Factor Multivariate Stochastic Volatility Via Wishart Processes

Alexander Philipov
Federal National Mortgage Association

Mark E. Glickman
Department of Health Services
Boston University School of Public Health

December 27, 2004

Abstract

This paper proposes a high dimensional factor multivariate stochastic volatility (SVOL) model in which factor covariance matrices are driven by Wishart random processes. The framework allows for unrestricted specification of intertemporal sensitivities, which can capture the persistence in volatilities, kurtosis in returns, as well as correlation breakdowns and contagion effects in volatilities. The factor structure allows addressing high dimensional setups used in portfolio analysis and risk management, as well as modeling conditional means and conditional variances within the model framework. Due to the complexity of the model, we perform inference using Markov chain Monte Carlo simulation from the posterior distribution. A simulation study is carried out to demonstrate the efficiency of the estimation algorithm. We illustrate our model on a data set that includes 88 individual equity returns and the two Fama-French size and value factors. With this application, we demonstrate the ability of the model to address high dimensional applications suitable for asset allocation, risk management and asset pricing.
1 Introduction

Most financial applications focus on the risk and return of financial portfolios and are concerned with the correlated movements of their assets. Recent advances in the financial literature have focused on the time-variation of asset covariances and correlations. Efforts to improve covariance forecasts have led to the development of time-varying volatility models, such as multivariate GARCH and Stochastic Volatility (SVOL). Multivariate models of time-varying volatility, however, have two major drawbacks that have prevented them from enjoying the success of their univariate counterparts. First, multivariate models can be highly parameterized, resulting in potential interpretation difficulties. Second, multivariate models bring a significant increase in the computational complexity of model fitting, in part due to complex model structures, and in part due to the increased number of parameters.

The literature on multivariate volatility models is extensive. Multivariate ARCH models were introduced by Engle and Kraft (1983) and further developed by Bollerslev, Engle, and Wooldridge (1988) (BEW), and by Engle and Kroner (1995) with their very popular BEKK model. Imposing an explicit factor structure in the multivariate ARCH models was first suggested by Engle, Ng, and Rothschild (1990). Factor structure finds motivation in asset pricing theory. It explains the co-movements of asset returns and reduces the dimensionality of the multivariate models. One most characteristic restriction of ARCH models is that all variation in the covariances is driven by return shocks.

Allowing the covariances to have their own stochastic component produces the multivariate stochastic volatility models studied by Harvey, Ruiz, and Shephard (1994), Mahieu and Schotman (1994), Jacquier, Polson, and Rossi (1995), etc. Multivariate SVOL with factor structure have been proposed by Aguilar and West (2000), Chib, Nardari, and Shephard (2002), and Han (2002). The common feature of these models is that the covariance matrix of the factors is assumed diagonal and the factor variances each follow their independent stochastic process. When the factors are not observable, model estimation requires specific restrictions (Geweke and Zhou (1996), Aguilar and West (2000)). Earlier factor SVOL specifications are extended in Aguilar and West (2000) by adding a time-varying vector of means. This model deals effectively with issues of overparameterization as well as of identifying the latent factors using the approach suggested by Geweke and Zhou (1996).

This paper proposes a factor multivariate SVOL model that assumes that the factor volatilities follow an unconstrained Wishart random process. Our model has close ties to Philipov and Glickman (2004), which proposes a class of multivariate SVOL models where the volatilities, unconditional on any factors, follows a Wishart process. Our model offers several advantages over existing factor multivariate models. First, the evolution of factor covariances is described by a general matrix-variate Wishart process. This modeling approach allows for interaction among the factors, their variances, and covariances, thus giving more support to the empirical evidence on observed priced factors. Second, our factor model can be seen as a special case of the more general multivariate SVOL model of Philipov and Glickman (2004). As such, it offers a natural framework for testing the asset-pricing restrictions imposed by factor-based models. We restrict our attention to observable factors, as this case allows
the factor covariances to follow an unconstrained Wishart process. We discuss some of
the difficulties incorporating unobserved factors later in the paper. Common variation in
security returns is assumed to be driven by a set of observable factors whose covariance
matrix follows a stochastic Wishart process. This factor multivariate SVOL model avoids a
number of restrictions common to other classes of multivariate SVOL models, allowing for
increased flexibility in modeling returns data.

We adopt a Bayesian framework for model development, and assume inference is carried
out using Markov chain Monte Carlo (MCMC) simulation from the posterior distribution.
Recent examples of MCMC simulation for SVOL models include Philipov and Glickman
(2004), Aguilar and West (2000), and Jacquier, Polson, and Rossi (1995). Parameter sum-
maries and forecasts are obtained using the empirical distribution of the samples obtained
from MCMC posterior simulation.

The remainder of this paper is organized as follows. In Section 2 we introduce the
factor multivariate stochastic volatility model and discuss the properties and interpreta-
tion of the model. This is followed in Section 3 by the derivation of the posterior distribu-
tion. In Section 4 we outline the steps to setting up MCMC posterior simulation. In Section 5,
we examine the performance of the MCMC algorithm on simulated returns data, and apply
our method to a data example involving 88 individual equity returns and two Fama-French
size and value factors. Finally, in Section 6 we discuss the utility of our approach, and
limitations and extensions of the basic model presented.

2 The Factor Model

Suppose that we have \( n \) asset returns, \( y_t \), and \( k \) factors, \( f_t \). We assume that the asset returns
follow a multivariate normal distribution, where

\[
\begin{align*}
   y_t \mid B, f_t, \Omega & \sim N(B f_t, \Omega) \\
   f_t \mid V_t & \sim N(0, V_t) \\
   V_t^{-1} \mid V_{t-1}, A, \nu, d & \sim \text{Wish}(\nu, S_{t-1})
\end{align*}
\]

where

\[
S_{t-1} = \frac{1}{\nu} A^{1/2} (V_{t-1}^{-1})^d A^{1/2},
\]

is the scale parameter for the Wishart process. In our model specification, \( V_t \) is a matrix
of factor volatilities, \( A \) is a symmetric positive definite matrix, \( d \) is a scalar persistence
parameter, \( \nu \) is the degrees of freedom parameter for the Wishart distribution of the factor
volatilities, \( B \) is a \( n \times k \) matrix of factor sensitivities, and \( \Omega \) is a \( n \times n \) matrix of idiosyncratic
risks for the assets. One feature of our model is that the conditional expectation of \( V_t^{-1} \) is
given by

\[
E(V_t^{-1}) = \nu S_{t-1} = \left( A^{1/2} \right) (V_{t-1}^{-1})^d \left( A^{1/2} \right)'.
\]

An interesting case of our model results when \( d = 1 \) and \( A \) is the identity matrix, in which
case the model for factor volatilities is a Wishart random walk. In general, we can regard the
factor model as a special case of the model proposed by Philipov and Glickman (2004), which imposes no restrictions on the covariance matrix. This model imposes specific restrictions on the covariance matrix of returns, namely it separates unconditional volatility, \( \Sigma_t \), into systematic and idiosyncratic components,

\[
\Sigma_t = B V_t B' + \Omega.
\] 

(4)

Interest often centers on making inferences for the \( \Sigma_t \), so it is useful to separate the components of variation in the volatility matrix.

The parameter \( A \) can be interpreted as a measure of intertemporal sensitivity. This matrix parameter reveals how each element of the current period covariance matrix depends on elements of the previous period covariance matrix. It is the parameter that in large part determines mean reversion characteristics on a multivariate level. The scalar parameter \( d \) denotes the overall strength of the intertemporal relationships. This parameter accounts for the presence of long memory or persistence, a phenomenon described for univariate series as today’s return having a large effect on the forecast variance many periods in the future (see Engle and Patton (2001)). The parameter \( d \) is theoretically bound between -1 and 1 (see Philipov and Glickman (2004)) to ensure stationarity of the covariance process. For practical purposes, values of \( d \) that are applicable to financial data are between 0 and 1, as we would rule out stochastic processes on the \( V_t \) which alternate between powers of inverses. A value of \( d \) close to zero would indicate a weak overall effect of current volatility on future values, i.e. shocks in returns are “forgotten” fast – within a few subsequent periods. Under this model specification, we do not need to impose constraints to ensure positive definiteness (as in early multivariate models) as this property is inherent in the quadratic form in (3) and in the Wishart process.

3 Bayesian analysis of factor SVOL model

Inference for our model is based on a Bayesian analysis that leads to the posterior distribution of model parameters. The posterior density is proportional to the product of the prior density and the likelihood function. We first write out the likelihood function, and then suggest an appropriate noninformative prior distribution, after which we derive the posterior density.

The likelihood function is proportional to the product of the joint density of returns \( y_t \) conditional on the factor sensitivities \( B \) and the idiosyncratic volatility \( \Omega \), and the Wishart
density of the factor volatility matrix, $V_t$.

$$L(Y \mid B, \Omega) = \prod_{t=1}^{T} \text{Wishart}_k(V_{t}^{-1} \mid \nu, S_{t-1})$$

$$\times \int \ldots \int N_n(y_t \mid Bf_t, \Omega) N_k(f_t \mid 0, V_t) df_t$$

$$\propto \prod_{t=1}^{T} (V_{t-1}^{-1})^\nu |V_t^{-1}|^{\nu-k-1} \exp \left( -\frac{\nu}{2} tr[A^{-1} (V_{t-1}^{-1})^{-d} V_t^{-1}] \right)$$

$$\times \prod_{t=1}^{T} \int \ldots \int \frac{1}{|\Omega|^{\nu/2}} \exp \left( -\frac{1}{2} (y_t - Bf_t)'\Omega^{-1}(y_t - Bf_t) \right)$$

$$\times \frac{1}{|V_t|^{1/2}} \exp \left( -\frac{1}{2} f'_t V_t^{-1} f_t \right) df_t$$

where $Y = (y_1, \ldots, y_T)$ is the collection of $T$ asset returns across all time periods, and the functions $\text{Wishart}(\cdot, \cdot)$ and $N(\cdot, \cdot)$ are Wishart and normal densities of the given arguments, assuming given parameter values.

We construct a noninformative prior distribution for the parameter set $(\Omega, B, A, d, \nu)$ that factors into independent densities. We examine the prior component for each parameter separately.

First, following [Gelman, Carlin, Stern, and Rubin (1995, p.408) and Box and Tiao (1992, p.425)], we assume

$$p(B) \propto \text{constant}. \quad (7)$$

Extending a univariate application of Jeffrey’s rule for the unknown volatility parameter into a multivariate case, [Box and Tiao (1992)] suggest a prior for $\Omega$ that is proportional to the information matrix,

$$p(\Omega) \propto |I(\Omega)|^{\frac{1}{2}} \quad (8)$$

where

$$|I(\Omega)| = |I(\Omega^{-1})| \left| \frac{\partial \Omega}{\partial \Omega^{-1}} \right|^{-2}$$

It is shown in [Box and Tiao (1992)] that

$$|I(\Omega^{-1})| \propto \left| \frac{\partial \Omega}{\partial \Omega^{-1}} \right| = |\Omega|^{n+1} \quad (9)$$

Therefore, the suggested prior density for $\Omega$ is proportional to the inverse of the determinant:

$$p(\Omega) \propto |\Omega|^{-\frac{1}{2}(n+1)} \quad (10)$$

For a more tractable analysis, we express the prior density for $A$ in terms of $A^{-1}$. A Wishart density for $A^{-1}$ is assumed with a $k \times k$ scale matrix $Q_0$ which is a positive definite
symmetric matrix, and can be decomposed as $Q_0 = (Q_0^{1/2} \left( Q_0^{1/2}\right)^{'}$. The value assigned to $Q_0$ will depend on expectations about the data. We suggest that $A^{-1}$ is centered on the identity matrix ($Q_0 = I$), which, by equation (3), results in the prior mean of $V_t^{-1}$ equal to the previous period’s covariance matrix, $(V_{t-1}^{-1})^d$. We also suggest a degrees of freedom parameter $\gamma_0 = k + 1$.

For the parameter $d$, a uniform density such as $p(d) = U_{[0,1]}(d)$ may be used. For $\nu$ a locally noninformative gamma density seems appropriate. Because $\nu$ must be greater than $k$, the dimension of the covariance matrix, the gamma density is shifted by $k$. Therefore, it would be appropriate to specify that $(\nu - k)$ has a gamma density.

The joint prior density for the parameters of the model would be the product of the densities specified above,

$$p(\Omega, A^{-1}, d, \nu) = p(B)p(\Omega) \times \text{ Wishart}(\gamma_0, Q_0) \times p(d) \times \text{Gamma}(\nu - k) \quad (11)$$

By Bayes theorem, the posterior distribution of the parameter set $(V^{-1}, A^{-1}, B, \Omega, d, \nu)$, is proportional to the product of the prior (11) and the likelihood (5). We suppress the time subscripts in the posterior density, $p(V^{-1}, A, B, \Omega, \nu, d \mid Y)$, to denote generically that the density is defined over the collection of all time periods.

Letting $F = (f_1, ..., f_T)$ represent the collection of $T$ factors across all time periods, the posterior distribution of the parameter array $(V^{-1}, A, B, \Omega, \nu, d)$ is given by

$$p(V^{-1}, A, B, \Omega, \nu, d \mid F, Y) \propto p(A^{-1}, B, d, \nu)L(V^{-1}, A, B, \nu, d \mid Y, F)$$

$$= p(\Omega) \text{ Wishart}(A^{-1} \mid \gamma_0, Q_0)p(d)\text{ Gamma}(\nu - k)$$

$$\times \prod_{t=1}^{T} \text{ Wishart}_{k}(V_t^{-1} \mid \nu, S_{t-1})N_n(y_t \mid Bf_t, \Omega)N_k(f_t \mid 0, V_t) \quad (12)$$

$$\propto \text{ Wishart}(\Omega^{-1} \mid \tau_0, U_0)\text{ Wishart}(A^{-1} \mid \gamma_0, Q_0)p(d) \times \text{ Gamma}(\nu - k)$$

$$\times \prod_{t=1}^{T} \left| (V^{-1}_{t-1})^d \right|^{-\frac{\nu}{2}} \left| V^{-1}_t \right|^{-\frac{(d-k-1)}{2}} \exp \left( -\frac{\nu}{2} tr[A^{-1} (V_{t-1}^{-1})^{-d} V_t^{-1}] \right)$$

$$\times \prod_{t=1}^{T} \frac{1}{|\Omega|^2} \exp \left( -\frac{1}{2} (y_t - Bf_t)\Omega^{-1}(y_t - Bf_t) \right) \frac{1}{|V_t|^2} \exp \left( -\frac{1}{2} f_t'(V_t^{-1}f_t) \right) \quad (13)$$

4 Model Fitting

It is difficult to fit the model in Section 2 using exact methods. Instead, we use MCMC to simulate parameters from the joint posterior distribution. We summarize features of the posterior distribution using the simulated values. We can also forecast future returns using the distribution of simulated parameter values. In order to do so, we create a Gibbs sampler, which consists of drawing simulated values from the conditional posterior densities of the parameters $V^{-1}, A^{-1}, B, \Omega, d, \nu$. For increased efficiency, we draw simulated values from the conditional posterior distributions of $B \text{ and } \Omega$ simultaneously. In this section we
outline the strategy for setting up the MCMC algorithm. Further details of the MCMC simulation appear in the appendix.

4.1 The Conditional Posterior Distribution for \( B \) and \( \Omega \)

To sample draws from the conditional posterior distribution of \( B \) and \( \Omega \), we first note that

\[
p(B, \Omega \mid \Sigma^{-1}, A, \nu, d, Y, F) \propto |\Omega|^{-\frac{1}{2} (n+1)} \prod_{t=1}^{T} \frac{1}{|\Omega|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (y_t - Bf_t)' \Omega^{-1} (y_t - Bf_t) \right) \tag{14}
\]

which can be rewritten as

\[
p(B, \Omega \mid Y, F) \propto |\Omega|^{-\frac{1}{2} (T+n+1)} \exp \left( -\frac{1}{2} tr \left[ \Omega^{-1} G \right] \right) \tag{15}
\]

where \( G = \sum_{t=1}^{T} (y_t - Bf_t)'(y_t - Bf_t) \). Recognizing that sampling the inverse, \( \Omega^{-1} \), is slightly more convenient, we have

\[
p(B, \Omega^{-1} \mid Y, F) = p(B, \Omega \mid Y, F) \left| \frac{\partial \Omega}{\partial \Omega^{-1}} \right| \tag{16}
\]

and using the result in (9), we obtain

\[
p(B, \Omega^{-1} \mid Y, F) \propto |\Omega^{-1}|^{\frac{1}{2} (T-n-1)} \exp \left( -\frac{1}{2} tr \left[ \Omega^{-1} G \right] \right). \tag{17}
\]

Recognizing that

\[
p(B, \Omega^{-1} \mid Y, F) = p(B \mid Y, F)p(\Omega^{-1} \mid B, Y, F), \tag{18}
\]

we first simulate from the conditional posterior distribution of \( B \), and then from the conditional posterior distribution of \( \Omega^{-1} \) given the drawn value of \( B \).

First, we show that \( B \) can be sampled from a distribution which is a matrix-variate extension of the multivariate Student \( t \)-distribution, and whose density is a product of multivariate \( t \)-densities

\[
p(B \mid Y) = p(b_1 \mid Y)p(b_2 \mid b_1, Y) \ldots p(b_n \mid b_1, b_2, \ldots, b_{n-1}, Y). \tag{19}
\]

Using the identity in Result 1 in Appendix A, and setting \( q \) equal to \( (T - n + 1) \), we integrate out \( \Omega^{-1} \) in the conditional posterior density \( (17) \). We obtain the conditional posterior distribution of \( B \),

\[
p(B \mid Y, F) = \int_{\Omega^{-1} > 0} p(B, \Omega^{-1} \mid Y, F) d\Omega^{-1} \propto |\Omega|^{-\frac{T}{2}} \tag{20}
\]
where \( \mathbf{R} \) is defined below. From this very general result we can derive the posterior distribution for the parameter matrix \( \mathbf{B} \) in the case of the linear multivariate factor SVOL model. This derivation follows Box and Tiao (1992). We can write each element of the matrix \( \mathbf{R} \):

\[
\mathbf{R}_{ij} = (y_i - \mathbf{b}_i \mathbf{F})'(y_j - \mathbf{b}_j \mathbf{F}) = (i, j = 1, ..., n) \quad (21)
\]

\[
= (y_i - \hat{\mathbf{b}}_i \mathbf{F})'(y_j - \hat{\mathbf{b}}_j \mathbf{F}) + (\mathbf{b}_i - \hat{\mathbf{b}}_i)' \mathbf{F} \mathbf{F}(\mathbf{b}_j - \hat{\mathbf{b}}_j) \quad (22)
\]

where \( \hat{\mathbf{b}}_i = (\mathbf{FF}')^{-1} \mathbf{Fy}_i \) is a \( k \times 1 \) vector with the least squares estimates of \( \mathbf{b}_i \) for asset \( i \), \( i = 1, ..., n \), and \( y_i \) is the \( i \)-th row of the return matrix \( \mathbf{Y} \). Denoting \( \hat{\mathbf{B}} \) to be the \( k \times n \) matrix concatenating the vectors \( \hat{\mathbf{b}}_i \), then for all returns, \( \mathbf{y}_t \), across all time observations, we have

\[
\mathbf{R} = \mathbf{D} + (\hat{\mathbf{B}} - \hat{\mathbf{B}}) \mathbf{F} \mathbf{F}'(\hat{\mathbf{B}} - \hat{\mathbf{B}})' \quad (23)
\]

where

\[
\mathbf{D} = (\mathbf{Y} - \hat{\mathbf{B}} \mathbf{F})(\mathbf{Y} - \hat{\mathbf{B}} \mathbf{F})' \quad (24)
\]

From the general result in Appendix A, we obtain the posterior distribution of \( \mathbf{B} \)

\[
p(\mathbf{B} \mid \mathbf{Y}, \mathbf{F}) \propto |\mathbf{D} + (\mathbf{B} - \hat{\mathbf{B}}) \mathbf{F} \mathbf{F}'(\mathbf{B} - \hat{\mathbf{B}})'|^{-\frac{T}{2}} \quad (25)
\]

In the case of one factor that is equal to unity, the matrix \( \mathbf{B} \) is an \( n \times 1 \) vector and we have

\[
\mathbf{b} = (b_1, ..., b_n)', \quad \hat{\mathbf{b}} = (\bar{y}_1, ..., \bar{y}_n)', \quad \mathbf{FF}' = T,
\]

\[
d_{ij} = \sum_{t=1}^{T} (y_{ti} - \bar{y}_i)(y_{tj} - \bar{y}_j) \quad i = 1, ..., n, \quad j = 1, ..., n \quad (26)
\]

from which we obtain the result

\[
p(\mathbf{b} \mid \mathbf{Y}) \propto |\mathbf{D} + T(\mathbf{b} - \hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})'|^{-\frac{T}{2}} \quad (27)
\]

\[
\propto |\mathbf{I} + T\mathbf{D}^{-1}(\mathbf{b} - \hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})'|^{-\frac{T}{2}}. \quad (28)
\]

We next implement the matrix identity in Result 2 in the appendix and obtain:

\[
p(\mathbf{b} \mid \mathbf{Y}) \propto |1 + T(\mathbf{b} - \hat{\mathbf{b}})'\mathbf{D}^{-1}(\mathbf{b} - \hat{\mathbf{b}})|^{-\frac{T}{2}} \quad (29)
\]

which is an \( n \)-dimensional \( t_n(\mathbf{b} \mid \hat{\mathbf{b}}, T^{-1}(T - n)^{-1}\mathbf{D}, T - n) \) distribution with \( (T - n) \) degrees of freedom and mean and variance parameters equal to \( \hat{\mathbf{b}} \) and \( T^{-1}(T - n)^{-1}\mathbf{D} \), respectively. If the number of returns, \( n \), or the number of factors, \( k \), is different from one, we cannot express the distribution of the factor sensitivities, \( \mathbf{B} \), as a multivariate \( t \)-distribution. In terms of the \( n \times k \) matrix of factor sensitivities, \( \mathbf{B} \), we can first apply the identity (see Appendix A):

\[
|\mathbf{I}_n + \mathbf{D}^{-1}(\hat{\mathbf{B}} - \mathbf{B}) \mathbf{F} \mathbf{F}'(\mathbf{B} - \hat{\mathbf{B}})'| = |\mathbf{I}_k + (\mathbf{FF}')((\mathbf{B} - \hat{\mathbf{B}})'\mathbf{D}^{-1}(\mathbf{B} - \hat{\mathbf{B}})| \quad (30)
\]
The above result allows us to conclude that we can represent each row of $B$ as a vector following a $k$-variate $t$-distribution, or equivalently, using the identity \[57\], we can represent each column of $B$ following an $n$-variate $t$-distribution. Thus, we can regard the general case \[25\] as a matrix-variate extension of of the multivariate $t$-distribution:

\[
p(B \mid Y, F) \propto t_{nk}(\tilde{B}, D^{-1}, (FF')^{-1}, T - (n + k) + 1) \quad (31)
\]

\[
p(B' \mid Y, F) \propto t_{kn}(\tilde{B}', (FF')^{-1}, D^{-1}, T - (n + k) + 1) \quad (32)
\]

The density function for the matrix-variate extension \[31\] is:

\[
p(B \mid Y, F) = C(n, k, \nu) |FF'|^{\frac{n}{2}} |D^{-1}|^{\frac{k}{2}} I_n + D^{-1}(B - \tilde{B})FF'(B - \tilde{B})^{-\frac{1}{2}(\nu + k + n - 1)} \quad (33)
\]

where $\nu = T - (n + k) + 1$ is the degrees of freedom parameter, and $C(n, k, \nu) = \left[ \Gamma\left(\frac{1}{2}nk\right) \Gamma_n\left(\frac{1}{2}(k + m)ight) \right]^{-1}$, where $\Gamma_n(x)$ is the generalized Gamma function defined in \[56\]. This constant applies to the multivariate extension $t_{nk}()$ in \[31\]. For $t_{kn}()$ in \[32\] the constant becomes $C(k, n, \nu) = \left[ \Gamma\left(\frac{1}{2}nk\right) \Gamma_k\left(\frac{1}{2}(n + m)ight) \right]^{-1}$.

We next examine the marginal and conditional distributions of subsets of rows of the parameter matrix $B$ with the purpose of representing the distribution of $B$ as a product of conditional densities. First we create two partitions of $B$ with $n_1$ and $n_2$ rows, $n_1 + n_2 = n$. We also partition the matrices $\tilde{B}$ and $D$:

\[
B = \left( \begin{array}{c} B_1 \\ B_2 \end{array} \right) \quad \tilde{B} = \left( \begin{array}{c} \tilde{B}_1 \\ \tilde{B}_2 \end{array} \right) \quad D = \left( \begin{array}{c} D_{11} \ D_{12} \\ D_{21} \ D_{22} \end{array} \right) \quad (34)
\]

Box and Tiao \[1992\] report the following results.

1. conditional on $B_1$, the subset $B_2$ is distributed as the extended multivariate $t$-distribution:

\[
B_2 \sim t_{kn_2}(\tilde{B}_2, H^{-1}, D_{22}^*, \nu + n_1) \quad (35)
\]

where

\[
H^{-1} = (FF')^{-1} + (B_1 - \tilde{B}_1)'D_{11}^{-1}(B_1 - \tilde{B}_1) \quad (36)
\]

\[
\tilde{B}_2 = \tilde{B}_2 + (B_1 - \tilde{B}_1)'D_{11}^{-1}D_{12} \quad (37)
\]

\[
D_{22}^* = D_{22} - D_{21}D_{11}^{-1}D_{12} \quad (38)
\]

2. $B_1$ is distributed as:

\[
B_1 \sim t_{nk}(\tilde{B}_1, (FF')^{-1}, D_{11}, \nu) \quad (39)
\]

Box and Tiao \[1992\] provide the following proof:

\[
(B - \tilde{B})'D^{-1}(B - \tilde{B}) = (B_1 - \tilde{B}_1)'D_{11}^{-1}(B_1 - \tilde{B}_1) + (B_2 - \tilde{B}_2)'D_{22}^{-1}(B_2 - \tilde{B}_2)
\]

\[
\downarrow
\]

\[
\left| I_k + (FF')(B - \tilde{B})D^{-1}(B - \tilde{B})' \right| = \left| I_k + (FF')(B_1 - \tilde{B}_1)'D_{11}^{-1}(B_1 - \tilde{B}_1) \right| \times
\]

\[
\times \left| I_k + H(B_2 - \tilde{B}_2)'D_{22}^{-1}(B_2 - \tilde{B}_2) \right| \quad (39)
\]
Substituting (39) into (33) we obtain the distribution of $B_2$ conditional on $B_1$:

$$p(B_2 \mid B_1, Y) \propto |I_k + H(B_2 - \tilde{B}_2)'D_{22}^{-1}(B_2 - \tilde{B}_2)|^{-\frac{1}{2}(\nu+k+n-1)}$$

(40)

$$\propto |I_{n_2} + D_{22}^{-1}(B_2 - \tilde{B}_2)H(B_2 - \tilde{B}_2)'|^{-\frac{1}{2}(\nu+n_1+k+n_2-1)}$$

(41)

$$= C(n_2, k, \nu + n_1) |H|^{\frac{n_2}{2}} |D_{22}|^{\frac{k}{2}} \times$$

$$\times |I_{n_2} + D_{22}^{-1}(B_2 - \tilde{b}_2)H(B_2 - \tilde{B}_2)'|^{-\frac{1}{2}(\nu+n_1+k+n_2-1)}$$

(42)

Using the above approach and also using the following identity for the marginal distribution of $B_1$:

$$p(B_1 \mid Y) = \frac{p(B_1)}{p(B_2 \mid B_1, Y)}$$

(43)

we obtain the marginal density of $B_1$:

$$p(B_1 \mid Y) \propto |H|^{-\frac{n_2}{2}} |I_k + (FF')(B_1 - \hat{B}_1)'D_{11}^{-1}(B_1 - \hat{B}_1)|^{-\frac{1}{2}(\nu+k+n-1)}$$

$$\propto |I_{n_1} + D_{11}^{-1}(B_1 - \hat{B}_1)(FF')(B_1 - \hat{B}_1)'|^{-\frac{1}{2}(\nu+n_1+k-1)}$$

(44)

Or the $n_1 \times k$ matrix of parameters $B_1$ follows an extended multivariate $t$-distribution:

$$B_1 \sim t_{n_1k} \left( \hat{B}_1, D_{11}, (FF'), \nu \right)$$

(45)

By setting $n_1 = 1$ we observe that a particular row of $B$ follows a $k$-dimensional multivariate $t$-distribution previously described in (31):

$$p(b_1 \mid Y) \propto \left| 1 + d_{11}^{-1}(b_1 - \hat{b}_1)(FF')(b_1 - \hat{b}_1) \right|^{-\frac{1}{2}(\nu+k)}$$

(46)

However, here we have fewer degrees of freedom due to the introduction of the additional $\frac{n(n-1)}{2}$ parameters in this multivariate setup. Alternatively, if we set $n_1 = n - 1$ and $n_2 = 1$, then the conditional distribution of $b_2 = b_n$ conditional on $B_1' = [b_1, ..., b_{n-1}]$ is:

$$b_n \sim t_k \left( \hat{b}_n, \nu + n - 1, d_{nn}^{-1}, \nu + n - 1 \right)$$

(47)

Therefore, if we partition $B$ $(n - 1)$ times we obtain the result that the distribution of $B$ is a product of multivariate $t$-distributions,

$$p(B \mid Y) = p(b_1 \mid Y)p(b_2 \mid b_1, Y) \cdots p(b_n \mid b_1, b_2, ..., b_{n-1}, Y)$$

(48)

Given $B$, sampling $\Omega^{-1}$ is straightforward. From (17), we obtain

$$\Omega^{-1} \sim p(\Omega^{-1} \mid B, Y, F) = Wishart_n(T, R^{-1})$$

(49)

where the degrees of freedom parameter is equal to the number of time observations, $T$, and the scale parameter $R^{-1}$ is a matrix function of $B$ which is defined in (23).
4.2 Conditional Posterior Distribution of $V_t$

For periods $t = 1, 2, ..., T - 1$ the conditional posterior distribution is proportional to:
\[
p(V_t^{-1} \mid \text{rest}) \propto \text{Wish}(V_t^{-1} \mid \nu, S_{t-1}) \times \mathcal{N}(0, V_t) \times \text{Wish}(V_{t+1}^{-1} \mid \nu, S_t)
\]
\[
\propto \exp(-\frac{1}{2} \text{tr}[S_{t-1}^{-1}V_t^{-1}]) \left| \sum_{t=1}^{T} \frac{1}{2} y_t' V_t^{-1} y_t \right| \exp(-\frac{1}{2} \text{tr}[S_t^{-1}V_{t+1}^{-1}])
\]
\[
\propto \text{Wish}(V_t^{-1} \mid \tilde{\nu}, \tilde{S}_{t-1}) \times f(V_t^{-1})
\]

where, $\tilde{\nu} = \nu(1 - d) + 1$, $\tilde{S}_{t-1} = S_{t-1}^{-1} + y_t y_t'$, and $f(V_t^{-1}) = \left| V_t^{-1} \right|^{\frac{1 - \nu}{2}} \exp(-\frac{1}{2} \text{tr}[S_t^{-1}V_{t+1}^{-1}])$ is a residual term. We assume a known initial period covariance matrix $V_0^{-1}$.

4.3 Conditional posterior distribution of $A^{-1}$

We obtain a conditional distribution for $A^{-1}$ that is a product of a Wishart prior and a Wishart likelihood, and is also a Wishart distribution:
\[
p(A^{-1} \mid \text{rest}) \propto \text{Wish}(\gamma_0, Q_0) \times \text{Wish}(\gamma, Q) \propto
\]
\[
\propto |A^{-1}|^{-\frac{m + k - 1}{2}} \exp(-\frac{1}{2} \text{tr}[Q_0^{-1} A^{-1}]) |A^{-1}|^{-\frac{\nu}{2}} \exp(-\frac{1}{2} \text{tr}(\nu Q^{-1} A^{-1}))
\]
\[
\propto \text{Wish}(A^{-1} \mid \tilde{\gamma}, \tilde{Q})
\]

where $\tilde{Q} = \tilde{Q}(V^{-1}, d, \nu)$ is the scale parameter, a function of $V^{-1}, d, \nu$, $\tilde{Q}^{-1} = \nu Q^{-1} + Q_0^{-1}$, $Q^{-1} = \sum_{t=1}^{T} (V_t^{-1})^{-\frac{d}{2}} V_t^{-1} (V_{t-1}^{-1})^{-\frac{d}{2}}$, $\tilde{\gamma} = \gamma + \gamma_0 - k - 1 = T \nu + \gamma_0$, $\gamma = T \nu + k + 1$ (we use the $|A^{-1}|^{-\frac{T \nu}{2}}$ term in the likelihood above to solve for $\gamma$). We are able to sample directly from the conditional distribution of $A^{-1}$ as a step in the Gibbs sampler.

4.4 Conditional posterior distributions of $d$ and $\nu$

The conditional distribution of $d$ is:
\[
p(d \mid \text{rest}) \propto \exp(d \psi - \frac{1}{2} \text{tr}[\nu A^{-1} Q(d)^{-1}])
\]

(52)

where $\psi = -\frac{\nu}{2} \sum_{t=1}^{T} \text{ln}(|\Sigma_t^{-1}|)$, and $Q^{-1}$ is the parameter defined in the previous section. While $d$ was treated as a fixed parameter in the previous section, now it is an argument of the function $Q^{-1}$.

The conditional distribution of $\nu$ is:
\[
p(\nu \mid \text{rest}) \propto e^{(\alpha - 1) \text{log}(\nu - k) - \beta(\nu - k)} \left( \frac{|\nu A^{-1}|^{\frac{\nu}{2}}}{2^{\nu k} \prod_{j=1}^{k} \Gamma \left( \frac{\nu + j - 1}{2} \right)} \right)^T
\]
\[
\times \prod_{t=1}^{T} |Q_t^{-1}|^{\frac{\nu}{2}} \prod_{t=1}^{T} \exp(-\frac{\nu}{2} \text{tr}[A^{-1} Q_t^{-1}])
\]

(53)
5 Results

We first evaluate our model on simulated data to assess convergence properties. The simulation exercise, which involves two observed factors and five asset returns, reveals that the true parameters are estimated without bias and after speedy convergence. We then apply our model to actual monthly returns data of 88 individual stocks using the two [Fama and French (1993)] factors. The data application reveals that the factors exhibit significant time-varying correlations and volatilities. In addition, the exact small sample inference provided by the MCMC simulation reveals that the value factor is not significant in our analysis based on posterior confidence intervals.

5.1 Simulation Results

The simulation exercise consists of two stages. First, known fixed values are determined for $A$, $d$, $\nu$, $\Omega$, $B$, and for the initial factor covariance matrix, $V_0$. These values represent the true model parameters. Using the true parameter values, covariance matrices are sampled once for each time period, according to the model. Using the true covariance parameters, we generate the sample of returns data. The second stage of the simulation exercise estimates the parameters using the data generated in stage one. The convergence of the Gibbs sampler is assessed by observing convergence plots of the sampled estimates in the burn-in stage. The estimates are based on Gibbs samples from fifty different trials. In each trial the Gibbs sampler was run over 1000 iterations, and then the values of the last draws of the Gibbs sampler was collected. The estimates are the mean values of the draws over the 50 trials of the simulation experiment. The results from estimating the model parameters in the simulation exercise are presented in Table 2. Figures 1 and 2 present graphical evidence of the accuracy of the estimation algorithm to back out the true model values. We can observe that the drawn values of the log-determinants of the intertemporal sensitivity matrix $A$ and the idiosyncratic risk matrix $\Omega$ are centered around the true model values. Their posterior sampling distributions are symmetric and bell-shaped. Figures 3 shows the mean of the sensitivities matrix for each draw for all 50000 draws of the simulation study along with the mean of the true model parameter. Figure 4 shows all Gibbs sampler draws from all 50 trials for each of the elements of the factor sensitivities matrix centered around the true model parameter values.

Overall, the simulation study shows that the factor volatility model parameters estimates are unbiased and efficiently estimated. The simulation study reveals sensitivity between the matrix parameter $A$ and the persistence parameter $d$. The sensitivity remains after imposing a diagonality constraint on $A$, which significantly reduces the number of model parameters. The interaction between the $A$ and $d$ parameters in the Gibbs sampler could prevent them from converging to their true values. However, the bias in one of the parameters is compensated by opposite bias in the other parameter, resulting in an unbiased estimate of the conditional expectation of the covariance matrix in time $t$. The Metropolis algorithm, which is used to draw the covariance matrix estimates only, takes the most CPU
time. Hence, the Metropolis efficiency is very important for the overall Gibbs sampler performance. The algorithm is assessed by analyzing acceptance ratios in the data application for all time periods. Metropolis acceptance ratios average 16%. Acceptance ratios decrease fast with the increase in model dimensions. For a 12-variate model, the average acceptance ratio is 0.5%.

5.2 Empirical Results

The factor multivariate SVOL model was implemented on return series data of 88 individual companies from the S&P500 which have long enough history. The data set contains 324 monthly observations, from January 1975 to December 2001. The two observable factors in the model are the Fama French market excess return and the value factors. The value factor represents the difference between growth stocks and value stocks, which are defined as high book-to-value and low book-to-value stocks, respectively. Descriptive statistics of the data are presented in Table 1. The factor data are available on the web at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

Figure 5 shows the first 10000 draws of the Gibbs sampler in the burn-in stage for the log-determinants of $A$ and $\Omega$ and the mean of $B$. The initial values of $A$ are far from the convergence values which helps observe the convergence of the parameter. Table 3 presents the estimates for the intertemporal sensitivity matrix $A$. It is interesting to observe that the two factor volatilities are strongly positively auto-correlated. However there is very little cross-correlation in the volatilities. Indeed the factors are constructed so that they unconditional correlation is zero. However, diagonality in $A$ does not restrict the volatility matrix to be diagonal, it only implies the the volatilities are unconditionally uncorrelated. Figure 10 shows the correlations implied by the estimated covariance matrices $V_t$ over the whole sample period. Since our model does not restrict covariances or correlations, we can observe significant time variation in these parameters. The average sample correlation is close to zero. However, there is significant time variation in the factor correlations. They vary from -0.7 to about 0.6. Such unrestricted variation in the correlations is a feature of our model that is not captured by alternative volatility models. The standard deviations of the two factors also exhibit significant time variation. The standard deviation of the second factor has smaller values on average but exhibits large spikes. The scalar persistence parameter $d$ was fixed at 0.5 as a way to eliminate the sensitivity between $A$ and $d$ revealed by earlier simulation studies. This constraint does not affect the estimation of the conditional expectation of the factor covariance matrices. The estimated factor sensitivities, the idiosyncratic risk, and the factor variance parameters can best be summarized graphically due to their large numbers. Figures 6 and 7 show the mean estimates of the posterior samples for the factor sensitivities of all assets to the two factors, with 5 and 95% bounds. Our MCMC approach to estimating the factor sensitivities enables exact finite sample inference compared to the asymptotic approximations usually employed in alternative estimation methods. We can observe in Figures 6 and 7 that the sensitivities to the market factor are significantly different from zero for most of the sample, while the book-to-market factor is not significant.
Figures 8 and 9 show the estimates of the standard deviations for the two factors derived from each draw of the factor covariance matrix in the Gibbs sampler.

6 Discussion

This paper presented the formulation, the estimation methodology, and data application results for a factor multivariate SVOL model in which the time-varying factor volatilities matrix is driven by a Wishart process. The model is part of a unified framework, in which univariate and multivariate SVOL model formulations are part of the same model. The general model formulation offers several advantages over previous multivariate SVOL models. First, it naturally extends scalar variances into covariance matrices rather than vectors of log-variances. Second, time variation of covariance matrices is not restricted by the sole variation in variances (assumed by existing MSVOL models). The Wishart distribution scale parameter, which underlies the conditional expectation of the asset covariance matrix, allows for both variances and correlations to evolve stochastically over time. Third, the general model allows for the conditional volatility of an asset to depend not only on its past volatility but also on past covariances with other assets, incorporating the observed contagion among asset returns into their covariance structure. Finally, the general setup of our multivariate SVOL model includes most of the existing SVOL models as special cases, allowing the researcher to test the restrictions imposed by previous SVOL models.

Our factor model can be seen as a special case of the more general multivariate SVOL model of Philipov and Glickman (2004). The general model allows easy imposition of constraints in which some popular existing stochastic volatility model can be recognized as special cases. A multivariate factor structure is easily incorporated in the model specification. In the context of existing factor multivariate models, the current model offers several advantages. First, the evolution of factor covariances is described by general ”matrix-variate” Wishart random process which allows for interaction among the factors, their variances, and covariances, thus giving more support to the empirical evidence on observed priced factors. As part of the general multivariate SVOL model of Philipov and Glickman (2004), the factor model in this paper offers a natural framework for testing the asset-pricing restrictions imposed by factor-based models. The factor structure allows addressing high dimensional setups used in portfolio analysis and risk management, as well as modelling conditional means and conditional variances within the model framework. We show that it is possible to implement factor multivariate SVOL models with very high dimensionality, making them feasible for practical applications in risk management, asset allocation, and portfolio optimization.

In this paper we deal exclusively with the case of observable factors. Many of the popular factor SVOL models deal with latent factors. We can see that in the case of unobservable factors, the model imposes additional restrictions on the factor variance structure that eliminate some desirable current features of our model. For example, an orthogonality constraint on the latent factors would restrict both unconditional and time-varying correlations among the factors to zero and would limit the model to estimating a vector of variances, as is the case with popular factor SVOL models. We believe that the model presented in this paper
is very useful in describing asset pricing applications where there are multiple observable market factors driving equity returns.
APPENDIX A

A.1 Results used in the Multivariate Factor Model

Result A.1.1 Let \( p(Z) \) be a Wishart\((q, \Theta)\) density function (Box and Tiao 1992):

\[
p(Z) \propto |Z|^{\frac{1}{2}q-1} \exp(-\frac{1}{2}trZ\Theta) \tag{54}
\]

after integrating \( Z \) out we obtain the identity:

\[
\int_{Z>0} |Z|^{\frac{1}{2}q-1} \exp(-\frac{1}{2}trZ\Theta) \, dZ = |\Theta|^{-\frac{1}{2}(q+n-1)} 2^{\frac{1}{2}n(q+n-1)} \Gamma_n \left( \frac{q+n-1}{2} \right) \tag{55}
\]

where \( \Gamma_q(b) \) is the generalized gamma function (Siegel, 1935):

\[
\Gamma_q(b) = \left[ \Gamma \left( \frac{1}{2} \right) \right]^{\frac{1}{2}p(p-1)} \prod_{\alpha=1}^{p} \Gamma \left( \frac{b + \alpha - p}{2} \right), \quad b > \frac{p-1}{2} \tag{56}
\]

Result A.1.2 A fundamental matrix identity:

\[
\begin{vmatrix} K \times K & I \times L & L \times K \times L \end{vmatrix} = \begin{vmatrix} L \times L & I \times K & K \times L \end{vmatrix} \tag{57}
\]

A.2 Derivation of the Distributional Results

A.2.1 Derivation of \( L(A \mid \text{rest}) \)

From (5) we derive:

\[
L(A \mid \text{rest}) = C^T \prod_{t=1}^{T} \exp \left( -\frac{\nu}{2} tr[A^{-1} (\Sigma_{t-1}^{-1})^{-d} \Sigma_t^{-1}] \right) =
\]

\[
= \frac{\left( \frac{1}{2} \right)^{\nu p} |A|^{-\nu \frac{p}{2}}}{2^{\frac{\nu p}{2}} \pi^{\frac{k(k-1)}{2}} \prod_{j=1}^{p} \Gamma \left( \frac{\nu_j + 1}{2} \right)} \exp \left( -\frac{\nu}{2} \sum_{t=1}^{T} tr[A^{-1} (\Sigma_{t-1}^{-1})^{-d} \Sigma_t^{-1}] \right) =
\]

\[
= |A|^{-\nu \frac{p}{2}} \exp \left( -\frac{1}{2} tr[A^{-1} \nu \sum_{t=1}^{T} (\Sigma_{t-1}^{-1})^{-d} \Sigma_t^{-1}] \right)
\]

A.2.2 The product of multivariate normal and wishart

This product is another wishart up to a normalizing constant:

\[
N(y_t | 0, \Sigma_t) \text{ Wishart}_k(\Sigma_t^{-1} | \nu, S_{t-1}) =
\]

\[
= |\Sigma_t^{-1}|^{\frac{\nu}{2}} \exp \left( -\frac{1}{2} y_t' \Sigma_t^{-1} y_t \right) \times
\]

\[
\times |S_{t-1}|^{-\frac{\nu}{2}} |\Sigma_{t-1}|^{\frac{\nu-k+1}{2}} \exp \left( -\frac{1}{2} tr[S_{t-1}^{-1} \Sigma_{t-1}] \right) =
\]

\[
= |S_{t-1}|^{-\frac{\nu}{2}} |\Sigma_{t-1}|^{\frac{\nu}{2}} |\Sigma_{t-1}|^{\frac{\nu-k+1}{2}} \exp \left( -\frac{1}{2} tr[y_t y_t' + S_{t-1}^{-1} \Sigma_{t-1}] \right) \propto
\]

\[
\alpha |\tilde{S}_{t-1}|^{-\frac{\nu-k+1}{2}} |\Sigma_{t-1}|^{\frac{\nu+1-k}{2}} \exp \left( -\frac{1}{2} tr[\tilde{S}_{t-1}^{-1} \Sigma_{t-1}] \right)
\]
A.2.5 Transforming

We have transformed the log scalar into a matrix.

We have:

\[ \rho \]

A.2.4 Expressing

\[ \zeta − \gamma = \exp k \]

In the last line we need to adjust the first term under the sum, dividing by \( k \).

\[ \exp (−\frac{1}{2}tr[(\Psi^{-1} + Q^{-1})] \] \[ \propto W ishart \]

where \( R^{-1} = \Psi^{-1} + Q^{-1} \), and the new degrees of freedom parameter, \( \gamma^* \) equals \( \gamma + \zeta - k - 1 \).

A.2.3 The product of two Wisharts

The product is proportional to another wishart density:

Let \( A \sim W ishart(\gamma, \Psi) \), and \( A \sim W ishart(\zeta, Q) \)

\[ p(A \mid \text{new parameters}) = \text{Wishart}(\gamma, \Psi)\text{Wishart}(\zeta, Q) = \]

\[ C^2 |\Psi Q|^{-\frac{1}{2}} |A|^{-\frac{1}{2} + \frac{\zeta - k - 1}{2}} \exp(−\frac{1}{2}tr[(\Psi^{-1} + Q^{-1})A]) \]

\[ \propto |R|^{-\frac{1}{2}} |A|^{-\frac{1}{2} + \frac{\gamma + \zeta - k - 1 - k - 1}{2}} \exp(−\frac{1}{2}tr[(\Psi^{-1} + Q^{-1})A]) \]

\[ \propto \text{Wishart}(\gamma^*, R) \]

where \( R^{-1} = \Psi^{-1} + Q^{-1} \), and the new degrees of freedom parameter, \( \gamma^* \) equals \( \gamma + \zeta - k - 1 \).

A.2.4 Expressing \( p(d \mid \text{rest}) \) as \( \exp(\text{trace}) \)

We have:

\[ p(d \mid \text{rest}) = \left( \prod_{t=1}^{T} |\Sigma_t^{-1}|^{-\frac{d}{2}} \right)^{d} \times \exp \left( −\frac{d}{2}tr[A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d}] \right) = \]

\[ = \exp \left( d \log \left( \prod_{t=1}^{T} |\Sigma_t^{-1}|^{-\frac{d}{2}} \right) \right) \times \exp \left( −\frac{d}{2}tr[A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d}] \right) = \]

\[ = \exp \left( −\frac{1}{2}tr \left( d \sum_{t=1}^{T} \log |\Sigma_t^{-1}| \right) \right) \times \exp \left( −\frac{d}{2}tr[A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d}] \right) = \]

\[ = \exp \left( −\frac{1}{2}tr \left( A^{-1} \sum_{t=1}^{T} \left( \frac{d}{k}A \log |\Sigma_t^{-1}| + \Sigma_t^{-1} (\Sigma_t^{-1})^{-d} \right) \right) \right) \]

In the last line we need to adjust the first term under the sum, dividing by \( k \), because we have transforme the log scalar into a matrix.

A.2.5 Transforming \( p(\nu \mid \text{rest}) \)

\[ p(\nu \mid \text{rest}) = \left( \frac{|\frac{1}{2}A|^{-\frac{\nu}{2}}}{2^{\nu k} \prod_{j=1}^{k} \Gamma^\nu \left( \frac{\nu + 1}{2} \right)} \right)^{\frac{T}{2}} \prod_{t=1}^{T} |(\Sigma_t^{-1})^{-d} \Sigma_t^{-1}|^{-\frac{\nu}{2}} \]

\[ \times \exp \left( −\nu \frac{1}{2}tr \left[ A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d} \Sigma_t^{-1} \right] \right) = \]

\[ = f(\nu)^T \exp \left( \nu \frac{1}{2} \sum_{t=1}^{T} \ln |(\Sigma_t^{-1})^{-d} \Sigma_t^{-1}| \right) \times \exp \left( −\nu \frac{1}{2}tr \left[ A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d} \Sigma_t^{-1} \right] \right) = \]

\[ = f(\nu)^T \exp \left( \frac{\nu}{2} \left( tr \left[ \sum_{t=1}^{T} \ln |(\Sigma_t^{-1})^{-d} \Sigma_t^{-1}| \right] - tr \left[ A^{-1} \sum_{t=1}^{T} (\Sigma_t^{-1})^{-d} \Sigma_t^{-1} \right] \right) \right) = \]

\[ = f(\nu)^T \exp \left( \frac{\nu}{2} \left[ A^{-1} \sum_{t=1}^{T} \left( \frac{1}{k}A \ln |(\Sigma_t^{-1})^{-d} \Sigma_t^{-1}| - (\Sigma_t^{-1})^{-d} \Sigma_t^{-1} \right) \right] \right) \]
which can be further modified:

\[
p(\nu \mid \text{rest}) = \left( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right)^{-T} \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \sum_{t=1}^{T} (\cdot) \right] \right) = \frac{\left( (2\nu \frac{1}{2})^{-T} \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right)}{\exp \left( \frac{\nu}{2} tr \left[ A^{-1} \sum_{t=1}^{T} (\cdot) \right] \right)} = \left( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right)^{-T} \times \exp \left( \frac{\nu}{2} \left( -T \ln \left| \frac{1}{2} A \right| \right) \right) \times \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \sum_{t=1}^{T} (\cdot) \right] \right) = \left( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right)^{-T} \times \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \sum_{t=1}^{T} \left( \frac{1}{k} A \ln \left| (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right| - (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right) \right) \right] \right)
\]

When computing the above function for different values of \( \nu \), the first term may get very large while the second practically zero and vice versa. In these cases the function will give a zero value. To get around this problem, we include the first term under the trace of the exponent:

\[
p(\nu \mid \text{rest}) = \left( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right)^{-T} \times \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \sum_{t=1}^{T} \left( \frac{1}{k} A \ln \left| (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right| - (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right) \right) \right] \right) = \exp \left( -T \ln \left( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) \right) \right) \times \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \sum_{t=1}^{T} \left( \frac{1}{k} A \ln \left| (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right| - (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right) \right) \right] \right)
\]

Call \( \prod_{j=1}^{k} \Gamma \left( \frac{\nu+1-j}{2} \right) = \varphi \), and \( \sum_{t=1}^{T} \left( A_{1}^{-1} \ln \left| (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right| - (\Sigma_{t-1}^{-1})^{-d} \Sigma_{t}^{-1} \right) = \Psi \). Then:

\[
p(\nu \mid \text{rest}) = \exp \left( -T \ln \varphi + \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \Psi \right) \right] \right) = \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \Psi \right) \right] \right) = \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \left| \frac{1}{2} A \right| \right) + \Psi \right) \right] \right) = \exp \left( \frac{\nu}{2} tr \left[ A^{-1} \left( \frac{1}{k} A \left( -T \ln \varphi + \ln \left| \frac{1}{2} A \right| \right) + \Psi \right) \right] \right)
\]
APPENDIX B

B.1 Algorithm for Sampling the Parameters of the Factor MSVOL

We sample the parameters, $B$, $\Omega$, $A$, $V_t$, $\nu$, and $d$. We specify initial values for the unobserved factors, $F$, the initial factor volatility matrix $V_0$, and the parameters of the prior distributions. The algorithm involves the following steps:

1. Draw the factor sensitivities matrix $B$ and the idiosyncratic volatility parameter $\Omega$ from their joint distribution (15), conditional on the data and the factors. This step includes two sampling substeps described below,
2. Draw the parameter $A$ for the distribution of the factor volatilities, $V_t$,
3. Draw the factor volatilities $V_t$, conditional on the parameters $V_{t-1}$, $A$, $\nu$, and $d$,
4. Draw the factors, $f_t$, conditional on the factor volatility $V_t$,
5. Draw the degrees of freedom parameter $\nu$, and
6. Draw the persistence parameter $d$

Step 1a. Sampling the factor sensitivities, $B$

This step involves the following operations:

1. Compute the mean factor sensitivities $\hat{B}$ using Ordinary Least Squares Regression
   (a) In the case of unobserved factors, we follow Geweke and Zhou (1996) and restrict the first asset sensitivities to factors $f_2, \ldots, f_k$ to zero. Therefore the first row of $\hat{B}$ will result from an OLS regression of asset one on factor one only.
   (b) the second row will result from an OLS regression of asset two on factors one and two and so on.
   (c) we restrict all sensitivities of the first $k$ assets to be positive.
2. Compute the error terms:
   \[ e = Y - \hat{B}F \]
3. Compute the squared deviations matrix, $D$, defined in eq. (24)
   \[ D = ee' \]
4. Compute the degrees of freedom parameter for the multivariate t-distribution of the first row (this parameter increases by 1 with sampling every subsequent row)
   \[ \nu_{t-dist}^{(1)} = T - (n + k) + 1 \]
   where $T$ is the number of time observations, $n$ is the number of returns, and $k$ is the number of factors
5. draw the first row of $B$ from the marginal distribution $p(b_1 \mid Y, F)$

$$b_1 \sim t_k(\hat{b}_1, \nu_{t-dist}^{(1)}, S_{t-dist}^{(1)})$$

where $\hat{b}_1$ is the first row of the Least Squares estimator, $\hat{B}$, $\nu_{t-dist}^{(1)}$ is the degrees of freedom parameter, and $S_{t-dist}^{(1)} = \frac{1}{\nu_{t-dist}^{(1)}} d_{11}(FF')^{-1}$ is the scale parameter of the $t$-distribution.

6. draw each subsequent row $i$ of $B$, conditional on $Y$, $F$, and the $(i-1) \times (i-1)$ matrix $B_1$ of previously drawn rows of $B$. We need to go through the following steps:

- Compute the matrix $H^{-1}$. This matrix is part of the scale parameter of the conditional $t$-distribution of row $i$.

$$H^{-1} = (FF')^{-1} + (B_1 - \hat{B}_1)'D_{11}(B_1 - \hat{B}_1)$$

where $D_{11}$ is a scalar in the second iteration and is augmented each following iteration to include one more row and column. For example, in the fourth iteration we have:

$$D = D_4 = \begin{pmatrix}
D_{11} &=& \begin{pmatrix} d_{11} & d_{12} & d_{13} \\
 d_{21} & d_{22} & d_{23} \\
 d_{31} & d_{32} & d_{33}
\end{pmatrix} & D_{12} &=& \begin{pmatrix} d_{14} \\
 d_{24} \\
 d_{34}
\end{pmatrix} \\
D_{21} &=& \begin{pmatrix} d_{41} & d_{42} & d_{43}
\end{pmatrix} & D_{22} &=& d_{44}
\end{pmatrix}$$

- Compute the mean of the multivariate $t$-distribution:

$$\tilde{b}_2 = \hat{b}_1 + (B_1 - \hat{B}_1)'D_{11}d_{12}$$

- Compute the degrees of freedom parameter

$$\nu_{t-dist}^{(i)} = \nu_{t-dist}^{(1)} + i - 1$$

- Compute the scale parameter of the multivariate $t$-distribution

$$S_{t-dist}^{(i)} = \frac{1}{\nu_{t-dist}^{(1)}} d_{ii} H^{-1}$$

- Draw $b_i = b_2$ from a multivariate $t$-distribution:

$$b_2 \sim t_k(\tilde{b}_2, \nu_{t-dist}^{(i)}, S_{t-dist}^{(i)})$$

- Update the parameters $D_{11}, D_{12}, D_{21}, D_{22}, \hat{B}_1, B_1$
Step 1b. Sampling the idiosyncratic volatility parameter, Ω
The posterior of the volatility parameter Ω was derived in equation (49). We directly sample from the Wishart distribution:

\[ \Omega^{-1} \sim p(\Omega^{-1} \mid B, Y, F) = \text{Wishart}_n(T, S(B)^{-1}) \]

Step 2. Sampling the parameter A
We sample the inverse A\(^{-1}\) from a Wishart distribution, directly:

\[ A^{-1} \mid V, \nu, d \sim \text{Wishart}_k(\tilde{\gamma}, \tilde{Q}) \]

Step 3. Sampling the systematic volatility parameter V
We sample the parameters V\(_t\) iteratively, following the methodology for the extended multivariate model

\[ p(V_{t-1}^{-1} \mid A^{-1}, V_{t-1}, \nu, d) \propto \text{Wishart}(V_t^{-1} \mid \nu, S_t) \times N(0, V_t) \times \text{Wishart}(V_{t+1}^{-1} \mid \nu, S_t) \]
\[ \propto \exp\left(-\frac{1}{2} tr[S_t^{-1} V_t^{-1}] \right) \left| V_t^{-1} \right|^{\nu(1-d)-k} \exp\left(-\frac{1}{2} f_t^T V_t^{-1} f_t \right) \exp\left(-\frac{1}{2} tr[S_t^{-1} V_{t+1}^{-1}] \right) \]
\[ \propto \text{Wishart}(V_t^{-1} \mid \tilde{\nu}, \tilde{S}_t) \times f(V_t^{-1}) \]

\[ p(V_T^{-1} \mid A^{-1}, V_{T-1}, \nu, d) \propto \text{Wishart}(V_T^{-1} \mid \nu, S_{T-1}) \times N(0, V_T) \propto \]
\[ \propto \exp\left(-\frac{1}{2} tr[S_{T-1}^{-1} V_T^{-1}] \right) \left| V_T^{-1} \right|^{\nu(1-d)-k} \exp\left(-\frac{1}{2} f_T^T V_T^{-1} f_T \right) \propto \]
\[ \propto \text{Wishart}(V_T^{-1} \mid \tilde{\nu}, \tilde{S}_{T-1}) \]

Step 4. Sampling the unobservable factors, given V
For each time period, draw the factors f\(_{t}\) from a k-variate normal distribution:

\[ f_t \mid V_t \sim N_k(0, V_t) \]

Step 5. Sampling the degrees of freedom parameter, \(\nu\)
Step 6. Sampling the persistence parameter, \(d\)
References


<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>St.Dev</th>
<th>S.E.</th>
<th>Skew.</th>
<th>Kurt.</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>0.011</td>
<td>0.049</td>
<td>0.003</td>
<td>-0.024</td>
<td>3.719</td>
<td>-0.147</td>
<td>0.178</td>
<td>0.325</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.005</td>
<td>0.027</td>
<td>0.002</td>
<td>-0.316</td>
<td>7.051</td>
<td>-0.128</td>
<td>0.106</td>
<td>0.233</td>
</tr>
<tr>
<td>W.PENN ELCTR</td>
<td>0.014</td>
<td>0.057</td>
<td>0.003</td>
<td>0.757</td>
<td>6.370</td>
<td>-0.158</td>
<td>0.333</td>
<td>0.491</td>
</tr>
<tr>
<td>AMERICAN TOB</td>
<td>0.015</td>
<td>0.068</td>
<td>0.004</td>
<td>-0.018</td>
<td>4.774</td>
<td>-0.248</td>
<td>0.278</td>
<td>0.526</td>
</tr>
<tr>
<td>AT&amp;T</td>
<td>0.011</td>
<td>0.072</td>
<td>0.004</td>
<td>0.284</td>
<td>6.141</td>
<td>-0.256</td>
<td>0.391</td>
<td>0.647</td>
</tr>
<tr>
<td>ARCHER DANIELS</td>
<td>0.014</td>
<td>0.085</td>
<td>0.005</td>
<td>0.153</td>
<td>3.137</td>
<td>-0.274</td>
<td>0.275</td>
<td>0.549</td>
</tr>
<tr>
<td>BRUNSWICK BALKE</td>
<td>0.018</td>
<td>0.112</td>
<td>0.006</td>
<td>0.285</td>
<td>4.601</td>
<td>-0.387</td>
<td>0.495</td>
<td>0.883</td>
</tr>
<tr>
<td>BURROUGHS ADDING</td>
<td>0.010</td>
<td>0.142</td>
<td>0.008</td>
<td>0.768</td>
<td>9.676</td>
<td>-0.550</td>
<td>0.818</td>
<td>1.368</td>
</tr>
<tr>
<td>COCA COLA</td>
<td>0.016</td>
<td>0.066</td>
<td>0.004</td>
<td>0.059</td>
<td>3.967</td>
<td>-0.191</td>
<td>0.257</td>
<td>0.448</td>
</tr>
<tr>
<td>CNSLD EDISON</td>
<td>0.018</td>
<td>0.059</td>
<td>0.003</td>
<td>1.330</td>
<td>11.250</td>
<td>-0.140</td>
<td>0.450</td>
<td>0.590</td>
</tr>
<tr>
<td>DETROIT EDISON</td>
<td>0.014</td>
<td>0.056</td>
<td>0.003</td>
<td>1.856</td>
<td>16.863</td>
<td>-0.131</td>
<td>0.484</td>
<td>0.616</td>
</tr>
<tr>
<td>DU PONT</td>
<td>0.013</td>
<td>0.070</td>
<td>0.004</td>
<td>0.223</td>
<td>3.751</td>
<td>-0.242</td>
<td>0.276</td>
<td>0.519</td>
</tr>
<tr>
<td>EASTMAN KODAK</td>
<td>0.007</td>
<td>0.074</td>
<td>0.004</td>
<td>-0.263</td>
<td>5.026</td>
<td>-0.343</td>
<td>0.231</td>
<td>0.575</td>
</tr>
<tr>
<td>ELECTRIC BOAT</td>
<td>0.015</td>
<td>0.072</td>
<td>0.004</td>
<td>0.138</td>
<td>5.353</td>
<td>-0.293</td>
<td>0.352</td>
<td>0.645</td>
</tr>
<tr>
<td>GM</td>
<td>0.012</td>
<td>0.078</td>
<td>0.004</td>
<td>0.111</td>
<td>4.428</td>
<td>-0.295</td>
<td>0.277</td>
<td>0.572</td>
</tr>
<tr>
<td>GOODRICH BF</td>
<td>0.013</td>
<td>0.097</td>
<td>0.005</td>
<td>-0.114</td>
<td>6.705</td>
<td>-0.430</td>
<td>0.479</td>
<td>0.910</td>
</tr>
<tr>
<td>INGERSOLL RAND</td>
<td>0.012</td>
<td>0.091</td>
<td>0.005</td>
<td>0.368</td>
<td>4.927</td>
<td>-0.326</td>
<td>0.394</td>
<td>0.720</td>
</tr>
<tr>
<td>IBM</td>
<td>0.013</td>
<td>0.075</td>
<td>0.004</td>
<td>0.261</td>
<td>3.958</td>
<td>-0.262</td>
<td>0.318</td>
<td>0.580</td>
</tr>
<tr>
<td>INTL HARVESTER</td>
<td>0.005</td>
<td>0.135</td>
<td>0.008</td>
<td>0.818</td>
<td>4.434</td>
<td>-0.345</td>
<td>0.560</td>
<td>0.905</td>
</tr>
<tr>
<td>KRESGE SS</td>
<td>0.006</td>
<td>0.100</td>
<td>0.006</td>
<td>0.627</td>
<td>8.774</td>
<td>-0.440</td>
<td>0.647</td>
<td>1.087</td>
</tr>
<tr>
<td>MAY DEPT STRS</td>
<td>0.016</td>
<td>0.081</td>
<td>0.004</td>
<td>0.250</td>
<td>3.434</td>
<td>-0.222</td>
<td>0.287</td>
<td>0.509</td>
</tr>
<tr>
<td>MAYTAG</td>
<td>0.014</td>
<td>0.089</td>
<td>0.005</td>
<td>-0.442</td>
<td>5.786</td>
<td>-0.468</td>
<td>0.270</td>
<td>0.738</td>
</tr>
<tr>
<td>PG&amp;E</td>
<td>0.011</td>
<td>0.064</td>
<td>0.004</td>
<td>-0.125</td>
<td>8.030</td>
<td>-0.287</td>
<td>0.328</td>
<td>0.615</td>
</tr>
<tr>
<td>PPLS GAS LT&amp; COKE</td>
<td>0.014</td>
<td>0.063</td>
<td>0.003</td>
<td>0.059</td>
<td>3.640</td>
<td>-0.179</td>
<td>0.204</td>
<td>0.383</td>
</tr>
<tr>
<td>PEPSI</td>
<td>0.018</td>
<td>0.068</td>
<td>0.004</td>
<td>-0.061</td>
<td>4.332</td>
<td>-0.284</td>
<td>0.228</td>
<td>0.512</td>
</tr>
<tr>
<td>PHILIP MORRIS</td>
<td>0.018</td>
<td>0.072</td>
<td>0.004</td>
<td>-0.209</td>
<td>4.309</td>
<td>-0.266</td>
<td>0.244</td>
<td>0.510</td>
</tr>
<tr>
<td>A T F</td>
<td>0.011</td>
<td>0.080</td>
<td>0.004</td>
<td>0.057</td>
<td>3.803</td>
<td>-0.273</td>
<td>0.292</td>
<td>0.566</td>
</tr>
<tr>
<td>SEARS ROEBUCK</td>
<td>0.012</td>
<td>0.082</td>
<td>0.005</td>
<td>-0.028</td>
<td>3.734</td>
<td>-0.304</td>
<td>0.279</td>
<td>0.583</td>
</tr>
<tr>
<td>STD OIL CA</td>
<td>0.015</td>
<td>0.069</td>
<td>0.004</td>
<td>0.677</td>
<td>5.815</td>
<td>-0.176</td>
<td>0.363</td>
<td>0.539</td>
</tr>
<tr>
<td>SUN OIL</td>
<td>0.012</td>
<td>0.083</td>
<td>0.005</td>
<td>0.351</td>
<td>3.555</td>
<td>-0.190</td>
<td>0.314</td>
<td>0.504</td>
</tr>
<tr>
<td>UNION OIL CA</td>
<td>0.014</td>
<td>0.086</td>
<td>0.005</td>
<td>0.353</td>
<td>4.190</td>
<td>-0.256</td>
<td>0.338</td>
<td>0.594</td>
</tr>
<tr>
<td>US STEEL</td>
<td>0.009</td>
<td>0.088</td>
<td>0.005</td>
<td>0.577</td>
<td>4.788</td>
<td>-0.262</td>
<td>0.363</td>
<td>0.625</td>
</tr>
<tr>
<td>US TOB</td>
<td>0.019</td>
<td>0.075</td>
<td>0.004</td>
<td>1.036</td>
<td>8.721</td>
<td>-0.168</td>
<td>0.522</td>
<td>0.690</td>
</tr>
<tr>
<td>WRGLEY WILLIAM</td>
<td>0.019</td>
<td>0.070</td>
<td>0.004</td>
<td>0.199</td>
<td>3.524</td>
<td>-0.185</td>
<td>0.297</td>
<td>0.482</td>
</tr>
<tr>
<td>INTERCONT RUBBER</td>
<td>0.017</td>
<td>0.119</td>
<td>0.007</td>
<td>0.367</td>
<td>4.116</td>
<td>-0.349</td>
<td>0.542</td>
<td>0.891</td>
</tr>
<tr>
<td>SO CALIF EDISON</td>
<td>0.012</td>
<td>0.066</td>
<td>0.004</td>
<td>-0.747</td>
<td>8.902</td>
<td>-0.369</td>
<td>0.259</td>
<td>0.628</td>
</tr>
<tr>
<td>GILLETTE</td>
<td>0.018</td>
<td>0.078</td>
<td>0.004</td>
<td>-0.085</td>
<td>5.157</td>
<td>-0.280</td>
<td>0.317</td>
<td>0.597</td>
</tr>
<tr>
<td>GOODYEAR TIRE&amp;RUBR</td>
<td>0.012</td>
<td>0.092</td>
<td>0.005</td>
<td>0.266</td>
<td>4.833</td>
<td>-0.332</td>
<td>0.394</td>
<td>0.726</td>
</tr>
<tr>
<td>HERSHEY</td>
<td>0.019</td>
<td>0.074</td>
<td>0.004</td>
<td>0.135</td>
<td>4.242</td>
<td>-0.249</td>
<td>0.269</td>
<td>0.518</td>
</tr>
<tr>
<td>KROGER</td>
<td>0.021</td>
<td>0.087</td>
<td>0.005</td>
<td>0.894</td>
<td>7.684</td>
<td>-0.204</td>
<td>0.566</td>
<td>0.770</td>
</tr>
<tr>
<td>GENERAL MILLS</td>
<td>0.015</td>
<td>0.061</td>
<td>0.003</td>
<td>0.266</td>
<td>2.866</td>
<td>-0.148</td>
<td>0.190</td>
<td>0.338</td>
</tr>
<tr>
<td>MCGRAW HILL</td>
<td>0.019</td>
<td>0.073</td>
<td>0.004</td>
<td>-0.023</td>
<td>4.299</td>
<td>-0.304</td>
<td>0.313</td>
<td>0.617</td>
</tr>
<tr>
<td>Company</td>
<td>Sector</td>
<td>NAIC</td>
<td>S&amp;P</td>
<td>Price</td>
<td>EPS</td>
<td>P/E</td>
<td>ROE</td>
<td>ROA</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------</td>
<td>------</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>
TABLE 2
Simulation Results

The table presents the true predetermined values of the parameters in the model that describe the data generating process, and the estimated parameters using the Bayesian estimation procedures. The estimates are based on Gibbs samples from fifty different trials. In each trial the Gibbs sampler was run over 1000 iterations, and then the values of the last draws of the Gibbs sampler was collected. The estimates are the mean values of the last draws over the 50 trials of the simulation experiment.

A. True pre-determined values for the model used in the generation stage

<table>
<thead>
<tr>
<th></th>
<th>$d = 0.5$</th>
<th>$\nu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0$</td>
<td>1.20</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>-0.50</td>
<td>0.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1.50</th>
<th>-0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_{model}$</td>
<td>5.07</td>
<td>-1.93</td>
</tr>
<tr>
<td></td>
<td>-1.93</td>
<td>6.21</td>
</tr>
<tr>
<td></td>
<td>-0.61</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>-2.21</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>1.47</td>
</tr>
</tbody>
</table>

B. Estimated values for the model parameters

$d, \nu$ fixed at true values

<table>
<thead>
<tr>
<th></th>
<th>$d = 0.5$</th>
<th>$\nu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0$</td>
<td>1.20</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>-0.50</td>
<td>0.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1.64</th>
<th>-0.46</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{B}$</td>
<td>-1.18</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>2.04</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>3.72</td>
<td>-7.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>5.07</th>
<th>-1.93</th>
<th>-0.61</th>
<th>2.05</th>
<th>0.81</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Omega}$</td>
<td>-1.93</td>
<td>6.21</td>
<td>0.52</td>
<td>-2.21</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>-0.61</td>
<td>0.52</td>
<td>4.63</td>
<td>0.42</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>-2.21</td>
<td>0.42</td>
<td>4.71</td>
<td>-1.44</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>1.47</td>
<td>-0.60</td>
<td>-1.44</td>
<td>6.59</td>
</tr>
</tbody>
</table>
TABLE 3
Empirical Results

The table presents the estimate of the intertemporal sensitivity matrix $\mathbf{A}$ for the two-factor covariance structure over 324 time observations.

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>0.0382</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0.0355 0.0410]</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.0001 0.0179</td>
</tr>
<tr>
<td></td>
<td>[-0.0015 0.0013] [0.0166 0.0193]</td>
</tr>
</tbody>
</table>
Figure 1. Histogram of the log of the determinant of $A$ over 50 000 draws in a simulation exercise. The vertical dark line represents the true model value for the log-determinant of $A$. In the simulation exercise true model values were specified for the parameters, base on which the covariance matrices and data were generated. Using the generated covariance matrices and data, the model parameters were estimated and the estimates were compared to the true values to assess the precision and efficiency of the MCMC algorithm.

Figure 2. Histogram of the log of the determinant of $\Omega$ over 50 000 draws in a simulation exercise. The vertical dark line represents the true model value for the log-determinant of $\Omega$. In the simulation exercise true model values were specified for the parameters, base on which the covariance matrices and data were generated. Using the generated covariance matrices and data, the model parameters were estimated and the estimates were compared to the true values to assess the precision and efficiency of the MCMC algorithm.
Figure 3. Histogram of the mean of all elements of the beta matrix $\mathbf{B}$ over 50,000 draws in a simulation exercise. The vertical dark line represents the true model value for the mean $\mathbf{B}$. In the simulation exercise true model values were specified for the parameters, based on which the covariance matrices and data were generated. Using the generated covariance matrices and data, the model parameters were estimated and the estimates were compared to the true values to assess the precision and efficiency of the MCMC algorithm.
Figure 4. Histogram plots of the elements of the factor sensitivities matrix estimated in the simulation exercise, against the true model values, represented by straight horizontal lines.
Figure 5. Plot of the log-determinant of the intertemporal sensitivity matrix $A$, the idiosyncratic covariance matrix, Omega, and the mean of all betas, over 10000 iterations of the burn-in stage of the Gibbs sampler. Also plotted are ACF graphs of a subsample of subsequent draws of the Gibbs sampler.
Figure 6. Box plot of the mean of the first factor beta for the 88 securities in the sample, with 5 and 95% bounds represented by vertical lines.

Figure 7. Box plot of the mean of the second factor beta for the 88 securities in the sample, with 5 and 95% bounds represented by vertical lines.
Figure 8. Time series plot of the posterior distribution of the standard deviations of the first factor for all sample periods. We show 5th percentile, mean, and 95th percentile.

Figure 9. Time series plot of the posterior distribution of the standard deviations of second factor for all sample periods. We show 5th percentile, mean, and 95th percentile.

Figure 10. Plot of the correlations between the two factors for all periods in the sample. We show 5th percentile, mean, and 95th percentile.