Business and Economics
Working Papers

BEWP 205/2014

Parsimony Inducing Priors for Large Scale State-Space Models

H. F. Lopes
R. E. McCulloch
R. S. Tsay

Parsimony Inducing Priors for Large Scale State-Space Models

H. F. Lopes, R. E. McCulloch and R. S. Tsay*

September 2, 2014

Abstract

State-space models are commonly used in the engineering, economic, and statistical literatures. They are flexible and encompass many well-known statistical models, including random coefficient autoregressive models and dynamic factor models. Bayesian analysis of state-space models has attracted much interest in recent years. However, for large scale models, prior specification becomes a challenging issue in Bayesian inference. In this paper, we propose a flexible prior for state-space models. The proposed prior is a mixture of four commonly entertained models, yet achieves parsimony in high-dimensional systems. Here "parsimony" is represented by the idea that in a large system, some states may not be time-varying. Simulation and simple examples are used throughout to demonstrate the performance of the proposed prior. As an application, we consider the time-varying conditional covariance matrices of daily log returns of 94 components of the S&P 100 index, leading to a state-space model with $94 \times 95/2$=4,465 time-varying states. Our model for this large system enables us to use parallel computing.

KEY WORDS: Bayesian modeling; Conditional Heteroscedasticity; Forward Filtering and Backward Sampling; Parallel Computing; Prior; random walk.

*Hedibert F. Lopes is Professor of Statistics and Econometrics, INSPER Institute of Education and Research, Rua Quatá 300, Vila Olímpia, São Paulo/SP, Brazil, 04546-042. E-mail: `hedibertFL@insper.edu.br`. Robert E. McCulloch is Katherine Dusak Miller Professor of Econometrics and Statistics, and Ruey S. Tsay is H. G. B. Alexander Professor of Econometrics and Statistics, The University of Chicago Booth School of Business, 5807 S. Woodlawn Ave, Chicago, IL 60637. Emails: `[mcculloch,ruey.tsay]@chicagobooth.edu`.

1 Introduction

State-space models, also known as *dynamic models*, are well established in many scientific areas ranging from signal processing to spatio-temporal modeling to marketing applications, to name only a few. See Migon, Gamerman, Lopes & Ferreira (2005) for a recent review of dynamic models. In the recent business and economics literature, these state-space structures have gained additional attention, particularly in macroeconomic and financial applications where they are used, respectively, when describing time-varying parameters (TVP) in vector autoregressive (VAR) models (Primiceri 2005) or in large-scale dynamic factor models (DFM); and time-varying variances and covariances in stochastic volatility (SV) models (Lopes & Polson 2010).

More specifically, the basic dynamics governing the state-space component, namely s_t, which can be a time-varying coefficient in the VAR model, a log-volatility in a SV model or a time-varying loading in a DFM, resembles a first order autoregressive, AR(1), model,

$$s_t = \alpha + \beta\, s_{t-1} + \tau\varepsilon_t,$$

where the errors $\varepsilon_1, \ldots, \varepsilon_T$ are independent and identically distributed, usually standard normals. Primiceri (2005), for example, models the US economy with a trivariate TVP-VAR model containing inflation rate, unemployment rate and short-term interest rate. The paper assumes $(\alpha, \beta) = (0, 1)$ when modeling the time-varying coefficients of the VAR, giving random walk dynamics. Similarly, the conditionally conjugate normal-inverse gamma prior, commonly used in the Bayesian state-space literature to model (α, β, τ^2), fails to properly account for common parsimonious/shrinkage cases.

We argue, and show in our applications, that limiting the evolution of state-space components to a random walk and/or using conditionally conjugate normal-inverse gamma priors for (α, β, τ^2) are both unrealistic practices. This is particularly troubling when dealing with large-scale systems where several hundreds, or thousands, of coefficient are essentially flat-line, rendering the random walk hypothesis meaningless. One of our main goals, extensively discussed in Section 2, is to propose a general mixture prior structure that allows us to entertain and investigate different kinds of state evolution within the simple AR(1) framework. More specifically, we will focus our attention on parsimonious/shrinkage cases, such as $(\alpha, \beta) = (0, 1)$ (random walk component), $\beta = 0$ (flat-lined component), $\alpha = \beta = 0$ (irrelevant state-space component) and $0 < \beta < 1$ (stationary component).

Another major contribution of the paper is the modeling of time-varying covariance matrices in large-scale financial time series of log-returns, where the above-mentioned parsimonious prior structure will play a major regularizing role by shrinking unnecessary (or flat-line) coefficients toward zero (or toward constants). More specifically, we will rewrite the time-varying covariances Σ_t of the multivariate normal log-returns via a Cholesky transformation $\Sigma_t = A_t H_t A_t'$ and, in turn, model the recursive conditional regression coefficients in the lower-triangular matrix A_t and the log conditional variances from the diagonal matrix H_t, both with the above state-space AR(1) structure and mixture prior. Section 3 provides extensive details regarding this Cholesky stochastic volatility (CSV) structure along with a customized MCMC scheme for posterior Bayesian inference that takes advantage of the parallel nature of the CSV model.

We illustrate our approach by a number of real and synthetic examples, including a real application on the estimation of log-volatilities in a state-space model based on realized volatilities and a CSV model with 94 of the 100 financial time series comprised in the S&P100 index.

2 Prior Specification for the State Equation

To facilitate discussion, we begin with the univariate state-space model

$$\begin{array}{ll} \text{Observation equation:} & y_t = f(x_t, s_t, \eta_t) \\ \text{State equation:} & s_t = \alpha + \beta\, s_{t-1} + \tau \varepsilon_t, \end{array}$$

where s_t is the latent (hidden) state-space variable. η_t and ε_t are independent random shocks in the observation and state equations respectively, usually Gaussian, and we observe the pair (x_t, y_t). In our examples, we will consider two specifications for the observation equation: *i)* $y_t = x_t\, s_t + \eta_t$, a dynamic regression with a time-varying coefficient s_t, and *ii)* $y_t = \exp(s_t/2)\, \eta_t$, a standard stochastic volatility model.

Our goal is to specify a prior on (α, β, τ) that allows us to investigate different kinds of state evolutions within the simple AR(1) framework. In addition, we will specify a prior for s_0, the initial state.

In this paper, the posterior will be computed using the Gibbs sampler:

$$[(s_0, s) \mid (\alpha, \beta, \tau), y, x] \quad \text{and} \quad [(\alpha, \beta, \tau) \mid (s_0, s)],$$

where $y = (y_1, y_2, \ldots, y_T)$, $x = (x_1, x_2, \ldots, x_T)$, and $s = (s_1, s_2, \ldots, s_T)$. That is, we draw the states given the AR(1) parameters and the AR(1) parameters given the states. The likelihood for the draw of the AR(1) parameters is that of a simple linear regression. In developing a flexible non-conjugate prior for (α, β, τ), our computational scheme will take advantage of the simple normal regression likelihood.

2.1 A Mixture Prior for AR Parameters

In this section we present a mixture prior for (α, β, τ). The basic notions our prior must be able to express are *i)* we may want τ small, and *ii)* the following four cases are of particular interest:

Case (1): $(\alpha, \beta) = (0, 1)$ - (random walk component)

Case (2): $\beta = 0$ - (flat-lined component)

Case (3): $(\alpha, \beta) = (0, 0)$ - (irrelevant state-space component)

Case (4): $0 < \beta < 1$ - (stationary component).

Our prior mixes over these four cases. We put zero prior weight on $\beta < 0$.

Case (1) corresponds to the classic "random-walk" prior. With τ small, this prior succinctly expresses the notion that the state evolves smoothly and may "wander". Many applications assume Case (1). Case (2) says the state simply varies about a fixed level α. With very small τ this is practically equivalent to a fixed value for the state. Case (3) says that the state is fixed near zero, which is often a possibility of particular interest. For example, if the state is a regression coefficient then the corresponding variable has no effect. Case (4) allows the state to vary in a stationary fashion.

Note that in some cases the posterior is largely identified by the prior. A near constant state can be achieved with $(\alpha, \beta) = (0, 1)$ (Case (1)) or $(\alpha, \beta) = (\alpha, 0)$ (Case (2)), given τ small, and the data does not care how you do it. Depending on the application, the user may choose to weight different mixture components. For example, if we are only extrapolating a few periods ahead, $\beta \approx 1$ may be fine. If, however, we are brave enough to predict farther ahead, we may be more comfortable going with $\beta < 1$, if the data allows it. As usual, the mixture representation allows us to push the inference in desired directions, without imposing it.

In Section 3 we consider the problem of modeling high dimensional multivariate stochastic volatility. This large, complex model consists of thousands of univariate state-space models.

In this application we found it essential to be able to flexibly consider the possibility that many of the states are constant over time. This gives us a kind of adaptive parsimony in that a constant state greatly simplifies the model.

We have chosen not to consider the case $\tau = 0$. There is no practical advantage in considering τ to be zero as opposed to small. The key is to be able to draw $(\alpha, \beta, \tau) \,|\, (s_0, s)$ jointly and let our prior beliefs about τ and α depend on β.

In order to make a joint draw and have a minimally restricted choice of prior specification, we put β and τ on a bivariate grid. We then restrict ourselves to a normal density for $p(\alpha \,|\, \beta, \tau)$. Given this normal prior and normal likelihood, we can integrate out α analytically to obtain $p(\beta, \tau \,|\, s_0, s)$ which is used to draw from the bivariate grid. Given a draw of (β, τ), we are in a conditionally conjugate situation so that the draw of $\alpha \,|\, \beta, \tau, s_0, s$ is just a normal draw.

To specify a prior for τ on a grid of values, we first choose minimum and maximum values τ_{min} and τ_{max}. Using n grid points, we have evenly spaced values $(t_1, t_2, \ldots, t_n)$ with $t_1 = \tau_{min}$ and $t_n = \tau_{max}$. We let $P(\tau = \tau_{min}) \equiv p_{min}$. For $i > 1$, $P(\tau = t_i) \propto \exp(-c_\tau \, |t_i - \tau_{min}|)$. Thus, our τ prior has the four hyper-parameters $(\tau_{min}, \tau_{max}, p_{min}, c_\tau)$. This prior is very simple. We pick an interval, and then our choice of c_τ determines the degree to which we push τ towards smaller values. The commonly used prior for τ is the inverted chi-squared: $\tau^2 \sim \nu \, \lambda / \chi^2_\nu$. We found it very difficult to choose values for ν and λ that gave consistently good results. Both the left and right tails of the inverted chi-squared distribution are difficult to manage.

To specify a prior for $\beta \in (0, 1)$, on a grid of points $(b_1, b_2, \ldots, b_n)$, we let $p(\beta = b_i) \propto n(b_i \,|\, \bar{\beta}, \sigma^2_\beta)$, where $n(\cdot \,|\, \bar{\beta}, \sigma^2_\beta)$ denotes a normal density with mean $\bar{\beta}$ and standard deviation σ_β.

Therefore, our full mixture prior has the form

$$\begin{aligned}
p(\alpha, \beta, \tau) &= p_{01} \, p(\tau \,|\, \beta = 1) \, \delta_{\{\alpha=0, \beta=1\}} \\
&+ p_{00} \, p(\tau \,|\, \beta = 0) \, \delta_{\{\alpha=0, \beta=0\}} \\
&+ p_{u0} \, p(\tau \,|\, \beta = 0) \, p(\alpha \,|\, \beta = 0, \tau) \, \delta_{\{\beta=0\}} \\
&+ p_{uu} \, p(\beta) \, p(\tau \,|\, \beta \neq 0) \, p(\alpha \,|\, \beta),
\end{aligned}$$

where p_{01}, p_{00}, p_{u0}, and p_{uu} are the mixture weights of our four components. The notation δ_x represents the distribution such that x happens for sure. p_{01} is the probability that $(\alpha, \beta) = (0, 1)$, p_{00} is the probability that $(\alpha, \beta) = (0, 0)$, p_{u0} is the probability that $\beta = 0$ and α is unrestricted, and p_{uu} is the probability that $\beta \in (0, 1)$ and α is unrestricted. $p(\tau \,|\, \beta)$

denotes a discrete distribution on a grid as discussed above. We allow for the possibility that the choice of the parameters $(\tau_{min}, \tau_{max}, p_{min}, c)$ could depend on β. We may want to suggest smaller values of τ when $\beta = 0$. For example, in all our applications, the choice of c_τ given $\beta = 0$ is twice the value used for non-zero β.

As discussed previously, our computational approach constrains us to choose a normal distribution for $\alpha \,|\, \beta, \tau$. However, we are free to let the parameters of normal depend on β and τ in any way. We use,

$$\alpha \,|\, \beta \sim N(0, \sigma_\alpha^2\,(1 - \beta^2)).$$

When $\beta = 0$ we simply have $\alpha \sim N(0, \sigma_\alpha^2)$. As β increases, we shrink our prior down towards the case where $\alpha = 0$ at $\beta = 1$.

2.2 The Initial State

In many applications, zero is a value of particular importance for the state because it represents a model simplification. An important example is that of a time varying regression coefficient. To shrink the initial state s_0 towards zero we use a mixture prior along the lines of that used by George & McCulloch (1992) for variable selection:

$$\begin{aligned} s_0 &\sim \gamma\, N(0, (cw)^2) + (1 - \gamma)\, N(0, w^2) \\ \gamma &\sim \text{Bernoulli}(p_*), \end{aligned}$$

where c is a large positive real number, w is small, and p_* is a hyper-parameter denoting the prior knowledge about the initial state. A small p_* favors zero initial state and $p_* = 0.5$ shows no preference. The variable γ is a latent variable. When $\gamma = 0$, the state is shrunk heavily towards zero and when $\gamma = 1$, the state may be large. The basic Gibbs sampler is augmented with the draw

$$[\gamma \,|\, (s_0, s), (\alpha, \beta, \tau)] \equiv [\gamma \,|\, s_0].$$

Conditional on a draw of γ, we have a normal prior for the initial state with mean zero and standard deviation w in the case $\gamma = 0$ and standard deviation (cw) in the case $\gamma = 1$. This facilitates our draws of the states.

2.3 Simulated Mixture Prior Examples

In this section we illustrate our prior on (α, β, τ) in the simple normal dynamic linear model (NDLM),

$$
\begin{aligned}
y_t &= x_t\, s_t + \sigma\, \eta_t, \\
s_t &= \alpha + \beta\, s_{t-1} + \tau\, \varepsilon_t,
\end{aligned}
$$

where η_t and ε_t are independent and identically distributed $N(0, 1)$. We simulate series of length $T = 200$ with $\sigma = 0.1$, $s_0 = 0$, and $x_t \sim N(0, 9)$. The state is a time-varying regression coefficient. As discussed in Sections 2.1 and 2.2, the posterior is computed using the Gibbs sampler with full conditionals given by *i)* $[(s_0, s) \mid (\alpha, \beta, \tau), \gamma, y, x]$, *ii)* $[(\alpha, \beta, \tau) \mid (s_0, s), \gamma]$, and *iii)* $[\gamma \mid s_0]$.

In all examples, we use $(p_*, w, c) = (0.5, 0.1, 10)$ for the γ prior, $(\bar{\beta}, \sigma_\beta) = (1.0, 1.0)$ for the β prior, and $\sigma_\alpha = 2.0$ for the α prior. To specify $p(\tau \mid \beta)$ we let $\tau_{max} = 0.15$ and $p_{min} = 0.5$ for all β. If $\beta > 0$, $\tau_{min} = 0.005$ and $c_\tau = 100$ while if $\beta = 0$, $\tau_{min} = 0.001$ and $c_\tau = 200$. The prior thus suggests smaller τ when $\beta = 0$ which effectively leads to a constant state around α. The mixture weights are give by $p_{01} = 0.50$, $p_{00} = 0.15$, $p_{u0} = 0.15$, $p_{uu} = 0.20$. We have a 50% prior probability of the random walk prior and 30% chance of a constant state. There is a 20% that we have a time-varying stationary state.

Figure 1 displays the prior. The top two panels are density smooths of prior draws of β and τ, respectively. The density smooths naturally "jitter" the draws (add a bit of normal noise) so that our mixture of discrete and continuous distributions can be displayed as a single continuous distribution. The marginal for β displays our preference for expressing a smooth state with either $\beta \approx 0$ or $\beta \approx 1$ with more weight being given to the vicinity of 1. The prior for τ expresses our desire for small values. Again, this is driven by our desire for a smooth state. The two τ modes reflect the choice of a smaller τ_{min} when $\beta = 0$. In this case the two modes are not very separated so this aspect of our prior has little practical effect. If we separated these modes more dramatically, we could use this aspect for our prior to help identify $\beta \approx 0$ versus $\beta \approx 1$ by saying you can only have a really small τ if $\beta \approx 0$. The long right tail of our τ prior allows the data to push the inference towards larger values if needed.

The bottom two panels of Figure 1 display the joint prior of (α, β). The bottom left panel displays contours from a bivariate smooth of draws of (α, β). The bottom right panel is a scatterplot of jittered draws of (α, β). In the bivariate distribution we can see our preference for $(\alpha, \beta) \approx (0, 1)$ or $(\alpha, \beta) \approx (0, 0)$ with more weight given to the first pair of values. As

β decreases, the conditional prior for α becomes more spread out. The contours appear to tighten as β approaches 0 because the choice $(\bar{\beta}, \sigma_\beta) = (1, 1)$ puts weight on larger β.

Figure 1 about here.

Figure 2 displays the results from three different data simulations. Each row corresponds to a different simulation scenario. In the first row results are for data simulated with $(\alpha, \beta, \tau) = (0, 1, 0.04)$, in the second row we used $(\alpha, \beta, \tau) = (0, 0.8, 0.1)$, and in the third row we used $(\alpha, \beta, \tau) = (0.5, 0, 0.01)$. Thus, in the first row our state follows a random walk, in the second row the state is time varying but stationary, while in the third row the state is essentially constant at 0.5.

The first column of plots in Figure 2 shows the simulated states (small circles) and the posterior mean of the state draws (with a solid line). In each row we see that the posterior mean nicely smooths the true states and that the essential nature of the state is quite different reflecting our three scenarios. The second, third, and fourth columns of Figure 2 display time series plots of MCMC draws of α, β, and τ respectively. In each plot a dashed horizontal line indicates the true value of the parameter.

In each case the posterior nicely captures the true value. In some cases, our mixture prior has some interesting shrinkage effects. In the case of the first scenario (row 1), virtually all of our posterior draws have $\alpha = 0$ and $\beta = 1$. Thus, while our prior probability of the random walk case was 0.5, our posterior probability is very close to one. For the second data set our posterior for β mixes between the two components $\beta = 1$ and $\beta \in (0, 1)$. The posterior probabilities of the components are 0.55 and 0.45 compared to prior probabilities 0.5 and 0.2. Given the data we are not sure whether β is one or not, but we know it is not zero. In the third scenario our posterior mixes between $\beta = 0, \alpha \approx 0.5$ and $\beta \in (0, 1)$. The posterior probability that $\beta = 0$ and $\alpha \neq 0$ is 0.52 compared with the prior probability of 0.15. The posterior for τ also reflects our prior. In both the second and third scenarios our prior pushes the posterior towards small τ but still covers the true values.

Figure 2 about here.

2.4 Realized Volatility with the Mixture Prior

In this section we apply our mixture prior to an analysis of the log series of daily realized volatility of Alcoa stock from January 2, 2003 to March 7, 2004 for 340 observations. The realized volatility is calculated using the intraday 10-minute log returns. It is well known that realized volatility is subject to the impact of market micro-structure noises. See, for instance, Bandi & Russell (2008) and Zhang, Mykland & Ait-Sahalia (2005). Consequently, we entertain the model

$$y_t = s_t + \sigma\eta_t \quad \text{and} \quad s_t = \alpha + \beta\, s_{t-1} + \tau\varepsilon_t,$$

for the log series of Alcoa realized volatility. Here the shock $\sigma\eta_t$ denotes the impact of market micro-structure noises and the state s_t is simply the level of the log volatility. The AR(1) state equation with error term $\tau\varepsilon_t$ allows for volatility to be time-varying.

Our first step in the analysis is to standardize the data to have zero mean and standard deviation one. Given this standardization, we can use the same prior as in Section 2.3. Figure 3 is the time series plot of the standardized data and the posterior mean of the state. The states smooth the series. In this particular case the posterior probability of the mixture component with $\beta = 1$ is 0.99. The prior probability was .5. Thus, we have strong support for the random walk state specification, which is in good agreement with empirical characteristics of asset volatility. For example, consider the VIX index of the Chicago Board Options Exchange (CBOE), which is the most widely used daily volatility index in the U.S. The log VIX series fails to reject the null hypothesis of a unit root in hypothesis testing. This analysis of realized volatility demonstrates that the mixture prior of Section 2 can easily produce reasonable results.

Figure 3 about here.

3 Cholesky Stochastic Volatility

In this section we show the impact of our mixture prior in a much larger set up where thousands of state variables might evolve over time according to an AR(1) process. More specifically, we are interested in the case where $y_t = (y_{1t}, \ldots, y_{qt})'$ denotes a q-dimensional vector of financial time series observed at time t and consider posterior inference regarding the (possibly large) covariance matrices Σ_t driven by the observation equation:

$$y_t|F_{t-1} \sim N(0, \Sigma_t), \tag{1}$$

where F_{t-1} denotes the information available at time $t-1$. Without loss of generality, we assume that any mean structure of y_t has been subtracted out as part of a larger MCMC algorithm.

The main focus is on modeling the dynamic behavior of the conditional covariance matrix Σ_t, which is known as the volatility matrix in finance. Two challenges arise in the multivariate context. Firstly, the number of distinct elements of Σ_t equals $q(q+1)/2$. This quadratic growth has made the modeling Σ_t computationally very expensive and, consequently has created, up to a few years ago, a practical upper bound for q. The vast majority of the papers available in the literature employed a small q or use highly parametrized models to simplify the computation. For instance, Engle (2002) and Tse & Tsui (2002) proposed dynamic conditional correlation (DCC) models where the time evolution of correlations is essentially driven by a pair of parameters. We argue that that such models unrealistically over-simplify the complexity of the covariance dynamics. Secondly, the distinct elements of Σ_t cannot be modeled independently since positive definiteness has to be satisfied. Section 3.1 briefly reviews the literature on multivariate stochastic volatility models, while Section 3.2 introduces our proposed Cholesky stochastic volatility (CSV) model.

3.1 Brief Literature Review

There are at least three ways to decompose the covariance matrix Σ_t.

Correlations and Standard Deviations. In the first case, the covariance matrix is decomposed as

$$\Sigma_t = D_t R_t D_t$$

where D_t is a diagonal matrix consisting of the standard deviations, i.e. $D_t = \text{diag}(\sigma_{1t}, \dots, \sigma_{qt})$ with σ_{it} being the volatility of y_{it}, and R_t is the correlation matrix. The above two challenges remain in this parametrization, i.e. the number of parameters increases with q^2 and R_t has to be positive definite.

Factor Analysis. In the second case, a standard factor model is used to produce

$$\Sigma_t = \beta_t H_t \beta_t' + \Psi_t$$

where β_t is the $q \times k$ matrix of factor loadings and is block lower triangular with diagonal elements equal to one. Ψ_t and H_t are the diagonal covariance matrices of the specific factors and common factors, respectively. This is the *factor stochastic volatility* (FSV) model of Harvey, Ruiz & Shephard (1994), Pitt & Shephard (1999), Aguilar & West (2000), and, more recently, Lopes & Migon (2002), Chib, Nardari & Shephard (2006), Han (2006), Lopes & Carvalho (2007) and Philipov & Glickman (2006*a*), to name just a few. Philipov & Glickman (2006*a*) extended the FSV model by allowing H_t to follow a Wishart random process and fit a 2-factor FSV model to the covariance of the returns of $q = 88$ S&P500 companies. Han (2006) fitted a similar FSV model to $q = 36$ CRSP stocks. Chib et al. (2006) analyzed $q = 10$ international weekly stock index returns (see also Nardari & Scruggs, 2007). Lopes & Carvalho (2007) extended the FSV model to allow for Markovian regime shifts in the dynamic of the variance of the common factors and apply their model to study $q = 5$ Latin America stock indexes.

A Cholesky Approach. In this paper we take a third alternative that decomposes Σ_t via a Cholesky decomposition as

$$\Sigma_t = A_t H_t A_t'$$

where $A_t H_t^{1/2}$ is the lower triangular Cholesky decomposition of Σ_t. H_t is a diagonal matrix, the diagonal elements of A_t are all equal to one and, more importantly, the lower diagonal elements of A_t are unrestricted since positive definiteness is guaranteed. In the next section we show that there will be $q(q+1)/2$ dynamic linear models to be estimated and $3q(q+1)/2$ static parameters. When $q = 30$, for example, there are 465 latent states at each time and 1395 static parameters.

The basic advantage of the Cholesky approach over the Factor approach is that no restrictions are imposed on the structure of the Σ_t. The prior developed in above Section 2 coupled with the compuational approach developed below enables us to search for simplifying structure in a large system without imposing it.

3.2 Time-Varying Triangular Regressions

In this section we lay out our basic parametrization of the time-varying covariance structure in terms of linear regressions. Recall that $y_t \sim N(0, \Sigma_t)$ and $\Sigma_t = A_t H_t A_t'$ where $A_t H_t^{1/2}$ is the lower triangular Cholesky decomposition of Σ_t. The matrix A_t is lower triangular with ones in the main diagonal and $H_t = \text{diag}(\omega_{1t}^2, \ldots, \omega_{qt}^2)$. Therefore,

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

Let the (i,j)th element of the lower triangular matrix A_t^{-1} be $-\phi_{ij}$ for $i > j$, while the diagonal (i,i) element is one. It follows that the joint normal distribution for y_t given F_{t-1}, that is $N(0, \Sigma_t)$, can be rewritten as a set of q recursive conditional regressions where

$$y_{1t} \sim N(0, \omega_{1t}^2) \tag{2}$$

and, for $i = 2, \ldots, q$,

$$y_{it} \sim N(\phi_{i1t} y_{1t} + \phi_{i2t} y_{2t} + \cdots + \phi_{i(i-1)t} y_{(i-1)t}, \omega_{it}^2). \tag{3}$$

Once ϕ_{ijt}s and ω_{it}^2s are available, so are A_t^{-1} (and A_t), H_t and, consequently, $\Sigma_t = A_t H_t A_t'$. To make Σ_t fully time-varying without any restrictions, we simply make each parameter in the regression representation time-varying. More precisely,

$$\phi_{ijt} \sim N(\alpha_{ij} + \beta_{ij}\, \phi_{ij(t-1)}, \tau_{ij}^2) \tag{4}$$

for $i = 2, \ldots, q$ and $j = 1, \ldots, i-1$, and

$$d_{it} \sim N(\alpha_i + \beta_i\, d_{i(t-1)}, \tau_i^2) \tag{5}$$

for $d_{it} = \log(\omega_{it}^2)$ and $i = 1, \ldots, q$, where τ_{ij}^2 and τ_i^2 are hyper-parameters. It is understood that the aforementioned distributions are all conditional on the available information F_{t-1}.

The actual parameters we work with are the ϕ_{ijt}s and d_{it}s. These parameters are our state variables in the state equations (4) and (5), while the recursive conditional regressions (or simply *triangular regressions*) are our observation in the observation equations (2) and (3). Our *Cholesky stochastic volatility* (CSV) model comprises equations (2) to (5).

The Cholesky decomposition approach has been studied elsewhere. Uhlig (1997) and Philipov & Glickman (2006*b*), for example, proposed models for the covariance matrix based on the temporal update of the parameters of a Wishart distribution (see also Asai & McAleer, 2009). Uhlig (1997) models $\Sigma_t^{-1} = B_{t-1}^{-1} \Theta_{t-1} (B_{t-1}^{-1})' \nu / (\nu + 1)$, where $\Theta_{t-1} \sim Beta\,((\nu + pq)/2, 1/2)$, $B_t = A_t H_t^{1/2}$ and *Beta* denotes the multivariate Beta distribution (Uhlig 1994). See also Triantafyllopoulos (2008) for a similar derivation in the context of multivariate dynamic linear models. Philipov & Glickman (2006*b*) model $\Sigma_t^{-1} \sim W(\nu, S_{t-1}^{-1})$, where $S_{t-1}^{-1} =$

$\frac{1}{\nu}(C^{1/2})(\Sigma_{t-1}^{-1})^d(C^{1/2})'$, such that $E(\Sigma_t|\Sigma_{t-1},\theta) = \nu(C^{-1/2})(\Sigma_{t-1})^d(C^{-1/2})'/(\nu - q - 1)$. The parameter d controls the persistence in the conditional variance process. A constant covariance model arises when $d = 0$, so $E(\Sigma_t) = \nu C^{-1}/(\nu-q-1)$ and C plays the role of a precision matrix. When $d = 1$ and $C = I_q$, it follows that $E(\Sigma_t) = \Sigma_{t-1}$ so generating random walk evolution for the conditional covariance. See Dellaportas & Pourahmadi (2011) for a similar model for time-invariant A and H_t following GARCH-type dynamics. Uhlig (1997) models daily/current prices per tonne of aluminum, copper, lead and zinc, i.e. $q = 4$, exchanged in the London Metal Exchange. Philipov & Glickman (2006*b*) fit their model to returns data on $p = 5$ industry portfolios. Dellaportas & Pourahmadi (2011) model exchange rates of the US dollar against $q = 7$ other country/regions.

A thorough review of the multivariate stochastic volatility literature up to a few years is provided in Asai, McAleer & Yu (2006) and Lopes & Polson (2010). See also Bauwens, Hafner & Laurent (2012).

3.3 Posterior Inference

We detail here the Markov chain Monte Carlo algorithm for posterior computation of our CSV model introduced above. Let q denote the number of series and T denote the number of observations on each time series. Let $Y_i = \{y_{it}\}_{t=1}^T$ and $d_i = \{d_{it}\}_{t=1}^T$, $i = 1, 2, \ldots, q$. Let $\phi_{ij} = \{\phi_{ijt}\}_{t=1}^T$, $i = 2, 3, \ldots, q$, $j = 1, 2, \ldots, (i-1)$. That is, Y_i is the time series of observations on the i^{th} variable, d_i is the time-varying state corresponding to the residual variance of the regression of y_{it} on $y_{jt},\ j < i$, and ϕ_{ij} is the time-varying state corresponding to the regression coefficient of y_{it} on y_{jt}. See Equation (3). Let d_{i0} and ϕ_{ij0} denote initial states.

With $p(\cdot)$ denoting a generic probability density function, the full joint distribution of everything we need to think about is then given by the product of the following four hierarchical terms:

i. Likelihood function: $\prod_{i=2}^q p(Y_i \mid Y_1, \ldots, Y_{i-1}, d_i, \phi_{i1}, \ldots, \phi_{i(i-1)}) \times p(Y_1|d_1)$,

ii. (d, ϕ) *states:* $\prod_{i=1}^q p(d_i \mid \alpha_i, \beta_i, \tau_i, d_{i0}) \prod_{j<i} p(\phi_{ij} \mid \alpha_{ij}, \beta_{ij}, \tau_{ij}, \phi_{ij0})$,

iii. AR parameters: $\prod_{i=1}^q p(\alpha_i, \beta_i, \tau_i) \prod_{j<i} p(\alpha_{ij}, \beta_{ij}, \tau_{ij})$, and

iv. Initial states: $\prod_{i=1}^q p(d_{i0}) \prod_{j<i} p(\phi_{ij0})$,

where it is understood that $\prod_{j<i} = 1$ when $i = 1$. The terms $p(\alpha, \beta, \tau)$ and $p(s_0)$ (with i and j dropped and s standing for d or ϕ) denote our prior on the parameters of the autoregressive specification of the state evolution and our prior on the initial state, respectively. The choice of this prior is a key component of our approach and is discussed in Section 2.

Our Markov chain Monte Carlo is a (large-scale) Gibbs sampler where we (efficiently) draw from the following full conditional distributions (with $\circ$ denoting "everything else"):

i. d states: $(d_{i0}, d_i) \mid \circ$,

ii. ϕ states: $(\phi_{ij0}, \phi_{ij}) \mid \circ$,

iii. d AR parameters: $(\alpha_i, \beta_i, \tau_i) \mid \circ$, and

iv. ϕ AR parameters: $(\alpha_{ij}, \beta_{ij}, \tau_{ij}) \mid \circ$.

The key property in this potentially large system is that, in the conditionals above, the states and parameters for a given equation are independent of the states and parameters of the other equations. This is readily seen in the structure of the full joint distributions given above. Thus, to draw d_i, we simply compute $\widetilde{y}_{it} = y_{it} - \sum_{j<i} \phi_{ijt}\, y_{jt}$ and use standard methods developed for univariate stochastic volatility given the model:

$$\begin{aligned} \widetilde{y}_{it} &\sim N(0, \exp\{d_{it}/2\}), \\ d_{it} &\sim N(\alpha_i + \beta_i\, d_{i(t-1)}, \tau_i^2). \end{aligned}$$

Similarly, the draw of ϕ_{ij} reduces to the analysis of a basic dynamic linear model (DLM) for $\widetilde{y}_{ijt} = y_{it} - \sum_{k<i, k\neq j} \phi_{ikt}\, y_{kt}$:

$$\begin{aligned} \widetilde{y}_{ijt} &\sim N(\phi_{ijt}\, y_{jt}, \exp\{d_{it}/2\}), \\ \phi_{ijt} &\sim N(\alpha_{ij} + \beta_{ij}\, \phi_{ij(t-1)}, \tau_{ij}^2). \end{aligned}$$

The draws of the AR parameter also reduce to consideration of a single state,

$$(\alpha_i, \beta_i, \tau_i) \mid \circ \equiv (\alpha_i, \beta_i, \tau_i) \mid (d_{i0}, d_i),$$

$$(\alpha_{ij}, \beta_{ij}, \tau_{ij}) \mid \circ \equiv (\alpha_{ij}, \beta_{ij}, \tau_{ij}) \mid (\phi_{ij0}, \phi_{ij}).$$

Thus, all the ϕ_{ij} draws reduce to simple applications of FFBS and all of the d_i draws reduce to those of the univariate stochastic volatility model. We use the method of Kim, Shephard and Chib (1998), again based on FFBS, for the univariate stochastic volatility model.

In order to keep the entire system manageable for large q, we use a univariate DLM for each ϕ in each equation rather than running a multivariate FFBS to jointly draw all the ϕ series for a given equation. This approach avoids a great many high-dimensional matrix operations. Potentially, this could put dependence into our chain depending upon the application. This does not seem to be a severe problem in our examples.

Thus, the whole thing boils down to repeated applications of the basic Gibbs sampler that cycles through $(s_0, s) \,|\, (\alpha, \beta, \tau)$ and $(\alpha, \beta, \tau) \,|\, (s_0, s)$, where s denotes a state series and s_0 the initial state. Since we need to put a strong prior on (α, β, τ) there is unavoidable dependence in the basic chain. Because of this dependence, we have found it useful to draw (α, β, τ) jointly as discussed in Section 2.1.

3.4 Parallel Processing

One of the strengths of the proposed CSV framework is that the triangular representation of the model naturally leads to parallelization in the MCMC scheme. More specifically, the $T \times i$-dimensional state-space matrix $(d_i, \phi_{i1}, \ldots, \phi_{i,i-1})$ and the $3 \times i$-dimensional parameter matrix $(\alpha_i, \beta_i, \tau_i, \alpha_{i1}, \beta_{i1}, \tau_{i1}, \ldots, \alpha_{i,i-1}, \beta_{i,i-1}, \tau_{i,i-1})$ corresponding to the i-th recursive conditional regression can be drawn independently from the other recursive conditional regressions.

However, it is well known that sampling d_i (log-volatilities) is more computationally expensive (more time consuming) than sampling ϕ_{ij}. In fact, for a small to moderate i, it is likely that the computational burden is due to d_i almost exclusively. Let c_d, c_ϕ and c_θ be the computational cost (in seconds, for instance) to draw the T-dimensional vectors d_i and ϕ_{ij} and the 3-dimensional vectors $\theta_i = (\alpha_i, \beta_i, \tau_i)$, for any i and j (see full conditional distributions in Section 3.3). Usually c_θ is negligible when compared to c_d and c_ϕ. The cost to draw the states from recursive conditional regression i is $c_i = c_d + (i-1)c_\phi + ic_\theta$, and the total cost is

$$c = \kappa_1(q)c_d + \kappa_2(q)c_\phi + \kappa_3(q)c_\theta$$

where $\kappa_1(q) = q$, $\kappa_2(q) = q(q-1)/2$ and $\kappa_3(q) = q(q+1)/2$. Similarly, the total cost of running regressions $i_a + 1$ to i_b ($i_b - i_a$ regressions) is

$$c_{i_a:i_b} = \Delta\kappa_1^{ab} c_d + \Delta\kappa_2^{ab} c_\phi + \Delta\kappa_3^{ab} c_\theta$$

where $\Delta\kappa_j^{ab} = \kappa_j(i_b) - \kappa_j(i_a)$, for $j = 1, 2, 3$. Assume that computation can be split between two parallel processors. Due to the imbalance between (mainly) c_d and c_ϕ (and c_θ), it is not immediately obvious which recursive conditional regression i_1 will make $c_{1:i_1} = c_{(i_1+1):q} = c/2$.

Similarly, what are the optimal i_1 and i_2 when three processors are available? In general, for m processors, the goal is to find the cut-offs $(i_1, i_2, \ldots, i_{m-1})$ such that the cost within each group of recursive conditional regressions is the same:

$$c_{1:i_1} = c_{(i_1+1):i_2} = \cdots = c_{(i_{m-2}+1):i_{m-1}} = c_{(i_{m-1}+1):q} = c/m.$$

The search for the cut-offs is performed recursively with i_1 selected from $\{1, \ldots, q\}$ such that $c_{1:i_1} < c/m$ and $c_{1:(i_1+1)} > c/m$, i_2 selected from $\{i_1 + 1, \ldots, q\}$ such that $c_{1:i_2} < 2c/m$ and $c_{1:(i_2+1)} > 2c/m$, and so forth.

Figure 4 provides an illustration when there are $q = 100$ time series and up to $m = 20$ processors. The costs $(c_d, c_\phi, c_\theta) = (30, 2, 0)$ are based on actual run times (in seconds) for $T = 2516$ time points and 10,000 MCMC draws. It takes 15 times longer to draw d_i than it does to draw ϕ_{ij}. These costs were based on our code running in a 2.93 GHz Intel Core 2 Duo processor. For $m = 1$ processor, the total cost is 215 minutes. For $m = 2$ processors, $i_1 = 67$ and the cost per processor is 107 minutes. For $m = 3$ processors, $(i_1, i_2) = (52, 79)$ and the cost per processor is 71 minutes. For $m = 4$ processors, $(i_1, i_2, i_3) = (44, 67, 84)$ and cost per processor is 53 minutes. For $m = 20$ processors, cost per processor is about 11 minutes.

Figure 4 about here.

3.5 A Simulated Example

To gain insight into the proposed analysis, in this section we present results from a simple simulated example. We let $q = 3$, Σ_0 be the identity and Σ_1 be the covariance matrix corresponding to standard deviations of $\sigma_1 = 4$, $\sigma_2 = 1$, $\sigma_3 = 0.25$, and correlations $\rho_{12} = \rho_{23} = 0.9$, $\rho_{13} = 0.81$. We then let $\Sigma_t = (1 - w_t)\,\Sigma_0 + w_t\,\Sigma_1$ where w_t increases from 0 to 1 as t goes from 1 to T. At each t we draw $y_t \sim N(0, \Sigma_t)$. We simulate $T = 500$ tri-variate observations.

Results obtained using the prior given in Section 2.3 are shown in Figure 5. We also present results obtained from a second prior having

$$p_{01} = 0.85, \quad p_{00} = 0.05, \quad p_{u0} = 0.05, \quad p_{uu} = 0.05, \quad \tau_{max} = 0.05, \quad c_\tau = \{200, 400\}, \tag{6}$$

where $c_\tau = 200$ when $\beta > 0$ and $c_\tau = 400$ when $\beta = 0$. Results obtained using the second prior in Equation (6) are shown in Figure 6. The second prior puts more weight on the random walk component and favors smaller τ by decreasing τ_{max} and increasing c_τ. The same (α, β, τ) prior was used for each of the six state series (three d state series and three ϕ state series).

The time-varying standard deviations are plotted in the diagonal panels and the (i, j) correlation is plotted in the (i, j) lower panel below the diagonal. The very smooth line is the true value, the lighter smooth line is the posterior mean of the parameter (σ_{it} or ρ_{ijt}) and the dashed line is the estimate obtained by computing the standard sample covariance of a moving window of observations. The moving window for estimation of Σ_t, includes all available observations within 50 time periods of t.

In both cases the posterior mean seems to nicely smooth the evidence from the data, with the tighter second prior giving smoother results.

Figures 5 and 6 about here.

3.6 Many Assets From the S&P100 Index

In this section we use asset returns from firms making up the S&P100 index in order to illustrate the procedure with large q. We first consider a selection of returns on $q = 20$ of the firms and use the second prior from Section 3.5 given in Equation (6). We use this prior because with larger q, more smoothing may be desireable.

Figure 7 plots the posterior means of the σ_{it} and ρ_{ijt} series. The top panel shows the 20 standard deviations series and the bottom panel shows the $20(19)/2 = 190$ correlations series. There is no simple way to plot so much information, but even with the many series, we can see that there is substantial time variation in both the standard deviations and the correlations. From the $\{\sigma_{it}\}$ series we see that there is an overall pattern to their behavior over time. For example, the volatility is generally lower at the end of the time period. However, there is substantial variation across assets (across i) both in the overall level of volatility and the amount of time variation. Similarly, there are time periods where ρ_{ijt} is relatively large for most (i, j) pairs but some pairs behave quite differently from the rest.

Figure 7 about here.

Figure 8 plots the posterior means of the states with the d states in the top panel and the ϕ states in the bottom panel. The top panel shows the time variation in the residual variances which vary markedly over time. The bottom panel shows that most of the ϕ series have relatively little time variation and are centered near zero. However, a few of the ϕ_{ijt} series do vary substantially over time. This figure shows how our Bayesian model, with our particular prior choice, seeks a parsimonious representation of a very high-dimensional problem.

Of course, the amount of "parsimony", "smoothness", or "regularization" inevitably is heavily influenced by our choice of prior. Figure 9 shows the posterior means of the states obtained when we use yet another prior given by

$$p_{01} = 0.85, \quad p_{00} = 0.05, \quad p_{u0} = 0.05, \quad p_{uu} = 0.05, \quad \tau_{max} = 0.02, \quad c_\tau = \{300, 600\} \tag{7}$$

where again the two values for c_τ correspond to $\beta > 0$ and $\beta = 0$. This figure looks like a smoothed version of Figure 8. The "flat-line" appearance of many of the ϕ states is striking. The corresponding standard-deviation-correlation plot is, again, a smoothed version of Figure 7.

Figures 8 and 9 about here.

Figures 10 and 11 plot the posterior means of the portfolio weights for the global minimum variance portfolio using the prior in Equation (6). We see, from Figure 10, that the time variation in the standard deviations and correlations may be of real practical importance in that the corresponding portfolio weights change over time substantially. Figure 11 shows that at the final time period, T, most of the weights are negligible, while being positive for about half a dozen assets.

Figures 10 and 11 about here.

Figure 12 plots two different estimates of a time-varying standard deviation. Again we used the $q = 20$ series but in in the second run we reversed the order of the series. We are plotting the posterior means of σ_{1t}, the standard deviation of the first series in the original order, and $\tilde{\sigma}_{qt}$, the standard deviation of the *same series* in the model where it is last in our Cholesky decomposition. These two fits result from different priors, so there is no reason that they be identical. However, their similarity is striking. The Cholesky approach requires us to pick an order for the time series. In some cases, this may be a convenient way to express prior information. For example, if one series represented returns on the market (or some "factor") we may want to put it early in the list. In some applications, a natural ordering may not be apparent. Figure 12 suggests that we do not have to worry too much about getting the "right" order.

Figure 12 about here.

Figure 13 reports results for $q = 94$ assets using the prior in Equation (7). Six of the return series from the 100 companies have missing values, due to inclusion and exclusion to the index, leaving us with 94. In this case there are 94 standard deviation series (σ_{it}) and $94(93)/2 = 4{,}371$ correlation pairs (ρ_{ijt}) so it becomes quite difficult to present the results. The top panel displays results for the σ_{it} while the bottom panel displays the ρ_{ijt}. The two panels have the same format. The solid gray band gives pointwise quartiles for the posterior means. Thus, in the top panel, the gray band is the middle 50% of the 94 standard deviation posterior means $\hat{\sigma}_{it}$ for each fixed t and in the bottom panel it is the middle 50% of the 4,371 correlation estimates for each fixed t. The thick solid (black) lines give 95% intervals. We can see that with 94 series we observe the same overall patterns we saw with $q = 20$.

We also randomly picked 20 of the $\{\hat{\sigma}_{it}\}$ series to plot in the top panel and 20 of the $\{\hat{\rho}_{ijt}\}$ series to plot in the bottom panel. These plots, along with the size of the 95% intervals, indicate the while there is an overall pattern over time, there are substantial differences amongst the $\{\hat{\sigma}_{it}\}$ across i (assets) and the $\{\hat{\rho}_{ijt}\}$ across (i, j) (pairs of assets).

Figure 13 about here.

4 Conclusion

Clearly, applying large-scale state-space models in general and modeling time-varying covariance matrices in particular are very difficult when the dimension q is large. In most approaches to volatility modeling available in the literature, the set of possible Σ_t is highly restricted. In this paper we present a method which works for large q without restricting the support of the prior for the Σ_t. The key aspects of our approach are (i) a very general and flexible specification for the prior on the parameters, and, (ii) restricting our model and prior so the computations may be done in parallel. The prior allows us to "softly-restrict" or "regularize" our inference. Although our prior has positive support on any $\{\Sigma_t\}$ sequence, we can put in the prior belief that they do not change too quickly. A key insight of this paper is that we may also want to put in the prior belief that many of the states describing how one series relates to another (the ϕ) are essentially constant over time. Even for relatively small q, we find the prior a useful tool for exploring the possible dynamic structure.

References

Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and variance matrix discounting for portfolio allocation', *Journal of Business and Economic Statistics* **18**, 338–357.

Asai, M. & McAleer, M. (2009), 'The structure of dynamic correlations in multivariate stochastic volatility models', *Journal of Econometrics* **150**, 182–192.

Asai, M., McAleer, M. & Yu, J. (2006), 'Multivariate stochastic volatility: a review', *Econometric Reviews* **25**, 145–175.

Bandi, F. M. & Russell, J. R. (2008), 'Microstructure noise, realized variance, and optimal sampling', *Review of Economic Studies* **75**, 339–369.

Bauwens, L., Hafner, C. & Laurent, S. . (2012), Volatility models, *in* L. Bauwens, C. Hafner & S. Laurent, eds, 'Handbook of Volatility Models and Their Applications', John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 1–45.

Chib, S., Nardari, F. & Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**, 341–371.

Dellaportas, P. & Pourahmadi, M. (2011), 'Cholesky-GARCH models with applications to finance (to appear)', *Statistics and Computing* **19**, 237–271.

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.

George, E. I. & McCulloch, R. E. (1992), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **79**, 677–83.

Han, Y. (2006), 'Asset allocation with a high dimensional latent factor stochastic volatility model', *The Review of Financial Studies* **19**, 237–271.

Harvey, A. C., Ruiz, E. & Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

Lopes, H. F. & Carvalho, C. M. (2007), 'Factor stochastic volatility with time varying loadings and markov switching regimes', *Journal of Statistical Planning and Inference* **137**, 3082–3091.

Lopes, H. F. & Migon, H. S. (2002), 'Comovements and contagion in emergent markets: stock indexes volatilities', *Case Studies in Bayesian Statistics* **6**, 285–300.

Lopes, H. F. & Polson, N. G. (2010), Bayesian inference for stochastic volatility modeling, *in* K. Bocker, ed., 'Rethinking Risk Measurement and Reporting: Uncertainty, Bayesian Analysis and Expert Judgement', RiskBooks, pp. 515–551.

Migon, H. S., Gamerman, D., Lopes, H. F. & Ferreira, M. A. R. (2005), Bayesian inference for stochastic volatility modeling, *in* D. Dey & C. R. Rao, eds, 'Handbook of Statistics: Bayesian Thinking, Modeling and Computation', Vol. 25, Elsevier, pp. 553–588.

Nardari, F. & Scruggs, J. T. (2007), 'Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and apt pricing restrictions', *Journal of Financial and Quantitative Analysis* **42**, 857–892.

Philipov, A. & Glickman, M. E. (2006*a*), 'Factor multivariate stochastic volatility via wishart processes', *Econometric Reviews* **25**, 311–334.

Philipov, A. & Glickman, M. E. (2006*b*), 'Multivariate stochastic volatility via wishart processes', *Journal of Business and Economic Statistics* **24**, 313–328.

Pitt, M. & Shephard, N. (1999), Time varying covariances: A factor stochastic volatility approach, *in* J. B. *et al*, ed., 'Bayesian statistics 6', London: Oxford University Press.

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

Triantafyllopoulos, K. (2008), 'Multivariate stochastic volatility with bayesian dynamic linear models', *Journal of Statistical Planning and Inference* **138**, 1021–1037.

Tse, Y. K. & Tsui, A. K. C. (2002), 'A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business and Economic Statistics* **20**, 351–362.

Uhlig, H. (1994), 'On singular wishart and singular multivariate beta distributions', *The Annals of Statistics* **22**, 395–405.

Uhlig, H. (1997), 'Bayesian vector autoregressions with stochastic volatility', *Econometrica* **65**, 59–73.

Zhang, L., Mykland, P. A. & Ait-Sahalia, Y. (2005), 'A tale of two time scales: determining integrated volatility with noisy high-frequency data', *Journal of the American Statistical Association* **100**, 1394–1411.

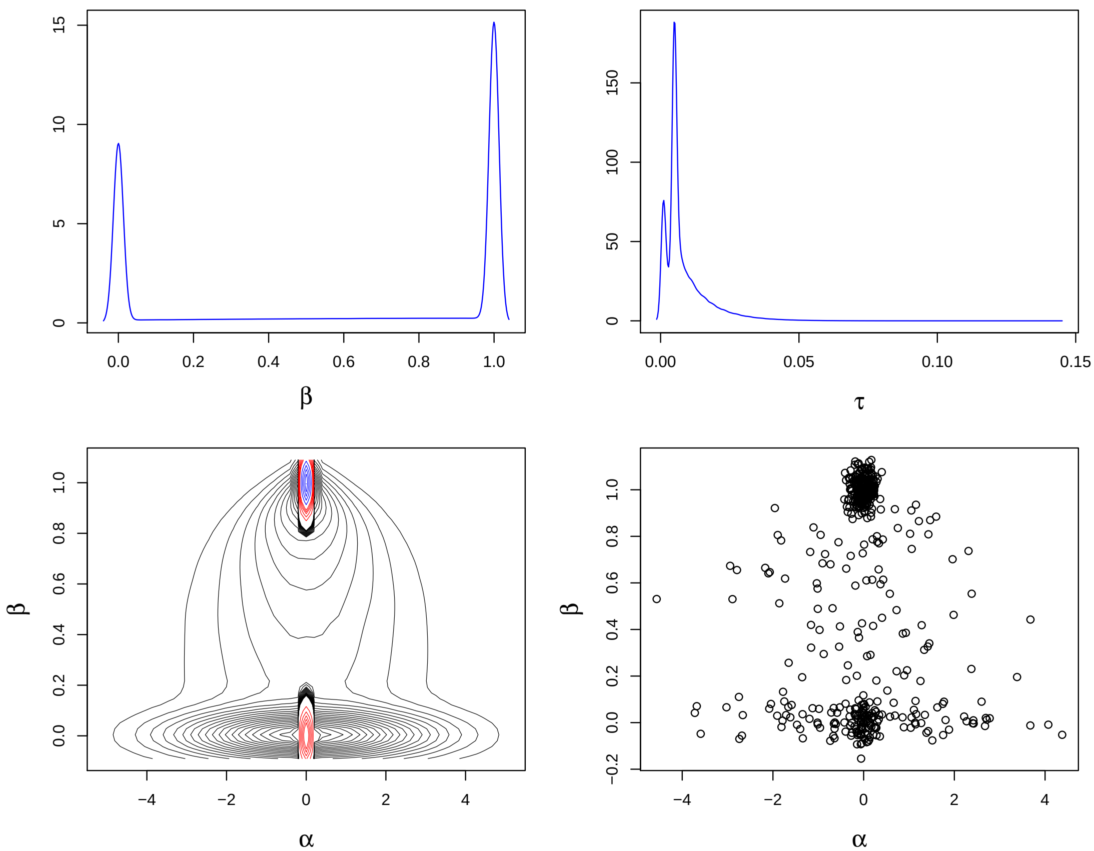

Figure 1: Prior used in Section 2.3: The top two panels are density smooths of prior draws of β and τ. The bottom left panel displays contours from a bivariate smooth of draws of (α, β). The bottom right panel are jittered draws of (α, β).

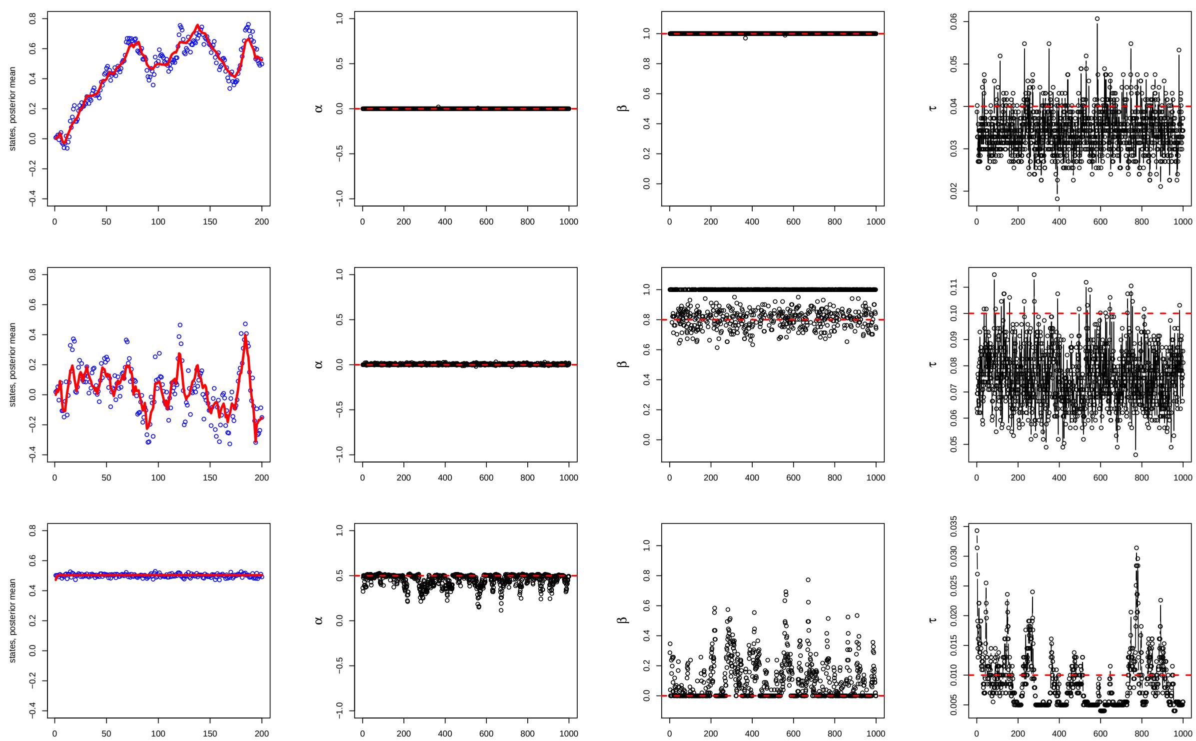

Figure 2: Each row shows the results for a different simulation. The plot in the first column shows the simulated states (the plotted points) and the posterior mean (the solid curve). The second, third, and fourth columns of plots show time-series of MCMC draws from the posteriors of α, β, and τ respectively. Dashed horizontal lines represent the true values of (α, β, τ).

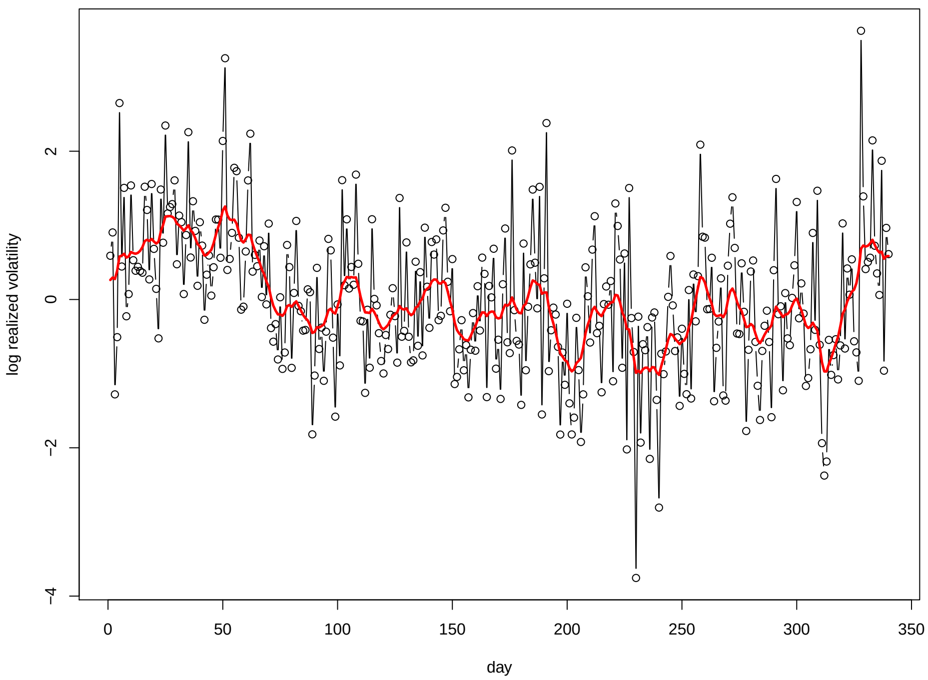

Figure 3: Log realized volatility for Alcoa stock (standardized), with the posterior mean of the state.

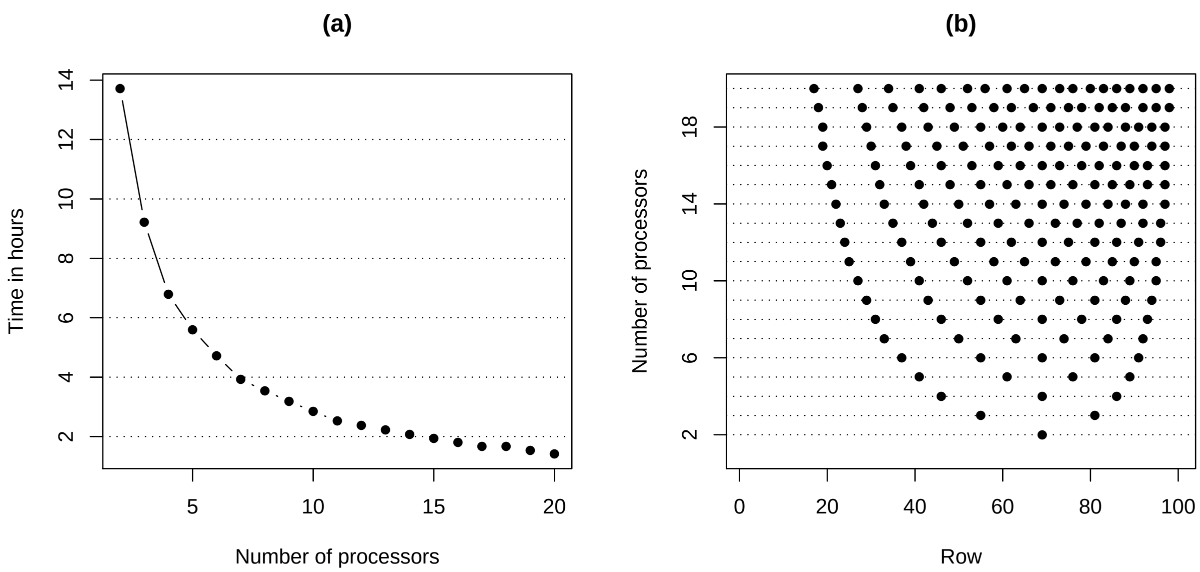

Figure 4: *Multiple processors*: In panel (a) we plot the number of processors vs. the total time in minutes to run 50,000 iterations for a 100×100 ($q = 100$) time varying covariance matrix with $T = 2,516$. It takes 15 times longer to draw a d state than it does to draw a ϕ state. Code was run on a 2.93 GHz Intel Core 2 Duo processor. With 20 processors, the time is about 11 minutes. In panel (b) we have the number of processors on the vertical axis and each set of points along the dotted lines indicate how the 100 conditional regressions in the Cholesky decomposition are allocated to the different processors. For example, when $m = 2$ the cut-off is regression $i_1 = 67$, i.e. the first processor runs regressions 1 to 67 while the second processor runs regressions 68 to 100.

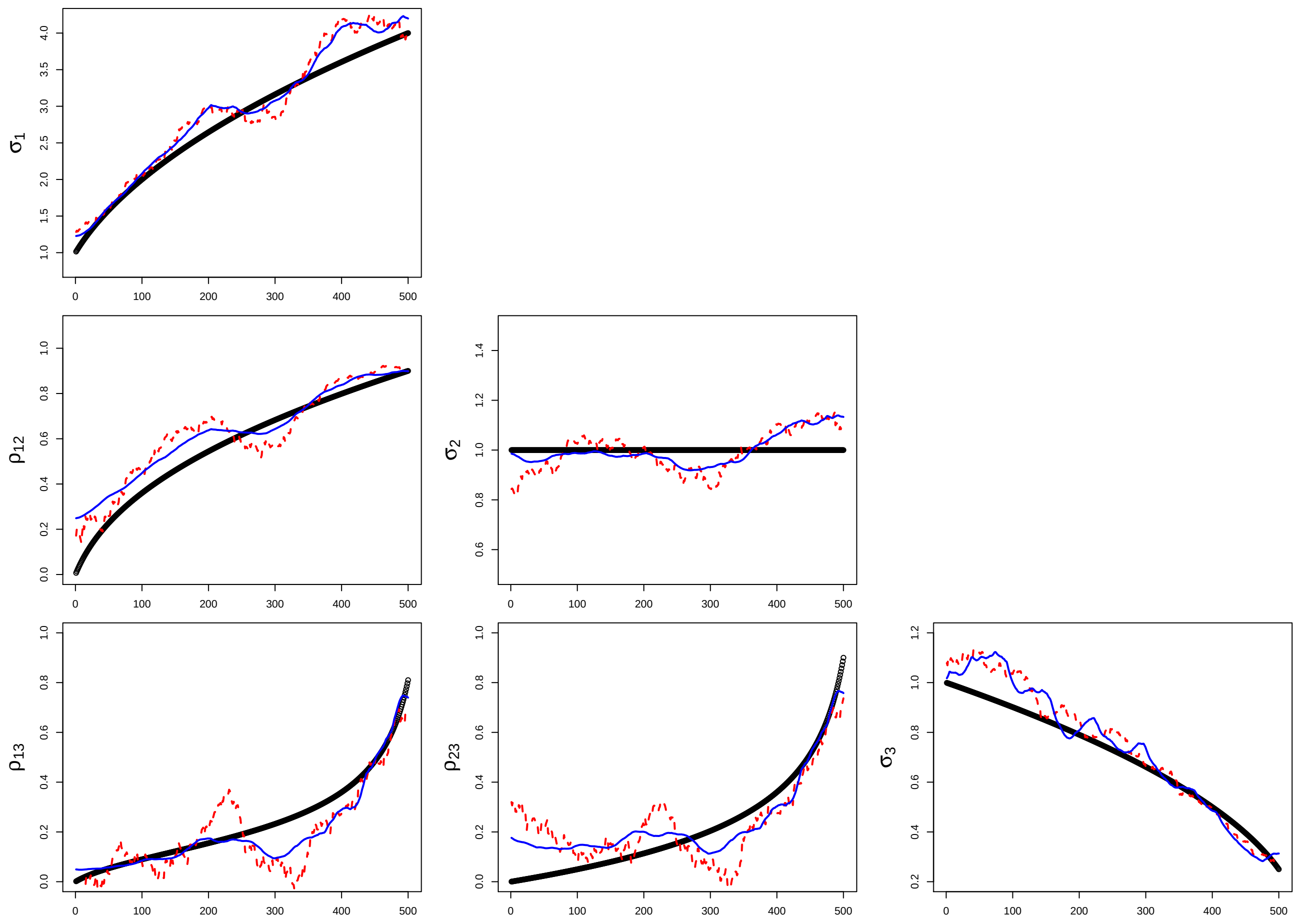

Figure 5: *Simulated example with prior given in Section 2.3*: Time-varying standard deviations are plotted in the diagonal panels and correlations are plotted in the lower panel below the diagonal. *True values:* very smooth black curves. *Posterior means:* lighter smooth blue curves. *Moving window estimates:* red dashed curves.

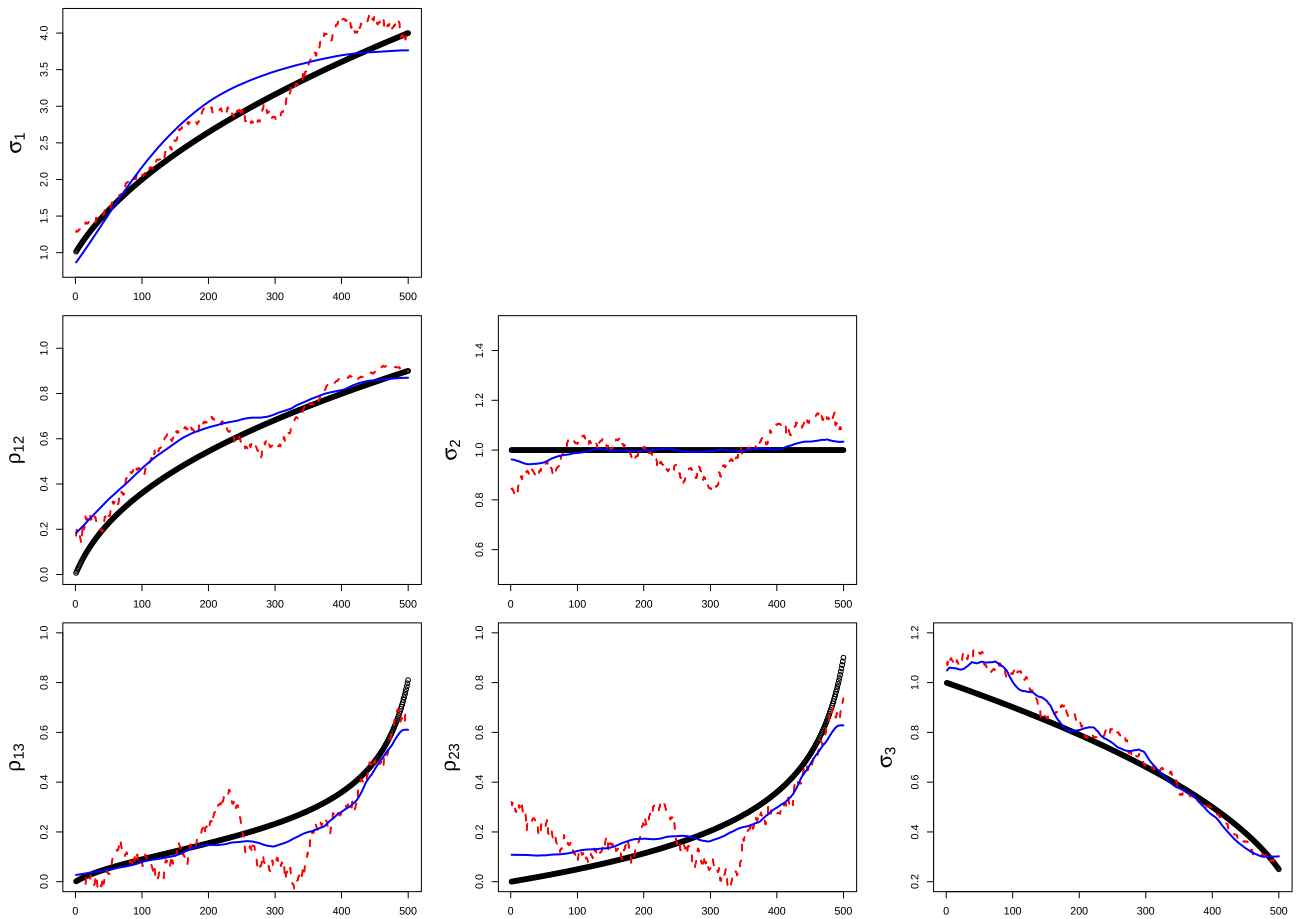

Figure 6: *Simulated example with prior in Equation (6)*: Time-varying standard deviations are plotted in the diagonal panels and correlations are plotted in the lower panel below the diagonal. *True values:* very smooth black curves. *Posterior means:* lighter smooth blue curves. *Moving window estimates:* red dashed curves.

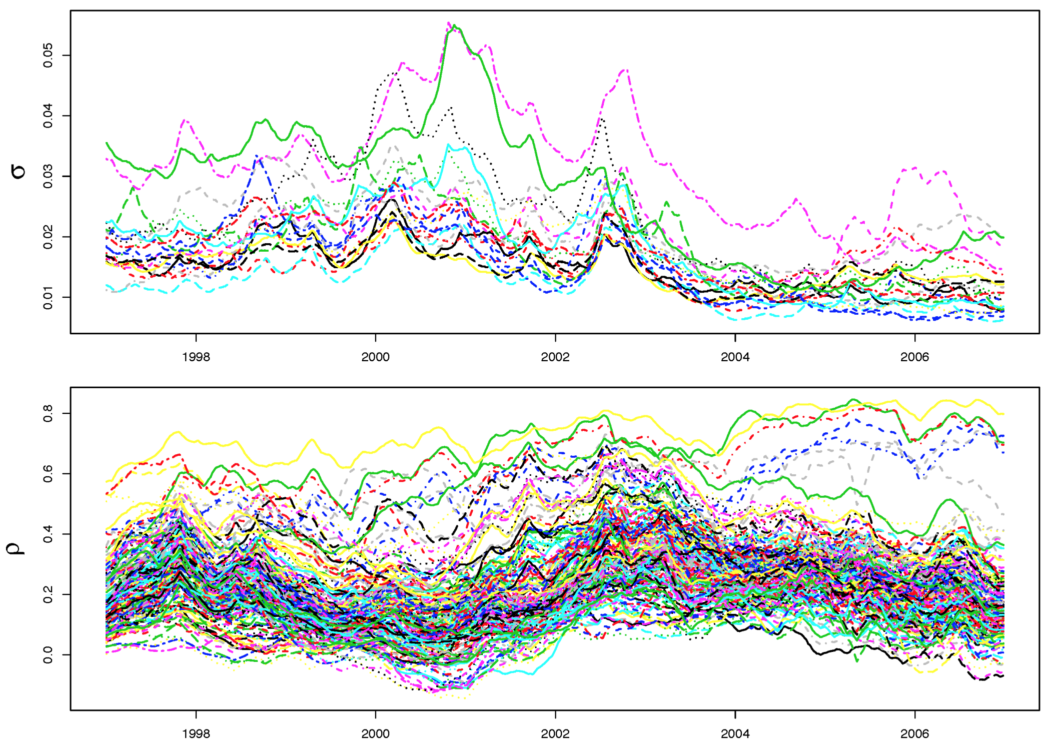

Figure 7: *S&P100 data, q = 20, prior in Equation (6)*: Posterior means of time-varying standard deviations and correlations.

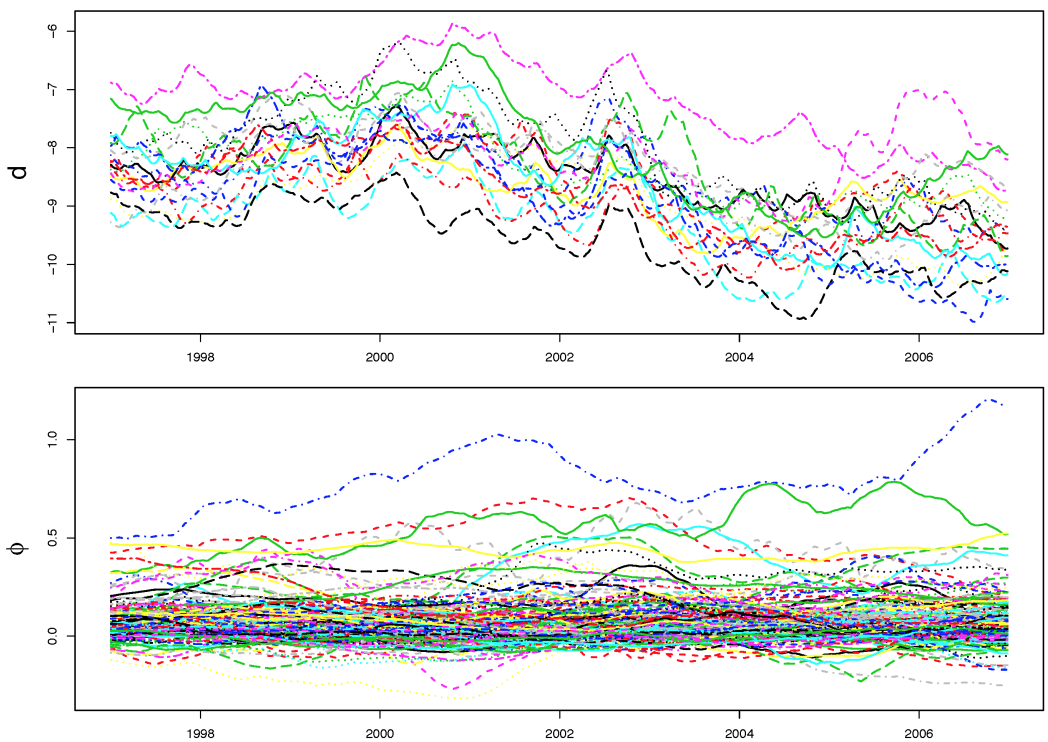

Figure 8: *S&P100 data, $q = 20$, prior in Equation (6)*: Posterior means of d and ϕ states.

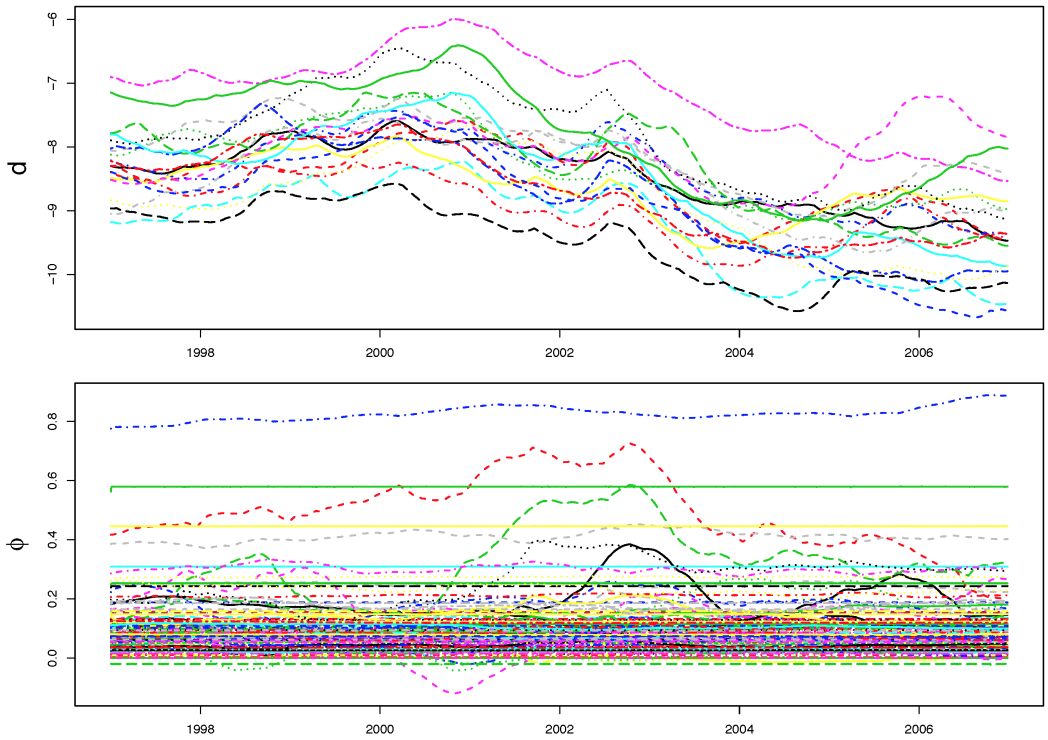

Figure 9: *S&P100 data, $q = 20$, prior in Equation (7)*: Posterior means of d and ϕ states. $q = 20$.

Figure 10: *S&P100 data, $q = 20$, prior in Equation (6)*: Posterior means of portfolio weights for the global minimum variance portfolio.

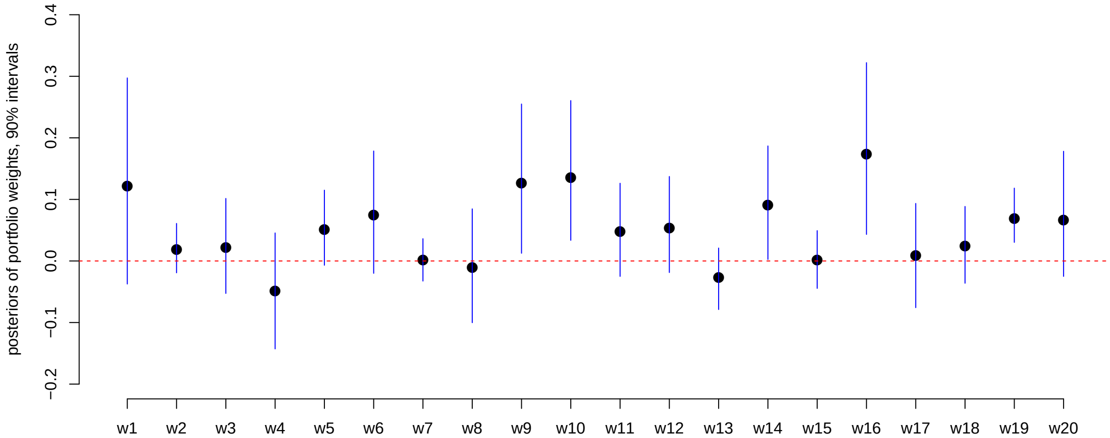

Figure 11: *S&P100 data, $q = 20$, prior in Equation (6)*: Posterior 5th, 50th and 95th percentiles of portfolio weights for the global minimum variance portfolio at the final time period T.

Figure 12: *S&P100 data, $q = 20$, prior in Equation (6)*: Comparison of the posterior means of the time-varying standard deviation of a series with the original order of the series and the order reversed.

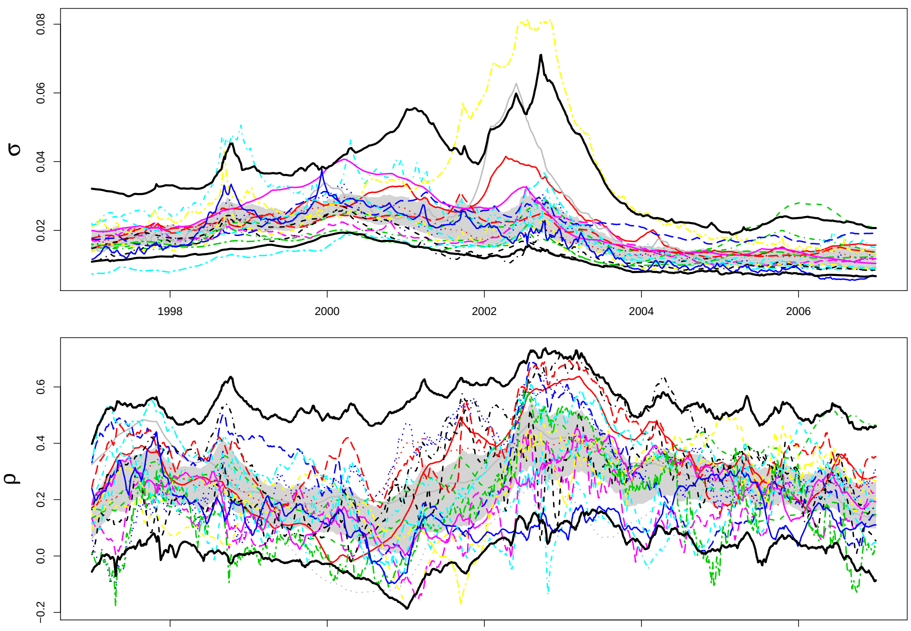

Figure 13: *S&P100 data, $q = 94$, prior in Equation (7)*: Top panel displays $\hat{\sigma}_{it}$. Bottom panel displays $\hat{\rho}_{ijt}$.