Constructing weekly returns based on daily stock market data: A puzzle for empirical research?

Baumöhl, Eduard and Lyócsa, Štefan

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

26 December 2012

Constructing weekly returns based on daily stock market data: A puzzle for empirical research?

Eduard Baumöhl* – Štefan Lyócsa

Faculty of Business Economics in Košice

University of Economics in Bratislava, Slovak Republic

* E-mail contact: eduard.baumohl@euke.sk

Abstract

The weekly returns of equities are commonly used in the empirical research to avoid the non-synchronicity of daily data. An empirical analysis is used to show that the statistical properties of a weekly stock returns series strongly depend on the method used to construct this series. Three types of weekly returns construction are considered: (i) Wednesday-to-Wednesday, (ii) Friday-to-Friday, and (iii) averaging daily observations within the corresponding week. Considerable distinctions are found between these procedures using data from the S&P500 and DAX stock market indices. Differences occurred in the unit-root tests, identified volatility breaks, unconditional correlations, ARMA-GARCH and DCC MV-GARCH models as well. Our findings provide evidence that the method employed for constructing weekly stock returns can have a decisive effect on the outcomes of empirical studies.

Keywords: stock markets, weekly returns, statistical properties

JEL Classification: C10, C80, G10

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11.

Introduction

There are several types of non-synchronicity in daily stock market data. The "non-synchronous trading effect I" arises from the fact that the national stock exchanges are subject to different national, religious, and other holidays, unexpected events, and other occasions. Consequently, within a given period, we will obtain a different number of observations.[1] Another type of non-synchronicity (the so-called "non-synchronous trading effect II") induced by daily data stems from the fact that the national stock exchanges operate in different time zones. Thus, the information set included in the closing prices of equities in Tokyo is obviously distinct from the information set at the end of the same trading day in New York. There is also a third source of non-synchronous trading that is worth mentioning – the last trades of individual stocks may occur at different times.

In many empirical studies regarding the financial markets, the analysis of weekly returns (of individual stocks or market indices) is conducted to avoid the non-synchronous trading effect. Unfortunately, the method for data construction is rarely specified. In this paper, we will show that the method for constructing the weekly series is important. Three methods for constructing the weekly returns are compared (Wednesday-to-Wednesday, week averages and Friday-to-Friday) on a dataset of two stock market indices, namely the US S&P500 and the German DAX30.

We provide empirical evidence of the contrasting statistical properties between the three methods for constructing the weekly returns. We show distinct differences based on descriptive statistics, unit-root tests, correlations, the identification of breaks in volatility, ARMA-GARCH and DCC MV-GARCH models. It may be expected to find some small differences across the returns constructed by different approaches, but as we show, in some cases the results are changed dramatically. Thus, the method for constructing the weekly returns may be partially responsible for the contradictory findings in many empirical works.

[1] One simple matching procedure for daily data was proposed by Baumöhl – Výrost (2010) that ensures that the computed returns will be consecutive. Some easier methods for data imputation may be applied as well (e.g., repeat the last known price).

1. Data and methodology

Our dataset comprises the daily closing prices of the two stock market indices: the US S&P500 (*SPX* henceforth) and the German DAX30 (*DAX* henceforth). The daily data are obtained from Datastream and cover the period from 5 January 1998 to 29 July 2012.

We will apply three different methods for constructing the weekly returns:

- Wednesday-to-Wednesday approach – the returns (r_t) are computed from the Wednesday closing prices P_t, i.e., $r_t = \ln(P_t/P_{t-1})$. In the cases in which Wednesdays were not active trading days, the closing values from the next date with valid prices from the sequence of the nearest days is used: Tuesday, Thursday, Monday, and Friday. More days are usually unnecessary; at least one trading day is always active in a given week. The series obtained by this procedure will be denoted as ***w/w***.
- Week averaging approach – the representative price for the selected week is obtained by averaging all of the available daily closing prices in the given week. The series obtained by this procedure will be denoted as ***ave***.
- Friday-to-Friday approach – the closing prices correspond to Fridays or to the last known closing price in a given week. We have observed that when requesting weekly data, several data providers give the closing prices that correspond to the last trading day of a given week (e.g., Datastream, finance.yahoo.com). In addition, no missing observations are retrieved because at least one price is always available in a given week (if Friday was not an active trading day, Thursday's closing price is imputed and so on). The series obtained by this procedure will be denoted as ***f/f***.

We obtained 760 observations of weekly returns within the selected time span. We decided to split the entire sample into two subsamples (leaving each subsample with 380 observations) to show that our findings are not necessarily related to a specific time period. The 1st sample covers the period from 18 January 1998 to 17 April 2005, and the 2nd covers the period from 24 April 2005 to 29 July 2012 (the dates correspond to the Sundays of the given week). The following figure shows the price evolution of the two indices.

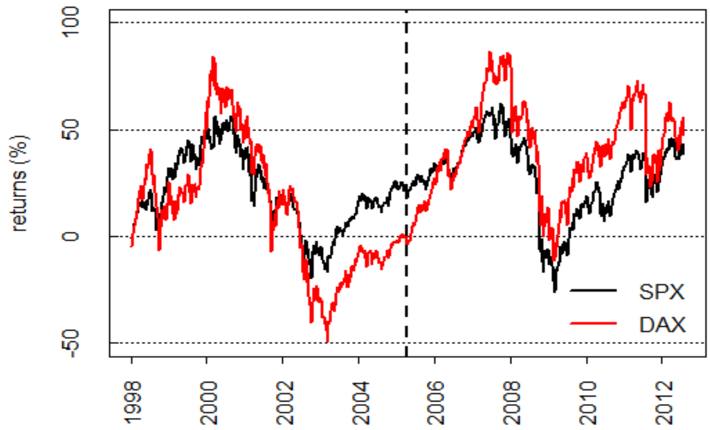

Figure 1: Cumulative returns of the *SPX* and the *DAX*

Note: The dashed vertical line splits the sample into two subsamples. The base date is set to 5 January 1998.

The statistical properties of the 12 resulting series (*SPX_w/w, SPX_ave, SPX_f/f, DAX_w/w, DAX_ave*, and *DAX_f/f* for both subsamples) were compared using the following methods:

- Standard descriptive statistics including autocorrelation and partial autocorrelation functions and normality tests (Anderson – Darling, Shapiro – Wilk and Jarque – Bera).
- Unit-root tests (ADF-GLS and KPSS tests).
- Pearson correlation coefficients.
- ICSS algorithm proposed by Inclán – Tiao (1994) and the κ_2 statistics presented by Sansó et al. (2004) using the critical values from the response surface regressions.[2]
- ARMA(1,1)-GARCH(1,1) models.
- Bivariate DCC MV-GARCH model proposed by Engle – Sheppard (2001) and Engle (2002).

2. Results

This section is divided into three subsections: in Section 2.1, we examine the basic descriptive statistics and unit-root (stationarity) testing; this discussion is followed by Section 2.2, which presents the correlations and the volatility breaks; and in Section 2.3, the ARMA-GARCH models are estimated together with the dynamic conditional correlations (DCC).

[2] The methodological details are omitted for the sake of brevity. Aside from the original work, one may also find further details and the application of this algorithm in Výrost et al. (2011).

2.1 Descriptive statistics and unit-root testing

Table 1 presents some basic descriptive statistics. Some notable differences can be observed. Averaging the daily closing prices to obtain a single week price obviously leads to less volatile series. The maximal and minimal returns differ quite extensively between the three types of weekly returns construction. Also note that the *SPX_w/w* for the 1st subsample has positive skewness while for two remaining procedures, it is negative. Normality is rejected in all cases. The autocorrelation structure of the series also appears to be different because the signs are changing and the values of the autocorrelation coefficient vary significantly. For example, the difference in the AR(1) coefficient between the *DAX_w/w* and the *DAX_ave* is approximately ≈ 0.27. Except for one case, the AR(1) coefficients for *f/f* and *w/w* are negative, while for the *ave*, they are positive, which most likely reflects the fact that the averages smear out the volatility.

Table 1: Descriptive statistics

	1st subsample					
	SPX_w/w	*SPX_ave*	*SPX_f/f*	*DAX_w/w*	*DAX_ave*	*DAX_f/f*
Mean	0.0004	0.0005	0.0006	-0.0001	-0.0001	0.0000
Std.	0.0254	0.0213	0.0256	0.0381	0.0313	0.0361
Min	-0.0904	-0.0854	-0.1233	-0.1522	-0.1358	-0.1392
Max	0.1018	0.0831	0.0749	0.1715	0.1177	0.1289
Skew	0.0517	-0.1496	-0.4875	-0.3056	-0.3133	-0.1646
Kurt	4.3378	4.8207	5.3273	5.2127	4.5292	4.2059
ACF(1)	-0.0850	0.1030	-0.0790	-0.1180	0.1580	0.0580
PACF(2)	-0.0220	-0.0230	0.0380	0.0470	-0.0060	-0.0200
AD	1.1939	1.2076	1.5797	2.4105	1.6239	0.9904
SW	0.9856	0.9787	0.9726	0.9676	0.9787	0.9866
JB	32.2914	44.3186	106.9139	90.5444	36.7213	24.4840
	2nd subsample					
	SPX_w/w	*SPX_ave*	*SPX_f/f*	*DAX_w/w*	*DAX_ave*	*DAX_f/f*
Mean	0.0005	0.0005	0.0005	0.0013	0.0013	0.0013
Std.	0.0262	0.0219	0.0284	0.0325	0.0270	0.0346
Min	-0.1645	-0.1528	-0.2008	-0.1680	-0.1387	-0.2435
Max	0.0964	0.0706	0.1136	0.1094	0.0779	0.1494
Skew	-1.1220	-1.4282	-0.9159	-1.1368	-1.2048	-1.1290
Kurt	9.2133	10.7203	10.6946	7.4105	7.0035	11.2702
ACF(1)	-0.0480	0.1480	-0.0640	-0.0940	0.1270	-0.1110
PACF(2)	-0.0430	0.0060	0.0660	0.0010	0.0280	0.1210
AD	6.5849	7.2980	5.9248	5.7818	5.6553	5.5895
SW	0.9154	0.9030	0.9182	0.9269	0.9278	0.9105
JB	1353.4340	2206.4200	1788.8190	738.3846	635.0392	1897.3580

In the next step, we apply the ADF-GLS unit-root test and the KPSS stationarity test. Covariance stationarity plays a crucial role in time-series econometrics because using non-stationary data may lead to spurious results. Table 2 summarizes the results from the ADF-GLS test and Table 3 summarizes the results from the KPSS test.

Table 2: Results from the ADF-GLS unit-root test

	1st subsample											
	SPX_w/w	*lag*	*SPX_ave*	*lag*	*SPX_f/f*	*lag*	*DAX_w/w*	*lag*	*DAX_ave*	*lag*	*DAX_f/f*	*lag*
$\tau_{\mu,s}^{GLS}$	-3.331	16	-3.259	16	**-0.795**	11	**-1.555**	10	**-1.669**	10	-5.016	9
$\tau_{\mu,m}^{GLS}$	-4.271	10	-4.143	10	**-0.689**	14	**-1.555**	10	**-1.669**	10	-5.016	9
$\tau_{\mu,O}^{GLS}$	-6.111	5	-9.039	2	-4.035	2	**-1.555**	10	-1.669	10	-5.016	9
$\tau_{\mu,u}^{GLS}$	-6.694	6	-6.627	6	-1.168	9	-2.057	7	-2.229	7	-18.038	0
	2nd subsample											
	SPX_w/w	*lag*	*SPX_ave*	*lag*	*SPX_f/f*	*lag*	*DAX_w/w*	*lag*	*DAX_ave*	*lag*	*DAX_f/f*	*lag*
$\tau_{\mu,s}^{GLS}$	**-1.327**	16	-2.958	16	-3.194	16	-4.106	14	-4.130	14	-2.935	15
$\tau_{\mu,m}^{GLS}$	**-1.327**	16	-3.342	14	-3.721	14	-4.106	14	-4.629	13	-3.056	14
$\tau_{\mu,O}^{GLS}$	**-1.327**	16	-2.958	16	-3.194	16	-6.685	5	-4.130	14	-5.904	5
$\tau_{\mu,u}^{GLS}$	-3.146	7	-12.246	1	-20.453	0	-6.685	5	-17.038	0	-5.904	5

Note: We considered four lag selection procedures for the ADF-GLS test: the sequential procedure of Ng – Perron (1995) $\tau_{\mu,s}^{GLS}$, the MAIC as in Ng – Perron (2001) $\tau_{\mu,m}^{GLS}$, the MAIC as in Perron – Qu (2007) $\tau_{\mu,O}^{GLS}$ and a method where the number of lags was determined by testing for no autocorrelation in the residuals of the auxiliary regression $\tau_{\mu,u}^{GLS}$. The bold statistics denotes those where the null hypothesis of the unit-root is rejected at least at a 10% significance level. The critical values for $\tau_{\mu,s}^{GLS}$ and $\tau_{\mu,m}^{GLS}$ are obtained from Cook – Manning (2004). The critical values for $\tau_{\mu,u}^{GLS}$ are calculated from the response surfaces of Cheung – Lai (1995, Table 1). ‡ emphasizes that only the critical values for α = 0.05 and 0.10 were known. For the $\tau_{\mu,O}^{GLS}$, we used the asymptotical critical values as in Elliott et al. (1996).

Although it is possible that one of the four testing procedures may provide different results, our main concern regards the situations where the same procedure yields opposite suggestions on the stationarity property of the weekly return series. While it appears (deductively) obvious that the returns should be mean stationary, these results can mislead the empirical researcher to false conclusions. Table 2 provides some contradictory results. Perhaps the most interesting result is that obtained for the $\tau_{\mu,s}^{GLS}$ and $\tau_{\mu,O}^{GLS}$ test statistics in the 2nd sample for the *SPX* returns. Even for the same lag lengths in the auxiliary regressions, the test statistics differ considerably between the employed time series.

Table 3: Results from the KPSS stationarity test

1st subsample											
SPX_w/w	*BW*	*SPX_ave*	*BW*	*SPX_f/f*	*BW*	*DAX_w/w*	*BW*	*DAX_ave*	*BW*	*DAX_f/f*	*BW*
0.1758	2	0.1843	4	0.2166	7	0.1661	3	0.1688	3	0.1690	1
2nd subsample											
SPX_w/w	*BW*	*SPX_ave*	*BW*	*SPX_f/f*	*BW*	*DAX_w/w*	*BW*	*DAX_ave*	*BW*	*DAX_f/f*	*BW*
0.1321	4	0.1294	3	0.1068	7	0.1327	1	0.1477	3	0.1193	5

Note: The KPSS test procedure (as described in Hobijn et al., 2004) was conducted with the automatic bandwidth selection of Newey – West (1994). "BW" is the bandwidth parameter. The quadratic spectral kernel is applied, but the results remain qualitatively the same as with the Bartlett kernel. Only the intercept is included in the test procedure.

The KPSS test provides more stable results without any conflicting evidence of stationarity/non-stationarity. Despite the differing test statistics and bandwidths, it is consistently suggested that all of the series are stationary.

2.2 Correlations and volatility breaks

We further proceed to a correlation analysis (see Table 4) where the differences between the series of returns that were constructed in these three various manners are perhaps the most convincing. The most notable distortion is visible between the *w/w* and the *f/f* returns. In these cases, the correlations are smaller by the amounts of 0.3972, 0.3278, 0.3246, and 0.3389 in comparison to the correlations between the *w/w* and the *ave* returns. Note that we are actually addressing the same series (*SPX* or *DAX*) and only our approximations of weekly returns are computed in alternative ways. One would naturally expect high correlations; in principle, they should be very close to 1. We may thus conclude that the obtained results are, in fact, surprising.

There are differences in the correlations between the *SPX* and the *DAX* as well; the correlations range from 0.7212 (*SPX_f/f – DAX_ f/f*) to 0.7957 (*SPX_ave – DAX_ ave*) in the 1st subsample and from 0.7978 (*SPX_w/w – DAX_ w/w*) to 0.8709 (*SPX_f/f – DAX_ f/f*) in the 2nd subsample. Some further evidence is provided in Section 2.3.

Table 4: Correlation matrix

1st subsample						
	SPX_w/w	*SPX_ave*	*SPX_f/f*	*DAX_w/w*	*DAX_ave*	*DAX_f/f*
SPX_w/w	–	0.9192	0.5220	0.7513	0.7304	0.4883
SPX_ave	0.9192	–	0.6808	0.7310	0.7957	0.5979
SPX_f/f	0.5220	0.6808	–	0.3680	0.4878	0.7212
DAX_w/w	0.7513	0.7310	0.3680	–	0.9219	0.5941
DAX_ave	0.7304	0.7957	0.4878	0.9219	–	0.7259
DAX_f/f	0.4883	0.5979	0.7212	0.5941	0.7259	–
2nd subsample						
	SPX_w/w	*SPX_ave*	*SPX_f/f*	*DAX_w/w*	*DAX_ave*	*DAX_f/f*
SPX_w/w	–	0.8946	0.5700	0.7978	0.7933	0.5681
SPX_ave	0.8946	–	0.7406	0.7686	0.8690	0.7190
SPX_f/f	0.5700	0.7406	–	0.4531	0.5852	0.8709
DAX_w/w	0.7978	0.7686	0.4531	–	0.9234	0.5845
DAX_ave	0.7933	0.8690	0.5852	0.9234	–	0.7364
DAX_f/f	0.5681	0.7190	0.8709	0.5845	0.7364	–

To identify the breaks in volatility, we employ the ICSS algorithm proposed by Inclán – Tiao (1994) and the κ_2 statistics presented by Sansó et al. (2004). First, we need residuals with no autocorrelation. For that reason, we fit simple autoregressive models where the number of AR(p) terms is determined according to a Ljung – Box test of no autocorrelation in the resulting residuals; we chose the first model where the null of no autocorrelation was not rejected. The maximum number of lagged terms was set to 8.[3]

Table 5 contains the identified breaks in volatility. In the 1st subsample for the *SPX* series, the breaks are the same for *SPX_w/w* and *SPX_ave*. For the *SPX_f/f*, the identified break occurred one week later. According to the algorithm, no breaks occurred for the *DAX* series in the 1st subsample. These results would be quite satisfactory because no crucial distinction in the volatility breaks is observed for the series. However, the situation is markedly different in the 2nd subsample, which was expected to be more volatile because it covers the recent financial crisis. Three breaks in volatility are identified for the *SPX_w/w* and the *SPX_ave* series, and their position is not the same. Additionally, for the *SPX_f/f*, no breaks are found. Similar results are found in the case of the *DAX* series, where four breaks are identified in *DAX_ave*, but no breaks are found in the remaining two *DAX* series.

Table 5: Identified breaks in volatility

	1st subsample	**2nd subsample**
SPX_w/w	23.3.2003	15.7.2007; 14.9.2008; 22.3.2009
SPX_ave	23.3.2003	22.7.2007; 9.3.2008; 26.7.2009
SPX_f/f	30.3.2003	NULL
DAX_w/w	NULL	NULL
DAX_ave	NULL	14.9.2008; 19.4.2009; 8.11.2009; 4.12.2011
DAX_f/f	NULL	NULL

Once again, using the three methods for constructing the weekly returns means that we are actually working with three different time series (with different breaks in volatility).

[3] We also considered the possibility of estimating models for the maximal suggested AR orders for all of the *SPX* and *DAX* series, thus over-fitting some of the series (e.g., when the AR(8) model is suggested for *SPX_w/w*, we fit the AR(8) models for *SPX_ave* and *SPX_f/f* as well, even though a smaller number of AR terms is needed to remove the autocorrelation in these two series). However, the results were qualitatively the same.

2.3 GARCH models and conditional correlations

Arguably the most popular tool in finance and financial econometrics is the GARCH model. We have therefore decided to estimate the standard ARMA(1,1)-GARCH(1,1) models for each of the series (see the detailed results in the Appendix).

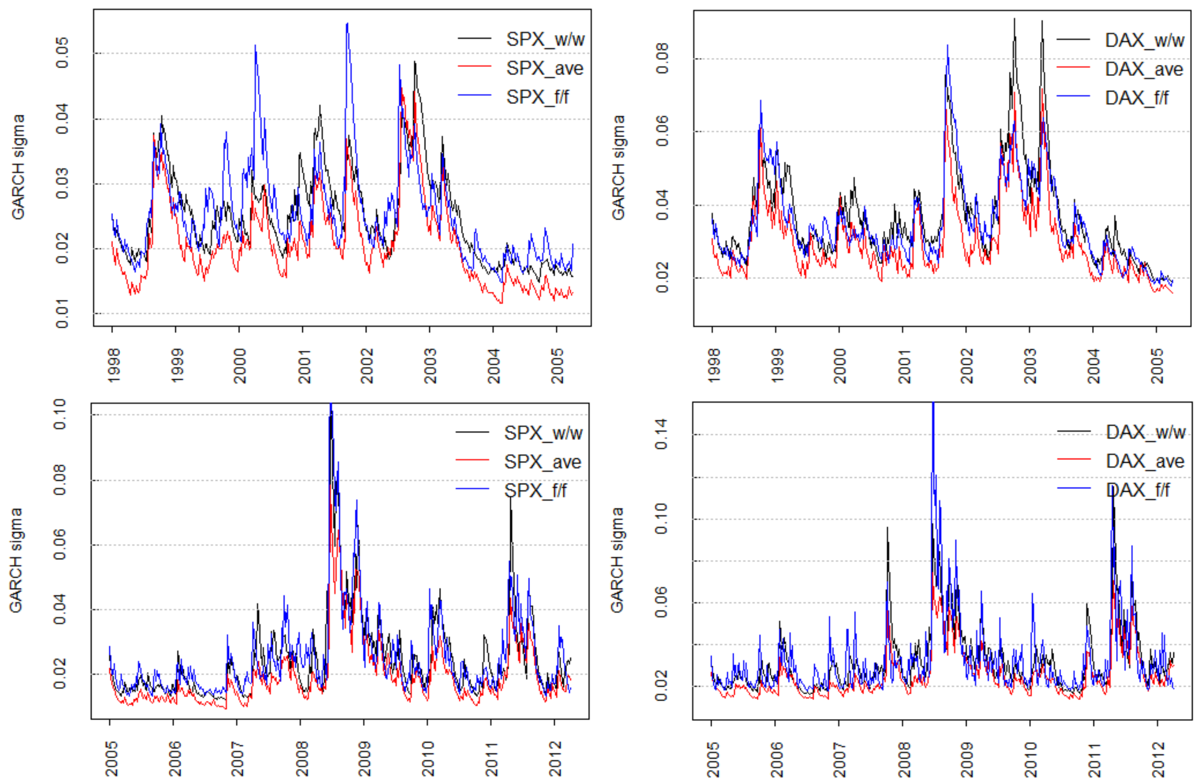

Figure 2: Conditional standard deviations from the ARMA-GARCH models

Note: The charts on the top correspond to the 1st subsample and those at the bottom correspond to the 2nd subsample. To preserve the best possible visibility, the Y-axes are not equally scaled.

Based on the previous results, it is expected that the fitted GARCH model will not be the same. The persistence in volatility ($\alpha_1 + \beta_1$) is similar between the three types of weekly returns; more distortion is visible in the mean equations (ARMA models). Nevertheless, the estimated coefficients are not the same and thus the estimated conditional volatility will also not be the same (see Figure 2) although the dynamics appear to be very similar. Not surprisingly, the conditional volatility is usually lower for the *ave* series, higher for the *w/w*, and much higher for the *f/f* series. During periods of high volatility, the magnitude of the conditional volatility appears to be much different that it is in situations where the accuracy of volatility forecasting is the most needed.

However, in our last step, we estimate the bivariate DCC models to visualize the evolution of the correlations between our three types of weekly returns. We already know that the unconditional constant correlations are markedly different (see Section 2.2). Figure 3 shows that the correlations of the *w/w* and the *ave* with the *f/f* series are lower and much more volatile. However, the correlations between the series constructed by the *w/w* and the *ave* methods are approximately 0.9. Because the *f/f* is the type of weekly returns that are provided by most databases, we believe that these returns are the most commonly utilized in the empirical research.

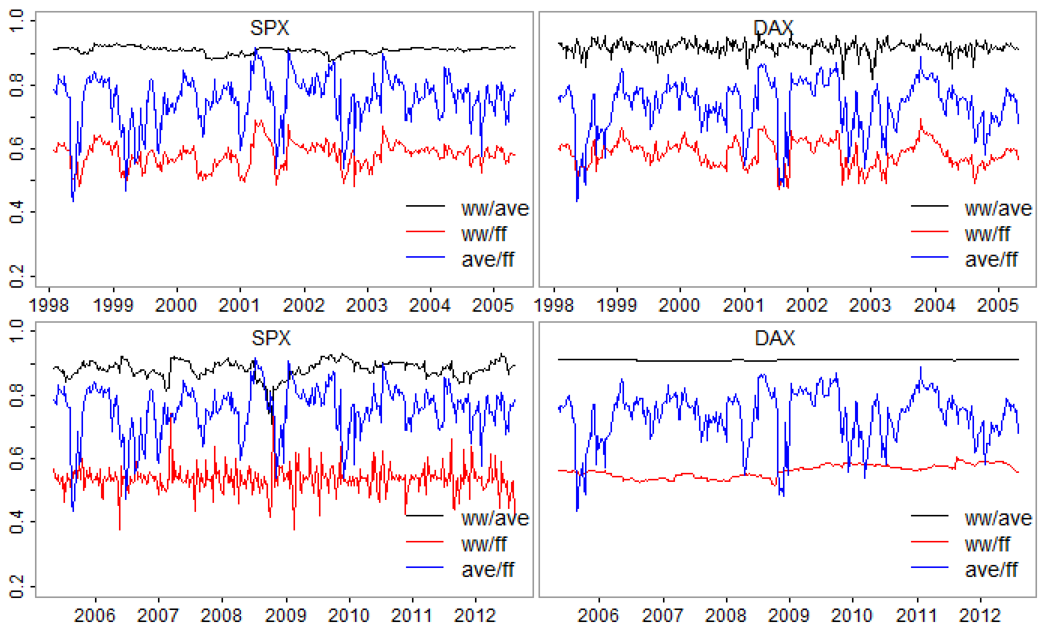

Figure 3: Dynamic conditional correlations between three types of weekly returns

Some differences are also visible in the conditional correlations computed between the *SPX* and the *DAX* series (see Figure 4). In the 1st subsample, these differences are not as strong as in the 2nd subsample. Note that according to the Engle – Ng (1993) Sign Bias test (see Appendix), some asymmetric univariate GARCH models could fit the data better, but we wanted to preserve the same model specification. Particularly in the 2nd sample, one could draw different conclusions regarding the development of the co-movements between these two markets. With the *w/w* data, we observed sharp drops in the correlation during the outburst of the financial crisis that started in the second half of 2007. These drops are not visible in the other two series and these results could have a significant impact on the vast amount of research in the field of stock market integration and contagion (in short, contagion may be defined as a significant increase in correlations during a shock).

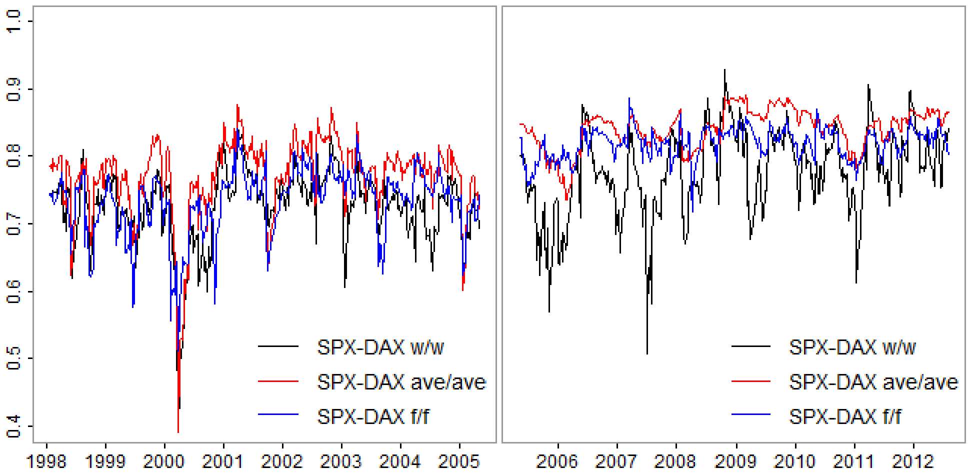

Figure 4: Dynamic conditional correlations between the corresponding *SPX* and *DAX* series

3. Conclusion

We found evidence of qualitatively distinctive statistical properties in the weekly stock returns related to the method used for their construction. In summary, we found interesting differences in the following:

- the autocorrelation structure of the series,
- the unconditional correlations between the series,
- the number of volatility breaks identified for the series,
- the coefficients of the mean equations of the ARMA-GARCH models,
- the magnitude of the conditional volatilities during periods of high volatility,
- the dynamic conditional correlations.

We also provided evidence that the choice of the return series might influence the conclusions drawn about the development of co-movements between the US and the German stock markets. All of these results indicate that the construction of the weekly return series deserves attention.

To this point, we have not discussed the implications of the different methods used to construct the weekly return series. The *f/f* method provided the most diverse returns and was the least correlated with the *ave* and the *w/w* methods. In this series, one can expect significant day-of-the-week effects. It is exactly for that reason that the *w/w* method is often used. Note that the right choice might depend on the goal of the analysis. When we wish to forecast volatilities on a weekly basis, in some instances even the *f/f* series might be useful.

If we want to find the representative price for a given week, the *ave* method appears to be the logical choice. The averaged daily data within the corresponding week should contain more information on the price behavior than any other method considered in this paper. The *ave* smears out the volatility of the daily returns and is less prone to one-time daily effects, which might contaminate the *w/w* series. Intuitively, this method appears to be the most promising. Within our samples, the *ave* was highly correlated with the *w/w* (the alternative) but it also had larger correlations with the *f/f* series than with the *w/w*. Even though our results are strongly sample based, we believe that they will convince the empirical researcher that a careful consideration of the method used for constructing the weekly returns is worth the effort.

References

[1] Baumöhl, E. – Výrost, T. 2010. Stock Market Integration: Granger Causality Testing with respect to Nonsynchronous Trading Effects. *Czech Journal of Economics and Finance*, 60(5), p. 414-25.

[2] Cheung, Y. W. – Lai, K. S. 1995. Lag Order and Critical Values of a Modified Dickey-Fuller Test. *Oxford Bulletin of Economics and Statistics*, 57(3), p. 411-19.

[3] Cook, S. – Manning, N. 2004. Lag optimisation and finite-sample size distortion of unit root tests. *Economics Letters*, 84(2), p. 267-74.

[4] Engle, R. F. – Ng, V. K. 1993. Measuring and Testing the Impact of News On Volatility. *Journal of Finance*, 48(5), pp. 1749-78.

[5] Engle, R. F. – Sheppard, K. 2001. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Paper 8554.

[6] Engle, R. F. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business Economic Statistics*, 20(3), pp. 339-50.

[7] Hobijn, B. – Franses, P. H. – Ooms, M. 2004. Generalizations of the KPSS-test for stationarity. *Statistica* Neerlandica, 58(4), p. 483-502.

[8] Inclán, C. – Tiao, G. 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427), p. 913-23.

[9] Newey, W. K. – West, K. D. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies,* 61(4), p. 631-53.

[10] Ng, S. – 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(429), p. 268-81.

[11] Ng, S. – Perron, P. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica,* 69(6), p. 1519-54.

[12] Perron, P. – Qu, Z. 2007. A simple modification to improve the finite sample properties of Ng and Perron`s unit root tests. *Economics Letters,* 94(1), p. 12-19.

[13] Sansó, A. – Arragó, V. – Carrion, J. 2004. Testing for change in the unconditional variance of financial time series. *Revista de Economiá Financiera*, 4, p. 32-53.

[14] Výrost, T. – Baumöhl, E. – Lyócsa, Š. 2011. On the relationship of persistence and number of breaks in volatility: new evidence for three CEE countries. MPRA Working Paper 27927, University Library of Munich, Germany.

Appendix

Table 6: ARMA-GARCH estimation (1st subsample)

	Estimate	SE	*t*-stat	*p*-value
SPX_w/w				
mu	0.0014	0.0011	1.3598	0.1739
ar1	-0.0022	0.0083	-0.2687	0.7882
ma1	-0.0983	0.0545	-1.8043	0.0712*
omega	0.0000	0.0000	1.3226	0.1860
alpha1	0.1147	0.0372	3.0816	0.0021***
beta1	0.8602	0.0463	18.5804	0.0000***
Q stat(10)	13.1600	LM(10)	8.2684	
Q^2 stat(10)	8.7610	SB test	4.6160	
SPX_ave				
mu	0.0018	0.0010	1.8420	0.0655*
ar1	-0.5060	0.1212	-4.1750	0.0000***
ma1	0.6186	0.1028	6.0160	0.0000***
omega	0.0000	0.0000	1.1766	0.2394
alpha1	0.1445	0.0596	2.4246	0.0153**
beta1	0.8390	0.0641	13.0980	0.0000***
Q stat(10)	9.0140	LM(10)	3.3976	
Q^2 stat(10)	3.4320	SB test	12.6123***	
SPX_f/f				
mu	0.0015	0.0011	1.3520	0.1764
ar1	-0.3464	0.1485	-2.3327	0.0197**
ma1	0.2419	0.1420	1.7035	0.0885*
omega	0.0000	0.0000	1.1759	0.2396
alpha1	0.1491	0.0598	2.4924	0.0127**
beta1	0.8134	0.0805	10.1097	0.0000***
Q stat(10)	11.7800	LM(10)	5.8459	
Q^2 stat(10)	6.0330	SB test	3.1506	

	Estimate	SE	*t*-stat	*p*-value
DAX_w/w				
mu	0.0016	0.0016	1.0095	0.3127
ar1	-0.1602	0.1318	-1.2154	0.2242
ma1	0.0827	0.1560	0.5301	0.5961
omega	0.0001	0.0000	1.3767	0.1686
alpha1	0.1792	0.0497	3.6048	0.0003***
beta1	0.7916	0.0536	14.7803	0.0000***
Q stat(10)	8.6720	LM(10)	6.7200	
Q^2 stat(10)	7.4890	SB test	2.9552	
DAX_ave				
mu	0.0017	0.0016	1.0442	0.2964
ar1	-0.2175	0.1449	-1.5012	0.1333
ma1	0.3861	0.1281	3.0150	0.0026***
omega	0.0001	0.0000	1.8113	0.0701*
alpha1	0.2194	0.0622	3.5300	0.0004***
beta1	0.7303	0.0688	10.6086	0.0000***
Q stat(10)	7.3110	LM(10)	10.1400	
Q^2 stat(10)	9.2190	SB test	7.5650*	
DAX_f/f				
mu	0.0009	0.0017	0.5337	0.5936
ar1	-0.1747	0.3856	-0.4530	0.6506
ma1	0.1897	0.4409	0.4302	0.6671
omega	0.0000	0.0000	0.9268	0.3540
alpha1	0.1612	0.0550	2.9302	0.0034***
beta1	0.8153	0.0700	11.6490	0.0000***
Q stat(10)	11.6400	LM(10)	4.9730	
Q^2 stat(10)	5.2420	SB test	7.7580*	

*Note: "SE" denotes robust standard errors; "Q stat (10)" represents the Ljung – Box test Q statistic of standardized residuals at lag 10; "Q^2 stat (10)" represents the Ljung – Box Q^2 test statistic of squared standardized residuals at lag 10; "LM (10)" is the ARCH LM test at lag 10 and "SB test" contains the test statistic of joint hypothesis from the Sign Bias test of Engle – Ng (1993). The significance codes are denoted as *, **, and *** for the 10%, 5%, and 1% significance levels, respectively.*

Table 7: ARMA-GARCH estimation (2nd subsample)

	Estimate	SE	*t*-stat	*p*-value		Estimate	SE	*t*-stat	*p*-value
SPX_w/w					**DAX_w/w**				
mu	0.0024	0.0010	2.4001	0.0164**	mu	0.0048	0.0017	2.8729	0.0041***
ar1	0.0151	0.1450	0.1042	0.9170	ar1	-0.4593	0.3464	-1.3261	0.1848
ma1	-0.1289	0.1260	-1.0227	0.3065	ma1	0.3511	0.3662	0.9587	0.3377
omega	0.0000	0.0000	2.1392	0.0324**	omega	0.0001	0.0001	1.3151	0.1885
alpha1	0.3275	0.1356	2.4146	0.0158**	alpha1	0.3306	0.1461	2.2623	0.0237**
beta1	0.6588	0.0972	6.7795	0.0000***	beta1	0.6518	0.1092	5.9717	0.0000***
Q stat(10)	7.4450	LM(10)	5.5746		Q stat(10)	12.6100	LM(10)	5.5870	
Q^2 stat(10)	5.6540	SB test	6.0333		Q^2 stat(10)	5.6470	SB test	11.3441***	
SPX_ave					**DAX_ave**				
mu	0.0021	0.0010	2.1944	0.0282**	mu	0.0041	0.0014	2.8212	0.0048***
ar1	-0.1936	0.3108	-0.6229	0.5333	ar1	0.2007	0.2036	0.9856	0.3243
ma1	0.3651	0.2971	1.2291	0.2190	ma1	-0.0128	0.2039	-0.0627	0.9500
omega	0.0000	0.0000	2.3287	0.0199**	omega	0.0001	0.0000	2.6327	0.0085***
alpha1	0.2525	0.0987	2.5571	0.0106**	alpha1	0.2564	0.0940	2.7265	0.0064***
beta1	0.7103	0.0810	8.7732	0.0000***	beta1	0.6942	0.0728	9.5305	0.0000***
Q stat(10)	9.8820	LM(10)	2.2750		Q stat(10)	12.3800	LM(10)	5.0910	
Q^2 stat(10)	2.5740	SB test	8.3715**		Q^2 stat(10)	4.4860	SB test	7.9551**	
SPX_f/f					**DAX_f/f**				
mu	0.0032	0.0010	3.1416	0.0017***	mu	0.0056	0.0012	4.6691	0.0000***
ar1	0.5326	0.3731	1.4276	0.1534	ar1	-0.8890	0.0597	-14.8882	0.0000***
ma1	-0.6117	0.3371	-1.8144	0.0696*	ma1	0.8499	0.0563	15.0908	0.0000***
omega	0.0001	0.0000	2.4550	0.0141**	omega	0.0002	0.0001	2.5988	0.0094***
alpha1	0.3638	0.1363	2.6695	0.0076***	alpha1	0.6179	0.2859	2.1611	0.0307**
beta1	0.5782	0.1068	5.4115	0.0000***	beta1	0.3156	0.1722	1.8332	0.0668*
Q stat(10)	5.7870	LM(10)	8.3570		Q stat(10)	14.1000*	LM(10)	8.9689	
Q^2 stat(10)	8.3950	SB test	14.3349***		Q^2 stat(10)	9.1980	SB test	7.8910**	

*Note: "SE" denotes the robust standard errors; "Q stat (10)" represents the Ljung – Box test Q statistic of standardized residuals at lag 10; "Q^2 stat (10)" represents the Ljung – Box Q^2 test statistic of squared standardized residuals at lag 10; "LM (10)" is the ARCH LM test at lag 10 and the "SB test" contains test statistics of the joint hypothesis from the Sign Bias test of Engle – Ng (1993). The significance codes are denoted as *, **, and *** for the 10%, 5%, and 1% significance levels, respectively.*