DOCUMENTOS DE TRABAJO

BILTOKI

D.T. 2011.04

Wavelet multiple correlation and cross-correlation:
A multiscale analysis of euro zone stock markets.

Javier Fernández-Macho

Documento de Trabajo BILTOKI DT 2011.04
Editado por el Departamento de Economía Aplicada III (Econometría y Estadística)
de la Universidad del País Vasco.

Depósito Legal No.: BI-2095/2011
ISSN: 1134-8984
URN: RePEc:ehu:biltok:201104

WAVELET MULTIPLE CORRELATION AND CROSS-CORRELATION: A MULTISCALE ANALYSIS OF EURO ZONE STOCK MARKETS

JAVIER FERNÁNDEZ-MACHO

ABSTRACT. Statistical studies that consider multiscale relationships among several variables use wavelet correlations and cross-correlations between pairs of variables. This procedure needs to calculate and compare a large number of wavelet statistics. The analysis can then be rather confusing and even frustrating since it may fail to indicate clearly the multiscale overall relationship that might exist among the variables. This paper presents two new statistical tools that help to determine the overall correlation for the whole multivariate set on a scale-by-scale basis. This is illustrated in the analysis of a multivariate set of daily Eurozone stock market returns during a recent period. Wavelet multiple correlation analysis reveals the existence of a nearly exact linear relationship for periods longer than the year, which can be interpreted as perfect integration of these Euro stock markets at the longest time scales. It also shows that small inconsistencies between Euro markets seem to be just short within-year discrepancies possibly due to the interaction of different agents with different trading horizons. On the other hand, multiple cross-correlation analysis shows that the French CAC40 may lead the rest of the Euro markets at those short time scales.

Key words: Euro zone, MODWT, multiscale analysis, multivariate analysis, stock markets, returns, wavelet transform.
JEL Classification: C32, C58, C87, G15.

1. INTRODUCTION

This paper extends wavelet methodology to handle multivariate time series (or, more generally, multivariate ordered variables of two- or three-dimensional support such as spatial data). As their names imply, the wavelet multiple correlation and cross-correlation try to measure the overall statistical relationships that might exist at different time scales among a set of observations on a multivariate random variable. The proposal is justified by noting how the alternative of using standard wavelet correlation analysis usually needs to calculate, plot and compare a large number of wavelet correlation and cross-correlation graphs. For example, in many wavelet studies where the relationships among several variables are considered the wavelet correlation is used between pairs of variables. This needs visualizing graphically the wavelet

Version: v2.0; 20110611.

Javier Fernández-Macho is Professor of Econometrics, Dpt. of Econometrics and Statistics, & Instituto de Economía Pública, University of the Basque Country, Agirre Lehendakari etorb. 83, E48015 BILBAO, Spain (email: javier.fernandezmacho@ehu.es).

correlation values pairwise along the wavelet scales. So if we have n series then we would end up with $n(n-1)/2$ wavelet correlation graphs and J times as many cross-correlation graphs, where J is the order of the wavelet transform. This soon can be quite exhausting and confusing. Besides, at the end, the whole set of graphs most probably will not give a clear indication about the type of overall correlation there exists within the set of series.

In contrast, the proposed wavelet multiple correlation, and similarly its companion wavelet multiple cross-correlation, consists in one single set of multiscale correlations which are not only easier to handle and interpret but also may provide a better insight of the overall statistical relationship about the multivariate set under scrutiny.

All this will be illustrated with the application of the proposed wavelet multiple correlation and cross-correlation in the multiscale analysis of daily returns obtained from a set of eleven Eurozone stock markets during a recent nine year period. In this relation, we may point out how correlation among European stock markets, as a measure of their integration, has attracted quite some interest in the economic and financial literature, especially so ever since the creation of the European Monetary Union (EMU) (see, *e.g.*, Fratzscher, 2002; Yang et al., 2003; Hardouvelis et al., 2006; Syllignakis, 2006; Bartram et al., 2007, and others). However, none of these studies take into account the fact that stock markets involve heterogenous agents that make decisions over different time horizons and operate on different time scales (Gençay et al., 2002, p.10, Gallegati and Gallegati, 2007, Gallegati, 2008). On the other hand, the relatively large number of markets to be analyzed may render pairwise multiscale comparisons pointless in practice, which is the reason why this type of market analysis may find useful the wavelet multiple correlation and cross-correlation proposed here.

The paper is organized as follows. Section 2 defines the proposed wavelet multiple correlation and cross-correlation, whilst Section 3 provides sample estimators for these quantities and establishes their large sample theory. Section 4 gives approximate confidence intervals that can be used for estimation and testing purposes. Simulation results on the validity of the previous results are presented in Section 5. Finally, Section 6 shows the empirical results and Section 7 presents the main conclusions.

2. DEFINITION

Let $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ be a multivariate stochastic process and let $W_{jt} = (w_{1jt}, x_{2jt}, \ldots, w_{njt})$ be the respective scale λ_j wavelet coefficients obtained by applying the maximal overlap discrete wavelet transform (MODWT) (Gençay et al., 2002; Percival and Walden, 2000) to each x_{it} process.

The wavelet multiple correlation (WMC) $\varphi_X(\lambda_j)$ can be defined as one single set of multiscale correlations calculated from X_t as follows. At each wavelet scale λ_j, we calculate the square root of the regression coefficient of determination in that linear combination of variables w_{ijt}, $i = 1, \ldots, n$, for which such coefficient of determination is a maximum. In practice, none of these auxiliary regressions need to be run since, as it is well known, the coefficient of determination corresponding to the regression of a variable z_i on a set of regressors $\{z_k, k \neq i\}$, can most easily be obtained as $R_i^2 = 1 - 1/\rho^{ii}$, where ρ^{ii} is the i-th diagonal element of the inverse of the complete correlation matrix P. Therefore, $\varphi_X(\lambda_j)$ is obtained as

$$\varphi_X(\lambda_j) = \sqrt{1 - \frac{1}{\max \operatorname{diag} P_j^{-1}}}, \tag{1}$$

where P_j is the $(n \times n)$ correlation matrix of W_{jt}, and the $\max \operatorname{diag}(\cdot)$ operator selects the largest element in the diagonal of the argument.

Since the R_i^2 coefficient in the regression of a z_i on the rest of variables in the system can be shown to be equal to the correlation between the observed values of z_i and the fitted values $\widehat{z}_i$ obtained from such regression, we have that $\varphi_X(\lambda_j)$ can also be expressed as

$$\begin{aligned} \varphi_X(\lambda_j) &= \operatorname{Corr}(w_{ijt}, \widehat{w}_{ijt}) \\ &= \frac{\operatorname{Cov}(w_{ijt}, \widehat{w}_{ijt})}{\sqrt{\operatorname{Var}(w_{ijt}) \operatorname{Var}(\widehat{w}_{ijt})}}, \end{aligned} \tag{2}$$

where w_{ij} is chosen so as to maximize $\varphi_X(\lambda_j)$ and $\widehat{w}_{ij}$ are the fitted values in the regression of w_{ij} on the rest of wavelet coefficients at scale λ_j. Hence the adopted name of 'wavelet multiple correlation' for this new statistic. Expression (2) will be useful later in determining the statistical properties of an estimator of $\varphi_X(\lambda_j)$.

It may also be interesting to point out how a multiple correlation statistic is known to be related to the first eigenvalue of the correlation matrix, which indicates the (proportion of) variance of the variables accounted for by a single underlying factor. In fact when all pairwise correlations are positive, this first eigenvalue is approximately a linear function of the average correlation among the variables (Yanai and Ichakawa, 2007; Friedman and Weisberg, 1981; Mayer, 1976).

Finally, allowing a lag τ between observed and fitted values of the variable selected as the criterion variable at each scale λ_j we may also define the wavelet multiple cross-correlation (WMCC) as

$$\begin{aligned}\varphi_{X,\tau}(\lambda_j) &= \operatorname{Corr}(w_{ijt}, \widehat{w}_{ijt+\tau}) \\ &= \frac{\operatorname{Cov}(w_{ijt}, \widehat{w}_{ijt+\tau})}{\sqrt{\operatorname{Var}(w_{ijt})\operatorname{Var}(\widehat{w}_{ijt+\tau})}}.\end{aligned}$$

Of course, for $n = 2$ the WMC and WMCC will coincide with the standard wavelet correlation and cross-correlation. This is because $\operatorname{Cov}(w_{1jt}, \widehat{w}_{1jt}) = \widehat{\beta}_j \operatorname{Cov}(w_{1jt}, w_{2jt})$ and $\operatorname{Var}(\widehat{w}_{1jt}) = \widehat{\beta}_j^2 \operatorname{Var}(w_{2jt})$, where $\widehat{\beta}_j$ is the estimated coefficient in the regression of w_{1jt} on w_{2jt} at scale λ_j. Therefore, $\varphi_X(\lambda_j) = \operatorname{Corr}(w_{1jt}, \widehat{w}_{1jt}) = \operatorname{Corr}(w_{1jt}, w_{2jt}) = \rho_X(\lambda_j)$ and, similarly, $\varphi_{X,\tau}(\lambda_j) = \operatorname{Corr}(w_{1jt}, \widehat{w}_{1jt+\tau}) = \operatorname{Corr}(w_{1jt}, w_{2jt+\tau}) = \rho_{X,\tau}(\lambda_j)$.

3. Estimation

Let $X = \{X_1 \dots X_T\}$ be a realization of the multivariate stochastic process X_t, for $t = 1 \dots T$. Applying a MODWT of order J to each of the univariate time series $\{x_{i1} \dots x_{iT}\}$, for $i = 1 \dots n$, we would obtain J length-T vectors of MODWT coefficients $\widetilde{W}_j = \{\widetilde{W}_{j0} \dots \widetilde{W}_{j,T-1}\}$, for $j = 1 \dots J$.

From (1) the wavelet multiple correlation of scale λ_j is seen to be a nonlinear function of all the $n(n-1)/2$ wavelet correlations of $X_t = (x_{1t}, x_{2t}, \dots, x_{nt})$ at that scale. Alternatively, it can also be expressed in terms of all the wavelet covariances and variances for X_t as in (2).

Therefore, a consistent estimator of the wavelet correlation based on the MODWT is given by

$$
\begin{aligned}
\widetilde{\varphi}_X(\lambda_j) &= \sqrt{1 - \frac{1}{\max \operatorname{diag} \widetilde{P}_j^{-1}}} \\
&= \operatorname{Corr}(\widetilde{w}_{ijt}, \widehat{\widetilde{w}}_{ijt}) \\
&= \frac{\operatorname{Cov}(\widetilde{w}_{ijt}, \widehat{\widetilde{w}}_{ijt})}{\sqrt{\operatorname{Var}(\widetilde{w}_{ijt}) \operatorname{Var}(\widehat{\widetilde{w}}_{ijt})}},
\end{aligned} \tag{3}
$$

where we note that, following Gençay et al. (2002), the wavelet covariances and variances can be estimated as

$$
\operatorname{Cov}(\widetilde{w}_{ijt}, \widehat{\widetilde{w}}_{ijt}) = \bar{\gamma}_j = \frac{1}{\widetilde{T}_j} \sum_{t=L_j-1}^{T-1} \widetilde{w}_{ijt} \widehat{\widetilde{w}}_{ijt} \tag{4a}
$$

$$
\operatorname{Var}(\widetilde{w}_{ijt}) = \bar{\delta}_j^2 = \frac{1}{\widetilde{T}_j} \sum_{t=L_j-1}^{T-1} \widetilde{w}_{ijt}^2 \tag{4b}
$$

$$
\operatorname{Var}(\widehat{\widetilde{w}}_{ijt}) = \bar{\zeta}_j^2 = \frac{1}{\widetilde{T}_j} \sum_{t=L_j-1}^{T-1} \widehat{\widetilde{w}}_{ijt}^2 \tag{4c}
$$

where $\widetilde{w}_{ij}$ is such that the regression of $\widetilde{w}_{ij}$ on the set of regressors $\{\widetilde{w}_{kj}, k \neq i\}$ maximizes the coefficient of determination, $\widehat{\widetilde{w}}_{ij}$ denotes the corresponding fitted values and $L_j = (2^j - 1)(L - 1) + 1$ is the number of wavelet coefficients affected by the boundary associated with a wavelet filter of length L and scale λ_j so that $\widetilde{T}_j = T - L_j + 1$ is the number of coefficients unaffected by the boundary conditions.

Similarly, a consistent estimator of the wavelet multiple cross-correlation $\widetilde{\varphi}_{X,\tau}(\lambda_j)$, can be calculated as

$$
\begin{aligned}
\widetilde{\varphi}_{X,\tau}(\lambda_j) &= \operatorname{Corr}(\widetilde{w}_{ijt}, \widehat{\widetilde{w}}_{ijt+\tau}) \\
&= \frac{\operatorname{Cov}(\widetilde{w}_{ijt}, \widehat{\widetilde{w}}_{ijt+\tau})}{\sqrt{\operatorname{Var}(\widetilde{w}_{ijt}) \operatorname{Var}(\widehat{\widetilde{w}}_{ijt+\tau})}}.
\end{aligned}
$$

The large-sample distribution of the sample wavelet multiple correlation $\widetilde{\varphi}_X(\lambda_j)$ can be established along similar lines as for the standard single wavelet correlation in Gençay et al. (2002). In our present multivariate case, we note from (3) that $\widetilde{\varphi}_X(\lambda_j)$ is a nonlinear function of all the sample wavelet covariances and variances which, in turn, are just sample moments of vectors of MODWT coefficients. Therefore, the estimator can be written as a function of the three moments in (4):

$$\widetilde{\varphi}_X(\lambda_j) = f(\bar{\gamma}_j, \bar{\delta}_j, \bar{\zeta}_j) = \frac{\bar{\gamma}_j}{\bar{\delta}_j\, \bar{\zeta}_j}.$$

We may now apply the continuous mapping theorem to establish that

$$\sqrt{\widetilde{T}_j}\left(\widetilde{\xi}_j - \xi_j\right) \sim \mathcal{N}(0, V_j), \tag{5a}$$

where $\mathrm{abs}(\xi_j) = \varphi_X(\lambda_j)$,

$$V_j = df_j' S_j(0)\, df_j \tag{5b}$$

with df_j as the gradient vector of $f(\gamma_j, \delta_j, \zeta_j)$, and

$$S_j(0) = \begin{pmatrix} S_{\gamma^2,j}(0) & S_{\delta\gamma,j}(0) & S_{\zeta\gamma,j}(0) \\ S_{\delta\gamma,j}(0) & S_{\delta^2,j}(0) & S_{\zeta\delta,j}(0) \\ S_{\zeta\gamma,j}(0) & S_{\zeta\delta,j}(0) & S_{\zeta^2,j}(0) \end{pmatrix}$$

where *e.g.* $S_{\delta\gamma,j}(0)$ is the spectral density function of the product of scale λ_j wavelet moments $\delta_j\gamma_j$ evaluated at the zero frequency, etc. (*cf.* Whitcher, 1998).

4. Confidence intervals

In principle, we can start from the asymptotics obtained in the previous section and use standard procedures in order to construct a confidence interval (CI) for the wavelet multiple correlation $\varphi_X(\lambda_j)$ based on a folded and truncated normal distribution. In practice, however, obtaining the corresponding critical values, let alone calculating the spectral density functions involved in the computation of $\widetilde{V}_j$, can be rather cumbersome. A more feasible alternative can be obtained by using Fisher (1915)'s transformation, since it is a well known normalizing and variance-stabilizing transformation for the otherwise non-Gaussian sample correlation (see, *e.g.*, Johnson et al., 1995, p.571).

Fisher's transformation is defined as $\mathrm{arctanh}(r)$, where $\mathrm{arctanh}(\cdot)$ is the inverse hyperbolic tangent function, and its use in the construction of a CI for a population correlation is based on the fact that if (X,Y) has a bivariate normal distribution with $\rho = \mathrm{Corr}(X,Y)$, then the transformed sample correlation coefficient calculated from T independent pairs of observations can be shown to (approximately) be normally distributed with mean $\mathrm{arctanh}(\rho)$ and variance

$(T-3)^{-1}$ (Fisher, 1921, Johnson et al., 1995, p.572). In our case, we apply the result to the sample wavelet multiple correlation coefficient $\widetilde{\varphi}_X(\lambda_j)$ as follows:

Theorem 1. *Let $X = \{X_1 \ldots X_T\}$ be a realization of a multivariate Gaussian stochastic process $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ and let $\widetilde{W}_j = \{\widetilde{W}_{j0} \ldots \widetilde{W}_{j,T-1}\} = \{(\widetilde{w}_{1j0} \ldots \widetilde{w}_{nj0}), \ldots, (\widetilde{w}_{1j,T/2^j-1} \cdots \widetilde{w}_{nj,T/2^j-1})\}$, $j = 1 \ldots J$, be vectors of wavelet coefficients obtained by applying a MODWT of order J to each of the univariate time series $\{x_{i1} \ldots x_{iT}\}$ for $i = 1 \ldots n$. Let $\widetilde{\varphi}_X(\lambda_j)$ be the sample wavelet correlation obtained from* (1). *Then,*

$$\widetilde{z}_j \overset{a}{\sim} \mathscr{FN}(z_j, (T/2^j - 3)^{-1}),$$

where $z_j = \operatorname{arctanh}(\varphi_X(\lambda_j))$, $\widetilde{z}_j = \operatorname{arctanh}(\widetilde{\varphi}_X(\lambda_j))$ and $\mathscr{FN}$ stands for the folded normal distribution[1].

The demonstration is straightforward since X being Gaussian implies that, at each scale λ_j, the sample wavelet coefficients in $\widetilde{W}_j$ are also Gaussian and, in turn, this means that $\widehat{\widetilde{w}}_{ij}$, which is a linear combination of $\widetilde{w}_{1j}, \ldots, \widetilde{w}_{nj}$, must also be Gaussian. Therefore, we have from (2) that $\widetilde{\varphi}_X(\lambda_j)$ is a correlation coefficient between observations from two Gaussian variates, of which $T/2^j$ are (asymptotically) serially uncorrelated (note that this is the number of wavelet coefficients from a DWT that can be shown to decorrelate a wide range of stochastic processes; Craigmile and Percival, 2005). Applying Fisher's result[2] to ξ_j such that $\operatorname{abs}(\xi_j) = \operatorname{arctanh}(\widetilde{\varphi}_X(\lambda_j))$ the theorem follows. $\square$

Therefore, an approximate $100(1-\alpha)\%$ CI for the true value of $\varphi_X(\lambda_j)$ is

$$CI_{(1-\alpha)}(\varphi_X(\lambda_j)) = \tanh\left[\widetilde{z}_j - c_2/\sqrt{T/2^j - 3}\,;\, \widetilde{z}_j + c_1/\sqrt{T/2^j - 3}\right], \tag{6}$$

where the folded normal critical values $c1, c2$ are such that $\Phi(c_1) + \Phi(c_1 - 2z^0) = 1 - \alpha/2$ and $\Phi(c_2) + \Phi(c_2 + 2z^0) = 2 - \alpha/2$, with $\Phi(\cdot)$ as the standard gaussian probability distribution function and $\tanh(z^0) = \varphi_X^0(\lambda)$ as the value of some wavelet multiple correlation as set under certain null hypothesis. This can be used in practice to construct a confidence interval as well

[1] That is, the probability distribution of $\operatorname{abs}(\xi)$ such that ξ is normally distributed with the said mean and variance. It coincides with the noncentral chi distribution with 1 degree of freedom and noncentrality parameter $\lambda = (T/2^j - 3)^{-1/2} \operatorname{arctanh}(\varphi_X(\lambda_j))$ (see, *e.g.*, Johnson et al., 1995, ch.29).

[2] Note that $\operatorname{sgn}(\operatorname{arctanh}(\cdot)) = \operatorname{sgn}(\cdot)$ and $\widetilde{\varphi}_X(\lambda_j)$ never takes negative values.

as for testing hypothesis about wavelet correlations amongst a multivariate set of observed variables X.

For example, two typical cases of interest to test are whether the variables in X are (1) uncorrelated $H_0 : \varphi_X(\lambda) = 0$ and (2) almost perfectly correlated $H_0 : \varphi_X(\lambda) \to 1$. In the former case, we want to test $H_0 : z^0 = 0$, therefore we would set $c_1 = \phi^{-1}_{1/2-\alpha/4}$ and $c_2 = \phi^{-1}_{1-\alpha/4}$ in (6), where ϕ^{-1}_p is the $100p\%$ point of the standard normal distribution. The relevant test would check whether the lower bound $\widetilde{z}_j - \phi^{-1}_{1-\alpha/4}/\sqrt{T/2^j - 3} > 0$ and therefore we reject that X are uncorrelated. On the other hand, in the second case, we have that $H_0 : z^0 \to \infty$, therefore $c_1 = c_2 = \phi^{-1}_{1-\alpha/2}$, (that is, like for a typical two-sided gaussian test statistic). In this case we would check whether $\widetilde{z}_j - \phi^{-1}_{1-\alpha/2}/\sqrt{T/2^j - 3} > 0.99$ (say) in which case we may infer that X are almost perfectly correlated.

5. Simulations

The rationale for the CI in (6) is the analogy with Fisher's result for the usual bivariate correlation coefficient. We now want to check whether this is still correct when working with multiple wavelet correlation coefficients calculated from more that two variables as in (1).

For this purpose, we run a simulation exercise consisting in drawing 1000 bootstrap samples of size $T = 2454$ from a multivariate Gaussian distribution with mean $\mu = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}'$ and variance $\Sigma = \left[\begin{pmatrix} 2 & 1 & .5 & 1 \end{pmatrix}' \; \begin{pmatrix} 1 & 1 & .2 & 1 \end{pmatrix}' \; \begin{pmatrix} .5 & .2 & 1 & 0 \end{pmatrix}' \; \begin{pmatrix} 1 & 1 & 0 & 3 \end{pmatrix}' \right]$. We then calculated, for the bivariate, three-variate and four-variate cases respectively, the sample moments and quantiles reported in Table 1. We observe that the calculated values for the multiple correlation coefficient $\widetilde{\varphi}_X$ and its $CI_{95\%}$ reflect quite closely the bootstrap distribution in all the three cases whilst, on the other hand, the bootstrap standardized values z_{boot} confirm that $\widetilde{\varphi}_X$ comes from a Gaussian distribution with variance $\sqrt{T-3}$ without further correction related to the number of variates.

6. Eurozone returns

In this section we illustrate the usage of the advocated wavelet multiple correlation with data from the eleven main Eurozone stock markets as follows (arbitrarily ordered by nominal GDP

TABLE 1. Simulation results

		$d=2$	$d=3$	$d=4$
$\widetilde{\varphi}_X$:		0.7022	0.7355	0.7662
	$CI_{95\%}$ lower bound	0.6815	0.7167	0.7494
	$CI_{95\%}$ upper bound	0.7217	0.7531	0.7821
$\widetilde{\varphi}_{X\text{boot}}$:				
	mean	0.7019	0.7354	0.7662
	std. dev.	0.0104	0.0094	0.0082
	quantile(.025)	0.6811	0.7162	0.7504
	quantile(.975)	0.7211	0.7532	0.7826
z_{boot}:				
	mean	-0.0080	0.0045	0.0139
	std. dev.	1.0176	1.0135	0.9890
	skewness	0.0054	0.0124	0.0462
	kurtosis	2.7428	2.7315	2.8630
	J-B p-value	0.2374	0.2065	0.5

$CI_{95\%}$ for $\widetilde{\varphi}_X$ calculated as from (6) with $\alpha = 5\%$, where d is the number of variates. $\widetilde{\varphi}_{X\text{boot}}$ are bootstrap samples of size $T = 2454$ for boot $= 1,\ldots,1000$. $z_{\text{boot}} = \sqrt{(T-3)}\Big(\operatorname{arctanh}(\widetilde{\varphi}_{X\text{boot}}) - \operatorname{arctanh}(\widetilde{\varphi}_X)\Big)$ are the standardized values of $\widetilde{\varphi}_{X\text{boot}}$. J-B is the Jarque and Bera (1987) test statistic of the null hypothesis that z_{boot} comes from a normal distribution.

---

of the country where they operate): DAX (Germany), CAC40 (France), FTSE/MIB30 (Italy), IBEX35 (Spain), AEX25 (Netherlands), NBEL20 (Belgium), ATX20 (Austria), FTSE/ASE20 (Greece), OMXH25 (Finland), PSI20 (Portugal) and ISEQ-Overall (Ireland). The data were collected daily (closing prices) from January 4, 2000 to May 29, 2009[3]. The analysis was conducted using daily stock market returns, *i.e.*, $R_{it} = \log(S_{it}/S_{i,t-1}) = \Delta \log S_{it}$, where S_{it}, $i = 1\ldots 11$, $t = 2\ldots 2455$, are the corresponding stock market index values. Therefore, the total number of observations used is 27005 trading days, containing thus a large amount of information that may not be easy to convey using standard procedures.

In order to calculate the proposed wavelet multiple correlation we need to apply, first of all, the Maximal Overlap Discrete Wavelet Transform (MODWT) to each of the daily stock market returns series (Percival and Walden, 2000). The MODWT is similar to the Discrete

[3]As published by Yahoo `http://finance.yahoo.com`, Euroinvestor `http://www.euroinvestor.co.uk/`, Marketwatch `http://www.marketwatch.com/` and Enet `http://archive.enet.gr/finance/finance.jsp`.

Wavelet Transform (DWT). However, the choice is not arbitrary since the MODWT has some advantages over the classical DWT. To start with, the MODWT can handle any sample size T, whilst the DWT of level J restricts the sample size to a multiple of 2^J. On the other hand, MODWT (wavelet and scaling coefficients) are invariant to circularly shifting the time series under study and its multiresolution detail and smooth coefficients are associated with zero phase filters, two properties that the DWT does not hold. Finally, the MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on DWT, which in turn makes it more suitable when calculating wavelet correlations (Percival and Mofjeld, 1997; Gençay et al., 2002; Percival and Walden, 2000).

In the application, we decomposed the daily stock market returns applying the MODWT with a Daubechies least asymmetric (LA) wavelet filter of length $L = 8$, commonly denoted as $LA(8)$ (Daubechies, 1992; Gençay et al., 2002). This filter appears to be favored mostly in the financial literature (Percival and Walden, 2000; Ranta, 2010). The maximum decomposition level J is given by $\lfloor \log_2(T) \rfloor$ (Gençay et al., 2002; Percival and Walden, 2000), which, in the present case, means a maximum level of 11. Since the number of feasible wavelet coefficients gets critically small for high levels, we chose to carry out the wavelet analysis with $J = 8$ so that eight wavelet coefficients and one scaling coefficient were produced for each daily returns series, *i.e.* $\widetilde{w}_{i1}, \ldots, \widetilde{w}_{i8}$ and $\widetilde{v}_{i8}$ respectively.

We may note that for all families of Daubechies compactly supported wavelets the level j wavelet coefficients are associated with changes at the effective scale $\lambda_j = 2^{j-1}$ (Gallegati, 2008). On the other hand, as the MODWT utilizes approximate ideal band-pass filters with bandpass given by the frequency interval $[2^{-(j+1)}, 2^{-j})$ for $j = 1 \ldots J$, inverting the frequency range we have that the corresponding time periods are $(2^j, 2^{j+1}]$ time units (Whitcher et al., 2000). This means that, with 20 daily data per month, the scales $\lambda_j, j = 1 \ldots 8$, of the wavelet coefficients are associated to periods of, respectively, 2–4 days (which includes most intraweek scales), 4–8 days (including the weekly scale), 8–16 days (fortnightly scale), 16–32 days (monthly scale), 32–64 days (monthly to quarterly scales), 64–128 days (quarterly to biannual scale), 128–256 days (biannual scale) and 256–512 days (annual scale).

Figure 1 shows the wavelet multiple correlation obtained. We observe that the multiple correlations are all quite high, starting at nearly 0.95 for intraweek periods and increasing as the time scale increases, reaching values near 1 at the longest time scales. This means that, when periods of time longer than the year are considered, the existence of an exact linear relationship between Eurozone stock markets cannot be ruled out. This can be interpreted as perfect integration between Euro stock markets in the sense that the returns obtained in any of them can be totally determined by the overall performance in the other markets at horizons longer than the year. In other words, discrepancies between markets are not only small, but also they get dissipated within time horizons smaller than the year. Upon further inspection we also observe that the increase of multiple correlation breaks down for periods between a month and a quarter where they actually momentarily decrease. This means that, together with the otherwise obvious higher daily discrepancies, the differences between Euro markets appear to concentrate along these medium-term time horizons, and may possibly point out to the actuation of different agents across the Euro markets with different trading horizons. We may note in passing that it would be quite hard to reach or justify this conclusion using the standard analysis that relies on the visualization of all the 55 wavelet correlation graphs between pairs of variables. (*cf.* Ranta, 2010, p.29, where he compares 4 stock market returns only).

Figure 2 shows the wavelet multiple cross-correlations obtained for the different wavelet scales with leads and lags up to one month and a half (30 trading days). Each wavelet scale plot shows in its upper-right corner the variable that maximizes the multiple correlation against a linear combination of the rest of variables and, thus, signals a potential leader or follower for the whole system. In our case the data selected CAC40 as such potential leader or follower across all wavelet levels.

As with the contemporaneous multiple correlations, there is a clear tendency to increase multiple correlation as the time horizon gets longer. On the other hand, almost all cross-correlations appear significant at all leads and lags for all levels, with the exception of some leads (negative lags) between 20 and 25 days for the second, third and fourth wavelet scales that are not significant or just marginally significant at the 5% statistical level while the corresponding positive lags are clearly significant. As a consequence, there is a slight asymmetry (right-skewness) that may indicate that CAC40 has a slight inclination to lead the rest of the Euro markets for time scales between one week and one month.

7. CONCLUSIONS

This paper presents two new statistical tools, the wavelet multiple correlation and the wavelet multiple cross-correlation, that may be useful in the wavelet analysis of multivariate time series (or other multivariate ordered data such as multivariate spatial data, etc.) The wavelet multiple correlation consists in one single set of multiscale correlations each of them calculated as the square root of the regression coefficient of determination in that linear combination of wavelet coefficients for which such coefficient of determination is a maximum. The wavelet multiple cross-correlation is obtained similarly by allowing a certain number of lags between observed and fitted values from the same linear combination as before at each of the wavelet scales. We may note that the alternative of using standard wavelet correlation analysis would need to calculate, plot and compare a large number of wavelet correlation and cross-correlation graphs.

Figures 1 and 2 offer some graphical examples of these tools as obtained in the wavelet analysis of a set of 11 times series, namely the returns from the main Eurozone stock markets during a recent period of 2455 trading days.

The wavelet multiple correlation analysis reveals the existence of a nearly exact linear relationship between Eurozone stock markets for periods of time longer than the year, which can be interpreted as perfect integration between Euro stock markets at the longest time scales. It also shows that small inconsistencies between Euro markets seem to be just short and medium term discrepancies that occur as consequence of the interaction of different agents across the Euro markets with different trading horizons in mind. On the other hand, multiple cross-correlation analysis shows that CAC40 may have a small inclination to lead the rest of the Euro markets at those short/medium time scales.

We may finally point out that all these results would be quite hard to establish using the standard wavelet analysis that relies on the visualization of all the 55 wavelet correlation graphs between pairs of variables and they serve to illustrate some of the potential of these new tools in the multiscale analysis of multivariate data.

Supplemental material

The wavemulcor R computer package has been written to facilitate the computation of the wavelet multiple correlation and cross-correlation. It can be obtained from The Comprehensive R Archive Network (CRAN) at `http://cran.r-project.org/web/packages/wavemulcor/index.html` or directly from the author upon request.

References

Bartram, S., Taylor, S., Wang, Y., 2007. The Euro and European financial market dependence. Journal of Banking & Finance 31, 1461–1481.

Craigmile, P., Percival, D., 2005. Asymptotic decorrelation of between-scale wavelet coefficients. IEEE Transactions on Information Theory 51 (3), 1039–1048.

Daubechies, I., 1992. Ten lectures on wavelets. Society for Industrial Mathematics.

Fisher, E. A., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10, 507–521.

Fisher, E. A., 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1, 3–32.

Fratzscher, M., 2002. Financial market integration in Europe: on the effects of EMU on stock markets. International Journal of Finance & Economics 7, 165–193.

Friedman, S., Weisberg, H. F., 1981. Interpreting the first eigenvalue of a correlation matrix. Educational and Psychological Measurement 41, 11–21.

Gallegati, M., 2008. Wavelet analysis of stock returns and aggregate economic activity. Computational Statistics & Data Analysis 52, 3061–3074.

Gallegati, M., Gallegati, M., 2007. Wavelet variance analysis of output in G-7 countries. Studies in Nonlinear Dynamics & Econometrics 11 (3), 1435.

Gençay, R., Selçuk, F., Whitcher, B., 2002. An introduction to wavelets and other filtering methods in finance and economics. Academic Press.

Hardouvelis, G., Malliaropulos, D., Priestley, R., 2006. EMU and European stock market integration. The Journal of Business 79, 365–392.

Jarque, C. M., Bera, A. K., 1987. A test for normality of observations and regression residuals. International Statistical Review 55, 163–172.

Johnson, N. L., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions. Vol. 2. Wiley, New York.

Mayer, E. P., 1976. A measure of the average intercorrelation. Educational and Psychological Measurement 35, 67–72.

Percival, D., Mofjeld, H., 1997. Analysis of subtidal coastal sea level fluctuations using wavelets. Journal of the American Statistical Association 92, 868–880.

Percival, D., Walden, A., 2000. Wavelet methods for time series analysis. Cambridge University Press, reprinted 2006.

Ranta, M., 2010. Wavelet multiresolution analysis of financial time series. Ph.D. thesis, University of Vaasa, Finland, acta Wasaensia 223, Statistics 5.

Syllignakis, M., 2006. EMU's impact on the correlation across the European stock markets. International Research Journal of Finance and Economics 6, 201–219.

Whitcher, B., 1998. Assessing nonstationary time series using wavelets. Ph.D. thesis, University of Washington.

Whitcher, B., Guttorp, P., Percival, D., 2000. Wavelet analysis of covariance with application to atmospheric time series. Journal of Geophysical Research 105, 941–962.

Yanai, H., Ichakawa, M., 2007. Factor analysis. In: Rao, C., Sinharay, S. (Eds.), Psychometrics. Vol. 26 of Handbook of Statistics. North Holland, Amsterdam, Ch. 9.

Yang, J., Min, I., Li, Q., 2003. European stock market integration: does EMU matter? Journal of Business Finance & Accounting 30, 1253–1276.

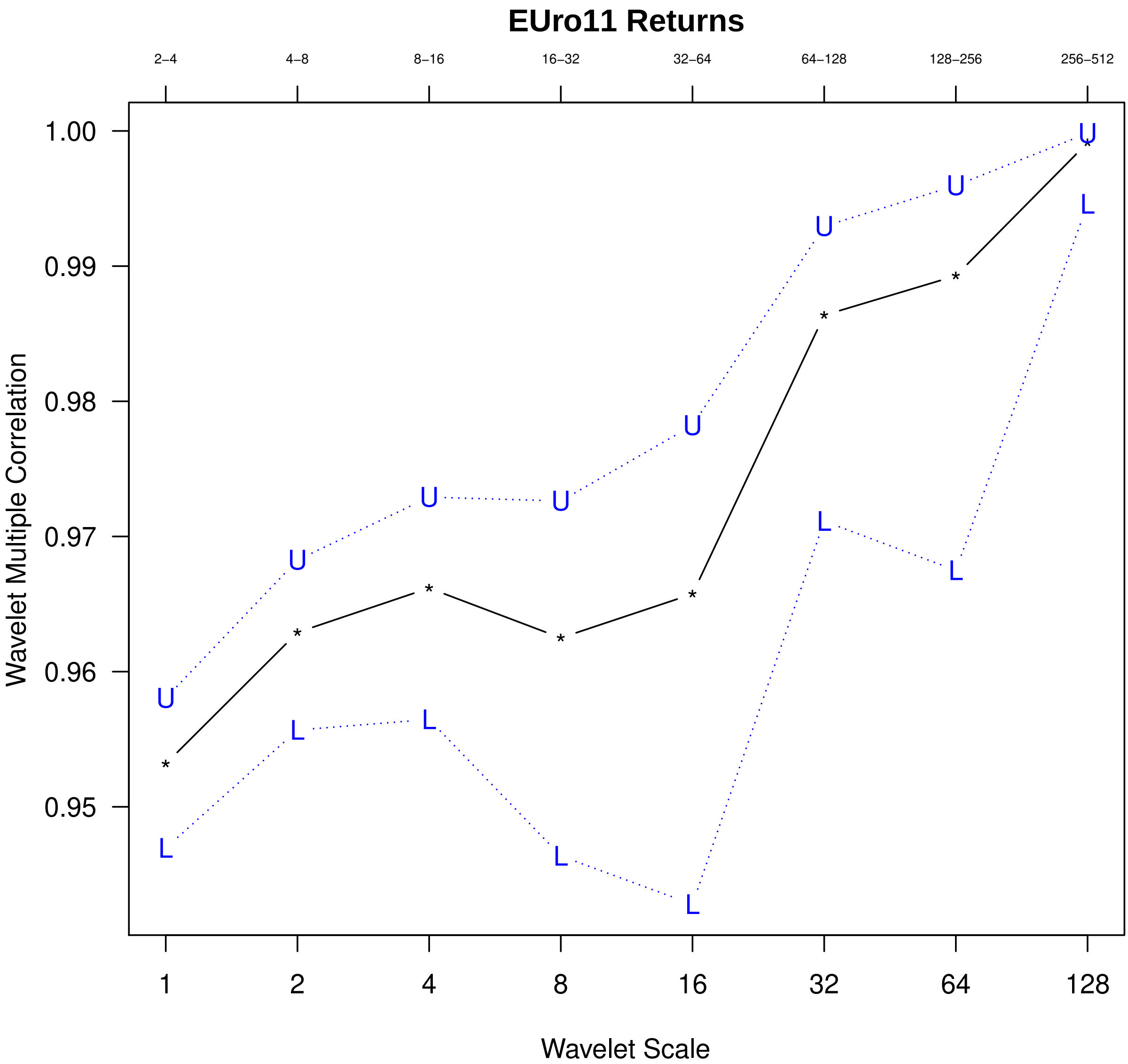

FIGURE 1. Wavelet multiple correlation for the main Eurozone stock markets. The dotted lines correspond to the upper and lower bounds of the corresponding 95% confidence interval.

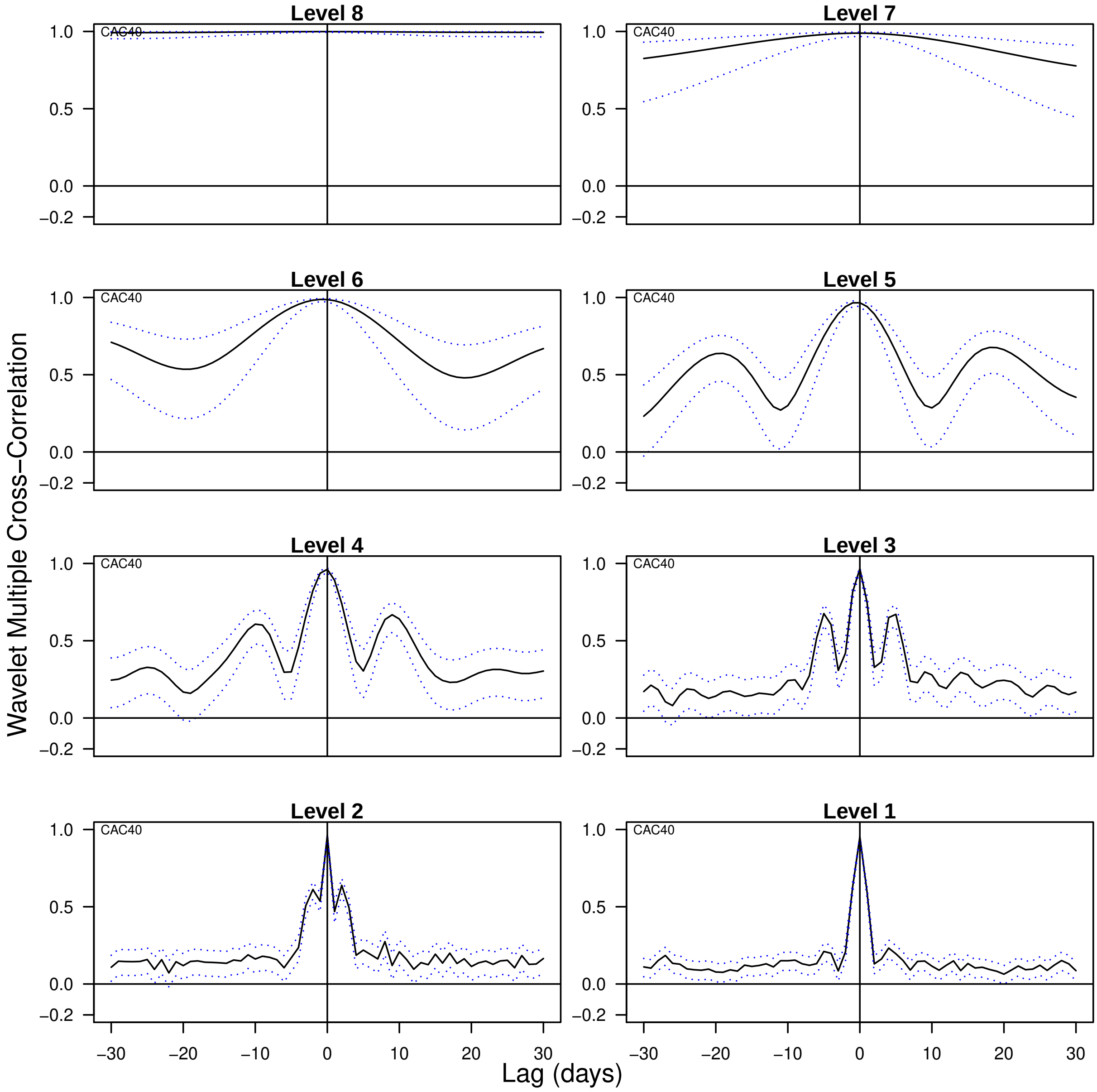

FIGURE 2. Wavelet multiple cross-correlations for the main Eurozone stock markets at different wavelet scales (the upper-right corner signals the market acting as potential leader/follower). The dotted lines correspond to the upper and lower bounds of the corresponding 95% confidence interval.