The instability of the correlation structure of the SP 500

Lyócsa, Štefan and Výrost, Tomáš and Baumöhl, Eduard

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[a] – Tomáš VÝROST[b] – Eduard BAUMÖHL[c]**

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

[a] *Department of Business Informatics and Mathematics*

[b] *Department of Finance and Accounting*

[c] *Department of Economics*

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

Tajovského 13, 04130 Košice, Slovakia

* *Corresponding author. E-mail: eduard.baumohl@euke.sk*

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 ($\boldsymbol{\rho}_t$, $\leq$). 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 $\mathbf{C}_t$ 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_t) is usually constructed from $\mathbf{C}_t$, which retains all N vertices but only N–1 edges[1] (e.g. Onnela *et al.*, 2003). Denote $E(t)$ the set of edges of MST_t, then $\sigma(t) = |E(t) \cap E(t-1)|/(N-1)$ measures the *ICS* represented by *MST*s (Onnela *et al.*, 2003). However, some information from $\mathbf{C}_t$ 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 approach[2].

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 $\mathbf{C}_t$. 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 $d_{i,j} = (2(1-\rho_{i,j}))^{0.5}$.
[2] The Threshold Asset Graph in Tse *et al.* (2010) solves some issues but still ignores some information from $\mathbf{C}_t$.

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 $R_{t,i} = \ln(P_{t,i}/P_{t-1,i})$ 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 $k_{\max} = \left\lfloor 12(T/100)^{1/4} \right\rfloor$. 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 $k_{\max}$ 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]:

$$
\begin{aligned}
&R_t = \beta_1 + \beta_2 DU_t + z_t \\
&\left(1 - \sum_{k=1}^{p} \phi_k L^k\right) z_t = \left(1 + \sum_{l=1}^{q} \theta_l L^l\right) \varepsilon_t
\end{aligned}
\tag{1}
$$

where $\varepsilon_t \sim \mathrm{N}(0,\sigma^2)$, 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 $0.1T$ 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 $\mathbf{r}_t = (\varepsilon_{i,t},\varepsilon_{j,t})^T$ obtained from the previous step, it is assumed that:

$$\mathbf{r}_t \mid \Omega_{t-1} \sim N(\mathbf{0},\mathbf{H}_t) \quad (2)$$

$$\mathbf{H}_t \equiv \mathbf{D}_t\mathbf{C}_t\mathbf{D}_t \quad (3)$$

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:

$$\mathbf{C}_t = diag\{\mathbf{Q}_t\}^{-1/2}\mathbf{Q}_t diag\{\mathbf{Q}_t\}^{-1/2} \quad (4)$$

$$\mathbf{Q}_t = \left(1-\sum_{p=1}^{P}\alpha_p - \sum_{q=1}^{Q}\beta_q\right)\overline{\mathbf{Q}} + \sum_{p=1}^{P}\alpha_p\left(\mathbf{s}_{t-p}\mathbf{s}_{t-p}^T\right) + \sum_{q=1}^{Q}\beta_q\,\mathbf{Q}_{t-q} \quad (5)$$

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}\,q_{j,j,t}}},\ i,j = 1,2,\ldots,n; i \neq j \quad (6)$$

where $\mathbf{s}_t$ are standardized residuals, $\overline{\mathbf{Q}}$ 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 α_p and β_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 ($\hat{\sigma}^2_{GK,t}$) with jump adjustment j_t (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 $j_t = ln(O_t)–ln(C_{t-1})$:

$$\hat{\sigma}^2_{GK,t} = 0.5(h_t - l_t)^2 - (2\ln 2 - 1)c_t^2 + j_t^2 \quad (7)$$

Second, we estimated the conditional variance ($\hat{\sigma}^2_{GARCH,t}$) 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 ($\boldsymbol{\rho}_{t-1}$, $\leq$) and ($\boldsymbol{\rho}_t$, $\leq$). Both have $N(N-1)/2$ components, with $\{\rho_{i,j}: i < j, i = 1,2,\ldots,N-1, j = 2,\ldots,N\}$. Coefficients in both sets can be ranked separately with ranks $R\rho_{i,j,t-1}$ and $R\rho_{i,j,t}$. The ordered pair of observations ($R\rho_{i,j,t-1}$, $R\rho_{i,j,t}$) corresponds to the ranks of correlation coefficients between stocks i and j. If $R\rho_{i,j,t-1} \neq R\rho_{i,j,t}$, 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 $\Delta\hat{\sigma}^2_{GK,t} = \hat{\sigma}^2_{GK,t} - \hat{\sigma}^2_{GK,t-1}$ or $\Delta\hat{\sigma}^2_{GARCH,t} = \hat{\sigma}^2_{GARCH,t} - \hat{\sigma}^2_{GARCH,t-1}$) 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:

$$K_t = \alpha_0 + \alpha_1 DU_t + \alpha_2 \Delta\hat{\sigma}^2_{GK,t} + \alpha_3 ST_t + e_t \tag{8}$$

$$K_t = \alpha'_0 + \alpha'_1 DU_t + \alpha'_2 \Delta\hat{\sigma}^2_{GARCH,t} + \alpha'_3 ST_t + e'_t \tag{9}$$

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.

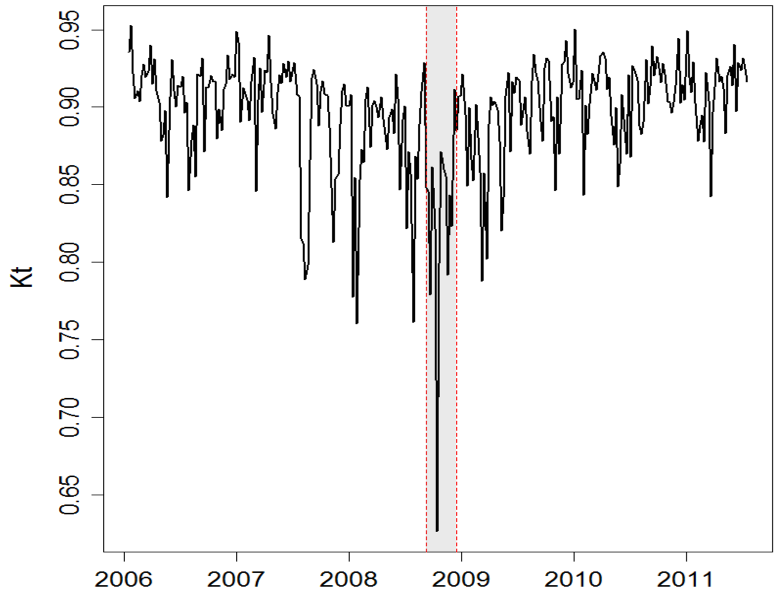

Figure 1: Instability of Correlation Structure

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.

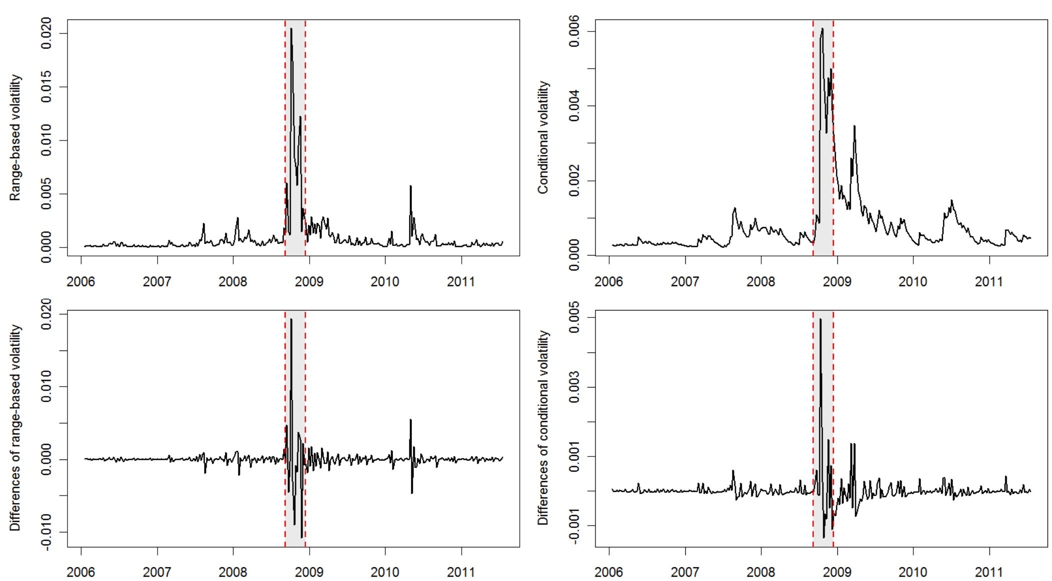

Figure 2: Stock market volatility estimates and their differences

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

Dependent K_t	*Intercept*	*DU_t*	*ST_t*	$\Delta\hat{\sigma}^2_{GK,t}$	$\Delta\hat{\sigma}^2_{GARCH,t}$	*R^2 adj.*	*F-stat*
Panel A	**sample (11.01.2006 - 20.07.2011, T = 289)**						
Equation 8	0.9153^{***}	0.0149^{***}	-0.0089^{***}	1.2280		0.1650	7.1394^{***}
	(0.0082)	(0.0054)	(0.0029)	(1.1067)			
Equation 9	0.9057^{***}	0.0162^{***}	-0.0058^{**}		-5.1487^{***}	0.4718	162.2073^{***}
	(0.0070)	(0.0048)	(0.0022)		(0.3225)		
Panel B	**sample (11.01.2006 - 03.09.2008, T = 139)**						
Equation 8	0.9241^{***}		-0.0093^{***}	10.3061		0.0880	4.3638^{**}
	(0.0088)		(0.0034)	(8.8294)			
Equation 9	0.91802^{***}		-0.0073^{**}		-135.336^{***}	0.2261	27.4709^{***}
	(0.0096)		(0.0035)		(22.4598)		
Panel C	**sample (24.12.2008 - 20.07.2011, T = 135)**						
Equation 8	0.8838^{***}	0.0320^{***}	-0.0031	0.2573		0.2086	5.7249^{***}
	(0.0112)	(0.0098)	(0.0019)	(2.2402)			
Equation 9	0.8735^{***}	0.0377^{***}	-0.0015		-5.3921^{***}	0.6280	101.7606^{***}
	(0.0069)	(0.0049)	(0.0016)		(0.4053)		

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 $\Delta\hat{\sigma}^2_{GK,t}$ 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

Cook, S. and Manning, N. (2004) Lag optimization and finite-sample size distortion of unit root tests, *Economics Letters*, **84**, 267–274.

Ding, L., Miyake, H. and Zou, H. (2011) Asymmetric correlations in equity returns: a fundamental-based explanation, *Applied Financial Economics*, **26**, 389–399.

Engle, R. F. (2009) High dimension dynamic correlations, in *The Methodology and Practice of Econometrics: Festschrift for David Hendry*, J. L. Castle & N. Shephard, Oxford University Press, Oxford, pp. 122–148.

Engle, R. F. and Sheppard, K. (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, NBER Working Paper No. 8554, University of California, San Diego.

Garman, M. B. and Klass, M. J. (1980) On the Estimation of Security Price Volatilities from Historical Data, *Journal of Business*, **53**, 67–78.

Lee, J. and Strazicich, M. (2004) Minimum LM Unit Root Test with One Structural Break. Working Paper No. 04–17, Appalachain State University, Boone.

Molnár, P. (2010) Properties of range-based volatility estimators, in *Econophysics Colloquium*, Academia Sinica, Taipei.

Ng, S. and Perron, P. (1995) Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag, *Journal of the American Statistical Association*, **90**, 268–281.

Onnela, J. P., Chakraborti, A., Kaski, K., Kertész, J. and Kanto, A. (2003) Dynamics of market correlations: Taxonomy and portfolio analysis, *Physical Review E*, **68**, 056110.

Tse, Ch. K., Liu, J. and Lau, F. C. M. (2010) A network perspective of the stock market, *Journal of Empirical Finance*, **17**, 659–667.