

Multivariate Stochastic Volatility with Large and Moderate Shocks

Marwan Izzeldin[†], Mike G. Tsionas[†], and Panayotis G. Michaelides^{*§}

[†]Department of Economics, Lancaster University

[§]National Technical University of Athens, Greece & Systemic Risk Centre, London
School of Economics

January 16, 2019

Abstract

This paper proposes a multivariate stochastic volatility model where shifts in volatility are endogenously driven by large return shocks. The proposed model generalizes the univariate stochastic volatility model of [Dendramis et al. \(2015\)](#) to a multivariate context. Allowing for multivariate dependence permits the volatility of common return factors to jointly affect individual stock returns volatility. The model is further extended to allow for endogenous thresholds that depend on covariates. Model-selection priors are introduced and the new techniques are applied using data from the FTSE-100.

Keywords: Multivariate Stochastic Volatility; Bayesian analysis; Particle filtering; Large Shocks.

1 Introduction

Stochastic volatility modeling is becoming commonplace in applied econometrics due to the wide availability of Markov Chain Monte Carlo (MCMC) methods and the use of scalable and modular techniques provided by [Chib et al. \(2006\)](#). Although multivariate stochastic volatility (MSV) models are less widely used, they continue to be developed; see [Lopes and Carvalho \(2007\)](#) and [Chib et al. \(2009\)](#) and the references therein. [Dendramis et al. \(2015\)](#) proposed a univariate stochastic volatility model which allows for breaks in volatility that are driven by large stock return shocks and where the shocks are identified using threshold parameters. The model is important as it allows inferences in stochastic volatility models when market news arrive exogenously, thus requiring no prior assumptions on the part of the user. The role of large shocks on volatility has also been examined in [Kapetanios and Tzavalis \(2010\)](#) in relation to changes in parameters of linear models. Important studies in the MSV field include [Harvey et al. \(1994\)](#), [Aguilar and West \(2000\)](#), [Liesenfeld and Richard \(2003\)](#), [Gouriéroux et al. \(2009\)](#), [Chan et al. \(2005\)](#), [Chib et al. \(2006\)](#), [Philipov and Glickman \(2006\)](#) and [Asai and McAleer \(2009\)](#). See, [Asai et al.](#)

The authors are grateful to the editor, an associate editor and two referees for the very useful comments and suggestions which greatly helped to improve the paper.

* *Corresponding Author:* E-mail: pmichael@central.ntua.gr

(2006) and Yu and Meyer (2006a) for a review and Clark (2011) and Clark and Ravazzolo (2015) for applications of MSV to macroeconomics forecasting. Also see Loddo et al. (2011) where the George and McCulloch (1997) stochastic search variable selection approach is applied to MSV models.

The contribution of this paper is to extend the univariate model in Dendramis et al. (2015) to the multivariate case along the following dimensions. First, we consider a factor model for returns: the model is based on a factor structure which is quite popular in many financial applications such as the construction of optimal portfolios. The factors are dynamic and characterised by stochastic volatility and differential effects of large positive and negative shocks. Second, in our set-up, it is not only individual large shocks that matter for the dynamic factors, but the large shocks present in all the other factors. Given a large number of returns, a parsimonious model for their multivariate volatility is provided by the assumption that each return's volatility depends on its own past and that of the volatilities factors. To allow for this, we adopt model-selection priors to deal with the proliferation of parameters when the number of stocks is large. Third, we allow for endogeneity of large shocks so that they depend on certain covariates. We also allow factor volatility to have an impact on return volatility, so as to capture dynamics in the correlations. We also demonstrate that systematic model selection, using predictive Bayes factors, can be performed by comparing various special cases of the MSV model. Our empirical framework relies on Bayesian methods using the particle filtering MCMC methodology to perform the computations.

Our model can be viewed as i) an extension of Dendramis et al. (2015) to the multivariate case, and ii) an extension of multivariate stochastic volatility models to allow for endogenously driven structural breaks. We argue that both are important extensions in modelling multivariate financial time series. First, MSV models abstract from structural breaks and fail to distinguish between the impact of large and moderate shocks. For state of the art modelling breaks in MSV models see Chib et al. (2009). Second, univariate models as in Dendramis et al. (2015) cannot be used with multiple time series that are driven by common factors. Regarding the first point, we argue that it is important to extend MSV models to allow for structural breaks, with a distinction between large and moderate shocks as well as endogenous thresholds that are explained by certain covariates. As is true for Dendramis et al. (2015), shifts in volatility are stochastic which distinguish the model from Markov-chain and threshold volatility models with fixed magnitude breaks in volatility. Moreover, threshold volatility models assume a known value of the threshold parameter. Dendramis et al. (2015) allow for both negative and positive shocks to be estimated from the data. We follow their approach for the multivariate case whereby we allow for

both types of shocks. We also allow for the thresholds values to be estimated from the data.

The rest of this paper unfolds as follows. Section 2 outlines the model and its generalisation to the multivariate case. Sections 3 and 4 outline the priors and the posterior. Section 5 outlines the modified model specification which accommodates endogenous shocks with variable magnitudes. Sections 6 and 7 present both simulation and empirical evidence. Section 8 provides a discussion on the generalised impulse response functions. Section 9 concludes. Technical appendices that outline the particle filtering approach and the particle Metropolis adjusted Langevin filters are provided in technical appendices A and B while robustness analysis to demonstrate the insensitivity of posterior results is presented in Figures of Appendix A.

2 The model

Suppose we have a vector time series of returns $y_t = (y_{t1}, y_{t2}, \dots, y_{tp}), t = 1, \dots, T$. Following standard practice, e.g. [Chib et al. \(2006\)](#), we adopt a factor model of the form:

$$y_t = Bf_t + u_t, \quad (1)$$

where $f_t = (f_{t1}, \dots, f_{tk})'$ are unobserved factors, B is a $p \times k$ matrix of coefficients subject to the identification restrictions $b_{ij} = 0, i < j$ and $b_{ii} = 1$, and u_t is a vector of innovations. We assume:

$$\begin{bmatrix} u_t \\ f_t \end{bmatrix} \Big| V_t, D_t \sim \mathcal{N}_{p+k} \left(0, \begin{bmatrix} V_t & 0 \\ 0 & D_t \end{bmatrix} \right). \quad (2)$$

The elements of V_t and D_t are:

$$\begin{aligned} V_t &= \text{diag}\{\exp(h_{t1}), \dots, \exp(h_{tp})\}, \\ D_t &= \text{diag}\{\exp(h_{t,p+1}), \dots, \exp(h_{t,p+k})\}, \end{aligned} \quad (3)$$

Each element follows an autoregressive process:

$$h_{tj} = \mu_j + \phi_j(h_{t-1,j} - \mu_j) + \eta_{tj}, \quad \eta_{tj} \Big| \sigma_j \sim \text{i.i.d. } \mathcal{N}(0, \sigma_j^2), \quad j = 1, \dots, p+k. \quad (4)$$

To this point, this is a standard multivariate stochastic volatility (MSV) model, see [Kim et al. \(1998\)](#).

To extend [Dendramis et al. \(2015\)](#) we first assume that:

$$h_{tj} = \mu_{tj} + \phi_j(h_{t-1,j} - \mu_{tj}) + \eta_{tj}, \quad \eta_{tj} \Big| \sigma_j \sim \text{i.i.d. } \mathcal{N}(0, \sigma_j^2), \quad j = 1, \dots, p+k, \quad (5)$$

and

$$\mu_{tj} = \mu_{t-1,j} + I_{tj}\zeta_{tj}, \quad \zeta_{tj} \Big| \sigma_{\zeta_j} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta_j}^2), \quad j = 1, \dots, p+k, \quad (6)$$

where:

$$I_{tj} = \begin{cases} 1, & \text{if } \eta_{tj} > \eta^R \text{ or } \eta_{tj} < \eta^L, \\ 0, & \text{otherwise,} \end{cases} \quad j = 1, \dots, p+k. \quad (7)$$

Here η^R and η^L represent threshold values which determine whether we have a “large shock” in the volatility of returns and the factors. When shocks are moderate, we have $\mu_{tj} = \mu_{t-1,j}$, but when they exceed the thresholds we have $\mu_{tj} = \mu_{t-1,j} + \zeta_{tj}$. One feature of the model in [Dendramis et al. \(2015\)](#) is that no distinction is made between shocks that exceed η^R and shocks that are below η^L . Thus, we modify the shock process as follows:

$$\begin{aligned} \mu_{tj} &= \mu_{t-1,j} + I_{tj}^R \zeta_{tj}^R + I_{tj}^L \zeta_{tj}^L, \\ \zeta_{tj}^L | \sigma_{\zeta^L j}^2 &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^L j}^2), \zeta_{tj}^R | \sigma_{\zeta^R j}^2 \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^R j}^2), j = 1, \dots, p+k, \end{aligned} \quad (8)$$

where:

$$I_{tj}^R = \begin{cases} 1, & \text{if } \eta_{tj} > \eta^R, \\ \text{otherwise,} \end{cases} \quad j = 1, \dots, p+k, \quad (9)$$

$$I_{tj}^L = \begin{cases} 1, & \text{if } \eta_{tj} < \eta^L, \\ \text{otherwise,} \end{cases} \quad j = 1, \dots, p+k. \quad (10)$$

For shocks that exceed η^R we have: $\mu_{tj} = \mu_{t-1,j} + \zeta_{tj}^R$. For shocks that are below the threshold η^L we have: $\mu_{tj} = \mu_{t-1,j} + \zeta_{tj}^L$, where the innovations $\zeta_{tj}^L, \zeta_{tj}^R$ are allowed to have different variances, viz. $\sigma_{\zeta^L j}^2$ and $\sigma_{\zeta^R j}^2$. We impose the restriction:

$$\eta^R > \eta^L. \quad (11)$$

In this model, we have common thresholds η^R, η^L for all volatilities of returns and factors. Three modifications are possible, the third of which is substantive:

1. We can adopt different threshold values η_j^R, η_j^L , $j = 1, \dots, p+k$.
2. Returns and factor volatilities can fall into G groups which have common thresholds η_g^R, η_g^L , $g = 1, \dots, G$, $G < p+k$.
3. Treating volatilities of returns and their factors symmetrically is not of much use. Specifically, it makes sense to argue that it is the volatility of the common factors that affects the volatility of returns and their threshold values.

Note that $j = 1, \dots, p$ refers to return volatility and $j = p+1, \dots, p+k$ to factor volatility. This involves

a substantive modification of (4) in the following form:

$$\text{return volatility : } h_{tj} = \mu_{tj} + \phi_j(h_{t-1,j} - \mu_{tj}) + \sum_{i=1}^k \gamma_{ij}(h_{t,p+i} - \mu_{t,p+i}) + \eta_{tj}, \quad (12)$$

$$\begin{aligned} \eta_{tj} | \sigma_j &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_j^2), \quad j = 1, \dots, p. \\ \text{factor volatility : } h_{tj} &= \mu_{tj} + \phi_j(h_{t-1,j} - \mu_{tj}) + \eta_{tj}, \quad \eta_{tj} | \sigma_j \sim \text{i.i.d. } \mathcal{N}(0, \sigma_j^2), \quad j = p+1, \dots, p+k. \end{aligned} \quad (13)$$

In the above model, current period volatility of the factors (viz. $h_{t,p+i}$, $i = 1, \dots, k$) has a direct effect on the volatility of returns. In addition, new parameters γ_{ij} are introduced. The volatilities of factors follow autoregressions whose means, like the means of returns assuming they exist, follow the process in (8).

The significant characteristic of this model is the introduction of dependence in the volatility of returns.

Introducing dependence in multivariate stochastic volatility models or multivariate GARCH-type models has always proven problematic. One of the innovations of this paper is to show that if there are indeed a few factors that explain stock returns, then the volatility of these factors should be informative for conditional variances. Jungbacker and Koopman (2006) proposed a model with a single volatility factor, that is given a vector of returns y_t parametrized as $y_t|h_t$, where h_t follows a standard univariate stochastic volatility model. Their motivation was, for the most part, computational convenience, as they used importance sampling to implement maximum likelihood estimation. See also Quintana and West (1987). The models featured in their paper are based on those developed by Pitt and Shephard (1999a) who followed Kim et al. (1998). The models are based on $y_t|f_t, h_t$ where f_t is a vector of factors and h_t contains log volatilities of both returns and factors. Their model appears in equations (2) and (3) of our paper except that the elements of h_t follow independent stochastic volatility process instead of a full VAR specification. The novel feature here is the formulation of μ_i s as in (6), with the definitions of (7)-(10) which accommodate Dendramis et al. (2015). Moreover, the factors are dynamic. Our model in (12) additionally accommodates factor volatility to have an effect on returns. As mentioned in Creal and Tsay (2015), ‘‘A factor structure reduces the number of parameters to estimate and provides simple expressions for the inverse and determinant of relevant matrices. This makes computations feasible in high-dimensions’’. Therefore, although we begin with univariate stochastic volatility models, we modify this in (12) by introducing dependence through the introduction of factor volatilities in each univariate process. In addition, making the threshold values depend on factor volatility is a reasonable approach as making them function of volatilities of all returns would be somewhat excessive. The question of whether we need this, can be answered by looking at the large number of parameters implied by unrestricted multivariate stochastic volatility (SV) models. Although there are efficient MCMC procedures to access

the posterior distribution, this does not mean that a simplification (dependence of univariate process on volatilities of a number of factors) would not provide a substantial improvement in terms of statistical efficiency. This has been our main motivation as it is a good compromise between (i) fully unrestricted multivariate SV models, and (ii) univariate SV models.

The γ_{ij} s are weights (factor loadings) that determine how factor volatilities impact stock volatility. As in vector autoregressive models these parameters lack structural interpretation. However, they determine impulse response functions which can be easily interpreted. “Since γ_{21} can be different from zero, the volatility of the second asset is allowed to be Granger caused by the volatility of the first asset. Consequently, both the returns and volatilities are cross-dependent. However, the cross-dependence of volatilities is realized via Granger causality and volatility clustering jointly. Furthermore, when both γ_{12} and γ_{21} are nonzero, a bilateral Granger causality in volatility between the two assets is allowed” (Yu and Meyer, 2006b, page. 365). To generalize this model, we adopt the specification in (8) for the factor volatilities:

$$\begin{aligned} \text{factors : } \mu_{tj} &= \mu_{t-1,j} + I_{tj}^R \zeta_{tj}^R + I_{tj}^L \zeta_{tj}^L, \\ \zeta_{tj}^L | \sigma_{\zeta^L j}^2 &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^L j}^2), \quad \zeta_{tj}^R | \sigma_{\zeta^R j}^2 \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^R j}^2), \quad j = p+1, \dots, p+k. \end{aligned} \quad (14)$$

As the number of factors is not expected to be large, we allow for factor-specific thresholds, η_j^L and η_j^R :

$$I_{tj}^R = \begin{cases} 1, & \text{if } \eta_{tj} > \eta_j^R, \\ \text{otherwise,} & \end{cases} \quad j = p+1, \dots, p+k, \quad (15)$$

$$I_{tj}^L = \begin{cases} 1, & \text{if } \eta_{tj} < \eta_j^L, \\ \text{otherwise,} & \end{cases} \quad j = p+1, \dots, p+k. \quad (16)$$

For the volatility of returns, the model can be generalized so that we have idiosyncratic and factor-specific shocks. Conditional on $\zeta_{ti}^R, \zeta_{ti}^L, i = p+1, \dots, p+k$ we have:

$$\begin{aligned} \text{returns : } \mu_{tj} &= \mu_{t-1,j} + I_{tj}^R \zeta_{tj}^R + I_{tj}^L \zeta_{tj}^L + \sum_{i=1}^k \delta_{ij}^R I_{t-1,p+i}^R \zeta_{ti}^R + \sum_{i=1}^k \delta_{ij}^L I_{t-1,p+i}^L \zeta_{ti}^L, \\ \zeta_{tj}^L | \sigma_{\zeta^L j}^2 &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^L j}^2), \quad \zeta_{tj}^R | \sigma_{\zeta^R j}^2 \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\zeta^R j}^2), \quad j = 1, \dots, p. \end{aligned} \quad (17)$$

In this model, factor volatility shocks that breach either the upper threshold or the lower threshold have a direct effect on μ_{tj} for all returns ($j = 1, \dots, p$) through the parameters δ_{ij}^R and δ_{ij}^L . The model requires the introduction of additional $2pk$ parameters. We also examine a model where return-specific thresholds $\eta_j^L, \eta_j^R, j = 1, \dots, p$ are allowed.

3 Priors

As the model involves a large number of parameters, we must carefully design our priors. For example, given that factor-specific large shocks are allowed in (17) it is unlikely that we may need return-specific

large shocks. Of course, this is an empirical matter, but the implication of adopting it is that $\eta_{tj}^L = \eta_{tj}^R, \eta_{tj}^R = \eta_{tj}^L, \quad j = 1, \dots, p$. The following accommodates this belief:

$$\begin{aligned} \eta_{tj}^L | \omega_\eta, \eta^L &\sim \mathcal{N}(\eta^L, \omega_\eta^2 s^2), \quad j = 1, \dots, p. \\ \eta_{tj}^R | \omega_\eta, \eta^R &\sim \mathcal{N}(\eta^R, \omega_\eta^2 s^2), \quad j = 1, \dots, p. \end{aligned} \quad (18)$$

where the parameter $\omega_\eta = 1$, s denotes the sample median of the standard deviation of returns, and we impose the restriction in (11). Parameters ϕ_j and σ_j^2 in (12) are likely to be similar for returns ($j = 1, \dots, p$). Parameters δ_{ij}^L and δ_{ij}^R for returns ($j = 1, \dots, p, \quad i = 1, \dots, k$) are potentially too many when there are large number of returns. Denote these parameters generically by δ_{ij} . We adopt a model-selection prior, as in George and McCulloch (1993, 1997); see also Ntzoufras et al. (2000) of the form:

$$\begin{aligned} \delta_{ij} &= 0, \text{ with probability } P. \\ \delta_{ij} | \omega_\delta, P &\sim \mathcal{N}(0, \omega_\delta^2), \text{ with probability } 1 - P, \end{aligned} \quad (19)$$

where P is a Bernoulli random variable with parameter $\mathcal{P} = \frac{1}{2}$. This setting is justified given our prior ignorance about this parameter and its likely values.

For parameters γ_{ij} in (12) we adopt exactly the same prior. For parameters ϕ_j we assume:

$$\phi_j | \bar{\phi}, \omega_\phi \sim \mathcal{N}(\bar{\phi}, \omega_\phi^2), \quad 0 \leq \phi_j < 1, \quad j = 1, \dots, p + k. \quad (20)$$

where $\bar{\phi} = 0.9$. That value is motivated from previous studies on stochastic volatility models and the fact that we expect substantial persistence on prior grounds. The parameter ω_ϕ^2 controls prior concentration about this central value. For scale parameters, σ_j^2 , we assume:

$$\frac{\bar{Q}}{\sigma_j^2} \sim \chi^2(\bar{n}), \quad j = 1, \dots, p + k. \quad (21)$$

where $\bar{n} = 1$ and $\bar{Q} = 0.001$, which is close to being ‘‘flat’’ but proper. We set $\omega_\phi = \omega_\delta = 1$. This choice is motivated by the need for informative but reasonably flat priors. As ϕ_j 's are expected to be between (0 and 1) and the choice $\omega_\phi = \omega_\delta = 1$, implies priors that are diffuse relative to the likelihood / posterior. With $\bar{\phi} = 0.9$ the ϕ_j 's are expected to be between -1.1 and 2.9 with prior probability 95%. This choice allows for priors that are informative and flexible. Finally for the different, non-zero elements of B which are collected into the vector β we assume $\beta | \omega_\beta \sim N(0, \omega_\beta^2 I)$ where $\omega_\beta = 10$.

4 Posterior

In this section, we present the posterior of the model under the most general assumptions, viz. we have return-specific thresholds $\eta_j^L, \eta_j^R, j = 1, \dots, p$ and, of course, factor-specific thresholds $\eta_j^L, \eta_j^R, j =$

$p + 1, \dots, p + k$. We denote all unknown “structural” or “deep” parameters of the model by $\theta \in \Theta \subseteq \mathbb{R}^D$, where D is the dimensionality of the parameter space. This parameter vector includes β , i.e. the non-zero elements of B , σ_j^2 , $\sigma_{\zeta_j^L}^2$, $\sigma_{\zeta_j^R}^2$, δ_{ij}^L , δ_{ij}^R , and the thresholds η_j^L and η_j^R . In order to facilitate Sequential Monte Carlo as μ_{tj} s are stochastic and time-varying, we change (14) and (17) as follows:

$$\text{factors : } \mu_{tj} = \mu_{t-1,j} + I_{tj}^R \zeta_{tj}^R + I_{tj}^L \zeta_{tj}^L + \xi_{tj}, \quad j = p + 1, \dots, p + k. \quad (22)$$

$$\text{returns : } \mu_{tj} = \mu_{t-1,j} + I_{tj}^R \zeta_{tj}^R + I_{tj}^L \zeta_{tj}^L + \sum_{i=1}^k \delta_{ij}^R I_{t-1,p+i}^R \zeta_{ti}^R + \sum_{i=1}^k \delta_{ij}^L I_{t-1,p+i}^L \zeta_{ti}^L + \xi_{tj}, \quad j = 1, \dots, p, \quad (23)$$

where $\xi_{tj} | \bar{\sigma}_\xi \sim \mathcal{N}(0, \bar{\sigma}_\xi^2)$, $j = 1, \dots, p + k$, and $\bar{\sigma}_\xi^2$ is a constant set equal to 10^{-5} . This constant value is common to all factors and returns, and facilitates writing the posterior in standard form. To proceed, let $Y = [y_1, \dots, y_T]$ denote the data and $\Sigma_t = \begin{bmatrix} V_t & 0 \\ 0 & D_t \end{bmatrix}$ where these elements are defined in (3). Under these assumptions, the posterior distribution of the model can be expressed in the following form:

$$\begin{aligned} & p(\theta, \mathbf{h}_t, \boldsymbol{\mu}_t, \boldsymbol{\zeta}^L, \boldsymbol{\zeta}^R, \mathbf{I}_t^R, \mathbf{I}_t^L | Y) \propto p(\theta) \cdot \\ & \prod_{t=1}^T |\Sigma_t|^{-1/2} \exp \left\{ -\frac{1}{2} (y_t - Bf_t)' \Sigma_t^{-1} (y_t - Bf_t) \right\} \cdot \\ & \prod_{j=1}^{p+k} \sigma_j^{-T/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \sum_{j=p+1}^{p+k} \frac{1}{\sigma_j^2} \left(h_{tj} - \mu_{tj} - \phi_j(h_{t-1,j} - \mu_{t-1,j}) - \sum_{j'=1}^k \gamma_{ij'} (h_{t,p+i} - \mu_{t,p+i}) \right)^2 \right\} \cdot \\ & \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \sum_{j=1}^p \frac{1}{\sigma_j^2} (h_{tj} - \mu_{tj} - \phi_j(h_{t-1,j} - \mu_{t-1,j}))^2 \right\} \cdot \prod_{t=1}^T \prod_{j=1}^{p+k} h_{tj}^{-1} \cdot \\ & \prod_{j=p+1}^{p+k} \prod_{\{t: I_{tj}^L=1\}} \Phi \left(\frac{\eta_j^L - h_{tj} - \mu_{tj} - \phi_j(h_{t-1,j} - \mu_{t-1,j}) - \sum_{j'=1}^k \gamma_{ij'} (h_{t,p+i} - \mu_{t,p+i})}{\sigma_j} \right) \cdot \\ & \prod_{j=1}^p \prod_{\{t: I_{tj}^R=1\}} \left\{ 1 - \Phi \left(\frac{\eta_j^R - h_{tj} - \mu_{tj} - \phi_j(h_{t-1,j} - \mu_{t-1,j})}{\sigma_j} \right) \right\} \cdot \\ & \prod_{j=1}^{p+k} \left[\sigma_{\zeta_j^L}^{-T/2} \sigma_{\zeta_j^R}^{-T/2} \cdot \exp \left\{ -\frac{1}{2\sigma_{\zeta_j^L}^2} \sum_{t=1}^T (\zeta_{tj}^L)^2 - \frac{1}{2\sigma_{\zeta_j^R}^2} \sum_{t=1}^T (\zeta_{tj}^R)^2 \right\} \right] \cdot \\ & \bar{\sigma}_\xi^{-Tk/2} \exp \left\{ -\frac{1}{2\bar{\sigma}_\xi^2} \sum_{t=1}^T \sum_{j=p+1}^{p+k} (\mu_{tj} - \mu_{t-1,j} - I_{tj}^R \zeta_{tj}^R - I_{tj}^L \zeta_{tj}^L)^2 \right\} \cdot \\ & \bar{\sigma}_\xi^{-Tp/2} \exp \left\{ -\frac{1}{2\bar{\sigma}_\xi^2} \sum_{t=1}^T \sum_{j=1}^p (\mu_{tj} - \mu_{t-1,j} - I_{tj}^R \zeta_{tj}^R - I_{tj}^L \zeta_{tj}^L - \sum_{j'=1}^k \delta_{ij'}^R I_{t-1,j'}^R \zeta_{t,j'}^R - \sum_{j'=1}^k \delta_{ij'}^L I_{t-1,j'}^L \zeta_{t,j'}^L)^2 \right\} \cdot \end{aligned} \quad (24)$$

In (24) $\mathbf{h}_t, \boldsymbol{\mu}_t, \boldsymbol{\zeta}^L, \boldsymbol{\zeta}^R, \mathbf{I}_t^R, \mathbf{I}_t^L$ each denote a vector that respectively contains $\{h_{tj}\}, \{\mu_{tj}\}, \{\zeta_j^L\}, \{\zeta_j^R\}, \{I_{tj}^R, I_{tj}^L\}$; and where $p(\theta)$ is the prior of the “structural” or “deep” parameters of the model as defined in the previous section. In the second line, we have the likelihood contribution of the factor model. In the third and fourth lines we have the likelihood contributions of the stochastic volatility terms. In the fifth and sixth lines we have the probabilities of error terms η_{tj} either exceeding R_j or being below L_j . In the seventh line we have the likelihood contributions of errors in ζ_{tj} that are associated with mean shifts, μ_{tj} . In the final two lines, we have the likelihood contributions from mean shifts in (22) and (23). The indicator functions I_{tj}^R and I_{tj}^L are defined in (15) and (16).

Our MCMC procedure works through the following steps:

1. Integrate out μ_t and h_t using the Auxiliary Disturbance Particle Filter (ADPF) of Hall et al. (2014),

see Technical Appendix, part B¹.

2. Draw I_t^L and I_t^R from their respective posterior conditional distributions.
3. Draw θ using the procedure for “structural parameters” following (B.5) in the Technical Appendix, part B.

The particle filter in step 1, draws $\{\mu_{tj}, t = 1, \dots, T\}$ and $\{h_{tj}, t = 1, \dots, T\}$ jointly and is applied separately for each $j = 1, \dots, p + k$. This is not a problem in high-end computing clusters where all $p + k$ latent variables (of each type) can be allocated to different cores, and where the final results can be considered jointly to form (24)².

5 A New Model: Endogenous magnitude of shocks

In the absence of large shocks, it is possible for the MSV model to produce large volatilities which, given the considerable uncertainty, cannot be parameterized without allowing for structural breaks. The question we address in this section is as follows: ‘Is it possible to improve the model by allowing for an endogenous explanation of thresholds?’ If the answer is positive, the model is capable of forecasting to some extent, whether a change in the structure is forthcoming.

The idea of endogenous thresholds is, as indicated below, equivalent to the notion that the probability of exceeding the lower or upper threshold depends on a vector of variables, z_t . In our case this means that information about factors and volatilities may be useful for predicting large absolute returns. To cater for this, we endogenize extreme movements in returns, otherwise market news must be treated exogenously. A similar approach has been used by Peracchi and Rossetti (2013) and Taylor and Yu (2016). An alternative interpretation is provided in the discussion following equation (26). The new model presented in this section is as follows.

Consider (16) and (15) and define the latent variables:

$$I_t^L = \begin{cases} 1, & L_t^* \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad I_t^R = \begin{cases} 1, & R_t^* \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

¹For robustness, we have also used the Particle Metropolis Adjusted Langevin Filter, see Nemeth and Fearnhead (2014), Technical Appendix, Part C. Using the same number of particles per iteration we obtained almost identical results.

²Computations were performed at the High End Computing Cluster (HEC) of Lancaster University. The combined facility offers over 5,000 CPU cores, 23 TB aggregate of memory, 70TB of high performance filestore and 1.5PB of medium performance filestore. A number of nodes offer Nvidia GPU cards, which support CUDA and OpenCL applications. The cluster operating system is Scientific Linux, with job submission handled by Son of Grid Engine (SoGE). All computations were performed using `fortran77` with extensive use of subroutines from the IMSL library and additional software from `netlib`. In our dataset, computation took, approximately, 22 minutes of CPU time.

We parameterize the latent variables as follows:

$$\begin{aligned} L_t^* &= z_t' \psi_L - \varepsilon_{Lt}, \quad \varepsilon_{Lt} \sim \mathcal{N}(0, 1), \\ R_t^* &= z_t' \psi_R - \varepsilon_{Rt}, \quad \varepsilon_{Rt} \sim \mathcal{N}(0, 1), \end{aligned} \quad (26)$$

where $z_t' = [f_t, f_{t-1}, h_{t-1}, h_{t-1}^f]$, $h_t = p^{-1} \sum_{j=1}^p h_{tj}$ is average volatility across all p stock returns, and h_t^f is the average factor volatility ($h_t^f := k^{-1} \sum_{i=1}^k h_{t,p+i}$). For the new parameters, we assume a prior: $[\psi_L', \psi_R']' | \omega_\psi \sim \mathcal{N}(0, \omega_\psi^2 I)$, and we set $\omega_\psi = 10$. The probability that the lower and upper threshold are exceeded is $\Phi(z_t' \psi_L)$ and $\Phi(z_t' \psi_R)$, respectively.

As the use of particle filter MCMC allows considerable flexibility, the model can be estimated using the techniques outlined earlier. Our interest focuses on the improvement afforded in terms of marginal likelihood or the recursive (predictive) Bayes factor. The evidence is provided in Figure 1. Although the Bayes Factor³ is an established tool for model comparison and hypothesis testing (Geweke (2007)), the approach suffers from a number of drawbacks. Not only it is sensitive to the priors (Kass and Raftery (1995)), the Jeffreys-Lindley's paradox (Poirier (1995); Robert (2001)) is also a concern, especially when we test a point null against a general alternative. In our case, the use of the Bayes Factor is justified by its application; i.e. to quantify the support for one model over another, rather than for point null hypothesis testing.

6 Simulation evidence

In this section simulation evidence is presented for the model with endogenous thresholds, as introduced in the previous section. With **Simulation A**, our objective is to examine the effectiveness of the particle MCMC Metropolis-Hastings algorithm (PMCMC) approach. This leads to discussion: (i) on the convergence of the PMCMC approach and (ii) on the choice of the number of particles and iterations. Simulation A adopts previously stated prior specifications, which are relatively flat in comparison with the likelihood. This implies that the data dominate the priors. **Simulation B** is a further exercise to show the effect of the prior on the inference results. Simulation B is conditional on the choice of the number of particles and iterations, where we examine different priors in a systematic way. In both Simulations A and B we specify the data generating process by setting the parameters at certain values. The specification is as follows:

³The Bayes Factor approach requires the evaluation of the marginal likelihood of the problem, which often proves to be a quite difficult problem to solve in itself and is not always easy to compute (Li et al. (2015)). Indeed, for an excellent presentation of a number of alternative strategies and the development of a new test statistic, which could be seen, asymptotically, as the Bayesian alternative to the Lagrange multiplier test, see Li et al. (2015) and the references therein.

i) We set $p = 100$ for the number of time series and $k = 1$ for the number of factors. We specify different values of T as 500, 1,000, 2,000 and 4,000.

ii) For the different non-zero elements of matrix B in (1) we assume that they are generated independently from a uniform distribution in the interval (0.1, 0.8).

iii) We then specify Parameters in (12)-(17). In relation to (12): parameters $\gamma_{ij}, i = 1, \dots, k, j = 1, \dots, p$ are drawn randomly and independently from a $\mathcal{N}(0.1, 1)$ distribution; parameters $\sigma_j, j = 1, \dots, p+k$ are drawn independently from a uniform distribution in the interval (0.1, 0.5); parameters $\phi_j, j = p+1, \dots, p+k$ are drawn independently from a uniform distribution in the interval (0.2, 0.9).

iv) In (14) parameters $\sigma_{\zeta^L j}^2$ and $\sigma_{\zeta^R j}^2$ ($j = 1, \dots, k$) are set to 0.1.

v) In (15) and (16) parameters η_j^L and η_j^R ($j = p+1, \dots, p+k$) are generated according to the endogenous threshold model (25) and (26), see step (vii).

vi) Parameters δ_{ij}^L and δ_{ij}^R from (17) are generated independently from a $\mathcal{N}(0, 0.1^2)$ distribution.

vii) Elements of parameter vectors ψ_L and ψ_R from (25) and (26) are generated independently from a $\mathcal{N}(0, 0.3^2)$ distribution.

Our simulation results are presented in Table 1. Our primary interest is with the mean-squared-error (MSE) of the parameters. As parameter estimates, we take the posterior means of draws from PMCMC. The reason we consider the MSE is to examine whether it is reduced by approximately half when the sample size is doubled (which is the reason we selected the particular time horizons T). As p and k are large, we report the *maximum value* of MSE as a conservative (worst case) scenario for the behaviour of parameters. The numbers reported in Table 1 show no difference in results between $NP = 10^6$ and $NP = 10^7$. As such we can safely use $NP = 10^6$ and $NI = 40,000$, where NI is the number of PMCMC iterations.

For Simulation B in Table 2, we change the baseline prior in a systematic way to examine sensitivity to prior assumptions. Using $NP = 10^6$, we examine sensitivity of posterior means when the scale parameters ω are changed. Specifically, the baseline values are scaled by $\frac{1}{4}, \frac{1}{2}, 2$ and 4 . We run MCMC for each new set of hyperparameters, recording the new posterior means and posterior standard deviations. We then compare these with the posterior means in the baseline model. We report only the maximum percentage change in posterior means and posterior standard deviations (where full results are available on request). Table 2 presents a concise summary of the results, where the maximum absolute percentage changes in posterior means and posterior standard deviations are, respectively, 0.53% and 0.33%. Where there are

differences, these are not substantial, so implying that posterior results from the baseline specification are robust to changes in the prior.

7 Data and Results

We consider all daily returns of the FTSE100 index⁴ from the Bloomberg's database for the period 01/03/2005 to 06/26/2015. First, we report predictive Bayes factors corresponding to our various models for different time periods. The predictive Bayes factors are computed from the predictive likelihoods of the particle filtering procedure⁵. The results are reported in Table 2. For simplicity in presentation, we maintain $k = 1$ factors throughout. As the evidence for the existence of a larger number of factors is quite weak, this assumption turns out to be inconsequential in the light of the data.

Regarding Predictive Bayes factors we have proceeded as follows. Conditional on the data y_t , the predictive distribution is

$$p(y_{t+1}|y_t) = \int_{\Theta} p(y_{t+1}|y_{1:t}, \theta) p(\theta|y_{1:t}) d\theta. \quad (27)$$

We use the notation $y_{1:t}$ to denote data available up to and including period t . Moreover, the parameter space is denoted by Θ . Of course, access to the posterior predictive is precluded as the integral can be high-dimensional and unavailable in closed form (θ also includes latent variables but we omit the time index for simplicity.) If an ergodic sample from the posterior $p(\theta|y_{1:t})$ is available, and we denote it by $\{\theta^{(s)}, s = 1, \dots, S\}$, then it is clear that we can approximate the posterior predictive in (27) as follows: Given a draw $\theta^{(s)}$ from its posterior ($p(\theta|y_{1:t})$), a draw from the posterior predictive is: $y_{t+1}^{(s)}|y_{1:t}, \theta^{(s)} \sim y_{t+1}|y_{1:t}, \theta^{(s)}, s = 1, \dots, S$. This distribution conditions on all parameters and latent variables and in many models, including ours, it is amenable to straightforward simulation, see Geweke and Amisano (2010). The posterior predictive distribution takes parameter uncertainty into account and, of course, it also accounts for uncertainty with respect to the unobserved value y_{t+1} . Therefore, the approximation is:

$$p(y_{t+1}|y_{1:t}) \cong S^{-1} \sum_{s=1}^S p(y_{t+1}|y_{1:t}, \theta^{(s)}). \quad (28)$$

For the entire data set, $y_{1:T}$, the marginal likelihood or "evidence" is:

$$p(y_{1:T}) = \int L(\theta; y_{1:T}) p(\theta) d\theta, \quad (29)$$

⁴For a description please see [London Stock Exchange](#).

⁵For alternatives see [Omori et al. \(2007\)](#).

viz. the normalizing constant of the posterior distribution, where $L(\theta; y_{1:T})$ is the likelihood function and $p(\theta)$ is the prior. One can define a marginal likelihood using data only up to period t , which we denote by $p(y_{t+1}|y_{1:t})$.

So far our discussion has considered a single model. Suppose however we have models $\mathcal{M}_1, \dots, \mathcal{M}_M$ and the posterior predictive for each model is denoted by $p(y_{t+1}|y_{1:t}, \mathcal{M}_m), m = 1, \dots, M$. In (27) the relevant posterior of parameters is $p(\theta|y_{1:t}, \mathcal{M}_m), m = 1, \dots, M$, and the posterior predictive is denoted by:

$$p(y_{t+1}|y_t, \mathcal{M}_m) = \int_{\Theta_m} p(y_{t+1}|y_{1:t}, \theta_m, \mathcal{M}_m)p(\theta|y_{1:t}, \mathcal{M}_m)d\theta_m, m = 1, \dots, M. \quad (30)$$

The same changes have to be made in (28). The different models are allowed to have different parameters and, therefore different parameter spaces. Evidently, we also have:

$$p(y_{1:T}|\mathcal{M}_m) = \int_{\Theta_m} L(\theta_m; y_{1:T}, \mathcal{M}_m)p(\theta_m|\mathcal{M}_m)d\theta_m, m = 1, \dots, M, \quad (31)$$

where $p(\theta_m|\mathcal{M}_m)$ is the prior for the parameters of model $\mathcal{M}_m, m \in \{1, \dots, M\}$. The expression in (31) can be accurately approximated using:

$$p(y_{1:t+1}|\mathcal{M}_m) = S^{-1} \sum_{s=1}^S p(y_{t+1}|y_{1:t}, \theta^{(s)}, \mathcal{M}_m), m = 1, \dots, M. \quad (32)$$

However, in this expression, $\theta^{(s)}$ is drawn from the posterior of θ conditional on $y_{1:t}$ so multiple MCMC runs must be executed. In general, however, we are not interested in the computation of posterior predictive densities for the whole sample but for only a few in-sample observations and/or a few observations in a hold-out sample. The predictive Bayes factor for two competing models, using data available only up to date t , is defined as follows:

$$PBF_{t+1}^{m:m'} = \frac{p(y_{t+1}|y_{1:t}, \mathcal{M}_m)}{p(y_{t+1}|y_{1:t}, \mathcal{M}_{m'})}, m \neq m' \in \{1, \dots, M\}. \quad (33)$$

Since marginal likelihood is available from the standard output of particle filtering, accurate approximations to (33) can be computed. Extensions to multi-step forecasting and predictive distributions are straightforward (Geweke and Amisano (2010), equations (10)-(12)). Finally, it is sufficient to compare predictive Bayes factors relative to a benchmark, as we can easily show that:

$$PBF_{t+1}^{m:m'} = \frac{PBF_{t+1}^{m:m'}}{PBF_{t+1}^{m:m''}} \cdot \frac{PBF_{t+1}^{m:m''}}{PBF_{t+1}^{m:m'}}, m \neq m' \neq m'' \in \{1, \dots, M\}. \quad (34)$$

All our computations are based on 120,000 MCMC iterations, the first 20,000 of which are discarded to mitigate possible start-up effects. The particle filter is implemented using 10^6 particles. The algorithm

we use is the PMCMC procedure which filters out the unobserved volatilities and then exploits MCMC to draw from the conditional posterior distribution of parameters like ϕ_j , γ_{ij} , δ_{ij} , σ_j^2 . See Technical Appendix A for further details. Convergence is tested (successfully) using Geweke (1992) diagnostics. In particular, we use the Hall et al. (2014) filter described in (A.6)-(A.7) of Appendix A along with a Metropolis Adjusted Langevin Approach procedure to validate the results. We summarize our models in Table 3.

In Figures 2a-2b, we report the marginal posterior densities⁶ of ϕ_j , γ_{ij} , δ_{ij}^R . Marginal posterior densities of δ_{ij}^R and δ_{ij}^L are reported in panels (a) and (b) of Figure 3. To conserve space we report marginal posterior densities for $i = 1$ and $j = 1, \dots, 20$ in the different sub-panels. The stocks were randomly selected as there are 10,000 parameters in $[\delta_{ij}^L]$ and $[\delta_{ij}^R]$. For other stocks the results were qualitatively and quantitatively the same. As expected, persistence parameters (ϕ_j) are close to unity. Due to model-selection priors, the marginal posterior densities of γ_{ij} , δ_{ij}^R and δ_{ij}^L are bimodal (or highly asymmetric) with a mode around zero. Evidently, asymptotic Gaussian inference would be misleading in these cases.

In the left panel of Figure 3a, we report marginal posterior densities of δ_{ij}^R for randomly selected 20 stocks. In the right panel we report marginal posterior densities of δ_{ij}^L . Evidently, most of them are randomly selected bimodal. Moreover, with high potential probability these parameters assume reasonable values for posterior densities of δ_{ij}^L and δ_{ij}^R . For the remaining 80 stocks see, the figures Appendix.

Next, we report the results for average stochastic volatility (ASV) and average correlation coefficients (ACC). These are computed using non-overlapping windows of size $N = 20$. Filtered volatility is computed in a similar fashion for each MCMC draw, and is then averaged. The results are reported in Figure 4. Note that the sub-prime crisis period is from 01/12/2007 to 30/06/2009.

Clearly, the MSV model with large shocks provides a superior in-sample fit for volatility (Figure 4) as it tracks better the 20-day-window sample volatility. Not only is this direct evidence in favour of the new model, it is impressive how well it tracks in-sample volatility during the sub-prime crisis (observations 40 to 60 roughly, as in Figure 4). The MSV model without large shocks shows a jump in volatility with a significant lag of about 20 days; and it features artificial spikes after the sub-prime crisis which contradicts the pattern of in-sample volatility.

⁶Detailed posterior statistics for all model parameters are available as a separate Appendix on request.

As in Table 3, the Bayes factors (BFs) for the MSV model with large shocks and unlike those of the MSV model without large shocks, favor the first model; especially so during the crisis period, see Figure 1. Bayes factors favoring endogeneity are reported in Figure 5 where we observe considerable improvement, particularly during the turmoil that was generated from the sub-prime crisis and overall. This shows that endogeneity improves substantially in the MSV model with exogenous large shocks. Moreover, the BFs are higher for the model with endogenous shocks during the sub-prime crisis. Predictive BFs against an MSV model without large shocks reported in Figure 1 favor the model which allows for large shocks. Specifically, from Figure 5, the predictive BFs favors the new model by a factor of 10 to 40, particularly during the sub-prime crisis. From Figure 5 endogeneity is favored by a factor of 40 to 50 on average. It is noteworthy that predictive BFs against the simple MSV model follow the general pattern of in-sample volatility after scaling as in Figure 1. The BFs are typically 10 times larger during periods of tranquillity, but 40 times higher during the turmoil. The higher predictive BFs during the crisis indicate the superiority of the proposed model. This implies that new information has an impact on volatility kept arriving post 2008 and that the proposed model is capable of capturing such effects ⁷.

Table 4 presents the sensitivity analysis of the BFs with respect to the prior. We examine 10,000 different priors: Given the baseline prior, parameters are multiplied by constants $c_j \in (10^{-5}, 10^5)$ which are generated, randomly from independent uniform distributions. Posterior analysis using PMCMC is conducted for each of the priors and the BFs are computed using the marginal likelihood. The results show, again, that the BFs favor the MSV with large shocks. The results also show that the BFs in favour of endogeneity are quite robust in terms of the prior.

In Figure 6 we report the marginal posterior densities of the threshold parameters η^L and η^R . In the MSV model with common thresholds, we report their marginal density. For the MSV model with different thresholds, η_j^L and η_j^R , we report the marginal posterior density of the average across all FTSE100 stocks, generated from each iteration. The two MSV models differ substantially in terms of posterior estimates of the thresholds. This is a consequence of the different assumptions used (i.e. whether the thresholds are common or different among shocks in the MSV models). Posterior densities of the thresholds for the

⁷Such pieces of information include, but are not limited, to announcements of the ECB and Euro Working Group, following the economic developments in southern Europe.

factor are also substantively different as illustrated in Figure 7.

In Table 5 we report BFs corresponding to different number of factors ($n = 1, \dots, 10$). The BFs strength favour a model with $k = 5$ factors (BFs being 12.75, 14.61, 16.71) relative to a model with $n = 1$ factor. Remarkably, Bayes factors increase over the period 12/05/2008 up to 22/10/2009. These increases can be attributed, naturally, to the sub-prime crisis and the increase of turbulence and volatility in the markets. Clearly, BFs always favour model V against all other models.

Finally, we report posterior statistics for certain parameters of the model in Table 7⁸. A notable feature of the results is that inefficiency factors (IF, defined as $IF = 1 + \sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s of MCMC draws for a given parameter) are generally lower than those reported in Dendramis et al. (2015).

8 Generalized impulse response functions

Part of the attraction of the newly modified model is that allows the effect of moderate and large shocks to be analysed through the Generalized Impulse Response Function (GIRF) on volatility, see Koop et al. (1996) and Pesaran and Shin (1998). The GIRF is defined as:

$$GIRF_j(N, u_t, s_t) = E(h_{j,t+N}|u_t, s_t) - E(h_{j,t}|s_t), \quad (35)$$

where s_t is the state vector. In the computation of this function, there are a number of differences relative to Dendramis et al. (2015). First, computing $E(h_{j,t+N}|u_t, s_t)$ is a natural by-product of the particle filtering algorithm as we average across MCMC draws. Second, computing the term $E(h_{j,t}|s_t)$ requires some care as we cannot draw u_t from its prior in (2) and average $E(h_{j,t+N}|u_t, s_t)$ over the draws of u_t . In fact, we have to draw from the conditional distribution $u_t|s_t$ which is revealed if we follow the auxiliary particle filtering procedure of Hall et al. (2014) described in the Appendix. This approach is particularly convenient as the states are available in analytic form. Third, as we cannot fix the threshold values at their posterior means, we have to take into account this uncertainty. The same is true for all static parameters involved in the computation of the term $E(h_{j,t}|s_t)$. Our results are presented in Figure 8 for large positive shocks, and in Figure 9 for large negative shocks.

From the posterior means of return thresholds in Figure 6, we notice that (when we allow for different thresholds) these are, approximately, (-0.06 and 0.07) or (-6% and 7%). For the factors they are, approxi-

⁸Posterior statistics for all other parameters are available upon request as a separated Appendix.

mately (-7% and 8%). Large negative and large positive shocks have quite different effects in terms of the GIRFs reported in Figures 8 and 9. Both are quantitatively and statistically important up to an horizon of $N = 60$ days. For horizons of $N = 1$ to $N = 10$ days the negative shock shows significantly more persistence which is numerically important even after $N = 40$ or $N = 60$ days. The effect of positive shocks (Figure 9) is practically zero after $N = 60$ days. In addition, the posterior densities of large positive and large negative shocks are quite different in shape, suggesting that important asymmetric effects are present in returns.

Regarding the GIRFs of the ordinary positive and negative shocks (Figures 10 and 11), these are heavily concentrated around zero in terms of their posterior densities. So, it appears unlikely for ordinary shocks to have important effects on the volatility of stock returns. Furthermore, the fact that the effects are not the same provides additional motivation for the application of our model i.e., one should, indeed allow for a differential effect of moderate shocks versus large positive and large negative shock (the later being differential as well). Indeed, the positive impact of large negative shocks ‘bad news’ on volatility gives confirmation to the leverage hypothesis. Although that same conclusion is reached in [Dendramis et al. \(2015\)](#), the advantage of our method is in at least two areas: (i) It forecasts volatility in multivariate stock return time series in a more precise way; (ii) It allows for a more effective construction of portfolios using multivariate data.

9 Conclusion

In this paper, we propose a multivariate stochastic volatility model (MSV) that generalizes the univariate stochastic volatility model of [Dendramis et al. \(2015\)](#), where the emphasis is placed on the impact of large shocks or structural breaks in volatility. In MSV models this mechanism is absent. Univariate models cannot be used where multiple time series are driven by common factors. MSV models that abstract from structural breaks and fail to distinguish between large and moderate shocks are inadequate in representing the data. In a generalized model where we allow for endogenous shocks, we find that Bayes factors for the MSV model with large shocks provide strong evidence that the new model is superior to MSV models that do not allow for large and moderate shocks. This is especially evident during the sub-prime crisis period.

References

- Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.
- Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, pages 697–725.
- Asai, M. and McAleer, M. (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, 150(2):182–192.
- Asai, M., McAleer, M., and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2-3):145–175.
- Cappé, O., Moulines, E., and Rydén, T. (2005). Inference in Hidden Markov Models. *Springer Series in Statistics, Berlin*, pages 565–601.
- Casarin, R., Marin, J.-M., et al. (2009). Online data processing: comparison of bayesian regularized particle filters. *Electronic Journal of Statistics*, 3:239–258.
- Chan, D., Kohn, R., and Kirby, C. (2005). Multivariate stochastic volatility with leverage. *Econometric Reviews*, 25(2-3):245–274.
- Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371.
- Chib, S., Omori, Y., and Asai, M. (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series*, pages 365–400. Springer.
- Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341.
- Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, 30(4):551–575.
- Creal, D. D. and Tsay, R. S. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2):335–345.

- Dendramis, Y., Kapetanios, G., and Tzavalis, E. (2015). Shifts in volatility driven by large stock market shocks. *Journal of Economic Dynamics and Control*, 55:130–147.
- Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*, pages 3–14. Springer.
- Fearnhead, P. (2008). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statistics and Computing*, 18(2):151–171.
- Flury, T. and Shephard, N. (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory*, 27(05):933–956.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica sinica*, pages 339–373.
- Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. In Bayesian Statistics 4 (eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith), Oxford: Oxford University Press.
- Geweke, J. (2007). Bayesian model comparison and validation. *The American Economic Review*, 97(2):60–64.
- Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230.
- Gordon, N. (1997). A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):353–358.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET.
- Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

- Hall, J., Pitt, M. K., and Kohn, R. (2014). Bayesian inference for nonlinear structural time series models. *Journal of Econometrics*, 179(2):99–111.
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.
- Jungbacker, B. and Koopman, S. J. (2006). Monte Carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews*, 25(2-3):385–408.
- Kapetanios, G. and Tzavalis, E. (2010). Modeling structural breaks in economic relationships using large shocks. *Journal of Economic Dynamics and Control*, 34(3):417–436.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.
- Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.
- Li, Y., Liu, X.-B., and Yu, J. (2015). A Bayesian chi-squared test for hypothesis testing. *Journal of Econometrics*, 189(1):54–69.
- Liesenfeld, R. and Richard, J.-F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10(4):505–531.
- Lin, M. T., Zhang, J. L., Cheng, Q., and Chen, R. (2005). Independent particle filters. *Journal of the American Statistical Association*, 100(472):1412–1421.
- Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice*, pages 197–223. Springer.
- Loddo, A., Ni, S., and Sun, D. (2011). Selection of multivariate stochastic volatility models via Bayesian stochastic search. *Journal of Business & Economic Statistics*, 29(3):342–355.
- Lopes, H. F. and Carvalho, C. M. (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137(10):3082–3091.

- Nemeth, C. and Fearnhead, P. (2014). Particle Metropolis adjusted Langevin algorithms for state space models. *arxiv.org*.
- Ntzoufras, I., Forster, J. J., and Dellaportas, P. (2000). Stochastic search variable selection for log-linear models. *Journal of Statistical Computation and Simulation*, 68(1):23–37.
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.
- Peracchi, F. and Rossetti, C. (2013). The heterogeneous thresholds ordered response model: Identification and inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3):703–722.
- Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29.
- Philipov, A. and Glickman, M. E. (2006). Multivariate stochastic volatility via Wishart processes. *Journal of Business & Economic Statistics*, 24(3):313–328.
- Pitt, M. K., dos Santos Silva, R., Giordani, P., and Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151.
- Pitt, M. K. and Shephard, N. (1999a). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.
- Pitt, M. K. and Shephard, N. (1999b). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.
- Poirier, D. J. (1995). *Intermediate statistics and Econometrics: A Comparative Approach*. MIT Press, Cambridge, MA.
- Quintana, J. M. and West, M. (1987). An analysis of international exchange rates using multivariate DLM's. *The Statistician*, pages 275–281.
- Ristic, B., Arulampalam, S., and Gordon, N. J. (2004). *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house.
- Robert, C. (2001). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, second edition edition.

- Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.
- Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.
- Taylor, J. W. and Yu, K. (2016). Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):1069–1092.
- Yu, J. and Meyer, R. (2006a). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25(2-3):361–384.
- Yu, J. and Meyer, R. (2006b). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25(2-3):361–384.

Tables

Table 1: Results for Simulation A. T denotes the sample size, NP denotes the number of particles and NI the number of iterations

NP	NI	γ_{ij}	σ_j^2	ϕ_j	$\sigma_{\zeta^L_j}^2 \& \sigma_{\zeta^R_j}^2$	$\delta_{ij}^L \& \delta_{ij}^R$	$\psi_L \& \psi_R$	γ_{ij}	σ_j^2	ϕ_j	$\sigma_{\zeta^L_j}^2 \& \sigma_{\zeta^R_j}^2$	$\delta_{ij}^L \& \delta_{ij}^R$	$\psi_L \& \psi_R$
		T = 500						T = 1000					
10^3	5000	0.292	0.277	0.245	0.233	0.189	0.263	0.213	0.187	0.200	0.179	0.135	0.174
	10000	0.287	0.261	0.220	0.217	0.177	0.244	0.185	0.172	0.182	0.182	0.177	0.166
	20000	0.231	0.222	0.215	0.201	0.163	0.231	0.177	0.144	0.144	0.144	0.12	0.182
	40000	0.231	0.222	0.215	0.201	0.163	0.23	0.177	0.144	0.144	0.145	0.12	0.183
10^4	5000	0.290	0.262	0.232	0.228	0.170	0.242	0.211	0.185	0.204	0.175	0.130	0.171
	10000	0.283	0.255	0.218	0.215	0.165	0.233	0.182	0.171	0.183	0.181	0.116	0.162
	20000	0.230	0.219	0.208	0.198	0.151	0.216	0.175	0.142	0.141	0.142	0.118	0.179
	40000	0.230	0.219	0.208	0.198	0.151	0.216	0.175	0.142	0.141	0.142	0.118	0.179
10^6	5000	0.266	0.244	0.219	0.232	0.166	0.217	0.257	0.272	0.188	0.190	0.155	0.192
	10000	0.244	0.232	0.202	0.206	0.144	0.202	0.231	0.244	0.200	0.188	0.132	0.188
	20000	0.215	0.202	0.185	0.173	0.132	0.188	0.203	0.133	0.084	0.092	0.085	0.105
	40000	0.215	0.203	0.187	0.173	0.132	0.185	0.206	0.133	0.085	0.091	0.085	0.104
10^7	5000	0.266	0.244	0.219	0.232	0.166	0.217	0.257	0.272	0.188	0.190	0.155	0.192
	10000	0.244	0.232	0.202	0.206	0.144	0.202	0.231	0.244	0.200	0.188	0.132	0.188
	20000	0.215	0.202	0.185	0.173	0.132	0.188	0.203	0.133	0.084	0.092	0.085	0.105
	40000	0.215	0.203	0.187	0.173	0.132	0.185	0.206	0.133	0.085	0.091	0.085	0.104
		T = 2000						T = 4000					
10^3	5000	0.181	0.121	0.144	0.101	0.119	0.115	0.109	0.083	0.144	0.089	0.073	0.087
	10000	0.174	0.118	0.132	0.092	0.105	0.092	0.104	0.081	0.132	0.074	0.062	0.073
	20000	0.103	0.094	0.091	0.095	0.095	0.104	0.089	0.077	0.091	0.082	0.091	0.093
	40000	0.103	0.094	0.091	0.095	0.095	0.103	0.089	0.077	0.091	0.081	0.092	0.092
10^4	5000	0.172	0.119	0.137	0.098	0.103	0.089	0.108	0.08	0.135	0.081	0.071	0.081
	10000	0.165	0.112	0.129	0.091	0.098	0.097	0.103	0.077	0.121	0.068	0.067	0.077
	20000	0.101	0.091	0.084	0.087	0.091	0.094	0.084	0.073	0.087	0.078	0.087	0.088
	40000	0.101	0.091	0.084	0.088	0.091	0.094	0.084	0.073	0.086	0.078	0.086	0.088
10^6	5000	0.233	0.232	0.172	0.177	0.147	0.115	0.182	0.155	0.166	0.122	0.122	0.103
	10000	0.212	0.214	0.184	0.104	0.122	0.115	0.177	0.133	0.161	0.082	0.092	0.098
	20000	0.124	0.092	0.071	0.072	0.065	0.083	0.071	0.061	0.044	0.055	0.051	0.062
	40000	0.124	0.091	0.071	0.072	0.066	0.083	0.072	0.061	0.044	0.055	0.051	0.062
10^7	5000	0.233	0.232	0.172	0.177	0.147	0.115	0.182	0.155	0.166	0.122	0.112	0.103
	10000	0.212	0.214	0.184	0.104	0.122	0.115	0.177	0.133	0.161	0.082	0.092	0.098
	20000	0.124	0.092	0.071	0.072	0.065	0.083	0.071	0.061	0.044	0.055	0.051	0.062
	40000	0.124	0.091	0.071	0.072	0.066	0.083	0.072	0.061	0.044	0.055	0.051	0.062

Table 2: Results for Simulation B. MAPCPM stands for "maximum absolute percentage change in posterior means". MAPCPSD stands for "maximum absolute percentage change in posterior s.d."

Change ω_s	MAPCPM	MAPCPSD
1/4	0.53%	0.25%
1/2	0.22%	0.12%
2	0.14%	0.33%
4	0.40%	0.42%

Table 3: Alternatives Models

Model	Description	Equations
I	Simple MSV without thresholds	(5), (6), (7)
II	Common thresholds for factor and return volatility	(9), (10)
III	Different thresholds for factor and return volatility	(14), (12)
IV	Alternative MSV model with common thresholds	(12)
V	Alternative MSV model with different thresholds	Variation of (12)
VI	Alternative MSV model with grouped thresholds	Variation of (12)

Table 4: Prior Sensitivity of Bayes Factors

Bayes Factor	% Deviation from baseline		
	median	minimum	maximum
In favour of MSV with large shocks and against MSV without large shocks	1.24%	0.42%	3.21%
In favour of endogeneity	0.77%	0.32%	1.44%

Table 5: Bayes factors for different numbers of factors. Results in red suggest that we should proceed with $k = 5$ common factors.

Model	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 7$	$k = 10$
IV	1.00	1.81	2.35	3.44	12.77	3.31	1.05
V	1.00	1.27	1.81	2.82	14.61	4.45	2.20
VI	1.00	2.35	3.44	4.77	16.71	3.12	2.44

Table 6: Bayes factor for alternatives models.

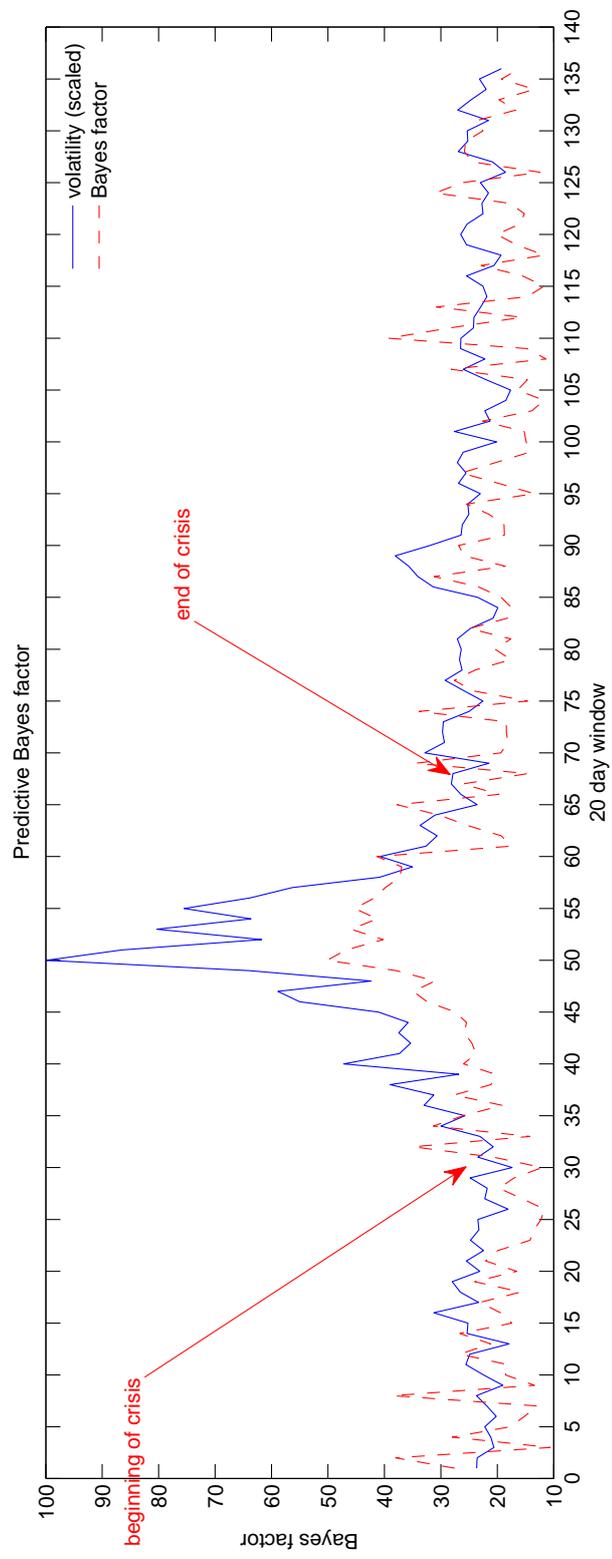
Model	Description	1/1/2005 - 11/05/2008	12/05/2008 - 21/10/2009	22/10/2009 - 24/6/2015
I	Simple MSV without thresholds	1	1	1
II	Common thresholds for factor and return volatility	3.717	6.212	6.553
III	Different thresholds for factor and return volatility	8.218	12.251	11.334
IV	Alternative MSV model with common thresholds	17.356	21.515	18.817
V	Alternative MSV model with different thresholds	23.716	35.988	35.585
VI	Alternative MSV model with grouped thresholds	21.453	17.554	19.425

Table 7: Posterior Statistics. The results refer to equations (26). Posterior means are reported with posterior standard deviations in parentheses. IF is the inefficiency factor based on 2,500 lags. When IF=1 it means we have i.i.d sampling from the posterior. IF is defined as $IF = 1 + \sum_{s=1}^{\infty} \varrho_s$ where ϱ_s is the sample autocorrelation at lag s of MCMC draws for a given parameter. We have $h_t = p^{-1} \sum_{j=1}^p h_{tj}$ which is average volatility across all p stock returns, and h_t^f is the average factor volatility ($h_t^f := k^{-1} \sum_{i=1}^k h_{t,p+i}$). For ψ_L and ψ_R , we assume a prior: $[\psi'_L, \psi'_R]' \sim \mathcal{N}(0, \omega_\psi^2 I)$, and we set $\omega_\psi = 10$. The probability that the lower threshold is exceeded is $\Phi(z'_t \psi_L)$ and the probability that the upper threshold is exceeded is, similarly, $1 - \Phi(z'_t \psi_R)$.

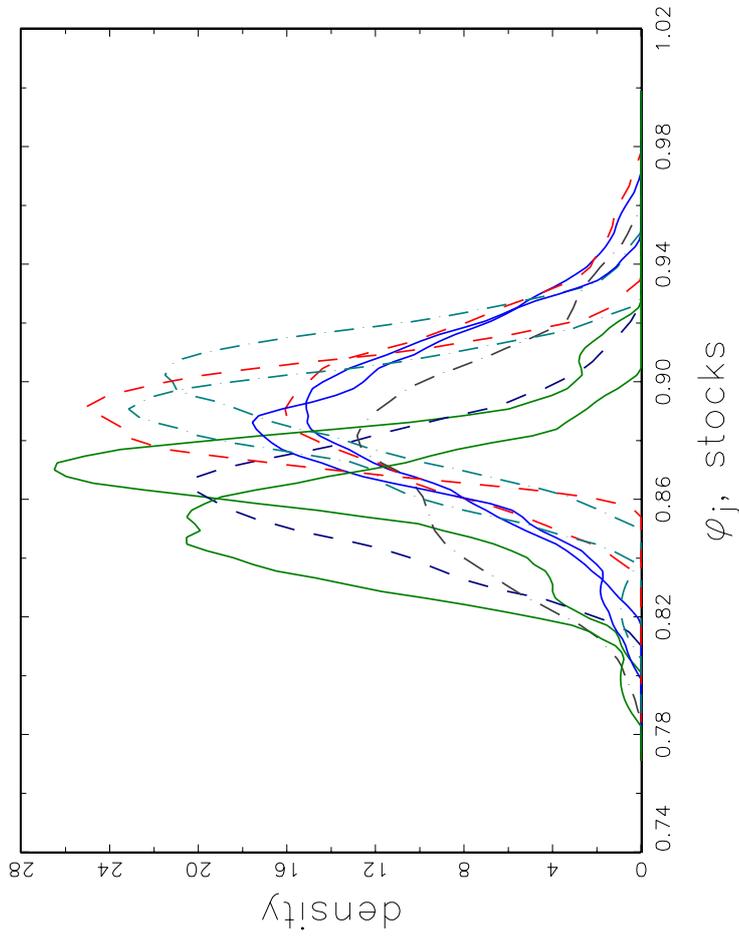
	L_t^*	R_t^*
constant	0.012	0.023
	(0.0017)	(0.0012)
	IF = 5.15	IF = 7.47
f_t	0.014	0.017
	(0.0025)	(0.002)
	IF=12.32	IF=8.45
f_{t-1}	0.0015	0.0021
	(0.0002)	(0.0004)
	IF=3.15	IF=2.88
h_{t-1}	0.022	0.0014
	(0.0013)	(0.0003)
	IF=5.38	IF=4.34
h_{t-1}^f	0.0017	0.0012
	(0.0002)	(0.0003)
	IF=2.81	IF=3.77

Figures

Figure 1: Predictive Bayes factors against MSV without large shocks. We take the beginning of the crisis as December 1, 2007 and the end of the crisis as June 30, 2009



(a) Representative marginal posteriors of ϕ_j (stocks)



(b) Representative marginal posteriors of γ_{ij} (stocks)

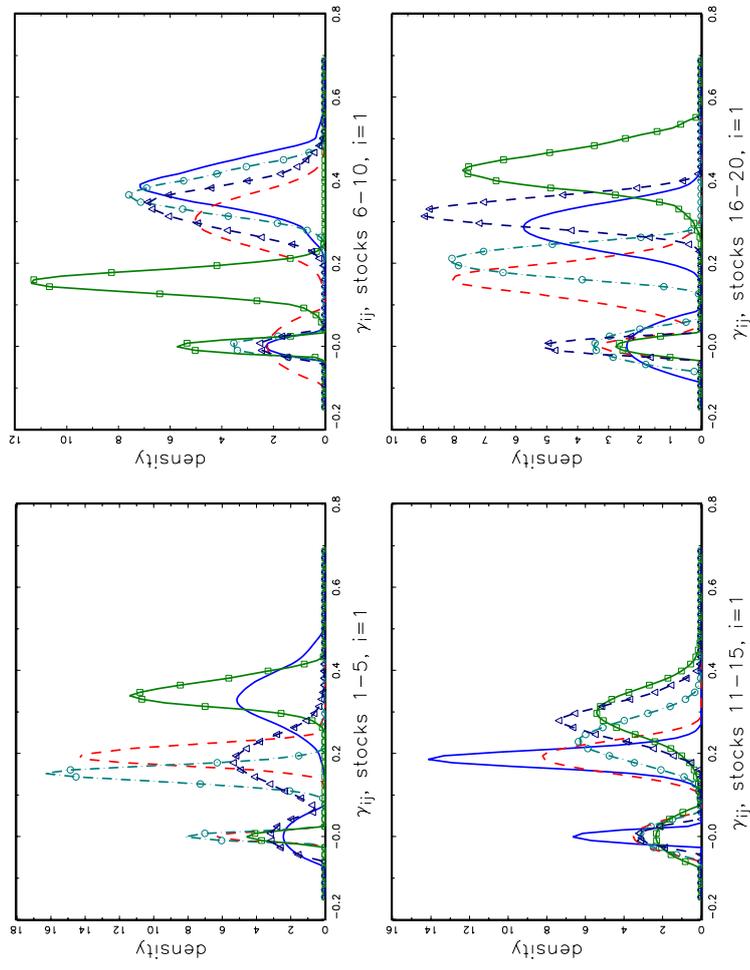
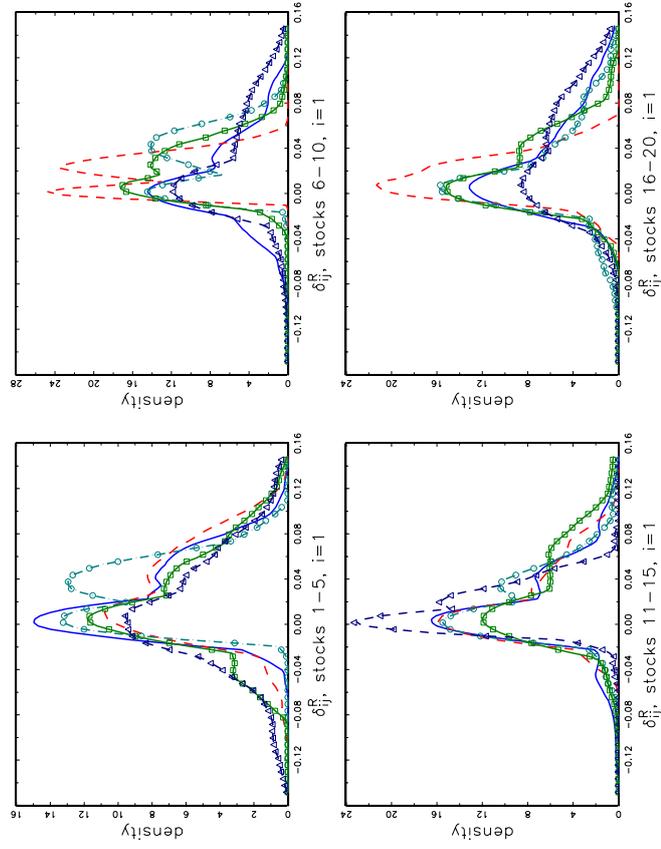


Figure 2: Representative Marginal Posteriors of ϕ_j and γ_{ij} (stocks)

(a) Representative marginal posteriors of δ_{ij}^R (stocks)



(b) Representative marginal posteriors of δ_{ij}^L (stocks)

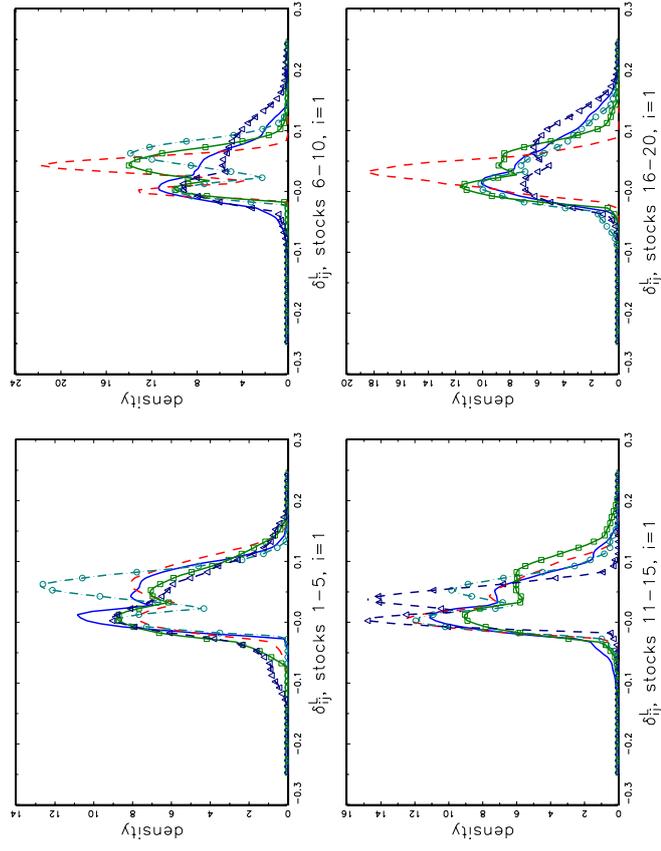


Figure 3: Representative Marginal Posteriors of δ_{ij}^R and δ_{ij}^L (stocks)

Figure 4: Average volatility from different models

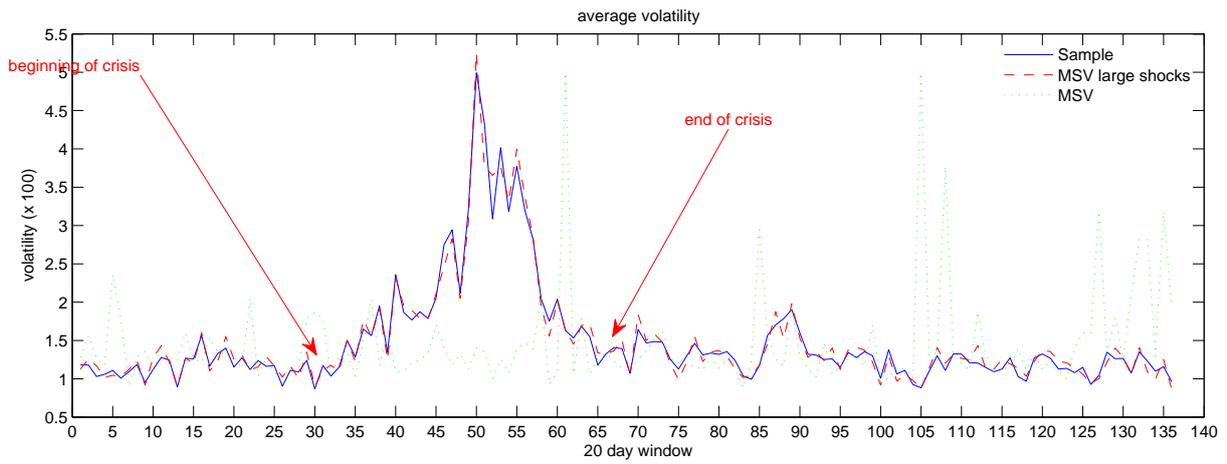


Figure 5: Predictive Bayes factors in favor of endogeneity

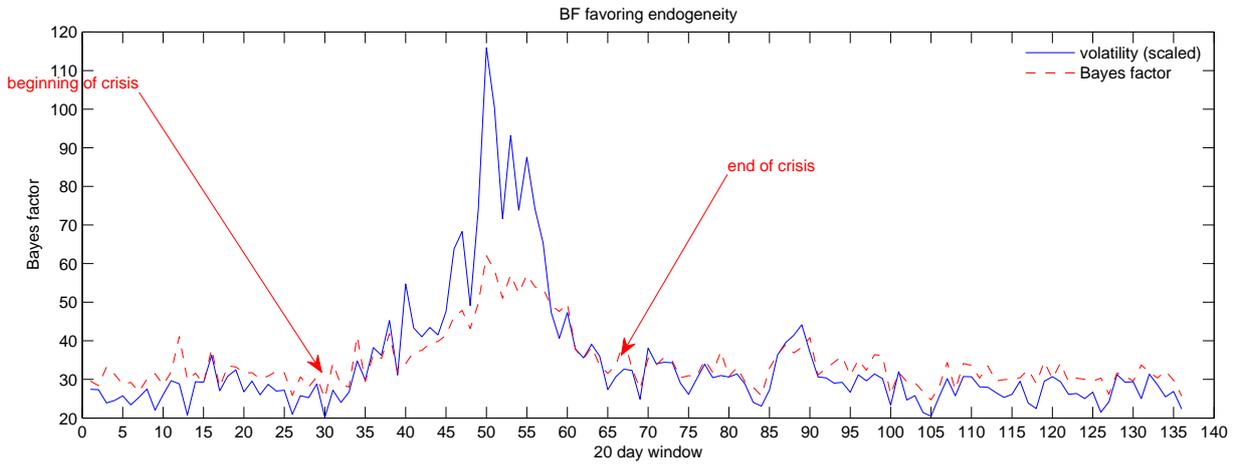


Figure 6: Marginal posterior densities of thresholds (returns)

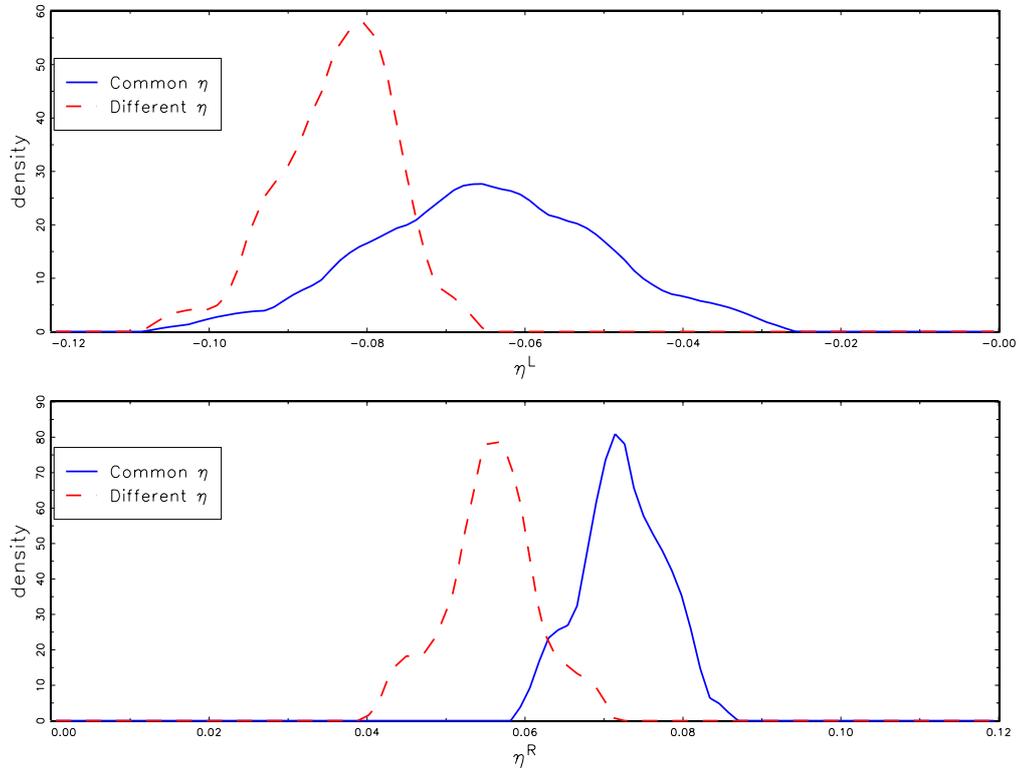


Figure 7: Marginal posterior densities of thresholds (factor). Case A: Common thresholds in returns. Case B: Grouped thresholds in returns. Case C: Different thresholds in returns. In Case B, from the overall marginal likelihood of the model, three different groups are selected.

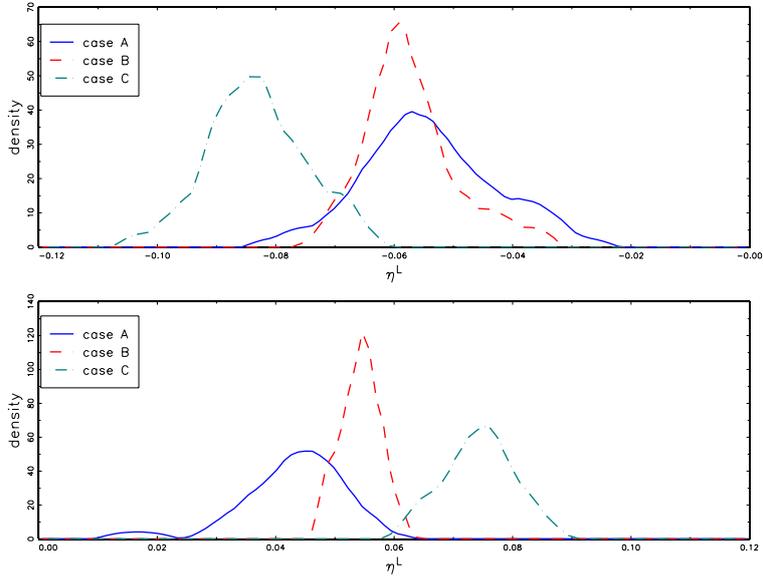


Figure 8: GIRF, large negative shocks. GIRF is average of GIRFs for individual stocks from the MSV model allowing for stock-specific thresholds. The figure shows the effect (GIRF) of shocks on volatility.

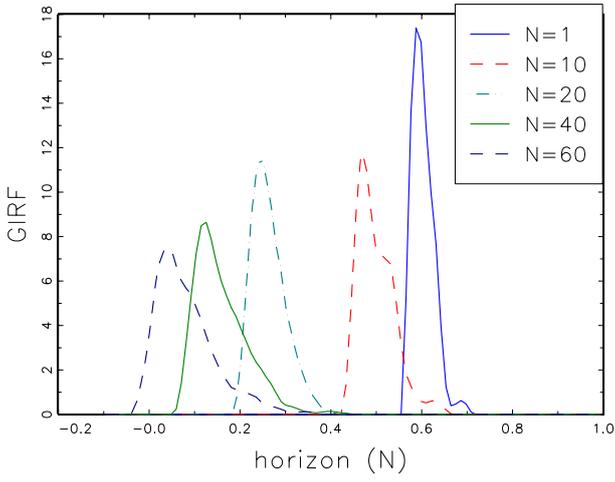


Figure 9: GIRF, large positive shocks. GIRF is average of GIRFs for individual stocks from the MSV model allowing for stock-specific thresholds. The figure shows the effect (GIRF) of shocks on volatility.

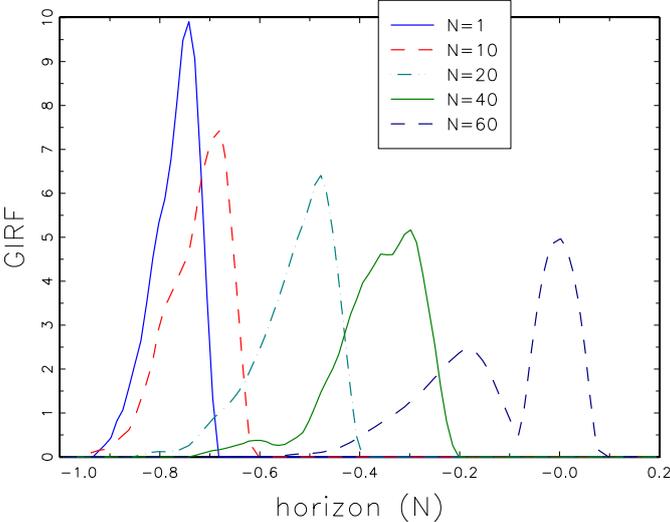


Figure 10: GRF, ordinary positive shocks. GIRF is average of GIRFs for individual stocks from the MSV model allowing for stock-specific thresholds. The figure shows the effect (GIRF) of shocks on volatility.

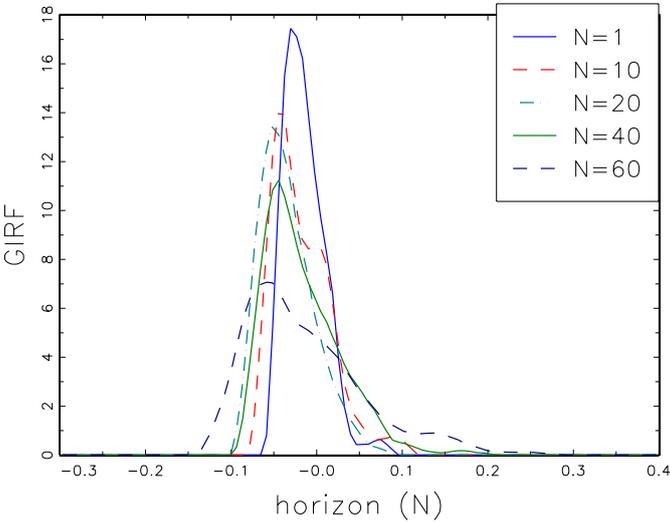
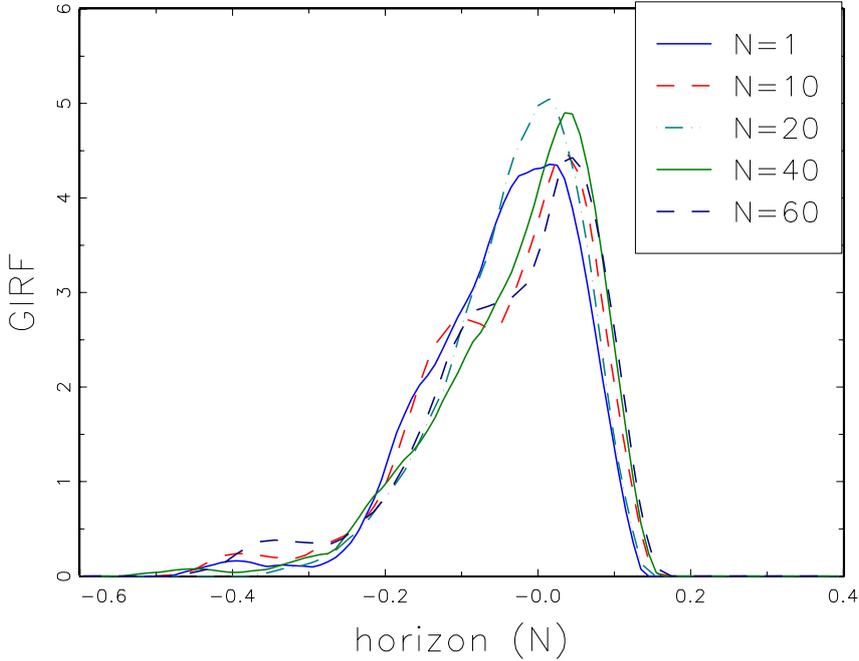


Figure 11: GIRF, ordinary negative shocks. GIRF is average of GIRFs for individual stocks from the MSV model allowing for stock-specific thresholds. The figure shows the effect (GIRF) of shocks on volatility.



TECHNICAL APPENDIX

A Particle filtering

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

$$y_T \sim p(y_T|x_T), \quad s_t \sim p(s_t|s_{t-1}), \quad (\text{A.1})$$

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 $\{s_t^{(i)}, i = 1, \dots, N\}$ with probability weights $\{w_t^{(i)}, i = 1, \dots, N\}$ where $\sum_{i=1}^N w_t^{(i)} = 1$. The predictive density can be approximated by:

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

and the final approximation for the filtering density is:

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

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

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

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 et al. \(2009\)](#).

[Andrieu and Roberts \(2009\)](#), [Flury and Shephard \(2011\)](#), and [Pitt et al. \(2012\)](#) provide the Particle Metropolis-Hastings (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 $L^c = \hat{p}_N(Y|\theta^c)$. We then set $\theta^{(j+1)} = \theta^c$ with

the Metropolis - Hastings probability:

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

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

Hall et al. (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 than the standard particle filtering algorithm to approximate $p(Y|\theta)$. From Pitt and Shephard (1999b) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t|s_{t-1})$. Whenever this is not possible, Hall et al. (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:

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

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

If it is possible to 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 (1999b)). A Gaussian approximation can be used 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 $p(y_t|s_{t-1}) = \mathcal{N}(\mathbb{E}(y_t|s_{t-1}), \mathbb{V}(y_t|s_{t-1}))$. 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 $m = 1, \dots, M$. 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:

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

Then we can construct a mixture approximation:

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

where $\Sigma_m = -[\nabla^2 l(\hat{u}_t^m)]^{-1}$ and $\lambda_m \propto \exp \{l(u_t^m)\}$ with $\sum_{m=1}^M \lambda_m = 1$. 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 as Hall et al. (2014) indicate, 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 $t = 0, \dots, T - 1$ given samples $s_t^k \sim p(s_t | Y_{1:t})$ with mass π_t^k for $k = 1, \dots, N$.

- 1) For $k = 1, \dots, N$ 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 $k = 1, \dots, N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t^i}^i(ds_t)$.
- 3) For $k = 1, \dots, N$ 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 $k = 1, \dots, N$ compute

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

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

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

B Particle Metropolis adjusted Langevin filters

Nemeth 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:

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

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 $\{s_{t-1}^i, i = 1, \dots, N\}$ and weights $\{w_{t-1}^i, i = 1, \dots, N\}$ which form a discrete approximation for $p(s_{t-1}|y_{1:t-1}, \theta)$ we have the approximation:

$$\hat{p}(s_{t-1}|y_{1:t-1}, \theta) \propto \sum_{i=1}^N w_{t-1}^i f(s_t|s_{t-1}^i, \theta). \quad (\text{B.2})$$

See Andrieu et al. (2010) and Cappé et al. (2005) for reviews. From (B.2) Fearnhead (2008) makes the important observation that the joint probability of sampling particle s_{t-1}^i and state s_t is:

$$\omega_t = \frac{w_{t-1}^i g(y_t|s_t, \theta) f(s_t|s_{t-1}^i, \theta)}{\xi_t^i q(s_t|s_{t-1}^i, y_t, \theta)}, \quad (\text{B.3})$$

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

$$\xi_t^i q(s_t|s_{t-1}^i, y_t, \theta) \simeq c g(y_t|s_t, \theta) f(s_t|s_{t-1}^i, \theta), \quad (\text{B.4})$$

and c is the normalizing constant in (B.1).

In the MALA algorithm of Roberts and Rosenthal (1998),⁹ we form a proposal:

$$\theta^c = \theta^{(s)} + \lambda z + \frac{\lambda^2}{2} \nabla \log p(\theta^{(s)}|Y_{1:T}), \quad (\text{B.5})$$

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

$$\alpha(\theta^c|\theta^{(s)}) = \min \left\{ 1, \frac{p(Y_{1:T}|\theta^c) q(\theta^{(s)}|\theta^c)}{p(Y_{1:T}|\theta^{(s)}) q(\theta^c|\theta^{(s)})} \right\}. \quad (\text{B.6})$$

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

$$\nabla \log p(Y_{1:T}|\theta) = E [\nabla \log p(s_{1:T}, Y_{1:T}|\theta) | Y_{1:T}, \theta], \quad (\text{B.7})$$

⁹The 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)

which corresponds to the expectation of:

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

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:

$$\alpha_t^i = \alpha_{t-1}^{\kappa_i} + \nabla \log g(y_t|s_t^i, \theta) + \nabla \log f(s_t^i|s_{t-1}^{\kappa_i}, \theta). \quad (\text{B.8})$$

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

$$\alpha_{t-1}^i \sim N(m_{t-1}^i, V_{t-1}). \quad (\text{B.9})$$

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

$$m_{t-1}^i = \delta \alpha_{t-1}^i + (1 - \delta) \sum_{i=1}^N w_{t-1}^i \alpha_{t-1}^i, \quad (\text{B.10})$$

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:

$$m_t^i = \delta m_{t-1}^{\kappa_i} + (1 - \delta) \sum_{i=1}^N w_{t-1}^i m_{t-1}^i + \nabla \log g(y_t|s_t^i, \theta) + \nabla \log f(s_t^i|s_{t-1}^{\kappa_i}, \theta), \quad (\text{B.11})$$

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t}|\theta) = \sum_{i=1}^N w_t^i m_t^i. \quad (\text{B.12})$$

As a rule of thumb [Nemeth 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:

$$p(Y_{1:T}|\theta) = p(y_1|\theta) \prod_{t=2}^T p(y_t|Y_{1:t-1}, \theta), \quad (\text{B.13})$$

where

$$p(y_t|Y_{1:t-1}, \theta) = \int g(y_t|s_t) \int f(s_t|s_{t-1}, \theta) p(s_{t-1}|Y_{1:T-1}, \theta) ds_{t-1} ds_t, \quad (\text{B.14})$$

provides, in explicit form, the predictive likelihood.

A APPENDIX FIGURES

In this Appendix we provide marginal posterior densities of parameters δ_{ij}^R and δ_{ij}^L for all stocks $j = 1, \dots, 100$ and five randomly selected stocks $i \in \{2, \dots, 100\}$. The reason is that all possible cases are 10,000 and visual presentation is rather difficult.

For $i = 1$ representative marginal posterior densities of these parameters have been presented in Figures [A.3](#) for $i = 1, \dots, 20$ randomly selected stocks.

In Figures [A.1-A.5](#) we present posterior densities of parameters δ_{ij}^R for randomly selected stock i but all stocks $j = 1, \dots, 100$.

In Figures [A.6-A.10](#) we present posterior densities of parameters δ_{ij}^L for (the same) randomly selected stock i but all stocks $j = 1, \dots, 100$.

Qualitatively and quantitatively, these marginal posterior densities are essentially like Figures 3a and 3b although the positive and negative posterior means of δ_{ij}^R and δ_{ij}^L , respectively, are more pronounced, reinforcing the findings in the paper.

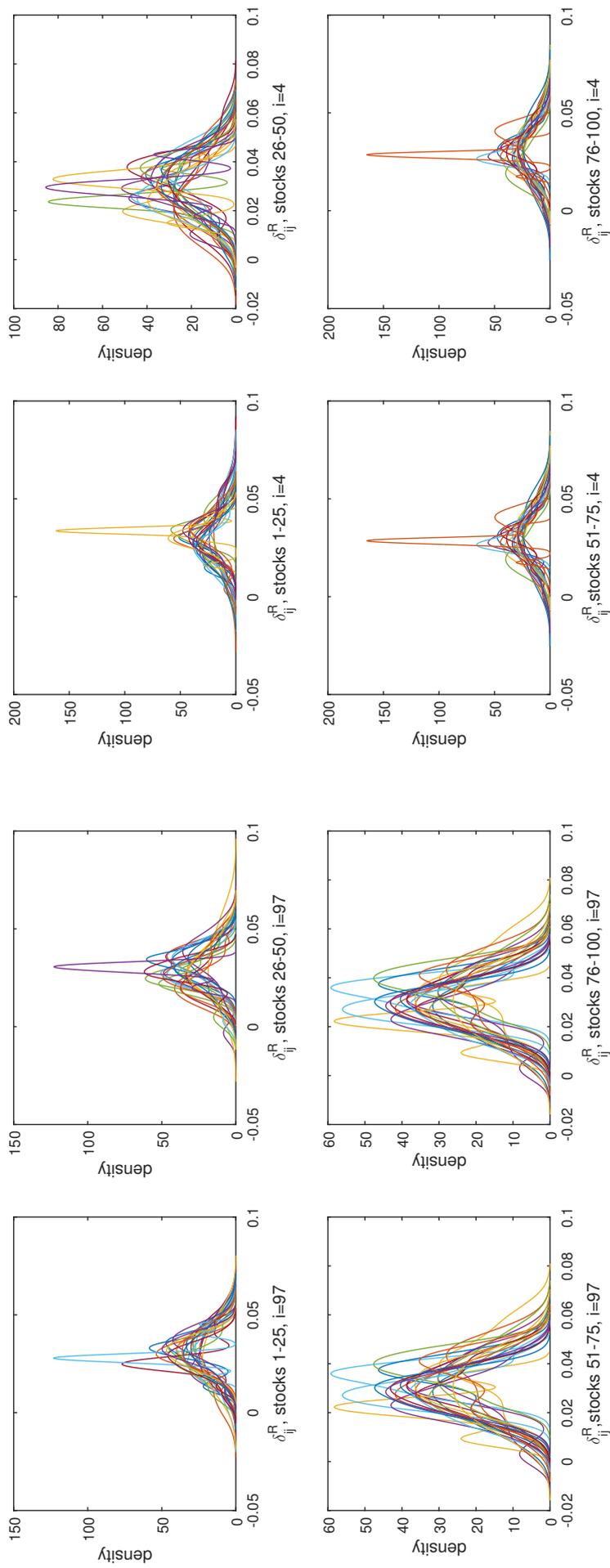


Figure A.1: Marginal posterior densities of δ_{ij}^R for all stocks $j = 1, \dots, 100$ and $i = 97$.

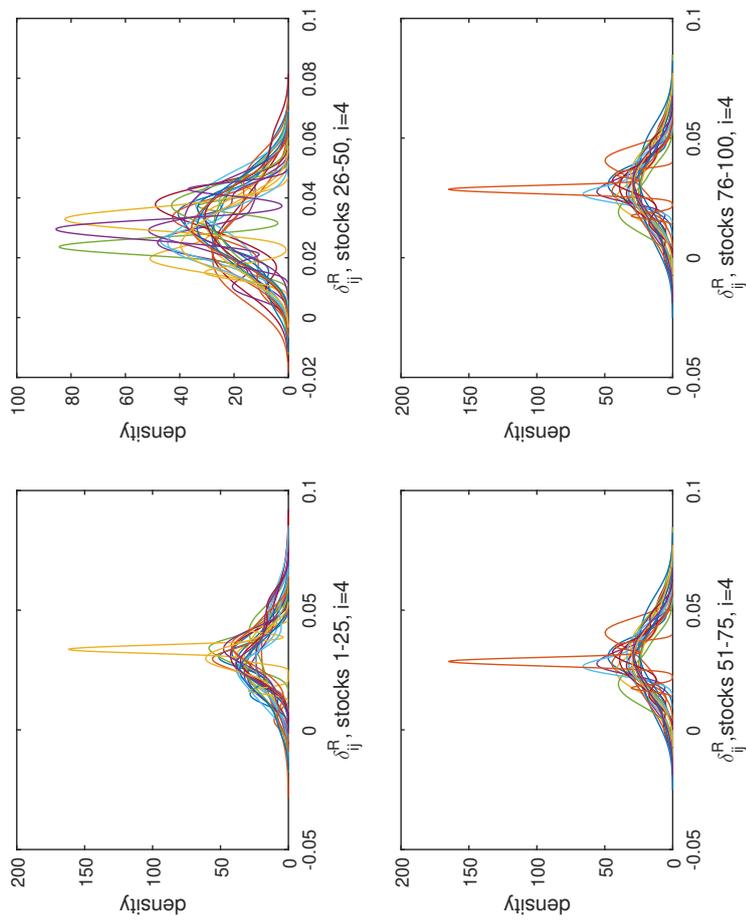


Figure A.2: Marginal posterior densities of δ_{ij}^R for all stocks $j = 1, \dots, 100$ and $i = 4$.

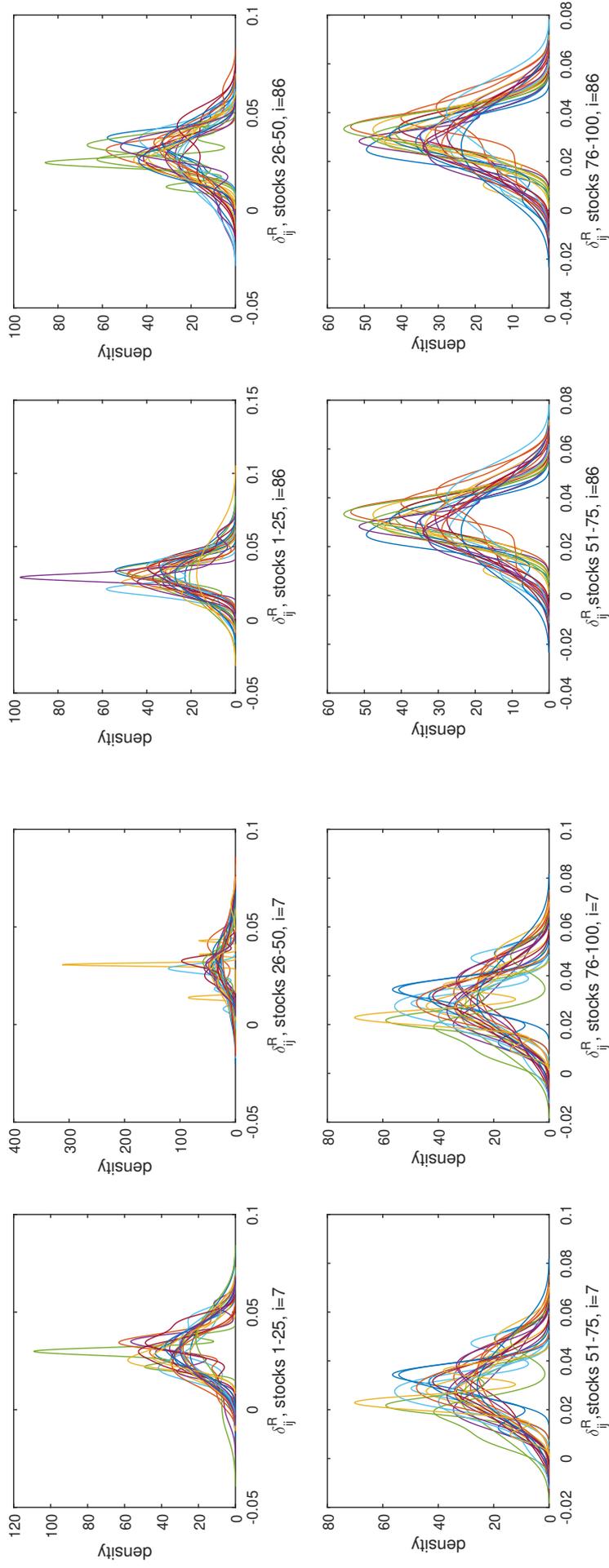


Figure A.3: Marginal posterior densities of δ_{ij}^R for all stocks $j = 1, \dots, 100$ and $i = 7$.

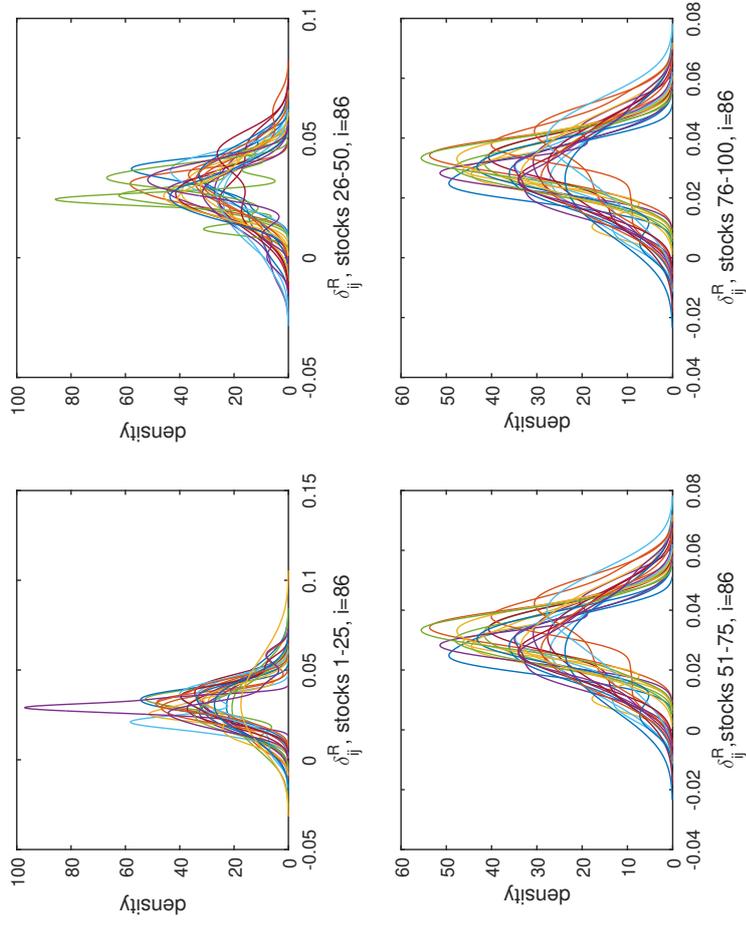


Figure A.4: Marginal posterior densities of δ_{ij}^R for all stocks $j = 1, \dots, 100$ and $i = 86$.

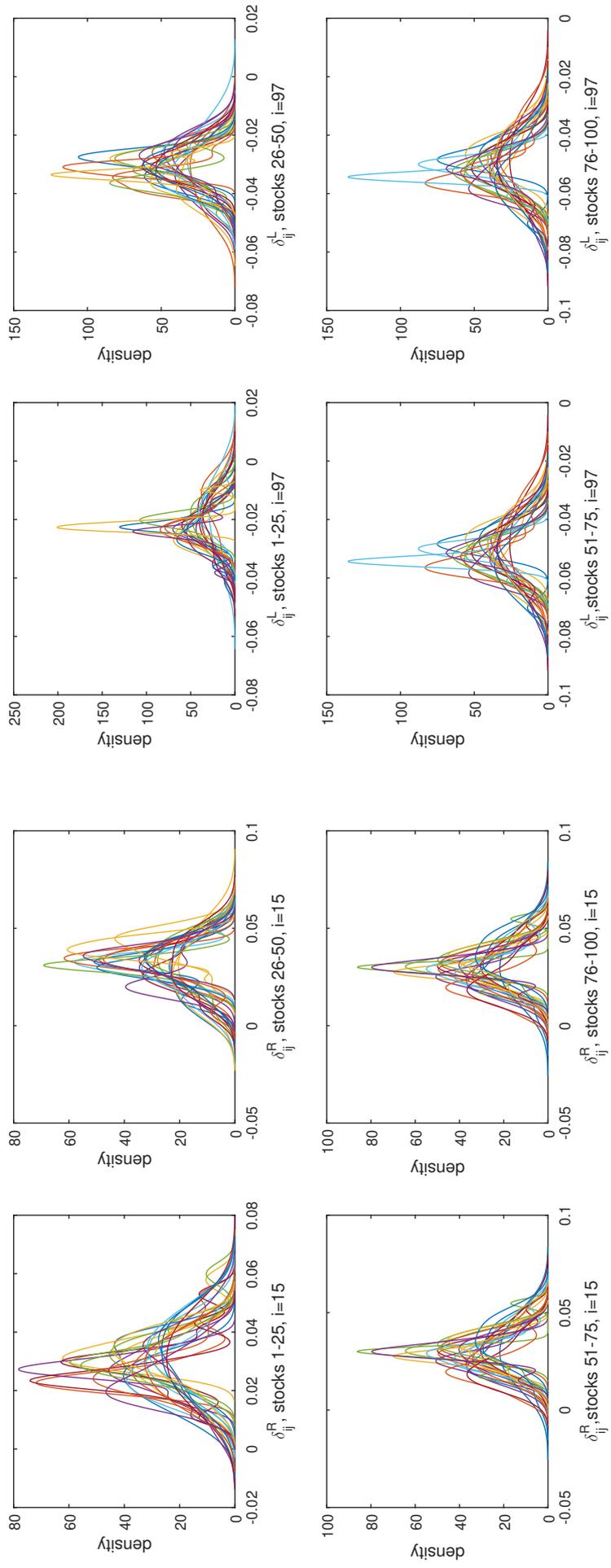


Figure A.5: Marginal posterior densities of δ_{ij}^R for all stocks $j = 1, \dots, 100$ and $i = 15$.

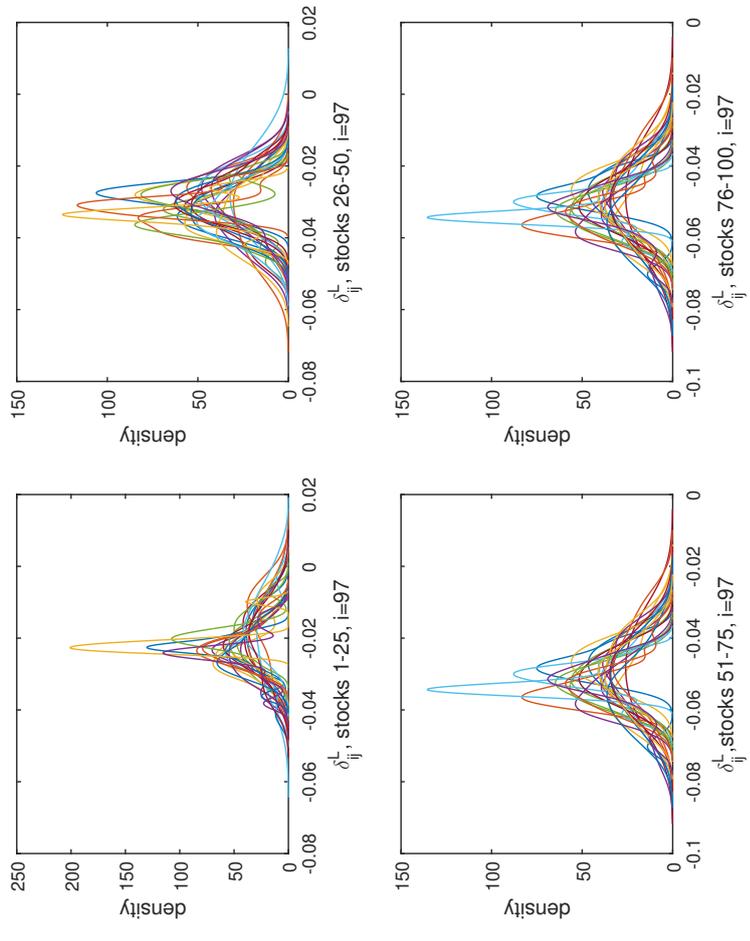


Figure A.6: Marginal posterior densities of δ_{ij}^L for all stocks $j = 1, \dots, 100$ and $i = 97$.

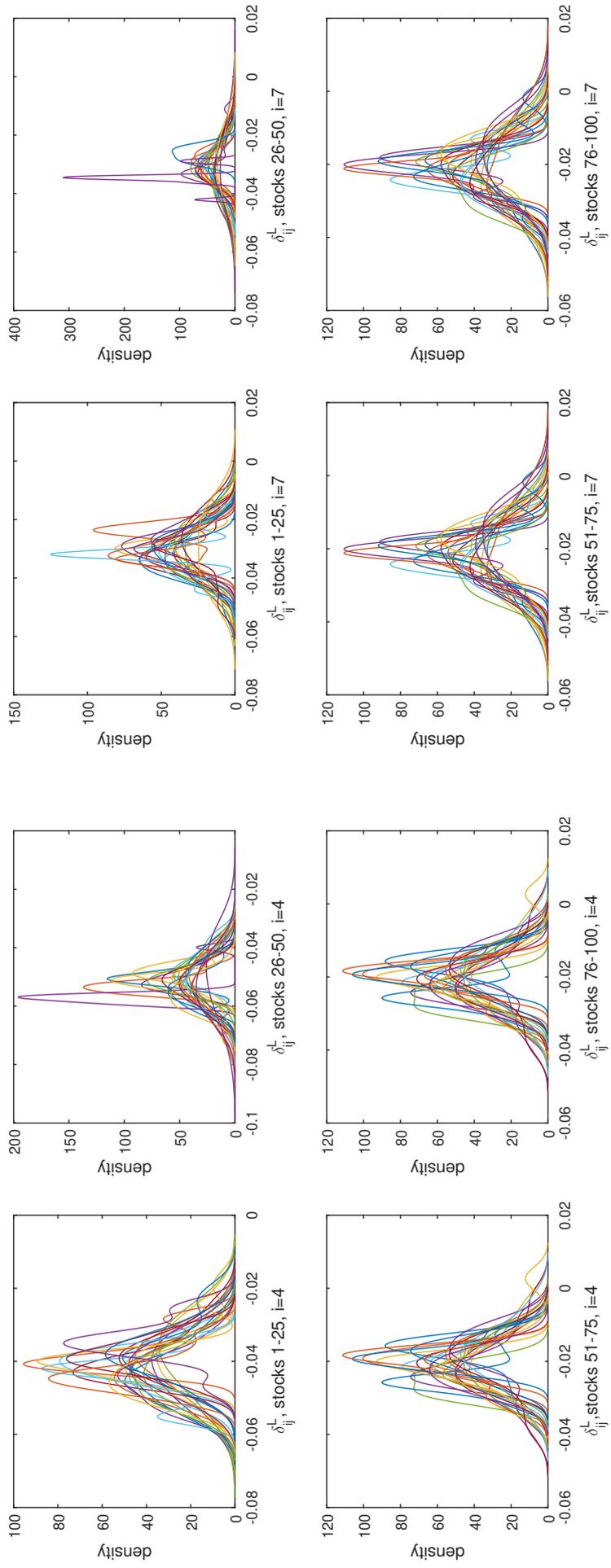


Figure A.7: Marginal posterior densities of δ_{ij}^L for all stocks $j = 1, \dots, 100$ and $i = 4$.

Figure A.8: Marginal posterior densities of δ_{ij}^L for all stocks $j = 1, \dots, 100$ and $i = 7$.

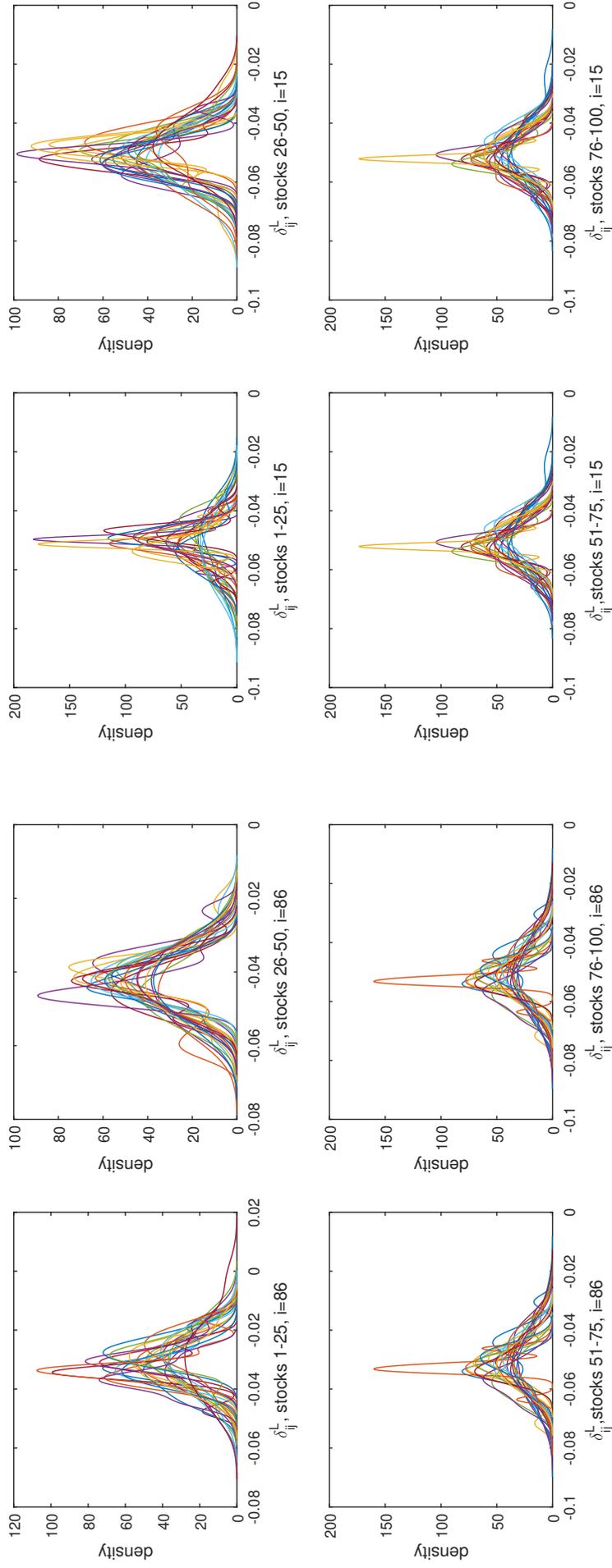


Figure A.9: Marginal posterior densities of δ_{ij}^L for all stocks $j = 1, \dots, 100$ and $i = 86$.

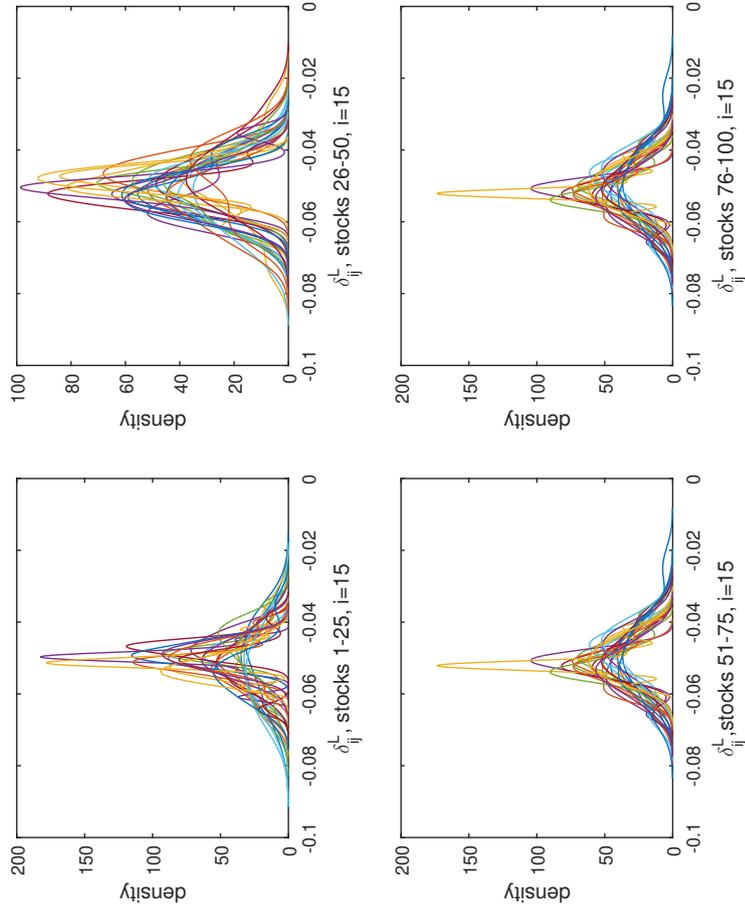


Figure A.10: Marginal posterior densities of δ_{ij}^L for all stocks $j = 1, \dots, 100$ and $i = 15$.