The instability of the correlation structure of the S&P 500

Štefan Lyócsa and Tomáš Výrost and Eduard Baumöhl

Faculty of Business Economics in Košice, University of Economics in Bratislava

17. October 2011
The instability of the correlation structure of the S&P 500

Štefan LYÓCSA - Tomáš VÝROST - Eduard BAUMÖHL*

Abstract
Using weekly returns of S&P 500 constituents, we study the time-varying correlation structure during the period of 2006 to mid-2011. Contrary to most of the previous correlation studies of many assets, we do not use rolling correlations but the DCC MV-GARCH model with the MacGyver strategy proposed by Engle (2009). We find empirical evidence that the correlation structure tends to change significantly during the periods of high volatility and market downturns.

Key words: correlation structure, dynamic conditional correlations, range-based volatility, conditional volatility, MacGyver strategy

JEL classification: C32, G1
I. Introduction

Volatility and correlation between assets are crucial for option pricing and portfolio and risk management. We study the changes in the correlation structure between stock returns of current and past constituents of the S&P 500 index and their relationship to market volatility. We define the correlation structure as a totally ordered set of correlation coefficients \( (\rho, \leq) \). If this ordering changes in time \( t+1 \), the correlation structure changed. This has important implications for portfolio analysis. For example, in a basic Markowitz model, when the correlation structure changes, so will the estimated weights in the portfolio.

Empirical studies of the relationship between market volatility and market correlations have shown that, in periods of high volatility, correlations between stock portfolio returns tend to increase (e.g. Ding et al., 2011). This is unfortunate, as diversification is needed most when uncertainty is high. The aim of this study is to examine empirically the relationship between changes in stock market volatility and the instability of correlation structure (ICS).

Similar approaches in measuring the ICS can be found in the literature on stock market networks. The correlation matrix \( C_t \) of asset returns may be represented as a complete graph, with the vertices being stocks and the weighted edges the relationships between stock returns. As the topological properties of this graph are not that interesting, a Minimum Spanning Tree (MST) is usually constructed from \( C_t \), which retains all \( N \) vertices but only \( N-1 \) edges¹ (e.g. Onnela et al., 2003). Denote \( E(t) \) the set of edges of MST, then \( \sigma(t) = |E(t) \cap E(t-1)|/(N-1) \) measures the ICS represented by MSTs (Onnela et al., 2003). However, some information from \( C_t \) is being ignored because \( N(N-1)/2 \) edges are squeezed into a graph with \( N-1 \) edges (Tse et al., 2010). We, therefore, do not follow this approach².

It is well known that correlation coefficients are distorted due to heteroskedasticity in the data. More importantly, when there are structural breaks accompanied by an increase in volatility, estimated correlations are biased upwards. Contrary to the standard approach based on rolling correlations in stock market network studies, we use the DCC MV-GARCH model of Engle and Sheppard (2001) to calculate \( C_t \). To mitigate the known negative bias in high-dimension DCCs, we use the MacGyver strategy of Engle (2009).

¹ Technically, the correlation matrix is first transformed to a distance matrix with elements \( d_{ij} = (2(1-\rho_{ij}))^{0.5} \).
² The Threshold Asset Graph in Tse et al. (2010) solves some issues but still ignores some information from \( C_t \).
II. Data and Methodology

Our analysis is based on weekly stock price returns calculated from Wednesday-to-
Wednesday\(^3\) closing prices \((P_t)\) of companies that were listed as S&P 500 constituents at some
point within 3 January 2006 to 22 July 2011; thus, \(T = 290\). The period was chosen because of
the high volatilities coupled with the latest recession. As only companies with all daily prices
available for the given period were considered, we started with \(N = 516\) companies.

**Correlations**

Each series \(i = 1, 2, \ldots, 516\) of stock returns \(R_{t,i} = \ln\left(P_{t,i}/P_{t-1,i}\right)\) was subject to unit root
testing. First, we used the DF-GLS test without the trend component, with MAIC lag
selection, where the maximum lag order \(k_{\text{max}} = \left\lfloor 12(T/100)^{1/4} \right\rfloor\). The finite sample critical
values (for \(T = 250\)) were taken from Cook and Manning (2004). When the null hypothesis of
unit root could not be rejected, we continued with the Lee and Strazicich (2004) test with one
break in the level of the series. The number of the augmented terms was chosen according to
Ng and Perron (1995) by first selecting the maximum lag order to \(k_{\text{max}}\) and then reducing the
number of lags until the coefficient on the last lag remained significant. The trimming
parameter for break detection was set at 0.1. Sample size and break dates specific critical
values were obtained from a Monte Carlo simulation with 2500 replications.

Next, we estimated the mean equations for each return series to obtain residuals
without autocorrelation. We used an ARMAX model in the following form\(^4\):

\[
R_t = \beta_1 + \beta_2 D U_t + z_t \\
\left(1 - \sum_{k=1}^{p} \phi_k L^k \right) \epsilon_t = \left(1 + \sum_{l=1}^{q} \theta_l L^l \right) \epsilon_t \tag{1}
\]

where \(\epsilon_t \sim \text{N}(0, \sigma^2)\), \(z_t\) is the stationary ARMA component and \(DU_t = 1\) for \(t > T_b\) and
\(DU_t = 0\) otherwise, with \(T_b\) denoting the break date. If in the previous step the series was
determined to be stationary without a break, the \(\beta_2 DU_t\) was dropped. The orders \(p, q = 1, 2, \ldots, 7\) were selected according to AIC from a set of \(\{p, q\} \ldots\) for which the Ljung-Box test
on residuals from Equation 1 signalled nonrejection of the null hypothesis of no
autocorrelation for up to 0.1\(T\) lags.

---

\(^3\) This allows us to avoid possible day-of-the-week effects.

\(^4\) Series specific subscripts are omitted for brevity.
For the estimation of dynamic conditional correlations, we have followed the work of Engle and Sheppard (2001). For each pair of residuals \( r_t = (\varepsilon_{1,t}, \varepsilon_{2,t})' \) obtained from the previous step, it is assumed that:

\[
\begin{align*}
\mathbf{r}_t &\mid \mathbf{Q}_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) \\
\mathbf{H}_t &\equiv \mathbf{D}_t \mathbf{C}_t \mathbf{D}_t
\end{align*}
\]

(2) (3)

where \( \mathbf{H}_t \) is a decomposed variance-covariance matrix and \( \mathbf{D}_t \) is a diagonal matrix of time-varying SDs from univariate GED-GARCH models (of orders up to 4, to capture the ARCH effects). The usual GARCH restrictions for non-negativity and stationarity were imposed. \( \mathbf{C}_t \) is the time-varying correlation matrix:

\[
\mathbf{C}_t = \text{diag} \{ \mathbf{Q}_t \}^{-1/2} \mathbf{Q}_t \text{diag} \{ \mathbf{Q}_t \}^{-1/2}
\]

(4)

\[
\mathbf{Q}_t = \left(1 - \sum_{p=1}^\alpha \alpha_p - \sum_{q=1}^\beta \beta_q\right) \mathbf{Q} + \sum_{p=1}^\alpha \alpha_p (\mathbf{s}_{t-p} \mathbf{s}_{t-p}') + \sum_{q=1}^\beta \beta_q \mathbf{Q}_{t-q}
\]

(5)

\[
\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}}, \ i, j = 1, 2, K, n; i \neq j
\]

(6)

where \( \mathbf{s}_t \) are standardized residuals, \( \mathbf{Q} \) is the unconditional correlation matrix in dynamic correlation structure \( \mathbf{Q}_t \) and \( \rho_{i,j,t} \) are the DCCs.

Estimation of DCCs with large dimensions is problematic especially when \( N > T \). One possibility is to estimate only bivariate DCCs, which still seems a preferable alternative to the rolling correlations used in stock market network studies. We followed Engle (2009), who proposed a better approach, the MacGyver methodology, where \( \alpha_q \) and \( \beta_q \) are estimated as medians of corresponding coefficients obtained from bivariate DCC MV-GARCH models for all series.

**Stock market volatility**

Stock market volatility was estimated from S&P 500 data. First, a Garman and Klass (1980) range-based unconditional volatility (\( \sigma_{GK}^2 \)) with jump adjustment \( j_{it} \) (Molnár, 2010) was used. If \( O, H, L, C \) denote opening, highest, lowest and closing prices in the given week, then \( h_t = \ln(H_t) - \ln(O_t) \), \( l_t = \ln(L_t) - \ln(O_t) \), \( c_t = \ln(C_t) - \ln(O_t) \) and \( j_t = \ln(O_t) - \ln(C_{t-1}) \):

\[
\sigma_{GK}^2 = 0.5(h_t - l_t)^2 - (2\ln 2 - 1)c_t^2 + j_t^2
\]

(7)

Second, we estimated the conditional variance (\( \sigma_{\text{GARCH}}^2 \)) from S&P 500 weekly returns using a standard GARCH(1,1) model following the procedure described above.
**Correlation structure**

Two sets of correlation coefficients are defined as \((\rho_{t-1}, \leq)\) and \((\rho_{t}, \leq)\). Both have \(N(N-1)/2\) components, with \(\{\rho_{ij};\ i < j, \ i = 1, 2, ..., N-1, \ j = 2, ..., N\}\). Coefficients in both sets can be ranked separately with ranks \(R\rho_{ij,t-1}\) and \(R\rho_{ij,t}\). The ordered pair of observations \((R\rho_{ij,t-1}, R\rho_{ij,t})\) corresponds to the ranks of correlation coefficients between stocks \(i\) and \(j\). If \(R\rho_{ij,t-1} \neq R\rho_{ij,t}\), then ICS is present. To measure the extent of instability from time \(t-1\) to \(t\), we use Kendall’s Tau \((K_t)\) coefficient. ICS is observed if \(-1 \leq K_t < 1\), with lower values indicating greater instability.

As in Equation 1, we regressed the \(K_t\) on the level shift variable\(^5\) \(DU_t\), the difference of volatility estimate (either \(\Delta \sigma_{GK,t}^2 = \sigma_{GK,t}^2 - \sigma_{GK,t-1}^2\) or \(\Delta \sigma_{GARCH,t}^2 = \sigma_{GARCH,t}^2 - \sigma_{GARCH,t-1}^2\)) and a variable \(ST_t\), which takes into account the presence of a short-term downturn in stock markets. We assumed that ICS would be more significant for market downturns, partly because of higher volatility, but primarily due to the often unique and unexpected events accompanying downturns that are difficult to take into account when pricing assets (see also Ding et al., 2011). The variable \(ST_t\) counts the number of negative weekly returns \((R_t)\) in the last seven weeks\(^6\). The final regressions were:

\[
K_t = \alpha_0 + \alpha_1 DU_t + \alpha_2 \Delta \sigma_{GK,t}^2 + \alpha_3 ST_t + e_t
\]

(8)

\[
K_t = \alpha'_0 + \alpha'_1 DU_t + \alpha'_2 \Delta \sigma_{GARCH,t}^2 + \alpha'_3 ST_t + e'_t
\]

(9)

**III. Results and Discussion**

For some series, we were unable to fit ARMAX or GARCH models. Therefore, the final sample was \(N = 496\) series. From Figure 1, it can be seen that the largest changes in the correlation structure occurred at the end of 2008 (note that lower values indicates greater instability).

---

\(^5\) Using the unit root testing procedure described above, \(K_t\) was found to be stationary with a break in the level of the series.

\(^6\) The statistical significance of changes in volatility estimates was not sensitive to alternative calculations of the short-term variable, namely five and nine weeks.
The results in Table 1 – Panel A suggest that, at least in our sample, during the increase of conditional volatility, changes in the correlation structure were more apparent. In stock market downturns, this instability of correlation structure seems to increase further. The change of unconditional volatility was not significant. This suggests that more than just one week of increasing volatility is needed to observe significant changes in the correlation structure.
Visual inspection of Figure 2 indicates that our results might be driven by the extremely high volatility in the short period from 8 September 2008 to 15 December 2008 (the highlighted region in all plots). In Panels B and C, Equations 8 and 9 are estimated in subsamples prior to and after this extreme event. For both samples, only conditional variance was significant.

Table 1: Regression coefficient estimated from Equations 8 and 9

<table>
<thead>
<tr>
<th>Dependent $K_t$</th>
<th>Intercept</th>
<th>$DU_t$</th>
<th>$ST_t$</th>
<th>$\Delta\sigma^2_{GK,t}$</th>
<th>$\Delta\sigma^2_{GARCH,t}$</th>
<th>$R^2_{adj.}$</th>
<th>$F$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A</strong></td>
<td>sample (11.01.2006 - 20.07.2011, T = 289)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 8</td>
<td>0.9153***</td>
<td>0.0149***</td>
<td>-0.0089***</td>
<td>1.2280</td>
<td></td>
<td>0.1650</td>
<td>7.1394***</td>
</tr>
<tr>
<td>(0.0082)</td>
<td>(0.0054)</td>
<td>(0.0029)</td>
<td>(1.1067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 9</td>
<td>0.9057***</td>
<td>0.0162***</td>
<td>-0.0058**</td>
<td>-5.1487***</td>
<td></td>
<td>(0.3225)</td>
<td>0.4718</td>
</tr>
<tr>
<td>(0.0070)</td>
<td>(0.0048)</td>
<td>(0.0022)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td>sample (11.01.2006 - 03.09.2008, T = 139)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 8</td>
<td>0.9241***</td>
<td>-0.0093***</td>
<td>10.3061</td>
<td></td>
<td></td>
<td>0.0880</td>
<td>4.3638**</td>
</tr>
<tr>
<td>(0.0088)</td>
<td>(0.0034)</td>
<td>(8.8294)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 9</td>
<td>0.91802***</td>
<td>-0.0073**</td>
<td>-135.336***</td>
<td></td>
<td></td>
<td>(22.4598)</td>
<td>0.2261</td>
</tr>
<tr>
<td>(0.0096)</td>
<td>(0.0035)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C</strong></td>
<td>sample (24.12.2008 - 20.07.2011, T = 135)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 8</td>
<td>0.8838***</td>
<td>0.0320***</td>
<td>-0.0031</td>
<td>0.2573</td>
<td></td>
<td>0.2086</td>
<td>5.7249***</td>
</tr>
<tr>
<td>(0.0112)</td>
<td>(0.0098)</td>
<td>(0.0019)</td>
<td>(2.2402)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation 9</td>
<td>0.8735***</td>
<td>0.0377***</td>
<td>-0.0015</td>
<td>-5.3921***</td>
<td></td>
<td>(0.4053)</td>
<td>0.6280</td>
</tr>
<tr>
<td>(0.0069)</td>
<td>(0.0049)</td>
<td>(0.0016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: OLS robust estimation (HAC); SE are in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively. Sample size corresponds to the reductions due to $K_t$ and $DU_t$ variables. $K_t$ breaks at $T_b = 18.05.2009$ and was therefore dropped from regressions reported in Panel B. The $\Delta\sigma^2_{GK,t}$ was found to be stationary at 0.05% significance using the DF-GLS test.

We conclude that with the increase of conditional stock market volatility, the correlation structure becomes more unstable. This instability seems to be much stronger during extreme market conditions and is intensified further in short-term market downturns.

**Acknowledgement**

The support of the VEGA project no. 1/0826/11 is acknowledged.
References


