Data reduction by the Haar function: A case study of the Phillips Curve

Data reduction by the Haar function: A case study of the Phillips curve

Marco Gallegati [*] ***Meghnad Desai*** [¥] ***and James B. Ramsey*** [§]

Abstract

In this study we wish to argue that Phillips was a pioneer in using data reduction methods, which he may well be aware given his training as an electrical engineer and his published work in control theory. The unorthodox estimation procedure Phillips (1958) adopted in his original paper, which uses non-overlapping moving averages with unity weights, is examined using the Haar filter, the simplest type of wavelet basis function introduced by Haar in 1910. The application of the Haar wavelet transform to Phillips' original data shows that Phillips' six pairs of mean coordinates display a strikingly similarity with the Haar scaling coefficients which represent averages with a period greater than 16 years. This is consistent with Desai's (1975) intuition that the interpretation of the Phillips curve needed rethinking. Our data reduction procedure with the Haar wavelet basis reveals that the choice of sorting observations by ascending values of the unemployment rate is crucial for reaching the goal of estimating the eye-catching nonlinear hyperbolic shape of the wage-unemployment relationship, that would be otherwise linear. Interestingly, we show that the wavelet decomposition based on the Haar function can account not only for the facts characterizing the Phillips relationship up to the early 1960s, but also for two important facts which are among the most debated among policymakers: the downward shift of the Phillips curve and its flattening over time.

JEL: B22, C63, E24

Key words: Data reduction methods; Phillips' averaging procedure, Haar wavelet basis function.

*DISES, Faculty of Economics "G. Fuà", Polytechnic University of Marche, Piazzale Martelli 8, Ancona, Italy, tel:+390712207114, email: marco.gallegati@staff.univpm.it

¥ Affiliation not provided.

§ (Retired) Department of Economics, New York University, Mercer Street 267, New York, NY, james.ramsey@nyu.edu.

Introduction

In the pioneering years of the development of Econometrics in the early decades of the last century data were scarce. There were at best annual observations and not many of them either. The Klein-Goldberger Model of the US economy had fewer than twenty five observations, just 1929-1952. In those days, samples being small, you preserved every observation. T-statistics were used for testing significance since the samples were small. The caveat was added that as the samples became larger the confidence levels would improve.[1] Nowadays the data modelling context is very different. We have huge data sets. Techniques such as machine learning are required to cope with such large data sets. The challenge is not so much as to estimate a regression relationship with coefficient estimates and attached standard errors, but finding patterns in data which are very large. It is something of a surprise to discover that strategies of data reduction were pioneered earlier in the last century. Mathematicians such as Alfred Haar, Paul Levy, and Dennis Gabor were pioneers of methods of data compression. Signal transformation techniques were developed to obtain a reduced or "compressed" representation of the original data as an aid to finding patterns in data.[2]

In this study, we bring to attention data reduction strategies by an application to the classic study by Phillips (1958) of his justly famous "Phillips Curve". The fact that Phillips (1958) did not describe formally his sophisticated data reduction strategy has been source of controversy. For most of the literature Phillips' unconventional procedure is nothing more than a way to overcome the computational difficulties of estimating a nonlinear equation in the parameters (e.g. Gilbert, 1976, Wulwick, 1987, 1989, Wulwick and Mack, 1990, Lipsey, 2000, Forder, 2014). Desai (1975) was a rare instance of someone who did draw attention to Phillips' averaging procedure arguing that the procedure meant that the interpretation of the Phillips Curve needed rethinking. His contention has been largely ignored in the literature, the consensus being that Phillips used an averaging procedure to overcome the computational difficulties of estimating a nonlinear equation in the parameters at the time he wrote his paper.

We wish to argue that Phillips was a pioneer in using data reduction methods which he may well have been aware given his training as an electrical engineer and his published work in control theory (see Leeson, 2000). As noted by Gallegati et al. (2011), the unorthodox estimation procedure Phillips (1958) adopted in his original paper, by using non-overlapping moving averages with unity weights, represents a "crude" version of the simplest and oldest wavelet basis function developed by Haar in 1910[3] The application of the Haar discrete wavelet transform (DWT) to Phillips' original data for the 1861-1957 period allows to reveal the long-run nature of the wage-unemployment relationship and to determine whether and how Phillips's findings are affected by his unconventional data transformation procedure. In

[1] One of our authors- Desai- can testify to this from first hand experience.

[2] There is a long history since the days of David Hume about scepticism regarding discovering **causal patterns** in data. David Hume and in his tradition Freidrich Hayek took the view that you can only finds patterns in data but not infer causality (see Hayek, 1952).

[3] This is a sophisticated data reduction method that was available at the time Phillips wrote his paper, but almost certainly unknown to him.

particular, we are able to isolate the effects of the arbitrary choice of variable-width intervals, and the choice of sorting observations in ascending order of unemployment rate values for the computation of averages.

The application of the Haar DWT to Phillips' original data shows that Phillips' six mean coordinates display a striking resemblance with the Haar scaling coefficients corresponding to averages with period greater than 16 years. Using these scaling coefficients as a benchmark we show that the arbitrary selection of intervals is responsible for the regularity of the pattern formed by the averages, but not for the eye-catching hyperbolic shape of the wage-unemployment relationship, (see Wulwick, 1989). The nonlinear pattern is otherwise related to the ordering choice of values for the computation of averages. Indeed, when observations are time-ordered, a "simple" linear negative relation becomes clearly evident.[4]

Finally, in order to assess the statistical significance of the long-run relationship between wages and unemployment we apply several bivariate and multivariate tools of the continuous wavelet transform (CWT), the wavelet coherence and the partial wavelet coherence. The results show that the findings generally obtained at the aggregate level, such as those presented by Lipsey (1960), are determined by the strong and stable association between wages and unemployment at longer time scales. All in all, we argue that the failure to recognize the sophisticated features of the data reduction strategy used by Phillips (1958) has led to a neglect, not to say a misunderstanding, of Phillips's findings.

The structure of the paper is as follows. In section 2, after the description of the Haar filter, we analyze its similarity to Phillips' data reduction method. In Section 3 we apply the Haar wavelet transform to Phillips' original data set and examine the properties of Phillips's averaging procedure taking the Haar wavelet transform as a benchmark. Section 4 applies bivariate and multivariate CWT tools to detect the statistical significance of the wage-unemployment relationship at different time scales, and section 5 concludes.

2. Data reduction with the Haar wavelet basis function

Wavelet analysis allows to analyze a signal in multiple resolutions, where each component reflect a different frequency range associated to a specific time scale. In order to extract information from a signal at different scales and distinct times, wavelet analysis uses a collection of local basis functions, called wavelets, that are compactly supported, i.e. they have finite length, and are localized both in the time and the frequency domain. Since the wavelet transform can be rewritten as a convolution product, the transform can be interpreted as a linear filtering operation.

The principal idea behind using wavelet analysis as a data reduction method is that by applying the DWT to a time series we can keep much of the information that is in the original time series using some smaller set of data. Haar wavelets, proposed in 1910 by Alfred Haar, are the simplest form of wavelets. The Haar wavelet is a piecewise constant function on the real line that can take only three values: 1, -1 and 0. Therefore, Haar

[4] Similar results in terms of linearity of the relationship are provided in Shadman-Mehta (2000).

wavelets are the simplest orthonormal wavelet basis function with compact support (see Figure 1).[5] The Haar scaling and wavelet filters are given, respectively, by

$$H = (h_0, h_1) = 1/\surd 2,$$

and

$$G = (g_0, g_1) = (1/\surd 2, -1/\surd 2).$$

Three basic orthonormal properties characterize the Haar scaling and wavelet filters.

$$\sum_l h_l = 0 \quad \text{and} \quad \sum_l g_l = \surd 2$$

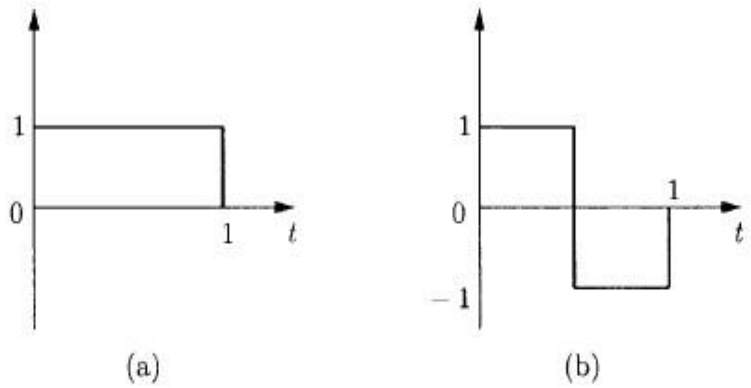
$$\sum_l h^2_l = 1 \quad \text{and} \quad \sum_l g^2_l = 1$$

$$\sum_l h_l h_{l+2n} = 0 \quad \text{and} \quad \sum_l g_l g_{l+2n} = 0 \qquad \text{for all integers } n \neq 0$$

The first property guarantees that *g* is associated with a difference operator, and thus identifies changes in the data, and that *h* may be viewed as a local averaging operator. The second property, unit energy, ensures that the coefficients from the wavelet transform preserves energy and, therefore, will have the same overall variance as the data.[6] The third property guarantees orthogonality to even shifts.

The Haar functions provide the two most elementary high-pass and low-pass filters. The wavelet filter *G*, with filter coefficients $g=[1/\surd 2,-1/\surd 2]$, by computing the difference between any two adjacent samples simply accomplishes differences. The scaling filter *H*, with filter coefficients $h=[1/\surd 2,1/\surd 2]$, represents a moving average filter because it essentially computes the average of successive pairs of non-overlapping values. Thus, for the Haar scaling filter the filtered signal is a weighted average of observations with the filter coefficients (h_0, h_1) used as weights. Therefore, the Haar wavelet transform is a series of averaging and differentiating operations.

Figure 1 – Haar scaling (left) and wavelet (right) filters

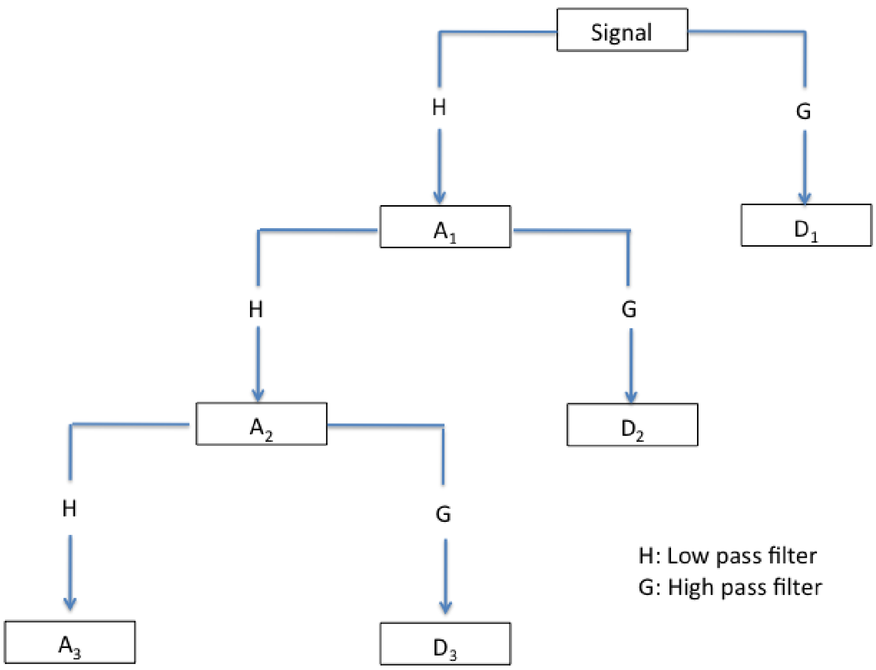


[5] The scaling and wavelet filters in Figure 1 assume unity values (as in Phillips' 'unorthodox' data transformation procedure).

[6] The normalization factor *√2* ensure that that the dilated and translated Haar function satisfies the second property in the wavelet definition.

Figure 2 shows how Mallat's (1989) pyramid algorithm breaks a signal down into different time scale components by recursively applying a sequence of filtering and downsampling steps. For the wavelet algorithm to decompose a signal into its different time scale components a dual pair of low-pass and high-pass filters is necessary at each scale level. The first is a non-overlapping moving average of the signal, the latter a non-overlapping moving difference. The first scale level uses a window width of two. Thus, the wavelet technique takes averages and differences on a pair of values of a signal, with the averages giving a coarse signal and the differences the fine details. This is done until the end of the signal. Then the algorithm shifts over by two values and calculates another average and difference on the next pair. At each successive scale level the window width is dilated (doubled) and, since wavelet algorithms are recursive, the smoothed data of the previous scale level, i.e. the averages, become the input for getting new (smoothed) approximations and detail components at coarser resolution levels. At each downsampling step, the wavelet algorithm decomposes a signal into two subsignals each with a length which is half the size of the input dataset. In the end, the application of the Discrete Wavelet Transform (DWT) to a dyadic length vector of observations ($N=2^J$ for some positive integer J) yields N wavelet coefficients, that is $N = N/2 + N/4, + ... + N/2^{J-1} + N/2^J + N/2^J$, where the number of coefficients at each scale level J is (inversely) related to the width of the wavelet function.

Figure 2 - Mallat filter scheme for a 3-level wavelet decomposiiton



The application of the DWT using the Haar (1910) wavelet filter for a 3-level decomposition, i.e. J=3, produces three vectors of wavelet detail coefficients, D_1, D_2 and D_3, and three vectors of scaling or approximation coefficients, A_1, A_2 and A_3. Table 1 presents the frequency domain interpretation of each detail and approximation level component using annual data. The detail levels D_1, D_2 and D_3 represent non-overlapping changes in the rate of change of money wage rates and in the unemployment rate at different frequency ranges, i.e. 2-4, 4-8 and 8-16 years, respectively. The approximation levels A_1 represents non-overlapping averages of the rate of change of money wage rates and unemployment rate greater than 4 years. Moreover, by adding D_2 and D_3 to the lower "smooth" component A_1 and A_2 we get, respectively, the additional levels of approximation A_2 and A_3 capturing fluctuations greater than 8 and 16 years. Finally, the last column in Table 1 denotes the number of values used in calculating the scaling coefficients at each scale level.

Table 1*: Frequency domain interpretation of multiresolution decomposition analysis with annual data for J=3*

Scale Level	Detail level, D_j	Years	Approximation level, A_J	Years	Window width
1	$\mathbf{D_1}$	2-4	A_1	from 4 to ∞	2
2	$\mathbf{D_2}$	4-8	A_2	from 8 to ∞	4
3	$\mathbf{D_3}$	8-16	$\mathbf{A_3}$	**from 16 to ∞**	**8**

According to the filtering literature, moving average is one of the varieties of discrete low-pass filter, with filtering effects being dependent on the window length. Hence, averaging corresponds to applying a low-pass filter to the signal. With the window size affecting the resolution level of the analyzed signal, using different averaging lengths is equivalent to viewing data at different resolution levels, as in multiresolution decomposition analysis: the longer the averaging window width, the lower is the frequency component extracted from the signal.

Phillips' (1958) 'unorthodox' data transformation procedure consists in reducing 53 observations to 6 average data points by first grouping observations into several variable-width arbitrarily selected intervals of the unemployment rate, and then computing for each interval the mean values of money wage inflation and the unemployment rate.[7] He yields six mean coordinates that represent pairs of non-overlapping moving averages with variable-width windows. In reducing the number of observations from 52 to 6, Phillips averaged into each interval a number of observations varying from 6 to 12 (i.e. 6, 10, 12, 5, 11, 9). Indeed, the number of values of money wage rates and the unemployment rate used on average by Phillips for computing the mean coordinates, the *six crosses*, roughly corresponds to the number of wavelet scaling coefficients at the approximation level A_3. Gallegati et al. (2011) suggest that Phillips' estimation methodology may be considered a "crude" version of the simplest wavelet basis function developed by Haar in 1910 as Phillips uses unity weights (as in the left panel of Figure 1).[8]

3. The application of the Haar DWT to Phillips's classic study

The Phillips Curve has been for nearly sixty years a Central tool for anti inflation Policy for central bankers and Finance Ministers. When A.W.H. Phillips published his results analysing

[7] His famous hyperbola is then fitted to these six mean coordinates through a procedure that combines least square estimation and graphical inspection (Gilbert, 1976).

[8] In this sense we can say that Phillips was, involuntarily, the first economist to use wavelets as a tool of analysis.

nearly a century of UK data (Phillips 1958), he did not indeed could not have anticipated its impact. Phillips himself was quite modest about his enterprise. He was an engineer by training who came to economics in post war Britain almost by accident as a demobilised military officer. He had got interested in macroeconomics and did pioneering work in application of control theory to economics and in time-series econometrics (Sleeman, 2011). The popularity of the Phillips Curve owes much to the attempt by Samuelson and Solow at the 1959 AEA conference to publish a US Phillips Curve (Samuelson and Solow, 1960). It was Milton Friedman in his Presidential address to the AEA who questioned the theoretical basis of the Phillips Curve (Friedman, 1968). The debate has not died down (see Leeson, 2000, for references).

There has been a persistent controversy as to whether the Phillips Curve tracing a relation between the rate of change of money wages and the rate of unemployment has been correctly specified (should it be real wages, should actual or expected inflation be an added variable), as well as its shape (should it be vertical rather than downward sloping). A different source of controversy has been about the unhortodox estimation procedure Phillips adopted in his original paper.

Phillips' averaging procedure did not invite many comments at the time of publication. It was presumed that he had reduced his data points from 52 to 6 because of the difficulty in those days of estimating an equation which was nonlinear in parameters as well as variables which is what Phillips had specified. Following Phillips's (1958) paper, his junior colleague Lipsey carried out a regression using all the data from 1861 to 1958. Lipsey choose a form which was linear in parameters but nonlinear in variables. This allowed OLS procedure to be used. His results confirmed Phillips's result (Lipsey, 1960).

Desai (1975) was the first person to highlight the importance of Phillips's procedure for data reduction. Desai argued that Phillips's averaging procedure meant that his estimated equation was not a short run structural relationship which could be given a causal interpretation. The averaging removed it from the time domain. The Phillips Curve as estimated by Phillips was a locus of long run equilibrium positions. As such, it could not be used for policy purposes. It was not possible according to Desai 'to slide down' the Phillips Curve. As explained by Phillips, along the Curve change in a Unemployment was zero. It was an equation for a singularity of a differential equation.

Desai's argument was challenged by Gilbert (1976) and more recently Lipsey (2000). The Economics profession in general were not persuaded by Desai's interpretation, the consensus being that Phillips averaging procedure was a a mere computational feature because estimation of his posited relationship was difficult in the Fifties of the last centuries when he wrote his paper (Gilbert, 1976, Wulwick, 1989, Wulwick, 1987, 1989, Wulwick and Mack, 1990, Lipsey, 2000, Forder, 2014). Others have reestimated Phillips Curve using recent econometric techniques (Shadman-Mehta, 2000).

Since both the estimated regression and the smooth hyperbolic curve are based on the six pairs of averaged values, the 'crude' statistical method used by Phillips can be considered crucial in getting his original results (Wulwick, 1987). Therefore, on the basis of its analogy with the Haar scaling filter, in this section we analyze the effects of Phillips' averaging procedure by applying the Haar DWT to Phillips' original dataset.[9]

Table 2 – Phillips' six mean coordinates and Haar A_3 scaling coefficients with observations sorted by ascending unemployment rate values (1861-1913)

Phillips' averages	*ur*	*1.516*	*2.351*	*3.483*	*4.49*	*5.954*	*8.372*
	dw	*5.058*	*1.547*	*0.848*	*0.346*	*-0.182*	*-0.350*
Haar A_3 coeffs	*ur*	*1.650*	*2.426*	*3.337*	*4.069*	*5.600*	*6.862*
	dw	*4.610*	*1.117*	*0.755*	*0.852*	*0.431*	*-0.991*

Note: The values in the row "Phillips' averages" are calculated by averaging observations sorted by ascending values of the unemployment rate for Phillips' (1958) intervals (0-2, 2-3, 3-4, 4-5, 5-7, 7-11). The values in the row "Haar A_3 coeffs" are computed using the Haar DWT wavelet transform (the 7th Haar A_3 scaling coefficient, not included here, is reported in Figure 3).

In Table 2 the row "Philips' averages" presents Phiilips' six pairs of mean coordinates of the unemployment rate and money wage rates for the period 1861-1913. Since the averaging length (or window width) determines the frequency resolution of the decomposition, the approximation component at the scale level J=3, by yielding 7 scaling coefficients each stemming from a window-width with length 8, can provide a useful benchmark for evaluating Phillips' averaging procedure. The row "Haar coeffs" shows the values of the Haar scaling coefficients at level A_3 with observations sorted by ascending values of the unemployment rate as in Phillips (1958).[10] For ease of interpretation the coefficient values shown in Table 2 are visually displayed in the left panel of Figure 3: filled blue circles represent A_3 level approximation coefficients, while Philips' averages are marked with a cross, as in his original paper. The positions of the filled blue circles and the crosses are quite well aligned in the left panel of Figure 3, except for values of the unemployment rate higher than 5-6%. Similar results[11] emerge in the right panel of Figure 3 where the same analysis is replicated for the whole period, 1861-1957.[12] Since the longer the averaging length the lower is the frequency component extracted using filtering methods, the result of Phillips' averaging procedure is to identify a "locus of long-run equilibrium points" (Desai, 1975, p.2) whose frequency resolution level corresponds to fluctuations greater than 16 years.[13]

---

[9] The same goal, that is repeating Phillips (1958) study, is performed by Wulwick and Mack (1990) using kernel regresion analysis.

[10] The wavelet domain thresholding algorithm SURE with soft thresholding has been used for signal denoising. With this method the optimal threshold selection values at different scale levels are based on the Stein's (1981) unbiased MSE risk estimate (SURE) threshold selection algorithm with soft thresholding function; noise structure: unscaled white noise. Therefore, the inverse wavelet transform is carried out via those thresholds and the resulted de-noised time series are decomposed via the Haar trasnform.

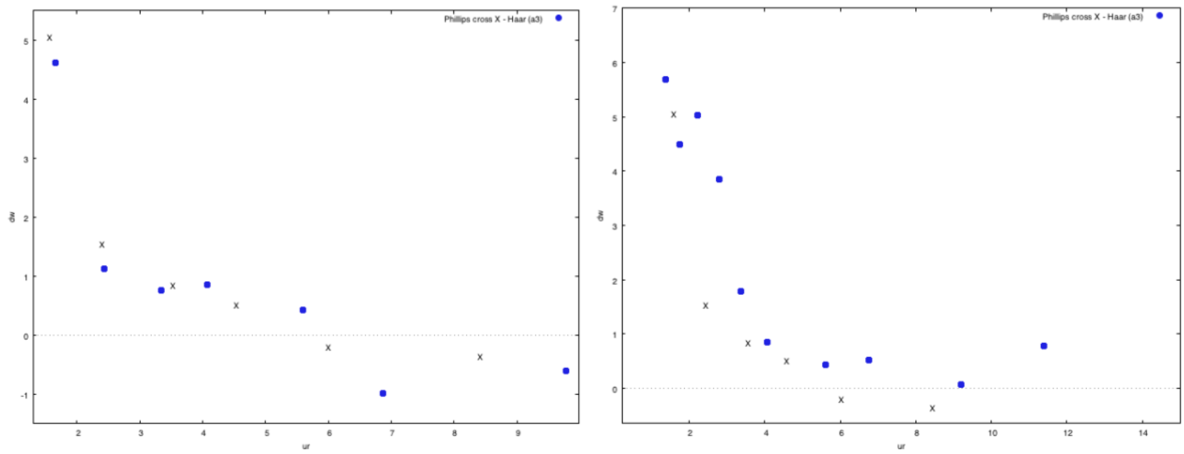
[11] The main difference being that A_3 coefficients are slightly shifted upward.

[12] In the right panel outliers have been excluded by limiting the range of x-axis and y-axis. Outliers have been defined as those values greater than 10% for the unemployment rate and greater than 15% for the money wage rate. Such values are all included in the 1918-1923 period.

[13] This is also consistent with Phillips' statement that "each cross (or mean coordinate) gives an approximation to the rate of change of wages which would be associated with the indicated level of unemployment if unemployment were held constant at that level (Phillips, 1958, p. 290).

In Figure 3 the comparison of Phillips' *six crosses* with the filled blue circles representing the Haar A_3 level approximation coefficients, that are based on fixed regular windows, allows to isolate the effect of Phillips' arbitrarily choice of intervals. Interestingly, the blue filled circles detect an irregular graph of averages, with several ranges of unemployment characterized by a positive relationship. Similar findings are provided by Wulwick (1989) that, after experimenting alternative intervals similar to Phillips' intervals, stated that "only Phillips' intervals resulted in the smooth *hyperbolic* graph of averages" (Wulwick, 1989, figs. 4 and 5, p.181-2). Therefore, while the shape of the wage-unemployment relationship does not seem to be affected by the use of fixed or variable window widhts, the regularity of the pattern formed by the averages is strictly related to the arbitrary selection of intervals made by Phillips.[14]

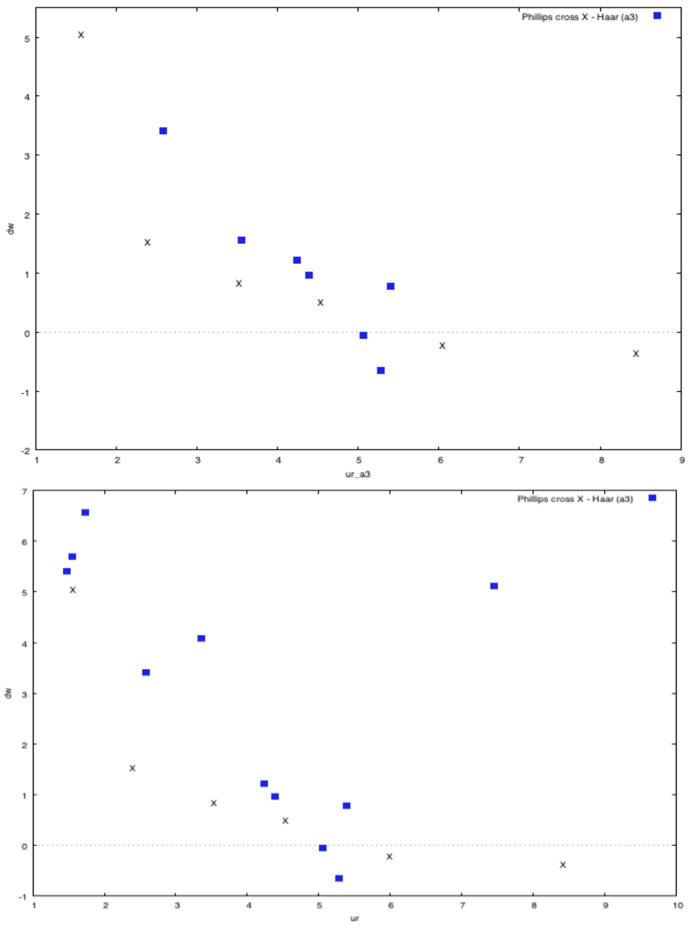
Figure 3 *– Philips' averages (x) and Haar A_3 scaling coefficients (•) with observations sorted by ascending unemployment rate values: 1861-1913 (upper panel) and 1861-1957 (lower panel).*



Note: In the right panel values higher than 15% for money wage rates and the unemployment rate are excluded because they can be considered as outliers (they refer to years between 1918 and 1923).

Figure 4 *– Philips' averages (x) and Haar A_3 scaling coefficients (•) with observations sorted by time: 1861-1913 (upper) and 1861-1957 (lower panel).*

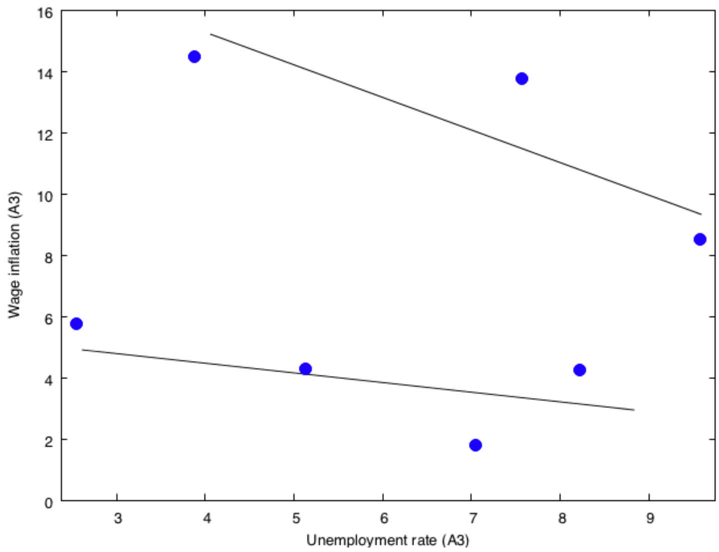
[14] Although it may be appear a secondary feature, it is not if we think that the failure in his 1955's book to draw or fit an eye-catching downward sloping convex curve on a scatter diagrams similar to Phillips' probably prevented from the attribution of the wage-unemployment relationship the label Brown- or Brown-Phillips curve (Corry, 2001, Button, 2018).



Note: In the right panel values higher than 15% for money wage rates and the unemployment rate are excluded because they can be considered as outliers (they refer to years between 1918 and 1923).

The Haar wavelet transform may also be useful in order to detect the effect of sorting observations according to increasing values of the unemployment rate. To this aim we apply the Haar wavelet transform to time ordered observations. Figure 4 shows the Haar A_3 level approximation coefficients (filled blue squares) and Philips' averages (crosses) for the periods 1861-1913 (upper panel) and 1861-1957 (lower panel) when observations are ordered by time. Interestingly, the pattern diplayed by the filled blue squares in both panels of Figure 4 contrasts strikingly with that delineated by Phillips' crosses. In particular, differently from the evidence presented in Figure 3 the filled blue squares identify a simple linear pattern for the wage-unemployment relationship. This finding suggests that the nonlinear pattern (shape) of the wage-unemployment relationship is dependent on grouping obervations into ascending values of the unemploment rate. What emerges from our analysis is that the ordering choice is highly influential for Phillips' results. Indeed, such ordering is responsible for obscuring the 'true' linear long-run relationship that is otherwise evident at the coarsest scale when observations are sorted through a time-ordered sequence.

Figure 5 – *Haar A_3 scaling coefficients (•) for the 1960-2016 period.*



Finally, we show how the proposed methodology can account for the Phillips' relationship many decades after the original findings by Phillips. Figure 5 shows the A_3 Haar scaling coefficients from wage changes and unemployment from the 1960s to present. Two separate clusters with respect to the level of nominal wage rate changes are clearly evident. The first, which includes three blue filled circles in the upper part of Figure 5, refers to values in the first part of the sample, i.e. until late-1980s, characterized by high wage inflation rates associated with both low and high unemployment rates. The latter, which includes the four blue filled circles in the lower area of Figure 5, refers to low values of wage inflation rates and are characteristics of the last three decades of the sample, i.e. from 1990s to present. The different slopes of the wage-unemployment relationship in the two periods, evidenced by two separate solid lines, show clearly that the relationship between the unemployment rate and nominal wage growth has moved lower and flatter over time.[15]

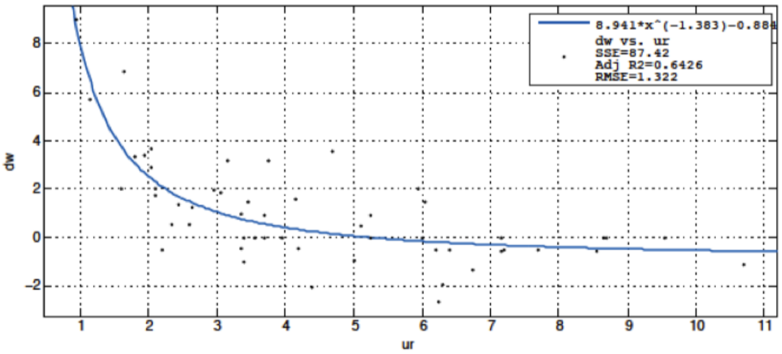
All in all, the findings presented in Figures 4 and 5 show that the wavelet decomposition based on the Haar function can account not only for the facts characterizing the Phillips relationship up to the early 1960s, but also for two important facts which are among the most debated among policymakers, namely that the curve has shifted downwards and has become flatter.

4. Aggregate or long-run relationship? Evidence from CWT tools

As firstly shown by Lipsey (1960, p.4 Fig.3), the pattern detected by Phillips for the wage-unemployment relationship during the 1862-1913 period may be easily reproduced by OLS estimation using aggregate data, as evidenced in Figure 6 where a power function is used as functional form. However, the aggregate pattern may be the result of different frequency-dependent relationship between unemployment and wages at different time scales.

Figure 6 – *Phillips curve estimation using a power function (1861-1913)*

[15] Proposed explanations for the downward shift of the Phillips curve include a lower natural rate of unemployment, lower or more anchored inflation expectations and changes in expectations of real pay growth (Cunliffe, 2017). By contrast, the integration of global value chains, increased competition and contestability of product and labour markets are among the reasons why the Phillips curve may have flattened (Carney, 2017).



This question may be explored using the CWT. The Haar DWT produces only a limited discrete number of translated and dilated versions of the wavelet basis, with scales and locations normally based on a dyadic arrangement (i.e., integer powers of two). When the transform operates on smooth continuous functions by decomposing the signals at all times and scales we get the CWT,[16] which produces the wavelet coefficients at every possible scale and time. Therefore, the CWT is a highly redundant transform that produces information in a two-dimensional format where each wavelet coefficient is represented by a pair of data, designing time, or location, and scale (Gencay et al. 2003).

The continuous wavelet transform (CWT) of a signal x(t) with respect to the wavelet function ψ is a function $W_x(s, u)$

$$W_{xy}(s,u) = \int_{-\infty}^{\infty} \psi_{(s,u)}(t) x(t) dt$$

where the wavelet basis, called "mother wavelet", defined as

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$$

is a function of two parameters s and u. The first is a scaling or dilation factor that controls the length of the wavelet, the latter is a location parameter that indicates where the wavelet is centered along the signal. The set of CWT wavelet coefficients, each representing the amplitude of the wavelet function at a particular position and for a particular wavelet scale, is obtained by projecting x(t) onto the family of "wavelet daughters" ψ(s,u) obtained by scaling and translating the "mother wavelet" ψ by s and u, respectively.

Let W_x and W_y be the continuous wavelet transform of the signals x(.) and y(.), their cross-wavelet power is given by $|W_{xy}|=|W_x W_y|$ and depicts the local covariance of two time series at each scale and frequency (see Hudgins et al. 1993). Being the product of two non-normalized wavelet spectra, the cross-wavelet can identify the significant cross-wavelet spectrum between two time series, although there is no significant correlation between

[16] A good introduction to the CWT and its associated univariate (wavelet spectrum), bivariate (wavelet coherence and phase) and multivariate (multiple and partial wavelet coherence) tools is provided in Aguiar-Conraria and Soares (2014). The interested reader may refer to that survey for more technical details.

them. The (squared) wavelet coherence is defined as the modulus of the wavelet cross spectrum normalized by the wavelet spectra of each signal,

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

where S is a smoothing operator (see Torrence and Webster, 1999). The squared wavelet coherence coefficient R^2_{xy} can be considered a direct measure of the local correlation between two time series at each scale. Hence, it can be used to assess how the degree of association between two series changes in the time-frequency plane, thus allowing to detect those time scales at which the relationship is significant from those at which it is not. The phase difference provides the relative phase between signals' variations, that is their lead/lag relationship.

The CWT provides several extensions to the bivariate case, represented by the partial wavelet coherence (PWC) and the multiple wavelet coherence (MWC), that allow to quantifying time-frequency multivariate relationships between variables. PWC is a technique, similar to partial correlation, that allows identifying the time-frequency relationship between two time series after eliminating the influence of a third common variable. MWC, like multiple correlation, allows to detect the time- and scale-specific effects of multiple independent variables on a dependent one.

Figure 7 shows the wavelet coherence between wage inflation and the unemployment rate (upper panel) and their partial wavelet coherence with inflation partialled out (lower panel) over the 1860-1960 period.[17] Time is recorded on the horizontal axis and periods, with their corresponding scales of the wavelet transform, on the vertical axis.[18] The outcome of the estimated wavelet coherence is in the form of a heatmap, which allows a straightforward interpretation of the degree of association between variables. The magnitude of each squared coherence coefficient is indicated by the color scale from dark blue (low coherence) to dark red (high coherence): the warmer the color, the higher the coherence power between the two series at that location in the time-frequency plane. Therefore, wavelet coherence maps can easily identify low- and high-coherence power regions in the time-frequency plane, that is areas where the degree of association between two time series is weak or strong. Phase difference are displayed as arrows on the wavelet coherence plot: right/left arrows indicate that both series are in phase/anti-phase, i.e. positively/negatively correlated.[19] Arrows pointing to the left-up indicate that wage is leading, while arrows pointing to the left-down show that unemployment is leading.[20]

[17] We use the Bank of England's collection of historical macroeconomic and financial statistics database (Thomas and Dimsdale, 2016).

[18] Hence, reading across the graph at a given value for the wavelet scaling one sees how the power of the projection varies over time at a given scale. Reading down the graph at a given point in time one sees how the power varies with the wavelet scale (Ramsey et al. 1995).

[19] A zero **phase difference** means that the two time series move together.

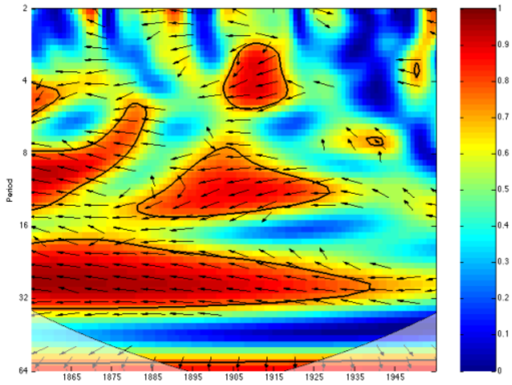
[20] The continuous wavelet transform, with other types of transforms, suffers from a distortion problem due to the finite time series length which affect wavelet transform coefficients at the beginning and end of the data series (Percival and Walden, 2000). In order to reduce the cone of influence before computing the CWT we

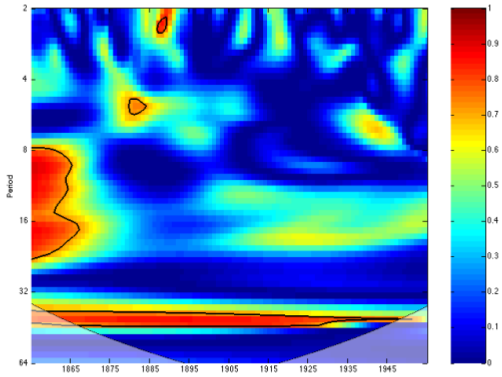
In the upper panel of Figure 7 four significant areas of high coherence are clearly detected, the most interesting being that visible at frequencies between 16 and 32 years for most of the estimation period, which denotes a stable and strong significant long-run relationship between the change in money wage rates and unemployment. The direction of arrows always indicates a negative relationship with money wage rates leading unemployment, except at the medium-run and business cycle frequencies in the late XIXth and early XXth century period when unemployment is leading wages.

Since the Phillips relationship necessarily involves other explanatory variables (e.g. Staiger et al. 2002) the findings presented in the upper panel of Figure 7 are subject to potential bias. Multivariate tools such as the partial wavelet coherence, which is an extension of the bivariate tool, allow to study the degree of correlation between two variables by taking into account the effect of control variables. The lower panel of Figure 7 shows the partial wavelet coherence between money wage rate changes and unemployment rate by partialling out the effect of consumer price inflation. After removing the effect of inflation two statistically significant high coherence regions are in evidence yet. The first is concentrated at the very beginning of the sample at scales beyond business cycle frequencies, i.e between 8 and 24 years. The latter provides evidence of a stable relationship throughout the sample at frequencies longer than 32 years. Therefore, even after considering the effect of control variables such as inflation, ther is evidence of a long-term relationship between unemployment and wages.

Figure 7 – *Wavelet coherence (upper panel) and partial wavelet coherence with respect to inflation (lower panel) between unemployment rate and wage inflation, 1860-1960*

apply the half-point symmetric extension mode at the left side with N=40 and use real data at the right side for a similar length.





Note: The wavelet coherence power is indicated by color coding: it ranges from dark blue (low coherence) to dark red (high coherence). A black contour line testing the wavelet power 5% significance level against a white noise null is displayed, as is the cone of influence, represented by a shaded area corresponding to the region affected by edge effects at the beginning and the end of the time series.

5. Conclusions

In this paper we draw attention to the Haar wavelet filter that is a very useful data reduction tool for modelling secular relationships by an application to the classic study by Phillips (1958). Phillips used a sophisticated data reduction method which he did not describe formally. This has led to a neglect, not to say a misunderstanding, of Phillips' averaging procedure. Against the consensus view that Phillips's estimation was a mere computational feature, Desai (1975) draw attention to Phillips' averaging procedure by arguing that the procedure meant that the interpretation of the Phillips Curve needed rethinking. However, his contention has been largely ignored in the literature. We wish to argue that Phillips was a pioneer in using data reduction methods which he may well have been aware given his training as an electrical engineer and his published work in control theory.

The application of the Haar DWT to Phillips' original data set allows to determine whether and how his results are affected by his 'unorthodox' data transformation procedure. We find that Phillips' six mean coordinates display a striking resemblance with the Haar coefficients of wages and unemployment representing averages with period greater than 16 years. This finding is consistent with Desai's (1975) view that Phillips unconventional estimates, based on fitting to averages calculated over periods long enough to eliminate business cycle effects (Hendry and Mizon, 2000), allows to abstract from the cyclical properties of the data (Hoover, 2014). Moreover, we find the choice of sorting observations by ascending values of the unemployment rate to be crucial for reaching the goal of estimating the eye-catching nonlinear hyperbolic shape of the wage-unemployment relationship, that would be otherwise linear. Finally, we show that the wavelet decomposition based on the Haar function can also account for two important phenomena characterizing the Phillips curve framework in the post-1960 period: the inward shift of the curve and the flattening of its slope over time.

References

AGUIAR-CONRARIA, L. and SOARES, M.J. (2014). The Continuous Wavelet Trans- form: Moving Beyond Uni- And Bivariate Analysis. Journal of Economic Surveys, 28 (2), 344–375.

BUTTON, K. (2018). A.J. Brown, 'Phillips' Curve', and Economic Networks in the 1950s. *Journal of the History of Economic Thought*, 40, 243-64.

BROWN, A.J. (1955). *The Great Inflation, 1939-1951.* Oxford: Oxford University Press.

CARNEY, M. (2017). [De]Globalisation and Inflation, IMF Michel Camdessus Central Banking Lecture .

CORRY, B. (2001). Some myths about Phillips' Curve. in P. Arestis, M. Desai and S. Dow (eds). *Money, Macroeconomics and Keynes: Essays in Honour of Victoria Chick, Volume 1*, London: Routledge.

CUNLIFFE, J. (2017). The Phillips curve: lower, flatter or in hiding? Bank of England speeche at Oxford Economic Society

DESAI, M. (1975). The Phillips Curve: A Revisionist Interpretation. *Economica*, 42, 1-19.

FORDER, J. (2014), Macroeconomics and the Phillips curve myth. Oxford: Oxford University Press.

GALLEGATI, MARCO, GALLEGATI, MAURO, RAMSEY, J.B. and SEMMLER, W. (2011). The US wage Phillips curve across frequencies and over time. *Oxford Bullettin of Economic and Statistics*, 73, 489-508.

GENCAY, R., SELCUK F. and WHITCHER B. (2003), Systemic risk and timescales, Quantitative Finance 3(2), 108-116

GILBERT, C.L. (1976). The Original Phillips Curve Estimates. *Economica*, 43, 51-57.

HAAR, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69, 331-371.

HAYEK, F.A. (1952), The Sensory Order: An Inquiry Into the Foundations of Theoretical Psychology, University of Chicago Press.

HENDRY, D.F. and G.E. MIZON (2000), The influence of A.W. Phillips on econometrics, In R. Leeson (Ed.) A. W. H. Phillips: Collected Works in Contemporary Perspective, Cambridge: Cambridge University Press.

HOOVER, K.D., (2014), The Genesis of Samuelson and Solow's Price-Inflation Phillips Curve, CHOPE Working Paper No. 2014-10.

HUDGINS, L., FRIEHE, C.A. and MAYER, M.E. (1993), Wavelet transforms and atmospheric turnulence, Phisycal Review Letters, 71 (20), 3279-3282.

KLEIN L.R. (2000), The Phillips Curve in Macroeconomics and Econometrics, In R. Leeson (Ed.) A. W. H. Phillips: Collected Works in Contemporary Perspective, Cambridge: Cambridge University Press.

LEESON, R. (2000), A. W. H. Phillips: Collected Works in Contemporary Perspective. Cambridge: Cambridge University Press

LIPSEY R.G. (1960). The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862-1957: A Further Analysis. *Economica*, New Series, 27, 1-31.

LIPSEY, R.G. (2000), *The famous Phillips Curve article*, In R. Leeson (Ed.) A. W. H. Phillips: Collected Works in Contemporary Perspective, Cambridge: Cambridge University Press.

MALLAT, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction Pattern Analysis*, 11, 674-693.

PERCIVAL, D.B., and WALDEN A.T. (2000), Wavelet Methods for Time Series Analysis, Cambridge: Cambridge University Press.

PHILLIPS, A.W.H. (1958). The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom. 1861-1957. *Economica* 25, 283-99.

PHILLIPS, A.W.H. and QUENOUILLE, M.H. (1960). Estimation, Regulation and Prediction in Interdependent Dynamic Systems. *Bulletin de l'Institute de Statistique*, 37, 335–43.

RAMSEY, J.B. and ZHANG, Z. (1995), The Analysis Of Foreign Exchange Data Using Waveform Dictionaries, Journal of Empirical Finance, 4, 341–372.

SAMUELSON, P.A. and SOLOW, R.M. (1960), Analytical Aspects of Anti-Inflation Policy, American Economic Review 50(2), 177-194.

SHADMAN-MEHTA, F. (2000), Does Modern Econometrics Replicate the Phillips curve?, In R. Leeson (Ed.) A. W. H. Phillips: Collected Works in Contemporary Perspective, Cambridge: Cambridge University Press.

SLEEMAN, A.G., (2011), The Phillips Curve: A Rushed Job? Journal of Economic PerspectiveS, 25 (1), 223–238.

STAIGER, D., STOCK, J.H. and WATSON, M.W. (2002), Prices, wages and the U.S. NAIRU in the 1990s, in A. Krueger and R. Solow (Ed.), The roaring nineties: Can full employment be sustained?, Russell Sage Foundation, New York.

STEIN, C.M. (1981), Estimation of the Mean of a Multivariate Normal Distribution, The Annals of Statistics, 9 (6), 1135–1151.

THOMAS, R and DIMSDALE, N. (2016) Three Centuries of Data - Version 2.3, Bank of England, http://www.bankofengland.co.uk/research/Pages/onebank/threecenturies.aspx

TORRENCE, C., and WEBSTER, P.J. (1998), The annual cycle of persistence in the El Nino-Southern Oscillation. Quarterly Journal of the Royal Meteorological Society, 124 (550), 1985– 2004.

WULWICK, N.J. (1987). The Phillips curve: Which? Whose? To do what? How? *Southern Economic Journal*, 54, 834-57.

WULWICK, N.J. (1989). Phillips' Approximate Regression. *Oxford Economic Papers*, New Series, 41, 1, *History and Methodology of Econometrics*, 170-188.

WULWICK, N.J. and MACK Y.P. (1990). A Kernel Regression of Phillips' Data. *Working Paper No. 40*, New York, Jerone Levy Economics Institute, Bard College.