0.998)*	*(0.992)*	*(1.010)*	*(0.270)*	*(0.349)*	*(0.401)*	*(0.283)*	*(0.407)*	*(0.354)*
	[0.973]	*[0.985]*	*[1.000]*	*[0.984]*	*[0.982]*	*[1.000]*	*[0.276]*	*[0.352]*	*[0.417]*	*[0.402]*	*[0.287]*	*[0.395]*
	0.977	0.970	0.965	0.977	0.981	0.987	0.282	0.293	0.215	0.210	0.180	0.171
GS10	*(1.018)*	*(1.017)*	*(1.039)*	*(1.030)*	*(0.995)*	*(1.030)*	*(0.025)*	*(−0.016)*	*(−0.053)*	*(−0.057)*	*(−0.004)*	*(0.030)*
	[1.000]	*[0.993]*	*[1.020]*	*[1.010]*	*[0.984]*	*[1.020]*	*[0.106]*	*[0.107]*	*[0.119]*	*[0.185]*	*[0.144]*	*[0.151]*
	0.855	0.813	0.811	0.822	0.845	0.882	1.913	1.982	1.820	1.778	1.633	1.515
Multivariate	*(0.905)*	*(0.884)*	*(0.892)*	*(0.916)*	*(0.924)*	*(0.967)*	*(1.653)*	*(1.701)*	*(1.573)*	*(1.224)*	*(1.049)*	*(0.851)*
	[0.902]	*[0.878]*	*[0.882]*	*[0.901]*	*[0.897]*	*[0.895]*	*[1.720]*	*[1.717]*	*[1.602]*	*[1.610]*	*[1.355]*	*[1.121]*
	$n=129$						$n=129$					
	0.617	0.515	0.504	0.561	0.674	0.775	0.582	0.611	0.600	0.455	0.422	0.310
PAYEMS	*(0.685)*	*(0.566)*	*(0.548)*	*(0.656)*	*(0.762)*	*(0.879)*	*(0.338)*	*(0.405)*	*(0.374)*	*(0.083)*	*(−0.447)*	*(−0.530)*
	[0.673]	*[0.582]*	*[0.557]*	*[0.662]*	*[0.741]*	*[0.885]*	*[0.340]*	*[0.416]*	*[0.385]*	*[0.120]*	*[0.103]*	*[−0.134]*
	0.893	0.820	0.815	0.807	0.782	0.727	0.352	0.487	0.489	0.510	0.610	0.482
CPIAUCSL	*(0.904)*	*(0.846)*	*(0.844)*	*(0.848)*	*(0.800)*	*(0.796)*	*(0.241)*	*(0.364)*	*(0.361)*	*(0.354)*	*(0.539)*	*(0.074)*	*(0.074)*
	[0.901]	*[0.835]*	*[0.832]*	*[0.835]*	*[0.791]*	*[0.790]*	*[0.322]*	*[0.357]*	*[0.351]*	*[0.345]*	*[0.510]*	*[0.106]*
	0.789	0.780	0.814	0.855	0.933	0.955	0.887	0.893	0.744	0.615	0.522	0.447
FEDFUNDS	*(0.885)*	*(0.911)*	*(0.920)*	*(1.022)*	*(1.034)*	*(1.075)*	*(0.715)*	*(0.577)*	*(0.489)*	*(0.445)*	*(0.100)*	*(0.269)*
	[0.865]	*[0.873]*	*[0.915]*	*[0.987]*	*[0.985]*	*[0.980]*	*[0.726]*	*[0.612]*	*[0.513]*	*[0.562]*	*[0.570]*	*[0.092]*
	0.813	0.820	0.884	0.895	0.904	0.922	0.332	0.330	0.280	0.275	0.261	0.230
INDPRO	*(0.896)*	*(0.928)*	*(0.957)*	*(0.996)*	*(1.002)*	*(1.020)*	*(0.116)*	*(0.036)*	*(−0.184)*	*(−0.320)*	*(−0.205)*	*(−0.210)*
	[0.881]	*[0.900]*	*[0.942]*	*[0.955]*	*[0.942]*	*[0.987]*	*[0.125]*	*[0.067]*	*[0.132]*	*[0.102]*	*[0.100]*	*[0.247]*
	0.813	0.821	0.835	0.922	0.944	0.986	0.225	0.220	0.197	0.144	0.132	0.130
UNRATE	*(0.836)*	*(0.851)*	*(0.880)*	*(0.949)*	*(0.981)*	*(1.026)*	*(0.122)*	*(0.102)*	*(0.078)*	*(0.050)*	*(0.034)*	*(0.010)*
	[0.830]	*[0.846]*	*[0.873]*	*[0.938]*	*[0.976]*	*[0.980]*	*[0.107]*	*[0.115]*	*[0.086]*	*[0.079]*	*[0.116]*	*[0.087]*
	0.915	0.910	0.922	0.935	0.965	0.970	0.387	0.415	0.419	0.422	0.439	0.420
PPIFGS	*(0.983)*	*(0.985)*	*(1.005)*	*(1.008)*	*(0.995)*	*(1.012)*	*(0.254)*	*(0.363)*	*(0.371)*	*(0.346)*	*(0.385)*	*(0.213)*
	[0.962]	*[0.973]*	*[0.986]*	*[0.980]*	*[0.982]*	*[0.981]*	*[0.262]*	*[0.108]*	*[0.109]*	*[0.352]*	*[0.392]*	*[0.310]*
	0.927	0.920	0.933	0.942	0.952	0.981	0.210	0.206	0.200	0.186	0.177	0.163
GS10	*(1.021)*	*(1.021)*	*(1.034)*	*(1.024)*	*(1.013)*	*(1.021)*	*(0.008)*	*(0.037)*	*(0.017)*	*(0.008)*	*(0.029)*	*(−0.033)*
	[1.010]	*[1.012]*	*[0.986]*	*[0.984]*	*[0.982]*	*[0.980]*	*[0.103]*	*[0.047]*	*[0.025]*	*[0.017]*	*[0.042]*	*[0.111]*
	0.872	0.865	0.860	0.893	0.904	0.933	1.844	1.912	1.910	1.895	1.417	1.218
Multivariate	*(0.902)*	*(0.883)*	*(0.885)*	*(0.922)*	*(0.932)*	*(0.967)*	*(1.667)*	*(1.666)*	*(1.593)*	*(1.216)*	*(1.002)*	*(0.713)*
	[0.900]	*[0.872]*	*[0.865]*	*[0.910]*	*[0.898]*	*[0.890]*	*[1.710]*	*[1.716]*	*[1.689]*	*[1.352]*	*[1.289]*	*[0.973]*

The figures in the table are the relative MSFEs and the differentials between the Average Predictive Likelihood (ALPL) and the benchmark, univariate AR(1) model. The last figures in each panel of the table represent the overall forecasting ability, measured using WMSFE and MVALPL, defined as in the previous tables.

All forecasts are generated as out-of-sample, using recursive estimates starting from July 1987 - see also the comments below Table A.

In each cell, results for this paper are the first number. Numbers in round brackets are taken from Table 6 in Koop et al. (2019), and correspond to the BCVAR-TVP-SV. Numbers in square brackets are based on applying the methodology developed by Carriero et al. (2019).

TABLE F

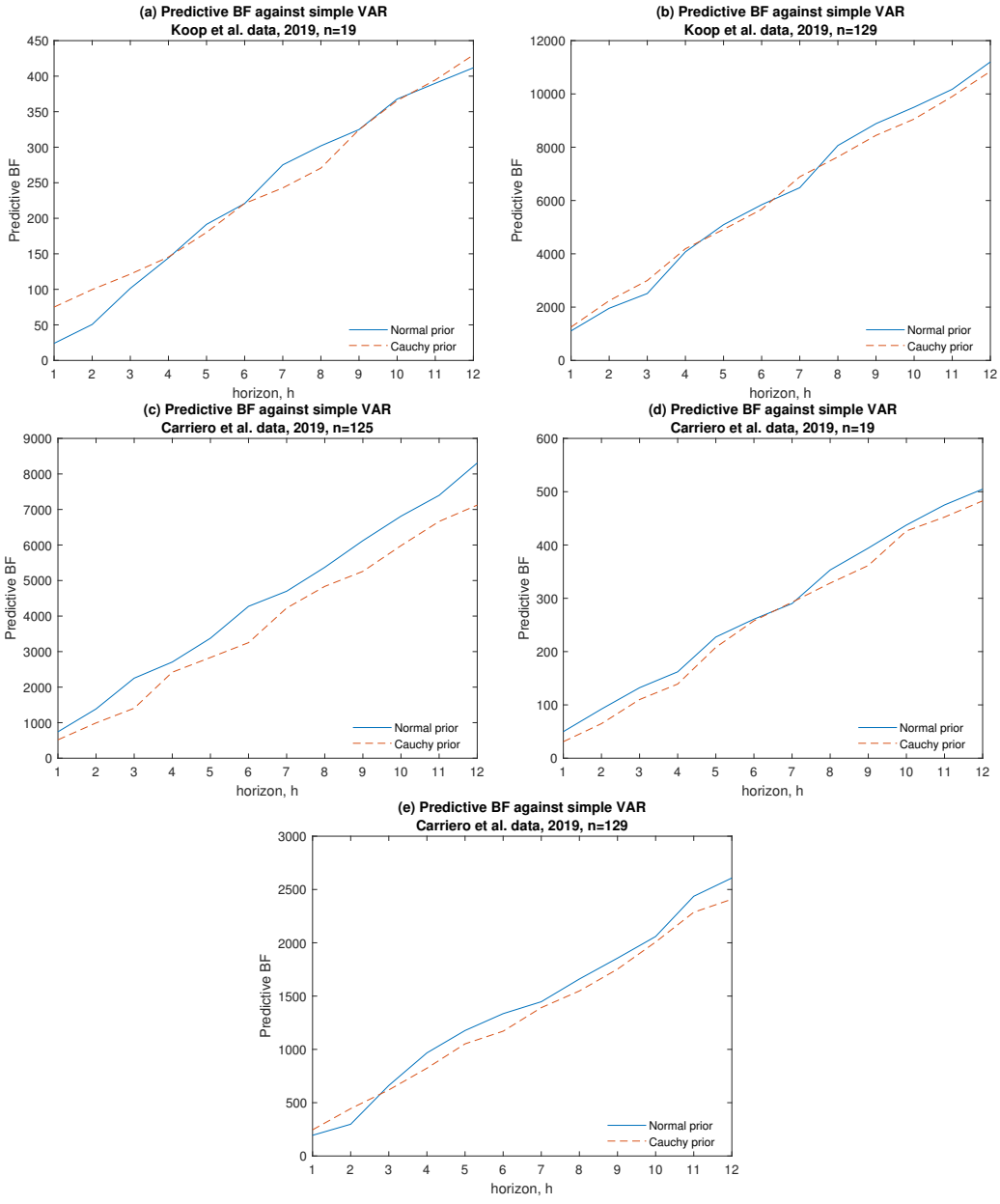
Out-of-sample forecasting performance at various horizons h - multivariate results, VARs with $n = 7$, $n = 19$ and $n = 129$.

	$n = 7$		$n = 19$		$n = 129$	
	$WMSFE$	$MVALPL$	$WMSFE$	$MVALPL$	$WMSFE$	$MVALPL$
	0.881	2.454	0.844	2.410	0.840	2.225
$h = 1$	(*0.917*)	(*2.047*)	(*0.905*)	(*1.653*)	(*0.902*)	(*1.667*)
	[*0.910*]	[*2.103*]	[*0.901*]	[*1.722*]	[*0.900*]	[*1.734*]
	0.845	2.450	0.841	2.433	0.832	2.215
$h = 2$	(*0.895*)	(*1.907*)	(*0.884*)	(*1.701*)	(*0.883*)	(*1.666*)
	[*0.891*]	[*1.920*]	[*0.870*]	[*1.835*]	[*0.875*]	[*1.712*]
	0.857	2.310	0.822	2.210	0.831	2.188
$h = 3$	(*0.901*)	(*1.845*)	(*0.892*)	(*1.573*)	(*0.885*)	(*1.593*)
	[*0.900*]	[*1.971*]	[*0.885*]	[*1.617*]	[*0.847*]	[*1.650*]
	0.872	2.287	0.885	2.280	0.890	2.261
$h = 6$	(*0.912*)	(*1.608*)	(*0.916*)	(*1.224*)	(*0.922*)	(*1.216*)
	[*0.910*]	[*1.735*]	[*0.905*]	[*1.476*]	[*0.913*]	[*1.457*]
	0.893	2.100	0.905	2.101	0.913	2.090
$h = 9$	(*0.936*)	(*1.385*)	(*0.924*)	(*1.078*)	(*0.932*)	(*1.002*)
	[*0.925*]	[*1.410*]	[*0.911*]	[*1.082*]	[*0.927*]	[*1.000*]
	0.914	1.776	0.922	1.872	0.930	1.914
$h = 12$	(*0.960*)	(*0.931*)	(*0.967*)	(*1.039*)	(*0.950*)	(*0.713*)
	[*0.973*]	[*0.961*]	[*0.975*]	[*1.210*]	[*0.942*]	[*0.955*]

As in Koop et al. (2019), we have considered one extra VAR model with $n = 7$ (containing only the seven variables of interest), alongside the ones with $n = 19$ and $n = 129$ variables.

The figures in the table are the weighted MSFEs (WMSFE) relative to the AR(1) benchmark, and the differentials between the multivariate Average Predictive Likelihood (WMALPL) and the benchmark, univariate AR(1) models. As in the previous tables, forecasts are generated as out-of-sample, using recursive estimates starting from July 1987. Similarly, results for this paper are the first number in each cell. Numbers in round brackets are taken from Table 7 in Koop et al. (2019), and are the best results for each variable and each h out of three specifications: a Bayesian compressed VAR with heteroskedasticity (denoted as $BCVAR_{sv}$ in Koop et al., 2019); a Bayesian compressed TVP-VAR with heteroskedasticity (denoted as $BCVAR_{tvp-sv}$); and, finally, the Bayesian VAR estimated with the methodology by Carriero et al. (2019) (denoted as $BVAR_{ccm}$). Numbers in square brackets are based on Carriero et al. (2019).

Fig 1: Predictive Bayes factors - fixed coefficients, homoskedastic VAR



.

Fig 2: Predictive Bayes factors - homoskedastic TVP-VAR

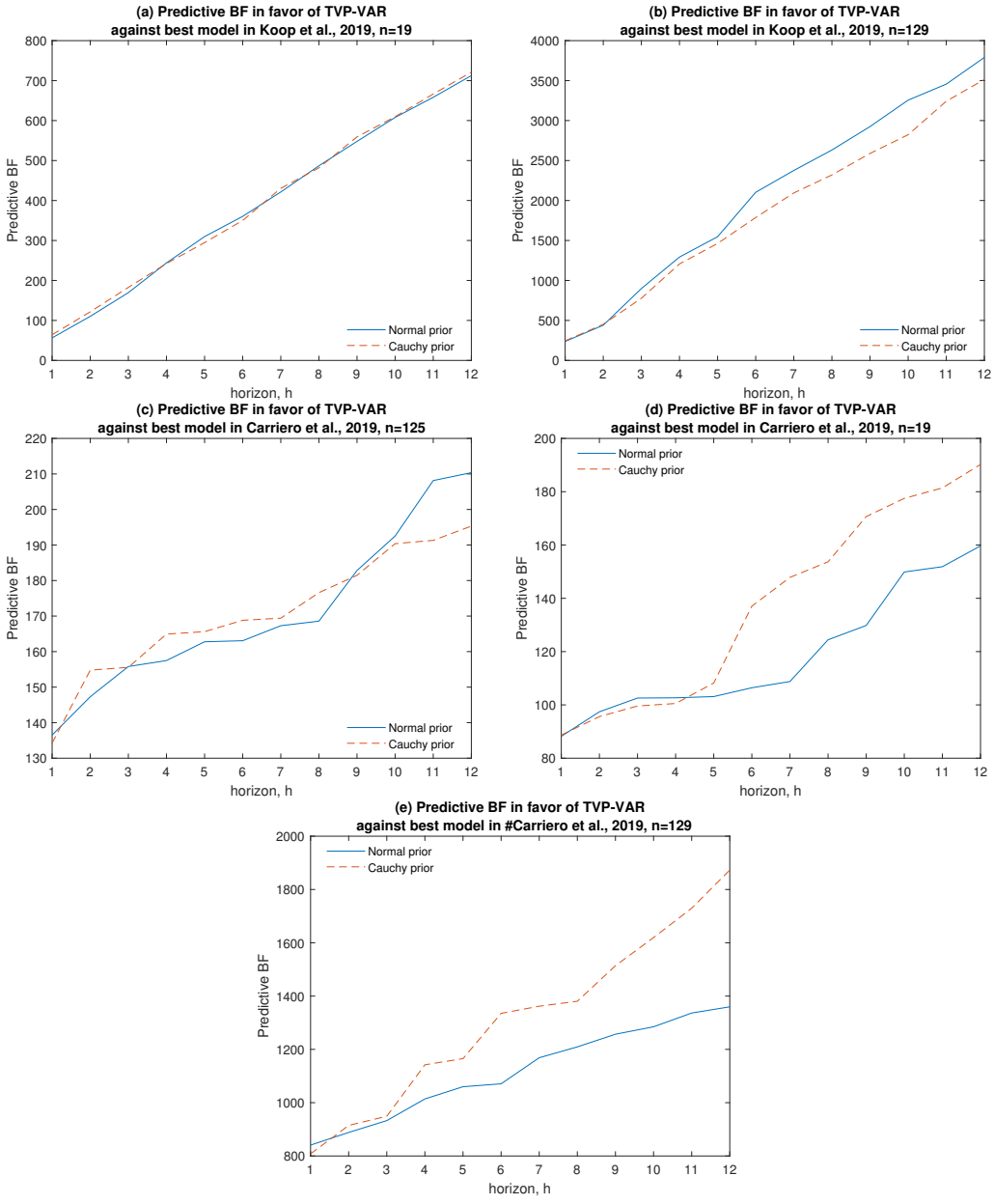
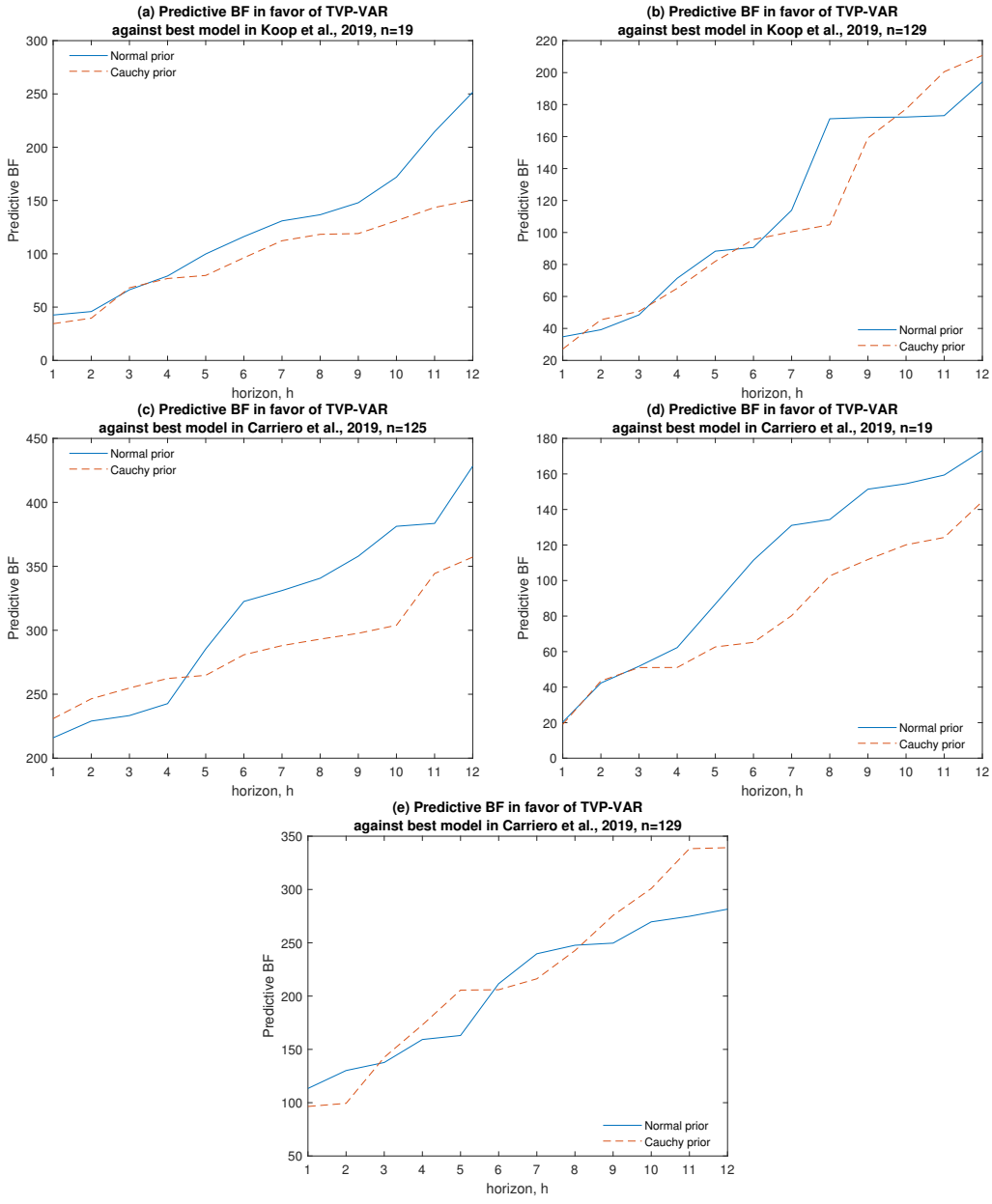


Fig 3: Predictive Bayes factors - heteroskedastic TVP-VAR



.

Fig 4: MCMC convergence diagnostics and computational efficiency.

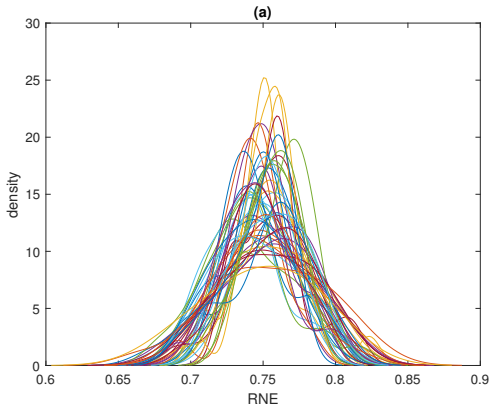


Fig 5: Normal priors - RNE

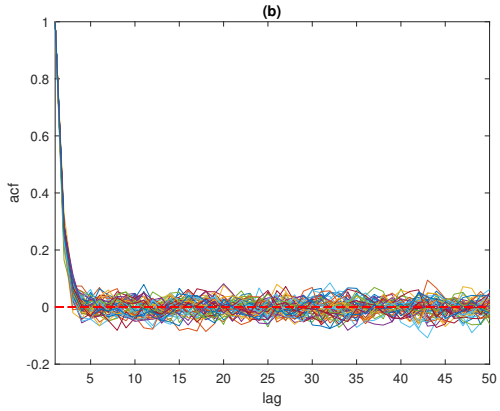


Fig 6: Normal priors - ACF

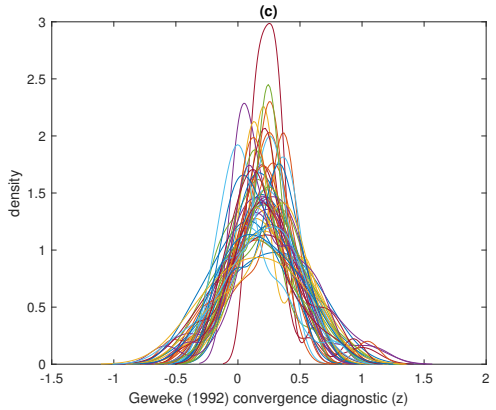


Fig 7: Normal priors - Geweke (1992) convergence diagnostic

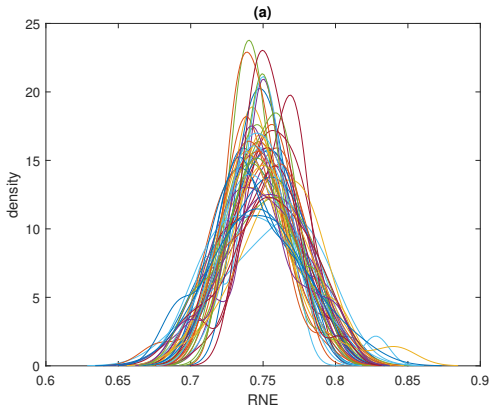


Fig 8: Cauchy priors - RNE

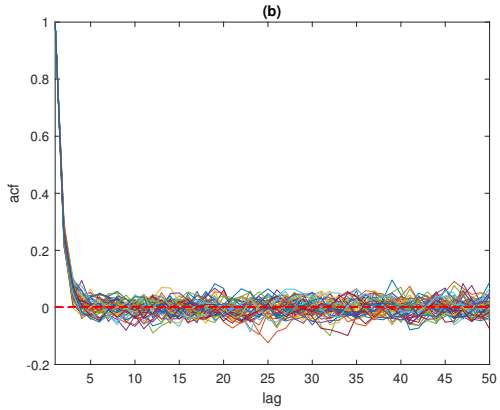


Fig 9: Cauchy priors - ACF

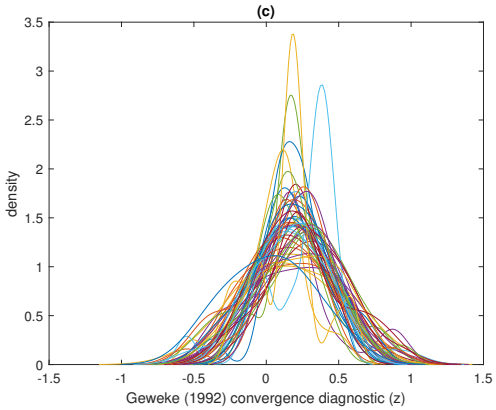
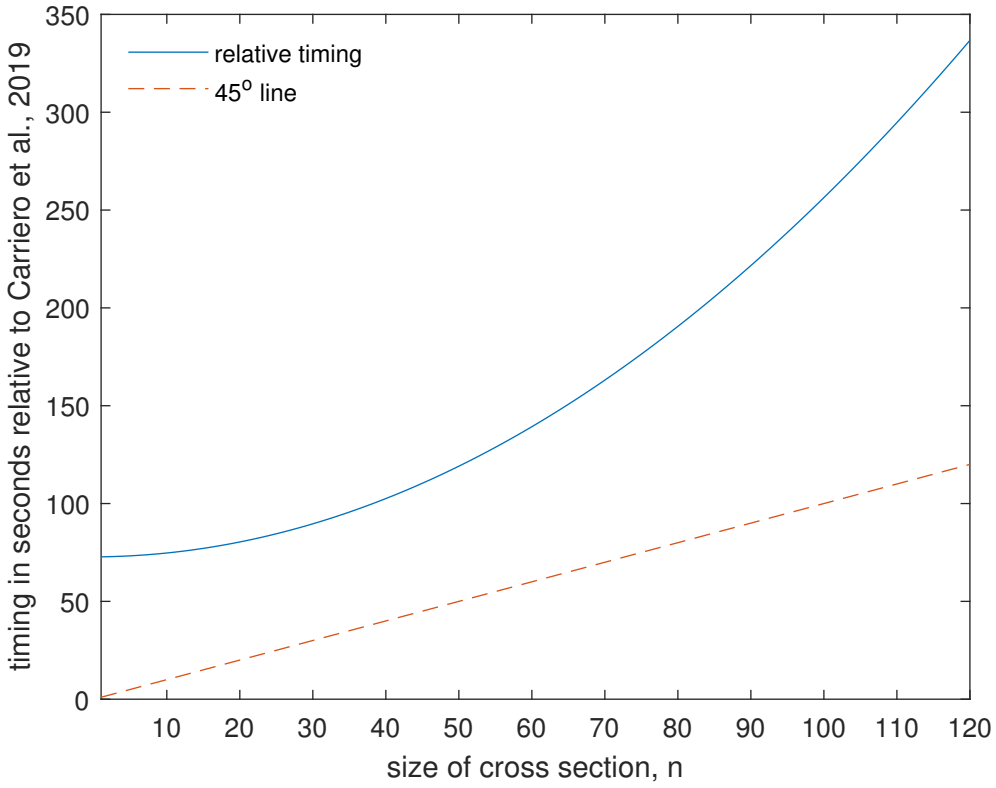


Fig 10: Cauchy priors - Geweke (1992) convergence diagnostic

ACFs are computed with skipping every other 5 draws. RNE (sub-panels *(a)-(d)*) is relative numerical efficiency (Geweke, 1992) which is equal to 1 for *i.i.d.* draws from the posterior (with skipping every other 5 draws). In sub-panels *(b)-(e)* we show ACFs for 50 randomly selected parameters (with skipping every other 5 draws). In sub-panels *(c)-(f)* we show Geweke (1992) convergence diagnostic (a z-statistic)

Fig 11: Timing (seconds) relative to Carriero et al. (2019)



Timings are for the High End Cluster of Lancaster University. The combined facility offers 9,900 cores, 50TB of aggregate memory, 6 Tesla V100 GPUs, 230TB of high performance filestore for general use and 4PB of medium performance filestore for GridPP data. The cluster operating system is CentOS Linux, with job submission handled by Son of Grid Engine (SGE). The service supports Fortran compilers. We use Fortran 77 making extensive of BLAS and LAPACK from NETLIB.

Fig 12: Predictive Bayes factors - TVP-VARMA

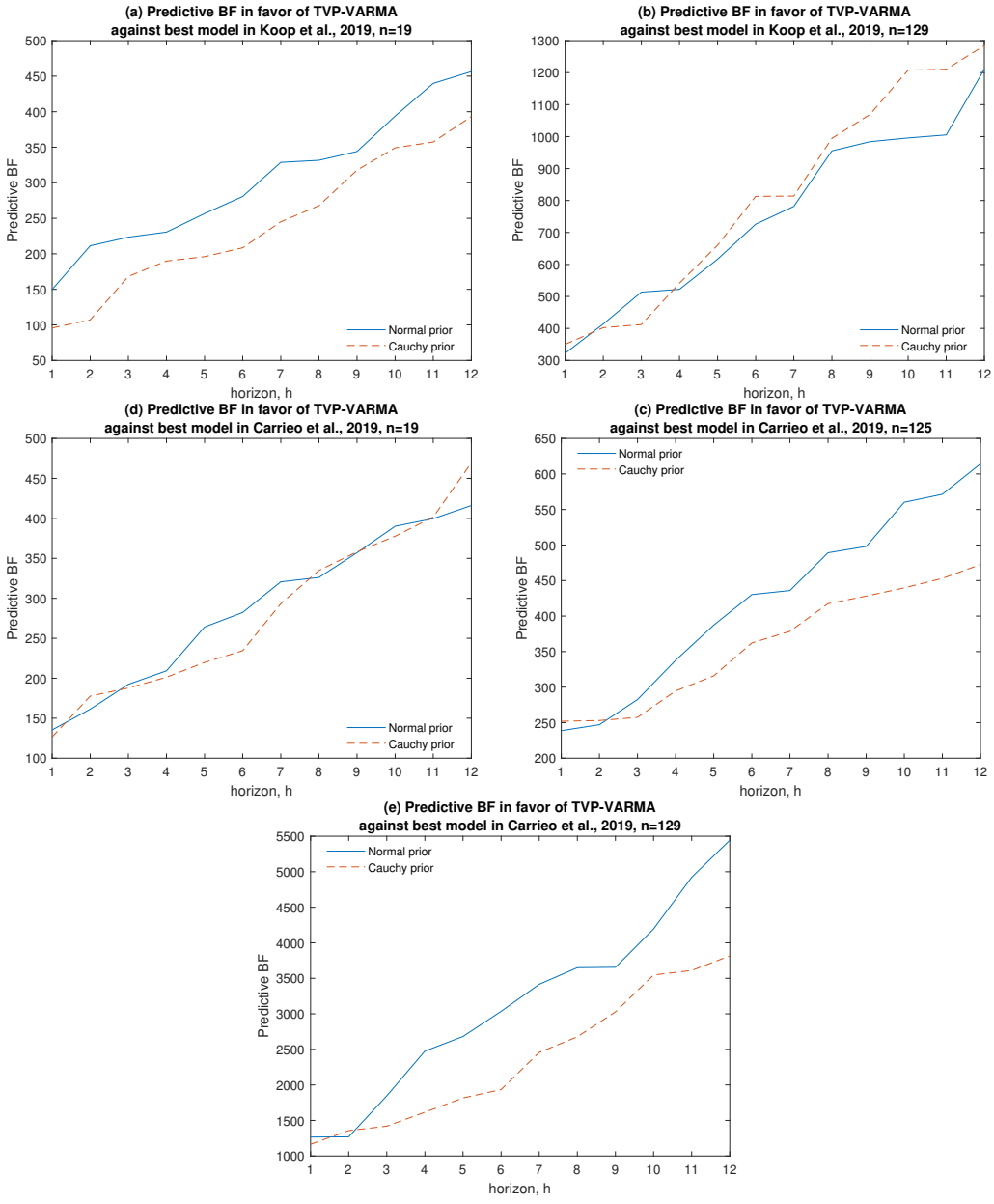
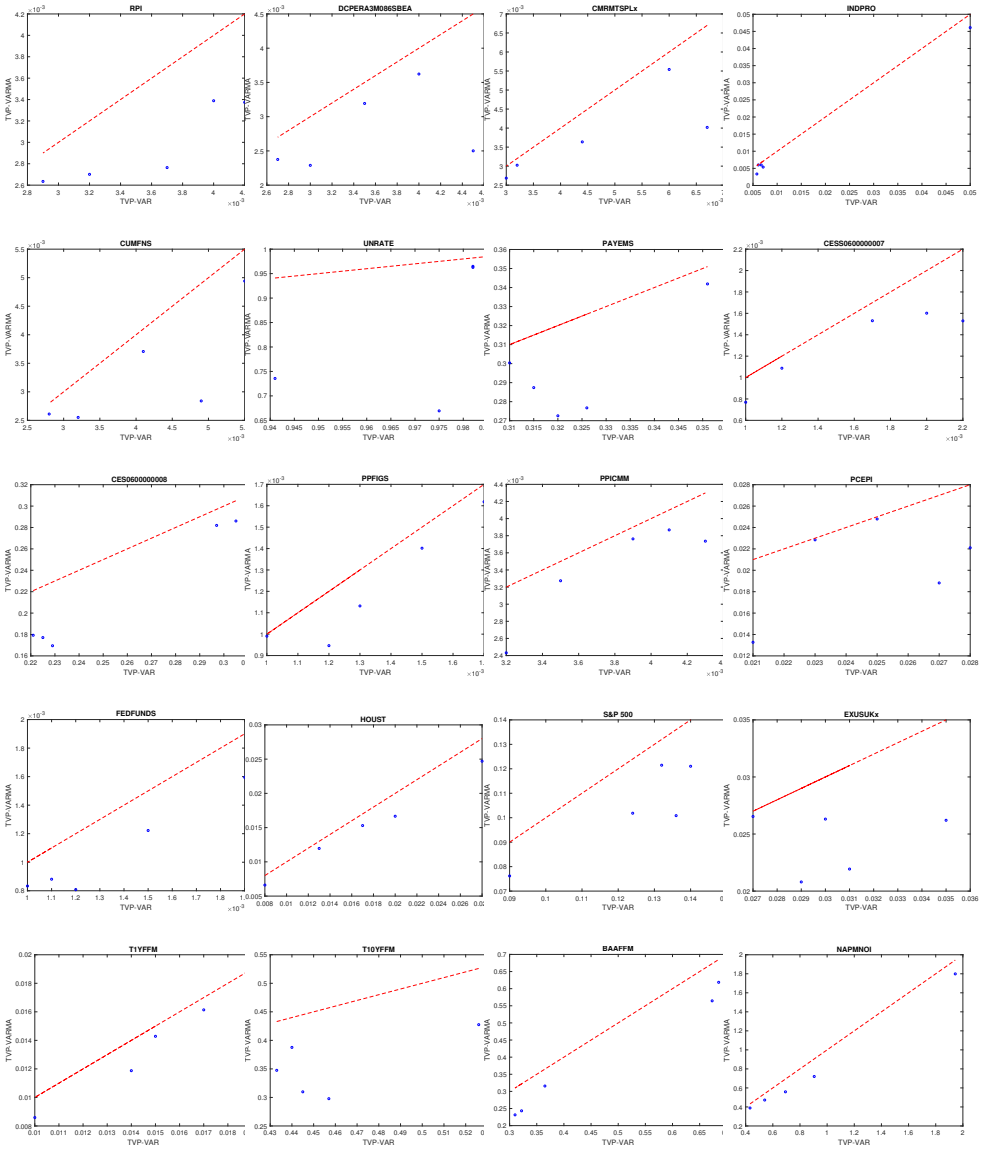
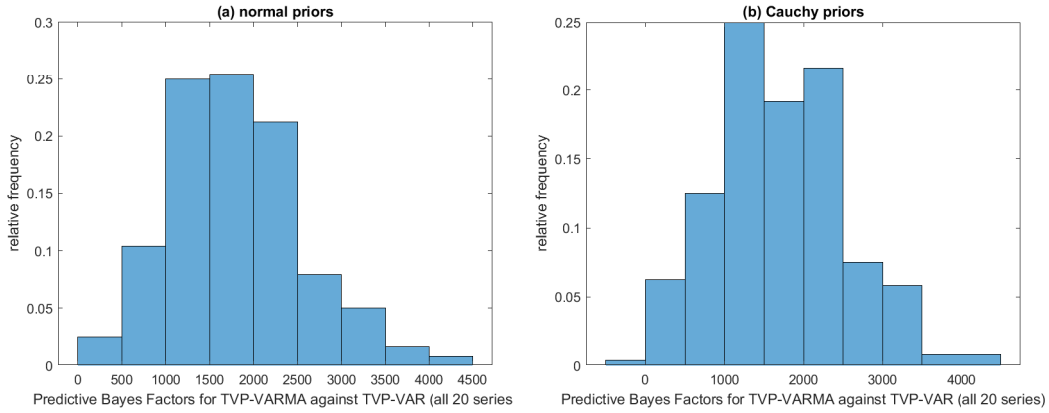


Fig 13: MSFE of TVP-VARMA relative to TVP-VAR, Carriero et al. (2019) data.



Circles represent MSFEs for h=1, 3, 6, 9, 12; the red dotted line represents the 45 degrees line.

Fig 14: Bayes factors for TVP-VARMA against TVP-VAR with Carriero et al. (2019), all 20 series and 12 horizons



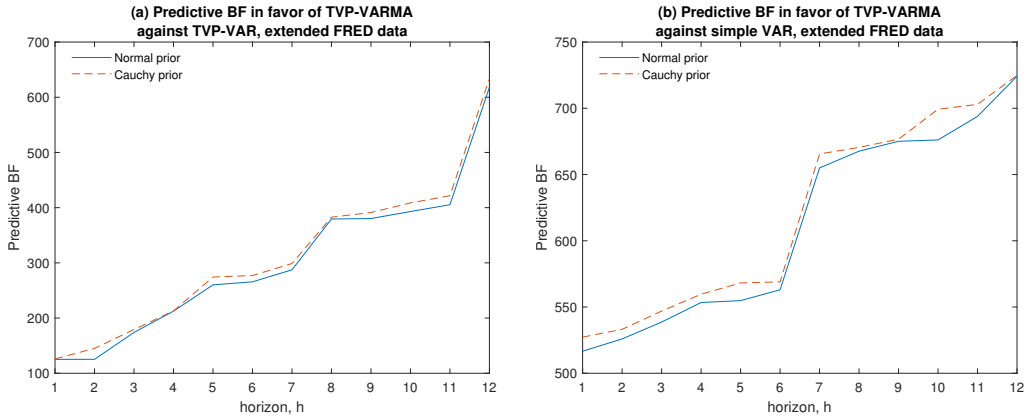
.

TABLE G
Out-of-sample forecasting performance from TVP-VARMA with $n = 129$ at various horizons.

Date	INDPRO	UNRATE	PAYEMS	FEDFUNDS	GS10	PPIFGS	CPIAUSL
March 2019	0.017	0.089	0.077	0.021	0.077	0.035	0.022
May	0.021	0.091	0.081	0.019	0.055	0.032	0.020
August	0.017	0.085	0.085	0.017	0.047	0.027	0.032
October	0.013	0.087	0.075	0.012	0.043	0.021	0.021
December	0.015	0.082	0.070	0.0091	0.040	0.019	0.017
January 2020	0.011	0.080	0.067	0.0085	0.040	0.015	0.015
March	0.010	0.079	0.061	0.0082	0.038	0.010	0.012
August	0.009	0.080	0.060	0.0077	0.037	0.009	0.009
December	0.007	0.077	0.058	0.0065	0.030	0.009	0.008

We use the same data as in Koop et al. (2019), extended until 2020:Q4. The figures in the table are the RMSEs for each series.

Fig 15: Predictive Bayes in favour of TVP-VARMA and against TVP-VAR during the Covid-19 outbreak.



.

TABLE H

Average log predictive scores for the number of factors $q = 0, 1, 2, 3, 4, 5$ across various specifications - comparison with Table 2 in Kastner and Huber (2020).

$q \rightarrow$	0	1	2	3	4	5
VAR(1)-FSV NG(0.1)$^{(a)}$	−10.55	−9.95	−9.31	−9.26	−9.47	−9.46
VAR(2)-FSV DL(1/k)$^{(b)}$	−10.50	−9.88	−9.22	−9.38	−9.28	−9.43
TVP-VAR *S1*	−9.44	−9.37	−9.21	−9.10	−9.01	−8.77
TVP-VAR *S2*	−9.44	−9.36	−9.20	−9.11	−9.00	−8.76
TVP-VARMA *S1*	−9.32	−9.12	−8.82	−8.77	−8.36	−8.13
TVP-VARMA *S2*	−9.32	−9.12	−8.82	−8.76	−8.35	−8.12

Results denoted as *(a)* and *(b)* are taken from Table 2 in Kastner and Huber (2020), and they correspond to their models with the best out-of-sample forecasting performance. Results are reported for various values of q.
Estimation and prediction is conducted on all $m = 215$ component series; the predictive density is then evaluated on the set of 11 variables of interest. Larger numbers indicate better joint predictive density performance.

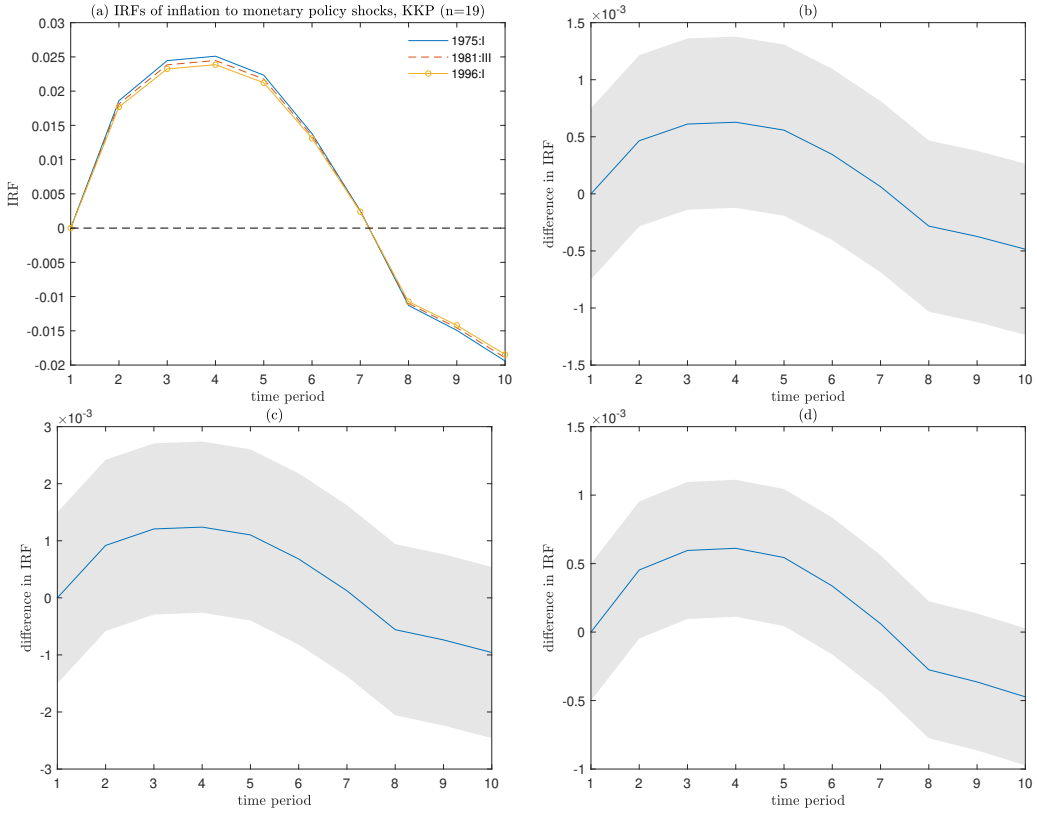
TABLE I

Average univariate log predictive scores for inflation (CPIAUCSL), short-term interest rates (FEDFUNDS), and output growth (GDPC96) with $q = 0.1.2$ factors. Comparison with Table 3 in Kastner and Huber (2020).

	CPIAUSCL			FEDFUNDS			GDPC96		
$q \rightarrow$	0	1	2	0	1	2	0	1	2
VAR(2)-FSV DL(1/k)$^{(b)}$	−1.030	−1.110	−1.130	−1.260	−1.260	−1.240	0.080	0.050	0.030
TVP-VAR *S1*	−0.970	−0.930	−0.090	−1.150	−1.200	−1.150	0.060	0.070	0.090
TVP-VAR *S2*	−0.970	−0.930	−0.090	−1.150	−1.200	−1.150	0.060	0.070	0.090
TVP-VARMA *S1*	−0.910	−0.893	−0.071	−1.100	−1.050	−1.050	0.070	0.090	0.090
TVP-VARMA *S2*	−0.910	−0.893	−0.071	−1.100	−1.050	−1.050	0.070	0.090	0.090

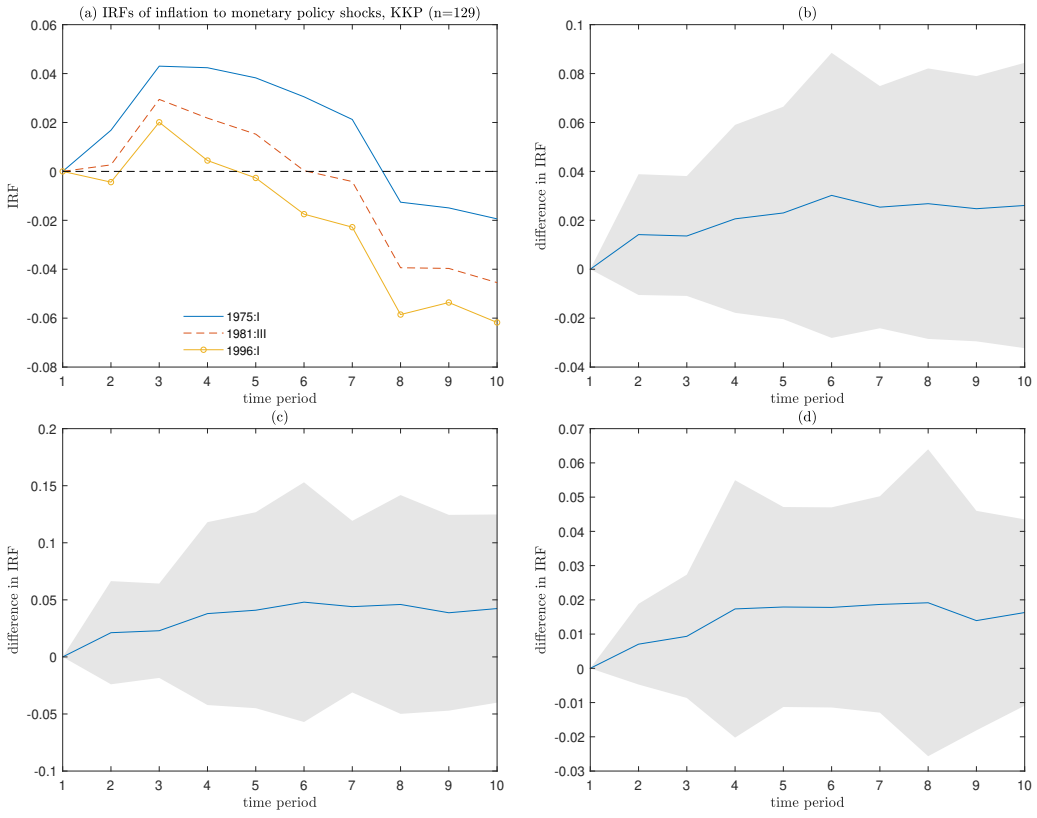
(a) is taken from Table 3 in Kastner and Huber (2020).

Fig 16: Impulse responses of inflation to monetary policy shocks - medium VAR ($n = 19$)



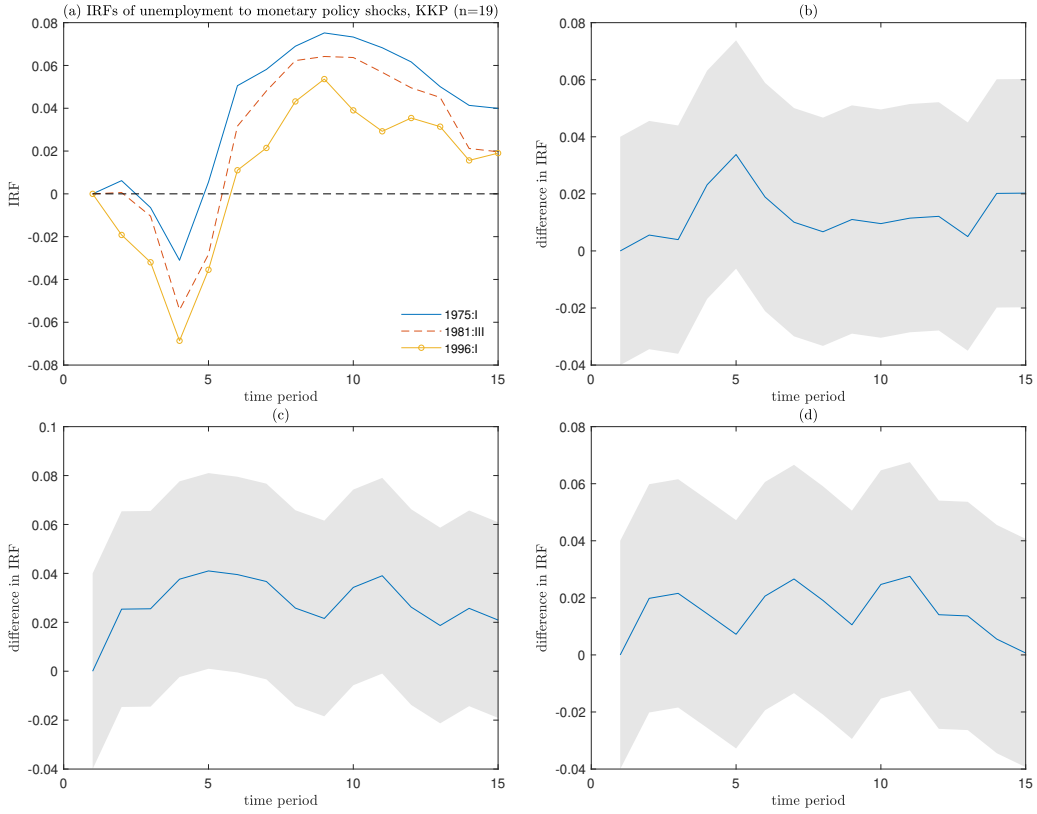
The grey area represents the 90% confidence region.

Fig 17: Impulse responses of inflation to monetary policy shocks - large VAR ($n = 129$)



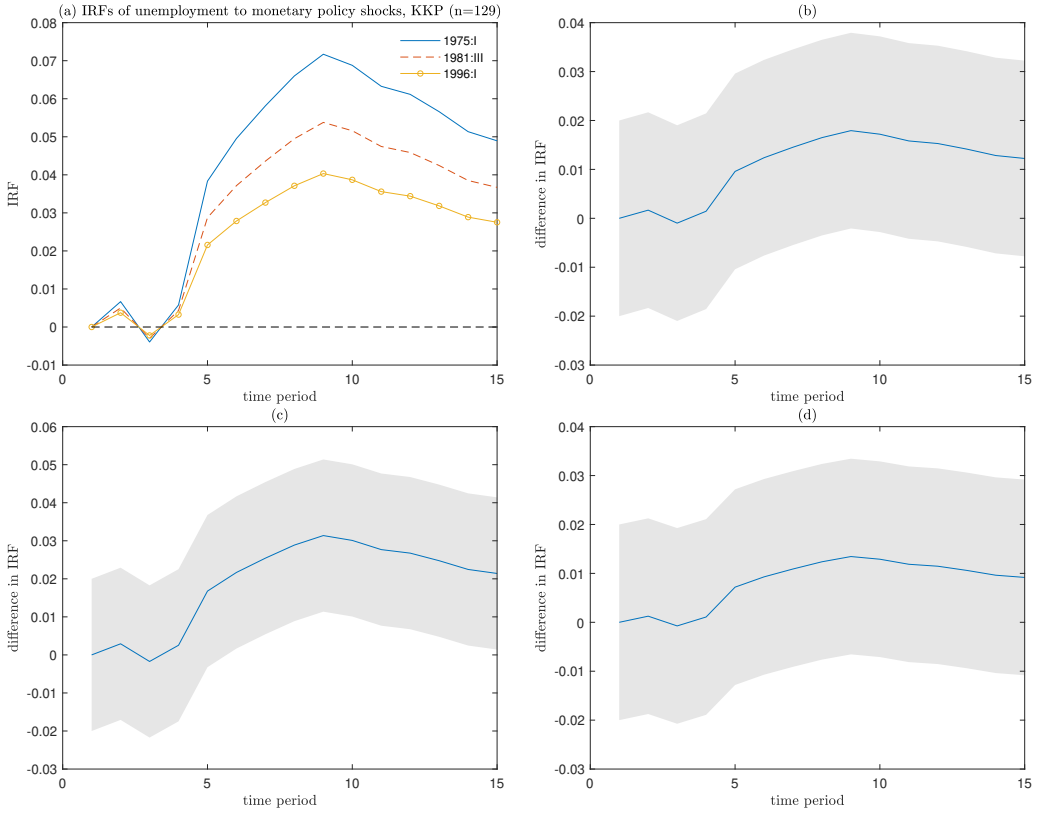
The grey area represents the 90% confidence region.

Fig 18: Impulse responses of unemployment to monetary shocks - medium VAR ($n = 19$)



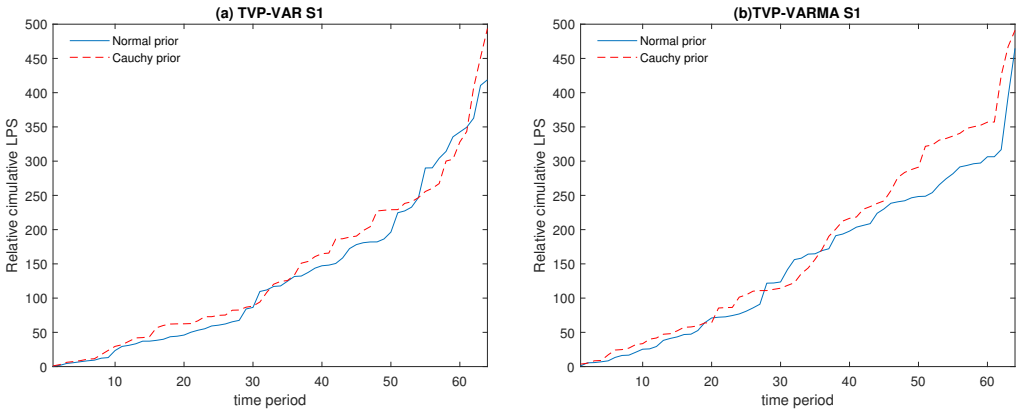
The grey area represents the 90% confidence region.

Fig 19: Impulse responses of unemployment to monetary shocks - large VAR ($n = 129$)



The grey area represents the 90% confidence region.

Fig 20: Relative Cumulative LPS (1990:Q1 - 2015.:Q4)



Cumulative log predictive scores, relative to a zero-mean model with independent stochastic volatility components for all component series; see also Figure 4 in Kastner and Huber (2020). Higher values correspond to better one-quarter-ahead density predictions up to the corresponding point in time. To assess the forecasting performance of our model, we conduct a pseudo out-of-sample forecasting exercise with initial estimation sample ranging from 1959:Q3 to 1990:Q2. Based on this estimation period, we compute one-quarter-ahead predictive densities for the first period in the hold-out (i.e. 1990:Q3). After obtaining the corresponding predictive densities and evaluating the corresponding log predictive likelihoods, we expand the estimation period and re-estimate the model. This procedure is repeated 100 times until the final point of the full sample is reached. The quarterly scores obtained this way are then accumulated.

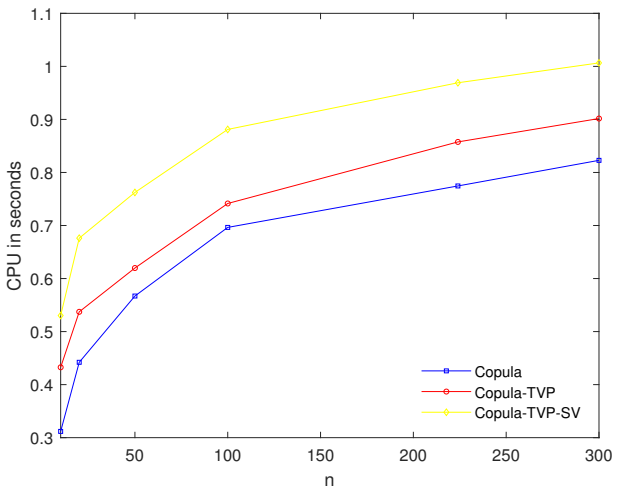
TABLE J
Simulation study - comparison with Table 1 in Kastner and Huber (2020).

	sparse				*intermediate*				*dense*			
n	10	20	50	100	10	20	50	100	10	20	50	100
T												
50	0.038*	0.019*	0.022*	0.025	0.049*	0.044	0.039*	0.033*	0.037*	0.030*	0.027*	0.025*
100	0.036*	0.017*	0.021*	0.023	0.047*	0.040*	0.035*	0.028*	0.033*	0.028*	0.024*	0.020*
150	0.033	0.017*	0.020*	0.024	0.043	0.038*	0.032*	0.026*	0.029*	0.026*	0.021*	0.018*
200	0.029*	0.015*	0.018*	0.020	0.041	0.031*	0.024*	0.020*	0.026*	0.022*	0.019*	0.017*
250	0.021*	0.012*	0.018†	0.020	0.038	0.026*	0.020*	0.016*	0.023*	0.020*	0.018*	0.016*

The numbers in the table are median values of the RMSE stemming from 10 simulations per setting, and can be contrasted with the results in Table 1 in Kastner and Huber (2020).

The symbol "*" denotes lower RMSE compared to all other results in Table 1 in Kastner and Huber (2020); "†" indicates the same forecasting ability as the best result in Kastner and Huber (2020). No symbol denotes that there is at least one result in Table 1 in Kastner and Huber (2020) which dominates our results.

Fig 21: CPU timings



Timings are in CPU seconds for a single draw from the posterior depending on the number of variables in the VAR, n. Timings are for a standard desktop i9 running fortran77 using the Lahey compiler. In all experiments, we have used $T = 224$ observations.

Appendix A: Further empirical evidence

In this Appendix, we report the full blown resampling algorithm, and also further output which complements Section 3.

A.1. Technical details

A.1.1. Univariate equations estimation: further details

Note that (2.3) entails that

$$\ln u_{i,t}^2 = h_{i,t} + \ln u_{i,t}^{*2}. \tag{A.1}$$

Conditional on $\beta_{i,t}$s, we have

$$\ln \left(y_{i,t} - \sum_{j=1}^{p} \beta_{i,t} y_{i,t-j} \right)^2 = h_{i,t} + \ln u_{i,t}^{*2}. \tag{A.2}$$

The model is linear in $h_{i,t}$. It is well known (see Kim et al., 1998) that using a Quasi-Maximum Likelihood estimator under the assumption that $\ln u_{i,t}^{*2}$ is Gaussian results in poor small-sample properties. Thus, we follow the approach suggested by Kim et al. (1998), and we approximate the distribution of $\ln u_{i,t}^{*2}$ using a mixture of normals with seven components. Thence, for each i, $h_{i,t}$s is sampled at once using the Kalman filter. In turn, conditionally on $h_{i,t}$, the model for $y_{i,t}$ has a linear state space representation in terms of $\beta_{i,t}$s. Therefore, for each i, we draw the entire vector $\beta_{i,t}$ at once, using again the Kalman filter. [16]

A.1.2. Sampling from the posterior: the MALA algorithm

Sampling from (2.14) can be done along similar lines as in the case of a fixed coefficient VARs, but with the complications arising from β_t being time-varying. We use the Metropolis Adjusted Langevin (MALA) algorithm by Roberts and Rosenthal (1998) (see also Girolami and Calderhead, 2011), which is likely to be more efficient than an ordinary Random Walk Metropolis algorithm in light of the large dimensionality of θ; we refer to Section 4 for an evaluation of CPU time. Further,

[16] We point out that an alternative approach is to use the Gibbs sampler to draw from the conditional posterior distribution $p\left(\beta_{i,t}|\{h_{i,\tau}, \tau \neq t\}, \{h_{i,t}\}, \{\mathbf{y}\}_{t=1}^{T}\right)$ but this approach, although simpler, results in slower convergence and higher autocorrelation in MCMC draws.

evidence on relative numerical efficiency reported in Section 4.1 shows that the resampling approach is quite efficient.

In order to illustrate the algorithm, we begin by defining the matrix

$$G\left(\widetilde{\theta}\right) = -\frac{\partial^2}{\partial\theta\partial\theta'} \ln L\left(\left\{\mathbf{y}_t\right\}_{t=p+1}^{T} |\theta, \left\{\beta_t\right\}_{t=1}^{T}\right)\Big|_{\theta=\widetilde{\theta}} \tag{A.3}$$

computed at a generic value $\widetilde{\theta}$. The likelihood $L\left(\left\{\mathbf{y}_t\right\}_{t=p+1}^{T} |\theta\right)$ is differentiable up to any order, within the whole parameter space, due to the normality assumption; thus, by the Schwarz Lemma, $G\left(\widetilde{\theta}\right)$ is symmetric for any $\widetilde{\theta}$ within the parameter space.

Based on the definitions above, the resampling scheme is as follows:

GC-Step 1 Initialise by drawing θ_0 from $p(\theta)$, and set $k = 0$.

GC-Step 2 Randomly generate $\widetilde{\theta}$ from the proposal density

$$q\left(\widetilde{\theta}|\theta_k\right) \sim N\left[m\left(\theta_k\right), \lambda^2 I_d\right]. \tag{A.4}$$

GC-Step 3 Compute the Metropolis acceptance probability

$$A\left(\widetilde{\theta}, \theta_k\right) = \min\left\{1, \frac{p\left(\widetilde{\theta}|\left\{\mathbf{y}_t\right\}_{t=1}^{T}\right)}{p\left(\theta_k|\left\{\mathbf{y}_t\right\}_{t=1}^{T}\right)} \frac{q\left(\theta_k|\widetilde{\theta}\right)}{q\left(\widetilde{\theta}|\theta_k\right)}\right\}, \tag{A.5}$$

where the marginal posterior is obtained by integrating out $\left\{\beta_t\right\}_{t=1}^{T}$ from the joint posterior in (2.14) with respect to the posterior of $\left\{\beta_t\right\}_{t=1}^{T}$.

GC-Step 4 Draw u from a uniform distribution in $[0, 1]$, defining the acceptance rule

$$\begin{aligned} &\text{if } u \leq A\left(\widetilde{\theta}, \theta_k\right) \Longrightarrow \theta_{k+1} = \widetilde{\theta} \\ &\text{if } u > A\left(\widetilde{\theta}, \theta_k\right) \Longrightarrow \theta_{k+1} = \theta_k \end{aligned}$$

GC-Step 5 Set $k = k + 1$ and return to Step 2.

We now present the proposal density. In (A.4), the scale parameter λ is discussed later on, and the mean $m\left(\theta_k\right)$ is given by

$$m\left(\theta_k\right) = \theta_k + \frac{1}{2}\lambda^2 \nabla \ln p\left(\theta_k|\left\{\mathbf{y}_t\right\}_{t=1}^{T}\right), \tag{A.6}$$

where "∇" refers to the gradient, which is computed with respect to θ and then specialised in the value θ_k (we use the same notation as Nemeth et al., 2016). In (A.6), the main difficulty is the computation of

$$\nabla \ln p\left(\theta_k | \{\mathbf{y}_t\}_{t=1}^T\right) = \nabla \ln L\left(\{\mathbf{y}_t\}_{t=1}^T | \theta_k\right) + \nabla \ln p\left(\theta_k\right). \tag{A.7}$$

Assuming - as is typical, see Nemeth et al. (2016) - that $\nabla \ln p\left(\theta_k\right)$ is known, this amounts to estimating $\nabla \ln L\left(\{\mathbf{y}_t\}_{t=1}^T | \theta_k\right)$. Note that by Fisher's identity (Cappé et al., 2009), it holds that

$$\nabla \ln L\left(\{\mathbf{y}_t\}_{t=1}^T | \theta_k\right) = E_{\{\beta_t\}_{t=1}^T}\left[\nabla \ln p\left(\{\mathbf{y}_t\}_{t=1}^T ; \{\beta_t\}_{t=1}^T | \theta_k\right)\right], \tag{A.8}$$

where $E_{\{\beta_t\}_{t=1}^T}$ denotes expectation taken with respect to $p\left(\{\beta_t\}_{t=1}^T | \{\mathbf{y}_t\}_{t=1}^T\right)$, with

$$\nabla \ln p\left(\{\mathbf{y}_t\}_{t=1}^T ; \{\beta_t\}_{t=1}^T | \theta_k\right) = \sum_{t=1}^T \nabla \ln p\left(\mathbf{y}_T | \{\mathbf{y}_t\}_{t=1}^{T-1} ; \{\beta_t\}_{t=1}^{T-1} ; \theta_k\right) + \nabla \ln p\left(\beta_T | \{\beta_t\}_{t=1}^{T-1} ; \theta_k\right). \tag{A.9}$$

Estimation of $\nabla \ln L\left(\{\mathbf{y}_t\}_{t=1}^T | \theta_k\right)$ uses the Rao-Blackwellised estimator proposed in Nemeth et al. (2016), as described below.

RB-Step 1 Initialise by sampling the particles $\beta_1^{(j)}$, $1 \leq j \leq M$, from $p\left(\beta_1\right)$, and set

$$w_1^{(j)} = \frac{p\left(\mathbf{y}_1 | \beta_1^{(j)}\right)}{\sum_{j=1}^M p\left(\mathbf{y}_1 | \beta_1^{(j)}\right)},$$

computing also the estimate

$$\nabla \ln \widehat{L}\left(\mathbf{y}_1 | \theta_k\right) = \nabla \ln p\left(\mathbf{y}_1 | \beta_1^{(j)} ; \theta_k\right) + \nabla \ln p\left(\beta_1\right).$$

RB-Step 2 For $t = 2, ..., T$, assume a set of weights $\left\{\xi_t^{(j)}\right\}_{j=1}^M$ and a proposal density $q\left(\beta_t | \beta_{t-1}^{(j)} ; \mathbf{y}_t ; \theta_k\right)$, and

(i) sample a set of indices $\{k_j\}_{j=1}^M$ from $1 \leq j \leq M$, with probabilities $\left\{\xi_t^{(j)}\right\}_{j=1}^M$;

(ii) define the updated weights

$$w_t^{(j)} = \frac{\widetilde{w}_t^{(j)}}{\sum_{j=1}^M \widetilde{w}_t^{(j)}},$$

where

$$\widetilde{w}_t^{(j)} = \frac{\widetilde{w}_{t-1}^{(k_j)} p\left(\mathbf{y}_t|\beta_t^{(j)};\theta_k\right) p\left(\beta_t^{(j)}|\beta_{t-1}^{(k_j)};\theta_k\right)}{\xi_t^{(k_j)} q\left(\beta_t^{(j)}|\beta_{t-1}^{(k_j)};\mathbf{y}_t;\theta_k\right)};$$

(iii) compute

$$m_t^{(j)} = \zeta m_{t-1}^{(k_j)} + (1-\zeta)\sum_{j=1}^{M} w_{t-1}^{(j)} m_{t-1}^{(j)} + \ln p\left(\mathbf{y}_t|\beta_t^{(j)};\theta_k\right) + \nabla \ln p\left(\beta_t^{(j)}|\beta_{t-1}^{(k_j)};\theta_k\right). \quad \text{(A.10)}$$

RB-Step 3 Compute

$$\nabla \ln \widehat{L}\left(\{\mathbf{y}_s\}_{s=1}^{t}|\theta_k\right) = \sum_{j=1}^{M} w_t^{(j)} m_t^{(j)}.$$

The output of the algorithm is the estimate $\nabla \ln \widehat{L}\left(\{\mathbf{y}_t\}_{t=1}^{T}|\theta_k\right)$, which can then be plugged in (A.7). As indicated by Nemeth et al. (2016), the algorithm also readily affords the computation of other important quantities such as the predictive likelihood, etc.

Finally, we note that when estimating VARMA models as in Section 4.1, we use essentially the same algorithm as above. The only difference is the proposal density employed in *GC-Step 2*, which in the case of a fixed parameter VARMA is defined as

$$q\left(\widetilde{\theta}|\theta_k\right) \sim N\left(\theta_k, \varepsilon^2 G^{-1}\left(\theta_k\right)\right), \quad \text{(A.11)}$$

where ε is chosen so as to pre-determine, roughly, the acceptance ratio in *GC-Step 4* of the algorithm, setting it to around $25-30$. [17]

A.2. Further results on predictive ability

In this section, we compare our methodology against the results in Koop and Korobilis (2013), also drawing a systematic analysis of the relative performance of several variants of our methodology. We use the same data as Koop and Korobilis (2013), namely $n = 25$ US macroeconomic variables (see Table A.7) running from 1959:Q1 to 2010:Q2. The focus of our exercise is the prediction of three series: inflation, GDP and interest rate. Given that all series are transformed into first differences in order to ensure stationarity, our model predicts the percentage change in inflation (the second log

[17] In order to carry out this step, we use $600,000$ replications, with a burn-in period of $100,000$ replications.

difference of CPI), GDP growth (the log difference of real GDP) and the change in the interest rate (the difference of the Fed funds rate). To ensure a fair comparison with Koop and Korobilis (2013), we have also demeaned all variables and then standardised them (we use the standard deviation calculated from 1959Q1 through 1969Q4). The forecasting horizon is 1970:Q1 till 2010:Q2.

Results are in Tables A.1-A.3, where we report the relative Mean Squared Forecast Errors (MSFE) when using the various VAR specifications to predict GDP, inflation and interest rate (respectively). The numbers in the tables are the MSFE relative to the TVP-VAR-DMA model in Koop and Korobilis (2013), which is therefore our baseline model.

Broadly speaking, results show that our methodology affords good forecasting ability especially for shorter horizons; a notable exception is the poor performance of the TVP-VAR for GDP, although using strategies *S1* and *S2* yields a marked improvement. Indeed, there is no clearly superior model, although the results seem to make a case for heteroskedastic VARs (nonetheless, homoskedastic VARs with GMCM show very good results). In general, using GMCM (and determining the G) works better than restricting G to 1 (as could be expected). Similarly, reducing the dimensionality of the copula model with either strategy *S1* or *S2* generally improves forecasting ability. Although Bayesian compression works well, it does not seem to yield a uniformly superior predictive performance than the univariate models proposed in equation (2.2). As a final point, a distinctive feature of Koop and Korobilis (2013) is that the authors propose to use "forgetting factors"(a procedure not dissimilar to an exponentially weighted moving average); thus, they avoid estimating the covariance matrix of the VAR and the covariance matrix of the time-varying coefficients. In our case, we are dealing with univariate models, and therefore we do not have to estimate covariance matrices.

A.3. Further evidence on VARMA models

In order to complement the findings in Section 4.1, we consider the estimation and the predictive ability of a VARMA model, applied to US macro data; our exercise is based on the one in Chan et al. (2016).

In this application, we do not consider time variation in the coefficients or in the volatilities: the purpose of our analysis is only to show the computational advantages of our procedure. We follow Chan et al. (2016), using the same dataset. The data are quarterly US macroeconomic variables, ranging from 1959:Q1 to 2013:Q4. All data are first-differenced to obtain stationarity, as is customarily recommended in this type of analysis (see Carriero et al., 2015) - see Table A.4 for a list of

variables employed. In order to ensure a meaningful comparison with Chan et al. (2016), we consider three models of increasing dimension, with $n = 3, 7$ and 12. In particular, for the model with $n = 3$ variables, we have used: Real GDP, CPI (All Items) and Effective Federal Funds Rate. For the case $n = 7$, the variables are the ones in the previous model, plus: Average Hourly Earnings: Manufacturing, M2 Money Stock, Spot Oil Price (WTI), and S&P 500 Index. Finally, for the model with $n = 12$, the additional variables are Real Personal Consumption, Housing Starts (total), Real GPDI, ISM PMI Composite Index and All Employees (Total nonfarm). [18]

In Table A.5, we compare models using the sum of the log predictive likelihood as a model selection criterion based on forecasting accuracy (see also an insightful contribution by Geweke and Amisano, 2014). [19] As can be noted, our methodology yields results which, broadly speaking, are as good as the ones in Chan et al. (2016); a distinctive advantage is that our procedure is simpler and quicker to implement (CPU time is always below 1 second using mainframe). Similarly to Chan et al. (2016), we note that VARMA models seem to offer better predictive power; yet, remarkably, our VAR(4) based on using the dimension reduction strategy denoted as *S2* is at least as good (in fact, marginally better) than both our VARMA$(4, 4)$, and the one in Chan et al. (2016). Interestingly, it can be noted that model averaging yields an even better result. This can be viewed as an indication that none of the models under consideration is correctly specified, which makes the case for model averaging.

We have also run a complementary exercise, in which we estimate the three models (with $n = 3, 7$ and 12 variables), and then consider the predictive likelihood for the variables that are common to the three specifications - that is, Real GDP, CPI (All Items) and Effective Federal Funds Rate. The results are presented in Table A.6, where it can be noted that the forecasting ability (slightly) improves as n increases, reinforcing the case for large VARs. Especially when $n = 12$ is considered, our approach to the estimation of VARMA$(4, 4)$ delivers the best predictive ability - again, this result should be read in conjunction with the decidedly lower CPU time of our approach.

Finally, we have also carried out impulse response analysis; whilst we do not report results (which are available upon request), we found that impulse response functions behave in a very similar way to those in Figures 1 and 2 in Chan et al. (2016).

[18] A complete description of the dataset is available from the authors.

[19] Details and formulas (also for other indicators) are available upon request.

A.4. Impulse responses from a large VAR - comparison with Giannone et al. (2015)

We compare our methodology with the results from Giannone et al. (2015), who estimate a large Bayesian VAR with $n = 23$ variables (see Table A.8 in the Supplement for a list of the variables), focusing primarily on the selection of an appropriate prior. We carry out a forecasting exercise, based on the same specifications as in Giannone et al. (2015); results - and a brief description of the exercise - are reported in Table A.9.

According to Table A.9, our approach affords very good predictive ability; as a general comment, forecasting ability seems to improve decidedly as the dimension of the VAR increases, which again makes the case for large VARs.

As well as comparing predictive ability, we have computed the Impulse Response Functions (IRFs) for all the variables in the large version of the VAR. Our results are in Figures A.3-A.5, which correspond exactly to IRFs reported in Figures 2-4 in Giannone et al. (2015). The pattern of the impulse responses is very similar, although there seem to be important differences in uncertainty (compare the 95% Bayes probability intervals for impulse responses). Note that we have followed exactly the same identification scheme as in Giannone et al. (2015), and therefore our findings are the same as in that paper. In particular, an increase in the Federal Funds rate results in a contraction of all variables related to economic activity (see e.g. the GDP and employment), in monetary aggregates, stock prices and prices (with a delay); thus, even in our case, there is no "price puzzle" contrary to small VARs.

Finally, Giannone et al. (2015) also compare the impulse response functions of their model against a "real" counterpart. To this end, they simulate data using a medium-sized Dynamic Stochastic General Equilibrium (DSGE) model, and compare the IRFs from the theoretical model with those from their hierarchical prior BVAR and a flat prior VAR. More specifically, Giannone et al. (2015) run a Monte Carlo exercise with 500 replications, generating time series of length $T = 200$ (quarters) for seven variables: output (Y), consumption (C), investment (I), hours worked (H), wages (W), prices (P) and the short-term interest rate (R). [20] We have carried out the same exercise, using exactly the same specification and parameterisation for the DSGE; likewise, for each variable, we have computed the IRFs at different horizons, and the MSE across replications. In particular, in Figure A.6, we report the ratio of the MSEs obtained using a flat prior VAR and our approach -

[20] The DSGE model is the same as in Justiniano et al. (2010), with the exception that the private sector conditions on monetary policy is set as in Christiano et al. (1996).

this can be directly compared with Figure 6 in Giannone et al. (2015), where, broadly speaking, similar outcomes are found. Even in our case such ratio is never lower than 1 in any of the cases considered, being in fact substantially higher in most cases. Indeed, depending on the horizon, all IRFs (except for the case of the short-term interest rate, R) can be at least three times as accurate. Note also that our IRFs for the short-term interest rate R are not as accurate as the others. This is found also in Giannone et al. (2015), who ascribe their results, in this case, to the need for a more sophisticated prior. We note however that, in our case, the MSE ratio is always greater than one even for R, possibly because our methodology does not rely mainly on the choice of a prior, but hinges around the usage of univariate models plus a copula. Based on Figure A.6, it is possible to conclude that the approach proposed in this paper works very well in practice.

TABLE A.1
Relative MSFE at various horizons h - predictions for GDP.

GDP	Forecast horizon							
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\beta_{T+h}\sim RW$	1.17	1.17	1.10	1.12	1.12	1.10	1.13	1.12
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\kappa=0.96$, $\alpha=0.99$	1.04	1.05	1.01	1.02	1.01	1.00	1.02	1.01
VAR, Heteroskedastic$^{(a)}$	1.10	1.10	1.03	1.04	1.05	1.01	1.06	1.04
VAR, Homoskedastic$^{(a)}$	1.13	1.03	1.03	1.05	1.08	1.06	1.10	1.08
TVP-VAR, this paper	1.12	1.14	1.15	1.17	1.20	1.20	1.22	1.25
TVP-VAR, using *S1*	0.90	0.92	0.94	0.94	0.96	0.98	1.00	1.02
TVP-VAR, using *S2*	0.89	0.91	0.92	0.92	0.90	0.94	0.92	0.97
VAR - Heteroskedastic, this paper ($G=1$)	1.01	0.99	0.99	0.97	0.98	0.98	1.01	1.03
VAR - Heteroskedastic, GMCM	0.91	0.93	0.95	0.99	1.02	1.04	1.07	1.09
VAR - Heteroskedastic, Bayes compression	0.84	0.87	0.87	0.90	0.90	0.92	0.92	0.93
VAR - Homoskedastic, this paper ($G=1$)	0.94	0.96	0.99	1.01	1.05	1.07	1.09	1.14
VAR - Homoskedastic, GMCM	0.89	0.91	0.91	0.92	0.94	0.95	0.97	0.97
VAR - Homoskedastic, Bayes compression	0.90	0.91	0.93	0.93	0.95	0.98	1.09	1.14

In each column, h denotes the horizon for which the prediction has been computed.
The first panel of the table contains the results for several models proposed in Koop and Korobilis (2013); specifically, the superscript "$^{(a)}$" refers to the models considered in Table 1 in Koop and Korobilis (2013). In the first row, "RW" denotes a random walk law of motion for the time-varying parameters; the parameters in the second row are defined in Koop and Korobilis (2013).
In the second panel of the table, we use the models proposed in this paper. In particular, in the model denoted as "GMCM", we use the mixed Gaussian copula model defined in (2.10); G has been selected equal to 4 based on the values of the marginal likelihood. In the row above, we have used $G=1$, with no model selection. In the row denoted as "Bayes compression", we have used the methodology proposed by Guhaniyogi and Dunson (2015), averaging across $10,000$ sets of weights, derived from marginal likelihoods converted into posterior probabilities.

TABLE A.2
Relative MSFE at various horizons h - predictions for inflation.

Inflation	Forecast horizon							
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\beta_{T+h}\sim RW$	1.03	1.02	1.00	1.01	1.00	1.00	1.00	1.02
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\kappa=0.96$, $\alpha=0.99$	1.03	1.02	1.03	1.04	1.00	1.00	1.02	1.00
VAR, Heteroskedastic$^{(a)}$	1.03	1.02	1.01	1.02	1.01	1.00	1.01	1.02
VAR, Homoskedastic$^{(a)}$	1.04	1.06	1.03	1.02	1.00	1.03	1.01	1.01
TVP-VAR, this paper	0.85	0.92	0.97	1.00	1.00	1.02	1.02	1.04
TVP-VAR, using *S1*	0.95	0.96	0.97	0.98	1.00	1.00	1.01	1.03
TVP-VAR, using *S2*	0.93	0.95	0.95	0.97	0.97	0.98	1.00	1.00
VAR - Heteroskedastic, this paper ($G=1$)	0.81	0.83	0.83	0.85	0.87	0.87	0.89	0.92
VAR - Heteroskedastic, GMCM	0.74	0.74	0.75	0.75	0.77	0.77	0.82	0.84
VAR - Heteroskedastic, Bayes compression	0.97	0.99	1.00	1.00	1.02	1.04	1.07	1.09
VAR - Homoskedastic, this paper ($G=1$)	1.01	1.03	1.03	1.05	1.07	1.07	1.09	1.13
VAR - Homoskedastic, GMCM	0.82	0.82	0.80	0.82	0.94	0.97	1.00	1.03
VAR - Homoskedastic, Bayes compression	1.05	1.05	1.07	1.09	1.10	1.10	1.12	1.15

The models considered in the table are the same as in Table A.1.

TABLE A.3
Relative MSFE at various horizons h - predictions for interest rates.

Interest Rate	Forecast horizon							
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\beta_{T+h}\sim RW$	1.11	1.03	1.02	1.02	1.02	1.02	1.01	0.99
TVP-VAR$^{(a)}$, $\lambda=0.99$, $\kappa=0.96$, $\alpha=0.99$	1.10	1.09	1.05	1.08	1.02	1.01	1.03	1.02
VAR, Heteroskedastic$^{(a)}$	1.10	1.01	1.01	1.02	1.01	1.01	1.03	1.03
VAR, Homoskedastic$^{(a)}$	1.11	1.07	1.11	1.11	1.03	1.03	1.09	1.08
TVP-VAR, this paper	0.93	0.95	0.95	0.97	1.00	1.00	1.01	1.03
TVP-VAR, using *S1*	0.90	0.92	0.93	0.94	0.95	0.99	1.02	1.03
TVP-VAR, using *S2*	0.91	0.93	0.93	0.95	0.97	0.97	0.97	1.03
VAR - Heteroskedastic, this paper ($G=1$)	0.90	0.92	0.92	0.94	0.94	0.96	0.98	0.99
VAR - Heteroskedastic, GMCM	0.83	0.85	0.86	0.86	0.86	0.88	0.91	0.93
VAR - Heteroskedastic, Bayes compression	0.80	0.80	0.81	0.83	0.83	0.85	0.87	0.90
VAR - Homoskedastic, this paper ($G=1$)	0.94	0.95	0.97	0.97	1.03	1.05	1.07	1.08
VAR - Homoskedastic, GMCM	0.87	0.91	0.91	0.93	0.96	0.99	1.01	1.04
VAR - Homoskedastic, Bayes compression	0.97	0.97	1.00	1.02	1.02	1.03	1.05	1.07

The models considered in the table are the same as in Table A.1.

TABLE A.4
List of variables employed in Chan et al. (2016)

GDP	GDP	GDP
CPI (All Items)	CPI (All Items)	CPI (All Items)
Effective Fed Fund Rate	Effective Fed Fund Rate	Effective Fed Fund Rate
	Average Hourly Earnings (Manifacturing)	Average Hourly Earnings (Manifacturing)
	M2	M2
	Spot Oil Price (WTI)	Spot Oil Price (WTI)
	S&P500 Index	S&P500 Index
		Real Personal Consumption
		Housing Starts (total)
		Real GDPI
		ISM PMI Composite Index
		All Employees (total nonfarm)

TABLE A.5
Sum of log predictive likelihoods for various specifications

	$n = 3$	$n = 7$	$n = 12$	RNE
Chan et al. (2016)				
$VARMA(4,4)^{(a)}$	-182.5	-401.9	-492.3	
$VARMA(4,4)^{(b)}$	-188.1	-406.0	-504.2	
$VAR(4)$	-187.1	-406.7	-496.9	
This paper				
$VAR(4)$	-187.1	-406.8	-496.9	0.423
$VARMA(4,4)$	-182.5	-401.9	-492.1	0.363
$VAR(4)$ *with* $S1$	-187.0	-406.7	-496.7	0.457
$VAR(4)$ *with* $S2$	-182.3	-401.7	-492.0	0.405
Model average	-179.4	-401.1	-490.0	0.443

The table contains the sums of the log predictive likelihood for various specifications - in panel "Chan et al. (2016)", we consider various VARMA specifications using the methodology proposed in Chan et al. (2016) - the superscripts "(a)" and "(b)" refer to two different prior specifications; in panel "This paper", we have considered various specifications based on our methodology.
When using the two strategies *S1* and *S2* described in Section 2.2.1, G has been selected by maximising the integrated likelihood as a selection criterion; in all cases, we this has led to the choice $G = 3$.
In the "Model average" row, we use a standard Bayesian model averaging, based on weights computed from the posterior model probabilities (details are available upon request); we have used $10,000$ sets of weights.
The column denoted as RNE contains the Relative Numerical Efficiency as defined in Geweke (1989).

TABLE A.6
Sum of log predictive likelihoods - predictions of GDP, CPI and interest rates

	$n = 3$	$n = 7$	$n = 12$
Chan et al. (2016)			
$VARMA(4,4)^{(a)}$	-182.5	-182.2	-181.1
$VARMA(4,4)^{(b)}$	-188.1	-185.4	-187.4
$VAR(4)$	-187.1	-187.2	-191.0
This paper			
$VAR(4)$	-183.1	-184.2	-189.0
$VARMA(4,4)$	-180.5	-180.2	-180.0

The table contains the sum of the log predictive likelihoods based on the predictive densities of the first three variables (Real GDP, CPI and Interest rate). All other specifications are the same as in Table A.5.
Note that we do not report the weighted average, since the posterior model probabilities favour only one model.

TABLE A.7
List of variables employed in Koop and Korobilis (2013)

GDP	Industrial production	US/UK exchange rate
CPI	Capacity utilisation	Real personal consumption expenditures
Fed Funds rate	Unemployment rate	Total nonfarm payroll
NAPM CPI	Housing starts	ISM Manifacturing (PMI composite)
Borrowing from Fed	Producer price index	ISM Manifacturing (New orders)
S&P500	Average hourly earnings	Output per hour
M2 money stock	M1 money stock	
Personal income	Spot oil price	
Real GDPI	10-year T-bill	

Fig A.1: Sampling distributions of the relative MSFEs (forecasting horizon: $h = 1$ periods) - the benchmark is the TVP-VAR-DMA model in Koop and Korobilis (2013)

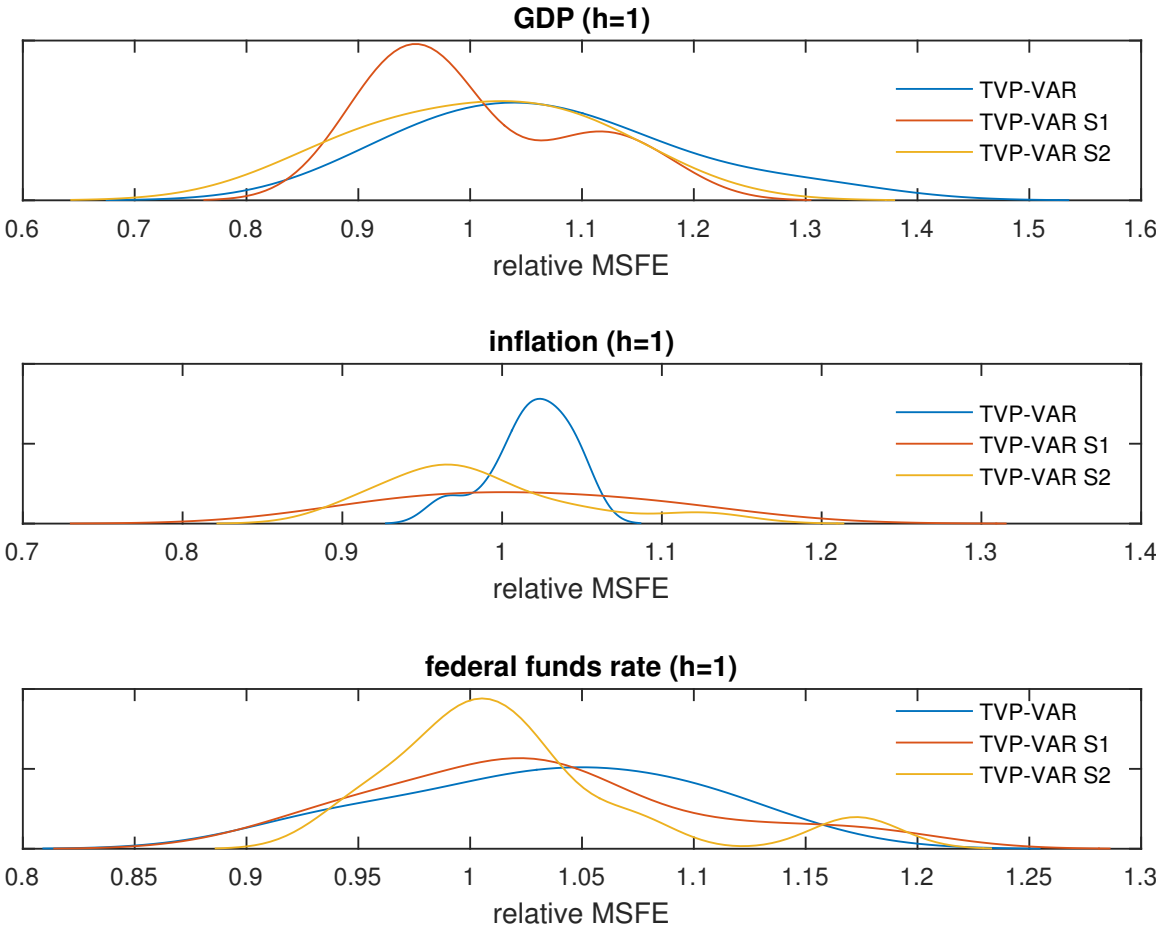


Fig A.2: Sampling distributions of the relative MSFEs (forecasting horizon: $h = 4$ periods) - the benchmark is the TVP-VAR-DMA model in Koop and Korobilis (2013)

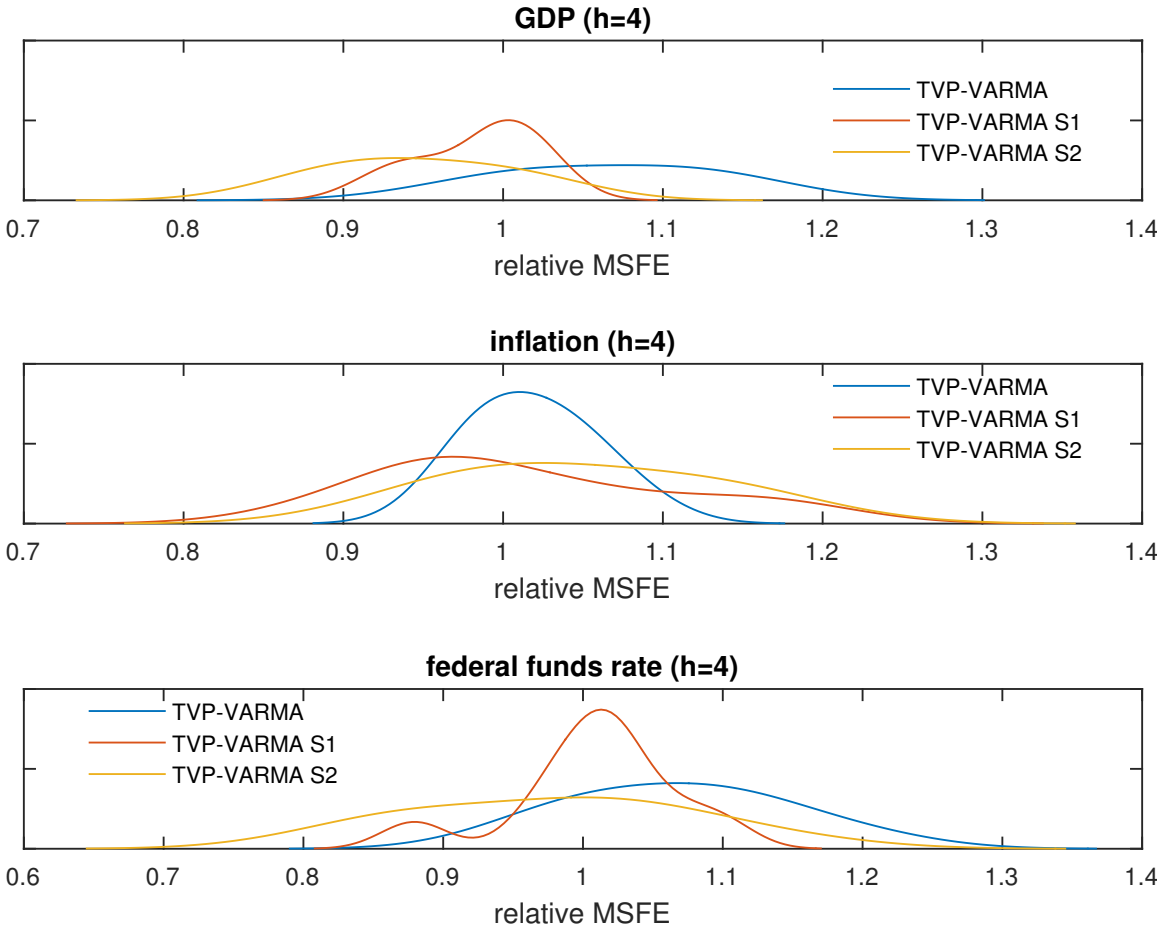


TABLE A.8
List of variables employed in Giannone et al. (2015)

Small BVAR	Medium BVAR	Large BVAR
Real GDP	Real GDP	Real GDP
GDP deflator	GDP deflator	GDP deflator
Federal Funds rate	Federal Funds rate	Federal Funds rate
	Real consumption	Real consumption
	Real investment	Real investment
	Residential investment	Residential investment
	Hours worked	Hours worked
	Real compensation per hour	Real compensation per hour
		Commodity Price
		Industrial Production
		Employment
		Unemployment
		CPI
		Non residential investment
		Personal consumption expenditures
		Gross private domestic investment
		Capacity utilisation
		Consumer expectations
		One year bond rate
		Five year bond rate
		SP500
		Effective exchange rate
		M2

TABLE A.9
Comparison with Giannone et al. (2015): Mean Squared Forecast Errors

		Small		Medium		Large		Factor Augmented Model	
Horizons	Variables	*BVAR*	*this paper*	*BVAR*	*this paper*	*BVAR*	*this paper*	*BVAR*	*this paper*
one quarter	Real GDP	9.61	4.71	7.97	5.13	8.18	4.14	7.29	3.75
	GDP Deflator	1.32	0.82	1.35	0.74	1.10	0.53	1.14	1.06
	Fed Funds Rate	1.04	0.45	1.03	0.83	1.00	0.44	1.25	1.10
one year	Real GDP	3.85	3.17	3.42	1.81	3.97	0.81	3.52	0.81
	GDP Deflator	1.45	0.77	1.58	0.72	0.96	0.58	1.01	0.88
	Fed Funds Rate	0.32	0.28	0.31	0.15	0.36	0.10	0.32	0.20

The table reports the Mean Squared Forecast Errors (MSFE) for each variable and each horizon. We compare the BVARs proposed in Giannone et al. (2015) against our methodology for three specifications (a small, medium and large VAR), and for the factor augmented model in Giannone et al. (2015) - see Table A.8 in the Supplement for a list of variables in each model.
The MSFE for the BVAR are the same as in Table 2 in Giannone et al. (2015); we refer to that paper for a description of models and methodologies. Under the heading "this paper" we have used our methodology based on a flat-prior VAR where each equation is reduced to an AR(1) model.
As in Giannone et al. (2015), the evaluation sample is 1975Q1 - 2008Q4 for the one quarter ahead forecasts and 1975Q4 - 2008Q4 for the one year ahead forecasts.

Fig A.3: Impulse responses for the VAR in Section A.4

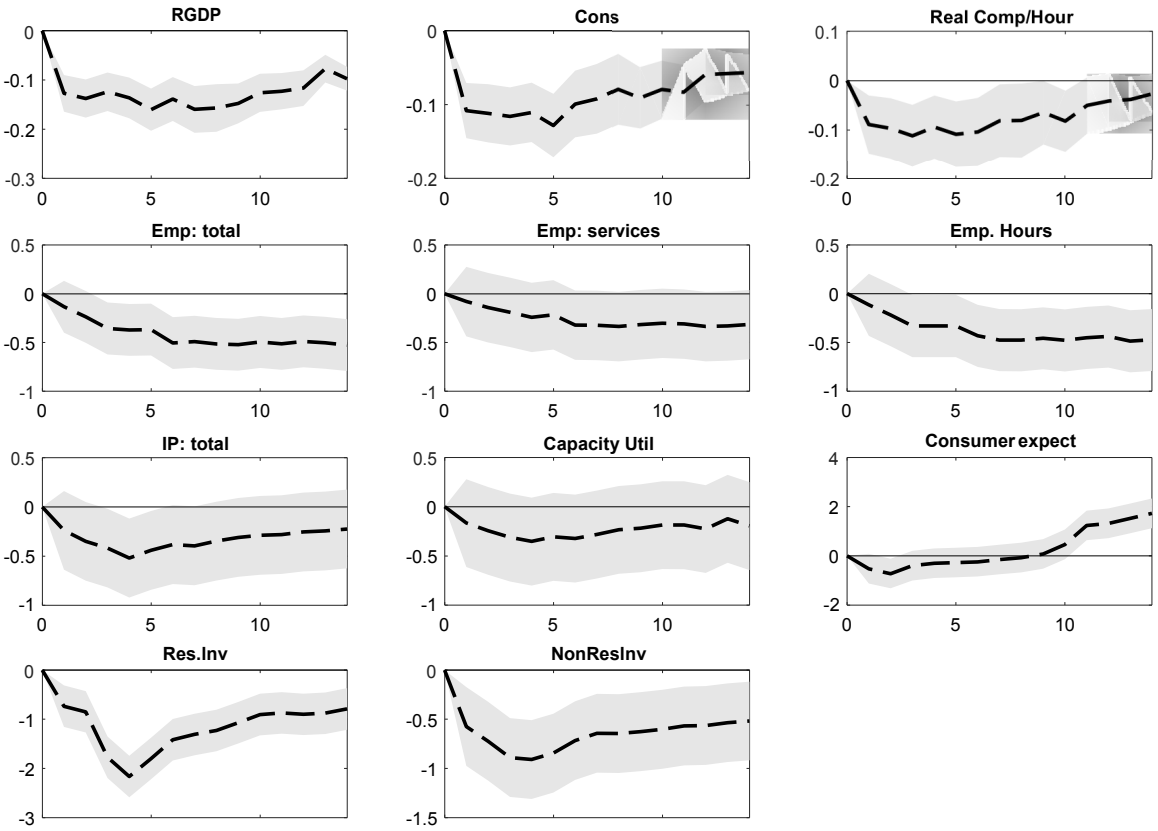


Fig A.4: Impulse responses for the VAR in Section A.4

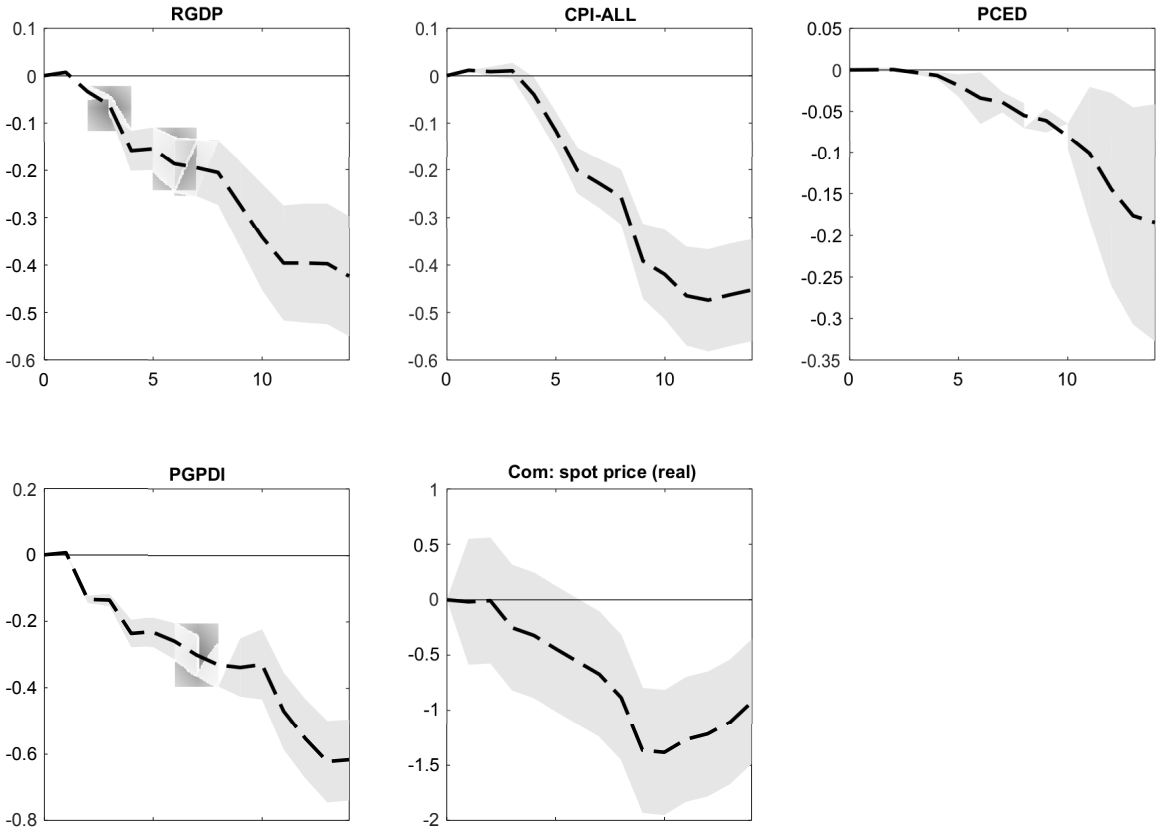


Fig A.5: Impulse responses for the VAR in Section A.4

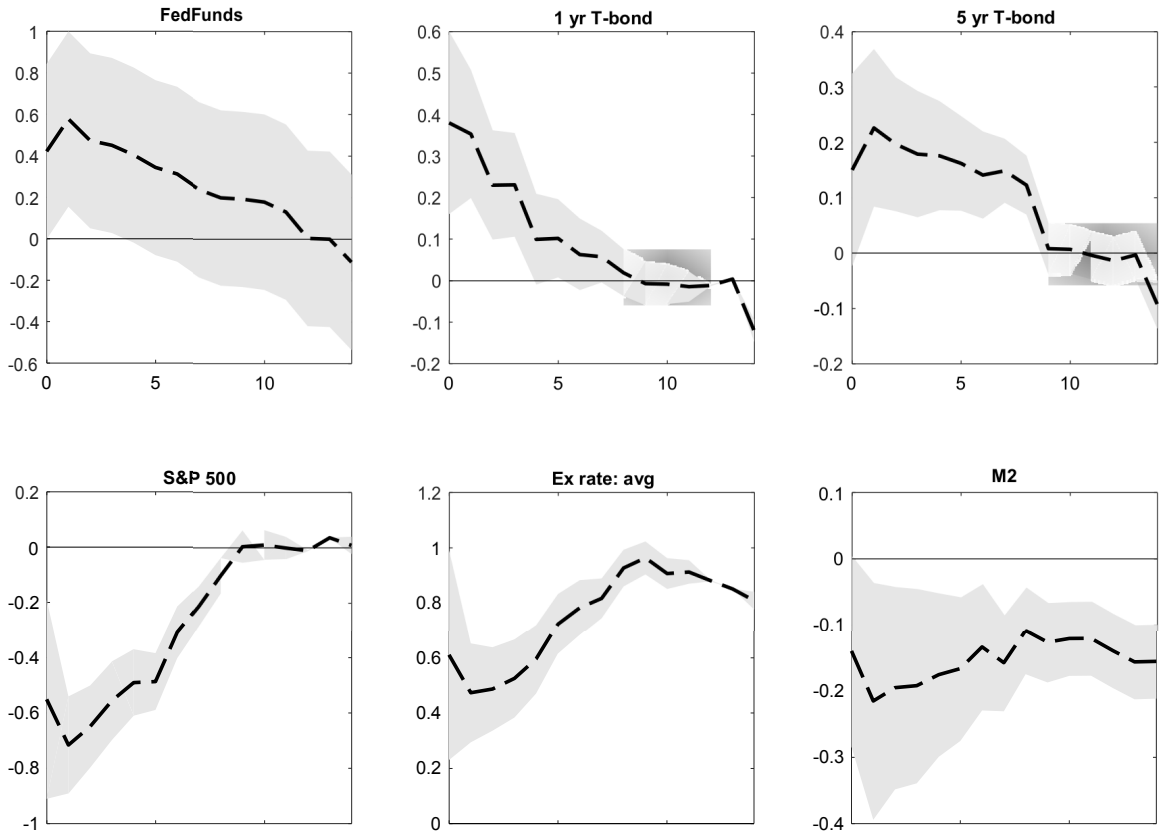


Fig A.6: Ratios of MSFEs for IRFs at various horizons