Multiscale Decomposition of Big Data Time Series for Analysis and Prediction of Macroeconomic Data

Livio Fenga

Italian National Institute of Statistics

lfenga@istat.it

Abstract. The problem of the extraction of the relevant information for prediction purposes – in a Big Data time series context – is tackled. This issue is especially crucial when the forecasting activity involves macroeconomic time series, i.e. when one is mostly interested in finding leading variables and, at the same time, avoiding overfitted model structures. Unfortunately, the use of big data can cause dangerous overparametrization phenomena in the entertained models. In addition, two other drawbacks should be considered: firstly, human–driven handling of big data on a case-by-case basis is an impractical (and generally not viable) option and secondly, focusing solely on the raw time series might lead to suboptimal results. The presented approach deals with these problems using a twofold strategy: *i*) it expands the data in time–scale domain, in the attempt to increase the likelihood of giving emphasis to possibly weak, relevant, signals and *ii*) carries out a multi-step dimension reduction procedure. The latter task is done by means of cross–correlation functions (whose employment will be theoretically justified) and a suitable objective function.

Keywords Autoregressive models; big data; distributed lag model; macroeconomic time series; multiresolution analysis; prediction; wavelet theory.

1 Introduction

Theoretical and applied research is nowadays focusing, with greater and greater intensity, on mathematical models aimed at predicting the future values taken by one or more time series. Economics is an area traditionally characterized by high degrees of uncertainty and reliable forecasting methods are always in demand. However,building,

testing and tuning–up a statistical prediction model for economic variables is an absolutely non trivial task. In fact, virtually all the economic–related processes reflect, to a different extent, the very many sources of instability nowadays World is affected from. Such disturbances can inject unwanted phenomena into the time series under investigation, e.g. in the form of noise, spikes, asymmetric cycles and other idiosyncratic components and irregularities. Under such conditions, even sophisticated forecasting models – whose construction often require heavy, conjoint efforts from different qualified actors (e.g. academic researchers, economists, statisticians and practitioners) – generally deliver suboptimal performances or can fail altogether. In such a situation, unconventional information of the type big data can be fruitfully employed. In fact, once extracted and properly processed, those data might possess valuable leading capabilities. In addition, they can show many other desirable properties such as: *i*) real-time availability; *ii*) free of charge accessibility; *iii*) great flexibility. However, sometimes, those advantages might turn into dangerous misleading factors. Tempted by the sometimes overwhelming quantity of available information, the analyst might use too much of it and include too many and/or logically unrelated variables in the entertained statistical models. Such an indiscriminate use of the available information set is against the parsimony principle and, as a result, not negligible to big amounts of uncertainty are injected into the analysis.

The strategy proposed in the present paper is a multi-step procedure aimed at coping with such a scenario; in fact, it is designed to extract a small subset of useful information (mainly in the form of wavelet components) employable in a parsimonious, statistically correct, linear prediction models.

2 The proposed method

As an evolving term, it is difficult to give an exact definition of big data. In general, this term relates to any huge amount of structured, semistructured and unstructured collection of (possibly multi-source) data, which can be computationally

exploited to the end of gaining useful information. Such data–sets can be extracted from official and unofficial sources. The former are usually archives maintained and disseminated by public organisms (ministers, central banks as well as national and supranational statistical bureaus) or private institutions (e.g. department of motor vehicles, research companies, transportation authorities). Unofficial sources are a much more complicated matter. In this case, in fact, the information is made available usually with no guarantees about its (potential) statistical usability. Being built and disseminated by many kinds of entities, by design outside the "statistical control" of certifying organisms, these data should be always taken with certain reservations. Clearly, it is always possible to conduct different types of tests to check their statistical properties (e.g. in terms of probabilistic structure, presence of dominant frequency components, linearity, coherence with similar phenomena) – or to adjust them (e.g. in case of outliers, trading day effects or periodic components). However, all these actions can be generally taken *ex post*. While there are many on-line big data sources – such as Google Trends, Google AdWords or Twitter data, just to cite a few, in the present paper we restrict our attention to one in particular, which stores the queries submitted to the search engine Google Search, i.e. Google Trends (for Italy the address is https://trends.google.it/trends/?geo=IT). In more details, by means of Google Trends (henceforth GT) data, we can measure the popularity of top search queries, across various regions and languages. Prior to their diffusion, the data are "normalized", (i.e. only the relative level of interest over time for a keyword phrase is made available). GT data are publicly and readily available free of charge and can be easily customized – other than for different configurations of the keywords of interest – also for geo-localization and time span. However, those data embody, to different extents, the sources of uncertainty above mentioned; in particular little or none is known about their quality control as well as the type and size of the population of reference, which, all the more so, changes over time. By design, GT data clearly rule out significant portions of a given population (i.e. those with no computer alphabetization or limited or no access to proper hardware equipments). Other uncontrollable sources of bias are technical – e.g. prolonged power outages, network failures, speed

and other network-related technological constraints – or even political (certain States regulate or prohibit WWW–based activities). Finally, there are always chances that Google Search queries are submitted by non-human entities, e.g. Artificial Intelligence algorithms.

2.1 Outline of the method

The proposed method, called *W–BD* (Wavelet Big Data), uses a multi–step procedure aimed at iteratively and progressively reducing the complexity of a given big-data set. Narrowed down to a workable number of variables, the search space is finally investigated using methods commonly employed in time series analysis. However, unlike many available dimension reduction procedures, in the case of the present paper the information set is initially expanded to include, along with the original (potentially interesting) variables extracted, also their wavelet decomposition. In more details, the focal point of *W–BD* method rests on the fact that the relevant, (usually tiny) portion of information is not only "hidden" in the time domain – represented by the columns of length T (the time span) of a huge dimension time series matrix of size $T \times N$ – but it is also assumed to be embedded in the time–scale domain. Therefore, the search space is double, being represented by both time *and* time–scale spaces; the former is explored in the usual theoretical framework of the times series analysis whereas the latter is investigated through a set of equations, designed to generate wavelet sequences and achieve the deconstruction of possibly complicated and/or non–linear dynamics. In particular, the method employed here – called Multi Resolution Analysis (MRA) (see, for example, **?**, **?** and **?**) – allows the decomposition of a given time series into its resolution levels, until a prefixed, bounded and arbitrary determined, maximum level.

W-BD method applies the MRA technique to a predefined, empirically chosen set of Google Trends time series, say $\Omega \subset \aleph$; $|\Omega| << |\aleph|$, being $\aleph$ the "big set" containing the totality of the GT time series (our population of reference) and the symbol $|\cdot|$ indicating the cardinality function. Once applied to Ω, MRA generates the set Ω'_q, of size

q. In practice, Ω'_q is an expanded version of Ω as it stores all the wavelet components the time series belonging to Ω are decomposed into. Therefore, it will be $|\Omega'_q| > |\Omega|$ by a factor of $(J+1)$, being J the number of the (arbitrarily chosen) wavelet decomposition levels. In general, Ω'_q is high dimensional, with possibly hundreds or thousands of elements and hopefully contains a low–dimensional subset containing only the relevant information. This subset is obtained by applying *ad hoc* dimension reduction procedures which are based on i) cross-correlation functions (whose employment in this particular context will be discussed and justified in Paragraph 4) between the different wavelet components in Ω_q and the time series of interest and ii) an objective function (discussed in Paragraph 6.1), designed to measure the forecasting ability of a multivariate predictor in comparison with its univariate counterpart. By means of the function sub i), a subset called Ω'' is generated whereas the function sub ii) is used to define a more refined subset, denoted by the symbol Ω°. Further, human–driven, steps based on the adopted target function – along with a through evaluation of a small number of competing equations used for the reconstruction of the underlying process – are needed to deliver the final forecasting model.

Despite the fact that *W-BD* is purposely designed to extract and employ wavelet decomposed time series (along with the lagged dependent variable), there are always chances that one of the following two scenarios takes place:

i) one or more series belonging to Ω (i.e. the set containing only the raw, non-decomposed series) and to Ω'_q (which contains the decomposed series) are selected. This case excludes the contemporaneous presence, in any of the entertained models, of a given series $x_{i,t}$ and its wavelet decomposition $\mathbb{W}_j(x_{i,t})$ at any level j.

ii) exclusively raw series are selected.

In other words, in i), the multivariate predictor is built using both decomposed and not decomposed time series. On the other hand, the case sub ii) refers to a (mostly

theoretical) case where no decomposed time series are found useful (and therefore *W-BD* method collapses to a standard multivariate model).

In the sequel, the set containing the non–decomposed time series (as in the case ii) will be denoted by the symbol $\overline{\Omega}$.

More formally, let Ω^* be the final subset containing the "best" variables, four different configurations of the problem are possible:

a) $\Omega^* \subset \overline{\Omega}$ – being $|\Omega^*| << |\overline{\Omega}| + |\Omega'_q| \equiv |\overline{\Omega}|$ – i.e. the final subset comprises only not decomposed time series (fact that rules *W–BD* out);

b) $\Omega^* \subset \Omega'_q$ – being $|\Omega^*| << |\overline{\Omega}| + |\Omega'_q| \equiv |\Omega'_q|$ – i.e. improvements are expected by employing solely decomposed time series;

c) $\Omega^* \subset \left[\overline{\Omega} \cup \Omega'_q\right]$ – being $|\Omega^*| << |\overline{\Omega}| + |\Omega'_q|$ – i.e. improvements are expected by using both original and decomposed time series;

d) $\Omega^* \equiv \emptyset$. In this case the current information set cannot support *W–BD*.

Under case c), to avoid redundancy, a statistical model must not embody time series belonging to both $\overline{\Omega}$ *and* Ω'_q. On the other hand, case a) falls into the more traditional regression framework, which obviously rules out *W–BD* method, as well as the case sub d). The most interesting case is the one under b) and its straightforward extension c).

To summarize, as a result of a first overview performed on the available big data set $\aleph$, the set Ω is extracted. Its multiresolution–driven expansion generates the set $\Omega'_q = \mathbb{W}(\Omega)$, with the subscript q referring to its size (the number of time–scale components), whereas the Wavelet transforming function (i.e. $\mathbb{W}$) is of the type Maximum Overlapping Discrete Wavelet Transform (MODWT) **?** and will be illustrated in the

next Section 3. A double step procedure is finally applied on Ω'_q to built a lower dimension set, which will be denoted by the symbol Ω°. As it will discussed later, Ω° is not the final subset (already defined with Ω^*), as it provides just one of the possible configurations the final set of regressors might take. This is consistent with the fact that the adopted target function gives only rough indications on the prediction capabilities of the selected variables, whose inclusion in a final statistical model must be carefully evaluated and tested.

Finally, lag selection of the dependent variable has not be considered yet, as it falls within the more standard domain of autoregressive order determination. The related procedure, as well as its integration in the *W–BD* scheme, will be discussed in Paragraph 3. For now, we only need to denote the information set represented by the dependent variable's (y) own past, i.e. $\{\widetilde{\Omega}_y \equiv y_{T-1}, y_{T-2}, \ldots, y_1\}$.

2.2 Statement of the problem

As already pointed out, big data should be looked at as complex information systems embedding useful structures, whose patterns can be hard to detect, given the high dimensionality of the problem. Through *W–BD*, the strong uncertainty associated with the use of big data (here of the type GT) is dealt with by extracting only the relevant information and employing it to enhance the forecasting power of a linear predictor. This is mainly done by (automatically) analysing the leading features of i) each of the time–scale decomposition levels the GT series have been broken into and ii) the raw time series belonging to $\overline{\Omega}$, with respect to y_t (i.e. our dependent variable, here in the form of a given macroeconomic time series). The prediction capabilities of the decomposed series are compared with those obtained by using an information set made up only by y's own past $\{\widetilde{\Omega}_y\}$. Formally, what we are looking for is a subset of $\{\Omega'_q\}$ and $\{\widetilde{\Omega_y}\}$ so that the following condition is verified:

$$MSPE(\mathbb{E}[y_{t+m}]|\{\Omega^* \cup \widetilde{\Omega}_y\}) < MSPE(\mathbb{E}[y_{t+m}]|\widetilde{\Omega}_y),$$

being $\{\Omega^*\}, \{\widetilde{\Omega}_y\}$ as above defined and $MSPE$ the Means Square Prediction Error, evaluated at forecast horizons $m = 1, 2, \ldots, M$. Any given raw series $x_{t,i}$ and its wavelet decomposition $\mathbb{W}_i(x_{t,i})$ are mutually exclusive (cannot be simultaneously belong to Ω^*). We are therefore interested in assessing the prediction performances achieved by conditioning the dependent variable to its own past *and* the information obtained from a suitable set of (lagged) wavelet components (or the correspondent raw series). In such a context, as it will be illustrated, the "natural" benchmark – which is adopted here – is the pure autoregressive model.

3 The analysis in time–scale

The degree of complexity characterizing many macroeconomic time series would greatly benefit from the outcomes of an analysis with arbitrary high levels of resolution in both time and frequency domains. Unfortunately, the Heisemberg uncertainty principle dictates that the more precisely a measurement is taken in frequency (time) domain the more the lack of resolution will result in time (frequency) domain. Therefore, if one wish to investigate the dynamical behaviour of a given realization of a stochastic process expressed at different frequencies, it is unavoidable to deal with the consequences of this principle and, as a result, to find a compromise. Fourier–type transforms are generally not an option since, under this set-up, time information is lost and it is not recoverable. Furthermore, by design Fourier spectrum is only suitable for stationary time series and shows no sensitivity towards transient relations or structural changes. A good compromise between considering the pure time series data (maximum time localization information) and looking at data through a Fourier analysis in frequency space (maximum frequency localization information) is the analysis of the type wavelet, as it is designed to retain some frequency localization and some time localization. In practice, through the wavelets, the spectral characteristics of a given time series can be estimated as a function of time. As a result, a sufficiently clear picture of the time evolution of its different periodic components is generally allowed. In more details, the wavelet theory deals with the Heisenberg uncertainty principle by means of the MRA, i.e. by analysing signals at different frequency bands with

different levels of resolutions: higher frequencies (showing good time resolution and poor frequency resolution) and lower frequencies (showing good frequency resolution and poor time resolution). In this sense, MRA is a useful tool for most real–life time series, i.e. composed of long-lasting, lower frequencies (approximations) and short-lasting higher frequencies (details).

Let $L^2(\mathbb{R})$ be the space of all squared integrable functions (otherwise known – e.g. in the case of signal processing – as space of function with finite energy) on $\mathbb{R}$ and $\psi = f(t) \in L^2(\mathbb{R})$ be a fixed generic function. Function $\psi = f(t)$ is a wavelet under the following two conditions: $\int_{-\infty}^{\infty} \psi(\cdot)du = 0$ and $\int_{-\infty}^{\infty} \psi^2(\cdot)du = 1$. Furthermore, to be a wavelet, ψ must also meet (among others) the admissibility condition. Let $\Psi(f) = \int_{-\infty}^{\infty} \psi(u)e^{-i2\pi fu}du$ be the transformation of ψ in the Fourier sense, this function is admissible under $0 < C_\Psi < \infty$, being $C_\psi = \int_0^{\infty} \frac{|\Psi(f)|^2}{f}df$. The last relation is important as it allows the reconstruction of a given function $f(x)$ from its (continuous) wavelet transformation. Wavelet functions have strong practical implications since they are subjected to the shifting and scaling operations, i.e. $\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})$, being $b \in \mathbb{R}$ and $a \in \mathbb{R}^+(a \neq 0)$, respectively the translation and the dilation parameter. The energy normalization factor $a^{-\frac{1}{2}}$ serves the purpose of equating scale–wise the values yielded by the square integrability of $\psi_{a,b}$. In other words, this function guarantees the decomposition of a square integrable function in terms of its dilated and translated wavelet so that the energy in the transform space can be equated to the signal energy. A given input signal $x = f(t) \in L^2(\mathbb{R})$ is continuously wavelet–transformed (CWT) according to: $W_x(a,b) =< x, \psi_{a,b} >= \int_{-\infty}^{\infty} x(t)\psi^*_{a,b}(t)dt = \frac{1}{a}\int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{a})dt$, being $W_x(a,b)$ the scaled and localized wavelet coefficients whereas with ψ^* and $< \cdot, \cdot >$ the complex conjugate of ψ and the scalar product in $L^2(\mathbb{R})$ are respectively denoted. This last relation is invertible and therefore the reconstruction of the original signal from its wavelet decomposition is always possible, through the so-called reconstruction of identity operation, i.e.: $x(t) = \frac{1}{C_\psi}\int_{\mathbb{R}}\int_{\mathbb{R}} W_x(a,b)\psi_{a,b}\frac{da\ db}{a^2}$. CWT is highly redundant and thus its discrete version (Discrete Wavelet Transform, DWT) – which associates wavelets to orthonormal bases in $\mathbb{R}^2$ – is often used instead. By signal dis-

cretization, a time series $x_t;\ t = 1, 2, \ldots T$, of sample size T and finite energy, can now be expressed using an orthonormal wavelet basis, i.e.

$$x(t) = \sum_m \sum_n d_m^n \psi_n^m(t), \tag{1}$$

being $\psi(t)$ the analyzing wavelet and d_n^m expressed as follows:

$$d_n^m = \int_{-\infty}^{\infty} x(t)\psi_n^m(t)dt. \tag{2}$$

The orthonormal basis functions are all the dilations (denoted by the Latin letter m) and translation (denoted by the Latin letter n) associated to this function. The set of basis functions can be expressed as $\psi_n^m(t) = 2^{\frac{m}{2}}\psi(2^m t - n)$. From (1), the contribution of x_t, ascribable to an admissible level m wavelet decomposition, is given by $d_m(t) = \sum_n d_n^m \psi_n^m(t)$. By virtue of this last equation, one can assess *i*) the dynamics of x_t across different scales and *ii*) the contribution of each of them as a fraction of the total energy. The computation of the quantities in (1) and (2) is crucial for *W–BD* method, as they make possible the decomposition of a time series. The method employed to this end, devised by Mallat **?**, is called *MRA*. This approach is designed to build orthogonal wavelet basis and, practically, results in an algorithm employing both wavelet functions and filter banks. The *MRA*–based decomposition of a given time series is achieved by sequentially performing its transformation into a series of details and approximations at progressively coarser levels. Let $\phi(t)$ be the scaling function and a_n^m the function associated with both the wavelet $\psi(t)$ and its detailing coefficients d_n^m (see Equations 1 and 2). At a given scale m, both a_n^m and d_n^m can be computed using $a_n^{(m+1)}$, i.e. the scaling coefficient at the next finer scale, i.e.: $a_n^m = \sum_l h[l-2n]a_l^{m+1}$ and $d_n^m = \sum_l g[l-2n]a_l^{m+1}$, being $h[n]$ and $g[n]$ respectively the low–pass and the high–pass filters. In other words, a_n^m and d_n^m are signals resulting from the convolution with the high pass – low pass filters followed by a downsampling of factor 2.

In the present paper, MRA has been performed through a modified version of DWT, i.e. the Maximum Overlapping DWT ($MODWT$). This is an highly redundant, non orthogonal method which has been chosen mainly for the following two reasons: *i)* it can work with time series of arbitrary length (in contrast to orthogonal DWT which requires the sample size to be an integer multiple of 2^J) and *ii*) the signals a_n and d_n^m are associated with 0–phase filters. Under MODWT the perfect recovery of the original signal is allowed. Let $\boldsymbol{x}$ be a T–dimensional vector representing the real valued input signal $\{x_t;\ t = 0, 2, \ldots, T-1\}$. For each admissible level J_0 the MODWT procedure results in length T, $J_0 + 1$ vectors, say $\boldsymbol{W}_j$ and $\boldsymbol{V}_j$, respectively storing the wavelet and the scaling coefficients. Following **?** these two vectors can be respectively written as

$$\mathbf{W}_j = \boldsymbol{\Lambda}_j \boldsymbol{x} \tag{3}$$

and

$$\mathbf{V}_j = \boldsymbol{\Theta}_j \boldsymbol{x}, \tag{4}$$

being both $\boldsymbol{\Lambda}_j$ and $\boldsymbol{\Theta}_j$ $T \times T$ matrices respectively storing the values of the wavelet and scaling filters. Equations 3 – 4 are respectively computed as follows: $W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}, x_{t-l \bmod N}$ and $V_{J,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l}, x_{t-l \bmod N}$, with $\left\{\tilde{h}_{j,l}\right\}$ and $\{\tilde{g}_{j,l}\}$ being the length L, level j equivalent wavelet and scaling filters for the transform of the type DWT, above denoted by the symbols $g[\cdot]$ and $h[\cdot]$. $MODWT$ is a filtering approach aimed at modifying the observed series x_t by artificially introducing an extension of it. In practice, unobserved samples $[x_{-1}, x_{-2}, \ldots, x_{T-2}, x_{T-1}, x_0]$ are assigned the observed values $[x_{T-1}, x_{T-2}, \ldots, x_0]$. This method considers the series as it were periodic and is known as using circular boundary conditions, where wavelet and scale coefficients are respectively given by:

$$\widetilde{W}_{j,t} = \sum_{l=0}^{T-1} \tilde{h^{\circ}}_{j,l}, x_{t-l \bmod T}, \tag{5}$$

$$\widetilde{V}_{J,t} = \sum_{l=0}^{T-1} \tilde{g^{\circ}}_{j,l}, x_{t-l \bmod T}, \tag{6}$$

being $\tilde{h}^{\circ}_{j,l}$ and $\tilde{g}^{\circ}_{j,l}$ respectively the periodization of $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$.

Let $\widetilde{\boldsymbol{\Lambda}}_j$ $\widetilde{\boldsymbol{\Theta}}_j$ two $T \times T$ matrices respectively storing the values of the wavelet and scaling filters computed in (5) and (6), the signal is recovered using the following equation:

$$\boldsymbol{x} = \sum_{j=1}^{J_0} \widetilde{\boldsymbol{\Lambda}}_j^T \widetilde{\boldsymbol{W}}_j + \widetilde{\boldsymbol{\Theta}}_{j_0}^{\boldsymbol{T}} \widetilde{\boldsymbol{V}}_{j_0} = \sum_{j=1}^{J_0} \widetilde{D}_j + \widetilde{S}_{j_0}. \tag{7}$$

For a given level $j \quad (j = 1, 2, ...J)$, MRA delivers the coefficients set D_j, which reflects signal local variations at the detailing level j, and the set S_{J_0}, accounting for the long run variations. Finally, in the present paper, $MODWT$–based MRA has been performed by means of the "Pyramid Algorithm" **?**. Being able to perform MRA using only O(N) multiplications (in contrast with a "brute force" approach which requires N^2 operations), it can be considered a computationally efficient algorithm. The detailed explanation of the "Pyramid Algorithm" is beyond the scope of the present paper (the interested reader is referred to **?** and **?**).

4 Detection of the leading components

We are interested in evaluating an expectation function of the type

$$\mathbb{E}([x(t+m)], y_t), \tag{8}$$

which expresses the linear statistical relationships between the time series of interest (y_t) and the generic time series x_t, which can belong either to the set comprising i) the original time series $\{\Omega\}$ (Subsection 4.1) or ii) to the set of the decomposed time series $\{\Omega'_q\}$ (Subsection 4.2).

4.1 The case of raw time series

The case sub i), dealing with time series belonging to Ω, poses no theoretical nor practical problems, as it falls within the well consolidated framework of the linear analysis of bivariate stochastic processes. In practice, the leading GT time series are searched using the traditional approach, i.e. by sequentially employ the empirical

cross–correlation function ($ECCF$) – computed until a predefined, arbitrary lag m – between the given GT variables and the variable of interest y_t. The use of such a function is particularly suitable for big data, as it is fast to compute and susceptible of being put in easy-to-interpret and handle format, either graphical (e.g. in the form of matrix of graphs) or tabular.

4.2 The case of wavelet components

The assessment of the leading capabilities of a decomposed time series is performed – as in the case of raw time series – through the analysis of the empirical cross correlation function. However, this time the focus is on the the cross correlation structure arising between the level j wavelet decomposition of a generic time series x_t – $x_t \in \Omega'_q \subset \mathbb{W}(\Omega)$ – and the original variable of interest y_t. To this end, an estimator of (8) is proposed through the following **theorem**:

let

$$\boldsymbol{\Gamma}^{x,y,m} = cov(x_{t+m}, y_t) \tag{9}$$

be the lag m cross-covariance sequence between two random vectors x and y, assuming the following conditions to hold, i.e.

a) (x, y) forms a jointly, second–order, discrete stationary process, i.e. satisfying the following conditions: i) each of the series are second-order stationary and ii) the autocovariance function depends only on the lag $m = 1, 2, \ldots;\ \forall t$;
b) $x = f(t)$ satisfy the above stated (see Section 3, page 5) admissibility conditions;
c) x's maximum level of decomposition, say J, satisfies the boundary conditions, $J \leq (K+1)$; $\quad T = 2^K$, $K \in \mathbb{Z}^+$,

the function $\boldsymbol{\Gamma}^{x,y,m}$ can be decomposed as:

$$\boldsymbol{\Gamma}^{x,y,m} = \sum_{j=1}^{J} \Gamma\left[w_j(x_t), y_t\right]_m + \Gamma\left[v_j(x_t), y_t\right]_m, \tag{10}$$

being $\Gamma\Big[w_j(x_t), y_t\Big]_m$ the cross covariance function computed between the wavelet coefficients and the unfiltered time series (the wavelet–signal cross covariance) at level j and scale 2^{j-1} and $\Gamma\Big[v_j(x_t), y_t\Big]_m$ the cross covariance function computed between the scaling coefficients and the unfiltered time series (the scaling–signal cross covariance) at level J and scale 2^J.

Proof Equation (10) has been derived using the results in **?**, where the validity of wavelt covariance decomposition is established component–wise in a $MODWT$–based MRA set-up. In more details, by denoting function (9) with $\Gamma_0^{x,y,m}$, where the subscript 0 serves the purpose of particularizing this function to the formal context as reported in **?**, the related decomposition takes the form

$$\boldsymbol{\Gamma}_0^{x,y,m} = \sum_{j=1}^{J} \Gamma\Big[w_j(x_t), w_j(y_t)\Big]_m + \Gamma\Big[v_J(x_t), v_J(y_t)\Big]_m.$$

Now, keeping in mind that – under asymptotic conditions $J \to \infty$ (i.e. the number of available scales goes to infinity) – it holds true that the sum of all available wavelet covariances yields the covariance between x_t and y_t. Therefore, using the results in **?**, Equation 22, page 180, it is

$$cov(x_t, y_t) = \Gamma^{x,y,m} = \sum_{j=1}^{\infty} \Gamma\Big[w_j(x_t), w_j(y_t)\Big]_{m=0}, \tag{11}$$

recalling that, under Gaussianity of x_t and y_t, the MODWT estimator of wavelet covariance is unbiased and asymptotically normally distributed.

Since MRA allows perfect recoverability of the original signal (7), i.e.

$$y_t = \sum_{j=1}^{J} w_j(y_t) + v_J(y_t), \tag{12}$$

the function $\Gamma^{x,y,m}$ can be asymptotically decomposed as follows:

$$\boldsymbol{\Gamma}^{x,y,m} = \sum_{j=1}^{J} \Gamma \Big[w_j(x_t) + v_J(x_t), y_t \Big]_m ; \quad J \to \infty.$$

Using the result in **?**, Equation 23, page 180, the finite sample counterpart of (10) can be expressed as follows: $\boldsymbol{\Gamma}^{x,y,m} = \sum_{j=1}^{J} \Gamma \Big[w_j(x_t), y_t \Big]_m + \Gamma \Big[v_J(x_t), y_t \Big]_m$. This ends the proof.

Due to the the MODWT's boundary conditions, by using all of the wavelet coefficients at all scales, biased estimations of the autocovariance function are unavoidable. However (see Equation 24, page 18 in **?**), for large T the 'biased' estimates of the cross–covariance function generated by $\Gamma^{x,y,m}$, should be approximately equal to those obtained from the wavelet-based estimator calculated without involving the MODWT coefficients.

The above stated condition a), validating Equation 9, must be carefully evaluated for the cross-covariance function (10) to be valid. However, we know that, in big data contexts, checking for such a condition on a case-by-case basis it is not a viable option. To circumvent this problem, the idea is to apply an "high" order autoregressive filter to the time series belonging to both Ω_q' and Ω and to work with the resulting white noise residuals. This technique, called prewithening, is suitable for big data contexts as can be easily implemented and automatically run. A detailed explanation of this method is beyond the scope of the present paper (the interested reader is referred, for example, to **?**, and, more recently, to **?**). Here, it is worth pointing out two facts about prewhitening: firstly it is here used only for variable selection purposes (i.e. to find the most predictive lag(s)) whereas the rest of the analysis is carried out using the original variables; secondly, this technique not only is a powerful bias reduction method but it is also a very fast procedure. In fact, prewhitening methods do not assume the specification of the "correct" model, being in general sufficient to find one able to generate a white (or close to white) noise residuals' sequence. This is the reason for which it is not uncommon to prewhite time series through simple (possibly

differenced) autoregressive models (thus avoiding more complicated structures, e.g. involving moving average operators).

Under prewhitening, (10) is rewritten in terms of white noise residual sequences as follows. Let:

a) the generic wavelet sequence $[w_j(x_t)]^*$ be the white noise residuals resulting from a "sufficiently" high (to ensure whiteness) $\hat{p}$ order autoregressive models, i.e. $[w_j(x_t)]^* = \phi(B)w_j(x_t)$, with B the backshift operator – i.e. $B^n w_t = w_{t-n}$ – and $\phi(B)$ the autoregressive operator, i.e. $\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \cdots - \phi_p(B^p)$, satisfying the usual stationary conditions;

b) y^* be the prewhitened version of y,

therefore, equation (10) can be expressed in terms of white noise series, i.e.

$$\boldsymbol{\Gamma}^{x^*,y^*,m} = \sum_{j=1}^{J} \Gamma\left[[w_j(x_t)]^*, y_t^*\right]_m + \Gamma\left[[v_J(x_t)]^*, y_t^*\right]_m.$$

The wavelet cross-covariance estimator can now be defined as follows:

$$\hat{\boldsymbol{\Gamma}}^{x^*,y^*,m} = \sum_{j=1}^{J} \begin{cases} \frac{1}{M} & \sum_{t=L_j-1}^{T-m-1} [w_j(x_{t+m)}]^*\ y_t^*;\ m = 0, 1, \ldots, M_j(T) - 1 \\ 0 & |m| \geq M_j(T), \end{cases}$$

with $M_j = T - L_j + 1$ and L being the wavelet filter length, as defined in (5) and (6). Finally the normalized version of (10), i.e. the MRA–based Empirical Cross Correlation Function ($mECCF$), reads as follows:

$$mECCF(x_t^*, y_t^*, m) = \frac{\hat{\boldsymbol{\Gamma}}^{x_t^*,y_t^*,m}}{\sigma^2_{x_t^*}\sigma^2_{y_t^*}}; \quad m = 1, 2, \ldots, M. \tag{13}$$

Function (13) will be extensively employed in the empirical study presented in the sequel (Section 7).

5 The dimension reduction procedures

In this section, the dimension reduction problem of the set Ω – i.e. the one containing the whole set of GT variables – is tackled. To this end, the already defined subsets – i.e: $\overline{\Omega}$, $\widetilde{\Omega}$, Ω'_q – will be separately considered.

5.1 Building the subset $\tilde{\Omega}$

Finding the autoregressive structure of the time series of interest is consistent with the nature of macroeconomic variables. In fact, they are generally characterized by a non negligible amount of persistence which can generate well defined patterns in their lagged probabilistic structures. In more details, the dependent variable is supposed to be a finite realization of a possibly infinite linear processes of the type: $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$; $t = 1, 2, \ldots, T$, being $\psi_0 = 1$, $\{\psi_j\}_{j=0}^{\infty}$ characterized by fast decaying coefficients and $\{\varepsilon_t\}$ generating a sequence of *iid* variables, so that $\mathbb{E}[\varepsilon_t] = 0$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. As it will clarified later, *W–BD* is a fast evaluation procedure requiring a simple statistical model to deliver useful information. Therefore, the available (finite) realization of Y_t, i.e. y_t, is modelled according to a $p - order$ autoregressive model ($AR(p)$) of the type: $y_t - \mu_x = \sum_{j=1}^{p} \varphi_t (y_{t-j} - \mu_x) + \varepsilon_t$; $t = 1, 2, \ldots, T$, being $\mu_y = E[y_t]$ and $\{\varepsilon_t\}$ as above detailed.

As it is well known, the order of the filter $\hat{p} = p(T)$ can be selected using a wide range of criteria both in time and frequency domains. The procedure chosen here is based on the Akaike Information Criterion (AIC), which is one of the most well known selector, successfully employed for many types of time series models, including the autoregressive ones. It is based on evaluating the "distance" between the estimated model and the true (unknown) model. Formally, it can be defined as follows: $AIC = -2 \max \log(L(\hat{\theta}|y)) + 2K$, being K the dimension of the model and $L(\hat{\theta}|y)$ the likelihood function. AIC selection strategy – commonly referred to as $MAICE$ (Minimum AIC Expectation) – is a procedure aimed at extracting, among the candidate models, a model order, say $\hat{p}$, satisfying $\hat{p} = \arg \min_{p \leq P_0} AIC(p)$. $MAICE$

procedure requires the definition of an upper bound P_0, as a maximum order the given time series can reach. The choice of this constant is *a priori* and arbitrary.

5.2 Building the subset Ω°.

The present subsection is devoted to the construction of the subset Ω°, that is the one containing only the relevant (i.e. leading) wavelet components. This task is performed through a eight–step procedure as illustrated below.

1 From $\aleph$ (the population of the GT time series), a (possibly very large) subset called Ω is extracted. Ω contains those variables which, in a broad sense, can be potentially useful for the explanation of the phenomenon at hand. This subset can be built in a qualitative fashion, e.g. based on brainstorming techniques or experts opinions. Also, any sorts of empirical evidences or outcomes from previous investigations, even if loosely related, can be fruitfully exploited. Quantitative techniques, based e.g. on big correlation matrices or clustering methods, can be employed as well;

2 set Ω is MRA–transformed element–wise, until an arbitrary, admissible, resolution level J – so that the set Ω'_q, with cardinality $(J+1) * |\Omega|$ is generated. Let $\{\Omega' \equiv x_{1,t}, x_{2,t}, \ldots, x_{N,t}\}$ – with $\{x_{i,t}; i = 1, \ldots, N; t = 1 \ldots, T\}$ – be the matrix containing the GT time series extracted in the previous step 1, subset Ω'_q is then defined as $\left\{\Omega'_q \equiv W_{t,i,j} \equiv W_j(x_{i,t});\ i = 1, \ldots, N;\ t = 1, \ldots T\right\}$; $\{\ j = 1, \ldots, J\}$, being W_j the level j wavelet component related to the series $x_{i,t}$;

3 $\left\{\Omega'_q\right\}$ is subjected to the prewhitened procedure, which has been outlined in Section 4.2. As a result, the set $\left\{^{pw}\Omega'_q\right\}$ is generated, i.e.:

$$\left\{^{pw}\Omega'_q\right\} = \mathcal{P}(\left\{\Omega'_q\right\}),$$

with the symbol $\mathcal{P}$ denoting the prewhitening function;

4 for each and every series in $\left\{{}^{pw}\Omega'_q\right\}$, the empirical cross correlation function ($ECCF$) (13) is computed with respect to the prewhitened version of the dependent variable, i.e. $\mathcal{P}(y_t)$. In symbols:

$$\hat{\Gamma}\Big(\mathcal{P}(w_{t,j}), \mathcal{P}(y_t), m\Big); \tag{14}$$

5 the series possessing leading properties under (14) – i.e. showing one or more empirical cross–correlations statistically significant $\forall m < 0$ – are detected, extracted and stored in the subset ${}^{pw}\Omega'' \subseteq {}^{pw}\Omega'_q$. Depending on the problem at hand, the last inequality can be relaxed to $m \leq 0$ (e.g. for nowcasting purposes, in the case of data – such as the GT series – which are available in real time and thus earlier than the updating disseminated at the end of the month of reference or in the following one);

6 ${}^{pw}\Omega''$ is back-transformed into its original (not prewhitened) form, by filtering back its elements. As a result, set Ω'' – is generated, i.e.:

$$\Omega'' \subset \Omega_q = \mathcal{P}^{-1}({}^{pw}\Omega'').$$

Ω'', therefore, encompasses all the leading wavelet components.

7 due to the potentially i) high cardinality of the set Ω'' and ii) the strong interdependency structures among its elements, an ulterior subsetting procedure is in order. The resulting subset – called Ω° – is the maximizer of an *ad hoc* target function $\mathcal{F}(\cdot)$, which will be detailed in Section 6.1. Function $\mathcal{F}(\cdot)$ is sequentially computed on J different competing subsets – $\Omega'''_j;\ j = 1, \ldots, N$, which are extracted from Ω''. The candidate subsets in Ω'''_j can be built randomly and/or on the basis of the knowledge of the phenomenon of interest so far accrued.

Formally:

$$\Omega^{\circ} = \underset{\Omega'''_j}{\operatorname{argmax}}\ \mathcal{F}(\Omega'''_j \subset \Omega''); \quad j = 1, \ldots, N. \tag{15}$$

However, Ω° is not the only subset considered to build the final subset $\Omega^{\star}$. In fact, it is crucial to emphasize how function $\mathcal{F}(\cdot)$ is designed to provide only a rough estimation of the gains obtainable by increasing the dimension of a linear model with additional, wavelet–based, information. Therefore, Ω° – as it will be outlined later – does not represent a globally optimal subset;

8 because of that, N $\mathcal{F}$–suboptimal subsets – i.e. $\{\ \Omega_{\circ,1}, \ldots \Omega_{\circ,N};\ N << J\}$ – so that $\{\Omega_{\circ,j}\} = \mathcal{F}(\Omega_j''')$ – are retained under the condition that their associated target functions $\mathcal{F}$ show values not "too far" from the optimal one (i.e. $\mathcal{F}(\Omega^{\circ})$). Practically, they need to be within a reasonable, predefined, threshold λ. All such N subsets are stored in a set denoted by the symbol $\Omega_{\bullet}$, which verifies the following condition:

$$\Omega_{\bullet}: \quad |\mathcal{F}(\Omega^{\circ}) - \mathcal{F}(\Omega_{\circ,j})| \leq \lambda;\ j = 1, \ldots, N. \tag{16}$$

$\Omega_{\bullet}$ is reasonably deemed to potentially contain exploitable information not accounted for by Ω°.

It is worth emphasizing that while step 1 is generally human driven, steps from 2 to 6 can be easily automatized. Finally, the crucial steps 7 and 8 can be carried out also with the help of suitable heuristic methods, e.g. of the type Genetic Algorithm (**?**, **?**).

5.3 Building the subsets $\overline{\Omega}$

Subset $\bar{\Omega}$ comprises those GT variables which are able to generate an amount of useful information greater than that achievable by using any levels (singularly or combined) of the related wavelet decompositions. *W–BD* method rests on the idea that wavelet decomposed time series possess better leading capabilities than the original ones. However, the case at hand – i.e. $\overline{\Omega} \neq \emptyset$ – cannot be *a priori* excluded. Consistently, the statistical model adopted for the evaluation of the benefits related to the GT

variables (presented in the next section 6), formally encompasses both these types of time series.

Recalling that with $\{x_{i,t}\}$ and $w_j(x_{i,t})$ the generic GT series and its j–level wavelet decomposition are respectively defined, it is

$$x_{i,t} \subset \overline{\Omega} \iff MSPE(\mathbb{E}[y_{t+m}] \mid x_{i,t} < MSPE(\mathbb{E}[y_{t+m}]|w_j(x_{i,t})), \quad \forall j, \qquad (17)$$

denoting with $\iff$ the logic symbol "if and only if". The searching procedure for these variables undergoes the same steps followed for the definition of subset Ω° reported in the previous Section 5.2.

Finally, in (17), the strict inequality has been chosen over its non strict counterpart to stress the fact that, even in the (mostly theoretical) case where a raw series and one (or a combination of more) of its wavelet component(s) is (are) perfectly interchangeable, the latter should be always preferred. This is for consistency with the idea that, in general, a decomposed time series is more informative (e.g. more easily interpretable or susceptible of comparative analysis) than its non decomposed counterpart.

6 The statistical model

To evaluate the prediction power of the series belonging to $\left[\Omega'' \cup \overline{\Omega}\right]$, a statistical model able to quickly and efficiently process those variables is in order. As already pointed out, the dependent variable's autoregressive structure (stored in the subset $\widetilde{\Omega}$) must also be evaluated and accounted for. The evaluation model chosen is of the type general dynamic regression, which can be formulated as in (18). Consistently with what referred to in point 7 (Section 5.2), this equation is not intended to provide immediately usable outcomes (e.g. predictions) but only results useful for the definitions of the final subset of variables Ω^*. In more details, (18) is a selection tool

serving the sole purpose of generating N residual variances for each of the N models generated by Ω_j'''; $j = 1, \ldots, N$.

For a model of the class general linear dynamic, the data are supposed to take the form of a vector time series process $z_t = (y_t, x_t')'$, being $x_t = (x_{1,t}, \ldots, x_{k,t})' \in \mathbb{R}^k$ and y_t the dependent variable (variable of interest). Recalling that the set of lagged dependent time series is denoted by $\widetilde{\Omega}$, we have that our generic (independent) time series variable x_t belongs (in a mutual exclusive way) to one of the following sets $\{\overline{\Omega}\}$, $\{\Omega''\}$ or $\{\widetilde{\Omega}\}$. Under this three-fold division, the general dynamic regression model takes the form

$$\hat{y}_t = \alpha_0 + \sum_{i=1}^{p} y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \beta_{j,p} x_{j,t-i}, \tag{18}$$

being $\{y_{t-m};\ m = 1, 2, \ldots, M;\ \in \widetilde{\Omega}\}$ and $\{x_t \in \overline{\Omega}_y \cup \Omega''\}$.

At this point, under their associated target functions $\hat{\mathcal{F}}(\cdot)$, different models (18) are

1. built according to different configurations of the variables in $\{\Omega_\bullet \cup \Omega^\circ\}$;
2. tested (e.g., for statistical significance, residuals autocorrelation structure and distribution);
3. calibrated (see Section 6.2)

so that the final, low dimension, competition set Ω^* is defined.

6.1 Estimation of the predictive power generated by Ω'' and $\overline{\Omega}$

The improvements achievable by employing GT data are assessed by evaluating the in sample performances delivered by both (18) and its univariate counterpart (19) – here employed as a benchmark – through Equation 20. In particular, Equation 19 is obtained by suppressing the last right side term in (18), so that it collapses to an autoregressive model, ie.e:

$$\hat{y}_t = \alpha_0 + \sum_{j=1}^{p} y_{t-j}. \tag{19}$$

Once determined through the MAICE procedure, recalled in the previous Section 5.1, the (possibly sparse) lag structure of Equation 19 is kept fixed across all the distributed lag models (18) attempted. In other words, only the series belonging to subsets Ω'' and $\overline{\Omega}$ are supposed to be recursively $\mathcal{F}$–tested. In symbols: $\mathcal{F}(\Omega'', \overline{\Omega})\Big|_{\bar{\Omega}=k}$.

The estimation of the predictive power of model (18) is performed by means of the objective function indicated by $\mathcal{F}_{M|U}(\cdot)$, which has been prposed in **?** as a slightly modified version of Akaike's Final Prediction Error (FPE) – derived by Bhansali **?**. Function $\mathcal{F}(\cdot)$ is a fast and effective way to measure the model predictability based on the estimation of the h–step ahead Mean Square Error of Prediction (MSPE) of the multivariate predictor versus its univariate counterpart. It is defined as follows:

$$^{(h)}\hat{\mathcal{F}}_{M|U} = 1 - \frac{FPE_{M(h)}}{FPE_{U(h)}}, \tag{20}$$

being

$$FPE_{U(h)} = \hat{\sigma}^2_U(h)(1 + \frac{p(U)}{T})$$

and

$$FPE_{M(h)} = \hat{\sigma}^2_M(h)(1 + \frac{p(M)}{T}),$$

where the subscripts M and U respectively refer to the multivariate and the univariate predictors whereas $\hat{\sigma}^2$ is the estimated OLS residual variance and p denotes the number of parameters (including the constant). The left superscript ^{h} in $^{(h)}\hat{\mathcal{F}}_{M|U}$, for simplicity of notation, will be omitted if not strictly necessary.

Recalling (15), the model generating the subset Ω° – denoted by the symbol M° – is the one satisfying the following maximization problem, i.e.:

$$\begin{aligned} \mathcal{M}^\circ(\Omega^\circ) = \underset{\Omega,\Omega'',\bar{\Omega}}{\text{maximize}} \quad & \hat{\mathcal{F}}_{M|U}(y) \\ \text{under} \qquad & \hat{\mathcal{F}}_{M|U}(y) > 0. \end{aligned} \tag{21}$$

The resulting subset Ω° – selected based on the $ECCFs$ – is likely to generate better forecasting results than the benchmark model. While this subset, on one hand, can provide precious insights on the variable of interest, on the other hand, as already pointed out, should not be used *tout court* in a statistical model, without further investigations. Therefore, once the target function has been maximized and the set Ω° stored, both the procedures detailed in 5.2 and 5.3 are rerun many times until N subsets $\Omega^\circ_j; j = 1, 2, \ldots, N$, satisfying the threshold expressed in (16), are generated. There are at least three reasons justifying this iterative approach. Firstly, given the possibly very high number of "good" $ECCFs$ (13), subset Ω'''s cardinality (but also $\bar{\Omega}$'s) should be constrained, in order to build a valid statistical model. This means, for Ω°, to undergo an ulterior subsetting procedure. However, the resulting time series could still generate low $\hat{\mathcal{F}}_{M|U}(y)$'s values. Such a situation is generally related to the high degrees of uncertainty conveyed by big data sets. In more details, it can happen that a variety of facts – mostly difficult to detect at a preliminary stage of an analysis involving big data (e.g. overparametrized or redundant models, lack of degree of freedom or simply poor predicting power of just one of the independent variables) – can disrupt the performance of an otherwise good set of predictors. Secondly, (20) is not designed to evaluate the goodness of a statistical model (e.g. statistical significance of its parameters, well–behaving residuals autocorrelation structures and their distribution). Thirdly, by examining the curve of the sorted values of $\hat{\mathcal{F}}(\cdot)$, it is possible to carry out a graphical sensitivity analysis. In fact, once a "sufficient" number of such values are available, one can draw conclusions on the level of interchangeability of the GT variables, by visually inspecting both shape and smoothness of this curve.

Clearly, the threshold λ (16) must be in agreement with the constraint expressed in Equation 21. This equation can be now finally reformulated in order to identify the $\mathcal{F}$–subsets of suboptimal (but worth to be evaluated) variables – called $\Omega^{\bullet}$, with $|\Omega^{\bullet}| = N$ – according to the predefined threshold λ (which accounts for the degree of "suboptimality" of the series), i.e.

$$\begin{aligned} \mathcal{M}^{\bullet}(\Omega^{\bullet}) = \underset{\Omega, \Omega''', \bar{\Omega}}{\text{maximize}} \quad & \hat{\mathcal{F}}_{M|U}(y) \\ \text{under} \quad & \hat{\mathcal{F}}_{M|U}(y) > 0 \\ \text{under} \quad & |\mathcal{F}(\Omega^{\circ}) - \mathcal{F}(\Omega_{\circ,j} \subset \Omega'')| \leq \lambda \end{aligned} \tag{22}$$

In practice, by means of Equation 22, subset $\Omega_{\circ}$ is expanded to include potentially valuable information which could not be adequately captured by (21) but that might be embedded in "slightly" suboptimal time series belonging to Ω''.

Finally, it is worth emphasizing that this procedure is very fast, being based on OLS estimators of the lagged time series coefficients. Even though the computation of (20) requires very little computer power, exhaustive searching routines – based on the iterative applications of Equation 20 to all the possible model configurations induced by the element in Ω'' – can be run in a reasonable time only if $|\Omega''|$ is "small".

6.2 Building the final models

Through the procedures explained in Sections 5 and 6, the information needed to build a fully operational statistical model is now available. In fact, both the multivariate (18) and the autoregressive (19) models generally could not be employed for any use apart from a rough subset selection. This is because neither (21) nor (22) are designed to deliver perfectly calibrated models but only to provide indications on the positive (if any) effects generated by GT external variables. In particular, through those equations the parameters statistical significance cannot be tested. It is only

when both the models are properly refined that the additional information conveyed by GT data can be fully exploited and the related benefits more precisely assessed.

The last step of *W–BD* is aimed at building the final, operational model, here denoted with the symbol $\mathcal{M}^* = \mathcal{M}^*(\Omega^*)$. Model $\mathcal{M}^*$ is of the type (18) and is built out of two, low–cardinality subsets, i.e. Ω° and $\Omega^\bullet$, respectively with cardinality 1 and $N << |\Omega''|$. To this end, a standard model selection procedure for autoregressive lag suppression and external variables (GT) subsetting have been applied to (18), which is thus built using different configurations of the time series in $\Omega_\circ$ and $\Omega^\bullet$. In particular, the well known Box–Jenkins (B-J) procedure has been adopted. The explanation of this method is beyond the scope of the present paper, therefore the interested reader can refer, for example, to **?** and **?**. B–J method is usually time consuming and requires a not negligible amount of experience to be properly carried out. However, this operation is now feasible for it is applied to a very small set of competing models (originated from a small set of predicting variables Ω° and $\Omega^\bullet$).

The benchmark autoregressive model undergoes the same two–stage identification procedure as the distributed lag model. In facts, it is firstly determined by $MAICE$ method and then fine–tuned through the B–J procedure. The reason for that is twofold: *a*) should (19) be carefully evaluated and properly built in the first place, the whole analysis would be biased against its multivariate counterparts, for which the same statistical attention would be infeasible or impractical; *b*) in many cases, one might be interested in processing a number of dependent variables related to the same phenomenon. In this case, therefore, the already extracted and processed information set can keep (at least part of) its validity. The phenomenon considered in the empirical experiment (presented in the following Section 7) is an example of such a situation. Here – being the variable of interest available for eight different age groups – one might be interested in an age–specific statistical analysis. In this case, a time consuming procedure – envisioning the construction of up to eight statistical models – should be carried out.

7 Empirical Expriment

In this section, *W–BD* method has been applied to the time series "Unemployment-to-population ratio", collected on a monthly basis by the Italian National Institute of Statistics (Istat) by interviewing a sample of approximately 77,000 households, representing 175,000 individuals Italian residents (even if they are temporarily abroad). The data – available at the following web-address: http://dati.istat.it/ – account for eight different age profiles, out of which we will concentrate our attention on the group "age 15 and over". For consistency with the *GT* data employed, the time span considered is January 2004–August 2018, for a total of 176 observations. The employed wavelet is of the type Daubechies with 8 vanishing moments. In more details, this experiment has been designed to simulate two common real-life activities, i.e., the