



BANK OF GREECE  
EUROSYSTEM

# Working Paper

Novel techniques for Bayesian inference in univariate  
and multivariate stochastic volatility models

Mike G. Tsionas

294

FEBRUARY 2022

WORKINGPAPERWORKINGPAPERWORKINGPAPERWO

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](http://www.bankofgreece.gr)

Published by the Bank of Greece, Athens, Greece  
All rights reserved. Reproduction for educational and  
non-commercial purposes is permitted provided that the source is acknowledged.

ISSN: 2654-1912 (online)

DOI: <https://doi.org/10.52903/wp2022294>

# NOVEL TECHNIQUES FOR BAYESIAN INFERENCE IN UNIVARIATE AND MULTIVARIATE STOCHASTIC VOLATILITY MODELS

Mike G. Tsionas  
Lancaster University

## Abstract

In this paper we exploit properties of the likelihood function of the stochastic volatility model to show that it can be approximated accurately and efficiently using a response surface methodology. The approximation is across the plausible range of parameter values and all possible data and is found to be highly accurate. The methods extend easily to multivariate models and are applied to artificial data as well as ten exchange rates and all stocks of FTSE100 using daily data. Formal comparisons with multivariate GARCH models are undertaken using a special prior for the GARCH parameters. The comparisons are based on marginal likelihood and the Bayes factors.

*Keywords:* Stochastic volatility; response surface; likelihood; Monte Carlo.

*JEL classifications:* C13; C15

*Acknowledgments:* The author wishes to thank Dimitris Malliaropoulos and an anonymous reviewer for useful comments on an earlier version. The views expressed in this paper are those of the author and do not necessarily reflect those of the Bank of Greece. This research was conducted in the context of the Bank's programme of cooperation with universities.

## Correspondence:

Mike G. Tsionas  
Department of Economics  
Lancaster University Management School  
Bailrigg, Lancaster, LA1 4YX, United Kingdom  
Tel. telephone: +44 (0)1524 592668  
Email: m.tsionas@lancaster.ac.uk

# 1 Introduction

Despite numerous advances in filtering and simulation inference in the stochastic volatility model (Jacquier, Polson and Rossi, 1994) is far from being standard. The reason is that MCMC and simulation techniques can behave unpredictably in certain data sets because of non-convergence, significant autocorrelation and / or tuning of certain parameters of the simulation. For different approaches see Chib, Nardari, and Shephard (2002), Danielsson (1994), de Jong and Shephard (1995), Kim, Shephard and Chib (1998), Knight, Satchell and Yu (2002), and Pitt and Shephard (1999). Excellent reviews are subsumed in Shephard (1996), Chib *et al* (2002) and Kim *et al* (1998). For alternative techniques based on the Hessian matrix of the log-posterior and Laplace approximations, see McCausland (2012) and Richard and Wang (2007).

In this paper we use a response surface methodology to obtain key components of the likelihood function in terms of the structural parameters of the stochastic volatility model. This is possible because the key components are expectations (with respect to a multivariate normal distribution) which are smooth and behave almost as linear functions of squared returns. The parameters of these functions depend on the structural parameters and this gives rise to the question of approximating them themselves. It turns out that simple functions of the structural parameters can be used to provide highly accurate approximations to order  $10^{-9}$  or better.

Response surface methodology was used primarily in connection to optimization in “black box” systems; see the classic paper of Box and Wilson (1951). Since then there have been quite a few applications in statistics; see Box, Hunter and Hunter (1978), Box and Draper (1987), Myers and Montgomery (1995) Khuri and Cornell (1996), Hood and Welch (1993), *inter alia*. The idea is that the output of a complex system or model can be approximated in terms of the underlying parameters or variables and, in turn, use interpolation to predict the output of the model on other values of the parameters or variables using interpolation.

Somewhat surprisingly we show that statistical inference in models with leverage or multivariate stochastic variance models does not require more approximations and the response surface techniques developed for the univariate case will suffice even in high-dimensional stochastic variance models. Clearly this opens up the way for routine implementations of highly accurate estimation and inference in this class of models requiring only interpolation or straightforward table look-up. Even this can be avoided since there are simple relationships between the structural parameters, the key components of the likelihood and certain coefficients related to the approximation of the latter.

## 2 Model and approximation

Consider the stochastic volatility (SV) model

$$y_t = \exp\left(\frac{1}{2}h_t\right) u_t, \quad (1)$$

$$h_t = \alpha + \rho h_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where  $u_t \sim iidN(0, 1)$  and  $\varepsilon_t \sim iidN(0, \sigma^2)$ ,  $t = 1, \dots, T$ . For the initial condition we assume  $h_0 = \frac{\alpha}{1-\rho}$ .

The likelihood function in terms of the parameters  $\theta = [\alpha, \rho, \sigma]'$  is

$$L(\theta; y^T) = (2\pi)^{-T} \sigma^{-T} \prod_{t=1}^T \exp\left\{-\frac{h_t}{2} - \frac{y_t^2 \exp(-h_t)}{2} - \frac{(h_t - \alpha - \rho h_{t-1})^2}{2\sigma^2}\right\} dh^T, \quad (3)$$

where generically  $y^T = [y_1, \dots, y_T]'$ . Using the last term and well known formulae for an AR(1) process we have:

$$h^T \sim N_T(\mu(\theta), V(\theta)). \quad (4)$$

Define the function  $f_t(h) = \exp\left[-\frac{1}{2}\{h + y_t^2 \exp(-h)\}\right]$ , for some  $t = 1, \dots, T$ . Then it follows

$$\log L(\theta; y^T) = -\frac{T}{2} \log(2\pi) + \sum_{t=1}^T \log \mathbb{E}[f_t(h_t) | y_{(t-1)}],$$

where  $y_{(t-1)}$  denotes the past history of  $y_t$ . The essence of the approximation proposed in the paper is to adopt different conditioning as a means to arrive at a convenient expression that can be automated in terms of computations. We propose, specifically, to work with the following expression:

$$\log L(\theta; y^T) = -\frac{T}{2} \log(2\pi) + \sum_{t=1}^T \log \mathbb{E} f_t(h_t), \quad (5)$$

where the expectation is taken with respect to the joint distribution of  $h^T$  in (4). The log of the expectation of  $f(h)$ , which arises also in Sandmann and Koopman (1998) and Kim et al. (1998) and are related to the density of a  $\log\chi_1^2$  variable. The expression has a particularly simple form for the plausible range of the parameters and can be approximated extremely accurately with a linear or quadratic function of  $y_t^2$ . These points are illustrated, for typical parameter values, in Figure 1. Figure

1a reports typical functions  $f_t(h)$  as a function of  $h$  for various values of  $y_t^2$ . In Figure 1b reported are typical functions  $\log \mathbb{E}f_t(h)$  as a function of  $y_t^2$  and various configurations of the structural parameters. Notice that in Figure 1b the range of  $y_t^2$  extends from zero to one which is more extended compared to the plausible values of squared returns.

This implies that for given parameters  $\theta$  the log of the expectations in (5) can be approximated accurately by

$$\log \mathbb{E}f(h_t) = d_{\theta 1} + d_{\theta 2}y_t^2 + d_{\theta 3}y_t^4 + w_\theta = z'_t d_\theta + w_\theta, \quad (6)$$

where  $z'_t = [1, y_t^2, y_t^4]$ ,  $d_\theta = [d_{\theta 1}, d_{\theta 2}, d_{\theta 3}]'$ ,  $w_\theta$  denotes the approximation error and the subscript  $\theta$  emphasizes the dependence of the approximating coefficients and approximation error on the true parameters. Although this result is useful in its own right it turns out that the parameters in  $d_\theta$  can be also approximated extremely accurately by a simple function of  $\theta$ , viz.,

$$d_\theta \approx \theta' \gamma_1 + \exp(\theta)' \gamma_2, \quad (7)$$

where  $\gamma_1$  and  $\gamma_2$  are parameters. To show the validity of this relationship we construct a grid  $\mathfrak{G} = \{\bar{\theta}_g, g = 1, \dots, N\}$  for the parameters of the stochastic volatility process in  $[-1, 1] \times [-0.9, 0.9] \times [0, 0.5]$ . For each parameter in  $\theta$  we use  $n$  equally spaced points so that the size of the grid is  $N = n^3$ . Suppose  $\Theta$  is the  $N \times 3$  matrix consisting of all values of the parameters in the grid  $\mathfrak{G}$ , and define the  $N \times 6$  matrix  $\mathfrak{C} = [\Theta : \exp(\Theta)]$ . Then we can write (7) as

$$\mathfrak{D} = \mathfrak{C}\gamma + v, \quad (8)$$

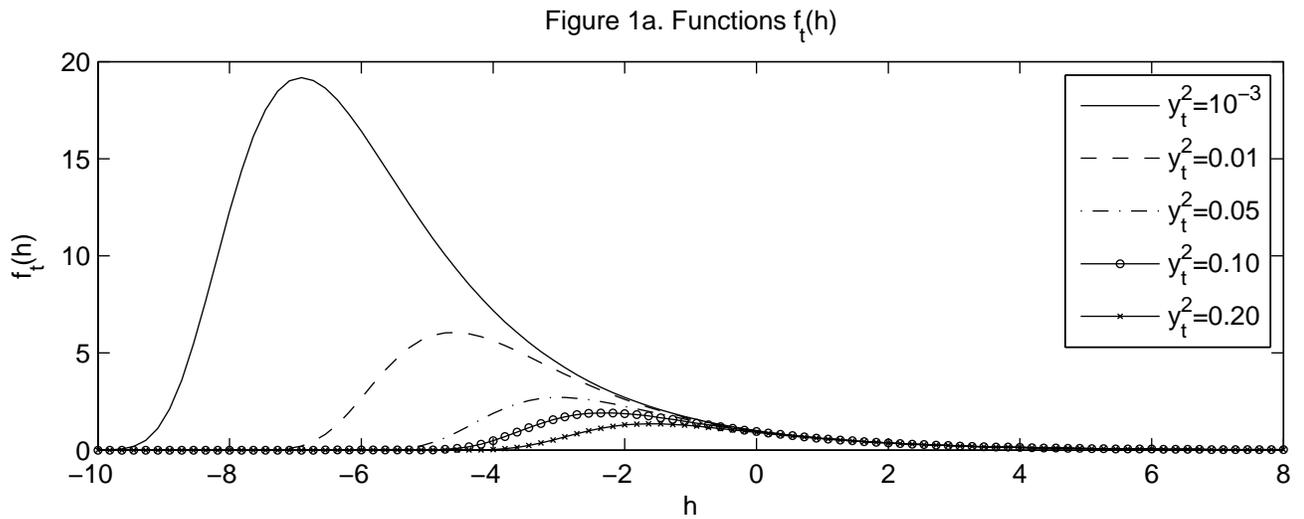
where  $\mathfrak{D}$  is the  $N \times 3$  matrix consisting of the values of  $d_\theta$  for  $\theta \in \mathfrak{G}$  and  $v$  denotes the error of approximation. The equations in (8) are in the form of multivariate regression with the same ‘‘regressors’’. To understand the construction we can describe it in three steps:

- (i) For any parameter  $\theta \in \mathfrak{G}$ , we can obtain a corresponding value of  $d_\theta$  from  $\mathfrak{D}$ , using (8).
- (ii) Given the coefficients  $d_\theta$  we use (6) to obtain the log expectations  $\log \mathbb{E}f(h_t)$ ,  $t = 1, \dots, T$ .
- (iii) The expectations are substituted in (5) and the log likelihood is approximated.

The natural question is in what sense this approximation is accurate since functions  $f_t(h)$  depend

Figure 1: Typical functions  $\mathbb{E}f_t(h)$  and  $\log\mathbb{E}f_t(h)$

In this Figure, we report typical functions  $\mathbb{E}f_t(h)$  and  $\log\mathbb{E}f_t(h)$  for various values of  $y_t^2$  as a function of  $h$ .



on the data through squared returns  $y_t^2$ . As we remarked these functions can be approximated in the worst case by quadratic functions as in (6) over the range of values  $y_t^2 \in Q = [0, 0.25]$ . Specifically:

(a) we will use 1,000 equally spaced values in the interval  $Q$  to fit (6) and obtain the coefficients  $d_\theta$  for any fixed  $\theta \in \mathfrak{G}$ . Since the functions in (6) are nearly linear it has been found that even 100 values can be used without distorting the accuracy of the approximation.

(b) The expectations in the left hand side of (6) are obtained using simulation from the  $T$ -variate normal distribution in (4). The number of draws is fixed to 10,000 but accurate results were obtained using as few as 1,000 draws.

(c) Next, (6) is estimated to obtain coefficients  $d_\theta$ . Finally,

(d) we use (8) to estimate the response surface between  $\theta$  and  $d_\theta$  and obtain the approximating coefficients  $\gamma$ .

From steps (a) through (d) we can obtain all the necessary elements that build up a response surface between the structural parameters  $\theta$  and the log expectation in (5). In Figure 2 we report a histogram of the median absolute residuals  $w_\theta$  from (6) for  $T = 1,500$  when 30 points are used for each parameter in the grid  $G$  for a total of 27,000 cases. All coefficients of determination ( $R^2$ ) are in excess of 0.99 and as can be seen from Figure 2 the errors of approximation are  $10^{-9}$  or better, for example  $10^{-11}$  and  $10^{-12}$  in certain cases.

We should remark that since our approximation is over values  $y_t^2 \in Q$  there is still some dependence of the entire approximation upon the sample size,  $T$ . The response surface methodology has to be implemented for given values of  $T$ . We omit the details to save space but a `Fortran` program is available to perform the computation. In Figure 3 we report response curves of the  $d_\theta$  coefficients versus the structural parameters  $\alpha$  and  $\rho$  for an intermediate value of  $\sigma = 0.242$ . In Figure 4 shown are the entire response surfaces in terms of the structural parameters  $\alpha$  and  $\rho$  for  $\sigma = 0.242$ .

### 3 Results

Results from artificial data, in the form of marginal posterior distributions of the structural parameters are reported in Figure 5. The data were generated from a model with  $\alpha = -0.1$ ,  $\rho = 0.9$  and  $\sigma = 0.1$  with sample size  $T = 1,500$ . MCMC was implemented using a Gibbs sampler with 110,000 passes the first 10,000 of which are omitted to mitigate the impact of start-up effects and are thinned every other 10th draw to mitigate autocorrelation. The updates for the posterior conditional distribution  $h_t|h_{t-1}, h_{t+1}, y^T$  are implemented using a rejection sampler for log-concave densities. Response surface Monte Carlo has been implemented using the same configuration for an adaptive Metropolis-Hastings algorithm that samples in the space of  $\theta = [\alpha, \rho, \sigma]'$ . The computation of  $\log \mathbb{E}f_t(h)$  is performed using the configurations described in steps (i)-(iii) and (a)-(d) in section 2. The marginal posteriors are surprisingly close but (after thinning) the Metropolis-Hastings algorithm using response surface behaves much better in terms of autocorrelation functions (acf), see Figure 5d. The maximum acf provides the point-wise maximum of acf across all structural parameters.

### 4 Leverage

The stochastic volatility model with leverage effect (SVL) is given by (1) and (2) under the alternative assumption that

$$\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} \sim iid N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \varphi \\ \varphi & 1 \end{bmatrix}, \quad (9)$$

Figure 2: Errors of approximation

In Figure 2 we report a histogram of the median absolute residuals  $w_\theta$  from (6) for  $T = 1,500$  when 30 points are used for each parameter in the grid  $G$  for a total of 27,000 cases. All coefficients of determination ( $R^2$ ) are in excess of 0.99 and as can be seen from Figure 2 the errors of approximation are  $10^{-9}$  or better, for example  $10^{-11}$  and  $10^{-12}$  in certain cases.

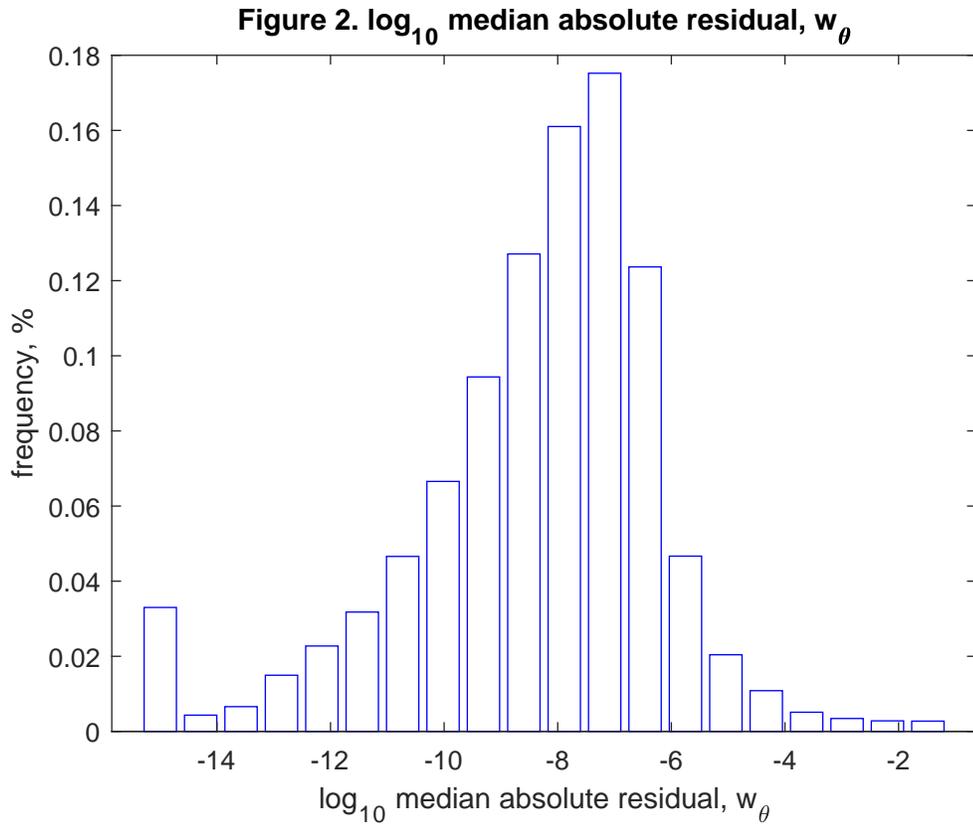


Figure 3: Response curves of  $d_\theta$  coefficients versus  $\alpha$  and  $\rho$  ( $\sigma = 0.242$ )

In this Figure we report response curves of the  $d_\theta$  coefficients versus the structural parameters  $\alpha$  and  $\rho$  for an intermediate value of  $\sigma = 0.242$ .

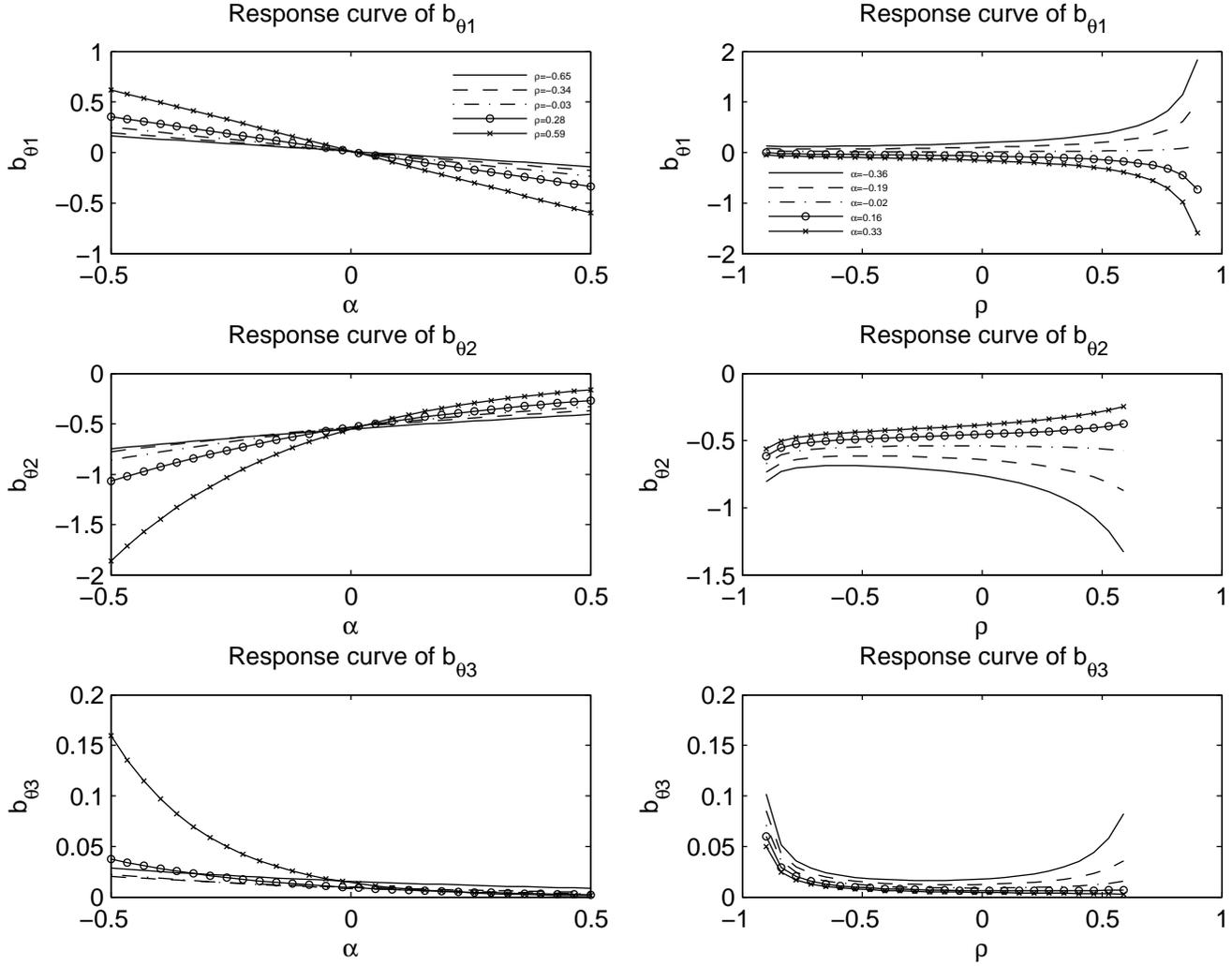
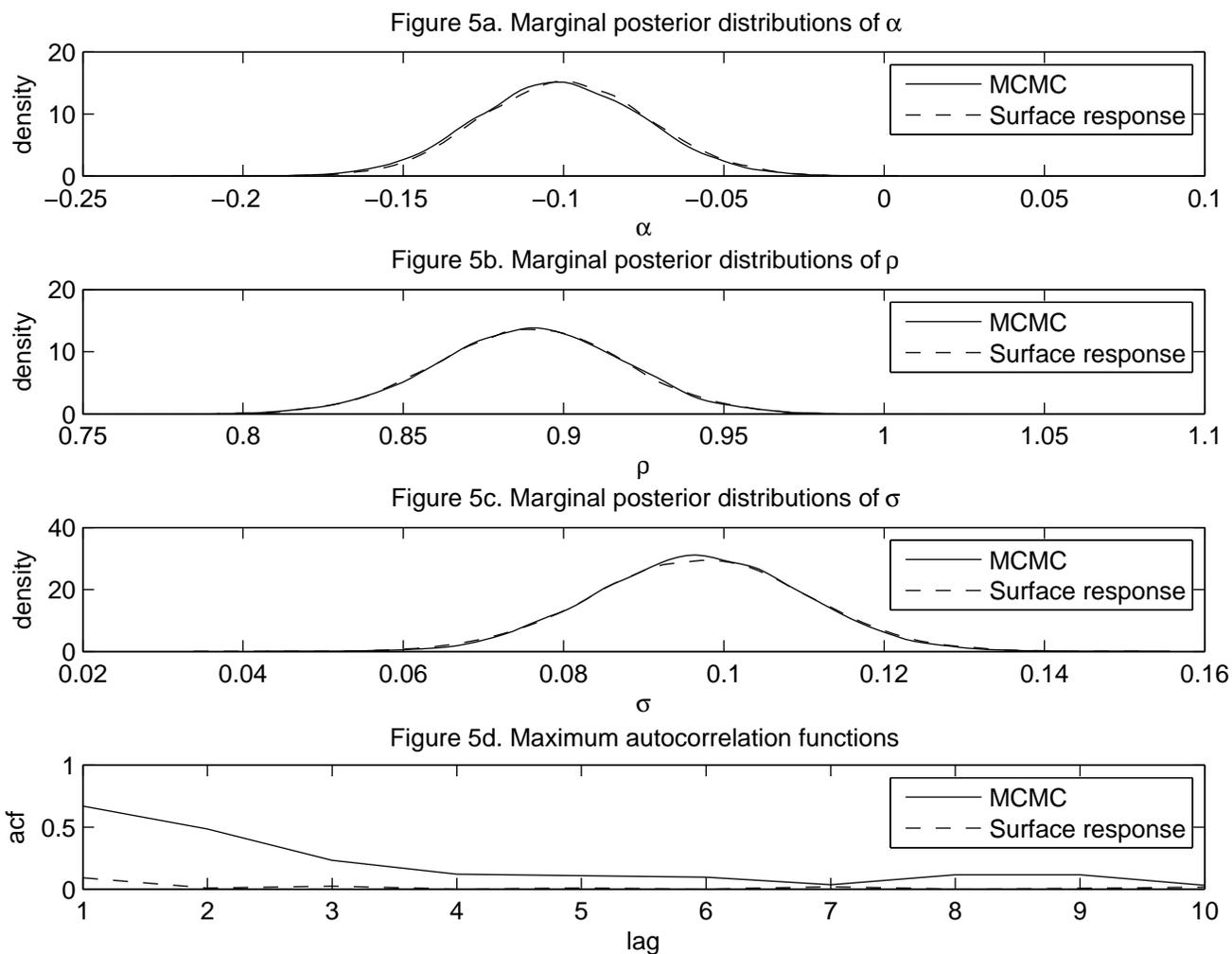


Figure 4: Response surfaces of  $d_\theta$  coefficients versus  $\alpha$  and  $\rho$  ( $\sigma = 0.242$ )

In this Figure we present the entire response surfaces in terms of the structural parameters  $\alpha$  and  $\rho$  for  $\sigma = 0.242$ .

Figure 5: Marginal posterior distributions (artificial data)

Results from artificial data, in the form of marginal posterior distributions of the structural parameters are reported in this Figure. The data were generated from a model with  $\alpha = -0.1$ ,  $\rho = 0.9$  and  $\sigma = 0.1$  with sample size  $T = 1,500$ . MCMC was implemented using a Gibbs sampler with 110,000 passes the first 10,000 of which are omitted to mitigate the impact of start-up effects and are thinned every other 10th draw to mitigate autocorrelation. The updates for the posterior conditional distribution  $h_t|h_{t-1}, h_{t+1}, y^T$  are implemented using a rejection sampler for log-concave densities. Response surface Monte Carlo has been implemented using the same configuration for an adaptive Metropolis-Hastings algorithm that samples in the space of  $\theta = [\alpha, \rho, \sigma]'$ . The computation of  $\log \mathbb{E} f_t(h)$  is performed using the configurations described in steps (i)-(iii) and (a)-(d) in section 2.



where  $\varphi$  is the correlation coefficient between asset return and its volatility. It follows from standard properties of the bivariate normal distribution that

$$\varepsilon_t = \varphi\sigma u_t + \sigma\sqrt{1-\varphi^2}\xi_t, \quad (10)$$

where  $\xi_t \sim iid N(0, 1)$ . The state equation (2) can be written as (Malik and Pitt, 2011)

$$h_t = \alpha + \rho h_{t-1} + \varphi\sigma y_{t-1} \exp(-\frac{1}{2}h_{t-1}) + \sigma\sqrt{1-\varphi^2}\xi_t, \quad (11)$$

and the new model is given by equations (1) and (11). If we define the parameter vector  $\theta = [\alpha, \rho, \sigma, \varphi]'$  the likelihood function of the model is

$$L(\theta; y^T) = \left(2\pi\sigma\sqrt{1-\varphi^2}\right)^{-T} \exp\left(-\frac{1}{2}\sum_{t=1}^T Q_t\right) dh^T, \quad (12)$$

where  $Q_t = h_t + y_t^2 \exp(-\frac{1}{2}h_t) + \frac{1}{\sigma^2(1-\varphi^2)}[h_t - \alpha - \rho h_{t-1} - \varphi\sigma y_{t-1} \exp(-\frac{1}{2}h_{t-1})]^2$ ,  $t = 1, \dots, T$ . Clearly the log likelihood in (12) can be written as

$$\log L(\theta; y^T) = -\frac{T}{2} \log(2\pi\sigma\sqrt{1-\varphi^2}) + \sum_{t=1}^T \log \mathbb{E} \exp[-\frac{1}{2}f(h_t)], \quad (13)$$

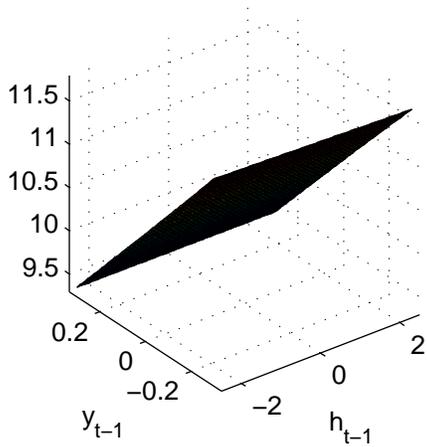
where, as before,  $f_t(h) = \exp[h + y_t^2 \exp(-\frac{1}{2}h)]$  and the expectation is taken with respect to the process in (11). The global approximation of the expectations in (13) requires taking account of the dependence of these expectations on the two variables  $y_{t-1}$  and  $h_{t-1}$ . Conditioning on them both  $y_t$  and  $h_t$  can be simulated from (1) and (11). Therefore the correct way to express the function  $f$  is really  $f_t(h; y_{t-1}, h_{t-1}) = \exp[h + y_t^2 \exp(-\frac{1}{2}h)]$  and the expectation in (13) is taken with respect to  $(y_t, h_t)|y_{t-1}, h_{t-1}$ ,  $t = 1, \dots, T$ . Compared to the simple SV model there are two differences. First, the expectation has to be approximated over  $y_{t-1}$  and  $h_{t-1}$ , rather than over  $y_{t-1}^2$  alone. Second, in the SV model with leverage the expectations in (13) has to be approximated term-by-term for all  $t = 1, \dots, T$ . In the SV model the term-by-term evaluation can be avoided at the cost of drawing from a  $T$ -variate normal distribution. Using 50,000 Monte Carlo simulations and 50 points for a grid on  $y_{t-1}$  and  $h_{t-1}$ , surface plots of the log expectation are provided in Figure 6. From the figure it is clear that the log expectation is roughly linear, indeed the values of  $R^2$  are very close to 1.

The reader will, of course, wonder how can it be possible to have a quadratic surface in the case of SV model but nearly a hyperplane in the SVL model. The reason is that the range of  $h_{t-1}$  has been

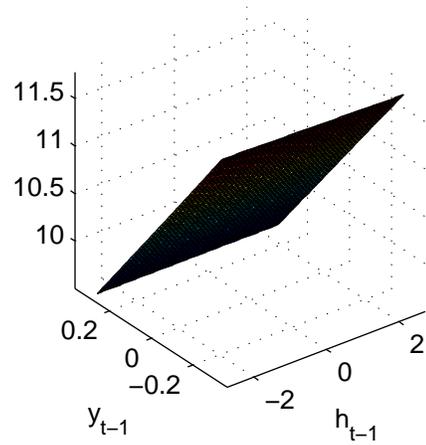
Figure 6: Values of  $\log \mathbb{E} \exp[-\frac{1}{2}f(h_t; y_{t-1}, h_{t-1})]$  for selected parameter values

Using 50,000 Monte Carlo simulations and 50 points for a grid on  $y_{t-1}$  and  $h_{t-1}$ , surface plots of the log expectation are provided in this Figure. From the figure it is clear that the log expectation is roughly linear, indeed the values of  $R^2$  are very close to 1.

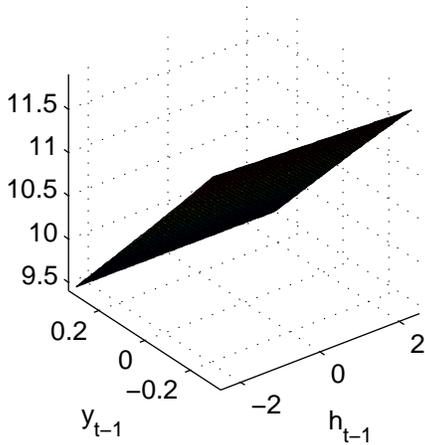
I.  $\alpha=-0.1, \rho=0.9, \sigma=0.1, \phi=0.7$



II.  $\alpha=-0.3, \rho=0.9, \sigma=0.2, \phi=0.25$



III.  $\alpha=-0.3, \rho=0.9, \sigma=0.2, \phi=0.5$



IV.  $\alpha=-0.3, \rho=0.9, \sigma=0.2, \phi=-0.5$

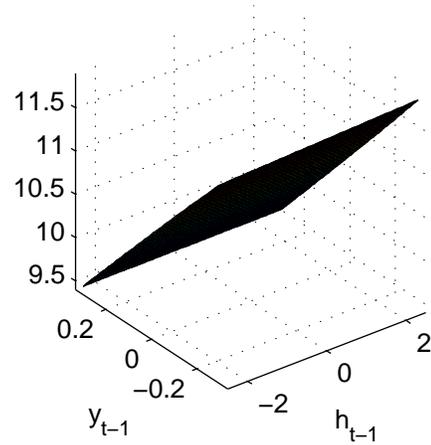
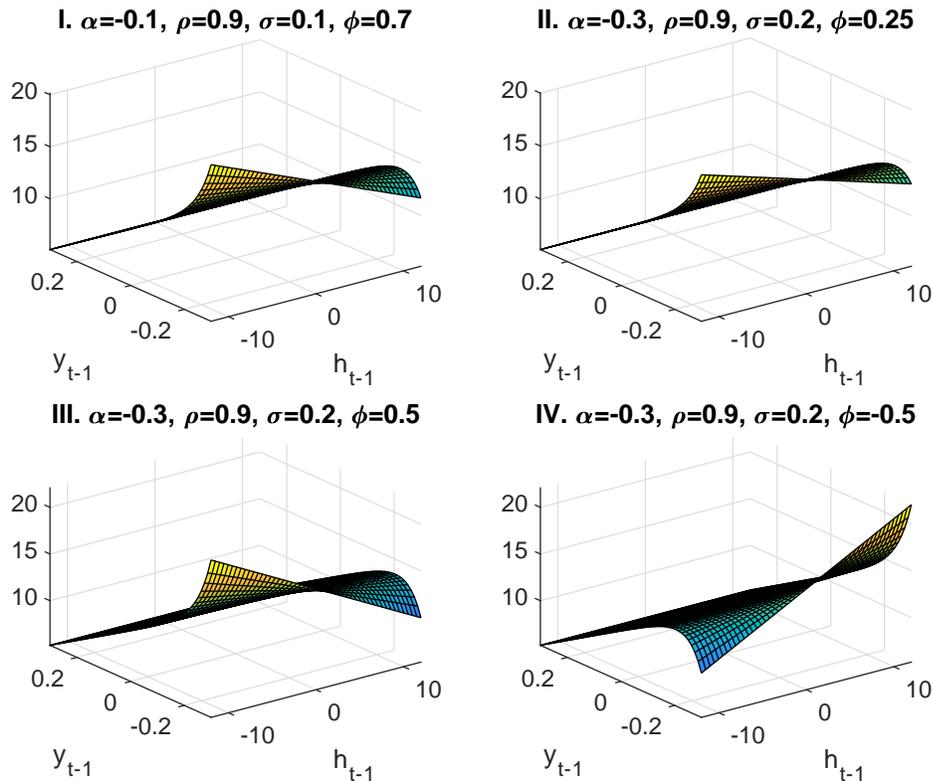


Figure 7: Values of  $\log \mathbb{E} \exp \left[ -\frac{1}{2} f(h_t; y_{t-1}, h_{t-1}) \right]$  for selected parameter values,  $h_{t-1} \in (-20, 20)$ .

Response surfaces when the the grid for  $h_{t-1}$  is extended over the interval  $(-12, 12)$ .



restricted to “reasonable” or suitable values. If we extend the grid for  $h_{t-1}$  over the interval  $(-12, 12)$  we have a different situation as shown in Figure 7.

Despite the fact that the surfaces are apparently nonlinear the values of  $R^2$  from a linear approximation are above 0.97 indicating that the approximation by hyperplanes is quite good. The approximation error can of course be improved considerably in this case by using more elaborate approximation methods. For each  $\theta \in \mathfrak{G}$  we have an “intermediate” configuration  $\mathcal{C}_\theta = \{y_{-1}, h_{-1}, \mathcal{E}_\theta\}$  where  $\mathcal{E}_\theta$  denotes the log expectation evaluated over a grid of values  $\mathcal{S} = \{y_{-1}, h_{-1}\}$  whose dimension is  $p^2 \times 1$ , and  $\mathcal{E}_\theta$  is also  $p^2 \times 1$ , defined by  $\mathcal{E}_\theta = \text{vec}(\mathcal{E}_\theta)$ , where  $\mathcal{E}_\theta = [\log \mathbb{E} \exp \{ -\frac{1}{2} f(h_t; y_m, h_n) \}, m, n = 1, \dots, p]$  is the  $p \times p$  matrix that is used to construct figures like 6 and 7. Since the configuration  $\mathcal{C}_\theta$  depends

on the parameter vector we have, for each parameter, a whole  $p \times p$  matrix that corresponds to the surface generated by the log expectation over the grid of values  $\{y_{-1}, h_{-1}\}$ .

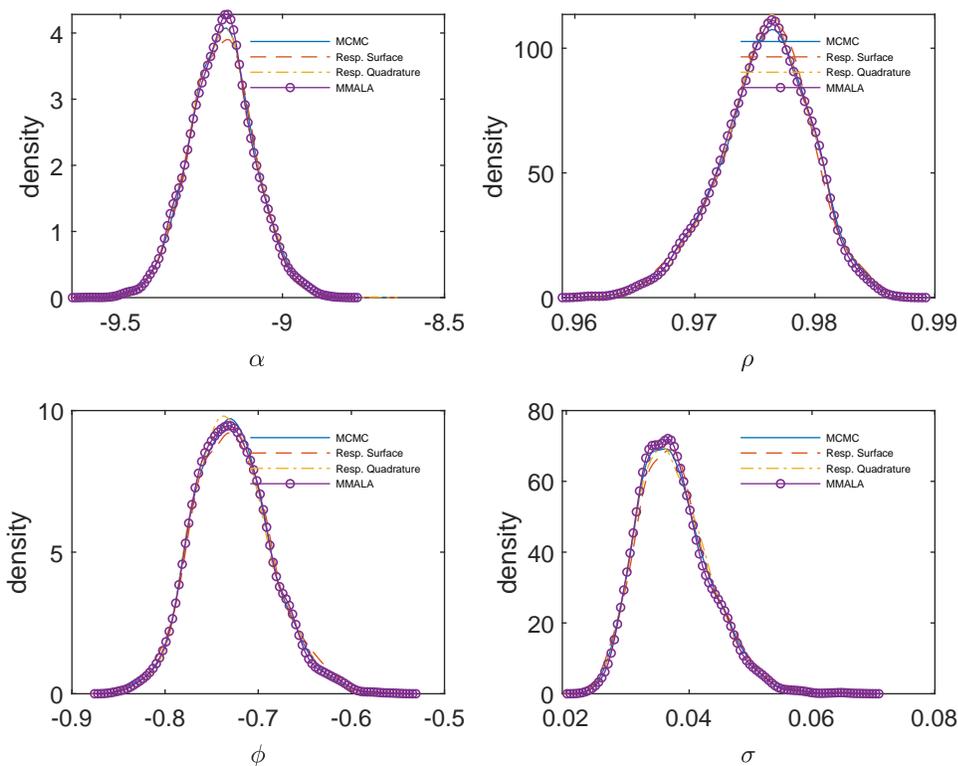
Define the final configuration as  $\mathcal{F}_{n,p} = (\mathfrak{B}, \{\mathcal{E}_\theta, \theta \in \mathfrak{B}\})$  where as in the previous section  $n$  denotes the number of points in the grid over the parameter space  $\Theta$ . In practice  $\mathcal{F}_{n,p}$  can be defined as a matrix whose dimension is  $n^4 p^2 \times 7$  where 7 stands for the four parameters  $(\alpha, \rho, \sigma, \varphi)$ , the two variables in state space  $(y_{t-1}$  and  $h_{t-1})$  and the final column contains the values of  $\mathcal{E}_\theta$ . Matrix or set  $\mathcal{F}_{n,p}$  contains all information we need to develop a global approximation to  $\log \mathbb{E} \exp \left\{ -\frac{1}{2} f(h_t; y_m, h_n) \right\}$  across the state space  $\mathcal{S}$ , and the parameter space  $\Theta$ , *independently of the specific data or the sample size*. Given matrix  $\mathcal{F}_{n,p}$  which is computed only once independently of the specific data or the sample size we can approximate the log expectation  $\log \mathbb{E} \exp \left\{ -\frac{1}{2} f(h_t; y_{t-1}, h_{t-1}) \right\}$  accurately using a variety of techniques, for example B-spline interpolation or simple table-look up. Although the initial set up cost of constructing the matrix  $\mathcal{F}_{n,p}$  is non-negligible (i) this is done only once for all possible SVL models and (ii) programming the construction of the matrix is quite simple in high-level languages like `WinGauss`, `R` or `Matlab`.

As an empirical application we use daily observations for the Dow-Jones industrial average over the period 1/2/1998 through 5/18/2012, a total of 3,577 observations. We use a SVL model with the slight difference that (2) is expressed in the “identified” form:  $h_t = \alpha(1 - \rho) + \rho h_{t-1} + \varepsilon_t$ . MCMC and MCMC with the aid of the Response Surface are implemented using 150,000 draws the first 50,000 of which are omitted and we thin every other 10th draw. Results are also presented using another method that we call “*Response Quadrature*”. Since the (log) likelihood can be expressed in (13) as a function of the four parameters in  $\theta$  with the aid of the Response Surface methodology, marginal posteriors densities and moments can be computed using trivariate and quadrivariate numerical integration techniques (Genz and Malik, 1980, NAG subroutine D01FCF). The following priors have been used:  $\alpha \sim N(0, 10^2)$ ,  $\varphi \sim U(-1, 1)$ ,  $\sigma \sim IG(10^{-3}, 10^{-3})$  and  $\rho \sim N(0, 1)$  truncated to the interval  $(-1, 1)$ . The results reported in Figure 8 are acceptably close and indicate that volatility is highly persistent and leverage is quite important empirically. The heading “MMALA” will be explained later.

The use of efficient (adaptive or non-adaptive) multivariate integration techniques in connection with Response Surface methods opens up the possibility of routine and accurate implementations of “exact” Bayesian inference in models with high-dimensional latent variables. In models where Monte Carlo maximum likelihood is possible it is usually also possible to write the likelihood in terms of functions that depend on conditional “sufficient statistics” such as  $y_{t-1}^2$  in SV or  $(y_{t-1}, h_{t-1})$  in the SVL. The

Figure 8: Marginal posterior densities of SVL, Dow-Jones data

In the empirical application we use daily observations for the Dow-Jones industrial average over the period 1/2/1998 through 5/18/2012, a total of 3,577 observations. We see a SVL model with the slight difference that (2) is expressed in the “identified” form:  $h_t = \alpha(1 - \rho) + \rho h_{t-1} + \varepsilon_t$ . MCMC and MCMC with the aid of the Response Surface are implemented using 150,000 draws the first 50,000 of which are omitted and we thin every other 10th draw. Results are also presented using another method that we call “Response Quadrature”. Since the (log) likelihood can be expressed in (13) as a function of the four parameters in  $\theta$  with the aid of the Response Surface methodology, marginal posteriors densities and moments can be computed using trivariate and quadrivariate numerical integration techniques (Genz and Malik, 1980, NAG subroutine D01FCF). The following priors have been used:  $\alpha \sim N(0, 10^2)$ ,  $\varphi \sim U(-1, 1)$ ,  $\sigma \sim IG(10^{-3}, 10^{-3})$  and  $\rho \sim N(0, 1)$  truncated to the interval  $(-1, 1)$ . The heading “MMALA” will be explained later.



reader may wonder why implementations based on Response Surface methods are better or preferable compared to Monte Carlo - based methods for a specific model (such as SV or SVL) that depend on the data, for example SVpack (Kim *et al.*, 1998).

*First*, Response Surface methods will be much faster by orders of magnitude (either in their MCMC or numerical integration variant).

*Second*, posterior sensitivity to prior assumptions can be accomplished routinely over a wide and fine grid of prior hyperparameters. This is an essential part of Bayesian communication in most applications.

*Third*, the marginal likelihood can be computed easily using the Laplace approximation.

*Fourth*, the approximation in terms of the parameters can be avoided and only the approximation with respect to conditional “sufficient statistics” can be used and MCMC or numerical integration may be used without increasing significantly the computational costs.

*Fifth*, the important issue is that the high-dimensional latent variable is integrated out of the likelihood using conditional sufficiency (as in Figures 1b and 6) and the resulting function is analyzed using either MCMC or adaptive integration since it is usually low dimensional. The situation we end up with is quite similar to a probit model whose log likelihood is

$$\log L(\theta; y^n, x^n) \doteq \sum_{i \in I_1} \log \Phi(x'_i \theta) + \sum_{i \in I_0} \log \Phi(-x'_i \theta),$$

where  $I_m = \{i : y_i = m \in \{0, 1\}\}$ ,  $\Phi$  is the standard normal distribution function and  $x_i$  is a  $k \times 1$  vector of explanatory variables. Since the likelihood is available in simple form, it is not necessary to use response surface methodologies to interpolate the parameters in terms of simpler functions.

*Sixth*, there is another MCMC technique that can be used in connection with response surface methodologies when the likelihood is expressed in terms of the structural parameters and the high-dimensional latent variable is integrated out (as for example in Figures 1b and 6). This is the method of Riemannian manifold MCMC of Girolami and Calderhead (2011) and especially their Manifold Metropolis Adjusted Langevin Algorithm or MMALA. For a SV model, as the authors explain in section 8.2 of their paper, drawing from  $p(h^T | \theta, y^T)$  requires the inversion of a sparse tridiagonal matrix which can be done in  $O(T)$  per iteration -instead of the prohibitive  $O(T^3)$  for general matrices. Expressions similar to our  $f_t(h)$  in (5) naturally arise also in the authors’ definition of the gradient of the log posterior with respect to the latent  $h^T$  (see p. 21). But even this  $O(T)$  operation can be

Table 1: Relative numerical efficiencies

method	$\rho$	$\phi$	$\alpha$	$\sigma$
MCMC	0.017	0.035	0.072	0.130
RS-MCMC	0.510	0.412	0.279	0.458
MMALA	0.290	0.401	0.256	0.410
RS-MMALA	0.325	0.520	0.300	0.433

completely avoided using a response surface as in Figure 1b which requires interpolation or table look-up in one dimension, an operation that is massively fast especially when the latent variables are highly correlated. MMALA can be applied to the posterior  $p(\theta|y^T)$  after integrating out the latent variables using the response surface technology. For the SV model this is not even necessary since there is a simple (LS) relation between the structural parameters  $\theta$  and the response-surface coefficients.

The results of applying MMALA to the SVL model for the Dow-Jones are also shown in Figure 8. But the impressive fact is perhaps that adaptive quadrature performs just as well. Relative numerical efficiencies (RNE, Geweke, 1992) are reported in Table 1.

With the exception of MCMC all algorithms perform well with some evidence that Response Surface (RS) - based techniques perform better. After skipping every other 10th draw the autocorrelation coefficients of all methods (with the exception of MCMC) up to lag 20 were trivial.

In terms of parallel computing, the response surface methodology can be used to fit several stochastic volatility models at once. This may be useful to obtain starting values for the model in the next section or, alternatively, in order to perform sequential analysis using batches of increasing size for the same data set. This is useful when the predictive likelihood is sought or different batches have to be analyzed using separate stochastic volatility models. In future research this will most likely be the case with “tall data”, an area that is rapidly evolving but is still focusing on i.i.d. data (Bardenet, Doucet and Holmes, 2015). Given the response surface coefficients, the analysis of a stochastic volatility model with 2,500 observations and 150,000 MCMC iterations takes less than two minutes on a modern desktop running `g77 Fortran` and it is trivial on mainframe computers.

## 5 Multivariate Stochastic Volatility

“*There is variation in the variation in the correlation for some pairs of series*” (Lopes *et al.*, 2011)

### 5.1 Background

Multivariate Stochastic Volatility (MSV) models have developed rapidly in recent years to account for possible co-movement in return covariances. The general model is

$$y_t \sim N_k(0, \Sigma_t), \quad (14)$$

where  $y_t$  denotes now a  $k \times 1$  vector of returns. Many models have been proposed in the literature to account for the temporal evolution of the covariance matrix  $\Sigma_t$ . One model places a factor structure, as  $\Sigma_t = \beta_t H_t \beta_t' + \Psi_t$  where  $\beta_t$  is a  $k \times m$  matrix and  $H_t, \Psi_t$  are diagonal. See Harvey, Ruiz and Shephard (1994), Pitt and Shephard (1999) and Philipov & Glickman (2006a,b). The other possibility is the Dynamic Conditional Correlation model of Engle (2002). The third model, called Cholesky Stochastic Volatility (CSV) is due to Lopes, McCulloch and Tsay (2011) and seems quite promising. In this model it is assumed that

$$\Sigma_t = A_t H_t A_t', \quad (15)$$

where  $H_t = \text{diag}(\exp(h_{t1}), \dots, \exp(h_{tk}))$  and  $A_t^{-1}$  is a lower triangular matrix with ones on the main diagonal and elements  $-\varphi_{t,ij}$  otherwise. Clearly, the Cholesky factor of  $\Sigma_t$  is  $A_t H_t^{1/2}$ , and

$$A_t^{-1} y_t \sim N_k(0, H_t). \quad (16)$$

It turns out from standard properties of the conditional distributions of the multivariate normal that

$$y_{t1} \sim N(0, \exp(h_{t1})), \quad (17)$$

$$y_{tm} \sim N\left(\sum_{j=1}^{m-1} \varphi_{t,jm} y_{tj}, \exp(h_{tm})\right), \quad m = 2, \dots, k. \quad (18)$$

Collect the latent variables in the vector  $\lambda_t = (h_{ti}, \varphi_{t,ij})$  whose dimension is  $d \times 1$  where  $d = \frac{k(k+1)}{2}$ . The temporal evolution of  $\Sigma_t$  in (15) can be modeled through:

$$\lambda_t = \mu + \delta \odot \lambda_{t-1} + \varepsilon_t, \quad (19)$$

where  $\varepsilon_t \sim N_d(0, \Omega)$ , where  $\Omega = \text{diag}(\omega_{(p \times 1)})$  is diagonal and  $\delta, \mu$  are  $p \times 1$ . We also use the more convenient notation

$$h_{ti} = \delta_{h0} + \delta_{h1} h_{t-1,i} + \varepsilon_{ht}, \quad i = 1, \dots, k, \quad (20)$$

$$c_{ti} = \delta_{c0} + \delta_{c1} c_{t-1,i} + \varepsilon_{ct}, \quad i = 1, \dots, \frac{k(k-1)}{2}, \quad (21)$$

where  $c_t = [c_{ti}, i = 1, \dots, \frac{k(k-1)}{2}]$  and  $c_t = \text{vec}(\varphi_{t,jm}, j, m = 2, \dots, k)$ .

How can it be possible to apply the RS technology in this instance? *As a matter of fact, Bayesian inference in the CSV model can be performed easily using tools developed for the univariate SV model.* This is, indeed, evident from (17) and (18) in which we have linear autoregressive models with univariate SV, independent across the return series. The only change that is required in connection with (1) is to consider the slightly more general model:

$$y_t = \beta + \exp\left(\frac{1}{2} h_t\right), \quad (22)$$

that is to introduce a constant term in the model. We can now tell the reader that in fact a constant has been present from the beginning, a fact that also applies to Figure 8. The purpose of the constant term is to accommodate the non-zero conditional means in (18), viz.,

$$\beta = \sum_{j=1}^{m-1} \varphi_{t,jm} y_{tj}.$$

Since the mean is a “conditionally sufficient” statistic for the posterior conditional distributions the  $\varphi_t$  sub-vector of  $\lambda_t$  can be subsumed in  $\beta$  in the course of MCMC. To be more precise, MCMC using the RS technology can be implemented easily as described in the Appendix.

## 5.2 Empirical application

As an empirical application of MSV we consider ten major currencies against the US dollar over the period July 3 1996 to May 21 2012. The currencies are Canadian dollar, Euro, Japanese yen, British pound, Swiss franc, Australian dollar, Hong-Kong dollar, New Zealand dollar, South Korean won and Mexican peso. The data is daily and we considered only the data for which complete observations are available. The data was converted to log differences and a vector autoregression of order 1 (favored by the Schwarz criterion) has been applied to obtain the residuals that constitute the data that we will use here.

In Figure 9 we report the temporal variances in logs (panel a), the temporal covariances (in panel b) and finally in panel c the percentage of total variance explained by the first few factors (principal components). Clearly there is a lot of time-varying volatility as well as co-movement which is also of time-varying character. The first two or three principal components account for a significant part of variability in the data. For this computation we have used 77 consecutive blocks each one containing 50 daily observations. We have experimented by varying the length from 10 to 200 to ensure that Figure 9 does not look qualitatively different.

What is probably impressive in Figure 9 is that along with the temporal variation in volatility and covariances there is also variation in the explanatory power of the factors - basically the first factor alone. As a second example in Figure 10 we provide results for all stocks of FTSE 100, minute data, 23-29/10/2009; the same as used in Plataniotis and Dellaportas (2012)<sup>1</sup>.

In panel (b) we report only 50 out of the 5,050 correlations for visual clarity. In this case as well, correlations vary considerably over time and so does the explanatory power of the first few factors which is, perhaps surprisingly, large. So are the intra-day correlations of returns. The temporal variation in the “explanatory power” can be given an alternative interpretation. If  $\mathbb{X}_{(T \times k)}$  denotes the matrix of returns then the  $k \times k$  sample cross-product matrix  $\mathbb{X}'\mathbb{X}$  contains useful information about the tails of the joint distribution. Using results from an important but relatively unnoticed paper of Meerschaert and Scheffler (1999) we can show that  $\frac{2 \log T}{\log \lambda_i} \xrightarrow{a.s.} \alpha_i$ ,  $i = 1, \dots, k$ , where  $\lambda_i$  is an eigenvalue of  $\mathbb{X}'\mathbb{X}$  and  $\alpha_i$  is the so-called tail index in the direction of the  $i$ th time series, provided  $\mathbb{E} \|\mathbb{X}\|^2 < \infty$  so that the classical central limit theorem applies and the eigenvalues are distinct. From that perspective the temporal variation in “explanatory power” can be interpreted as temporal variation in the tails

---

<sup>1</sup>The author is grateful to Petros Dellaportas and Tassos Plataniotis for providing their data.

Figure 9: Temporal variation in log variances, correlations and explanation by factors, ten currencies

In this Figure we report the temporal variances in logs (panel a), the temporal covariances (in panel b) and finally in panel c the percentage of total variance explained by the first few factors (principal components). Clearly there is a lot of time-varying volatility as well as co-movement which is also of time-varying character. The first two or three principal components account for a significant part of variability in the data. For this computation we have used 77 consecutive blocks each one containing 50 daily observations. We have experimented by varying the length from 10 to 200 to ensure that Figure 9 does not look qualitatively different.

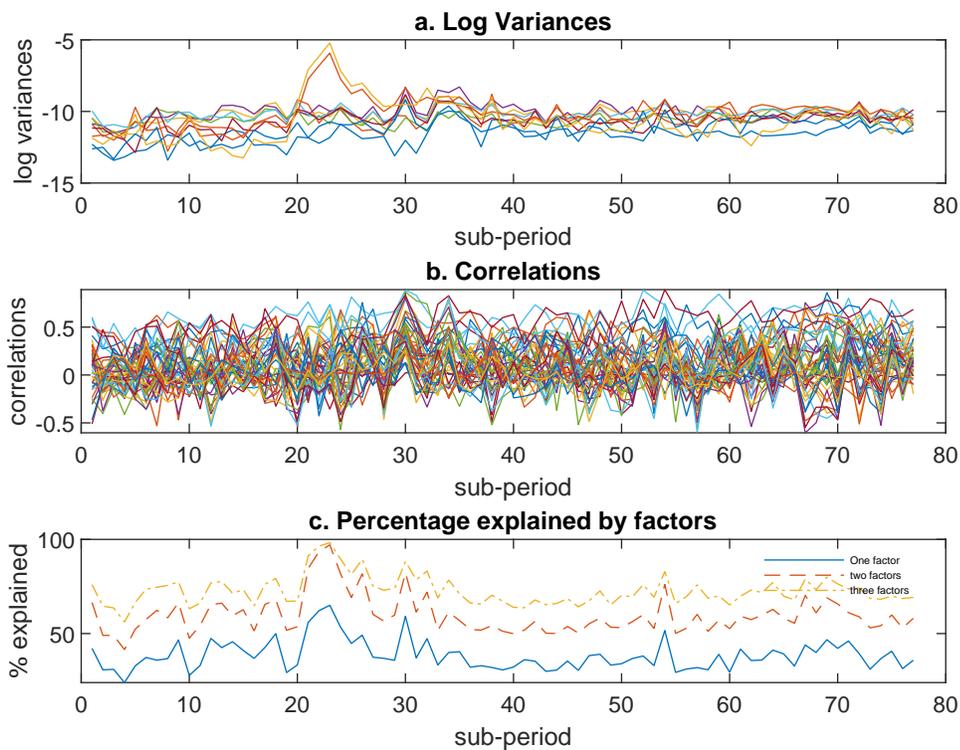


Figure 10: Temporal variation in log variances, correlations and explanation by factors, FTSE100

In this Figure we present marginal posterior densities results for all stocks of FTSE 100, minute data, 23-29/10/2009, the same as used in Plataniotis and Dellaportas (2012).

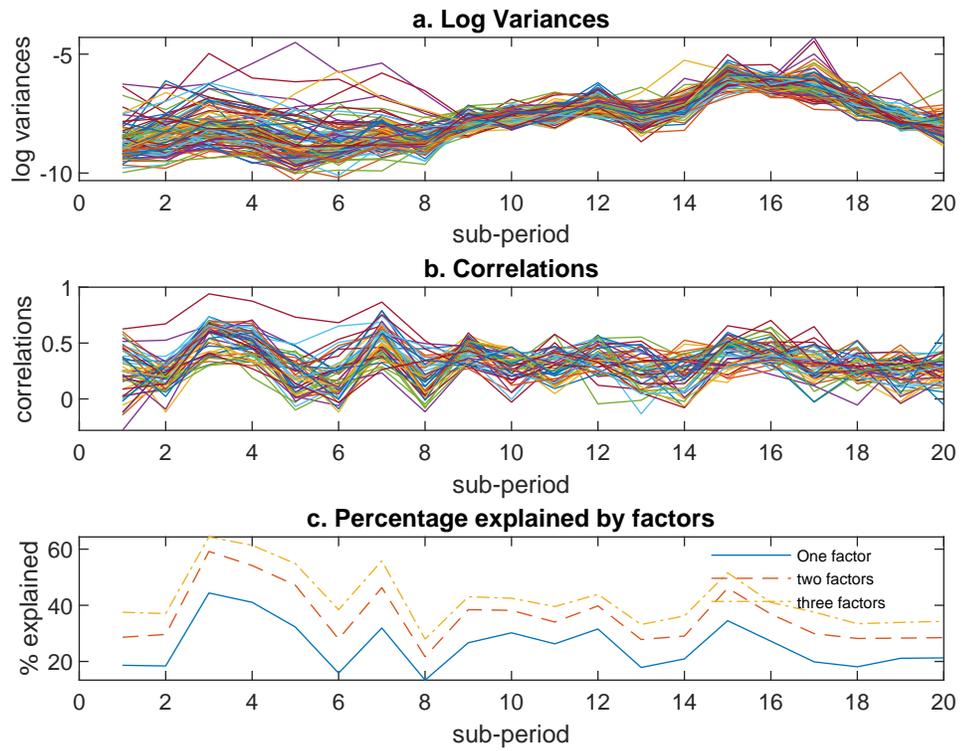
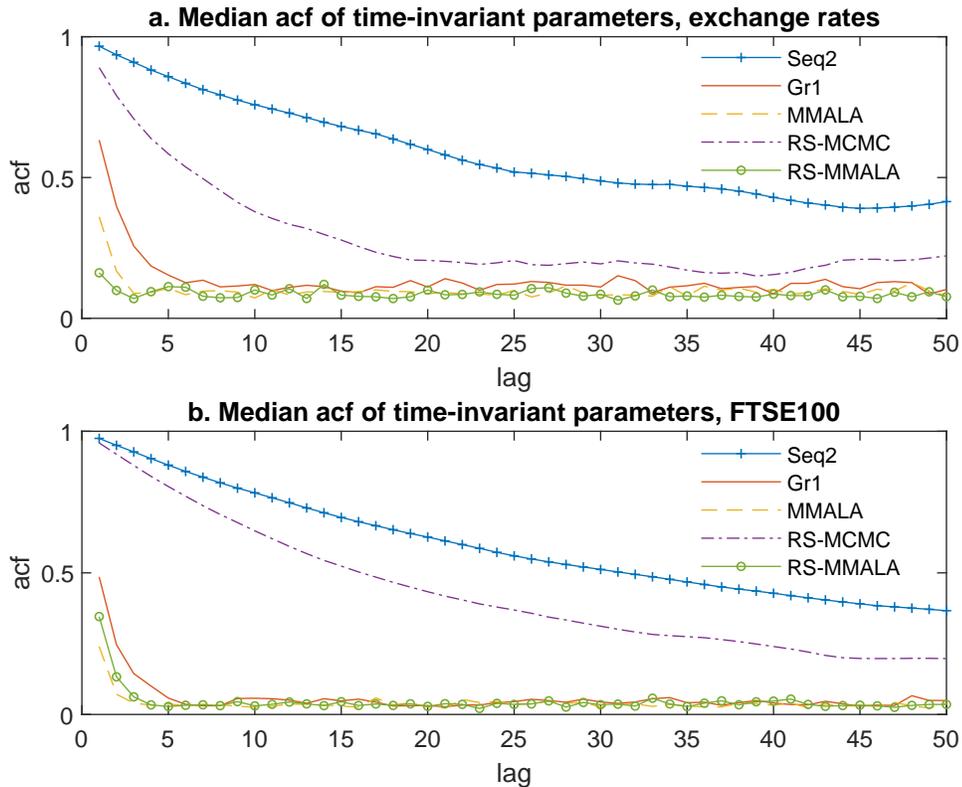


Figure 11: Median autocorrelation functions (acf) of the time-invariant parameters

In this Figure we report the autocorrelation functions (acf) of the time-invariant parameters, viz. the median (across all time-invariant parameters) acf.



of the joint distribution. It is, of course, well known that SV, GARCH and models that allow for leptokurtosis are related in a number of ways and models with fat tails can be considered *grosso modo* as approximate unconditional distributions of SV or GARCH models (de Vries, 1991).

For selected algorithms we report first the autocorrelation functions (acf) of the time-invariant parameters in Figure 11. For a meaningful summary we report the median (across all time-invariant parameters) acf. The definition of the various algorithms is provided in the Appendix.

From the Figure it is evident that **Seq2** is the most detrimental in terms of exploring the posterior distribution followed by **Gr1** which performed acceptably well in exchange rates but not so in FTSE100. MMALA and RS perform rather well. RS-MCMC performs better in the exchange rate data but RS-

MMALA is slightly better in the FTSE100 data. More information about the numerical performance of the different algorithms is obtained by looking at relative numerical efficiency (RNE). We monitor RNE for all parameters (latent or not)<sup>2</sup> and we report histograms in Figure 11 for the exchange rate data (left column) and the FTSE100 data (right column). Evidently, all algorithms are far from showing perfect performance (which corresponds to a value of one) but their behavior is not destructive in terms of exploring the posterior distribution. **Seq2** and **Gr1** perform worst again, as expected from their construction and the evidence in Figures 11 and 12.

Marginal posterior densities of constant terms and autoregressive parameters for volatility and correlations using the exchange rate data are reported in Figures 13 and 14. All results are based on RS-MMALA which seems to perform best in this application.

In Figures 15 and 16 shown are the marginal posterior densities of ten stocks from the FTSE100 for visual clarity. Certain marginal posterior distributions of constant terms and autoregressive parameters are clearly bimodal in many cases for both the stochastic variances and the correlations. This is, of course, due to the prior which “shrinks” the coefficients towards zero with high probability. The effect of the prior is “gentle” in the sense that the posteriors although concentrated around zero in certain cases, they do not have sharp point masses at zero.<sup>3</sup> The effect for both the exchange rate and the FTSE100 data is that many correlations are clearly non-zero and most of them show persistence. The persistence in the FTSE100 data is however clearly less than persistence in the exchange rate data but persistence in correlations is evidently more disparate and can be both positive and negative. Relative to Lopes et al. (2010) where most posterior densities are unimodal the datasets examined herein are less “informative” in the sense that prior information about “shrinkage” towards zero has a noticeable effect on the posteriors. Results not reported here show that marginal posterior densities of  $\omega_j$ s are unimodal and strongly reject the notion that they can be close zero. This is partly due to the value of  $\bar{c} = 10^{-6}$  in (24). Higher values of this parameter however produced marginal posteriors which were bimodal but away from zero suggesting, again, that stochastic volatility and stochastic correlations are important features of the data that cannot be dispensed. This is a first indication that MSV models perform better than multivariate GARCH models and should be preferred in the light of the data.

---

<sup>2</sup>We exclude the latent model indicators.

<sup>3</sup>That would change if the priors contained a point mass at zero instead of a normal component in (23).

Figure 12: Relative numerical efficiency

In this Figure we report RNE (relative numerical efficiency) for the exchange rate and FTSE100 data sets. RNE refers to median across all parameters excluding the latent indicators.

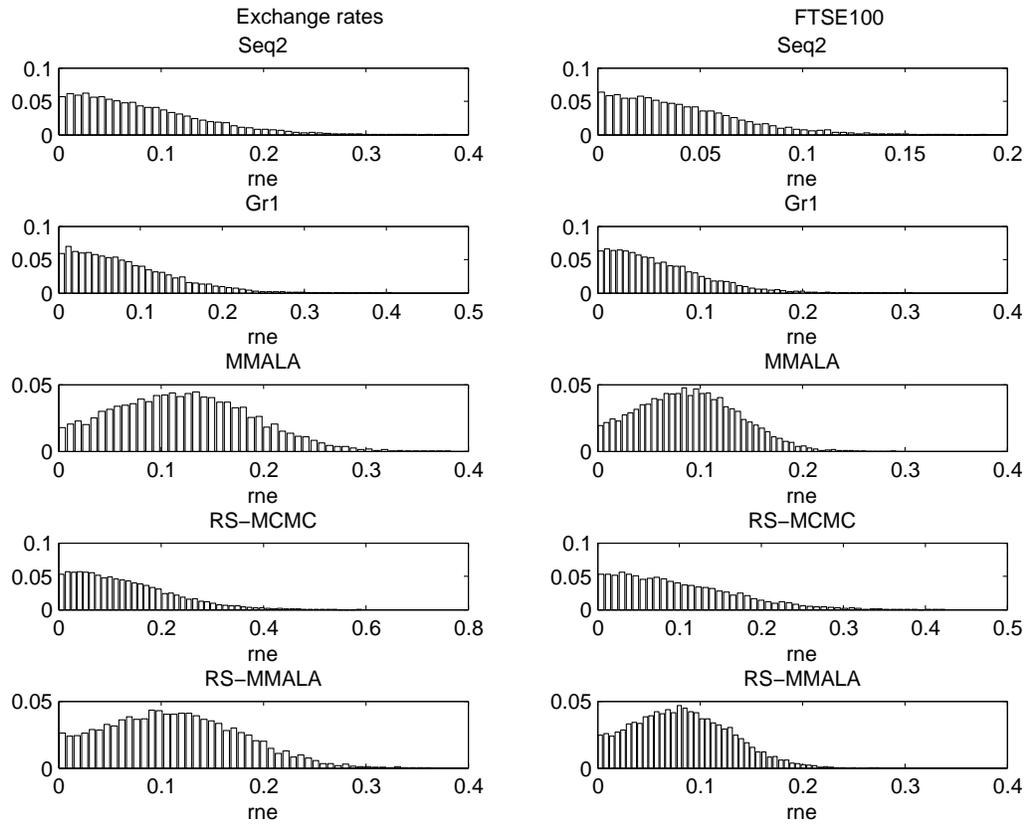


Figure 13: Marginal posterior densities of  $\delta_h$ , exchange rate data

In this Figure we report marginal posterior densities for constant and autoregressive terms of volatility using the exchange rate data set.

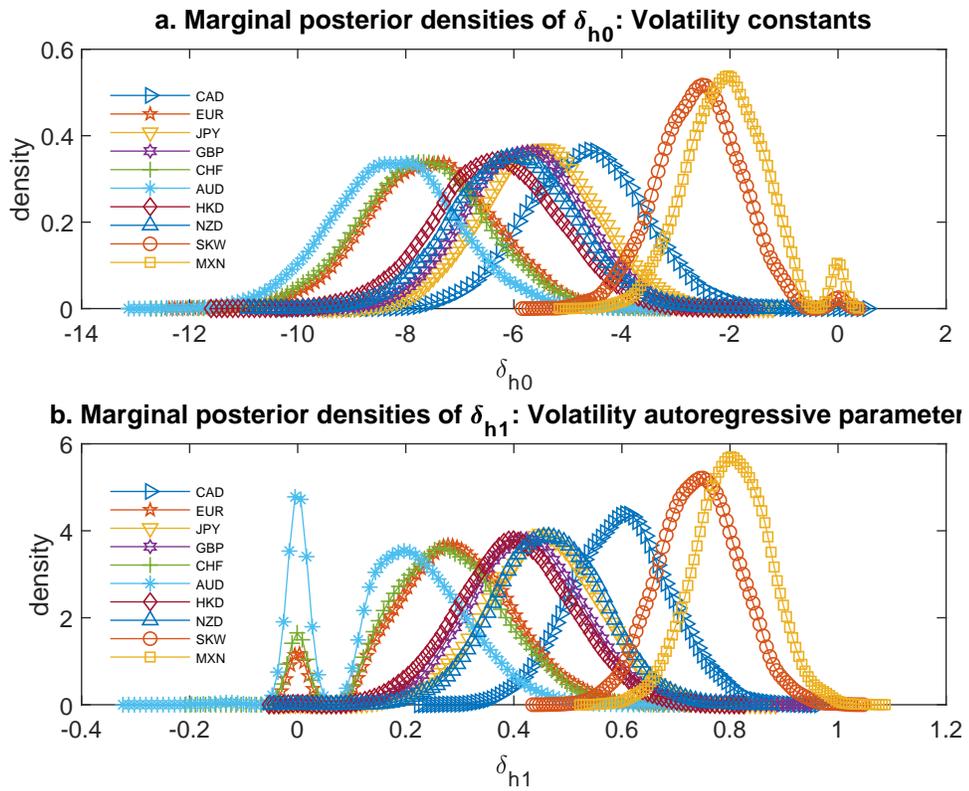


Figure 14: Marginal posterior densities of  $\delta_c$ , exchange rate data

In this Figure we report marginal posterior densities for autoregressive parameters of correlations using the exchange rate data set.

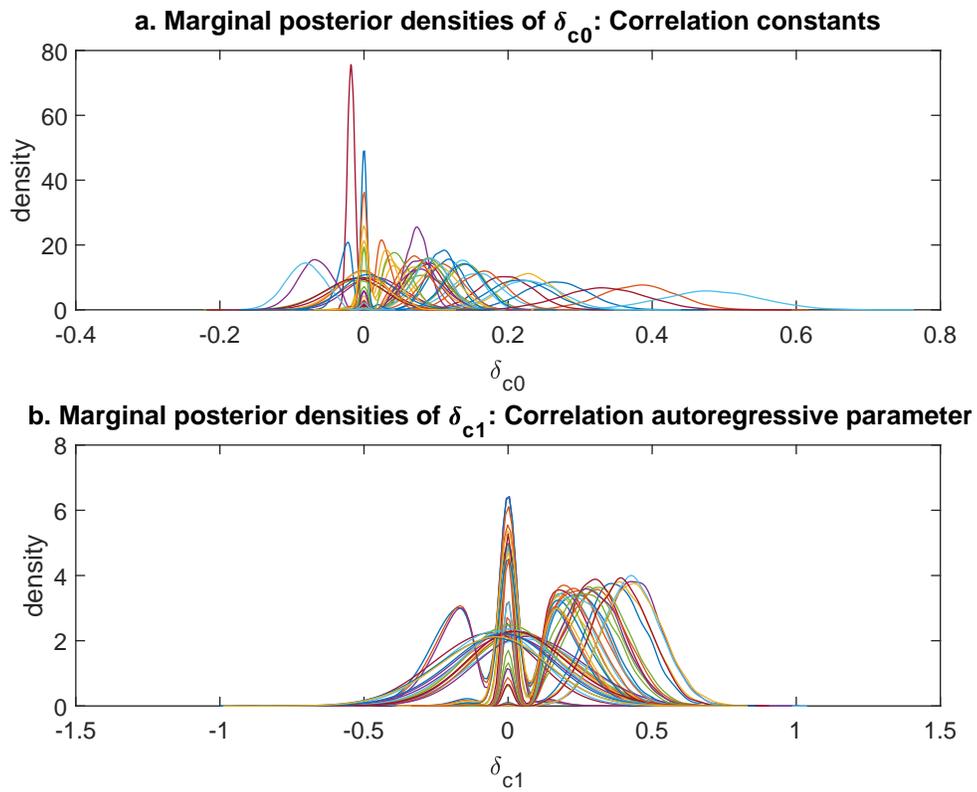


Figure 15: Marginal posterior densities of  $\delta_h$ , FTSE100 data

In this Figure we report marginal posterior densities for constant terms (upper panel) and autoregressive parameters (lower panel) for the multivariate stochastic volatility process using the FTSE100 data set.

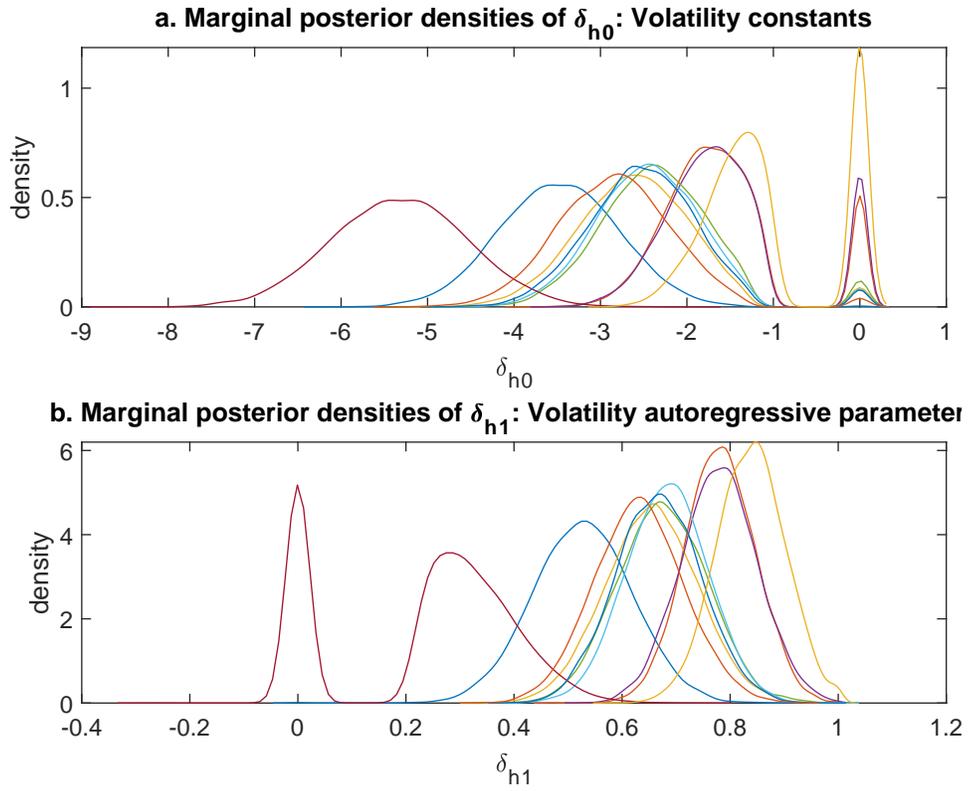
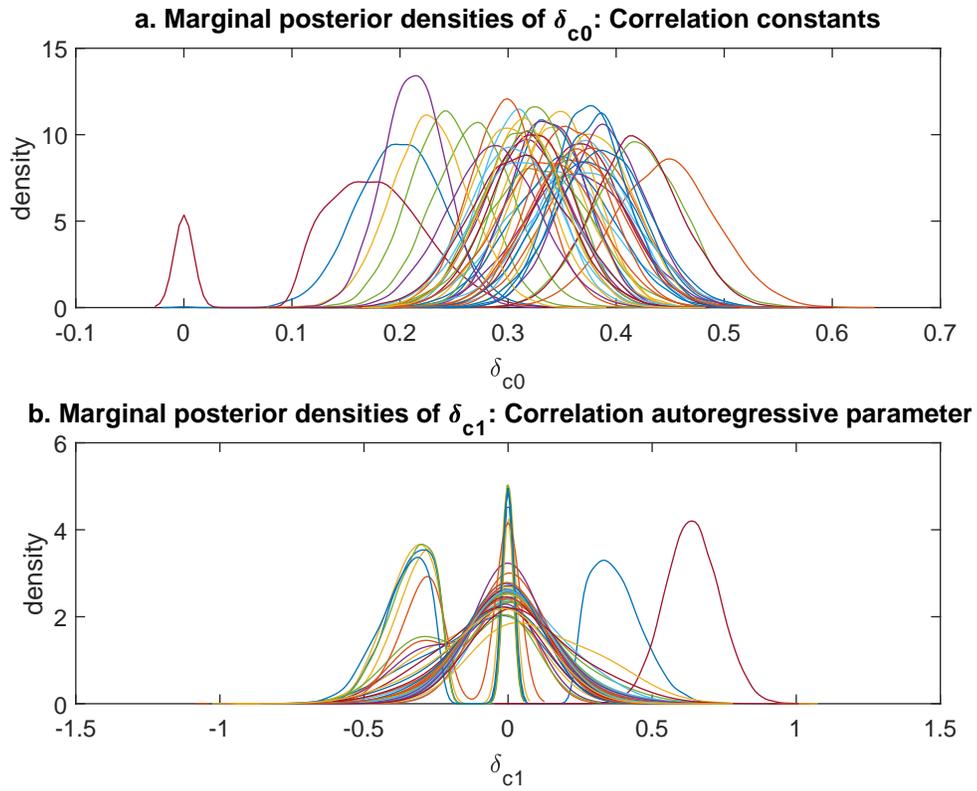


Figure 16: Marginal posterior densities of  $\delta_c$ , FTSE100 data

In this Figure we report marginal posterior densities for constant terms (upper panel) and autoregressive parameters (lower panel) for the correlation part of the multivariate stochastic volatility process using the FTSE100 data set.



Formal comparisons with multivariate GARCH models can be made using marginal likelihoods and Bayes factors. We use a multivariate GARCH model along the lines of the Engle and Kroner (1995) and BEKK formulation. In (15) we assume:

$$\Sigma_t = A_0 A_0' + A_1 y_{t-1} y_{t-1}' A_1' + B_1 \Sigma_{t-1} B_1', \quad (23)$$

where  $A_0$  is lower triangular,  $A_1, B_1$  are  $k \times k$  matrices which are left unrestricted and  $\Sigma_0$  is treated as parameter. Moreover, in relation to (15)  $A_t = A$  is a lower triangular matrix whose elements are treated as fixed parameters. A straightforward generic MCMC procedure is used to provide inferences for the model: Updates are component-wise and use uniform proposals whose bounds are adapted during the “burn-in” phase to achieve acceptance rates close to 25% using starting values from maximum likelihood estimation. Given posterior samples (obtained using 1,5 million iterations, a burn-in of length half a million and thinning every other tenth draw) the marginal likelihood is computed using the Laplace approximation of DiCiccio et al. (1997). Computation of the marginal likelihood for the MSV is facilitated by (17) and (18) along with step (i) and the accurate approximations in (6).

Since the choice of priors is critical in marginal likelihood computations we proceed as follows. Suppose  $p(\theta)$  is the prior of the MSV model and  $p(\beta)$  is the prior of the GARCH, where  $\beta$  is the vector of parameters in the GARCH model. The prior  $p(\theta)$  has been explained in (23) and (24) and we retain it as is since it is costly to change it at this stage. Suppose the prior of  $\beta$  is a mixture-of-normals of the form:  $\beta \sim \sum_{g=1}^G p_g \xi_g$ , where  $\xi_g \sim N(\bar{\beta}_g, \bar{V}_g)$ . Since prior simulation is feasible for both models consider a vector of summary statistics  $S(\tilde{y}^n; \lambda)$  that can be computed from the MSV and GARCH models respectively where  $\tilde{y}^n$  denotes in both cases a matrix of simulated data whose dimensions match those in the data, and  $\lambda$  denotes the parameters. Suppose  $\bar{V}_g = \bar{C}_g \bar{C}_g'$ ,  $g = 1, \dots, G$  and collect the parameters  $\bar{\Pi}_G = [\bar{\beta}_g', \text{vech}(\bar{C}_g)']$ ,  $g = 1, \dots, G$  where  $\text{vech}$  is the vectorization operator which excludes the zero elements of  $\bar{C}_g$ . Given synthetic data  $\tilde{y}^n$  from the MSV model the prior parameters of the GARCH are obtained as follows:

$$\bar{\Pi}_G = \underset{\mathbb{R}^n \times \Theta}{\text{argmin}} : \|S(\tilde{y}^n; \theta) - S(\tilde{y}^n; \bar{\Pi}_G)\|_{\Omega} p(\theta) p(\tilde{y}^n; \theta) d\tilde{y}^n d\theta, \quad (24)$$

where the norm  $\|x\|_{\Omega} = x' \Omega^{-1} x$  is used. For the summary statistics we use an extension of those developed by Ruiz (1994) for GMM estimation of univariate stochastic volatility models. In (25) we minimize the expected weighted distance of the summary statistics between MSV and GARCH where

the expectation is taken with respect to the prior of MSV parameters as well as the synthetic data. To solve this problem we have to decide the value of  $G$  and impose identification (avoid artificial multiple minima) by assuming  $p_1 \geq \dots \geq p_G$ . Then we follow the procedure below to compute the value of the objective function for a given  $\bar{\Pi}_G$ .

#### Computation of Objective Function

- (i) Draw parameters  $\theta^{(p)} \in \Theta$  from the prior,  $p(\theta)$ ,  $p = 1, \dots, P$ .
- (ii) Draw  $n_S$  synthetic data from  $p(y^n; \theta)$ , viz. the MSV model,  $\{\tilde{y}^{n,(s)}, s = 1, \dots, n_S\}$
- (iii) Compute  $d(\theta^{(p)}; \bar{\Pi}_G) = n_S^{-1} \sum_{s=1}^{n_S} \|S(\tilde{y}^{n,(s)}; \theta^{(p)}) - S(\tilde{y}^{n,(s)}; \bar{\Pi}_G)\|_{\Omega}$ .
- (iv) Compute  $d_{P,n_S}(\bar{\Pi}_G) = P^{-1} \sum_{p=1}^P d(\theta^{(p)}; \bar{\Pi}_G)$ .

Given the value of the objective function  $d_{P,n_S}(\bar{\Pi}_G)$  it can be minimized with respect to  $\bar{\Pi}_G$  using standard numerical procedures<sup>4</sup>. To minimize computational costs the matrix  $\Omega$  is selected as the covariance matrix of GMM applied to the actual data using as moment conditions those corresponding to the summary statistics. The reason is that in the course of simulations the GMM covariance matrix becomes numerically singular and regularization was found to affect significantly the final results. We use  $P = n_S = 1,000$ . Several other numerical issues arise as  $G$  increases beyond one or two and we decided it is best to approximate a solution in the following alternative way:

#### Alternative Approximate Solution

- (a) Use steps (i) and (ii) above.
- (b) For each  $\theta^{(p)}$  and each  $\tilde{y}^{n,(s)}$  use ML to fit the GARCH model obtaining the parameters  $\Pi_n^{(p,s)}$ .
- (c) For each  $p$ , track  $\hat{\Pi}_n^{(p)} = n_S^{-1} \sum_{s=1}^{n_S} \Pi_n^{(p,s)}$ ,  $p = 1, \dots, P$ .
- (d) Regress  $\{\hat{\Pi}_n^{(p)}, p = 1, \dots, P\}$  on covariates  $\{W^{(p)}, p = 1, \dots, P\}$  and obtain the coefficients  $b_{n,(P,S)}$ .

Drafting  $\Omega$ , the values of sample size and dimensionality of the data to the exchange rate series, we found that the two algorithms give different results so we decided to keep both for prior elicitation in the GARCH model. Specifically the first algorithm gives results that do not change significantly beyond  $G = 5$  groups provided we use the structured covariance matrices  $\bar{V}_g$ . The probabilities  $p_g, g > G$  tend to zero so we have some indication that this algorithm performs well and gives acceptable results. In step (b) of the alternative algorithm a standard implementation of conjugate-gradients is iterated to

<sup>4</sup>The procedure is superficially similar to indirect inference as developed by Gouriéroux *et al* (1993). However the objective here is to calibrate a common prior that makes sense under both GARCH and SV models.

convergence within  $10^{-6}$ . As we mention below we have found that various shortcuts do not work well; for example it is not possible to use less stringent convergence criteria (like 0.01) although it is quite encouraging that we can use *one iteration* of the Gauss-Newton method away from the first successful convergence of conjugate-gradient algorithm for a particular data set (given the same parameter values of the SV process).

The choice of the matrix of regressors in  $\{W^{(p)}, p = 1, \dots, P\}$  is motivated by considerations of parsimony, ease of use, fit but also the stationarity conditions that must be satisfied by the GARCH parameters. Stationarity is imposed numerically by rejecting parameter draws for which  $n_S^{-1} \sum_{s=1}^{n_S} \|y_n^{(p)} - y_0^{(p)}\| > \delta$  where the constant  $\delta = 10$  and the Euclidean norm is used. We opt for linear and quadratic functions of  $\theta^{(p)}$  in  $W^{(p)}$ . The algorithms use the same underlying sequence of standard (normal and uniform) random numbers to ensure that the objective functions and the mapping from  $\theta$  to  $\Pi$  are sufficiently smooth.

In Figure 17 we present average volatility and the Bayes factors in favor of the MSV model and against the multivariate GARCH for the exchange rate data. By average volatility we mean the simple average across all ten currencies of individual squared deviations from the median. For better visual presentation we plot every other 10th time period. The (predictive) Bayes factors are computed using predictive likelihoods, see Geweke (1999) and Geweke and Amisano (2008).

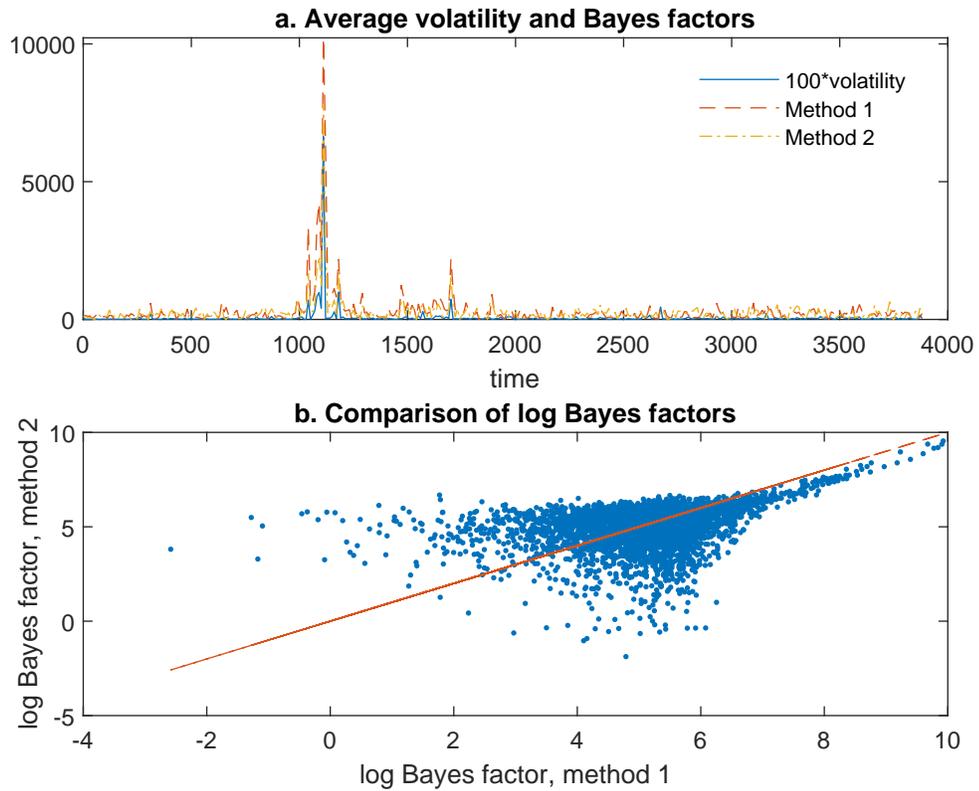
From the Bayes factors in panel (a) it is evident that they provide strong support for the MSV model against the multivariate GARCH particularly during episodes of high volatility. This shows that in all likelihood the multivariate GARCH leaves much to be desired. In panel (b) we compare the log Bayes factors of the two methods. The line corresponds to  $45^\circ$ . The Bayes factors are not the same although they are of the same order of magnitude. Their relationship is more complicated than the one predicted by straight lines or simple functions and probably indicates that the two methods provide different benchmark priors for the GARCH model<sup>5</sup>. Part of the problem is that (i) we have used a data-based selection for  $\Omega$  and (ii) we allowed only Gauss-Newton iteration away from a given parameter vector in method 2. We have not tested the performance of Bayes factors in large-scale multivariate models where these assumptions are removed but we have done so extensively in univariate models and models with  $k = 5$  as we explain below. As a general conclusion when we abstract from (ii) log Bayes factors differ by less than 1.5 which is well within the range in Figure 17b although non-trivial in itself. Regarding (i) the choice of  $\Omega$  does not seem extremely important: Using regularization to

---

<sup>5</sup>The initial variances of GARCH are treated as parameters.

Figure 17: Bayes factors in favor of MSV against GARCH

In this Figure we present average volatility and the Bayes factors in favor of the MSV model and against the multivariate GARCH for the exchange rate data. By average volatility we mean the simple average across all ten currencies of individual squared deviations from the median. For better visual presentation we plot every other 10th time period. The (predictive) Bayes factors are computed using predictive likelihoods, see Geweke (1999) and Geweke and Amisano (2008).



obtain  $\Omega$  as the covariance of GMM moment conditions the Bayes factors differ by less than 1.2 on the average provided the regularization parameter is sufficiently small; in practice less than about 10% of the smallest (numerically non-zero) eigenvalue of the original covariance matrix. Some relevant results are summarized in Table 2.

In Table 2 we report the difference in log Bayes factors (or percentage difference in Bayes factors) from different numerical approaches to the problem. The numbers are average differences and standard errors appear in parentheses. To minimize computational costs we have focused on models with  $k = 1$  and  $k = 5$  time series and  $n_y = n_\theta = 100$ . The Bayes factors are based on marginal likelihoods for the full sample using  $n = 1,500$  observations and individual stochastic variance processes for each of the  $k$  time series whose autoregressive parameters are 0.9, the standard deviations are 0.1 and the constant terms are  $0.3(1 - \rho)$ . In Table 2 `tol` refers to the tolerance in the conjugate-gradients algorithm and `1gauss.newton` is the algorithm that uses one Gauss-Newton iteration away from the first successful application of the conjugate-gradients algorithm iterated to full convergence which is taken as `tol` =  $10^{-6}$ . Moreover, `fixed.init` means that the initial conditions for  $\Sigma_0$  are taken to be fixed and equal to the sample covariance matrix. Data sets for which convergence takes too long (more than 100 iterations) are discarded and new data are generated. Generally the differences in Bayes factors are substantial when shortcuts are used but significantly less so when one Gauss-Newton iteration is used when the initial conditions are treated either as fixed or as parameters in the optimization. Therefore it seems reasonable to proceed using one Gauss-Newton iteration with fixed initial covariance matrix.

It would be a welcome addition to future research to examine the performance of more parsimonious multivariate GARCH specifications as alternatives to the full BEKK, for example  $\Sigma_t = A_0 A_0' + A_1 \odot y_{t-1} y_{t-1}' + B_1 \odot \Sigma_{t-1}$  where  $A_1 = \gamma \gamma'$  and  $B_1 = \delta \delta'$  for certain vectors  $\gamma$  and  $\delta$  which are likely to be useful in large-scale MSV models although we do not expect them to be accurate approximations unless the number of lags is significantly increased given the evidence in Monfardini (1997).

In the univariate case, let us consider typical sampling distributions of estimates  $\hat{\beta}$  and  $\hat{\gamma}$  in the GARCH model:

$$h_t = \alpha + \beta h_{t-1} + \gamma y_{t-1}^2,$$

under the various specifications of numerical algorithms in Table 2. For `1gauss.newton` the initial

Table 2: Difference in log Bayes factors for different numerical methods

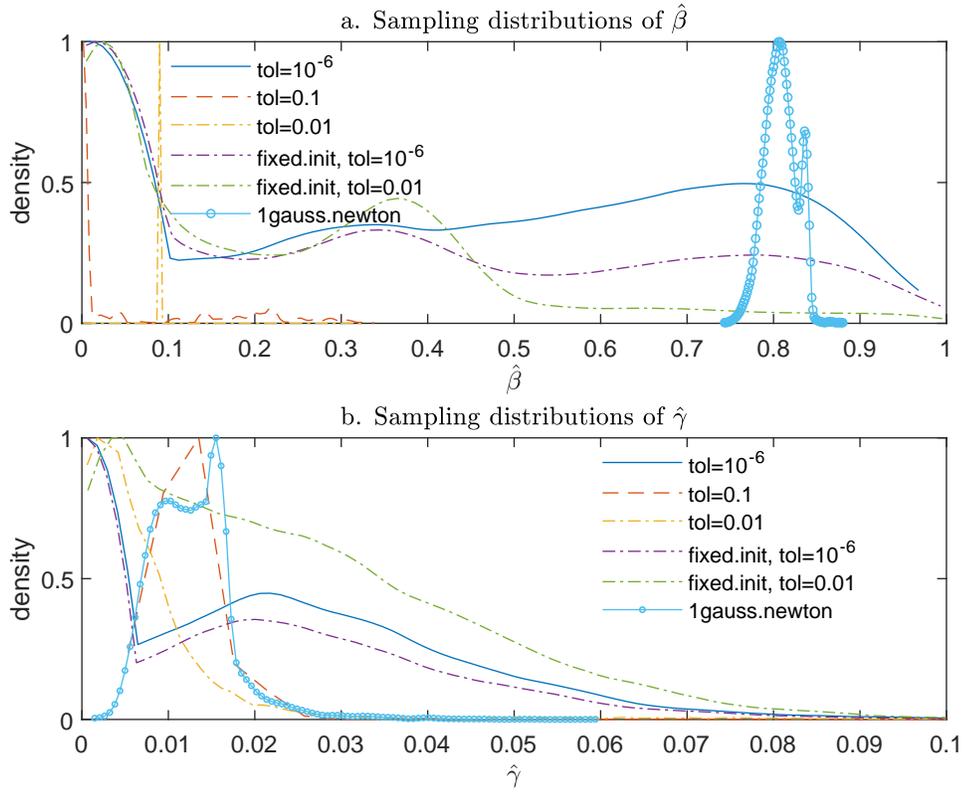
numerical method	difference in log Bayes factors
$k = 1$	
tol = 0.1	25.72 (4.45)
tol = 0.01	27.12 (4.50)
fixed.init, tol = 0.1	47.91 (15.12)
fixed.init, tol = 0.01	44.05 (16.31)
1gauss.newton	7.23 (5.17)
fixed.init, 1gauss.newton	5.12 (4.81)
$k = 5$	
tol = 0.1	89.67(22.15)
tol = 0.01	95.12(24.17)
fixed.init, tol = 0.1	82.11(31.12)
fixed.init, tol = 0.01	88.01(33.2)
1gauss.newton	12.15(4.33)
fixed.init, 1gauss.newton	10.50(4.02)

variance is treated as a parameter. In Figure 18 we present typical sampling distributions of estimates  $\hat{\beta}$  and  $\hat{\gamma}$  using 10,000 replications when the true generating process is a univariate stochastic variance model of the form  $h_t = 0.06 + 0.8h_{t-1} + 0.1\xi_t$ ,  $h_0 = 0.3$  and sample size 1,500. Overall there are significant differences among the various approximations and (perhaps consistent with the evidence in Table 2) `1gauss.newton` performs quite well taking into account the persistence of conditional variances from the SV process.

Clearly, the form of certain sampling distributions is due to the fact that we need to impose non-negativity constraints upon the GARCH parameters. Even if we iterate to convergence the conjugate-gradient algorithm does not distinguish sharply between the different parameter values particularly for  $\hat{\beta}$ . The fact that even one iteration of the Gauss-Newton algorithm performs extremely well is of course quite encouraging in this respect can probably be of further use in applications of indirect inference and not just the elicitation of common priors for the GARCH and SV models.

Figure 18: Sampling distributions of GARCH approximating parameters

In this Figure we present typical sampling distributions of estimates  $\hat{\beta}$  and  $\hat{\gamma}$  using 10,000 replications when the true generating process is a univariate stochastic variance model of the form  $h_t = 0.06 + 0.8h_{t-1} + 0.1\xi_t$ ,  $h_0 = 0.3$  and sample size 1,500. Overall there are significant differences among the various approximations and (perhaps consistent with the evidence in Table 2) `1gauss.newton` performs quite well taking into account the persistence of conditional variances from the SV process.



## 6 An application to time-varying parameters stochastic volatility models

We compare with the model of Chan (2007) which is a stochastic volatility-in mean-model (SVM) with time-varying parameters (TVP):

$$\begin{aligned}
 y_t &= \tau_t + \alpha_t e^{h_t} + \varepsilon_t^y, \quad \varepsilon_t^y \sim N(0, e^{h_t}), \\
 h_t &= \mu + \beta y_{t-1} + \phi(h_{t-1} - \mu) + \varepsilon_t^h, \quad \varepsilon_t^h \sim N(0, \sigma^2), \\
 \gamma_t &= \gamma_{t-1} + \varepsilon_t^\gamma, \quad \varepsilon_t^\gamma \sim N_2(0, \Omega), \\
 \gamma_t &\equiv [\alpha_t, \tau_t]'.
 \end{aligned} \tag{25}$$

Matrix  $\Omega$  is  $2 \times 2$  and  $\tau_t$  is a random-walk component in the mean level of  $y_t$ . In this model, we allow for the possibility of volatility feedback—volatility may impact the level of inflation. In addition, we also allow past inflation to affect the current inflation volatility. Chan (2017) described a method to apply MCMC in U.S. inflation. Chan’s (2017) method takes only 31 seconds to produce 10,000 draws. The response surface methodology depends on parameters

$$\theta = [\mu, \beta, \phi, \sigma^2, \text{vech}(\Omega)]. \tag{26}$$

This involves 6 parameters but in addition to  $h_t$  we have the latent variable  $\tau_t$  as well. We construct a response surface using values of the parameters in  $\theta$  *randomly* drawn  $M = 1,000$  times from the appropriate bounds (roughly the same as the posterior means and posterior standard deviations in Chan, 2017). Then we use an EM algorithm to estimate (25) and we save the estimated  $\theta$ , say  $\hat{\theta}^{(j)}$  for  $j = 1, \dots, M$  and the given sample size in Chan (2017) which is from 1948.Q1 to 2013.Q2,  $T = 262$ ). Relative to a full response surface methodology, here we take a shortcut as we use only  $M$  values of the parameters and we do not use MCMC at all. Computations for all  $M$  points can be parallelized easily and for the Chan (2017) data, timing was about 2 minutes on a standard desktop PC but under 10 seconds on a mainframe computer in which Chan’s (2017) MCMC took, nearly, 10 seconds as well since MCMC cannot be parallelized. In Table xx we report Chan’s results as well as ours (denoted RSM). Moreover, we denote

$$\Omega = \begin{bmatrix} \omega_\alpha^2 & \omega_{\alpha\tau} \\ \omega_{\alpha\tau} & \omega_\tau^2 \end{bmatrix}.$$

Table 3: Posterior moments

	Chan (2017)	RSM
$\mu$	0.121 (0.947)	0.120 (0.946)
$\beta$	0.003 (0.005)	0.003 (0.004)
$\phi$	0.963 (0.021)	0.962 (0.022)
$\sigma^2$	0.072 (0.020)	0.071 (0.021)
$\omega_\alpha^2$	0.044 (0.039)	0.044 (0.041)
$\omega_{\alpha\tau}$	0.011 (0.033)	0.011 (0.033)
$\omega_\tau^2$	0.117 (0.057)	0.119 (0.055)

Notes: Posterior standard deviations in parentheses.

Additionally, we use the “Alternative Approximation” discussed in the previous section. In Table 3 we report posterior means and posterior standard deviations using our RSM as well as the results taken from Chan (2017). Posterior moments are quite similar to those of Chan (2017). In Table 4, we report log marginal likelihood (LML) and numerical standard error (NSE). Although LMLs for different models are nearly the same, as expected, NSEs are an order of magnitude lower than the NSEs of Chan (2017). The reason’s is that use of MCMC in Chan (2017) produces highly autocorrelation and inefficiency factors (IF) range between 4 and 164, our results i.i.d draws from the posterior so, theoretically, the IF is 1. In practice, computed IFs ranged between 1.01 and 1.04.

Table 4: Estimated log marginal likelihoods for the unobserved components model (UC), the UC-SVM model with constant coefficients (UC-SVM-const), and the UC-SVM model with time-varying coefficients (UC-SVM).

	Chan (2017)		RSM	
	LML	NSE	LML	NSE
UC-SVM	-536.4	0.08	-536.5	0.005
UC-SVM-const.	-550.5	0.08	-550.4	0.002
UC	-550.4	0.03	-550.4	0.002

Notes: LML is log marginal likelihood and NSE is numerical standard error.

## Conclusions

In this paper we have provided simple but highly accurate response surface methodologies to estimate univariate and multivariate stochastic volatility models by either maximum likelihood or MCMC. The approximation is based on properties of the likelihood function, namely the approximation of a certain expectation with respect to a multivariate normal distribution. Results are presented in certain detail and the new methods are applied to artificial data as well as ten exchange rates and all stocks of FTSE100 using daily data.

## APPENDIX. Technical details for multivariate stochastic volatility model

Given the  $\mathcal{F}$  matrix from univariate SV response surface and parameters  $\theta = [\alpha', \varphi', \omega']'$ :

(i) Obtain the log likelihood in (17) and (18) for each  $m = 1, 2, \dots, k$ . The total log likelihood is the sum of the individual log likelihoods for each time series.

(ii) Apply any MCMC update scheme to obtain a new draw for  $\theta$ .

It is clear from the empirical application and the discussions in Lopes *et al* (2011) that the major burden in MCMC is drawing the latent variables in  $\lambda_t$ . Although this can be done in parallel it does

not mean that overall autocorrelation of the chain is unaffected. In Lopes *et al* (2011) the problem is that it is relatively time consuming to draw the  $h_t$ s from their posterior conditional distributions but then the drawing of  $\theta$  is quite easy. In the scheme described here we need an efficient way to draw  $\theta$  once the latent states have been integrated out; we note again that integration with respect to latent states is very easy using the RS technique.

With  $k = 20$  series in (14) we end up with  $p = 210$  elements in (19). The total number of parameters is 730. We think it is pointless to consider a general vector autoregression of the form  $\lambda_t = \alpha + \Phi\lambda_{t-1} + \varepsilon_t$  since in this case  $\Phi$  would contain 44,100 elements. With  $k = 50$  series we have 3,825 parameters. This is not by itself prohibitive but it precludes MCMC updates from standard distributions (like the normal, Wishart etc). So we have to devise special schemes to perform the computations. First of all we introduce “model selection priors” as in Lopes *et al* (2011). Suppose  $\theta = [\alpha', \varphi', \omega']' = [\gamma', \omega']$ . For each element of  $\gamma$  we assume

$$\gamma_j \stackrel{d}{=} \mathcal{B}\bar{\sigma}\xi + (1 - \mathcal{B})\bar{c}\zeta, \mathcal{B} \sim \text{Bernoulli}(p) \quad (\text{A.1})$$

with  $p = \frac{1}{2}$ ,  $c = 10^{-3}$ ,  $\bar{\sigma} = 10$ ,  $\zeta \sim N(0, 1)$  and  $\xi \sim N(\mu, 1)$ , where  $\mu = 1$  if  $\gamma_j$  is of the  $\varphi$ -variety and  $\mu = 0$  if it is of the  $\alpha$ -variety. This places a random-walk prior on the elements of the Cholesky matrix. For the elements of  $\omega$  we deviate somewhat from Lopes *et al* (2011) and assume

$$\omega_j \stackrel{d}{=} \mathcal{B}U(0, \bar{a}) + (1 - \mathcal{B})\bar{c}\zeta \quad (\text{A.2})$$

where  $\bar{a} = 10$ ,  $\zeta \sim N(0, 1)$  and  $\bar{c} = 10^{-6}$ . Since the  $\mathcal{B}$ s can be treated as latent indicators, (21) and (22) can be accommodated easily in MCMC. Due to the presence of the common “selector”  $\mathcal{B}$  these priors are not independent. In order now to devise new computational tools for the analysis of MSV-CSV models we notice that  $\theta$  appears in the likelihood of (14), (17)-(19) in a sequential form. Indeed, by numbering the constant terms and autoregressive coefficients in a more obvious way we have:

For the first time series,

$$y_{t1} \sim N(0, \exp(h_{t1})), h_{t1} = \alpha_1 + \rho_1 h_{t-1,1} + \sigma_1^{1/2} \xi_{t1}.$$

For the second and third series,

$$y_{t2} \sim N(\varphi_{t1} + \varphi_{t2}y_{t-1,1}, \exp(h_{t2})), h_{t2} = \alpha_2 + \rho_2 h_{t-1,2} + \sigma_2^{1/2} \xi_{t2},$$

$$y_{t3} \sim N(\varphi_{t3} + \varphi_{t4}y_{t-1,1} + \varphi_{t5}y_{t-1,2}, \exp(h_{t3})), h_{t3} = \alpha_3 + \rho_3 h_{t-1,3} + \sigma_3^{1/2} \xi_{t3}, \quad (\text{A.3})$$

and finally,

$$y_{tk} \sim N(\varphi_{t,k(M-k+1)} + \varphi_{t,k(M-k+2)}y_{t-1,1} + \dots + \varphi_{t,M}y_{t-1,k}, \exp(h_{tk})),$$

$$h_{tk} = \alpha_k + \rho_k h_{t-1,k} + \sigma_k^{1/2} \xi_{tk},$$

where  $M = \frac{(k-1)(k+2)}{2}$  is the total number of constant terms and autoregressive coefficients required

above, and  $\xi_{ti} \sim N(0, 1)$ ,  $i = 1, \dots, k$ . The parameters  $\tilde{\theta}_1 = (\alpha_1, \rho_1, \sigma_1)$  of the first equation can be obtained directly using the RS technique for univariate SV models. In the second equation we can do the same with  $\tilde{\theta}_2 = (\alpha_2, \rho_2, \sigma_2)$  provided we set  $\beta = \varphi_{t1} + \varphi_{t2}y_{t-1,1}$ . Given the values of  $(\phi_{t1}, \phi_{t2})$  this can be done easily. In fact the same is true for all equations and for the last one we can apply the RS technique with  $\tilde{\theta}_k = (\alpha_k, \rho_k, \sigma_k)$  conditional on the value of

$$\beta = \varphi_{t,k(M-k+1)} + \varphi_{t,k(M-k+2)}y_{t-1,1} + \dots + \varphi_{t,M}y_{t-1,k}$$

which in the CSV model acts as a “conditional sufficient statistic”.

Clearly the problem reduces to updating the parameters  $(\delta_i, \omega_i)$  of the processes

$$\varphi_{ti} = \delta_{i1} + \delta_{i2}\varphi_{t-1,i} + \omega_i^{1/2}\zeta_{ti}, \quad i = 1, \dots, k$$

where  $\zeta_{ti} \sim N(0, 1)$ . Conditional on the  $\varphi$ s, standard MCMC methods can be used to update the parameters  $(\delta_i, \omega_i)$  so the only remaining problem is how to draw the states  $(\varphi_{ti}, i = 1, \dots, k)$ . Let us illustrate the procedure for the second equation since the problem is similar for all equations. The conditional posterior distribution of the latent state is

$$p(\varphi_{ti}|\backslash\varphi_{ti}) \propto \exp\left\{-\frac{(\varphi_{ti} - \delta_{i1} - \delta_{i2}\varphi_{t-1,i})^2 + (\varphi_{t+1,i} - \delta_{i1} - \delta_{i2}\varphi_{t,i})^2}{2\omega_i^2}\right\} \mathfrak{f}_t(\varphi_{ti}; \beta_t, y_{i,t-1}^2),$$

for

$$\beta_t \equiv \varphi_{t1} + \varphi_{t2}y_{t-1,1}, i = 1, 2$$

and

$$\mathfrak{f}_t(\varphi_{ti}; \beta, y_{i,t-1}^2) \equiv \mathbb{E}f_t(\varphi_{ti}; \beta, y_{i,t-1}^2).$$

Here,  $\backslash\varphi_{ti}$  denotes conditioning on all other elements of state vectors and structural parameters. More generally for equation  $i$  the latent states are

$$\varphi_{t,(i)} \equiv (\varphi_{t,\nu}, \varphi_{t,\nu+1}, \dots, \varphi_{\nu+i-1}),$$

where  $\nu \equiv \frac{i(i-1)}{2}$ . For the second equation we need to draw latent states  $\varphi_{t,(2)} = (\varphi_{t1}, \varphi_{t2})$ ,  $t = 1, \dots, T$  and for the  $i$ th equation we have to draw  $i$  states.

In this form the states  $(\varphi_{ti}, t = 1, \dots, T)$  can be drawn sequentially by drawing from the first term and using the second term as acceptance probability in a Metropolis-Hastings update. Evaluation of the second term requires one interpolation or table look-up based on the value of  $\beta$ . Due to the sequential drawings of the states it is highly unlikely that this algorithm will explore the posterior distribution quickly and efficiently.

The posterior conditional for the entire state vector  $\varphi_i = (\varphi_{ti}, t = 1, \dots, T)$  is

$$p(\varphi_i|\backslash\varphi_i) \propto \exp\left\{-\frac{\sum_{t=1}^T (\varphi_{ti} - \delta_{i1} - \delta_{i2}\varphi_{t-1,i})^2}{2\omega_i^2} + \sum_{t=1}^T \log \mathfrak{f}_t(\varphi_{ti}; \beta_t, y_{i,t-1}^2)\right\}, i = 1, 2.$$

This form can be used to implement MMALA using the gradient  $\nabla \log p(\varphi_i|\backslash\varphi_i)$  and the Hessian  $\nabla^2 \log p(\varphi_i|\backslash\varphi_i)$ . An alternative is to find an approximation to the mode and the Hessian at the mode and use a Metropolis-Hastings update using a multivariate Student- $t$  with low degrees of freedom.

To conclude we have the following techniques to provide inferences in the context of the MSV model:

- (i) LMT, which is MCMC as proposed by Lopes, McCulloch and Tsay (2011).
- (ii) Seq1, where latent states are drawn *sequentially* from their posterior conditional distributions

$p(\varphi_{ti}|\backslash\varphi_{ti})$  using a MH update with acceptance probability provided by  $f_t(\varphi_{ti}; \beta, y_{i,t-1}^2)$ .

(iii) Seq2, as in (ii) with the difference that the MH update is implemented by drawing from a normal distribution centered at the approximate mode and variance equal to the negative inverse second derivative at the mode.

(iv) Gr1, when latent states are drawn *group-wise* from their posterior conditional distributions  $p(\varphi_i|\backslash\varphi_i)$ , using an acceptance MH update and a multivariate Student-t ( $\nu$ ) proposal where  $\nu = 10$ . The proposal is centered at the approximate mode and the scale matrix is derived from the Hessian at the mode.

(v) Gr2, as in (iv) with the difference that a simple random-walk MH update is used.

(vi) MMALA, which involves drawing the latent states using the Riemannian manifold approach and requires the gradient and Hessian of  $p(\lambda_i|\backslash\lambda_i)$  and all structural parameters.

(vii) RS-MCMC, which is MCMC using the response surface methodology and the updates we described above.

(viii) RS-MMALA, which is MMALA as in (vi) using the response surface methodology instead of relying on the (complete) gradient and Hessian of the log posterior.

In terms of implementation in LMT the major part is naming the coefficients correctly and keeping track of them. Preliminary experiments with a data set of ten major currencies and all stocks in FTSE100 (see next section) indicated no considerable differences between (Seq1, Seq2) and (Gr1, Gr2). The comparison clearly favored the group updates by far, notwithstanding the fact that a random-walk MH is used in Gr2. In RS-MCMC we need approximations to the mode of  $p(\varphi_i|\backslash\varphi_i)$  and its Hessian at the mode. The approximate mode is located using one Newton iteration from the current draw of the states to minimize computational costs. Given the RS the Hessian is easy to compute numerically. In another variant of RS-MCMC we use an adaptive component-wise random-walk MH using uniform proposals for each component / parameter. The bounds of the uniform proposals are adapted during the burn-in phase to target an acceptance rate between 20% and 30%. The two variants behaved similarly so we retained only the first version.

It must be emphasized that implementation of MMALA in the context of CSV requires *exclusively* (i) code that implements MMALA in the univariate case and (ii) using this code *sequentially* for all

time series. RS-MMALA also exploits the particularly convenient structure of CSV to provide quick and efficient updates of the latent states by simple interpolations.

## References

- Baba, Y., Engle, R., Kraft, D. and Kroner, K. (1990). Multivariate simultaneous generalized ARCH, mimeo.
- Bardenet, R., A. Doucet and C. Holmes, 2015, On Markov chain Monte Carlo methods for tall data, arXiv:1505.02827 [stat.ME]
- Box, G. E. P. and N. R. Draper. 1987. Empirical modelbuilding and response surfaces. NewYork: John Wiley & Sons.
- Box, G. E. P. and K. B. Wilson. 1951. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B* 13(1): 1–38.
- Chan, J. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics*, 35 (1), 17–28.
- Chib, S., Nardari, F., Shephard, N., 2002, Markov Chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* 108, 281-316.
- Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics* 61, 375–400.
- de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. *Biometrika* 82, 339–3.
- de Vries, C., 1991, On the relation between GARCH and stable processes. *Journal of Econometrics* 48, 313-324.
- DiCiccio, T.J., Kass, R.E., Raftery, A., Wasserman, L. (1997) Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association* 92: 903-915.
- Engle, R. F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339-350.
- Engle, R.F., Kroner, K., 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.

Genz, A. C., Malik, A. A., 1980, An Adaptive Algorithm for Numerical Integration over an N-dimensional Rectangular Region. *Journal of Computational and Applied Mathematics* 6, 295–302.

Geweke, J. (1992), “Evaluating the Accuracy of Sampling based Approaches to the Calculation of Posterior Moments,” in J.O. Berger, et al. (eds.), *Bayesian Statistics*, Vol. 4, pp. 169-194. Oxford: Oxford University Press.

Geweke, J., 1999, Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication, *Econometric Reviews* 18, 1-73 (with discussion).

Geweke, J., and G. Amisano, 2008, Comparing and evaluating Bayesian predictive distributions of asset returns, ECB working paper 969.

Girolami M., Calderhead B., 2011, Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society B.* (with discussion). 73, Part 2. pp 1-37.

Gourieroux C., Monfort, A., Renault, E. (1993) Indirect inference. *Journal of Applied Econometrics* 8, Supplement, S85-S118.

Harvey, A. C., Ruiz, E., Shephard, N., 1994, Multivariate stochastic variance models, *Review of Economic Studies* 61, 247-264.

Hood, S. J. and P. D. Welch. 1993. Response surface methodology and its application in simulation. In *Proceedings of the 1993 Winter Simulation Conference*, eds. G.W. Evans, M. Mollaghasemi, E. C. Russell and W. E. Biles, 115–122. Piscataway, New Jersey: IEEE Press.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* 12, 371–417.

Kim, S., Shephard, N., Chib, S., 1998, Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 361-393.

Khuri, A. I. and J. A. Cornell. 1996. *Response surfaces: designs and analyses*. New York: Marcel Dekker, Inc.

Knight, J. L., Satchell, S.E., Yu, J., 2002, Estimation of the stochastic volatility model by the empirical characteristic function method, *Australian & New Zealand Journal of Statistics* 44, 319–335.

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

Malik, S., Pitt, M. K., 2011, Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics* 165, 190-209.

McCausland, W.J., 2012, The HESSIAN method: Highly efficient simulation smoothing, in a

nutshell. *Journal of Econometrics* 168, 189-206.

Meerschaert, M.M., Scheffler, H.-P., 1999, Moment estimator for random vectors with heavy tails, *Journal of Multivariate Analysis* 71, 145-159.

Monfardini, C. (1997) Estimating stochastic volatility models through indirect inference, mimeo. European University Institute.

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

Myers, R. H. and D. C. Montgomery. 1995. Response surface methodology: process and product optimization using designed experiments. New York: John Wiley & Sons.

Philipov, A., Glickman, 2006b, Multivariate stochastic volatility via Wishart processes, *Journal of Business and Economic Statistics* 24, 313-328.

Pitt, M., Shephard, N., 1999. Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* 94, 590-599.

Plataniotis, A., Dellaportas, P., 2012, A multivariate stochastic volatility model, Department of Statistics, Athens University of Economics and Business, working paper.

Richard, J.-F., Z. Wang, 2007, Efficient high-dimensional importance sampling. *Journal of Econometrics* 141: 1385-1411.

Ruiz, E., 1994, Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63: 289-306.

Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87, 271-301.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), *Time Series Models with Econometric, Finance and Other Applications*. Chapman & Hall, London, pp. 1-67.

## BANK OF GREECE WORKING PAPERS

277. Anastasiou, D., Z. Bragoudakis and S. Giannoulakis, "Perceived vs Actual Financial Crisis and Bank Credit Standards: Is There Any Indication of Self-Fulfilling Prophecy?", January 2020.
278. Gibson, D. H., S. G. Hall, P. Petroulas, V. Spiliotopoulos and G. S. Tavlas, "The Effect of Emergency Liquidity Assistance (ELA) on Bank Lending During the Euro Area Crisis", January 2020.
279. Tavlas, G., "On the Controversy over the Origins of the Chicago Plan for 100 Percent Reserves", June 2020.
280. Milionis E. A, N.G. Galanopoulos, "A Study of the Effect of Data Transformation and «Linearization» on Time Series Forecasts. A Practical Approach", June 2020.
281. Gibson, D. H., S.G. Hall, D. GeFang, P. Petroulas, G.S. Tavlas, "Did the Absence of A Central Bank Backstop in the Sovereign Bond Markets Exacerbate Spillovers During the Euro-Area Crisis?", July 2020.
282. Gibson, D. H., S.G. Hall, G.S. Tavlas, "A Suggestion for A Dynamic Multi Factor Model (DMFM)", July 2020.
283. Balfoussia, H., H. D. Gibson, D. Malliaropoulos, D. Papageorgiou, "The Economic Impact of Pandemics: Real and Financial Transmission Channels", September 2020.
284. Avgeri, I., Y. Dendramis, H. Louri, "The Single Supervisory Mechanism and its Implications for the Profitability of European Banks", October 2020.
285. Avramidis, P., I. Asimakopoulos, D. Malliaropoulos, "Bank consolidation, firm-bank relationship and firms' strategic default decision: evidence from the tourism industry during the Greek economic crisis", January 2021.
286. Papageorgiou, D. and S. Tsiaras, "The Greek Great Depression from a neoclassical perspective", February 2021.
287. Lazaris, P., A. Petropoulos, V. Siakoulis, E. Stavroulakis and N. Vlachogiannakis, "Interest Rate Pass Through in the Deposit and Loan Products Provided by Greek Banks", February 2021.
288. Kollintzas, T., D. Papageorgiou and V. Vassilatos, "Implications of Market and Political Power Interactions for Growth and the Business Cycle I: Private Sector Equilibrium", March 2021.
289. Lolos, S., P. Palaios and E. Papapetrou, "The Greek Tourism-Led Growth Revisited: Insights and Prospects", June 2021.
290. Kakridis, A., "Nobody's Child: The Bank of Greece in the Interwar Years", July 2021.
291. Hondroyiannis, G. and D. Papaikonomou, "The Effect of Eurosystem Asset Purchase Programmes on Euro Area Sovereign Bond Yields During The Covid-19 Pandemic", July 2021.
292. Bragoudakis, Z. and D. Panas, "Investigating Government Spending Multiplier for the US Economy: Empirical Evidence using a Triple Lasso Approach", October 2021.
293. Sideris, D. and G. Pavlou, "Disaggregate Income and Wealth Effects on Private Consumption in Greece", November 2021