"Wavelet-based" variable selection methods and composite indicators: an application to the TCB-LEI

Marco Gallegati*

November 11, 2015

Abstract

Relationships between variables change over time, but also differ across scales. Wavelet analysis has the ability to provide both types of information in a unifying framework through time-frequency decomposition analysis. By exploiting wavelets' ability to examine the relationships between variables on a scale-by-scale, rather than aggregate, basis a wavelet-based methodology for the construction of composite indicators is applied to the Conference Board's leading index (TCB-LEI). with the aim of detecting early warning signals of business cycles turning points. The proposed methodology consists of constructing an overall composite index obtained by aggregating several scale-based sub-indexes, each corresponding to a different frequency band, where the individual component are selected on the basis of their statistical performance at each time scale. The application to the TCB-LEI using historical data shows that using the wavelet-based composite index it is possible to improve the performance of composite leading indicators by also retaining the reliability of the early warning signals provided.

Keywords: Wavelets; Composite Leading Indicators; TCB-LEI.
JEL: C1, C3, C5, E3

*Department of Economics and Social Sciences, Polytechnic University of Marche, Piazzale Martelli, 8, 60121 Ancona, Italy; e-mail: marco.gallegati@univpm.it

1 Introduction

Dealing with large datasets is not new in macroeconomics. Following Burns and Mitchell's (1946) classical definition of business cycles NBER business cycle analysts developed composite indexes to measure common movements in a set of macroeconomic variables. The linear tradition in macroeconomics and business cycle analysis, represented by the so-called "indicators approach" to business cycle analysis, involves the creation of composite leading, coincident, and lagging indexes by considering the pattern of comovements among hundreds of historical time series. The construction of composite indexes reflects the awareness that business cycles are complex phenomena with unique features, not only in terms of duration, amplitude, steepness, etc., but also with respect to their causes and consequences. Therefore, a wide coverage of the individual components in terms of different areas is considered essential in order to obtain a good representation of overall economic activity.[1].

In the last two decades the increasing availability of huge and detailed datasets has favored the development of researchers' interests in variable selection methods. A well known example is provided by the new strand of literature based on factor models with applications to traditional small-scale macroeconometric models. The idea of dynamic factor models is to provide an exhaustive summary of the large amount of available information by replacing large sets of variables with a small subset of them. In particular, factor models provide an alternative parsimonious modern statistical framework to extract information from a large number of potentially useful variables by capturing comovements among them through the estimation of a single common factor whose behavior is qualitatively similar to that of the set of contemporaneous economic variables.[2] For example, Stock and Watson's (2002) diffusion indexes extracted from dynamic factor models allow a parsimonious description of the dynamics common to the observed variables by taking a weighted average of these variables, with weights chosen so as to capture the largest possible amount of variation in the variables.

However, since the information content within such large aggregate datasets is mostly overlapping methods for extracting the information content in these data-rich environment are called for. Although in principle there are two ways to exploit the information content within this "soup" of largely overlapping macroeconomic variables, that is time and frequency domain analysis, data processing is generally performed using time domain methods. But the analysis in the time domain yields only part of the information

[1]Subject areas of selected individual indicators include production, construction, labour force, prices, monetary and financial, foreign trade, business and consumer surveys.

[2]Both time domain and frequency domain principal component analysis (of the dynamic factor model) have been developed by Stock and Watson (1989, 1991, 1993) and Forni and Rechlin (1996, 1997, 1998), Forni and Lippi (1997, 1998), Forni et al. (2000), respectively.

embedded in the data. In order to obtain the remaining part of the information the data need to be analyzed in the frequency domain.

Wavelet analysis has the ability to simultaneously provide both types of information in a unifying framework through time-frequency decomposition analysis, that is decomposing a signal into a set of time scale components, each associated to a specific frequency band and with a resolution matched to its scale. The time-frequency representation of the signal allowed by the wavelet transform can reveal patterns and features in the data beyond those evidenced from the usual aggregate perspective and may be useful in analyzing situations in which the degree of association between two time series is likely to change both over time and across frequencies.[3] Thus, wavelet multi resolution decomposition analysis allows to identify different relationships between variables at different time scales so that it is possible to distinguish those time scales at which there is a statistically significant relationship from those scales at which there is not.

These properties can be usefully employed to construct a "wavelet-based" composite indicator that, by efficiently using all available information, can potentially reduce the number of false signals, while retaining parsimony and reliability of the information provided (see Gallegati, 2014). Such a "wavelet-based" overall composite index is created by gathering the information obtained on a scale-by-scale basis[4] through the aggregation of several sub-indexes each corresponding to a different frequency band.

In this paper we examine to what extent the time-frequency decomposition properties of the wavelet transform can be usefully exploited for the construction of composite indicators with an application to the Conference Board's composite leading indicator for the United States (TCB-LEI). First, we perform a preliminary exploratory wavelet-based analysis of the data using some tools associated with the continuous wavelet transform (CWT), i.e. wavelet coherence and wavelet phase difference, with the aim to measure , respectively, the local correlation in the time-frequency space and the phase relationship between each individual TCB-LEI indicator and the reference series (i.e. the coincident economic indicator, TCB-CEI). Then, after decomposing each individual component series of the TCB-LEI into its time scale components, we select at every single time scale those individual indicators that display the best overall statistical performance to construct the corresponding scale-based sub-index. Finally, the scale-based

[3]That comovements among macroeconomic time series differ across frequencies has been firstly documented by Hannan (1963) and Engle (1974, 1978) using band spectrum regressions, then by Ramesy and Lampart (1988a, b) using regression over timescale decompositions. Recently, these findings have been confirmed with respect to the wage Phillips curve (Gallegati *et al.*, 2011), and the different informative content of interest rate spreads for future output and of stock and bond market prices for investments (see Gallegati *et al.*, 2013, and Gallegati and Ramsey, 2013, 2014, respectively).

[4]The statistical performance of each individual component at different time scales is used as a variable selection criterion of the different scale-based sub-indexes.

sub-indexes are aggregated to obtain the overall "wavelet-based" composite leading indicator, TCB-LEI^W.

We evaluate the performance of the "wavelet-based" composite leading indicator as to the TCB-LEI by investigating their length and consistency at NBER peaks and troughs. The results indicate that the wavelet-based composite leading indicator gives reliable signals of approaching turning points slightly in advance relative to the TCB-LEI with respect to NBER reference chronology. Moreover, the comparison between the components of the two indexes corresponding to the "major cycles", i.e. 4 to 8 years periodicities, indicates an average leading time of several months for the wavelet-based relative to the TCB-LEI derived measure.

The paper is organized as follows. Section 2 examines the motivation for using wavelets in the indicators approach by performing the exploratory research analysis of the data using continuous wavelet transform (CWT). Section 3 describes the methodology to obtain a wavelet-based composite index and show the construction of the wavelet-based composite leading indicator for the US. Section 4 evaluates the performance of the wavelet-based composite leading indicator TCB-LEI^W relative to the TCB-LEI and its derived measures and section 5 concludes.

2 The continuous wavelet transform

We begin this section by recalling the basic structure of the wavelet approach to modeling time series data. Wavelets are particular types of function $\psi(.)$ that are localized both in time and frequency domain and used to decompose a function $f(x)$, *i.e.* a surface, a series, etc.., in more elementary functions which include informations about the same $f(x)$. The essential characteristics of wavelets are best illustrated through the development of the continuous wavelet transform (CWT). At the simplest level wavelet analysis consists of a transformation of a signal obtained by projecting the signal, $x(t)$, onto a wavelet basis function ψ via

$$\begin{aligned} W_x(u,s) &= \int_{-\infty}^{\infty} x(t)\psi_{(u,s)}(t)dt \\ \psi_{(u,s)}(t) &= \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}) \end{aligned}$$

The CWT converts the series $x(t)$ into a set of wavelet coefficients, $W_x(u,s)$, each representing the amplitude of the wavelet function at a particular position and for a particular wavelet scale. More specifically, $W_x(u,s)$ is obtained by scaling and translating the the wavelet basis ψ, called "mother wavelet", which is a function of two parameters s and u, denoting the dilation (scale factor) and translation (time shift), respectively where s is a

scaling or dilation factor that controls the length of the wavelet and u a location parameter that indicates where the wavelet is centered (see Percival and Walden, 2000).

Let W_x and W_y be the continuous wavelet transform of two time series x and y with respect to the wavelet ψ:

- $|W_x|^2$ and $|W_x|^2$ represent the wavelet power spectra of x and y, respectively, which depict the local variance of x and y and measure the contribution to the total energy since the power is the absolute-value-squared of the wavelet coefficients;

- $|W_{xy}|$=$|W_x W_y^*|$ (where (*) indicates complex conjugate) the cross-wavelet spectrum of the two series and depicts the local covariance of the two time series at each scale and frequency (see Hudgins et al.,1993, Torrence and Compo, 1998, Grinsted et al., 2004).

Although useful for revealing potentially interesting features in the data, like characteristic scales, the wavelet power spectrum is not necessarily the best tool to deal with the time-frequency dependencies between two time-series. Indeed, even if two countries share a similar high power region, one cannot infer that their business cycles look alike. To detect and quantify relationships between variables, cross-wavelet tools like wavelet coherency and wavelet phase-difference have to be used. The absolute value squared of the smoothed cross-wavelet spectrum $|W_{xy}|$, normalized by the smoothed wavelet power spectra $|W_x|^2$ and $|W_x|^2$, gives the wavelet squared coherency,

$$R_{xy}^2 = \frac{|S(\lambda^{-1}W_{xy})|^2}{S(\lambda^{-1}|W_x)|^2)S(\lambda^{-1}|W_y)|^2)},$$

where S is a smoothing operator in both time and scale (see Torrence and Webster, 1999). The squared wavelet coherency coefficient R_{xy}^2, ranging between 0 and 1, provides an accurate measure of the local correlation between two time series in time-frequency domain (Chatfield, 1989) and is especially useful in highlighting the time and frequency intervals where two phenomena have strong interactions.

Finally, the wavelet coherency phase difference θ is the ratio between the imaginary and the real part of the wavelet coherency:

$$\theta_{xy} = arctan(\frac{\Im[S(\lambda^{-1}W_{xy})]}{\Re[S(\lambda^{-1}W_{xy})]}).$$

The phase of a given time-series x can be viewed as the position in the pseudo-cycle of the series and it is parameterized in radian ranging from $-\pi$ to π and can be useful to characterize the phase relationships between two

time series as a function of frequency, *i.e.* phase synchronization of two time series.

By showing how the power of the projection of the signal varies with the scale of observation the tools associated with the CWT can get indication about the underlying structure in the data, like characteristics scales, i.e. dominant scales of variation in the data, and (highly localized) patterns, beyond those usually evidenced by standard tools focussing on a unique time scale. As such, the CWT provides a powerful tool for performing exploratory data analysis (Tukey, 1977). These descriptive techniques, mostly graphical, find limited application in empirical works because of their incapacity to provide clear answers and the requirement of subjective judgements in interpreting results. CWT tools can avoid the main disadvantages of standard exploratory techniques by allowing the researcher interpretation and decisions processes not to rely anymore on a subjective data visualization process, by retaining at the same time their main advantages, that is to look for flexible ways to examine data without preconceived assumptions.

3 Exploratory data analysis of TCB-LEI components using CWT

In this section we perform a preliminary exploratory wavelet-based analysis of the component series of the TCB-LEI using some tools associated with the continuous wavelet transform (CWT). The most important gain from using the multi-resolution decomposition properties of wavelet transform for the construction of composite indicators is that it allows to efficiently process all available information by exploiting the different informative content of individual indicators at different time scales. More specifically, the time scale decomposition property of wavelets allows identification of different relationships between variables on a scale-by-scale basis so that it is possible to separate those time scales at which the relationship is statistically significant from those scales at which it is not.

We perform the exploratory research analysis of the data by using the wavelet coherency and its phase difference: the first measures the local correlation of two series, the latter the lead/lag relationship between two variables in time-frequency space. Among the several types of wavelet families available we employ the widely used wavelet, that is the Morlet wavelet. Being a complex wavelet, the Morlet wavelet produces complex transforms, and thus can provide us with information on both amplitude and phase,[5] and has optimal joint time frequency concentration, as it attains the minimum possible uncertainty of the corresponding Heisenberg box. The Morlet

[5] The main usefulness from using complex-valued wavelets like Morlet wavelet lies in its ability to also provide the phase information, that is a local measure of the phase delay between two time series as a function of both time and frequency.

wavelet is defined as

$$\psi_\eta(t) = \pi^{-\frac{1}{4}} exp(i\omega_0\eta)exp(-\frac{\eta^2}{2}).$$

where $-1/4$ is a normalization term, $\eta = t/\lambda$ is the dimensionless time parameter, t is the time parameter, λ is the scale of the wavelet and ω is the frequency parameter. We use the value $\omega_0 = 6$ since this particular choice provides a good balance between time and frequency localization (see Grinsted *et al.* 2004) and also simplifies the interpretation of the wavelet analysis because the wavelet scale, λ, is inversely related to the frequency, $f \approx 1/\lambda$.

Our dataset includes the individual components of the demi-last recent versions of the TCB-LEI: Average weekly hours, manufacturing (AWH); Average weekly initial claims for unemployment insurance (ICUI); Manufacturers new orders, consumer goods and materials (NOCG); Manufacturers new orders, nondefense capital goods excluding aircrafts (NOND); ISM Manufacturing, New Orders Index (ISM); Building permits, new private housing units (BP); Stock prices, 500 common stocks (SP500); Money supply (M2); Interest rate spread, 10-year Treasury bonds less Federal funds (YS); University of Michigan Index of consumer expectations (UMCE).

The plots of the wavelet squared coherency for each component series of the TCB-LEi vs the coincident economic index (TCB-CEI), along with the wavelet phase are shown in Figures 1 to 5.[6] The wavelet coherence between any two economic time series effectively charts the strength of their correlation by frequency and over time. The wavelet coherency power is indicated by color coding: it ranges from blue (low coherency) to red (high coherency), with significant regions associated with warmer colors (red, orange and bright green). Regions of high coherency between two time series detect areas in the time-frequency space where two phenomena have a significant interaction and the strength of the relationship is strong. The thick black contour lines denote regions of statistically significant correlation whereas the cone of influence, represented by a shaded area, corresponds to the region affected by edge effects at the beginning and the end of the time series. Following Grinsted et al. (2004) and Maraun and Kurths (2004) critical values were calculated as the 95^{th} percentile of the empirical distribution of the simulated wavelet coherencies.

The wavelet phase difference between the two series charts lead/lag relationships by frequency. The wavelet phase is superimposed on wavelet coherency plots and indicated by arrows in regions characterized by high coherency. The direction of arrows shows the relative phasing of the two time

[6]Figures 1 to 5 have been obtained using the MatLab package developed by Grinsted *et al.* (1994). For a comprehensive list of wavelet-related software see http://www.amara.com/current/wavesoft.html.

series and can be interpreted as indicating a lead/lag relationship: right arrow means that the two variables are in-phase, while left arrow means that the two variables are in-antiphase. In addition, if the right arrow points up (down) it means that the TCB-CEI is lagging (leading) the TCB-LEI component series, while if the left arrow points down (up) means that the TCB-CEI is lagging (leading) the TCB-LEI component series.

Figures 1 to 5 about here

The wavelet coherence between the TCB-CEI and several TCB-LEI component are presented in Figures 1 to 5.The plots show that the correlations structures are quite different, with regions of statistically significant correlation unevenly distributed both across scales and over time periods. In particular:

- most of the high coherency regions are concentrated at scales corresponding to business cycle frequencies and in the middle of the sample (early 1970s-early 1990s);
- only a few component series display a statistically significant relationship with the reference series at the longest scales;
- at lower scales there is evidence of intermittent periods of high correlation throughout the sample;
- the phase relationship is not constant over time, nor across scales.

In spite of the noticeable differences striking similarities of the correlation structure emerge for AWH, ICUI, NOCG, ISM and NOND, where regions of high correlation are uniformly evident throughout the sample at higher scales, that is at scales greater than 16 months. Similar evidence is displayed by building permits and interest rate spread at scales between 32 (64) and 128 months over the whole sample. By contrast, the remaining variables show intermittent periods of high correlation. Regions of high, significant correlation for UMCE and M2 are evident at scales from 32 to 128 months until mid 1980s and early 1980s, respectively, and at scales greater than 128 months from mid 1980s for UMCE. The correlation structure of $M2$ is consistent with the recent revision of the TCB-LEI which replaces $M2$ with an index of financial conditions because starting from the early 1990s then real money supply $M2$ has ceased to be a useful leading indicator (see Levanon et al. 2010). In fact, between mid1980s and mid1990s-early2000s the statistically significant relationship is not present anymore. In addition, when M2 is again statistically significant from mid1990s and mid-2000s, on the 128 and 64 months band respectively, the phase difference is not constant for M2, changing from leading to slightly lagging. Finally, there is sparse evidence of periods of significant correlation for the SP500 at scales around 32 months until early 1980s, from 16 to 128 months after 2000s and at the highest scales throughout the sample.

All in all, the main findings emerging from the visual inspection of the coherency plots are twofold: the absence of regions with significant correlations at lower scales for any individual leading indicator and their quite different correlation structures. Such a different concentration of the statistical significant regions in the time-scale plane provides the main motivation for the construction of a wavelet-based composite leading indicator.

4 Constructing a wavelet-based composite index: an application to Conference Board's leading economic indicator (TCB-LEI)

The results from the CWT-based exploratory data analysis suggest that the components series of TCB-LEI display different informative content at different time scales and that such informative content may change over time. As shown in Gallegati (2014) these findings can be used to develop a wavelet-based approach to composite indicators which differs from the traditional methodology as to the construction and data selection procedures. The result is a composite indicator obtained by combining those components that are likely to contain the most useful information. In particular, the overall composite index is constructed by combining several sub-indexes obtained on a scale-by-scale basis, where each sub-index is a weighted combination of different variables selected on the basis of their statistical performance at a specific time scale.

Let I_1, I_2,...., I_N be the N "reliable" indicators available for the construction of the composite index. By applying a J-level multi resolution decomposition analysis we can provide a complete decomposition of each individual indicator I_i into a smoothed version of the original signal and a set of detail information at different scales:

$$I_i \approx S_J[I_i] + D_J[I_i] + ... + D_j[I_i] + ... + D_2[I_i] + D_1[I_i] \tag{1}$$

where $S_J[I_i]$ contains the "smooth component" of the signal, and $D_j[I_i]$, with $j = 1, 2, ..J$, the detail signal components at ever-increasing levels of detail. At each level of decomposition $j = 1, 2, ..., J$ we can construct a "scale-based" composite sub-index CI_{D_j} by a weighted aggregation of the $k \leq$n statistically significant "reliable" indicators at that scale

$$CI^{D_j} = \omega_{1,j} D_j[I_1] + \omega_{2,j} D_j[I_2] + \omega_{3,j} D_j[I_3] + ... + \omega_{k,j} D_j[I_k] \tag{2}$$

where ω_{ij} is the weight of each indicator i at scale j. Finally, by aggregating the j "scale-based" composite sub-indexes CI_{D_j} we can obtain the wavelet-based composite index CI^W, that is

$$CI^W = CI^{S_J} + CI^{D_J} + ... + CI^{D_j} + ... + CI^{D_1} \quad (3)$$

4.1 Scale-by-scale selection of leading index components

So far we have considered in wavelet methods only continuously labeled decompositions. In practice, in order to construct a wavelet-based composite indicator we needs to develop discrete analogs of these techniques. Therefore we move to the discussion of the discrete wavelet transform (DWT). The DWT is based on similar concepts as the CWT, but is more parsimonious in its use of data (Gencay *et al.*, 2003).[7] The key difference between the CWT and the DWT lies in the fact that the DWT uses only a limited number of translated and dilated versions of the mother wavelet to decompose the original signal so that the information contained in the signal can be summarized in a minimum number of wavelet coefficients.

Before performing wavelet decomposition analysis a number of decisions must be made: what type of wavelet transform to apply, which family of wavelet filters to use, and how boundary conditions at the end of the series are to be handled. We perform the time scale decomposition analysis by using the maximal overlap discrete wavelet transform, MODWT, because of the practical limitations of DWT. Indeed, the MODWT is a non-orthogonal variant of the classical DWT that, unlike the orthogonal DWT, is translation invariant, as shifts in the signal do not change the pattern of coefficients and is not restricted to a dyadic sample length. The wavelet filter used in the decomposition is the Daubechies least asymmetric (LA) wavelet filter of length L = 8, or LA(8) wavelet filter, based on eight non-zero coefficients (Daubechies, 1992), the most widely used filter in economic applications. Finally, in order to calculate wavelet coefficient values near the end of the series boundary conditions are to be assumed. The series may be extended in a periodic fashion (periodic boundary condition) or in a symmetric fashion (reflecting boundary condition). We apply the reflecting boundary condition, where the original signal is reflected at its end point to produce a series of length $2N$ which has the same mean and variance as the original signal.

After application of the maximal overlap discrete wavelet transform to each transformed individual indicator.[8] we get, with a 6-level decomposition, six sets of N wavelet coefficients d_6, d_5, d_4, d_3, d_2, d_1 and a set of N scaling coefficients s_6 that provide information on the short- and long term features of the signal, respectively. Then by performing the synthesis operation,

[7] Since $W(u, s)$ is a function of two parameters, as such it contains a high amount of redundant information.

[8] Following the methodology applied by the Conference Board we calculate for each component of the leading index the month-to-month changes using the symmetric percent change formula, except for interest rate spread where simple arithmetic differences are calculated.

that consists in reassembling the original signal from the wavelet coefficients by using the inverse stationary wavelet transform, we yield reconstructed detail and smooth components as true constituents of the original signals. Specifically, with $J = 6$ we get five wavelet details vectors D_1, D_2, D_3, D_4, D_5, D_6 and one wavelet smooth vector, S_6, each associated with a particular frequency range. Specifically, as shown in Table 1, where the frequency resolution interpretation of wavelet decomposition scale levels for a 6-level decomposition and monthly data is reported, we have that:[9]

- detail levels D_1, D_2 and D_3, represent the very short-run dynamics of a signal (and contains most of the noise of the signal),
- detail levels from D_4 to D_6 captures fluctuations within the $16 - 128$ months frequency range (corresponding to business cycle frequency band),
- the smooth component S_6 captures oscillations with a period longer than 12 years which correspond to the low-frequency components of a signal.

Table 1: Frequency interpretation of MRD scale levels

Scale level, J	Detail level, D_j	Monthly resolution
1	D_1	2-4
2	D_2	4-8
3	D_3	8-16
4	D_4	16-32
5	D_5	32-64
6	D_6	64-128
7	S_6	>128

To select the set of individual indicators at each scale level we investigate the cyclical performance of each component series of the TCB-LEI with respect to the reference series, TCB-CEI, on a scale-by-scale basis using the results from wavelet cross-correlation analysis [10] In order to identify at

[9]Given that the level of the transform defines the effective scale λ_j of the corresponding wavelet coefficients, for all families of Daubechies compactly supported wavelets the level j wavelet coefficients are associated with changes at scale 2^{j-1}. Since scale 2^{j-1} corresponds to frequencies in the interval $f \in [1/2^{j+1}, 1/2^j]$, using monthly data scale 1 wavelet coefficients are associated to $2-4$ month periods, while scales 2 to 6 are associated to $4-8$, $8-16$, $16-32$, $32-64$ and $64-128$ month periods, respectively.

[10]This statistical method can provide a reliable measure of the average lead of the indicator as well as the fit of the regression relationship under the assumptions of a linear relationship between variables and absence of extreme values. However, both problems

any time scale which variables have better leading indicators properties for economic activity we use the correlation value at the peak and the location of the peak of the cross correlation function, since they both provide useful criteria for selecting candidate indicators. A high cross-correlation of the cyclical behavior indicates that the potential component series leads the reference series over the whole cycle, not only at turning points, and the location of the peak provides a reliable alternative indicator of the average lead time.

The results from cross-correlation analysis may be summarized as follows. There are notable differences in the strength and timing of the leading relationships across scales for each variable. At scales corresponding to frequencies greater than 16 months, that is scales D_4, D_5, D_6 and S_6, there is evidence of significant leading relationships for most of the individual indicators, especially at scales D_5 and D_6. By contrast, at scales corresponding to higher frequencies, that is scales D_1, D_2, D_3, there is no evidence of any leading behavior since the magnitude of the association between variables is generally close to zero at all leads; As to the average lead time of the various individual indicators at different scales some indicators are classifiable as "longer-leading" (Interest rate spread, ISM Manufacturing new orders index and Index of expectations) and others as "shorter leading" (Average weekly initial claims, Manufacturers new orders for consumer goods and materials and for nondefense capital goods), with the average lead time for indicators at different time scales increasing as the scale increases, from 4 periods at scale D_4 to 24 periods at scale S_6.

4.2 The wavelet-based composite leading index, TCB-LEI^W

The findings and analysis in previous sections indicate that it is possible to construct a parsimonious and "informationally efficient" composite leading index[11] by simply combining those sub-indexes corresponding to the highest scales, *i.e.* scales D_4, D_5, D_6 and S_6.

After the component series of each sub-index have been selected using wavelet cross-correlation results we apply Conference Board's methodology to construct the scale-based composite sub-indexes. Hence, we adjust the monthly contribution from each component through the equal variance weight method[12] and then add these standardized contributions to finally obtain the level of the composite index using the symmetric percent change formula.

are likely to be limited by performing cross-correlations calculations at time scale rather than aggregate level.

[11] Parsimony and efficiency are related to the ability of reducing redundant information in the construction of the composite index.

[12] Adjustments are based on the inverse of the standard deviation of the component contribution and are also normalized to sum to one.

The individual components selected from wavelet cross-correlation analysis that compose the leading sub-indexes, LEI_{S_6}, LEI_{D_6}, LEI_{D_5} and LEI_{D_4}, are listed in Tables 2 and 3. Tables 2 and 3 refer to different sub-samples, 1960:1-1990:4 and 1990:5-2014:3 respectively (see Levanon *et al.* 2010).

Table 2: Components of TCB-LEI^W sub-indexes until April 1990

LEI_{D_4} Components	Stdz. Factors
ISM Manufacturing, New Orders Index	0.1671
Stock prices, 500 common stocks	0.3068
Building permits, new private housing units	0.2148
Index of consumer expectations	0.3114
LEI_{D_5} Components	
ISM Manufacturing, New Orders Index	0.0925
Stock prices, 500 common stocks	0.1119
Building permits, new private housing units	0.0889
Money supply, M2	0.5946
Index of consumer expectations	0.1121
LEI_{D_6} Components	
ISM Manufacturing, New Orders Index	0.1571
Building permits, new private housing units	0.0722
Money supply, M2	0.5989
Index of consumer expectations	0.1717
LEI_{S_6} Components	
Manufacturers new orders, consumer goods and materials	0.5324
Manufacturers new orders, nondefense capital goods	0.4676

The wavelet-based index TCB-LEI^W is then obtained by summing up the scale-based composite sub-indexes as follows:

$$TCB - LEI^W = \alpha LEI_{S_6} + \beta LEI_{D_6} + \gamma LEI_{D_5} + \delta LEI_{D_4} \tag{4}$$

where α, β, γ and δ are the standardization factors of each sub-index calculated according to the equal variance weight method as before.[13] The wavelet-based composite leading index TCB-LEI^W is shown in Figure 6 along with the Conference Board new leading index TCB-LEI. The graphical examination of the two composite leading indicators clearly indicates that movements of the TCB-LEI^W tend to precede quite regularly those of the

[13]Such values are, respectively, .3572, .3394, .2199 and .0834 for the pre-1990 period and .5049, .2437, .1507 and .1007 for the post-1990 period.

Table 3: Components of TCB-LEI^W sub-indexes from May 1990

LEI_{D_4} Components	Stdz. Factors
ISM Manufacturing, New Orders Index	0.2427
Stock prices, 500 common stocks	0.4455
Building permits, new private housing units	0.3119
LEI_{D_5} Components	
ISM Manufacturing, New Orders Index	0.2281
Stock prices, 500 common stocks	0.2760
Building permits, new private housing units	0.2192
Index of consumer expectations	0.2766
LEI_{D_6} Components	
ISM Manufacturing, New Orders Index	0.3917
Building permits, new private housing units	0.1801
Index of consumer expectations	0.4282
LEI_{S_6} Components	
Manufacturers new orders, consumer goods and materials	0.5324
Manufacturers new orders, nondefense capital goods	0.4676

TCB-LEI.[14]

Figure 6 about here

5 Comparing the performance of TCB-LEI^W and TCB-LEI at a) business cycle turning points, and b) in signalling recessions

In order to evaluate the performance of the new approach in terms of making leading indicators "more leading", we compare the ability of the "wavelet-based" composite leading indicator TCB-LEI^W to provide early signals of turning points in economic activity relative to the TCB-LEI. More specifically, we compare the performance of the two leading indicators using the historical dataset. Indeed, the comparison of the historical estimates is generally adopted as a first step also in the revision process of individual components periodically carried out (e.g. OECD, 2002, and Levanon *et al.* 2011).

We perform a "peak-and-trough" analysis that, along with the analysis of the cross-correlation function and the number of missing or extra cycles,

[14] The visual evidence presented in Figure 6 is confirmed by the cross-correlation analysis between TCB-LEI^W and TCB-LEI where the latter is leading.

are the criteria generally used to evaluate the performance of the individual components of a composite indicator (and the composite indicator itself). The results of the "peak-and-trough" analysis are presented in Table 4, where the homogeneity of leads, in terms of length and consistency of the leading times, at turning points, rather than over the whole sample period are examined. The chronology used for the application of "peak-and-trough" analysis is the official NBER business cycle chronology.

In Table 4 we show the number of leads in months and the difference in leading months at each turning point between the two indicators, TCB-LEI and TCB-LEI^W. The TCB-LEI^W gives signals of all turning points with a longer lead than TCB-LEI, with an average leading time at peaks of 1 year and a half, six months longer than TCB-LEI. This longer leading property with respect to peaks makes the TCB-LEI^W useful as an indication of the approaching end of the current expansionary phase. This feature is clearly detectable from visual inspection of Figure 1. Interesting results also emerge about the average leading time at troughs. Although the TCB-LEI^W provide signals of turning points only slightly in advance relative to the new-LEI, the similarity between the mean and median leading values at troughs and the halved standard deviation, suggest that the signals provided by the TCB-LEI^W about approaching turning points are even more reliable than those from TCB-LEI.

Table 4: Peak-and-trough analysis of the TCB-LEI^W and TCB-LEI (LEI, not M2)

Peaks	TCB-LEI	TCB-LEI^W	*Diff.*	Troughs	TCB-LEI	TCB-LEI^W	*Diff.*
				Febr 1961	-11	-9	-2
Dec 1969	-8	-11	3	Oct 1970	-7	-8	1
Nov 1973	-8	-13	5	Mar 1975	-2	-4	2
Jan 1980	-15	-23	8	Jul 1980	-2	-2	0
Jul 1981	-8	-8	0	Nov 1982	-10	-9	-1
Jul 1990	-18	-9	-9	Mar 1991	-2	-4	2
Mar 2001	-14	-19	5	Nov 2001	-1	-4	3
Dec 2007	-21	-43	22	Jun 2009	-3	-4	1
Mean	-12.8	-18.0	4.8	Mean	-4.8	-5.0	0.75
Median	-12.0	-13.0	5.0	Median	-2.5	-4.0	1.0
Std.Dev.	5.0	12.3	–	Std.Dev.	4.0	2.7	–

Note: Leads in months of TCB-LEI and TCB-LEI^W at peaks and troughs.
The numbers in column *Diff.* measure the difference in leading months between $TCB - LEI^W$ and TCB-LEI at turning points.

The (annualized) six-month growth rate of the composite leading indicator is generally used as an operational recession-warning rule to assess the likelihood of an approaching recession in the short-term by means of the so-called "Three Ds" rule. Indeed, a decline in the leading index is interpreted as a short-term recession warning when duration, deepness (depth) and diffusion of its six-month change (at annual rate) exceed certain specified thresholds. [15]

Figure 7 about here

The six-month growth rate of the Conference Board's leading economic index is shown in Figure 12 along with the six-month growth rate of the wavelet-based leading indicator TCB-LEI^W. The wavelet-based leading indicator seems to provide earlier signals of business cycle turning points as to the Conference Board composite leading economic index anticipating the turning points by several months. Indeed, the cross-correlation analysis between the derived measures of the two indexes shown in Figure 13 indicates that the wavelet-based leading index has an average leading time of about five months.

Figure 8 about here

The wavelet-based approach proposed in this paper, and in particular the selection of the individual components on a scale-by-scale basis, gives the leading composite indicator an inherent smoothing property. Indeed, since the shortest time scale components of a signal are likely to contain mostly noise, the absence of any leading properties at these scales makes the wavelet-based composite indicator smooth. Thus, in order to detect which of the two above mentioned features is responsible for the different performance of the wavelet-based index we need to separate the two effects. Hence, in order to determine whether such smoothness is responsible for the different performance of the two leading indicators we compare the TCB-LEI^W against a smoothed version of the TCB-LEI represented by its centered moving average over 12-months. The plot of the two signals in Figure 14 shows that the tendency of the TCB-LEI^W to shift direction in advance of the TCB-LEI index is not related to smoothing. Therefore we conclude that the different performance of the wavelet-based leading indicator can be referred to the property of the wavelet approach to include a scale-by-scale selection of each individual indicators in the overall composite index.

Figure 9 about here

[15]The thresholds are three months of consecutive decline, a downward movement of three-and-a-half percent over a six-month span and a diffusion index below 50 percent (TCB, 2000).

6 Conclusions

In this paper a wavelet-based approach for constructing composite indexes is applied to the Conference Board's LEI for the US. The proposed methodology allows the construction of a composite leading indicator which combines several sub-indexes whose individual components are selected on a scale-by-scale, rather than aggregate, basis. Our overall findings using historical data indicate that by making selective use of all available information, both in the frequency and time domain, it is possible to improve the performance of composite leading indicators, especially at peaks, by also retaining reliability of the early warning signals provided. In sum, we believe that composite indicators can largely benefit from wavelet's ability to fully exploit the different informative content of individual indicators at different time scales.

The methodology applied in this paper reveals several potential advantages and limits. Potential gains from the "wavelet-based" methodology in constructing composite indicators are first related to the difficulty to recognize signals in noisy data. Since noise is mostly contained at shortest time scales, that is at scales where the leading properties of individual indicators are generally absent, the wavelet-based composite indicator can improve the quality and reliability of early signals of cyclical turning points relative to the actual system of composite leading indicators in terms of number of false signals. Moreover, by reducing the number of individual components used in the construction of the composite indicator the proposed methodology can allow greater timeliness than any other composite indicator by including only series that according to publication schedules make the lag between the release and target (or reference) month to be minimum.

On the other hand, we recognize that in order to fully determine the usefulness of such a wavelet-based approach for developing early warning signals the evaluation of its performance with real time data is needed. Real time evaluation methods are of great interest for all practitioners and also a critical issue for the estimation of the components in the multiresolution wavelet decomposition analysis. Indeed, as new data become available, some revision in past estimations is likely to result if the new data does not match the data extrapolated out-of-the sample by the boundary condition used in wavelet decomposition. This aspect needs to be further investigated in future research for a thorough assessment of the usefulness of the wavelet-based approach to composite indicators.

7 References

Ashley R. and R. Verbrugge, (2006), "Mis-Specification and Frequency Dependence in a New Keynesian Phillips Curve," *Working Papers e06-12*, Virginia Polytechnic Institute and State University.

Ashley R. and R. Verbrugge, (2009), "Frequency Dependence in Regression Model Coefficients: An Alternative Approach for Modeling Nonlinear Dynamic Relationships in Time Series," *Econometric Reviews*, 28 (1-3), 4-20.

Burns A.F. and W, C. Mitchell, (1946), *Measuring Business Cycles*, National Bureau of Economic Research.

Conference Board, (2000), *Business Cycle Indicators Handbook*.

Daubechies, I., (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Engle, R.F., (1974), "Band Spectrum Regression," *International Economic Review*, 15 (1), 1-11.

Engle, R.F., (1978), "Testing Price Equations for Stability across Spectral Frequency Bands," *Econometrica*, 46 (4), 869-81.

Gallegati, Marco, (2008), "Wavelet analysis of stock returns and aggregate economic activity," *Computational Statistics and Data Analysis*, 52 (6), 3061-3074.

Gallegati Marco (2014), "Making leading indicators more leading: a wavelet-based method for the construction of composite leading indexes ", *Journal of Business Cycle Measurement and Analysis*, published on line.

Gallegati Marco (2014a), "Early Warning Signals of Financial Stress: A "Wavelet-Based" Composite Indicators Approach", in Schleer-van Gellecom, Frauke (Ed.) Advances in Non-Linear Economic Modeling, Vol. 17, Springer Series: Dynamic Modeling and Econometrics in Economics and Finance, Springer-Verlag, Heidelberg, 115-138.

Gallegati, Marco, Gallegati, Mauro, Ramsey, J.B., Semmler, W., (2011), "The US wage Phillips curve across frequencies and over time," *Oxford Bullettin of Economics and Statistics*, 73 (4), 489-508.

Gallegati Marco and J.B. Ramsey, 2012, Errors-in-Variables and the Wavelet Multiresolution Approximation Approach: A Monte Carlo Study, in Badi H. Baltagi, R. Carter Hill, Whitney K. Newey, Halbert L. White (eds.) *Advances in Econometrics: Essays in Honor of Jerry Hausman*, 29,149-171.

Gallegati Marco and J.B. Ramsey, (2013), Structural changes and phase variations: a re-examination of the q-model using wavelet exploratory analysis , *Structural Change and Economic Dynamics*, 25, 60-73;

Gallegati Marco, J. B. Ramsey and Semmler (2014), "Output and interest rate spreads: a wavelet decomposition analysis, *Computational Statistics and Data Analysis*, 76, 283-290.

Levanon G., J.C. Manini, A. Ozyildirim, B. Schaitkin and J. Tanchua (2011a), "Using a Leading Credit Index to Predict Turning Points in the U.S. Business Cycle", *Economics Program Working Paper Series* no.11-05, The Conference Board.

McGuckin R.H., Ozyildirim A., and V. Zarnowitz (2007), "A More Timely and Useful Index of Leading Indicators", *Journal of Business and Economic Statistics.*

Levanon G., A. Ozyildirim, B. Schaitkin and J. Tanchua (2011b), "Comprehensive Benchmark Revisions for The Conference Board Leading Economic Index for the United States", *Economics Program Working Paper Series* no.11-06, The Conference Board.

Moore, G. H., and J. Shishkin (1967), *Indicators of Business Expansions and Contractions*, New York: National Bureau of Economic Research, Inc.

Picchetti, Pablo (2008), Wavelet based leading indicators of industrial activity in Brazil, *mimeo*, CIRET Conference, Santiago.

Ramsey, J.B., Lampart, C., (1998a), "The decomposition of economic relationship by time scale using wavelets: money and income", *Macroeconomic Dynamics and Econometrics*, 2 (1), 49-71.

Ramsey, J.B., Lampart, C., (1998b), "The decomposition of economic relationship by time scale using wavelets: expenditure and income", *Studies in Nonlinear Dynamics and Econometrics*, 3 (4), 23-42.

Ramsey, J.B., Gallegati, Marco, Gallegati, Mauro, Semmler, W., 2010. "Instrumental variables and wavelet decomposition". Economic Modeling 27 (6), 1498-1513.

Stock, James H., and Mark W. Watson, "New Indexes of Coincident and Leading Economic Indicators," in O. Blanchard and S. Fischer (eds.), NBER Macroeconomics Annual (Cambridge, Mass.: MIT Press, 1989), 351-394.

Stock, James H., and Mark W. Watson, "A Probability Model of the Coincident Economic Indicators," in K. Lahiri and G.H. Moore (eds.), Leading Economic Indicators: New Approaches and Forecasting Records (Cambridge: Cambridge University Press, 1991), 63-89.

Stock, James H., and Mark W. Watson, "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience," in J.H. Stock and M.W. Watson (eds.), Business Cycles, Indicators and Forecasting (Chicago: University of Chicago Press for NBER, 1993), 255-284.

Stock, J.H., and M.W. Watson (2004), An Empirical Comparison of Methods for Forecasting Using Many Predictors".

Figures

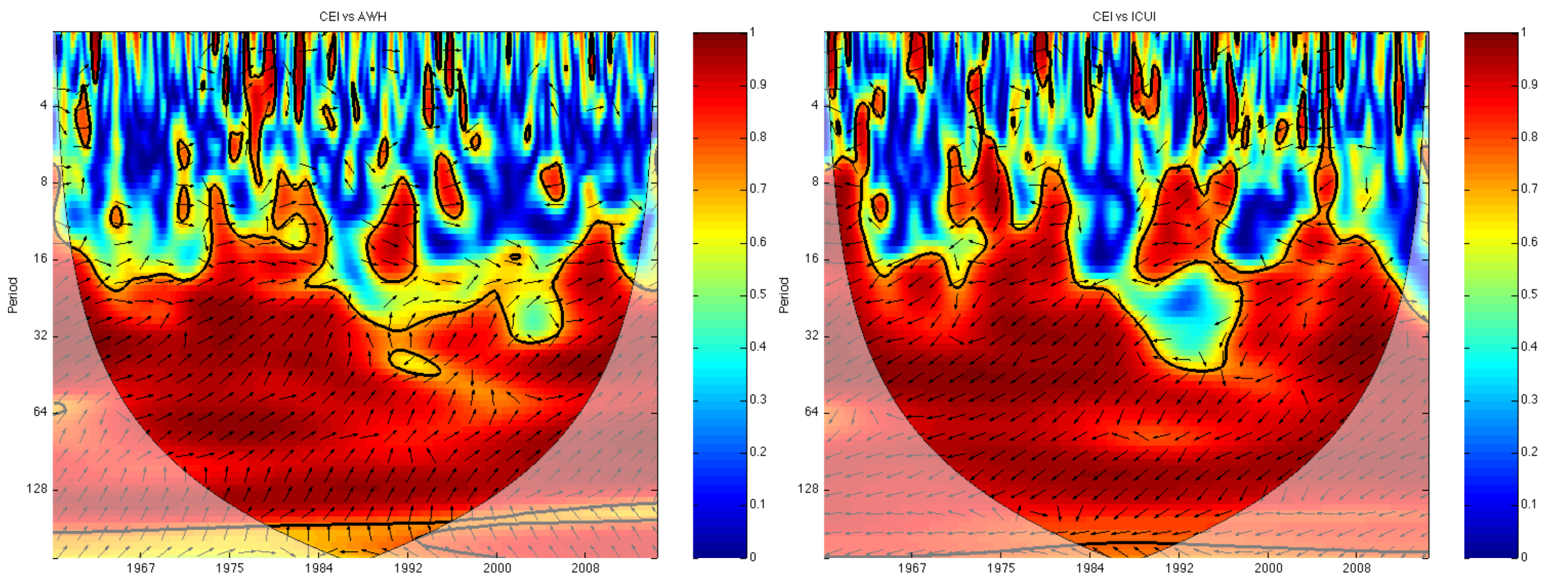

Figure 1: Wavelet coherence between TCB-CEI and Average Weekly Hours (left) and Average Weekly IUC (right)

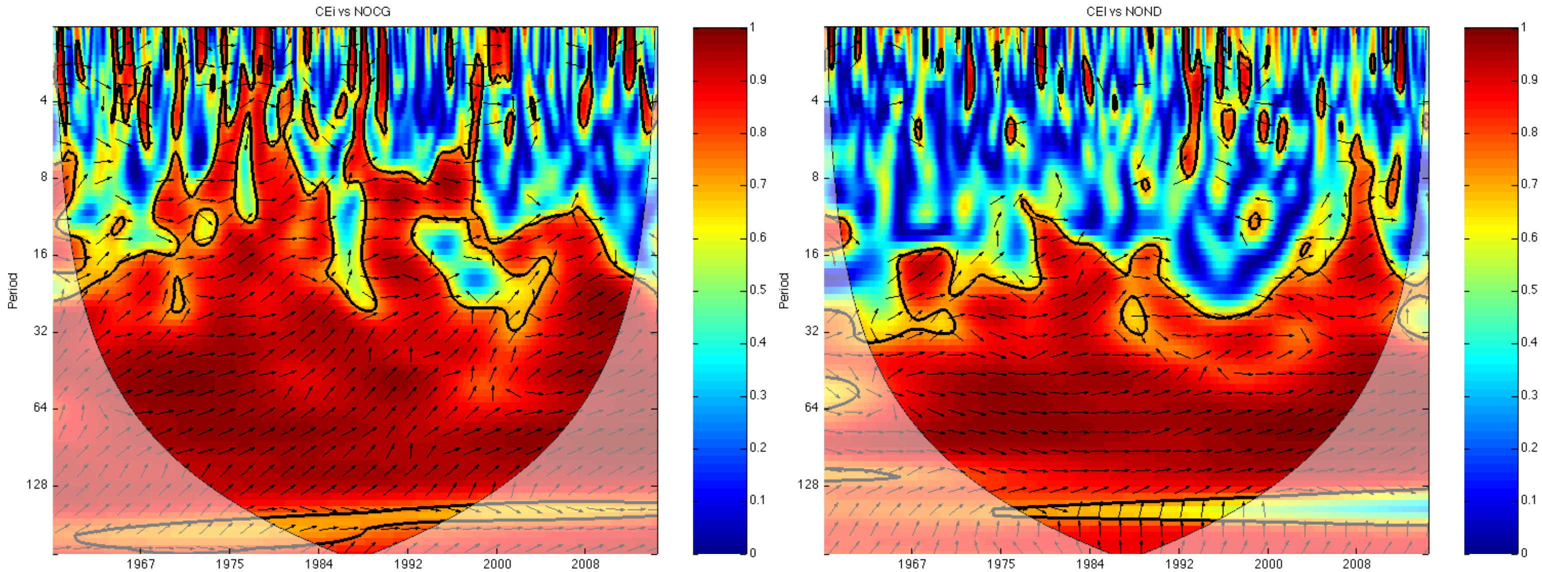

Figure 2: Wavelet coherence between TCB-CEI and NOCG (left) and between NOND (right)

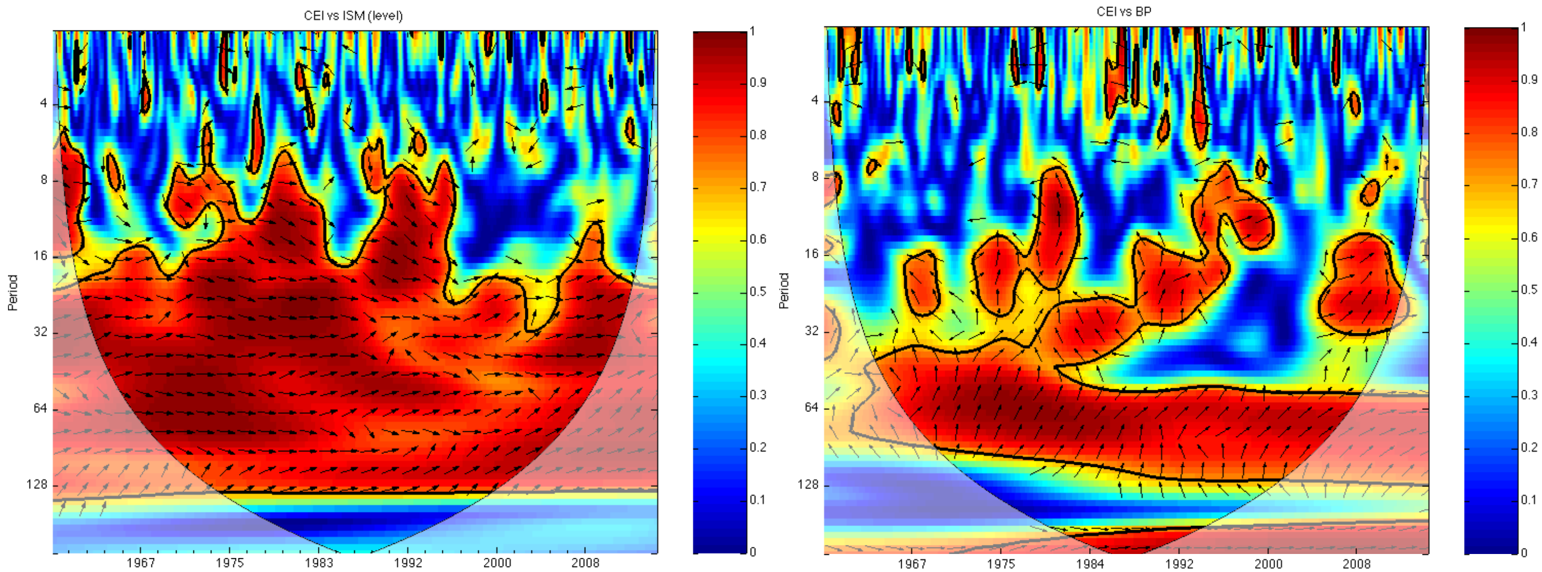

Figure 3: Wavelet coherence between TCB-CEI ISM (left) and between BP (right)

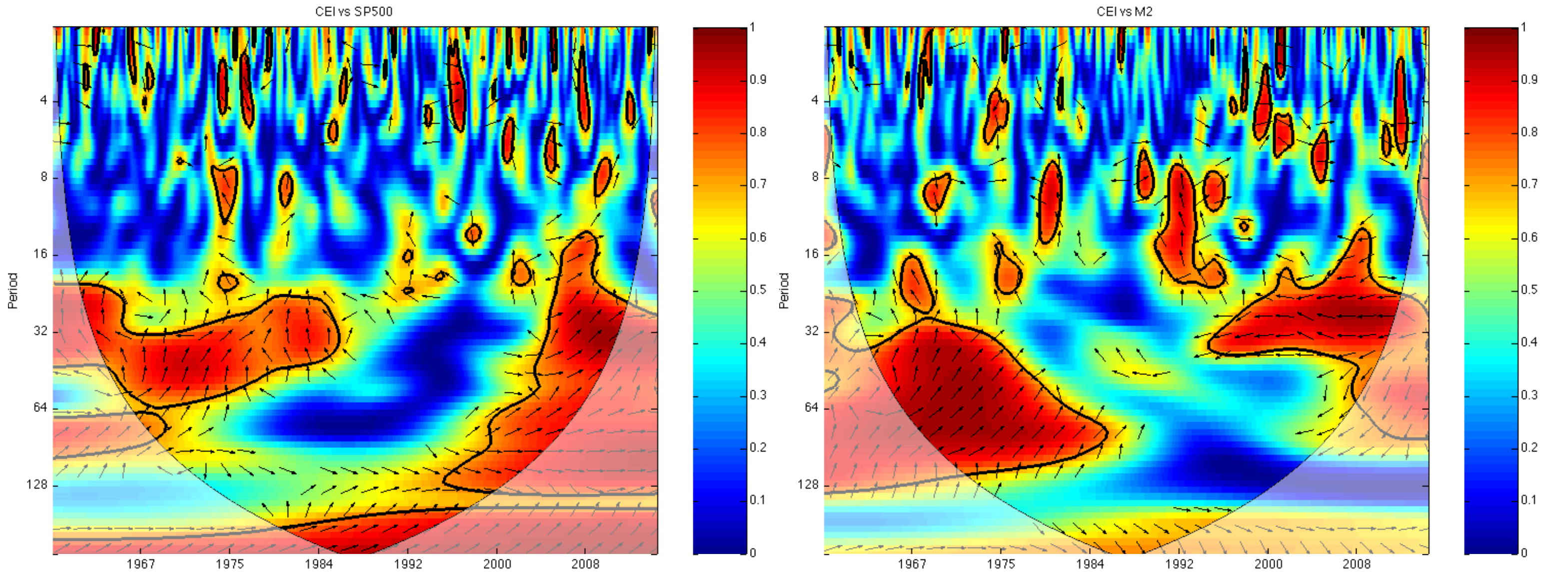

Figure 4: Wavelet coherence between TCB-CEI and SP500 (left) and M2 (right)

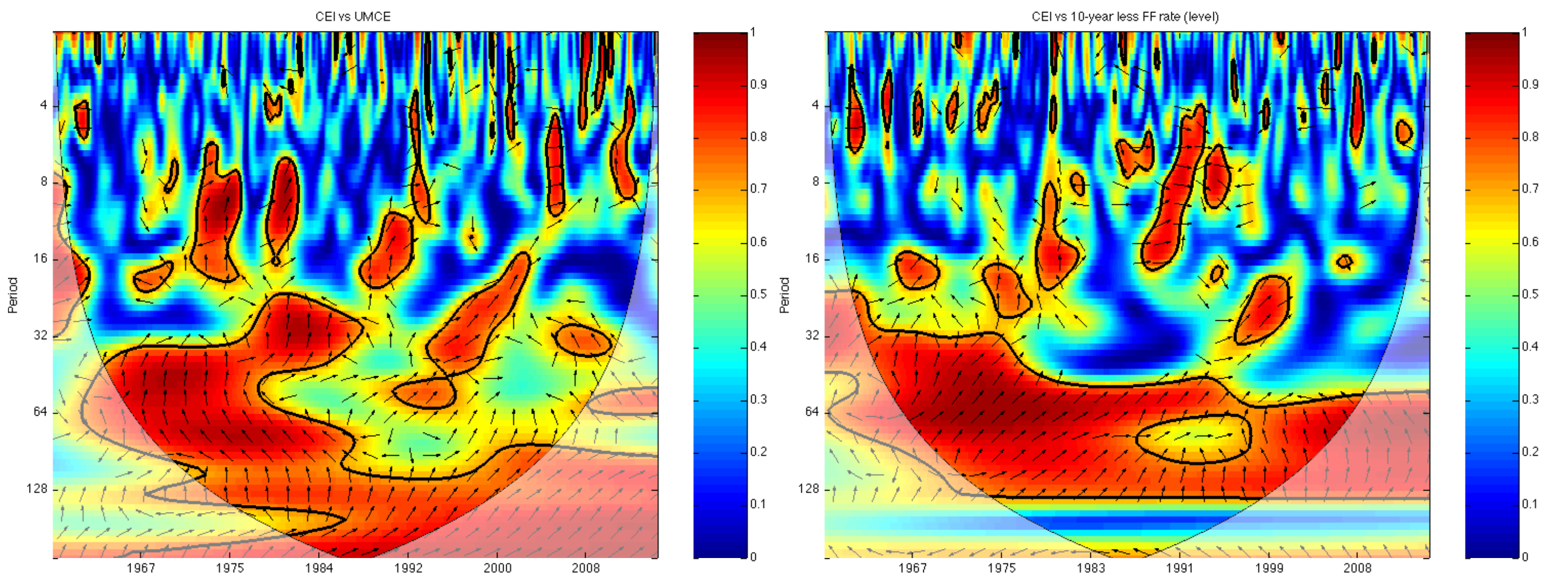

Figure 5: Wavelet coherence between TCB-CEI and UMCE (left) and YS (right)

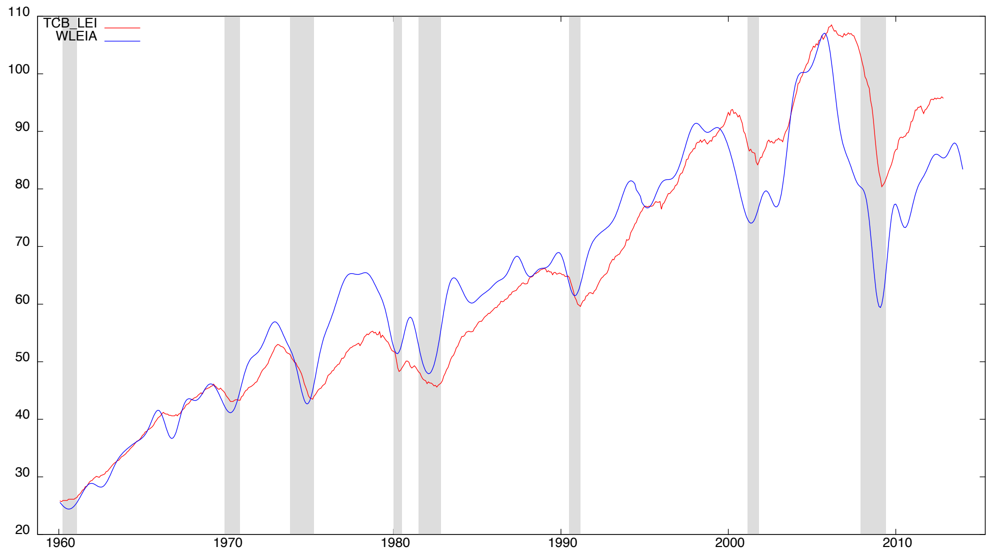

Figure 6: Wavelet-based leading composite indicator $TCB\!-\!LEI^{W}$ vs TCB-LEI

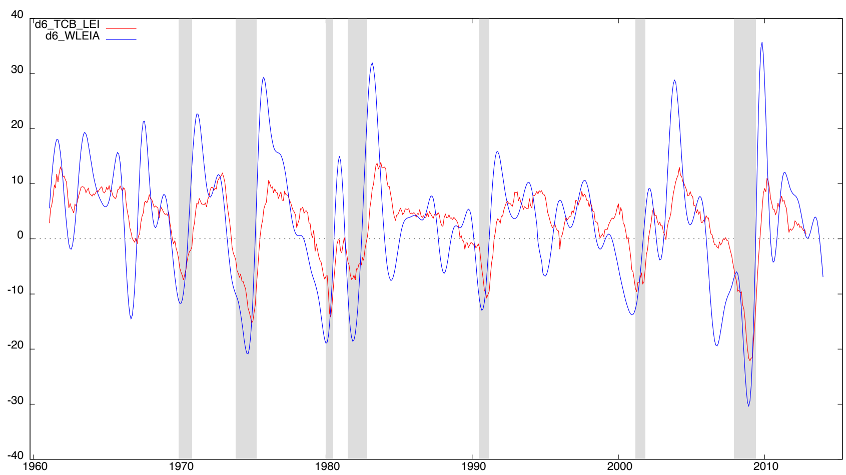

Figure 7: $TCB-LEI^{W}$ vs TCB-LEI: recession signals from 3D's

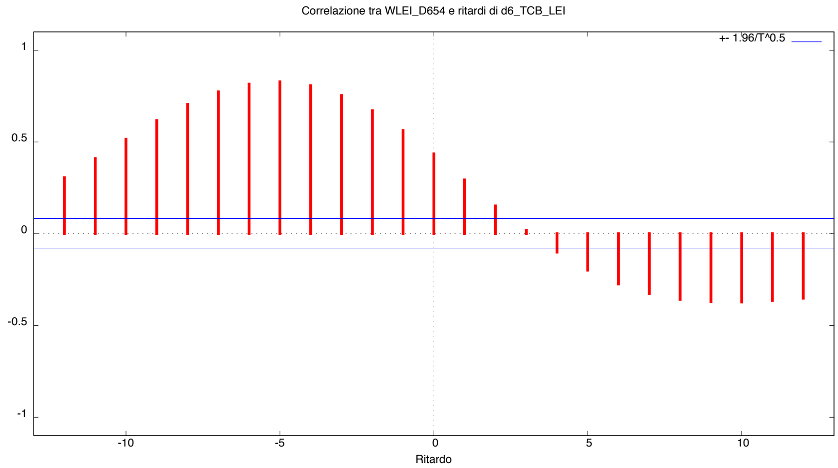

Figure 8: Cross-correlation between the six-month growth rate of TCB-LEI^W at $t-i$ and TCB-LEI at t.

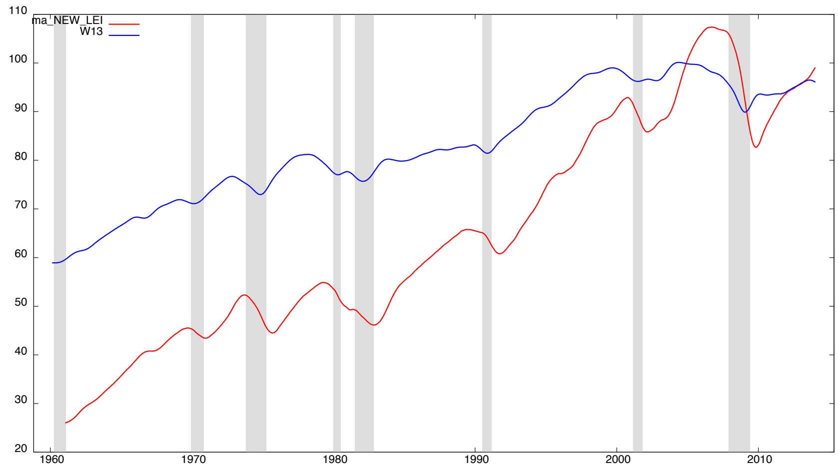

Figure 9: TCB-LEI^W (blue line) and the moving average of new-LEI (red line)