Modelling the Conditional Covariance Matrix in Stochastic Volatility Models with Applications to the Main Exchange Rates in Poland

1. Introduction

Multivariate models of asset returns are very important in financial applications. Asset allocation, risk assessment and construction of an optimal portfolio require estimates of the covariance matrix between the returns of assets (see e.g. Aguilar and West (2000), Pajor (2005a, 2005b)). Similarly, hedges require a covariance matrix of all the assets in the hedge.

There are two main types of volatility models for asset returns: the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) and the Stochastic Volatility (SV) families. The GARCH models define the time-varying covariance matrix as a deterministic function of past squared innovations and lagged conditional variances and covariances, whereas the conditional covariance matrix in the SV models is treated as an unobserved component that follows some separate multivariate stochastic process. The first multivariate SV model proposed in the literature by Harvey, Ruiz and Shephard (1994) allowed the variances of multivariate returns to vary over time, but constrained the correlations to be constant. Pitt and Shephard (1999) proposed a factor SV model, which allows a parsimonious representation of the time series evolution of covariances when the number of series being modelled is very large. Simple multivariate factor models for SV processes have been suggested, but not applied, by Jacquier, Polson and Rossi (1995, 1999). Tsay (2002) proposed the SV process based on the Cholesky decomposition of the conditional covariance matrix. A practical drawback of stochastic volatility

---

1 Research supported by a grant from Cracow University of Economics.
models is the intractability of the likelihood function. Because the conditional covariance matrix is an unobserved component the likelihood function is only available in the form of a multiple integral. Thus, estimating the parameters of SV models requires numerical methods based on Markov Chain Monte Carlo (MCMC) techniques.

The main goal of the paper is to compare the SV models differ in structure of conditional covariance matrix and in the number of latent processes. We consider the SV models with zero, constant or time-varying conditional correlation coefficient. The other aim of the paper is to check sensitivity of the results of Bayesian model comparison with respect to the ordering of financial instruments in the TSV model (proposed by Tsay (2002)). In order to compare five different SV-type specifications we build VAR(1) model with the disturbances following one of the competing bivariate SV specifications. These models are used to describe the main Polish exchange rates (the daily exchange rates of PLN/USD and PLN/DEM, 6.02.1996 – 31.12.2001, PLN/USD and PLN/EUR, 31.12.2001 – 31.12.2004). In order to obtain posterior distributions of the quantities of interest, we use Markov chain Monte Carlo (MCMC) methods, mainly the Metropolis-Hastings algorithm within the Gibbs sampler to simulate from the posterior distribution (see Gamerman (1997), Pajor (2003, 2005a, 2006) for details).

The structure of the article is as follows. Sections 2 focuses on the description of the competing bivariate SV models. Section 3 presents the posterior results connected with the model comparison. Finally, we give some conclusions in Section 4.

2. Competing Bivariate SV Models

Let \( x_j \) denote the price of asset \( j \) (exchange rate in our application) at time \( t \) for \( j = 1, 2 \) and \( t = 1, 2, \ldots, T \). The vector of growth rates \( y_t = (y_{1,t}, y_{2,t})' \), each defined by the formula \( y_{j,t} = 100 \ln (x_{t,j} / x_{j,t-1}) \), is modelled here using the basic VAR(1) framework:

\[
y_t - \delta = R(y_{t-1} - \delta) + \xi_t, \quad t = 1, 2, \ldots, T
\]

where \( T \) denotes the number of the observations used in estimation. In (1) \( \delta \) is a 2-dimensional vector, \( R \) is a 2x2 matrix of parameters, and \( \xi_t \) is a bivariate SV process. More specifically:

\[
\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \xi_{1,t} \\ \xi_{2,t} \end{bmatrix}, \quad t = 1, 2, \ldots, T.
\]

We assume that, conditionally on vector \( \Omega_{(0)} \) (consisting of model-specific latent variables) and the parameter vector \( \theta \), \( \xi_t \) follows a bivariate Gaussian distribution with mean vector \( 0_{[2 \times 1]} \) and covariance matrix \( \Sigma_t \), i.e.
\( \xi_t \mid \Omega(t), \theta_t \sim N(0_{(2 \times 1)}, \Sigma_t), \ t = 1, 2, \ldots, T \). Competing bivariate SV models are defined by imposing different structures on \( \Sigma_t \).

The elements of \( \delta \) and \( R \) are common parameters. We assume for them the multivariate standardised Normal prior \( N(0, I_6) \), truncated by the restriction that all eigenvalues of \( R \) lie inside the unit circle. These parameters and the remaining (model-specific) parameters are a prior independent.

2.1. Stochastic Discount Factor Model – SDF

The first specification considered here is the stochastic discount factor model (SDF) proposed by Jacquier, Polson and Rossi (1995). The SDF process is defined as follows:

\[
\xi_t = \epsilon_t / h_t, \quad \ln h_t = \phi \ln h_{t-1} + \sigma_h \eta_t, \quad \epsilon_t \sim iN(0_{(2 \times 1)}, \Sigma), \quad \eta_t \sim iN(0, 1),
\]

\( \epsilon_{j,t} \perp \eta_s, t, s \in \mathbb{Z}, j = 1, 2 \).

Here \( \{\epsilon_t\} \) is a sequence of independent and identically distributed normal random vectors with mean vector zero and constant covariance matrix \( \Sigma \). Thus, we have

\[ \xi_t \mid \Omega(t), \theta_t \sim N(0_{(2 \times 1)}, h_t \Sigma), \text{ where } \Omega(t) = h_t. \]

The conditional covariance matrix of \( \xi_t \) is time varying and stochastic, but all its elements have the same dynamics governed by \( h_t \):

\[
\Sigma_t = h_t \Sigma = \begin{bmatrix} h_t \sigma_{11} & h_t \sigma_{12} \\ h_t \sigma_{21} & h_t \sigma_{22} \end{bmatrix}.
\]

(2)

Thus, the conditional correlation coefficient is time invariant:

\[
\rho_{12,t} = \rho = \frac{\sigma_{12}^t}{\sqrt{\sigma_{11}^t \sigma_{22}^t}}.
\]

(3)

In order to complete the Bayesian model, we have to specify a prior distribution on the parameter space. We assume the following prior structure:

\[ p(\phi, \sigma_h^2, \ln h_0, \Sigma) = p(\phi) p(\sigma_h^2) p(\ln h_0) p(\Sigma), \]

where we use proper prior densities of the following distributions:

\( \phi \sim N(0, 100) I_{(-1,1)}(\phi) \), \( \sigma_h^2 \sim IG(1, 0.005) \), \( \ln h_0 \sim N(0, 100) \), \( \Sigma \sim IW(2I, 2, 2) \).

The prior distribution for \( (\phi, \sigma_h^2)' \) is the same as in the univariate SV model (see Pajor (2003)). We impose stationarity of \( \ln h_t \) by truncating the prior for \( \phi \). This implies that the support of \( \phi \) is \((-1, 1)\) – the region of stationarity \( I((-1,1)) \) denotes the indicator function of the interval \((-1,1))\). The symbol \( IG(v_0, s_0) \) denotes the inverse Gamma distribution with mean \( s_0/v_0 \) (for \( v_0 > 0 \)) and variance \( s_0^2/[(v_0 - 1)^2(v_0 - 2)] \) (thus, here the prior mean for \( \sigma_h^2 \) does not exist, but \( \sigma_h^2 \) has a Gamma prior with mean 200 and variance 40000). The symbol \( IW(B, d, 2) \)
denotes the two-dimensional inverse Wishart distribution with \(d\) degrees of freedom and parameter matrix \(B\). \(\ln h_0\) is treated as an additional parameter and estimated jointly with other parameters. The prior distribution used are relatively noninformative.

### 2.2. Basic Stochastic Volatility Model – BSV

Next, we consider the basic stochastic volatility process (BSV), where \(\xi_{1,t}\) and \(\xi_{2,t}\) follow independent univariate SV processes:

\[
\xi_t \mid \Omega_{(2)}, \theta_i \sim N(0_{(2 \times 1)}, \Sigma_i),
\]

where

\[
\Sigma_i = \text{Diag}(h_{1,j}, h_{2,j}).
\]

The conditional variance equations are

\[
\begin{align*}
\ln h_{1,j} - \gamma_{11} &= \phi_{11}(\ln h_{1,j-1} - \gamma_{11}) + \sigma_{11}\eta_{1,t}, \\
\ln h_{2,j} - \gamma_{22} &= \phi_{22}(\ln h_{2,j-1} - \gamma_{22}) + \sigma_{22}\eta_{2,t},
\end{align*}
\]

where \(\eta_t = (\eta_{1,t}, \eta_{2,t})'\), \(\eta_t \sim iiN(0_{(2 \times 1)}, I_2)\), \(\Omega_{(2)} = (h_{1,j}, h_{2,j})'\).

For the parameters we use the same specification of prior distribution as in the univariate SV model (see Pajor (2003)), i.e. \(\gamma_{jj} \sim N(0, 100)\), \(\phi_{jj} \sim N(0, 100)\).

In this case, the conditional correlation is equal to zero. Many studies find that this assumption is not supported by most financial data. Thus, there is a need to extend the BSV model to incorporate time-varying correlations.

### 2.3. Bivariate JSV(2) Model

Now, we propose a SV process based on the spectral decomposition of the matrix \(\Sigma_i\). That is

\[
\Sigma_i = P \Lambda_i P',
\]

where \(\Lambda_i = \text{Diag}(\lambda_{1,j}, \lambda_{2,j})\) is the diagonal matrix consisting of all eigenvalues of \(\Sigma\) and \(P\) is the matrix consisting of the eigenvectors of \(\Sigma\). For series \(\{\ln \lambda_{1,j}\}\) \((j = 1, 2)\), similarly as in the univariate SV process, we assume standard univariate autoregressive processes of order one, namely

\[
\begin{align*}
\ln \lambda_{1,j} - \gamma_{11} &= \phi_{11}(\ln \lambda_{1,j-1} - \gamma_{11}) + \sigma_{11}\eta_{1,t}, \\
\ln \lambda_{2,j} - \gamma_{22} &= \phi_{22}(\ln \lambda_{2,j-1} - \gamma_{22}) + \sigma_{22}\eta_{2,t},
\end{align*}
\]

where \(\eta_t = (\eta_{1,t}, \eta_{2,t})'\) and \(\eta_t \sim iiN(0_{(2 \times 1)}, I_2)\), \(\Omega_{(3)} = (\lambda_{1,j}, \lambda_{2,j})'\).

Log transformation for \(\lambda_{j,t}\) is used to ensure the positiveness of \(\Sigma\). The matrix \(\Sigma\) is positive definite if \(\lambda_{j,t} > 0\) for \(j = 1, 2\), which is achieved by modelling \(\ln \lambda_{j,t}\) instead of \(\lambda_{j,t}\). If \(|\phi_{jj}| < 1\) \((j = 1, 2)\) then \{\ln \lambda_{1,j}\} and \{\ln \lambda_{2,j}\} are...
stationary and the JSV(2) process is a white noise. In addition, $P$ is an orthogonal matrix, i.e. $PP' = I$. Without loss of generality, we can assume that

$$ P = \begin{bmatrix} p_{11} & \sqrt{1-p^2_{11}} \\ \sqrt{1-p^2_{11}} & -p_{11} \end{bmatrix}, \quad p_{11} \in (0, 1]. $$

Using equation (5), we obtain the conditional covariance matrix of $\xi_t$, which can be written as:

$$ \Sigma_t = \begin{bmatrix} \lambda_{1,t}p^2_{11} + \lambda_{2,t}(1-p^2_{11}) & (\lambda_{1,t} - \lambda_{2,t})p_{11}\sqrt{1-p^2_{11}} \\ (\lambda_{1,t} - \lambda_{2,t})p_{11}\sqrt{1-p^2_{11}} & \lambda_{2,t}p^2_{11} + \lambda_{4,t}(1-p^2_{11}) \end{bmatrix}. \quad (6) $$

Consequently, using (6), we obtain the conditional correlation coefficient, which is time-varying and stochastic if $p_{11} \neq 1$:

$$ \rho_{12,t} = \frac{(\lambda_{1,t} - \lambda_{2,t})p_{11}\sqrt{1-p^2_{11}}} {\sqrt{\lambda_{1,t}^2 p^2_{11}(1-p^2_{11}) + \lambda_{4,t}^2 \lambda_{2,t}^2}} \quad \text{for each } t = 1, 2, \ldots, T. \quad (7) $$

For the model-specific parameters we take the following prior distributions:

$$ \gamma_{ij} \sim N(0, 100), \quad \phi_{ij} \sim N(0, 100)\mathcal{I}_c, \quad \sigma^2_{ij} \sim IG(1, 0.005), \quad \ln \lambda_{j,0} \sim N(0, 100), j = 1, 2; \quad p_{11} \sim U(0, 1) \quad \text{(i.e. uniform over (0, 1)).} $$

Note that if $p_{11}=1$, then we obtain the BSV model, but we formally exclude this value.

### 2.4. Bivariate JSV(3) Model

In the JSV(2) model the structure of the conditional covariance matrix is based on two separate latent variables. The next specification uses three separate latent processes (thus called JSV(3)). In the definition of the JSV(2) model we replace $p_{11}$ by a process $p_{11,t}$ with value in (0,1). Thus, we have:

$$ \ln \lambda_{1,t} - \gamma_{11} = \phi_{11}(\ln \lambda_{1,t-1} - \gamma_{11}) + \sigma_{11}\eta_{11,t}, $$

$$ \ln \lambda_{2,t} - \gamma_{22} = \phi_{22}(\ln \lambda_{2,t-1} - \gamma_{22}) + \sigma_{22}\eta_{22,t}, $$

$$ w_i - \gamma_{21} = w_{i-1} - \gamma_{21} + \sigma_{21}\eta_{21,t}, \quad w_i = \ln[p_{11,t} / (1-p_{11,t})], $$

$$ \eta_{t} = (\eta_{11,t}, \eta_{22,t}, \eta_{21,t}), \quad \eta_{t} \sim iN(0_{3 \times 1}, I_3), \quad \Omega_{t(i)} = (\lambda_{1,t}, \lambda_{2,t}, p_{11,t})'. $$

Now the number of the latent processes is equal to the number of distinct elements of the conditional covariance matrix. Here we have:

$$ \Sigma_t = \begin{bmatrix} \lambda_{1,t}p^2_{11} + \lambda_{2,t}(1-p^2_{11}) & (\lambda_{1,t} - \lambda_{2,t})p_{11}\sqrt{1-p^2_{11}} \\ (\lambda_{1,t} - \lambda_{2,t})p_{11}\sqrt{1-p^2_{11}} & \lambda_{2,t}p^2_{11} + \lambda_{4,t}(1-p^2_{11}) \end{bmatrix}. \quad (8) $$

We assume the following prior distributions: $\gamma_{ij} \sim N(0, 100), \quad \phi_{ij} \sim N(0, 100)\mathcal{I}_c, \quad \sigma^2_{ij} \sim IG(1, 0.005), \quad \ln \lambda_{j,0} \sim N(0, 100), j, i \in \{1,2\}, i \geq j; \quad w_0 \sim N(0, 100).$

### 2.5. Bivariate TSV model
Note that, the JSV(3) model does not allow the covariance to evolve over time “independently” of the variances (see equation (8): each element of $\Sigma_t$ depends on all latent variables). The next specification (proposed by Tsay (2002), thus called TSV) uses the Cholesky decomposition of the conditional covariance matrix:

$$\Sigma_t = L_t G_t L_t^T,$$

(9)

where $L_t$ is a lower triangular matrix with unitary diagonal elements, $G_t$ is a diagonal matrix with positive diagonal elements:

$$L_t = \begin{bmatrix} 1 & 0 \\ q_{21,t} & 1 \end{bmatrix}, \quad G_t = \begin{bmatrix} q_{11,t} & 0 \\ 0 & q_{22,t} \end{bmatrix}.$$

Series $\{q_{21,t}\}$, and $\{\ln q_{jj,t}\} (j=1,2)$, analogous to the univariate SV, are standard univariate autoregressive processes of order one, namely

$$\ln q_{11,t} - \gamma_{11} = \phi_{11} (\ln q_{11,t-1} - \gamma_{11}) + \sigma_{11} \eta_{11,t},$$

$$\ln q_{22,t} - \gamma_{22} = \phi_{22} (\ln q_{22,t-1} - \gamma_{22}) + \sigma_{22} \eta_{22,t},$$

$$q_{21,t} - \gamma_{21} = \phi_{21} (q_{21,t-1} - \gamma_{21}) + \sigma_{21} \eta_{21,t},$$

where $\eta_t = (\eta_{11,t}, \eta_{21,t}, \eta_{22,t})'$ and $\eta_t \sim iN(0, I_3), \Omega(5) = (q_{11,t}, q_{22,t}, q_{21,t})'$. From the decomposition in (9), we have:

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t}^2 & \sigma_{12,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t}^2 & \sigma_{22,t} \\ \sigma_{21,t} & \sigma_{22,t} & \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} q_{11,t} & q_{11,t} q_{21,t} \\ q_{11,t} q_{21,t} & q_{11,t} q_{21,t} + q_{22,t} \end{bmatrix}.$$

Consequently, the conditional correlation coefficient between $\xi_t$ and $\xi_{2,t}$ is as follows:

$$\rho_{12,t} = q_{21,t} \sqrt{\frac{q_{11,t}}{q_{11,t}^2 + q_{22,t}^2}}$$

for each $t = 1, 2, ..., T$. (10)

We make similar assumptions about the prior distributions as previously. In particular: $\gamma_t \sim N(0, 100), \phi_t \sim N(0, 100) I_{l=1,1}(\phi), \sigma_{ij}^2 \sim IG(1, 0.005), \ln q_{ii,t} \sim N(0, 100) I_{i,j \geq j}, q_{21,t} \sim N(0, 100)$.

A major drawback of this process is that the conditional variances are not modeled in a symmetric way, thus the explanatory power of model may depend on the ordering of financial instruments.

3. Empirical Results

In order to compare competing bivariate SV – type specifications we use two sets of financial data: the growth rates of the PLN/USD and PLN/DEM, which Osiewalski and Pipień (2004, 2005) analysed using bivariate GARCH – type specifications and the growth rates of the PLN/USD and PLN/EUR. The first data set represents the daily exchange rate of the German mark against the Polish zloty and the US dollar against the Polish zloty from February 5, 1996 to
December 31, 2001 (1482 modelled observations for each series). The second data set consists of two daily exchange rate series, namely, the euro against the Polish zloty and the US dollar against the Polish zloty from February 2, 2002 to December 31, 2004 (758 modelled observations for each series). The data were downloaded from the website of the National Bank of Poland.

Table 1. Logs of Bayes factors in favour of JSV(3) model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TSV&lt;sub&gt;USD_DEM&lt;/sub&gt; (TSV&lt;sub&gt;USD EUR&lt;/sub&gt;)</td>
<td>3</td>
<td>18</td>
<td>5.326</td>
<td>0.400</td>
</tr>
<tr>
<td>TSV&lt;sub&gt;DEM USD&lt;/sub&gt; (TSV&lt;sub&gt;EUR USD&lt;/sub&gt;)</td>
<td>3</td>
<td>18</td>
<td>24.303</td>
<td>0.263</td>
</tr>
<tr>
<td>JSV(3)</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>JSV(2)</td>
<td>2</td>
<td>15</td>
<td>20.505</td>
<td>9.168</td>
</tr>
<tr>
<td>BSV</td>
<td>2</td>
<td>14</td>
<td>128.487</td>
<td>47.711</td>
</tr>
<tr>
<td>SDF</td>
<td>1</td>
<td>12</td>
<td>97.370</td>
<td>21.078</td>
</tr>
</tbody>
</table>

The decimal logarithms of the Bayes factors in favour of JSV(3), calculated for the two data sets using the Newton and Raftery’s (1994) method, are shown in Table 1. Because in the TSV specification the conditional variances are not modelled in a symmetric way, we consider two cases: TSV<sub>USD DEM</sub> (respectively TSV<sub>USD EUR</sub>) and TSV<sub>DEM USD</sub> (respectively TSV<sub>EUR USD</sub>). These models differ in ordering of elements in \( y_t \). In the TSV<sub>USD DEM</sub> (respectively TSV<sub>USD EUR</sub>) model \( y_{1,t} \) denotes the daily growth rate of the PLN/USD exchange rate at time \( t \), \( y_{2,t} \) is the daily growth rate of the PLN/DEM (respectively PLN/EUR) exchange rate at time \( t \). In the TSV<sub>DEM USD</sub> (respectively TSV<sub>EUR USD</sub>) model the ordering of components in \( y_t \) is contrary to previous one. The empirical results show (see Table 1) that the explanatory power of TSV model depends on the ordering of components in \( y_t \). The TSV<sub>USD DEM</sub> model is about 19 orders of magnitude more probable a posterior than the TSV<sub>DEM USD</sub> model. Furthermore, the TSV model with “wrong” ordering of financial time series (i.e. TSV<sub>DEM USD</sub>) fits the data worse than the JSV(2) model, which describe the three distinct elements of the conditional covariance matrix by two separate latent processes. Thus, the explanatory power depends not only on the number of latent processes, but also the structure of the conditional covariance matrix.

We see that for both data sets the JSV(3) model wins our model comparison. But it is important to stress that the JSV structure is difficult to use in higher dimensions. The results indicate that the data reject the constant or zero conditional correlation hypothesis, represented by the SDF and BSV specifications. The BSV and SDF models are inadequate – they are much worse than the TSV and JSV(2), JSV(3) models. In case of the growth rates of the PLN/USD and PLN/DEM the decimal log of the Bayes factor of the BSV model relative to the JSV(3) model is 128. Assuming equal prior model probabilities, the SDF model (with the constant conditional correlation) is about
31 orders of magnitude more probable a posterior than the BSV model, but about 77 orders of magnitude worse than the JSV(2) model and about 92 orders of magnitude worse than the $\text{TSV}_{\text{USD,DEM}}$ model.

The ranking obtained for the growth rates of the PLN/USD and PLN/EUR ($T=758$ observations) is different. The models with as many latent processes as there are conditional variances and covariances receive practically all posterior probability mass. The JSV(2) model, with the number of latent processes equal to the dimension of the modelled time series, is about 9 orders of magnitude less probable a posterior than the JSV(3) model. The $\text{TSV}_{\text{EUR,USD}}$ fits the data worse than $\text{TSV}_{\text{USD,EUR}}$, but not as poorly as the JSV(2) model. We see that the longer series confirm very clearly inadequacy of the BSV and SDF models; the two SV specifications with zero or constant conditional correlation coefficient are strongly rejected. The distances (measured by the Bayes factor) between the best model and the BSV and SDF models become smaller when we use the shorter time series.

Of course, our model comparison relies on the prior distributions for the parameters of the models. It seems that these prior distributions are not very informative - they are quite diffuse.

4. Conclusions

In this article we used the main Polish exchange rates to compare various bivariate SV-type specifications using their Bayes factors. We considered five bivariate SV models, including the specification with zero, constant and time-varying conditional correlation. The competing bivariate stochastic volatility models differ in assumption on conditional correlation and in the number of latent processes. The results indicate that the most adequate specifications are those that allow for time-varying conditional correlation and that have as many latent processes as there are conditional variances and covariances. The empirical results show that the explanatory power of TSV model depends on the ordering of modelled financial instruments.

References


