Study of Relationships Between Time Series by Co-spectral Analysis

Francesco Domenico d'Ovidio[1] and Najada Firza[1,2](✉)

[1] Department of Economics and Finance, University of Bari "Aldo Moro", Largo Abbazia Santa Scolastica, Bari, Italy
najada.firza@uniba.it
[2] Catholic University Our Lady of Good Counsel, Tirana, Albania

Abstract. In the literature, cross correlation is often used to analyze the relationships between time series. When the objective of the study is to determine the antecedents and consequences between two or more data sets, cross-correlation analysis is a useful but not sufficient tool to provide information. This paper presents an exploratory technique for investigating relationships between time series. The chosen technique can provide information about priority or precedence within a series, which may establish dependency relationships that cross-correlation analysis cannot identify. The proposed solution is based on the transformation of the observed data into Fourier series and the joint analysis of two series by co-spectral analysis. An example is presented to demonstrate the application of time series analysis to the study of financial data. The article emphasizes the usefulness of these techniques for subsequent analysis.

Keywords: Time series · Cross correlations · Fourier transform · Spectrogram · Co-spectral analysis · Cross gain

1 Introduction

In time series analysis, it is common to evaluate the relationships between them using cross-correlation, which measures the similarity of two observed data series over time as a function of a time shift or translation applied to one of them.

However, if the goal of analysis is not the generic inter-relationship between two or more sets of data, but rather the determination of which among them might be antecedent and which might be consequent (implying concepts that are often preparatory to causality, though not necessarily implying it), the study of cross-correlations proves to be a necessary tool, but not sufficient to provide the required information directly or as an exploration of the different studied phenomena.

To address the issue of identifying antecedent and consequent time series for targeted analysis, we followed the recommendations of a 50-year-old study (Delvecchio 1974), paying attention for this purpose to co-spectral analysis. This technique is based on transforming observed data into Fourier series and analyzing them jointly with techniques that operate in the frequency domain instead of the temporal domain, where cross-correlation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024, LNCS 14813, pp. 239–254, 2024.
https://doi.org/10.1007/978-3-031-64605-8_17

analysis is typically used. Fourier techniques are well-known in Signal Theory, but their exploratory value in the economic and social sciences may be underestimated.

This study first explains the mathematical and methodological foundations of spectral analysis (Sect. 2). Then, Sect. 3 delves into the specifics of co-spectrum analysis. Section 4 provides a detailed examination of the application of this technique to the study of financial series. Finally, Sect. 5 presents some concluding remarks.

2 Time Series Analysis and Fourier Series Transformation

To determine the statistical significance of cyclic variations in a stationary time series [1], the analysis of the correlogram can be used. The correlogram is the series of autocorrelation coefficients rh = C /Ch 0, where Ch (h = 0, 1, 2,…, m) are the lag autocovariances of the detrended series, which is assumed to be stationary. The eigenvariances for each lag h, in the case of stationary time series with zero mean can be best determined as described in Bendat and Piersol (1966, 2004):

$$C_h = \frac{1}{s-h}\sum_{t=1}^{s-h}(x_t - \overline{x})(x_{t+h} - \overline{x}) = \frac{1}{s-h}\sum_{t=1}^{s-h} x_t x_{t+h}$$

To achieve stationary data series, it is typically necessary to adjust the trend by taking the difference with respect to an analytic function or by using prime differences. The second technique is preferred because it allows for obtaining a self-stationary series in the mean without complex functional analysis and without altering the structure of polynomial-type evolutionary series 2 (see, e.g., Yaglom 1958);

2.1 Verification of Stationarity of Time Series

Various tests can be used to evaluate the stationarity of the observed series, which is a necessary condition for time domain analysis, and its opposite, integration. In this study, two tests were used in addition to graphical observation of the differentiated series. The most well-known test for this topic is the Augmented Dickey-Fuller (ADF) test, which was also the first test proposed on this topic (see, Dickey and Fuller 1979; Said and Dickey 1984). This test compares a stationary AR(p) to a stochastic process (random walk) after excluding deterministic components. To apply the Augmented Dickey-Fuller (ADF) test, an auxiliary regression is conducted using the T-statistic (at a prefixed α level of significance). The number of lags of the differenced variable must be predetermined and set as the limit of the summation of differenced terms in the regression function

$$\Delta y_t = \mu + \beta t + \varphi y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta y_{t-j} + \varepsilon_t$$

where μ is a constant, β is the coefficient of a hypothetical residual trend, and φ is the coefficient of the unit root of the original series while the coefficients δj represent the

[1] A time series is considered "stationary in the strong sense" (or *narrow sense*) when its mean, variance and all moments of third order and higher are independent of time.

autoregression of the differentiated series. The number of lags k can be selected using Akaike's, Schwarz's, or Hannan-Quinn's information criterion: different models can be estimated, and the value of k that results in the minimum value of the chosen statistic should be chosen (Cottrell and Lucchetti 2016, 2017).

The ADF test is unidirectional with the basic hypothesis H0: $\varphi = 0$ and alternative hypothesis H1: $\varphi < 0$. When testing the null hypothesis, yt must be differentiated at least once to be considered stationary. However, under the alternative hypothesis, yt is already stationary and does not require further differentiation[2]: therefore, largely and significantly negative values of the DFτ test statistic (and thus p-values $< \alpha$) imply the rejection of the basic hypothesis. To compare with the critical values tabulated by Dickey and Fuller using simulation techniques, it is necessary to standardize the test statistic by dividing it by its s.e.: $DF_\tau = \hat{\varphi}/s.e.(\hat{\varphi})$.

This test is highly flexible because the shape of its distribution (and therefore the set of critical values for the test) depends on the constraints placed on some of the parameters of the auxiliary regression (Hamilton 1994).

The full model described in the previous equation is applicable when the time series is assumed to be nonstationary in mean and variance, meaning it has a trend with quite different cyclical or conjunctural variations. If nonstationarity in variance is assumed with a nonzero stationary mean, the β parameter should be set to zero. If the mean is instead zero and nonstationarity in variance is assumed, the μ parameter should also be zero. Note that by setting all parameters δj to zero (i.e., by setting $k = 0$, i.e., no lag), transforms the ADF test into the Dickey-Fuller test for 1st-order integration.

The KPSS test (Kwiatkowski et al. 1992), while less powerful than the ADF test, resolves some of its application uncertainties: ADF gives unreliable results if the time series is neither stationary nor integrated, meaning that the differentiated series is also nonstationary. This test, in fact, has as its basic assumption the stationarity of the studied series and not its integration: its rationale is that if the elements of the series are expressed as $y_t = \mu_y + u_t$, then the sample mean of y_t is a consistent estimator of μ_y if (and only if) u_t represents a stationary process with zero mean. Additionally, the long-run variance of u_t is a finite number.

In practice, after estimated the u_t with $e_t = y_t - \bar{y}$, one should calculate their empirical ĉ autocovariances from order -m to order m: the bandwidth m must be large enough to allow short-term persistence of the et but not too large relative to the length T of the series: the authors of the GRETL recommend m values of $k = INT[(4 \times T/100)0{,}25]$ or not much larger (Cottrell and Lucchetti 2017).

In this work, $m = 2 \times k = 16$ was chosen.

Starting from the autocovariances ĉ, we can estimate the long-run variance with

$$\overline{\sigma}^2 = \sum_{i=-m}^{m} \left(1 - \frac{|i|}{m+1}\right) \cdot \hat{c}_i.$$

[2] The *rationale* for the ADF test is that if the series is integrated, then the lagged level of the series $\{y_{t-1}\}$ will not provide relevant information to predict changes in y_t beyond that provided by the lagged differences $\{\Delta y_{t-j}\}$. In this case, therefore, $\varphi = 0$ may be posited, which is precisely the basic assumption H_0.

The KPSS test statistic is

$$\eta = \frac{\sum_{t=1}^{T}\left(\sum_{i=1}^{t} e_i\right)}{T^2\bar{\sigma}^2}$$

and, under the basic hypothesis, its distribution is asymptotic and independent of disturbance parameters, as defined by the authors using simulation methods. The assumption of stationarity should be rejected if the empirical value of the test exceeds the tabulated asymptotic critical value for the predetermined significance level ($\eta_{0,10} = 0.347$; $\eta_{0,05} = 0.463$; $\eta_{0,025} = 0.574$; $\eta_{0,01} = 0.739$).

2.2 Correlograms and Spectrograms

In autocorrelation analysis, it is recommended to limit the maximum shift m to no more than s/3, where s is the number of terms in the time series (Malinvaud 1971). This ensures good accuracy, even though shifts equal to half the length of the series itself can be placed in series of conspicuous numerosity.

Empirical autocorrelation coefficients have a useful property: they are invariant, meaning they do not depend on the units of measurement or point of origin. In the case of incorrelation, they follow an approximately normal distribution law (Hannan 1960), allowing for the drawing of confidence intervals given by $\pm 2\,\widehat{\sigma}$ or $\pm 3\,\sigma$) (see Kendall 1973; Kendall and Stuart 1976), within which these coefficients should be included when the residuals distribute randomly (e.g., a self-stationary series should have coefficients that are all nonsignificant, except possibly for the first one).

Correlogram analysis is helpful in analyzing cycles (Malinvaud 1971), but it may not provide precise indications. In the presence of time cycles of different frequencies, each autocorrelation coefficient is affected by all existing cycles, and partial autocorrelation function (PACF) analysis may not always be sufficient to solve this problem. The purpose of the partial autocorrelation function is to evaluate each autoregressive coefficient net of the effect of the previous ones: this reduces the possibility of considering significant components of 'spurious autoregression', which could be caused by the combination of two or more autoregressive components. For example, consider a series that has significant autoregressive components for lag 2, lag 5, and lag 10. The problem is to identify how much of the component with lag 10 is due to the 'beating' between the component with lag 2 and that with lag 5. Moreover, when dealing with series that have a small or large number of terms, the significance of the coefficients may be underestimated or overestimated, respectively. These flaws can make the results considerably less useful, or otherwise less clear. Therefore, the applicability of the method must be evaluated on a case-by-case basis.

However, there are statistical methodologies that provide more robust information, such as spectrogram analysis also known as spectral analysis (Malinvaud 1971; Hamilton 2020). These techniques are well-known in various fields of scientific research and can be easily applied to the phenomenon discussed in this note, as well as many others.

For instance, periodogram and spectrogram enable the estimation of the variance in series explained by different cycles of varying frequencies, which together generate

the series itself. These are defined in the so-called 'frequency domain', which differs in properties and developments from the 'time domain' where source data and even autocorrelation functions are defined.

There is a clear relationship between these functions and the representation of the various spectral densities, as correlogram and spectrogram can provide similar information. However, spectral analysis is particularly useful when multiple cyclic or erratic components are present in the series and can be advantageously used in conjunction with autocorrelation analysis.

The robustness of the solutions identified by spectral analysis is not dependent on the amount of available data, since this technique is based on purely mathematical transformations and computations, rather than assumptions about the generating processes of the series, making it almost entirely nonparametric (IBM Corp. 2012). Furthermore, according to the assumptions of Fourier analysis, the correlation between the sine functions used is always zero: therefore, all individual terms of a periodogram are mathematically and statistically independent of each other. Additionally, it is always possible to reconstruct the original series from the spectral coefficients, which imples that this analysis is lossless.

Regarding the methodological references for spectral analysis, it is important to note that we define each kth term of the periodogram of a stationary stochastic process {Xt} (such as the residual series considered in this study) with the notation:

$$I_k = a_k^2 + b_k^2$$

The coefficients for Ik, as proposed by Malinvaud (1971), are given by:

$$a_k = \sqrt{\frac{2}{s}} \sum_{t=1}^{s} x_t \cos \frac{2\pi t}{s}, b_k = \sqrt{\frac{2}{s}} \sum_{t=1}^{s} x_t sen \frac{2\pi t}{s}$$

in this context, we will have:

$$a_0 = \frac{1}{\sqrt{s}} \sum_{t=1}^{s} x_t, a_p = 1/\sqrt{s} \cdot \sum_{t=1}^{s} (-1)^t x_t, b0 = bp = 0.$$

Ik also represents the contribution that the sinusoidal component of period s/k makes to the sum of squares of xt. Assuming that the observed series has a zero mean, we can develop this expression to obtain $I_k = 2 \sum_{h=-s}^{s} C_h \cos \frac{2\pi kh}{s}$, where Ch represents the shift autocovariances of the series. This equivalence does not apply to Io (which is null by assumption) nor, if the number of terms in the series is even, to the pth term of the periodogram: $I_p = \sum_{h=-s}^{s} C_h \cos(\pi h)$.

However, it is generally not appropriate to assign significance to every peak in the periodogram. To reduce variance and disturbance components, it is preferred to apply smoothing functions, known as 'windows'. These windows correspond, generally, to moving averages of three or more terms of the periodogram.

The *spectral density function* (Malinvaud 1971; Bendat and Piersol 1966, 2004) is the result of the periodogram transformation at point k, and its Cartesian representation

is precisely the *spectrogram*. For interpretation purposes, it is important to consider only spectra of sufficient magnitude, which represent points of relative maxima in the spectrogram. This is because the spectral density is the result of a smoothing function of the k^{th} term of the periodogram, which can be likened to a centered moving average.

The following text includes graphs that display spectra at the null frequency (although this frequency corresponds, mathematically, to a period of time that is unquantifiable but tends towards infinity). The purpose of displaying this information is to show the overall trend of the spectrogram and, more importantly, the actual contribution made by each individual component to the variability of the residual series. According to Stuart's theorem, the ordinate value of the spectrum measuring the contribution is expressed in terms of the deviance of the series:

$$\frac{f(0)}{2}+\sum_{k=1}^{m-1} f(\omega_k)+\frac{f(\omega_m)}{2} = DEV(X).$$

3 Cross-Correlation and Co-spectral Analysis

In probability and statistics, given two stochastic processes $X = \{X_t\}$ and $Y = \{Y_t\}$, *cross-covariance* is a function that returns the covariance of each process with the other at pairs of time points (t, s). This means that if $E(X_t) = \mu_{x(t)}$ and $E(Y_t) = \mu_{y(t)}$, the cross-covariance can be calculated as follows:

$$Cov(X_t, Y_s) = C_{xy}(t, s) = E[(X_t - \mu_{x(t)})\,(Y_s - \mu_{y(s)})] = E(X_t\, Y_s) - \mu_{x(t)}\,\mu_{y(s)}.$$

If X and Y are weakly stationary stochastic processes, it follows that:

$E(X_t) = \mu_x$, $E(Y_s) = \mu_y$, $Cov(X_t, X_s) = E(X_t\, X_s) - \mu^2_x$, $Cov(Y_t, Y_s) = E(Y_t\, Y_s) - \mu^2_y$;

However, to consider a *stationary* (second-order) *stochastic bivariate process*, it is necessary to confirm that the cross-moments $Cov(X_t, Y_s)$ also depend solely on (t-s). This condition is met when $Cov(X_t, Y_s) = E(X_t\, Y_s) - \mu_x\,\mu_y$, to prevent any distortion of the results (Harris 1995; Johansen 1995, 2000; Cottrell and Lucchetti 2017).

Keep in mind that the cross-covariance function is not an even function: generally, $Cov(X_t, Y_s) \neq Cov(X_s, Y_t)$. It is *not positive semidefinite* and, unlike autocovariance, does not exhibit a maximum for null lag ($h = t–s = 0$), but its maximum value can occur at any lag (Battaglia 2007). If one wishes to compare the relationships between different pairs of stationary stochastic processes, it is necessary to construct a standardized index with respect to the maximum, similar to the correlation coefficient. This index measures *cross-correlation*:

$$r_{xy}(h) = \frac{Cov(X_t, Y_{t+h})}{\sqrt{Var(X_t)\cdot Var(Y_{t+h})}} = \frac{C_{xy}(h)}{\sqrt{\sigma_x^2\cdot\sigma_y^2}}$$

Thus, cross-correlation is a measure of the similarity between two time series, and a function of the time between individual observations. It is important to note that the

amplitude values of cross-correlation are not fully normalized[3], meaning that the maximum correlation (positive or negative) will be less than 1 in absolute value. Additionally, the cross-correlation function results only display half of the correlation coefficients to be studied (times t, t + 1, t + 2,…, t + h). To examine the other half, simply reverse the order of the series (cross-correlate B with A instead of A with B).

When cross-correlation is not applicable, such as when the cointegration test fails or when there are too few series to apply statistical tests with sufficient power, co-spectral analysis (also known as *cross-spectral analy*sis) can still provide reliable additional information. This technique, like spectral analysis, operates in the frequentist domain and is a quasi-parametric method.

In fact, the (univariate) spectrograms of two processes $\{X_t\}$ and $\{Y_t\}$ describe each series as a combination of cyclic components (of frequency ω between 0 and π) that are uncorrelated with each other. It is shown that any linear relationships between $\{X_t\}$ and $\{Y_t\}$ only act between components of equal frequency (Battaglia 2007; Bendat and Piersol 1966); therefore the amplitudes of the harmonics of $\{X_t\}$ and $\{Y_t\}$ are mutually uncorrelated for all components of different frequencies.

The joint spectral density function of $\{X_t\}$ and $\{Y_t\}$ can be obtained by taking the Fourier transform of the cross-covariance. However, it is typically represented by separating the real and imaginary parts due to its complex values:

$$S_{xy}(\omega) = c(\omega) + iq(\omega).$$

The co-spectral density (real part of the previous expression) is the even function $c(\omega)$, meaning that $f(x) = f(-x)$, and measures the correlation of the in-phase frequency components of the two series. The quadrature spectrum[4] is the odd function $iq(\omega)$, meaning that $f(x) = -f(-x)$, and corresponds to the correlation of the out-of-phase components.

Co-spectral analysis provides further information about the relationships between the series in the frequentist domain.

- first, *coherence*, which is a normalized form of the cross-spectrum measuring the overall correspondence between the spectra of the two related series. In practice, this is referred to as the 'correlation between spectra':

$$K_{xy}(\Omega) = \frac{S_{xy}(\Omega)}{\sqrt{S_x(\Omega) \cdot S_y(\Omega)}}.$$

$K_{xy}(\omega) = 1$ indicates a high degree of similarity between the frequency component ω in both signals, with a perfect linear relationship if applicable for each ω. Conversely, $K_{xy}(\omega) = 0$ indicates no similarity at all, and if this is the case for each ω, the series

[3] The cross-correlation coefficients are proportional to the relative covariances. Therefore, their maximum functional is $|r_{xy}(h)|<1$, as they are not even functions. Additionally, at the null lag, the cross-correlation will be exactly equal to the linear correlation coefficient between the two variables: $r_{xy}(0) = r_{xy}$.

[4] If the two series are mutually uncorrelated, i.e., $r_{xy}(h) = 0$ for each h, then $c(\omega) = iq(\omega) = 0$; on the other hand, if Y = Xt, then $S_{xy}(\omega) = S_x(\omega)$, and thus we fall into the univariate case, being $c(\omega) = S_x(\omega)$ and $iq(\omega) = 0$.

will be completely uncorrelated, resulting $r_{xy}(h) = 0$ for each h[5]. Sometimes, the *square coherence* is calculated as the square of the ratio mentioned above. This can be interpreted similarly to the well-known *index of determination*, which is the square of the linear correlation coefficient. Consistency and squared coherence are insensitive to linear transformations of the processes $\{X_t\}$ and $\{Y_t\}$; therefore, the information they provide applies not only to the stationary series describing them but also to the non-stationary series of which the former are a linear transformation.

However, interpreting coherence values independently is inadvisable. For instance, when the spectral density estimates in both series are very small, large coherence values can be generated (due to a very small divisor in the calculation of coherence values), even if there are no strong cyclic components in both series in their respective frequencies.

- In the second instance, the *crossed amplitude* (also known as 'phase amplitude') should be considered:

$$A_{xy}(\omega)=\sqrt{c(\omega)^2 + {}_i q(\omega)^2}$$

This index measures the extent to which each frequency component of a given series is influenced by the components of the other series.

- The *spectral gain* of the series is a noteworthy measure, calculated by the cross-amplitude value related to the estimated spectral density for one of the two series in the analysis. Accordingly, it is necessary to calculate *two* distinct gain values: $A_{xy}(\omega)/S_x(\omega)$ and $A_{xy}(\omega)/S_y(\omega)$. These values can be interpreted as OLS regression coefficients of the respective frequencies of the series. They serve as *confirmation* (in the frequentist domain) of the cross-correlation coefficients in the time domain.
- Finally, the *phase shift* (or *phase spectrum*) estimates are calculated by taking the arctangent of the ratio between the estimated quadrature spectrum and the co-spectral density. These estimates are typically represented by the Greek letter φ: $\varphi(\omega) = \arctan[q(\omega)/c(\omega)]$. The *phase shift* indicates the time difference between each frequency component of one series and those of the other series.

Due to the considerations outlined in this section, it is reasonable to conclude that the results of the co-spectral analysis, specifically the *gain* and *shift* values of the studied series, have intrinsic value in terms of analysis in the frequentist domain. They are also excellent complements to cross-correlation analysis, which provides clear and immediately identifiable results in the temporal domain. Co-spectral analysis does not suffer from some of the methodological constraints of cross-correlation analysis, such as the necessary cointegration conditions and a *substantial number of terms* in the data series.

Additionally, similar to the correlogram, some lagged terms may appear statistically significant due to 'spurious autoregression.' Similarly, certain components among the cross-correlations may only be significant due to the interaction of components with smaller delays (*beat*), which can be difficult to identify without partial cross-correlation information.

[5] This means that the result of co-spectral analysis can be used as confirmatory information for cross-correlation, which may be unreliable due to a lack of cointegration or low series numerosity. Additionally, since $K_{xy}(\omega) = K_{yx}(\omega)$ it is generally preferable to omit the indices and write simply $K(\omega)$.

This analysis procedure can be applied to various types of time series, including economic, social, financial, or physical-technical data. For example, see the following section for its application to financial series.

4 Application of the Proposed Technique to Financial Series Case

To apply the proposed study technique to real data, non-recent financial data was chosen. This was done to comply with the rules of use stipulated by the data providers, which often require long periods of embargo for their free use in research and scientific dissemination operations. The data covers the time span from January 2008 to September 2014 and includes the opening, high, and low prices, as well as the daily closing prices within each working week (5 days) for the following assets:

- *Gold Future*, a popular instrument among traditional investors, which has recently experienced periods of significant upward interest alternating with periods of bearish crisis, even long ones;
- *Copper Spot*, which concerns the prices of copper, a consumer good rather than an investment good like the former, and is therefore closely linked to the performance of the real economy and the industrial sector;
- the *Thomson-Reuters-Jefferies CRB*, which is a commodity index that replaced the former CRBMET. The CRBMET became less and less representative, and therefore was adjusted from 2010 onward;
- the *U.S. 10-yrs Treasure Note*, which is an asset that represents the yield on the main U.S. medium- to long-term government bond with a 10-year maturity. It is a useful indicator of the health of the U.S. economy and is likely to influence the choices of large investors;
- the *Standard & Poor's 500 index*, that is the world's leading stock market index, but it only covers U.S. companies;
- The *S&P 500 volatility index with a 30-day expiration* (representing the average value of the "risk premium" investors are willing to pay for options on the S&P 500, offering a forecast of the variability of the stock market over the next 30 days.)
- The S&P 500 volatility index with 60- and 90-day expiration.

The described series provide comprehensive coverage for the period under consideration, with minimal midweek interruptions, usually only for holidays. In the U.S. market, there are fewer holidays than in Italy and other countries.

Any missing data occurring in sequences of no more than 2–3 days were estimated using rectilinear interpolation of contiguous values (Figs. 1, 2 and 3).

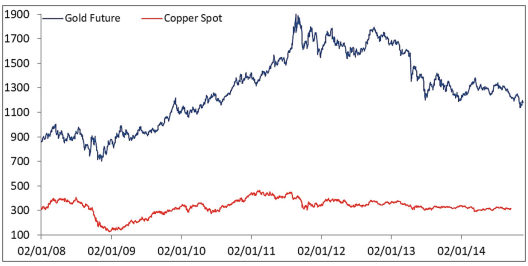


Fig. 1. Daily closures of Gold Future and Spot Copper.

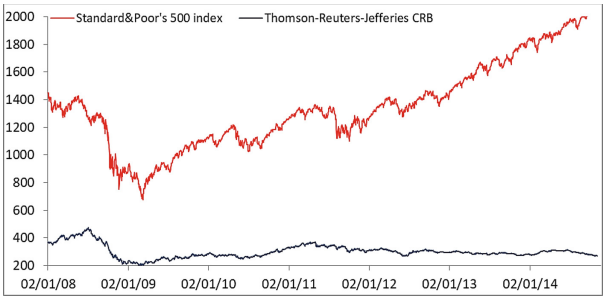


Fig. 2. Daily closures of the Standard&Poor's 500 and Thomson-Reuters-Jefferies CRB indices.

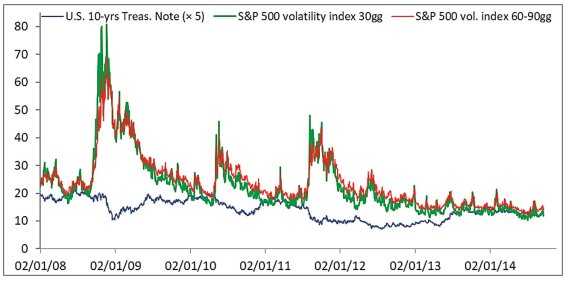


Fig. 3. Daily closures of U.S. 10-Year Treas. Note and S&P500 volatility index at 30 days and 60–90 days expiration.

The financial series exhibits various forms of trend, including both increasing and decreasing trends, as well as internal cycles of different cadence. This indicates non-stationarity on average, with significant differences in variance over time, particularly for Gold Future, Standard&Poor's 500, and the S&P 500 volatility index at both described time frames. It is important to note the non-stationarity in variance. To avoid spurious relationships between time-bound series, it is recommended to obtain stationary data series. This can be achieved by first applying a logarithmic transformation of the data (Box-Cox 1964) and then calculating their prime differences. The transformation used

is as follows:

$$y_t = ln(xt) - ln(xt-1) = ln[xt/(xt-1)].$$

As a result of this operation, the transformed series is assimilated into stationary processes of the second order (see Fig. 4).

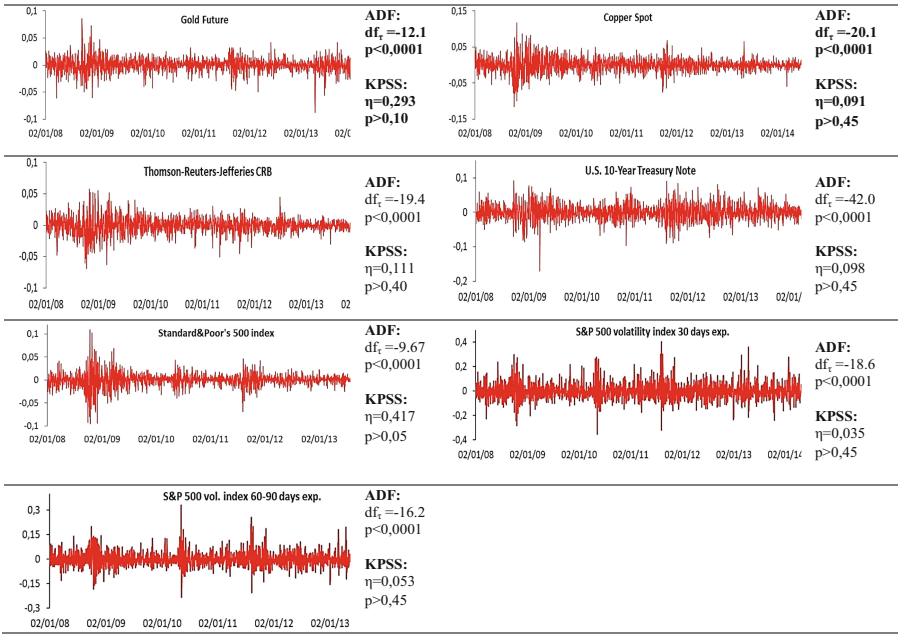


Fig. 4. Differential logarithmic series of daily asset closures. Statistical significance of ADF and KPSS tests for each asset.

This assumption was tested for each series by both ADF unit root test (with a significance level of $\alpha = 0.01$) and KPSS test with 16 lag periods (with a critical value $\eta_{0,025} =$ 0,574). The significance level $\alpha = .025$ was chosen instead of the more critical $\alpha = .01$ due to the lower power of the test, which increases the risk of accepting a false baseline hypothesis if the significance level is too small. However, the redundancy of statistics is justified by the constraint that only weak stationarity of the series is necessary for proper use of the analysis in the frequentist domain, as discussed in Sect. 2, while the ADF test is also limited by series integration.

Pairwise cointegration analysis was conducted using both techniques available in GRETL, after verifying the stationarity of each individual series with comforting findings: just the empirical KPSS test value for the Standard&Poor's 500 index series approached $\alpha = 0.05$, but it was not significant enough to raise doubts, considering its lower power. For the combinations obtained by setting the Gold Future series as a fixed term, the results of the Engle-Granger test and those of the Johansen test generally coincide perfectly. The basic hypothesis (integration of the residuals of the linear combination the first, presence of nonzero eigenvalues the other) is rejected with p-values < 0.001. An exception to the rule is the relationship between Gold Future indices and Standard & Poor's indices. The Johansen trace test shows that $.010 < p < .015$, which means that, for a significance level $\alpha = .01$, we can accept the basic hypothesis that at least one linear combination of the two series is nonstationary.

The analysis is limited to investigating the potential effects of the various indices on the Gold Future related index (and viceversa). Since causal relationships cannot be hypothesized, only antecedence/consequence of the data is considered. To achieve this, co-spectral analysis is used, specifically examining the gain function of Gold Future index with respect to each of the other financial indicators, using a 5-term Tukey-Hamming window. When the gain function indicates an antecedent relationship, it is accompanied by an analysis of the corresponding cross-correlation function with a lag of up to 20 days.

Figure 5 displays the gain functions taken from the co-spectral (bivariate) analysis between the Gold Future index and each of the other financial indices described above. This analysis may indicate the antecedent and consequent relationships between the compared indices.

The co-spectral gain function in the first image shows the degree of dependence between Copper spot (trend marked by the green line) and Gold Future (trend marked by the blue line). The largest gain is evident when considering Copper spot prior to Gold Future. Almost all the phases observed in the graph are lower if Copper spot is assumed as consequent. Therefore, it is recommended to assume Gold future as consequent to Copper spot.

In this brief exploratory-methodological analysis, we can observe significant relationships between Gold Future and U.S. 10-year Treasury Note through the co-spectral gain function. Additionally, Gold Future has significant relationships with the S&P 500 volatility at 30-day and S&P 500 volatility T 60–90-day indices.

Instead, the gain functions produced by the correlation between Gold Future and S&P500, as well as between Gold Future and Thomson-Reuters-Jefferies CRB, are not relevant: therefore, no antecedence/consequence can be assumed. This corresponds to the statistical independence of the two financial series when transitioning from the frequency domain to the time domain.

At this stage, it is appropriate to consider cross-correlation functions that relate only to the relationships clearly identified through the gain function. This will help to define relational lags. In cases where the gain function is not relevant, the cross-correlation function cannot adequately identify the relationships between the variables under study, and therefore it is not useful to study it.

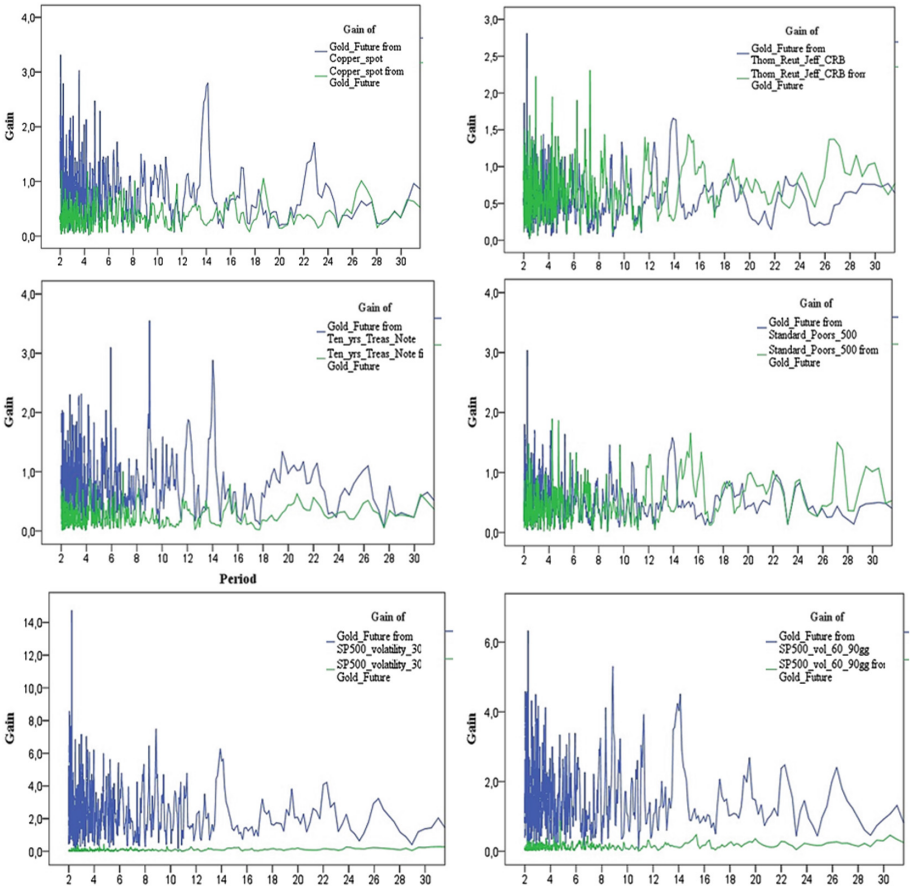


Fig. 5. Co-spectral gain functions between Gold Future and other financial indexes (linear scale period 1–30).

The graphs in Fig. 6 show the significance of a relationship with a kth lag. If the relationship is statistically significant, it indicates the influence of the identified series on the values of the Gold future k days later. This influence may be direct or inverse, depending on whether the lag is of positive or negative sign.

For instance, the cross-correlation between Copper spot and Gold Future indicates that the closing prices of the former asset have a direct influence on the day itself (at lag 0, a relationship that could still be spurious, i.e., both assets can be influenced by others), but also a slight but significant negative influence with lags 8 and 17. Borderline influences (positive, but almost insignificant) are also found at lags 6, 9, and 15.

The impact of the 10-year U.S. Treasury note asset on Gold Future is predominantly negative, albeit not very significant. This holds true for lag 0 (the same day) and even less so for lags 5, 8 and 19. However, there is a minor positive impact at lag 2 (i.e. 2 days later).

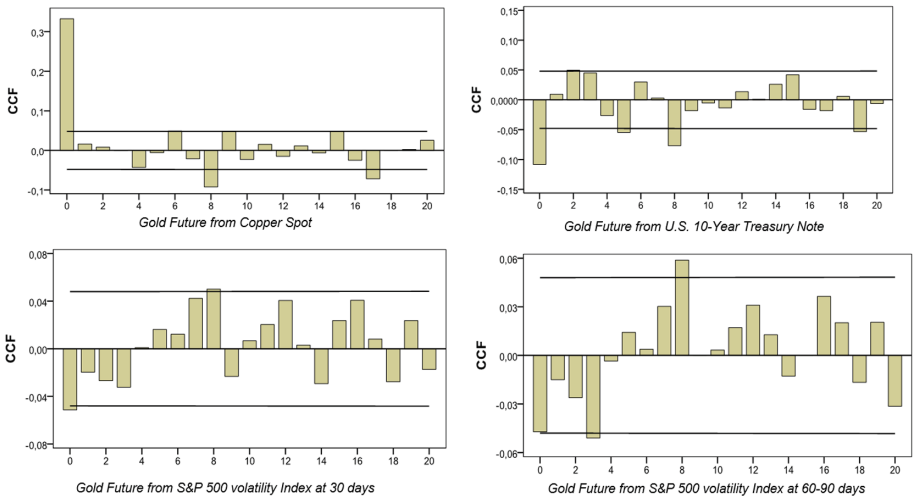


Fig. 6. Cross-correlation functions between Gold Future and its financial antecedent indices.

The influence of the 30-day S&P500 and 60–90-day S&P500 volatility indices on the Gold future is minimal and almost insensitive. However, there is a borderline negative influence with lag 0 (and in the second case, also with lag 2) and a positive influence with lag 8.

As argued and demonstrated here, it is difficult to interpret significant relationships when the gain function does not allow for the determination of antecedent and consequent series, as is the case with the Thomson-Reuters-Jefferies CRB and S&P500 indices. For instance, in the first graph of Fig. 7, we observe (without considering the correlation at lag 0) that, in addition to some relationships on the right side of the cross correlograms, there are significant relationships between the TRJ-CRB and Gold future with lags of 8 and 12, indicating a negative influence, and with lags of 1 and 15, indicating a slight positive influence.

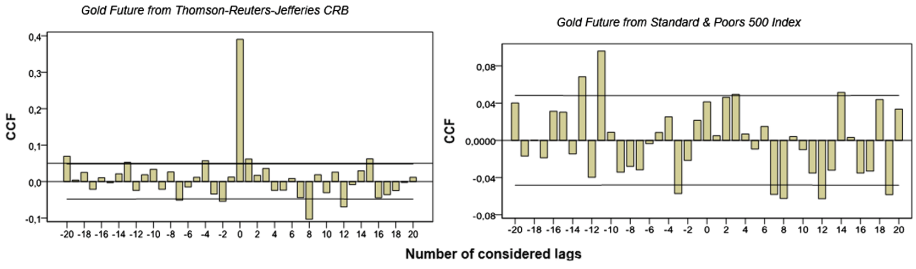


Fig. 7. Cross-correlation functions between Gold Future and financial indexes from which it has undetermined relationships.

The TRJ-CRB was also found to be minimally influenced by the Gold future, with negative influences at lags of –2 and –7, and a positive influence at lags of –4, –13, and –20.

Even more confusing turns out to be the relationship between the Gold Future and the S&P500, which not surprisingly is the most problematic pair of assets in the Johansen test.

5 Concluding Remarks

The work carried out has demonstrated the validity of the frequency approach in time series analysis. This has allowed for exploration of the different relationships that exist among highly relevant financial variables in the U.S. market. The empirical analysis has identified some interesting elements (years 2008–2014). The Gold Future is typically an affected asset rather than an influencer of changes in other financial market indices, even after making the series stationary. It is important to maintain objectivity and avoid biased language when discussing financial market indices.

A statistically significant relationship, albeit of slight magnitude, exists between the closing values of various indices and the Gold Future closing price 8 (working) days later.

In the next work we will focus on the frequency domain analysis which lacks inference characters and thus well-defined probabilities of replication of results with different time data sets.

Therefore, observation of more recent financial series could lead to different conclusions. However, early explorations of the 2015–2022 data confirm the observations treated by earlier data. These explorations are not still publishable and therefore cannot be used in this study.

References

1. Battaglia, F.: Statistical Forecasting Methods. Springer, Milan (2007)
2. Bendat, J.S., Piersol, A.G.: Measurement and Analysis of Random Data. J. Wiley, New York (1966)
3. Bendat, J.S., Piersol, A.G.: Random Data Analysis and Measurement Procedures, Wiley Series in Probability and Statistics (3rd Edition) (2004)
4. Box, G.E.P., Cox, C.: An Analysis of transformations. J. Am. Stat. Assoc. **65** (1964)
5. Cottrell, A., Lucchetti, R.: Gretl - Gnu Regression, Econometrics and Time-series Library, Samurai Media Limited (2016). ISBN: 978-9888406272
6. Cottrell, A., Lucchetti, R.: Gretl User's Guide, gretl documentation (2017). http://sourceforge.net/projects/gretl/files/manual/gretl-guide-a4.pdf/download
7. Delvecchio, F.: Structural changes in the marriage curve in Italy. J. Econ. Annals Econ. Year XXXVI (new series), no. 3–4, March–April 1974
8. Dickey, D.A., Fuller, W.A.: Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. **74**(366), 427–431 (1979). https://doi.org/10.2307/2286348
9. Engle, R.F., Granger, C.W.J.: Co-integration and error correction: representation, estimation, and testing. Econometrica **55**, 251–276 (1987)

10. Granger, C.W.J., Newbold, P.: Spurious regressions in econometrics. J. Econ. **2**(2), 111–120 (1974). https://doi.org/10.1016/0304-4076(74)90034-7
11. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
12. Hannan, E.J.: Time Series Analysis. Methuen, London (1960)
13. Harris, R.: Using Cointegration Analysis in Econometric Modelling. Prentice-Hall (1995)
14. IBM Corporation: IBM SPSS Forecasting 21 (2012). ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/en/client/Manuals/IBM_SPSS_Forecasting.pdf
15. Kendall, M.G.: Time-Series, Charles Griffin & Co. Ltd., London & High Wycombe (1973)
16. Kendall, M.G., Stuart, A.: The Advanced Theory of Statistics - Volume 3 - Design and Analysis, and Time-Series, Charles Griffin & Co. Ltd., London & High Wycombe (1976)
17. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.: Testing the null of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? J. Econ. **54**, 159–178 (1992)
18. Johansen, S.: Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford (1995)
19. Johansen, S.: Modelling of cointegration in the vector autoregressive model. Econ. Model. **17**, 359–373 (2000)
20. Malinvaud, E.: Statistical methods of econometrics. UTET, Turin (1971)
21. Oppenheim, A., Verghese, G.C.: Signals, Systems, and Inference, MIT OpenCourseWare (2010). https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-011-introduction-to-communication-control-and-signal-processing-spring-2010/readings/MIT6_011S10_notes.pdf
22. Said, S.E., Dickey, D.A.: Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika **71**(3), 599–607 (1984). https://doi.org/10.1093/biomet/71.3.599
23. Stuart, J.S.: Fourier Analysis. Methuen & Co, London (1961)
24. Trantner, C.T.: Integral Transforms in Mathematical Physics. Methuen London (1951)
25. Hamilton, J.D.: Time Series Analysis. Princeton University Press (2020)
26. Yaglom, A.M.: Correlation Theory of processes with random stationary n^{th} increments, Trans. Am. Math. Soc. Series 2, **8**, 87–141, Original (1955) (1958). Корреляционная теория процессов со случайными стационарными *n-ми* приращениями, Математический сборник, 37(79), 1: 141–196. https://gretl.sourceforge.net/index.html