



Alternatives to large VAR, VARMA and multivariate stochastic volatility models

Mike G. Tsionas



DECEMBER 2016 I

BANK OF GREECE Economic Analysis and Research Department – Special Studies Division 21, E. Venizelos Avenue GR-102 50 Athens Tel: +30210-320 3610 Fax: +30210-320 2432

www.bankofgreece.gr

Printed in Athens, Greece at the Bank of Greece Printing Works. All rights reserved. Reproduction for educational and non-commercial purposes is permitted provided that the source is acknowledged.

ISSN 1109-6691

ALTERNATIVES TO LARGE VAR, VARMA AND MULTIVARIATE STOCHASTIC VOLATILITY MODELS

Mike G. Tsionas Athens University of Economics and Business

ABSTRACT

In this paper, our proposal is to combine univariate ARMA models to produce a variant of the VARMA model that is much more easily implementable and does not involve certain complications. The original model is reduced to a series of univariate problems and a copula - like term (a mixture-of-normals densities) is introduced to handle dependence. Since the univariate problems are easy to handle by MCMC or other techniques, computations can be parallelized easily, and only univariate distribution functions are needed, which are quite often available in closed form. The results from parallel MCMC or other posterior simulators can then be taken together and use simple sampling - resampling to obtain a draw from the exact posterior which includes the copula - like term. We avoid optimization of the parameters entering the copula mixture form as its parameters are optimized only once before MCMC begins. We apply the new techniques in three types of challenging problems. Large timevector autoregressions (TVP-VAR) varying parameter with nearly 100 macroeconomic variables, multivariate ARMA models with 25 macroeconomic variables and multivariate stochastic volatility models with 100 stock returns. Finally, we perform impulse response analysis in the data of Giannone, Lenza, and Primiceri (2015) and compare, as they proposed with results from a dynamic stochastic general equilibrium model.

Keywords: Vector Autoregressive Moving Average models; Multivariate Stochastic Volatility models; Copula models; Bayesian analysis.

JEL Classifications: C11, C13.

Acknowledgements: The author is grateful to the Bank of Greece for its hospitality and funding the research in this paper. Excellent research assistance by Xingzhi Yao is gratefully acknowledged. The views of the paper are of the author and do not necessarily reflect those of the Bank of Greece.

Correspondence:

E. G. Tsionas Athens University of Economics and Business 76 Patission Str.,10434 Athens,Greece email: tsionas@otenet.gr

1 Introduction

Since the seminal contribution of Sims (1990, 1993) Vector Autoregressions (VAR) have become part of the standard arsenal in applied econometrics. In recent years many contributions have been made including and summarized in Koop and Korobilis (2010, 2013) and Kadiyala and Karlsson (1997).

Recently, estimation of large VARs and vector autoregressive moving average models (VARMA) has received increased attention in the literature. In the VAR front see, among many others, Banbura, Giannone and Reichlin, 2010, Carriero, Kapetanios and Marcellino, 2009, Koop, 2013, Koop and Korobilis, 2013, Korobilis, 2013, Giannone, Lenza, Momferatou and Onorante, 2014. For example, attention has been given to time-varying-parameter (TVP, e.g. Primiceri, 2005) VAR with stochastic volatility and Bayesian analysis of VARMA as alternatives to VARs in medium sized data sets, see Chan, Eisenstat and Koop (2015). Koop, Korobilis and Pettenuzzo (2015) apply the Bayesian compressed regression concept due to to Guhaniyogi and Dunson (2014) in large VARs with favorable results.

Multivariate Stochastic Volatility (MSV) models have also been proposed and estimated. With many time series these methods face serious challenges and shortcuts have been proposed. For example, in TVP-VAR covariance matrices are replaced by approximations (Koop and Korobilis, 2013). In VARMA models, certain priors aid in identification but the substantive issues are still outstanding (Chan and Eisenstat, 2013 and Chan, Eisenstat and Koop, 2015). The issue of estimating MSV models is also at large although considerable progress has been made by Creal and Tsay (2015) based on copula models, Lopes, McCulloch and Tsay (2011) using a stage-wise decomposition that relies on univariate stochastic volatility models, and Philipov and Glickman (2006a,b) using a Wishart formulation. Flores de Frutos and Serrano (2002) propose a new generalized least squares. They explicitly consider the stochastic structure of the error committed when lagged innovations are replaced with lagged residuals obtained from a long VAR. Their simulation results show that this method is similar to Mauricio's (1995) exact maximum likelihood.

In this paper, we propose alternative formulations of the original problem rather than new MCMC algorithms or special priors, identification restrictions etc. The original model is reduced to a series of univariate problems and a copula - like term (a mixture-of-normals densities) is introduced to handle dependence. Since the univariate problems are easy to handle by MCMC or other techniques, computations can be parallelized easily on high-performance computing facilities and only univariate distribution functions are needed, which are quite often available in closed form. The results from parallel MCMC or other posterior simulators can then be taken together and use simple sampling -

resampling to obtain a draw from the exact posterior which includes the copula - like term.

Our objective is to handle models whose sampling distribution is $\prod_{t=1}^{T} p(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\theta})$ when the dimensionality of both \mathbf{y}_t (say n) and the parameters $\boldsymbol{\theta}$ is large. The model can include dynamic latent variables and the restriction to one lag is, exclusively, for convenience in notation.

First, we reduce the model to a univariate specification of the form $y_{it} = g_i(y_{i,t-1}, \theta_i) + \varepsilon_{it}$ for some function g_i and error term ε_{it} , for each $i = 1, \ldots, n$. Second, dependence is taken into account using a copula - like term. Each θ_i can, presumably, be simulated easily using MCMC as we have a univariate model. Finally, the MCMC draws are re-weighted to obtain a sample from the exact copula posterior. We avoid optimization of the parameters entering the copula mixture form as its parameters are optimized only once before MCMC begins. We apply the new techniques in three types of challenging problems. First, large time-varying parameter vector autoregressions (TVP-VAR) with nearly 100 macroeconomic variables. Second, multivariate ARMA models with 25 macroeconomic variables. Third, MSV models with 100 stock returns. We should mention that handling VARMA models is quite cumbersome as it has been shown in Chan, Eisenstat and Koop, 2015). Certain invertibility conditions have to be imposed and this non-trivial at all as it involves, among other things, finding certain Kronecker indices associated with the model. For estimation of VARMA models see Metaxoglou and Smith (2007) who introduced a state-space representation for VARMA models that facilitates maximum likelihood estimation using the EM algorithm. This state-space representation is related to that in Ansley and Kohn (1983) (also Jones, 1980), which applies to VARMA models with missing data and observational errors.

A large part of the literature is based on Newbold (1974). Osborn (1977) and Hillmer and Tiao (1979) extend Newbold's (1974) method to vector MA models, and Nicholls and Hall (1979) extend it to the VARMA case. These formulations still rely on calculation of the inverse and determinant of a high-dimensional covariance matrix which is highly nonlinear in the parameters. Using operations on lower dimension matrices, Mauricio (1995, 1997) improves its computational efficiency.

In this paper, our proposal is to combine univariate ARMA models to produce a variant of the VARMA model that is much more easily implementable and does not involve complications such as invertibility, finding certain Kronecker indices etc. The univariate ARMA models are efficiently combined using a novel Multivariate Mixture-of-Normals copula. Similar ideas but in a non-parametric Bayesian context have been proposed by Nieto-Barajas and Quintana (2015), Nieto-Barajas et al. (2012), Taddy (2010), Basseti et al. (2014), Rodríguez and ter Horst (2008), Di Lucca et al. (2013).

The rest of the paper is organized as follows. The model and posterior simulation are discussed in Section 2. In Sections 3 and 4 we consider empirical applications: Large Bayesian VAR and VARMAs in section 3.1, and a larger VAR in Section 3.2. A Multivariate Stochastic Volatility is considered in Section 4.1. In Section 4.2 we take up a comparison with time-varying copulas. In section 5, we compare with Giannone, Lenza, and Primiceri (2015, GLP) who estimated a large Bayesian VAR in 22 variables and proposed a comparison of impulse response functions of a Bayesian VAR model with impulse response functions delivered from repeated simulations of a dynamic stochastic general equilibrium (DSGE) model. The final section offers some concluding remarks.

2 Models and posterior simulation

2.1 General techniques

A VAR model is given by:

$$y_t = \sum_{l=1}^{L} A_l y_{t-l} + u_t, \ t = L + 1, \dots, T,$$
(1)

where y_t is an $n \times 1$ time series. As the number of parameters is Ln^2 the VAR quickly generates a large number of parameters. When $u_t \sim N(0, \Sigma)$ we have another n^2 parameters in the error covariance matrix Σ .

Here, we describe a viable alternative. For a random variable \mathbf{y} taking values in \mathbb{R}^k , we know from Sklar's (1959) theorem that a joint density can be represented as:

$$p(\mathbf{y}) = \prod_{j=1}^{k} p_j(y_j) \cdot c(u_1, \dots, u_k),$$
(2)

where $u_j = P_j(y_j) = \int_{-\infty}^{y_j} p_j(\psi) d\psi$ and $c(u_1, \ldots, u_k)$ represents a copula density. If we think of $p_j(y_j)$ as approximation to the marginals, then given a copula we have:

$$\log p(\mathbf{y}) \simeq \sum_{j=1}^{k} \log p_j(y_j) + \log c(u_1, \dots, u_k) \equiv \log q(\mathbf{y}).$$
(3)

For the marginal distributions, in the context of a VAR model we can choose univariate AR processes:

$$y_{it} = \sum_{l=1}^{L} \beta_{il} y_{i,t-l} + v_{it}, \ v_{it} \sim \text{i.i.d.} N(0, \sigma_i^2), \ \forall i = 1, \dots, n.$$
(4)

The specification can be enriched to take account of special features of the data like, for example, different number of lags, fat tails or time-varying heteroskedasticity etc. For example, we may assume:

$$v_{it} \sim \text{indep.}\mathcal{N}(0, \sigma_{it}^2), \, \forall i = 1, \dots, n,$$
(5)

$$\sigma_{it}^2 = \alpha_i + \gamma_i \sigma_{i,t-1}^2 + \delta_i v_{i,t-1}^2, \tag{6}$$

a simple GARCH(1,1) process. As the processes in (4) and (5) are univariate, they can be modeled easily almost without restrictions. For the copula function in (2) or (3) we choose a flexible family based on a mixture of normals (henceforth, CMN). Given a multivariate distribution $f(x) = f(x_1, ..., x_d), x \in \mathbb{R}^d$, we can always model the joint distribution using a copula:

$$f(x) = \prod_{j=1}^{d} f_j(x_j)c(u),$$
(7)

where $u \in \mathscr{S}$, $\mathscr{S} = \{u \in \mathbb{R}^d | 0 \le u_j \le 1, j = 1, ..., d\}$, $u_j = F_j^{-1}(x_j), j = 1, ..., d$ is the inverse distribution function of x_j , $c(u) = c(u_1, ..., u_d)$ is a copula function and $f_j(x_j), j = 1, ..., d$ denotes the marginal densities. Of course, we have $u_j = F_j(x_j), j = 1, ..., d$. A flexible model for a copula is a mixture-of-normals:

$$c(u) = c(F_1(x_1), ..., F_d(x_d)) \equiv \psi(x; \alpha) = \sum_{g=1}^G p_g f_{N,d}(\mu_g, \Omega_g),$$
(8)

where $f_{N,d}(\mu_g, \Omega_g)$ is the density of a multivariate normal distribution with mean μ_g and covariance matrix $\Omega_g, g = 1, ..., G, p = [p_1, ..., p_G]'$ is a vector of mixing probabilities, and $\alpha \in \mathscr{A} \subseteq \mathbb{R}^{d_\alpha}$ denotes the entire vector of parameters in probabilities p, means $\mu = [\mu_1, ..., \mu_G]'$ and covariance matrices Ω_g .

Once a copula $\psi(x; \alpha)$ in (8) has been determined, the joint distribution in (7) can be determined easily. As an alternative to (8) we use a mixture of Student-*t* densities:

$$\psi(x;\alpha) = \sum_{g=1}^{G} p_g f_{t,d}(\mu_g, \Omega_g, \nu_{(g)}),$$
(9)

where $\nu = [\nu_{(1)}, ..., \nu_{(G)}]'$ is a vector of the degrees of freedom which becomes also an element of vector α .

Typically, the marginals and their distribution functions, $f_j(x_j; \beta_j)$ and $F_j(x_j; \beta_j)$, j = 1, ..., d, belong in some parametric family parametrized by the vector $\beta_j \in \mathbb{R}^{d_{\beta_j}}$, j = 1, ..., d. Typically, estimates $\hat{\beta}_j$ of β_j are obtained in a first stage where univariate analysis based on maximum likelihood, for example, is used to determined these estimates. In the second stage, we can formulate a likelihood function of the form:

$$L(\alpha; X) = \prod_{t=1}^{T} \psi(x_t; \alpha), \qquad (10)$$

where X denotes the entire data set, and ψ is given by (8) or (9). Given a prior $p(\alpha)$ we can formulate the posterior by Bayes' theorem:

$$p(\alpha|X) \propto L(\alpha; X) \cdot p(\alpha), \ \alpha \in \mathscr{A} \subseteq \mathbb{R}^{d_{\alpha}}.$$
 (11)

In turn, the posterior can be analyzed using a variety of Sequential Monte Carlo (SMC) or Markov Chain Monte Carlo (MCMC) techniques. The techniques we use are based on the Particle Filtering (PF) approach and they are provided in the Technical Appendix. Of course, the possibility is open to determine the parameters of the marginal models and the parameters of the copula in one step, using the likelihood function:

$$L(\theta; X) = \prod_{t=1}^{T} \left(\psi\left(x_t; \alpha\right) \cdot \prod_{j=1}^{d} f_j(x_{jt}; \beta_j) \right),$$
(12)

where $\theta = [\alpha', \beta'_1, ..., \beta'_d]' \in \Theta \subseteq \mathbb{R}^{d_{\theta}}$. Using a prior, $p(\theta)$, we can formulate the posterior:

$$p(\theta|X) \propto L(\theta; X) \cdot p(\theta), \ \theta \in \Theta.$$
 (13)

The substantial problem is how to deal with the dimensionality problems when d is large and, therefore, the covariance matrices $\Omega_g, g = 1, ..., G$ are high-dimensional. For parameters μ_g or ν_g such problems can be dealt with more easily as they are in \mathbb{R}^d and \mathbb{R} respectively so we can tolerate more easily their proliferation as d grows larger.

Our first strategy is the following. Suppose we leave Ω_1 unrestricted. For the remaining matrices we choose:

$$\Omega_g = h_g \Omega_{g-1} + V_g, \ g = 2, ..., G, \tag{14}$$

where $V_g = \text{diag}[v_{g1}, ..., v_{gd}]$, g = 2, ..., G are diagonal matrices $(V_1 = O)$ and $h_g > 0$, g = 2, ..., Gare scalars. For Ω_1 we can use the decomposition $\Omega_1 = C'_1 C_1$ where C_1 is lower triangular and we treat the elements $c_1 \equiv \left[c_{ij}, i \ge j, i, j = 1, ..., \frac{d(d+1)}{2}\right]$ along and below the main diagonal as unrestricted parameters. The parametrization is also in terms of h_g^* and v_{gj}^* where $h_g = \exp(h_g^*)$ and $v_{gj} = \exp(v_{gj}^*)$ (j = 1, ..., d, g = 2, ..., G) and $h_g^*, v_{gj}^* \in \mathbb{R}$, are unrestricted. These parameters are all elements of α or θ in (11) and (13) respectively. Whether or not the second-stage or single-stage posterior in (11) and (13) perform better is an empirical matter.

Our second strategy is the following. Again, we leave Ω_1 unrestricted. For any matrix V we have:

$$V = QQ' = E\Lambda^{1/2},\tag{15}$$

where E is the $d \times d$ matrix of eigenvectors and Λ is the $d \times d$ diagonal matrix of eigenvalues. Humphreys et al (2015) propose the approximation:

$$V_o = Q_o Q'_o + D, \tag{16}$$

where Q_o is a $d \times k$ matrix containing the k eigenvectors, i.e. columns, of Q that correspond to the k largest eigenvalues, and D is an $d \times d$ diagonal matrix. The approach of Humphreys et al (2015) reduces the data storage requirements to scale proportional to the trace length d rather than d^2 making the approach practical for storage.

In our setting Q_o corresponds to any of the matrices $\Omega_2, ..., \Omega_G$. Therefore we have

$$\Omega_g = Q_{o,g} Q'_{o,g} + D_g, \ g = 2, ..., G, \tag{17}$$

and we do not attempt to estimate $Q_{o,g}$. Instead we fill its elements with dk random numbers from $N(0, h_g^2)$ and we perform R searches over these random numbers along with the dimensionality parameter k in a fashion related to the Bayesian Compression (BC) literature (Guhaniyogi and Dunson, 2016). The d diagonal elements of D and h_g s are treated as unknown parameters and they are estimated. However, the problem is not so daunting any more as sufficient dimensionality reduction has taken place.

We call the two strategies above, S-1 and S-2 respectively.

3 Empirical Applications I

3.1 Large Bayesian VAR and VARMA

We follow the paper by Chan, Eisenstat and Koop (2015). They investigate the performance of their algorithm (using soft SSVS and a special prior) in a substantive empirical application involving quarterly US macroeconomic data in VARMAs of varying dimensions: n = 3, n = 7 and n = 12. The

data covers the quarters 1959:Q1 to 2013:Q4. All data is transformed to stationarity. For n = 3 the variables are Real Gross Domestic Product, Consumer Price Index: All Items, and Effective Federal Funds Rate. For n = 7 the variables are the ones we mentioned plus Average Hourly Earnings: Manufacturing, M2 Money Stock, Spot Oil Price: West Texas Interm., and S&P 500 Stock Price Index. For n = 12 the additional variables are Real Personal Consumption Exp., Housing Starts: Total, Real Gross Private Domestic Invest., ISM Manuf.: PMI Composite Index. In Table 1 we report sum of log predictive likelihoods for comparison with VARMA models for Chan, Eisenstat and Koop (2015).

Our version of the VARMA(4,4) performs at least as well as the one in Chan, Eisenstat and Koop (2015) and we see an improvement in the case with n = 12. Table 2 presents predictive likelihoods for the variables which are common to all models. This allows for a comparison of different dimensional VARMAs and VARs, at least in terms their ability to forecast inflation, output growth and the interest rate.

Again, our VARMA performs as well as the models in Chan, Eisenstat and Koop (2015) and we even have a slight improvement for n = 12. The VAR(4) models perform similarly: This fact is quite reassuring as we can be relatively confident that our approach compares favorably with more difficult to implement methods. Impulse response functions also behave quite similarly and they are even better compared to the ones reported in Figures 1 and 2 of Chan, Eisenstat and Koop (2015).

3.2 A larger VAR

A larger VAR comprised of 25 U.S. macroeconomic variables is used in Koop (2013). The data set runs from 1959:Q1 to 2010:Q2. Koop (2013) transforms all variables to stationarity. He investigates the performance of our approach in forecasting inflation, GDP and the interest rate which are the variables in our small TVP-VAR. The transformations are such that the dependent variables are the percentage change in inflation (the second log 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). He also standardized all variables by subtracting off a mean and dividing by a standard deviation. He calculates this mean and standard deviation for each variable using data from 1959Q1 through 1969Q4 (i.e. data before the forecast evaluation period). Our comparison with his results is reported in Table 3.

To implement the Heteroskedastic VAR we follow the same approach as in Koop (2013) for our marginal models and we do the same for the TVP-VAR model. Koop (2013) uses forgetting factors and does not estimate the covariance matrix of the VAR or the covariance matrix of the timevarying coefficients. Of course, in our univariate models we do not have such covariance matrices. Dependence is introduced via a Multivariate Gaussian Mixture model whose implementation in practice, is not particularly difficult.

4 Empirical Applications II

4.1 Multivariate Stochastic Volatility

For a comparison with existing MSV models we focus on McCausland, Miller and Pelletier (2012). The authors analyze daily returns of 10 currencies relative to the US dollar: the Swiss Franc (CHF), Euro (EUR), Australian Dollar (AUD), New Zealand Dollar (NZD), Mexican Peso (MXN), Brazil Real (BRL), British Pound (GBP), Canadian Dollar (CAD), Japanese Yen (JPY) and Singapore Dollar (SGD). The exchange rates are the noon spot rate obtained from the Federal Reserve Bank of New York. The sample covers the period from January 5, 1999 to December 31, 2008. The authors compute the log returns of the exchange rates and remove returns for those days when one or more of the markets was closed, giving 2503 observations for each return series.

McCausland, Miller and Pelletier (2012) propose and use a model with no factors (MSV-q0) and a model with one common factor (MSV-q1). Posterior means and s.d. of their parameters are reported in their Tables 10 and 11. In Table 4, we report percentage differences of our approach compared to theirs.

Unfortunately, relative numerical efficiency (RNE) in McCausland, Miller and Pelletier (2012) is quite small (on the average $2 \cdot 10^{-3}$ approximately). In our case with 55,000 draws the first 5,000 of which are discarded, RNE ranged from 0.212 (MXN) to 0.453 (EUR) which represents a substantial improvement. The (vast) improvement is that our sampling - resampling scheme draws practically i.i.d. from the posterior and any remaining autocorrelation is due to MCMC in the univariate models. As MCMC in the univariate models follows Kim, Shephard, and Chib (1998), autocorrelation is reduced substantially as the blocks of latent volatilities are sampled together.

4.2 A comparison with time-varying copulas

Creal and Tsay (2015) proposed copula models that have flexible dynamics and heavy tails yet remain tractable in high dimensions due to their factor structure. Their Bayesian estimation approach uses recent advances in sequential Monte Carlo methods (the particle Gibbs sampling) which can draw large blocks of latent variables efficiently and in parallel. As copulas rely on a vector $\mathbf{u}_t = [u_{t1}, \ldots, u_{tn}]', 0 \le u_{it} \le 1$, Creal and Tsay (2015) proposed the following model:

$$\mathbf{u}_{t} \sim p(\mathbf{u}_{t}, |\Lambda_{t}, \mathbf{Y}_{t}, \theta), \ t = 1, \dots, T,$$

$$\Lambda_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi} \left(\Lambda_{t} - \boldsymbol{\mu}\right) + \eta_{t}, \ \eta_{t} \sim \mathcal{N}(0, \Sigma),$$
(18)

where Λ_t is an unobserved state vector, $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tn})$ denotes the data, and θ contains all parameters of the model. From the marginals, Creal and Tsay (2015) calculate the probability integral transforms u_{it} and then use them as data to perform statistical inference in (18). Although the c.d.f's of stochastic volatility models are not known in closed-form, the authors show how to calculate the probability integral transforms using the particle filter. In this paper, we follow a similar approach.

Our approach is different in that we take the copula from the data *without* the need to perform further MCMC computations. We introduce here a variation of our basic formulation in (11):

$$p(\boldsymbol{\theta}|\boldsymbol{\mathcal{Y}}_{\tau}) \propto \left\{ \prod_{i=1}^{n} p(\mathbf{y}_{i,\tau}|\boldsymbol{\beta}_{i,\tau}) \cdot \prod_{i=1}^{n} p(\boldsymbol{\beta}_{i,\tau}|\boldsymbol{\alpha}_{\tau}) \right\} \cdot p(\boldsymbol{\alpha}_{\tau}) \cdot c(\mathbf{u}_{\tau};\boldsymbol{\alpha}_{\tau}), \tau = T^* + 1, \dots, T, \quad (19)$$

where T^* represents a certain minimum sample size, \mathcal{Y}_{τ} represents the data from observation 1 to τ and the parameters $\beta_{i,\tau}$ and α_{τ} are indexed in a similar way. Instead of building a dynamic copula as in Creal and Tsay (2015) we keep the same functional form of the copula density function and re-estimate the univariate models for horizons 1 to τ where $\tau = T^* + 1, \ldots, T$. The re-computations are easily done using MCMC in parallel. As an alternative, to estimate the model for 1 to $\tau + h$ we use sampling - resampling of the draws for 1 to τ . The exact value of h is an empirical matter, and we report some basic results below. In practice, we proceed backwards starting from t = T, then we take t = T - 1, and so on until $t = T - r := T^*$.

Creal and Tsay (2015) have collected daily equity returns and log differences in credit default swap (CDS) rates for 100 U.S. Corporations from January 2, 2008 to February 28, 2013. All 100 firms are components of the S&P 500 index. The CDS rate is the 5 year contract with the XR clause. The data on equity prices are from the Center for Research in Securities Prices (CRSP) and the data on the 5 year CDS are from the Markit Corporation. They restrict attention to days when equity markets are open. This makes for a cross-section of n = 200 series with T = 1299observations. From these nT total days, there are 2,487 observations that are randomly missing. As we do not have the CDS data we consider data from FTSE-100 for the same time period.

As in their paper, we model the marginal distribution for each of the n = 200 series using

univariate stochastic volatility models with leverage and skewed Student's-t errors for the conditional distribution subject to the qualifications in their footnote 9. The model is as follows:

$$y_{it} = W_{it}\beta_{y,i} + \gamma_{y,i}\delta_{it} + \delta_{it}^{1/2} \exp\left(\frac{h_{it}}{2}\right)\varepsilon_{y,it}, \ \varepsilon_{y,it} \sim \mathcal{N}(0,1), t = 1, \dots, T,$$

$$h_{i,t+1} = \mu_{h,i} + \phi_{h,i}(h_{t,i} - \mu_{h,i}) + \sigma_{h,i}\varepsilon_{h,it}, \ \varepsilon_{h,it} \sim \mathcal{N}(0,1),$$

$$\delta_{it} \sim IG\left(\frac{\nu_{y,i}}{2}, \frac{\nu_{y,i}}{2}\right), \ \operatorname{corr}(\varepsilon_{y,it}, \varepsilon_{h,it}) = \rho_i,$$

(20)

where $\beta_{y,i}$ are regression parameters, W_{it} is a vector of covariates, $\nu_{y,i}$ is the degrees of freedom of the Student-t distribution, $\gamma_{y,i}$ is a skewness parameter, and ρ_i represents leverage. We have successfully replicated Table 1 in Creal and Tsay (2015) containing posterior results for the univariate models in (20). Our interest in the comparison with the methodology of Creal and Tsay (2015) lies in filtered values of average conditional correlations and the simulated values $\tilde{y}_{it} = P^{-1}(u_{it})$. In the interest of brevity and space we provide as much information as we can in Table 5. For the comparison we use MCMC for each univariate model re-estimated for each observation from 100 to T. So, in terms of previous notation we have h = 0. Results for different values of h are available on request but are not reported in the interest of space: We have tried values of h=5,10,20,50 and 100 and the results were robust.

5 Impulse response analysis

In this section we use the same data as Giannone, Lenza, and Primiceri (2015, GLP) who estimated a large Bayesian VAR in 22 variables (see their Table 1, p. 441). In Figures 1-3 we report impulse responses that correspond exactly to Figures 2-4 in GLP (pp. 445-447). The pattern of impulse responses is similar which is in itself striking but there are important differences in uncertainty, viz. 95% Bayes probability intervals for impulse responses. Our findings can be summarized as follows, since we follow the same identification scheme as in GLP: "A 1-standard deviation (approximately 60 basis points) exogenous increase in the federal funds rate generates a substantial contraction in GDP, employment, and all other variables related to economic activity. Monetary aggregates also decrease on impact, indicating strong liquidity effects. Moreover, stock prices decline, the exchange rate appreciates, and the yield curve flattens. Prices decrease with a delay. Notice that with the exception of the CPI, the response to a monetary contraction, which is instead typical of VARs with small information sets" (GLP, p. 446).¹

GLP want also to compare their model in terms of impulse response functions with a "real" counterpart. As we do not have one, they use a dynamic stochastic general equilibrium model (DSGE) and assess the gains in terms of accuracy for estimating impulse responses to monetary policy shocks relative to flat-prior Bayesian VARs. In this experiment they adopt a medium-size DSGE model to simulate 500 artificial time series of length 200 (quarters) for seven variables, viz, output(Y), consumption (C), investment (I), hours worked (H), wages (W), prices (P) and the short-term interest rate (R). The DSGE model is the same as in Justiniano, Primiceri, and Tambalotti (2010) with the exception that the private sector conditions on monetary policy as in Christano et al. (2005). Here, we focus on the MSE of our copula-VARMA model and the model in GLP which was compared to a flat-prior Bayesian VAR (see their Figure 6, p. 449).

From Figure 6 in GLP (p. 449) we have the ratio of MSEs of flat-prior VAR versus their Bayesian VAR which is mostly greater than one for most horizons indicating that they approximate better the actual impulse response function of the DSGE model. Specifically, "[f]or each replication, we compute the overall error as the difference between the theoretical response and the estimated median response across variables and horizons. Then, for each variable and horizon, we take the average of the squared errors across replications (MSE). Figure 6 reports the ratio between the MSE for the flat-prior VAR and the hierarchical BVAR. Such a ratio is greater than 1 for most variables and horizons, indicating that the hierarchical BVAR yields substantial accuracy gains" (GLP, p. 449).

If $a = \frac{MSE_{VAR}}{MSE_{BVAR}}$ in GLP and $b = \frac{MSE_{VARMA}}{MSE_{BVAR}}$ in our study, the ratio $r = \frac{a}{b} = \frac{MSE_{VAR}}{MSE_{VARMA}}$ and it can be used to compare directly with the MSE in GLP. We report this ratio for various horizons in Figure 5. As the ratio turns out to be substantially greater than unity in many cases, we conclude that the approach proposed here works very well in practice according to the norm put forward by GLP. viz. by comparing directly the results obtained through repeated simulations of a DSGE model.

Concluding Remarks

In this paper we provide reasonable approximations to large VARMA and Mutivariate Stochastic Volatility Models (MSV) based on the idea that we can specify the marginal distributions as ARMA

 $^{^{1}}$ One difference in Figures 2-4 of GLP is that the gray areas refer to error bands for a flat-prior Bayesian VAR whereas in this study they reflect the 95% Bayes probability interval, computed numerically using simulation techniques.

or univariate stochastic volatility models and then use a Multivariate Mixture-of-Normals distribution to capture the dependence. We explain how we can fit multivariate mixture-of-normals in large dimensions by exploiting recent advances in covariance matrix representations and Bayesian Compression. The encouraging results that we obtain relative to existing work, are in line with the good performance of Multivariate Mixture-of-Normals in related fields, e.g. Geweke and Keane (2007) and Villani, Kohn and Giordani (2009). The new copula is found to work reasonably well relative to the exact VARMA or MSV models and its implementation using numerical MCMC techniques is not particularly difficult. The better performance of Multivariate Mixture-of-Normals copulas, especially in high dimensions, relative to their conventional counterparts, clearly implies that there are features in the data that are captured by the copula but not the original models (MSV or VARMA). In this sense, we can think of the copula models not only as approximations to the original models but rather as viable alternatives in modeling and forecasting. An application to the data by Giannone, Lenza, and Primiceri (2015) shows that the new model performs much better and is closer to results delivered by medium-sized DSGE models, a fact that reinforces our conclusion.

References

Ansley, C. F., R. Kohn, R. (1983) Exact likelihood of vector autoregressive-moving average processes with missing or aggregated data. Biometrika 70, 275–78.

Banbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions. Journal of Applied Econometrics, 25, 71-92.

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

Carriero, A., Clark, T. and Marcellino, M. (2015). Large Vector Autoregressions with asymmetric priors and time varying volatilities, manuscript.

Carriero, A., Kapetanios, G. and Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. International Journal of Forecasting, 25, 400-417.

Carriero, A., Kapetanios, G. and Marcellino, M. (2015). Structural analysis with multivariate autoregressive index models. Centre for Economic Policy Research Discussion Paper DP10801.

Chan, J.C.C., E. Eisenstat (2013). Gibbs Samplers for VARMA and Its Extensions. ANU Working Papers in Economics and Econometrics # 604.

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

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Journal of Political Economy 113, 1–45.

Creal, D.D., and R.S. Tsay (2015). High Dimensional Dynamic Stochastic Copula Models. Journal of Econometrics, in press.

Di Lucca M, Guglielmi A, Müller P, Quintana FA. 2013. A simple class of Bayesian nonparametric autoregression models. Bayesian Analysis 8, 1: 63–88.

Fang, K.-T., S. Kotz and K. W. Ng (1990). Symmetric multivariate and related distributions, Chapman and Hall.

Figueiredo, M.A.T. (2003). Adaptive sparseness for supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(9):1050-1059.

Flores de Frutos, R., G. R. Serrano (2002). A generalized least squares estimation method for VARMA models, Statistics 36(4), 303–316.

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

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014). Short-term inflation

projections: a Bayesian vector autoregressive approach. International Journal of Forecasting, 30, 635-644.

Giannone, D., Lenza, M. and Primiceri, G. (2015). Prior selection for Vector Autoregressions. Review of Economics and Statistics, 97, 436-451.

Guhaniyogi, R., D. B. Dunson (2014). Bayesian Compressed Regression, Journal of the American Statistical Association.

Hillmer, S. C. and Tiao, G. C. (1979) Likelihood function of stationary multiple autoregressive moving average models. Journal of the American Statistical Association 74, 652–60.

Humphreys, D.E., P.M. Harris, M. Rodríguez-Higuero, F.A. Mubarak, D. Zhao, K. Ojasalo (2015). Principal Component Compression Method for Covariance Matrices Used for Uncertainty Propagation. IEEE Transactions on Instrumentation and Measurement 64 (2), 356-365.

Justiniano, A., G. E. Primiceri, and A. Tambalotti, "Investment Shocks and Business Cycles," Journal of Monetary Economics 57 (2010), 132–145.

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

Jones, R. H. (1980) Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics 22, 389–95.

Lopes, H.F., McCulloch, R.E., Tsay, R.S., 2011, Cholesky stochastic volatility, August 2 2011, discussion paper.

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

Koop, G. (2013). Using VARs and TVP-VARs with Many Macroeconomic Variables. Strathclyde Discussion Paper 13-03.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. Journal of Applied Econometrics. 28, 177-203.

Koop, G., D. Korobilis (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, Foundations and Trends in Econometrics: Vol. 3: No 4, pp 267-358.

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

Koop, G., D. Korobilis, D. Pettenuzzo (2015). Bayesian Compressed Vector Autoregressions. Manuscript.

Korobilis, D. (2008). Forecasting with vector autoregressions with many predictors. Advances in Econometrics. 23, 403-431.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. Journal of Applied Econometrics 28, 204-230.

Madigan, D., Raghavan, N., Dumouchel, W. (2002). Likelihood-based data squashing: A modeling approach to instance construction. Data Mining and Knowledge Discovery, 6:173-190.

Mauricio, J. A. (1995) Exact maximum likelihood estimation of stationary vector ARMA models. Journal of the American Statistical Association 90, 282–91.

Mauricio, J. A. (1997) The exact likelihood of a vector autoregressive moving average model. Applied Statistics 46, 157–71.

McCausland, W.J., S. Miller and D. Pelletier (2012). Multivariate Stochastic Volatility. Manuscript.

Metaxoglou, K., A. Smith. (2007). Maximum Likelihood Estimation of VARMA Models Using a State-Space EM Algorithm. Journal of Time Series Analysis 28 (5), 666-685.

Newbold, P. (1974) The exact likelihood function for a mixed autoregressive moving average process. Biometrika 61, 423–426.

Nicholls, D. F. and Hall, A. D. (1979) The exact likelihood function of multivariate autoregressivemoving average models. Biometrika 66, 259–64.

Nieto-Barajas LE, Müller P, Ji Y, Lu Y, Mills GB. (2012). A time series DDP for functional proteomics profiles. Biometrics 68: 859–868.

Nieto-Barajas LE, Quintana, FQA (2016). A Bayesian Non-Parametric Dynamic AR Model for Multiple Time Series Analysis. Journal of Time Series Analysis 37 (5), 675-689.

Reich, B.J., Bondell, H., and Li, L. (2011). Sufficient dimension reduction via Bayesian mixture modeling. Biometrics, 67(3):886-895.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions (with discussion). In Bayesian Statistics 3 (Edited by J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 395-402. Oxford University Press, Oxford.

Osborn, D. R. (1977) Exact and approximate maximum likelihood estimators for vector moving average processes. Journal of the Royal Statistical Society (Series B) 39, 114–8.

Owen, A. (2003). Data squashing by empirical likelihood. Data Mining and Knowledge Discovery, 7:101-113.

Philipov, A., Glickman, 2006a, Factor multivariate stochastic volatility via Wishart processes, Econometric Reviews 25, 311-334.

Philipov, A., Glickman, 2006b, Multivariate stochastic volatility via Wishart processes, Journal of Business and Economic Statistics 24, 313-328. Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72, 821-852.

Rodríguez A, ter Horst E. (2008). Bayesian dynamic density estimation. Bayesian Analysis 3: 339–366.

Sims, C. (1980), Macroeconomics and reality. Econometrica 48, 1-48.

Sims, C. (1993), 'A nine variable probabilistic macroeconomic forecasting model'. In: J. Stock and M. Watson (eds.): Business Cycles Indicators and Forecasting. University of Chicago Press for the NBER, pp. 179–204.

Sklar, M. (1959). Fonctions de repartition a n dimensions et leurs marges. Publications Institute de Statistique, Univ. Paris 8, 229–231.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. American Statistician 46, 84-88.

Strelen, J.C., F. Nassaj (2007). Analysis and generation of random vectors with copulas. WSC '07 Proceedings of the 39th conference on Winter simulation, IEEE Press, 488-496.

Taddy M. (2010). Autoregressive mixture models for dynamic spatial poisson processes: application to tracking intensity of violent crime. Journal of the American Statistical Association 105: 1403–1417.

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

TECHNICAL APPENDIX

Particle filtering

The particle filter methodology can be applied to state space models of the general form:

$$y_T \sim p(y_t|x_t), \ s_t \sim p(s_t|s_{t-1}),$$
 (A.1)

where s_t is a state variable. For general introductions see Gordon (1997), Gordon et al. (1993), Doucet et al (2001) and Ristic et al. (2004).

Given the data Y_t the posterior distribution $p(s_t|Y_t)$ can be approximated by a set of (auxiliary) particles $\left\{s_t^{(i)}, i = 1, ..., N\right\}$ with probability weights $\left\{w_t^{(i)}, i = 1, ..., N\right\}$ where $\sum_{i=1}^{N} w_t^{(i)} = 1$. The predictive density can be approximated by:

$$p(s_{t+1}|Y_t) = \int p(s_{t+1}|s_t) p(s_t|Y_t) ds_t \simeq \sum_{i=1}^N p(s_{t+1}|s_t^{(i)}) w_t^{(i)}, \tag{A.2}$$

and the final approximation for the filtering density is:

$$p(s_{t+1}|Y_t) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|Y_t) \simeq p(y_{t+1}|s_{t+1}) \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}.$$
 (A.3)

The basic mechanism of particle filtering rests on propagating $\{s_t^{(i)}, w_t^{(i)}, i = 1, ..., N\}$ to the next step, viz. $\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i = 1, ..., N\}$ but this often suffers from the weight degeneracy problem. If parameters $\theta \in \Theta \in \Re^k$ are available, as is often the case, we follow Liu and West (2001) parameter learning takes place via a mixture of multivariate normals:

$$p(\theta|Y_t) \simeq \sum_{i=1}^N w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2 V_t),$$
(A.4)

where $\bar{\theta}_t = \sum_{i=1}^N w_t^{(i)} \theta_t^{(i)}$, and $V_t = \sum_{i=1}^N w_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t) (\theta_t^{(i)} - \bar{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0,1)$ as $a = (1-b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also Casarin and Marin (2007).

Andrieu and Roberts (2009), Flury and Shephard (2011) and Pitt et al. (2012) provide the Particle Metropolis-Hastimgs (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y|\theta)$ as $p(Y|\theta)$ is often not available in closed form.

Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say

 $L^j = \hat{p}_N(Y|\theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c|\theta^{(j)})$ yielding $L^c = \hat{p}_N(Y|\theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min\left\{1, \ \frac{p(\theta^c)L^c}{p(\theta^{(j)}L^j)} \frac{q(\theta^{(j)}|\theta^c)}{q(\theta^c|\theta^{(j)})}\right\},\tag{A.5}$$

otherwise we repeat the current draws: $\{\theta^{(j+1)}, L^{j+1}\} = \{\theta^{(j)}, L^j\}.$

Hall, Pitt and Kohn (2014) propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering (Pitt et al., 2012) used within PMCMC requires far fewer particles that the standard particle filtering algorithm to approximate $p(Y|\theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t|s_{t-1})$. When this is not possible, Hall, Pitt and Kohn (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have:

$$p(y_t|s_{t-1}) = \int p(y_t|s_t) p(s_t|s_{t-1}) ds_t,$$
(A.6)

$$p(u_t|s_{t-1}; y_t) = p(y_t|s_{t-1}, u_t)p(u_t|s_{t-1})/p(y_t|s_{t-1}).$$
(A.7)

If one can evaluate $p(y_t|s_{t-1})$ and simulate from $p(u_t|s_{t-1}; y_t)$ the filter would be fully adaptable (Pitt and Shephard, 1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t|s_{t-1})$ by matching the first two moments of $p(y_t|s_{t-1})$. So in some way we find that the approximating density $p(y_t|s_{t-1}) = N(\mathbb{E}(y_t|s_{t-1}), \mathbb{V}(y_t|s_{t-1}))$. In the second stage, we know that $p(u_t|y_t, s_{t-1}) \propto p(y_t|s_{t-1}, u_t)p(u_t)$. For $p(u_t|y_t, s_{t-1})$ we know it is multimodal so suppose it has M modes are \hat{u}_t^m , for $m = 1, \ldots, M$. For each mode we can use a Laplace approximation. Let $l(u_t) = log [p(y_t|s_{t-1}, u_t)p(u_t)]$. From the Laplace approximation we obtain:

$$l(u_t) \simeq l(\hat{u}_t^m) + \frac{1}{2}(u_t - \hat{u}_t^m)' \nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m).$$
(A.8)

Then we can construct a mixture approximation:

$$g(u_t|x_t, s_{t-1}) = \sum_{m=1}^M \lambda_m (2\pi)^{-d/2} |\Sigma_m|^{-1/2} \exp\left\{\frac{1}{2}(u_t - \hat{u}_t^m)' \Sigma_m^{-1}(u_t - \hat{u}_t^m)\right\},$$
(A.9)

where $\Sigma_m = -\left[\nabla^2 l(\hat{u}_t^m)\right]^{-1}$ and $\lambda_m \propto \exp\left\{l(u_t^m)\right\}$ with $\sum_{m=1}^M = 1$. This is done for each particle s_t^i . This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a

proposal for s_t directly from the measurement density $p(y_t|s_t)$ although Hall, Pitt and Kohn (2014) are quite right in pointing out that the state equation can be very informative.

In the standard particle filter of Gordon et al. (1993) particles are simulated through the state density $p(s_t^i|s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t|s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

- For t = 0, ..., T 1 given samples $s_t^k \sim p(s_t | Y_{1:t})$ with mass π_t^k for k = 1, ..., N.
- 1) For k = 1, ..., N compute $\omega_{t|t+1}^k = g(y_{t+1}|s_t^k)\pi_t^k, \ \pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^N \omega_{t|t+1}^i$.
- 2) For k = 1, ..., N draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t}^i(ds_t)$.
- 3) For k = 1, ..., N draw $u_{t+1}^k \sim g(u_{t+1}|\tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; u_{t+1}^k)$.
- 4) For $k = 1, \ldots, N$ compute

$$\omega_{t+1}^{k} = \frac{p(y_{t+1}|s_{t+1}^{k})p(u_{t+1}^{k})}{g(y_{t+1}|s_{t}^{k})g(u_{t+1}^{k}|\tilde{s}_{t}^{k}, y_{t+1})}, \pi_{t+1}^{k} = \frac{\omega_{t+1}^{k}}{\sum_{i=1}^{N} \omega_{t+1}^{i}}.$$
(A.10)

It should be mentioned that the estimate of likelihood from ADPF is:

$$p(Y_{1:T}) = \prod_{t=1}^{T} \left(\sum_{i=1}^{N} \omega_{t-1|t}^{i} \right) \left(N^{-1} \sum_{i=1}^{N} \omega_{t}^{i} \right).$$
(A.11)

Particle Metropolis adjusted Langevin filters

Nemeth, Sherlock and Fearnhead (2014) provide a particle version of a Metropolis adjusted Langevin algorithm (MALA). In Sequential Monte Carlo we are interested in approximating $p(s_t|Y_{1:t}, \theta)$. Given that:

$$p(s_t|Y_{1:t},\theta) \propto g(y_t|x_t,\theta) \int f(s_t|s_{t-1},\theta) p(s_{t-1}|y_{1:t-1},\theta) ds_{t-1},$$
(A.12)

where $p(s_{t-1}|y_{1:t-1},\theta)$ is the posterior as of time t-1. If at time t-1 we have a set set of particles $\{s_{t-1}^i, i=1,\ldots,N\}$ and weights $\{w_{t-1}^i, i=1,\ldots,N\}$ which form a discrete approximation for $p(s_{t-1}|y_{1:t-1},\theta)$ then we have the approximation:

$$\hat{p}(s_{t-1}|y_{1:t-1},\theta) \propto \sum_{i=1}^{N} w_{t-1}^{i} f(s_{t}|s_{t-1}^{i},\theta).$$
(A.13)

See Andrieu et al. (2010) and Cappe at al. (2005) for reviews. From (A.13) Fernhead (2007)

makes the important observation that the joint probability of sampling particle s_{t-1}^i and state s_t is:

$$\omega_t = \frac{w_{t-1}^i g(y_t | s_t, \theta) f(s | s_{t-1}^i, \theta)}{\xi_t^i q(s_t | s_{t-1}^i, y_t, \theta)},\tag{A.14}$$

where $q(s_t|s_{t-1}^i, y_t, \theta)$ is a density function amenable to simulation and

$$\xi_t^i q(s_t | s_{t-1}^i, y_t, \theta) \simeq cg(y_t | s_t, \theta) f(s_t | s_{t-1}^i, \theta), \tag{A.15}$$

and c is the normalizing constant in (A.12).

In the MALA algorithm of Roberts and Rosenthal $(1998)^2$ we form a proposal:

$$\theta^c = \theta^{(s)} + \lambda z + \frac{\lambda^2}{2} \nabla \log p(\theta^{(s)} | Y_{1:T}), \tag{A.16}$$

where $z \sim N(0, I)$ which should result in larger jumps and better mixing properties, plus lower autocorrelations for a certain scale parameter λ . Acceptance probabilities are:

$$a(\theta^{c}|\theta^{(s)}) = \min\left\{1, \frac{p(Y_{1:T}|\theta^{c})q(\theta^{(s)}|\theta^{c})}{p(Y_{1:T}|\theta^{(s)})q(\theta^{c}|\theta^{(s)})}\right\}.$$
(A.17)

Using particle filtering it is possible to create an approximation of the score vector using Fisher's identity:

$$\nabla \log p(Y_{1:T}|\theta) = E\left[\nabla \log p(s_{1:T}, Y_{1:T}|\theta)|Y_{1:T}, \theta\right],\tag{A.18}$$

which corresponds to the expectation of:

$$\nabla \log p(s_{1:T}, Y_{1:T}|\theta) = \nabla \log p(|s_{1:T-1}, Y_{1:T-1}|\theta) + \nabla \log g(y_T|s_T, \theta) + \nabla \log f(s_T|s_{T-1}, \theta) + \nabla \log f(s_T|s_T|s_T) + \nabla \log f(s_T|s_T|s_T) + \nabla \log f(s_T|s_T) + \nabla \log f(s_T|$$

over the path $s_{1:T}$. The particle approximation to the score vector results from replacing $p(s_{1:T}|Y_{1:T}, \theta)$ with a particle approximation $\hat{p}(s_{1:T}|Y_{1:T}, \theta)$. With particle i at time t-1 we can associate a value $\alpha_{t-1}^i = \nabla \log p(s_{1:t-1}^i, Y_{1:t-1}|\theta)$ which can be updated recursively. As we sample κ_i in the APF (the index of particle at time t-1 that is propagated to produce the *i*th particle at time t) we have the update:

$$\alpha_t^i = a_{t-1}^{\kappa_i} + \nabla \log g(y_t | s_t^i, \theta) + \nabla \log f(s_t^i | s_{t-1}^i, \theta).$$
(A.19)

²The benefit of MALA over Random-Walk-Metropolis arises when the number of parameters n is large. This happens because the scaling parameter λ is $O(n^{-1/2})$ for Random-Walk-Metropolis but it is $O(n^{-1/6})$ for MALA, see Roberts et al. (1997) and Roberts and Rosenthal (1998)

To avoid problems with increasing variance of the score estimate $\nabla \log p(Y_{1:t}|\theta)$ we can use the approximation:

$$\alpha_{t-1}^i \sim N(m_{t-1}^i, V_{t-1}). \tag{A.20}$$

The mean is obtained by shrinking α_{t-1}^i towards the mean of α_{t-1} as follows:

$$m_{t-1}^{i} = \delta \alpha_{t-1}^{i} + (1-\delta) \sum_{i=1}^{N} w_{t-1}^{i} \alpha_{t-1}^{i}, \qquad (A.21)$$

where $\delta \in (0, 1)$ is a shrinkage parameter. Using Rao-Blackwellization one can avoid sampling α_t^i and instead use the following recursion for the means:

$$m_t^i = \delta m_{t-1}^{\kappa_i} + (1-\delta) \sum_{i=1}^N w_{t-1}^i m_{t-1}^i + \nabla \log g(y_t | s_t^i, \theta) + \nabla \log f(s_t^i | s_{t-1}^{\kappa_i}, \theta),$$
(A.22)

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t}|\theta) = \sum_{i=1}^{N} w_t^i m_t^i.$$
 (A.23)

As a rule of thumb Nemeth, Sherlock and Fearnhead (2014) suggest taking $\delta = 0.95$. Furthermore, they show the important result that the algorithm should be tuned to the asymptotically optimal acceptance rate of 15.47% and the number of particles must be selected so that the variance of the estimated log-posterior is about 3. Additionally, if measures are not taken to control the error in the variance of the score vector, there is no gain over a simple random walk proposal.

Of course, the marginal likelihood is:

$$p(Y_{1:T}|\theta) = p(y_1|\theta) \prod_{t=2}^{T} p(y_t|Y_{1:t-1},\theta),$$
(A.24)

where

$$p(y_t|Y_{1:t-1},\theta) = \int g(y_t|s_t) \int f(s_t|s_{t-1},\theta) p(s_{t-1}|Y_{1:T-1},\theta) ds_{t-1} ds_t,$$
(A.25)

provides, in explicit form, the predictive likelihood.

Additional References

Andrieu, C., and G.O. Roberts. (2009) The pseudo-marginal approach for efficient computation. Ann. Statist., 37, 697–725. Andrieu, C., A. Doucet, and R. Holenstein (2010), "Particle Markov chain Monte Carlo methods," Journal of the Royal Statistical Society: Series B, 72, 269–342.

Cappé, O., E. Moulines, E., and T. Rydén (2005). Inference in Hidden Markov Models. Springer, Berlin.

Casarin, R., J.-M. Marin (2007). Online data processing: Comparison of Bayesian regularized particle filters. University of Brescia, Department of Economics. Working Paper n. 0704.

Doucet, A., N. de Freitas, and N. Gordon (2001). Sequential Monte Carlo Methods in Practice. New York: Springer.

Fearnhead, P. (2007). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Statistics and Computing, 18(2):151-171.

Flury, T., and N. Shephard, (2011) Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. Econometric Theory 27, 933-956.

Gordon, N.J. (1997). A hybrid bootstrap filter for target tracking in clutter. IEEE Transactions on Aerospace and Electronic Systems 33: 353–358.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE-Proceedings-F 140: 107–113.

Hall, J., M.K. Pitt, and R. Kohn (2014). Bayesian inference for nonlinear structural time series models. Journal of Econometrics 179 (2), 99–111.

Liu, J., M. West (2001). Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag.

Nemeth, C., P. Fearnhead (2014). Particle Metropolis adjusted Langevin algorithms for statespace models. Pre-print arXiv:1402.0694v1.

Pitt, M.K., R.S. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. J. Econom. 171(2), 134–151.

Rustic, B., S. Arulampalam, and N. Gordon (2004). Beyond Kalman Filters: Particle Filters for Applications. Norwood, MA: Artech House.

	n = 3	n = 7	n = 12	W
VARMA $(4,4)^{(a)}$	-182.5	-401.9	-492.3	
$VARMA(4,4)^{(a)}$	-188.1	-406.0	-504.2	
VAR(4)	-187.1	-406.7	-496.9	
this paper	-179.41	-401.12	-490.02	0.32
VAR(4)	-187.1	-406.8	-496.9	0.24
VARMA(4,4)	-182.5	-401.9	-492.1	0.29
VAR(4), S-1	-187.0	-406.7	-496.7	0.22
VAR(4), S-2	-182.3	-401.7	-492.0	0.21

Table 1. Sum of log predictive likelihoods

Notes: (a) These are different versions of the VARMA model corresponding to different prior (echelon) restrictions. The numbers in the first three rows are taken from Chan, Eisenstat and Koop (2015), Table 7. In the multivariate normal mixture we consider G = 3 based on values of the marginal likelihood. For the Linear Combination approach we average across 10,000 different sets of weights. The averaging is performed using weights derived from marginal likelihoods converted to posterior model probabilities. S-1 and S-2 correspond to the two alternative strategies in Section 2, related to Multivariate Mixture-of-Normals.

	n = 3	n = 7	n = 12	W
$VARMA(4,4)^{(a)}$	-182.5	-401.9	-492.3	
$VARMA(4,4)^{(a)}$	-188.1	-406.0	-504.2	
VAR(4)	-187.1	-406.7	-496.9	
this paper	-179.41	-401.12	-490.02	0.32
VAR(4)	-187.1	-406.8	-496.9	0.24
VARMA(4,4)	-182.5	-401.9	-492.1	0.29
VAR(4), S-1	-187.0	-406.7	-496.7	0.22
VAR(4), S-2	-182.3	-401.7	-492.0	0.21

Table 2. Sum of log predictive likelihoods

Notes: (a) These are different versions of the VARMA model corresponding to different prior (echelon) restrictions. The numbers in the first three rows are taken from Chan, Eisenstat and Koop (2015), Table 7. In the multivariate normal mixture we consider G = 3 based on values of the marginal likelihood. For the Linear Combination approach we average across 10,000 different sets of weights. The averaging is performed using weights derived from marginal likelihoods converted to posterior model probabilities. S-1 and S-2 correspond to the two alternative strategies in Section 2, related to Multivariate Mixture-of-Normals.

Table 3. Sum of log predictive likelihoods based on the predictive density of the three variables in the n = 3 case.

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

Notes: (a) These are different versions of the VARMA model corresponding to different prior (echelon) restrictions. The numbers in the first three rows are taken from Chan, Eisenstat and Koop (2015), Table 8.

	Forecast horizon						W		
	h = 1	h=2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	
TVP-VAR ^(a) , $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.03	1.04	1.02	1.06	1.07	1.07	1.06	1.10	
TVP-VAR ^(a) , $\lambda = 0.99, \beta_{T+h} \sim RW$	1.02	1.05	1.03	1.06	1.06	1.08	1.07	1.09	
TVP-VAR, this paper	1.01	1.02	1.03	1.04	1.04	1.05	1.06	1.06	0.32
TVP-VAR, S-1	1.01	1.02	1.02	1.03	1.03	1.04	1.05	1.05	0.25
TVP-VAR, S-2	1.02	1.02	1.03	1.04	1.04	1.05	1.06	1.07	0.20
VAR Heteroskedastic ^{(a)}	1.09	1.12	1.08	1.11	1.09	1.10	1.10	1.13	
VAR Heteroskedastic, this paper	1.09	1.11	1.08	1.10	1.09	1.10	1.11	1.13	0.35
VAR Heteroskedastic, mixture	1.09	1.10	1.05	1.12	1.08	1.12	1.11	1.14	0.27
VAR Heteroskedastic, lin.comb.	1.09	1.11	1.07	1.11	1.13	1.11	1.12	1.13	0.21
VAR Homoskedastic ^{(a)}	1.02	1.05	1.04	1.04	1.03	1.03	1.03	1.05	
VAR Homoskedastic, this paper	1.02	1.05	1.04	1.04	1.03	1.03	1.04	1.05	0.22
VAR Homoskedastic, mixture	1.02	1.05	1.06	1.05	1.03	1.05	1.04	1.06	0.19
VAR Homoskedastic, lin. comb.	1.02	1.05	1.05	1.05	1.06	1.07	1.04	1.05	0.13

Table 4: Relative Mean Squared Forecast Errors, GDP, Comparison with Koop (2013)

Notes: (a) These are taken from Koop (2013), Table 8. "RW" stands for "random walk". In the multivariate normal mixture we consider G = 4 based on values of the marginal likelihood. For the Linear Combination approach we average across 10,000 different sets of weights. The averaging is performed using weights derived from marginal likelihoods converted to posterior model probabilities.

Table 5: Posterior parameters for MSV-q1 model

(Comparison with McCausland, Miller and Pelletier, 2012. Reported results are % differences of first two posterior moments)

	post. mean	post. s.d.		post. mean	post. s.d.
CHF			EUR		
$\overline{\alpha}_i$	0.13%	0.071%	\overline{lpha}_i	0.15%	0.051%
A _{ii}	0.43	0.032	A_{ii}	0.31	0.024
A_{ij}	0.17	0.11	A_{ij}	0.22	0.15
σ_{ii}	0.32	0.17	σ_{ii}	0.15	0.07
ν_i	0.11	0.11	$ u_i $	0.15	0.13
σ_{lpha}	0.07	0.03	σ_{lpha}	0.04	0.04
AUD			NZD		
$\overline{\alpha}_i$	0.10%	0.09%	\overline{lpha}_i	0.15%	0.13%
A_{ii}	0.17	0.12	A_{ii}	0.13	0.11
A_{ij}	0.15	0.14	A_{ij}	0.14	0.12
σ_{ii}	0.09	0.12	σ_{ii}	0.15	0.07
ν_i	0.06	0.05	$ u_i $	0.07	0.08
MXN			BRL		
$\overline{\alpha}_i$	0.13%	0.21%	\overline{lpha}_i	0.18%	0.13%
A _{ii}	0.14	0.07	A_{ii}	0.11	0.11
A_{ij}	0.15	0.13	A_{ij}	0.10	0.15
σ_{ii}	0.14	0.11	σ_{ii}	0.14	0.12
ν_i	0.21	0.17	ν_i	0.18	0.13
GBP			CAD		
$\overline{\alpha}_i$	0.14%	0.12%	\overline{lpha}_i	0.11%	0.11%
A _{ii}	0.17	0.10	A_{ii}	0.15	0.09
A _{ij}	0.08	0.07	A_{ij}	0.12	0.10
σ_{ii}	0.15	0.10	σ_{ii}	0.10	0.10
ν_i	0.11	0.10	ν_i	0.12	0.11
JPY			SGD		
$\overline{\alpha}_i$	0.21%	0.33%	$\overline{\alpha}_i$	0.18%	0.10%
A _{ii}	0.15	0.15	A_{ii}	0.22	0.11
A _{ij}	0.11	0.20	A_{ij}	0.16	0.17
σ_{ii}	0.12	0.32	σ_{ii}	0.34	0.25
ν_i	0.13	0.15	ν_i	0.10	0.44

	5%	25%	50%	75%	95%		
\tilde{y}_{it} difference ^(a)	0.007	0.001	0.001	0.001	-0.002		
Gaussian mixture	0.006	0.001	0.001	0.001	-0.002		
Linear combinations	0.007	0.001	0.002	0.002	-0.003		
conditional correlations							
posterior mean for FTSE	0.0035	0.0017	0.0011	0.008	0.015		
posterior mean for S&P	0.0022	0.0010	0.007	0.009	0.0011		
conditional volatility							
KO, equity	0.006	0.003	0.003	0.003	0.005		
BA, equity	0.005	0.003	0.002	0.002	0.004		
AIG, equity	0.005	0.003	0.002	0.003	0.003		
GS, equity	0.003	0.002	0.002	0.003	0.005		

Table 6: Comparison with Creal and Tsay (2015)

Notes: (a) The comparison is with the factor model in Creal and Tsay (2015) using a particle filter that we implemented ourselves. For the Gaussian mixture

we used G = 5 components. For the linear combinations we used 10,000 random sets of weights.









BANK OF GREECE WORKING PAPERS

- 185. Adam, A., and T., Moutos, "Industry-Level Labour Demand Elasticities Across the Eurozone: Will There Be Any Gain After the Pain of Internal Devaluation?" July, 2014.
- 186. Tagkalakis, O.A., "Fiscal Policy, Net Exports, and the Sectoral Composition of Output in Greece", September 2014.
- 187. Hondroyiannis, G. and D., Papaoikonomou, "When Does it Pay To Tax? Evidence from State-Dependent Fiscal Multipliers in the Euro Area", October 2014.
- 188. Charalambakis, C. E., "On Corporate Financial Distress Prediction: What Can we Learn From Private Firms in a Small Open Economy?, November 2014.
- 189. Pagratis, S., E., Karakatsani and E. Louri, "Bank Leverage and Return on Equity Targeting: Intrinsic Procyclicality of Short-Term Choices", November 2014.
- 190. Evgenidis, A. and C., Siriopoulos, "What are the International Channels Through Which a US Policy Shock is Transmitted to the World Economies? Evidence from a Time Varying Favar", January 2015.
- 191. Louzis, D. P., and A.T., Vouldis, "Profitability in the Greek Banking System: a Dual Investigation of Net Interest and Non-Interest Income", February 2015.
- 192. Papaspyrou, S.T, "EMU 2.0 Drawing Lessons From the Crisis a New Framework For Stability and Growth", March 2014.
- 193. Litina, A and T, Palivos, "Corruption and Tax Evasion: Reflections on Greek Tragedy", June 2015.
- 194. Balfoussia, H. and H.D. Gibson, "Financial Conditions and Economic Activity: The Potential Impact of the Targeted Longer-Term Refinancing Operations (TLTROS)", July 2015.
- 195. Louzis, P. D., "Steady-State Priors and Bayesian Variable Selection in VAR Forecasting", July 2015.
- 196. Zografakis, S. and A., Sarris, "The Distributional Consequences of the Stabilization and Adjustment Policies in Greece During the Crisis, with the Use of A Multisectoral Computable General Equilibrium Model", August 2015.
- 197. Papageorgiou, D. and E. Vourvachaki, "The Macroeconomic Impact of Structural Reforms in Product and Labour Markets: Trade-Offs and Complementarities", October 2015.
- 198. Louri, H., and P. M. Migiakis, "Determinants of Euro-Area Bank Lending Margins: Financial Fragmentation and ECB Policies", October 2015.
- 199. Gibson, D. H, S.G. Hall, and G. S. Tavlas, "The effectiveness of the ECB's asset purchase programs of 2009 to 2012", November 2015.
- 200. Balfoussia, H and D. Malliaropulos, "Credit-less recoveries: the role of investment-savings imbalances", November 2015.
- 201. Kalyvitis, S., "Who Exports High-Quality Products? Some Empirical Regularities from Greek Exporting Firms", December 2015.
- 202. Papadopoulos, S., P. Stavroulias and T. Sager, "Systemic Early Warning Systems for EU15 Based on the 2008 Crisis", January 2016.
- 203. Papadopoulos, G., S. Papadopoulos and T. Sager, "Credit Risk Stress Testing for EU15 Banks: a Model Combination Approach", January 2016.
- 204. Louzis, P. D., "Macroeconomic Forecasting and Structural Changes in Steady States", March 2016.
- 205. Christodoulakis, N. and C. Axioglou, "Underinvestment and Unemployment: the Double Hazard in the Euro Area", April 2016.

- 206. Bardaka, C. I., "Structural and Cyclical Factors of Greece's Current Account Balances: A Note", May 2016.
- 207. Lothian, J. R., and G. S. Tavlas, "How Friedman and Schwartz Became Monetarists", May 2016.
- 208. Balfoussia, H. and H. D. Gibson, "Firm Investment and Financial Conditions in the Euro Area: Evidence from Firm-Level Data", June 2016.
- 209. Samantas, G.I., "On the Optimality of Bank Competition Policy", July 2016.
- 210. Malliaropulos, D., and P.M. Migiakis, "The Re-Pricing of Sovereign Risks Following the Global Financial Crisis", July 2016.
- 211. Asimakopoulos, I., P.K. Avramidis, D. Malliaropulos, N. G. Travlos, "Moral Hazard and Strategic Default: Evidence from Greek Corporate Loans", July 2016.
- 212. Tavlas, S.G., "New Perspectives on the Great Depression: a Review Essay", September 2016.
- 213. Papapetrou, E. and P. Tsalaporta, "Inter-Industry Wage Differentials in Greece: Rent-Sharing and Unobserved Heterogeneity Hypotheses", November 2016.
- 214. Gibson, D.H, S.G. Hall, and G.S. Tavlas, "Self-Fulfilling Dynamics: the Interactions of Sovereign Spreads, Sovereign Ratings and Bank Ratings during the Euro Financial Crisis", November 2016.
- 215. Baumann, U. and M. Vasardani, "The Slowdown in US Productivity Growth -What Explains it and Will it Persist?", November 2016.
- 216. Tsionas. G.M., "Alternative Bayesian Compression in Vector Autoregressions and Related Models", November 2016.