Physica A 391 (2012) 1097–1104

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets

Javier Fernández-Macho *

Department of Econometrics and Statistics, University of the Basque Country, Lehendakari Agirre 83, E48015 BILBAO, Spain

ARTICLE INFO

Article history:
Received 20 August 2011
Received in revised form 19 October 2011
Available online 11 November 2011

Keywords:
MODWT
Multiscale analysis
Multivariate analysis
Stock markets
Returns
Wavelet transform

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.

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. 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.

More specifically, the present methods are not just a more convenient way to establish the overall multiple relationships but also they have important advantages over the usual wavelet methods that use simple correlations and cross-correlations between all possible pairs of variables. For example, in many wavelet studies where the relationships among several variables are considered the wavelet correlation is used between pairs of variables (see Refs. [1–4] etc.) 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 for the analyst who, at the end, is faced with a very large set of graphs with potentially conflicting information.[1] Conversely, the proposed single

* Tel.: +34 946013729; fax: +34 946013754.
E-mail address: javier.fernandezmacho@ehu.es.

[1] As an illustration of this we may note that a total of 55 wavelet correlations graphs and 440 wavelet cross-correlations graphs would be required for the multivariate dataset analyzed in Section 6.

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.11.002

wavelet multiple correlation and cross-correlation graphs will give the analyst a clearer indication about the type of overall correlation that exists within the multivariate set at different timescales.

Besides this more correct assessment of the overall multivariate relationship, the proposed methods will also provide protection against spurious detection of correlation at some wavelet scales obtained from simple pairwise comparisons due to possible relationships with other variables within the multivariate set. Finally, we can also note that they provide certain protection against the typical inflation of type I errors due to the *experimentwise error rate* [5, p. 617] when dealing with all possible pairwise comparisons in a multiscale context.[2]

As a consequence of these advantages the present methods can provide more accurate results that allow for more predictive interpretations of the data.

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.*, Refs. [6–10] 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 [11–13, p.10]. 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 $X_t = (x_{1t}, x_{2t}, \dots, x_{nt})$ be a multivariate stochastic process and let $W_{jt} = (w_{1jt}, x_{2jt}, \dots, w_{njt})$ be the respective scale λ_j wavelet coefficients obtained by applying the maximal overlap discrete wavelet transform (MODWT) [11,14] 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}, $i = 1, \dots, n$, 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 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 square 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

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

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 [15–17].

[2] For example, doing all possible pairwise wavelet correlation significance tests for a given wavelet scale at the nominal $\alpha = 5\%$ significance level among 11 unrelated series (the same number as in the multivariate dataset analyzed in Section 6, *i.e.* 55 comparisons) would increase the overall chance of a Type I error to $1 - (1 - \alpha)^{55} = .94$. That is, a huge 94% chance of finding a significant correlation at the given wavelet scale somewhere among those 55 tests instead of the nominal 5%.