A Persistence Based Decomposition of Macroeconomic and Financial Time Series

Fulvio Ortu,[1] Andrea Tamoni,[2] and Claudio Tebaldi[3]

(January, 2010) October 14, 2011
Preliminary empirical results

Abstract

If the aggregate response of the economy to an exogenous shock is a superposition of effects which develop over different time scales, then the statistical estimation of low frequency components is difficult. In fact highly persistent shocks have generally low instantaneous volatility and are hidden by those shocks with high instantaneous volatility and fast decay. We refer to this situation as heterogeneity of persistence levels phenomenon. This paper introduces a new spectral approach which is applicable to the analysis of time series in the presence of persistence heterogeneity. A new linear decomposition of a time series is introduced which generalizes the Wold decomposition for stationary time series and the Beveridge-Nelson permanent transitory decomposition for non stationary integrated ones. In order to prove the relevance of this new methodology for financial valuation, we apply it to clarify some open issues which arise in the empirical analysis of gdp and inflation forecasting.

JEL Classification Codes: E32, E43, E44, G12.

Keywords: .

[1]Bocconi University and IGIER; fulvio.ortu@unibocconi.it.
[2]Bocconi University; andrea.tamoni@unibocconi.it.
[3]Bocconi University and IGIER; claudio.tebaldi@unibocconi.it.

1 Introduction

Shocks that impinge an economy can be classified along two competing dimensions: their size as measured by their instantaneous volatility and their persistence as measured by their half life. Short run risk is determined by transitory risk components with large volatility which are expected to have a fast decay over a characteristic time horizon determined by their half-life. Long-run risk instead is dominated by highly persistent components with small size. There are many empirical tests in economics and finance that are complicated by the presence of effects which develop over different time scales. In macroeconomics the chief measurement issue concerns how to separate data into trends and cycles. The empirical literature contains a wide variety of competing methods. The seminal contribution of Beveridge and Nelson (1981) proposed a non structural approach to this decomposition. More recently research on financial valuation has proved that also the risk-return trade-off profile which describes efficient investment opportunities in the market is strongly dependent on the holding period (Campbell and Viceira (2005) and Bandi and Perron (2008)). A formal and systematic analysis of the long term risk valuation can be found in Hansen and Sheinkman (2009a) and Hansen and Sheinkman (2009b). Their approach builds on a multiplicative version of the Beveridge and Nelson (1981) permanent-transitory decomposition.

This paper proposes a spectral (linear) decomposition of the economic shocks which provides an effective method to represent a time series as a linear combination of uncorrelated shocks (the innovations) which are classified by the time of their arrival, as in the standard Wold decomposition, and by an additional index which measures their level of persistence.

A specific component accounts for those shocks which do not show a decay within any time interval smaller than the observation sample. This component is the permanent component as defined by Beveridge and Nelson (1981) but the new decomposition produces an alternative more efficient scheme for its identification. Transitory shocks are further split in orthogonal components. This splitting provides an effective method to represent a time series as a linear combination of innovations classified by their level of persistence. Shocks with different levels of persistence and the same arrival time define the term structure of innovation shocks. The empirical relevance of the term structure of innovations is illustrated by analyzing its implications for the cycle measures of GDP and Inflation. A separate companion paper (see Ortu, Tamoni and Tebaldi (2011)) is devoted to the analysis of the Long Run Risks model in a Bansal and Yaron (2004) economy set-up.

While traditional linear time series analysis is based on Fourier spectral analysis, our extended approach is based on spectral multiresolution analysis (see Daubechies (1990), Daubechies (1992), Mallat (1989a) and Mallat (1989b) for an introduction to the topic). Far from being the first attempt to use a multiresolution approach to economic analysis,we mention among many the contributions of Ramsey and Lampart (1998) which analyze the relation between money and income decomposed across time and scales using a multiresolution and Gencay and Fan (2008), Gencay and Gradojevic (2009), Gencay, Selcuk and Whitcher (2001) for a specific analysis of unit roots and filtering using multiresolution analysis. A thorough discussion of the econometric issues involved in the detection of low frequency structural relations can be found in Muller and Watson (2008) and Muller and Watson (2009). The role of the Beverigde-Nelson decomposition for characterizing the nature of macroeconomic fluctuations, and its relation to other unobserved components models, are discussed in Watson (1986), Morley, Nelson and Zivot (2003), Proietti (2006), Oh, Zivot and Creal (2006), and Morley (2011), among others.

The paper is organized as follows. Section 2 discusses the motivations which lead to the necessity of a persistence based classification of shocks, Section 3 states the main result of the paper, the Persistence Based Decomposition (PBD). Section 4 analyzes the spectral origins and the relation of the PBD with previously known decompositions and basic asymptotic limit theory. Section 5 is devoted to applications of the decomposition to gdp and inflation forecasting . All the proofs are collected in the Appendix.

2 A Persistence Based Classification of Shocks

There are many measures of the degree of persistence of an economic shock. Some of them are based on the time series representation, other rely on a frequency representation. In this section a measure of persistence is constructed by making use of a multiresolution filter[4], a class of filters which achieves an optimal "tradeoff" between a time and a frequency representation. The filter preserves the advantages of a time series representation because the decomposition is non-anticipative and can be computed using only past observations and the filtered components are represented by a time series.

Simultaneously the Fourier spectrum of each component is optimally concentrated on a narrow band of frequencies, thus legitimating a classification based on the level of persistence, which uniquely determines the mean reversion timescale i.e. the inverse of the characteristic frequency.

[4]For the basic definitions and notational conventions employed we refer to Hayashi (2000).

A multiresolution filter can be simply obtained through the recursive application of a well known object in financial econometrics: the moving average filter. This class of filters is explicitly designed to remove some systematic "distortions" which complicate the statistical detection of the low frequency components of random signals. The starting point of this analysis is the following:

Definition 1 *The dyadic mean operator acting on the time series of observations up to time t, $x_t = \{x_{t-k}\}_{k \in 0,..,+\infty}$, is defined by:*

$$M : \mathbf{x}_t \to \boldsymbol{\pi}_t^{(1)} = \left\{ \pi_{t-k}^{(1)} \right\}_{k \in 0,..,+\infty}$$
$$\pi_{t-k}^{(1)} = \frac{x_{t-k} + x_{t-k-1}}{2}$$

Iterated application of the operator M to the original time series $\mathbf{x}_t$ defines the sequence of time series $\boldsymbol{\pi}_t^{(J)}$ whose elements can be computed recursively:

$$\pi_{t-2^J k}^{(J)} = \left(M \boldsymbol{\pi}_t^{(J-1)} \right)_{t-2^J k} = \frac{\pi_{t-2^J k}^{(J-1)} + \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2} \tag{1}$$

therefore the element $\pi_{t-2^J k}^{(J)}$ corresponds to the sample mean over a window of past observations with size 2^J.

From the point of view of spectral methods, a moving average filter is a low pass filter, in fact the elements of $\pi_t^{(J)}$ are 2^J-period moving averages of the original time series, thus fluctuations with characteristic time scale smaller than 2^J are averaged out and leave the elements of $\pi_t^{(J)}$ unaffected.

The effect of the moving average filter on a generic time series is easily visualized in terms of the Fourier spectrum of the time series. The top subplot of Figure 1 shows the Fourier spectrum of the aggregate consumption growth time series; the shadowed region in the bottom left panel identifies the part of the spectrum which survives after the first application of the moving average filter, namely the spectrum of $\left\{ \pi_{t-k}^{(1)} \right\}_{k \in \mathbb{N}}$. The unshadowed, high frequency part of the spectrum is removed by the application of the filter. In fact in the frequency representation, a 2^j-period moving average operator works as a low band pass filter which removes all those components whose frequency is larger than $2\pi f_{max}/2^j$ where $f_{max} = 2\pi/h$ is the maximum frequency appearing in the spectrum generated by the application of the Fourier transform to a time grid of observations with minimum spacing h.

[Insert Figure 1 about here.]

The illustration shows that the difference between $\pi_t^{(J-1)}$ and $\pi_t^{(J)}$ identifies the component of the original time series with half life with a Fourier spectrum localized in the finite interval of frequencies $\left[\frac{f_{max}}{2^J}\ \frac{f_{max}}{2^{J-1}}\right)$ corresponding to the interval of characteristic timescales $\left[2^{J-1}, 2^J\right)$. This intuition suggests the following:

Definition 2 *Assume the following notational convention: $\pi_t^{(0)} = x_t$. The time series $\delta^{(J)} = \left\{\delta_{t-2^J k}^{(J)}\right\}_{k\in\mathbb{N}}$ with elements*

$$\begin{aligned}\delta_{t-2^J k}^{(J)} &= \pi_{t-2^J k}^{(J-1)} - \frac{\pi_{t-2^J k}^{(J-1)} + \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2} \\ &= \frac{\pi_{t-2^J k}^{(J-1)} - \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2}\end{aligned} \tag{2}$$

is called the J-th detail component of the time series x_t. The index J is called the level of persistence of the component. 2^J is called the characteristic "timescale" of the J-th, component. In fact the half life of the $J-th$ component is shorter than 2^J and longer than 2^{J-1}.

Figure 1 inset 3 shows the spectrum of the details δ_t^1 as extracted by the time series for consumption growth. A simple recursive procedure is practically used for the iterative computation of the decomposition. The details at scale 2^j are formed by the first differences of a moving average over 2^j periods of the original time series. Continuing the exemplification in Figure 1 the insets in the second column show the Fourier spectra of each term (namely the details $\delta_t^{(1)}$ and $\delta_t^{(2)}$ and of the persistent component $\pi_t^{(2)}$ obtained by a decomposition truncated at level $J = 2$.

Thanks to the recursive nature of the definition of $\pi_t^{(J)}$ and $\delta_t^{(J)}$ it is immediate to verify that:

Corollary 3 *The element x_t of the original time series can be decomposed as:*

$$x_t = \sum_{j=1}^{J} \delta_t^{(j)} + \pi_t^{(J)} \tag{3}$$

Each element of the decomposition can be computed using only observations prior to time t.

We are now ready to introduce the following

Definition 4 *The "redundant" Persistence Based Decomposition (PBD hereafter) truncated at level J of $\mathbf{x}_t$ is the vector $\left(\delta_t^{(J)}, \pi_t^{(J)}\right)'$, where $\delta_t^{(J)}$ denotes the sequence of details $\delta_t^{(J)} \equiv \left\{\delta_t^{(j)}\right\}_{j=1,..,J}$ and $\pi_t^{(J)}$ the level J scale component.*

Following the notation of Renaud, Starck and Murtagh (2005), the above decomposition is qualified as "redundant". In fact suppose that the above PBD is repeated at any time t and the information is collected in the array of vectors $\left\{\left(\delta_{t-k}^{(J)}, \pi_{t-k}^{(J)}\right)'\right\}_{k=0,2^J-1}$. This array is redundant because its elements are not linearly independent. The redundant detail vectors allow to linearly reconstruct the original vector of observations $\left(x_{t-2^J+1}, ..., x_t\right)'$ in an infinite number of ways. This implies that in general the redundant details are correlated even if the original observations are not, i.e. a redundant representation of the time series generates spurious correlations in the data. An alternative, "decimated" decomposition can be defined by selecting a minimal subset of details which is necessary to invert the transformation and reconstruct the vector of observations $\left(x_{t-2^J+1}, ..., x_t\right)'$. Cumbersome linear algebra considerations, see e.g. Renaud et al. (2005)and references therein, show that the reduction to the "minimal subset" is obtained sampling details at scale j on a coarser grid with time spacing $2^j h$.

In light of the above observations we can state the following:

Definition 5 *The decimated PBD truncated at level J of $\mathbf{x}_t$ is given by the vector $\left(\delta_t^{d(J)}, \pi_t^{(J)}\right)$ where $\delta_t^{d(J)} \equiv \left\{\delta_{t-2^j k_j}^{(j)}\right\}_{j=1,..,J,\ k_j=0,.,2^{J-j}-1}$ and $\pi_t^{(J)}$.*

Observe that the collection of vectors $\left\{\left(\delta_{t-2^J k}^{d(J)}, \pi_{t-2^J k}^{(J)}\right)\right\}_{k\in\mathbb{N}}$ allows the exact reconstruction of the complete original time series and therefore contains the same information of the array of redundant PBD $\left\{\left(\delta_{t-k}^{(J)}, \pi_{t-k}^{(J)}\right)'\right\}_{k\in\mathbb{N}}$.

The construction of the "decimated" PBD is easily understood by making use of a matrix algebra. Consider the decomposition with maximum level $J = 2$. We first group the variables $\{x_{t-k}\}_{k\in\mathbb{N}}$ in (disjoint) blocks of length 2^2.

$$X_t^{(2)} = \begin{pmatrix} x_{t-3} \\ x_{t-2} \\ x_{t-1} \\ x_t \end{pmatrix}$$

The matrix $\mathcal{T}^{(2)}$ that maps the the block of 4 observations in the minimum vector of details is given by:

$$\mathcal{T}^{(2)} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{4}$$

Using equation (1) and (2) we get

$$\widetilde{X}_t^{(2)} = \begin{pmatrix} \pi_t^{(2)} \\ \delta_t^{(2)} \\ \delta_{t-2}^{(1)} \\ \delta_t^{(1)} \end{pmatrix} = \begin{pmatrix} \frac{x_t+x_{t-1}+x_{t-2}+x_{t-3}}{4} \\ \frac{x_t+x_{t-1}-x_{t-2}-x_{t-3}}{4} \\ \frac{x_{t-2}-x_{t-3}}{2} \\ \frac{x_t-x_{t-1}}{2} \end{pmatrix}$$

In this example the decimated set of details is $\delta_{2,t}^D = \left\{\delta_t^{(1)}, \delta_{t-2}^{(1)}, \delta_t^{(2)}\right\}$ while the redundant one is $\delta_{2,t} = \left\{\delta_t^{(1)}, \delta_t^{(2)}, \delta_{t-1}^{(1)}, \delta_{t-1}^{(2)}, \delta_{t-2}^{(1)}, , \delta_{t-2}^{(2)}, \delta_{t-3}^{(1)}, \delta_{t-3}^{(2)}\right\}$. The minimal property of the decimated PBD is proved by the invertibility of the matrix $\mathcal{T}^{(2)}$.

It has to be underlined that the redundant and decimated versions of the PBD are perfectly consistent and equally relevant for econometric analysis. The redundant PBD can be computed at any time t, i.e. it is updated with the highest frequency and all the incoming information is immediately embedded in it. For this reason, the redundant PBD seems to be useful to improve the out of sample forecasting ability.

On the other hand, by construction, each detail at level j, $\delta_{t-k}^{(j)}$, is formed by a moving average of the observations x_{t-k} over a window of width 2^j, hence a spurious autocorrelation patterns will appear. In particular, when regressor and regressand are given by the time series of "redundant" details, the statistical significance tests on the results obtained running a linear predictive regression require an explicit modification to take into account the correlation induced on residuals by the moving average effect over a window of 2^j. The "decimated" PBD removes this spurious pattern at the cost of a large reduction of the sample size, since innovations with persistence j are then localized on a coarser grid with spacing 2^j.

In conclusion the tension between redundant and decimated PBD reflects the fact that innovations on the component at level of persistence j are naturally adapted (localized) on a time grid with time step 2^j, thus their measures over shorter time intervals generate spurious correlations.

The tradeoff between a higher resolution and the absence of spurious correlations is in fact unavoidable in theory and in practice and has to be considered as part of the identification procedure which connects a structural model to the observations of factors in a mixed frequency framework.

The persistence based decomposition highlights an additional distortion that affects dynamic estimation of a model in the presence of heterogeneity of persistence levels. Consider the case where the original observations form a white noise sequence of unit variance innovations $\{\varepsilon_{t-\kappa}\}_{k\in\mathbb{N}}$, $\varepsilon_t \sim N(0,1)$. Then the computation of the decimated persistence based decomposition produces:

$$WN_t = \begin{pmatrix} \varepsilon_{t-3} \\ \varepsilon_{t-2} \\ \varepsilon_{t-1} \\ \varepsilon_t \end{pmatrix}, \ \mathcal{T}^{(2)} WN_t = \begin{pmatrix} \pi_t^{(2)} \\ \delta_t^{(2)} \\ \delta_{t-2}^{(1)} \\ \delta_t^{(1)} \end{pmatrix} = \begin{pmatrix} \varepsilon_t^{\pi}/2 \\ \varepsilon_t^{(2)}/2 \\ \varepsilon_{t-2}^{(1)}/\sqrt{2} \\ \varepsilon_t^{(1)}/\sqrt{2} \end{pmatrix}$$

The innovations ε_t^{π}, $\varepsilon_t^{(j)} \sim N(0,1)$ are again standard normal innovation processes: ε_t^{π} describes the innovation on the scale process while the sequence $\left\{\varepsilon_t^{(j)}\right\}_{j=1,..,J}$ describe detail components of a white noise $\varepsilon_t^{(j)} = (\varepsilon_t - \varepsilon_{t-2^j})/2^{j/2}$. Innovations of detail components with level of persistence j, $\varepsilon_t^{(j)}$, have standard normal distributions but their spectral densities are localized on the specific frequency band corresponding to half lives in the interval $\left(2^{j-1}, 2^j\right]$. The PBD of the innovation ε_t is then given by:

$$\varepsilon_t = \frac{\varepsilon_t^{(1)}}{\sqrt{2}} + \frac{\varepsilon_t^{(2)}}{2} + \frac{\varepsilon_t^{\pi}}{2}$$

A key observation is that the contribution to the period variance of the detail component $\delta_t^{(2)}$ is half the contribution of of the detail component $\delta_t^{(1)}$:

$$Var\left(\delta_t^{(2)}\right)/Var\left(\delta_t^{(1)}\right) = Var\left(\frac{\varepsilon_t^{(2)}}{2}\right)/Var\left(\frac{\varepsilon_t^{(1)}}{\sqrt{2}}\right) = \frac{1}{2}$$

hence the higher the level of persistence of the component, the lower its contribution to the instantaneous variance of the shock.

Hence a statistical analysis based on the short term, single period, variance decomposition will generally underweight the importance of high persistence shocks compared to the effect of transitory shocks. This observation provides a strong motivation to the use of a filtering approach to disentangle the low frequency components in order to avoid a sever error in variables problem (for related analysis see Ortu et al. (2011)).

This distortion is corrected by introducing the Haar transform see e.g. Renaud et al. (2005), a mild modification of the transformation $\mathcal{T}^{(J)}$. The elements of the exact Haar

Transformation matrix are obtained by rescaling the matrix elements of $\mathcal{T}^{(J)}$ in order to generate an equivalent linear transformation which has also the isometric property. The isometric Haar matrix for $J = 2$ is given by:

$$\mathcal{T}^{(2)}_{Haar} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} x_{t-3} \\ x_{t-2} \\ x_{t-1} \\ x_t \end{pmatrix} \tag{5}$$

Let $Y^{(2)}_t$ be the vector of the Haar transform

$$Y^{(2)}_t = \mathcal{T}^{(2)}_{Haar} X^{(2)}_t,$$

Comparing $Y^{(2)}_t$ with $\widetilde{X}^{(2)}_t$ it is immediate to verify that:

$$\widetilde{X}^{(2)}_t = \begin{pmatrix} \pi^{(2)}_t \\ \delta^{(2)}_t \\ \delta^{(1)}_{t-2} \\ \delta^{(1)}_t \end{pmatrix}, \quad Y^{(2)}_t = \begin{pmatrix} \left(\sqrt{2}\right)^2 \pi^{(2)}_t \\ \left(\sqrt{2}\right)^2 \delta^{(2)}_t \\ \left(\sqrt{2}\right)^1 \delta^{(1)}_{t-2} \\ \left(\sqrt{2}\right)^1 \delta^{(1)}_t \end{pmatrix}$$

and the application of the Haar Transform to the white noise sequence gives:

$$\mathcal{T}^{(2)}_{Haar} WN_t = \begin{pmatrix} \varepsilon^{\pi}_t \\ \varepsilon^{(2)}_t \\ \varepsilon^{(1)}_{t-2} \\ \varepsilon^{(1)}_t \end{pmatrix}$$

and the distortion factors disappear.

In conclusion the introduction of the decimated PBD and of the isometric transform highlight the two "distortions" which complicate the statistical filtering of the low frequency components of random signals: shocks at scale of persistence j are naturally adapted on a scale 2^j and their filtration at higher frequencies induces spurious correlation effects in the observation, the instantaneous variance underweights the contribution to long run integrated variance of high persistence components. In the above analysis the PBD was truncated at a finite maximum level J corresponding to a maximum averaging window of 2^J periods. In the next section we prove that the iteration of the above recursive scheme converges in the limit $J \to +\infty$ and produces the non truncated PBD.

3 The Persistence Based Decomposition of a Time Series

In this section we state the two main results of the paper. The next theorem characterizes the the full, non truncated PBD of the time series $\{x_{t-k}\}_{k\in\mathbb{N}}$ of observations of a stationary stochastic process.

Theorem 6 *Consider the PBD of a stationary time series $\{x_{t-k}\}_{k\in\mathbb{N}}$ with Wold decomposition $x_t = \mu + \psi(L)\varepsilon_t$. Then:*

1. *the sequence of random variables $\{\pi_t^{(J)}\}_{J=0}^{+\infty}$ converges a.s. to a constant equal to the mean:*
$$\pi_t^{(\infty)} \equiv \lim_{J\to+\infty} \pi_t^{(J)} = \mu$$
and the following decomposition holds for x_t:
$$x_t = \sum_{j=1}^{+\infty} \delta_t^{(j)} + \mu \tag{6}$$

2. *the variance of the rescaled permanent component $\pi_t^{(J)}$ converges to the long run variance.*
$$\lim_{J\to\infty} \sqrt{Var\left[\sqrt{2^J}\pi_t^{(J)}\right]} = \psi(1)$$

3. *the details $\delta_t^{(j)}$ as defined by eq.(2) are first differences of stationary processes and the corresponding time series have zero long run variance.*

Proof. The proof is provided in Appendix B. ■

The next theorem states the most general version of the PBD which applies to a (possibly) non stationary integrated process:

Theorem 7 *Consider an integrated time series $\mathbf{y}_t = \{y_{t-k}\}_{k\in 0,..,+\infty}$ such that $E\left[y_0^2\right] < +\infty$ and let the time series of the first differences $x_t = \Delta y_t$ admit the Wold representation $x_t = \mu + \psi(L)\varepsilon_t$ with $\sum_{j=0}^{+\infty} j\psi_j < +\infty$. Then:*

$$y_t - y_0 = \tilde{\pi}_t^{(\infty)} + \sum_{j=1}^{+\infty} \widetilde{\delta}_t^{(j)} \tag{7}$$

where:

1. *the details $\widetilde{\delta}_t^{(j)}$ at any level of persistence j are stationary and their Wold decompositions are given by*

$$
\begin{aligned}
\widetilde{\delta}_t^{(j)} &= -\sum_{k_j=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)} \widetilde{\psi}\right)_{j,k_j} \varepsilon_{j,t-2^j k_j} \quad \varepsilon_{j,t-2^j k_j} = \sum_{k=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k} \varepsilon_{t-k} \\
\left(\mathcal{T}_{Haar}^{(\infty)} \widetilde{\psi}\right)_{j,k_j} &= \sum_{k=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k} \widetilde{\psi}_k, \qquad \widetilde{\psi}_k \equiv \left(\sum_{j>k} \psi_j\right)
\end{aligned}
$$

where $\mathcal{T}_{Haar}^{(\infty)}$ [5] *is the (infinite) matrix which implements the Haar transform.*

2. *the time series of the asymptotic scale component equals a random walk plus a deterministic trend:*

$$
\tilde{\pi}_t^{(\infty)} = \mu t + \psi(1) \sum_{s=1}^{t} \varepsilon_s .
$$

3. *the sequence of stationary details $\widetilde{\delta}_t^{(j)}$ $j \in \mathbb{N}$ can be computed using the recursive definition (2) for the integrated time series $\mathbf{y}_t = \{y_{t-k}\}_{k \in 0,..,+\infty}$.*

Proof. The theorem is proved in Appendix C. ■

It is important to remark that the above decompositions result from a purely spectral classification and should not be confused with a parametric assumption. Drawing an analogy with the traditional statement of the Wold Theorem, the possibility to decompose a process as an infinite moving average of uncorrelated innovations does not preclude the use more parsimonious parametric models, in the same way the PBD does not prevent from adopting more traditional time series models whose evolution involves only a small number of details.

As a consequence the use of the above spectral decomposition must be motivated by proving in a structural approach that details with different level of persistence are different channels of transmission of independent information. In particular when the information conveyed by details with high level of persistence and relevant over the long run is relevant for the economic analysis, the PBD is necessary in order to remove the distortions analyzed in

[5] The Haar Transform matrix $\mathcal{T}_{Haar}^{(\infty)}$ can be formally defined as the direct limit:

$$
\lim_{J \to +\infty} \mathcal{T}_{Haar|\delta}^{(J)} = \mathcal{T}_{Haar}^{(\infty)}
$$

where $\mathcal{T}_{Haar|\delta}^{(J)}$ is the matrix which is obtained by $\mathcal{T}_{Haar}^{(J)}$, the matrix which implements the transform truncated at level J but restricted to the vector space spanned by details (without the first line).

Section 2. Hence the use of a multiresolution approach is necessary in order to produce a good description of low frequency high persistence economic factor's properties.

In the next section we provide a more extended discussion of the spectral definition of the PBD and its relation with well known decomposition approaches.

4 The Spectral Definition of the Persistence Based Decomposition

The above decomposition can be obtained by an application of the abstract Wold theorem (Wold 1938) which is reviewed in Appendix A.

Abstract Wold theorem states that any isometric operator F in an Hilbert space H produces an orthogonal decomposition of the Hilbert

$$\begin{aligned} H &= H_S \oplus H^{(\infty)} \\ H_S &= \mathcal{W} \oplus F\mathcal{W} \oplus F^2\mathcal{W} \oplus \ldots = \bigoplus_{n=0}^{\infty} F^n \mathcal{W} \end{aligned}$$

where $\mathcal{W}$, the wandering subspace is defined as:

$$\mathcal{W} \equiv H_S \ominus F H_S$$

and $H^{(\infty)}$ is defined by

$$H^{(\infty)} \equiv \cap_n F^n H$$

The abstract theorem produces the traditional Wold decomposition (see e.g. Brockwell and Davis) if the isometric operator is identified with the lag operator L and the Hilbert space H is identified with the Hilbert space of covariance stationary processes. The derivation of the traditional Wold decomposition for stationary time series from the abstract version is reviewed in the Appendix. Within our analysis it will be assumed that the stationary time series $\{x_{t-k}\}_{k\in\mathbb{N}}$ is regular, i.e. there's no non constant component which is orthogonal to all past innovations. Hence it admits an square summable standard Wold representation:

$$\begin{aligned} x_t &= \mu + \psi(L)\,\varepsilon_t \\ \sum_{j=0}^{+\infty} \psi_j^2 &< +\infty \end{aligned}$$

The PBD eq.(6) is obtained by an application of the abstract Wold Theorem to a different isometry operator acting on a different Hilbert space. The Hilbert space and the Isometric operator which are necessary in order to characterize the PBD are introduced in the next:

Definition 8 *Let*

$$\mathcal{H}\left(\mathbf{x}_t\right) = \left\{ Z = \sum_{k\in\mathbb{N}} \alpha_k x_{t-k},\ \sum_{k\in\mathbb{N}} \left|\alpha_{t-k}\right|^2 < +\infty, \quad \left\langle Z^1, Z^2 \right\rangle = \sum_{k\in\mathbb{N}} \alpha_k^1 \alpha_k^2 \right\}$$

Let the dyadic dilation operator centered at time t be defined by:

$$\begin{gathered} D : \mathcal{H}\left(\mathbf{x}_t\right) \rightarrow \mathcal{H}\left(\mathbf{x}_t\right) \\ \widetilde{\mathbf{x}} = \left\{\widetilde{x}_{t-k}\right\}_{k\in\mathbb{N}} \rightarrow \mathbf{x}_t^{(1)} = \left\{x_{t-k}^{(1)}\right\}_{k\in\mathbb{N}} = D\widetilde{\mathbf{x}} \\ x_{t-k}^{(1)} = \sqrt{2}\widetilde{x}_{t-2k} \end{gathered}$$

Then the rescaling operator, R, centered in t is defined by:

$$R = D \circ M \tag{8}$$

i.e. it is the composition of the dyadic dilation operator D with the dyadic mean M.

and it is possible to state the following:

Theorem 9 *R acts as an isometric operator in $\mathcal{H}\left(\mathbf{x}_t\right)$. By abstract Wold theorem it is possible to decompose $\mathcal{H}\left(\mathbf{x}_t\right)$ as follows:*

$$\begin{aligned} \mathcal{H}\left(\mathbf{x}_t\right) &= \bigoplus_{j=1}^{+\infty} R^j \mathcal{W}_t^R \oplus \mathcal{H}_{t,R}^{(\infty)} \\ \mathcal{W}_t^R &= R\mathcal{H}\left(\mathbf{x}_t\right) \ominus \mathcal{H}_t\left(\mathbf{x}_t\right) \end{aligned} \tag{9}$$

$R^j \mathcal{W}_t^R = \left\langle \delta_t^{(j)} \right\rangle$ is the linear subspace generated by the time series of details with level of persistence j.

$\mathcal{H}_{t,R}^{(\infty)}$ is the largest vector subspace of $\mathcal{H}_t\left(\mathbf{x}\right)$ spanned by eigenvectors of R with eigenvalue larger or equal than 1. It satisfies the fixed point equation:

$$R\mathcal{H}_{t,R}^{(\infty)} = \mathcal{H}_{t,R}^{(\infty)}$$

Proof. The theorem is proved in Appendix B. ■

In light of the above operator decomposition we are now ready to state the following convergence result under suitable parametric assumptions.

Theorem 10 *Consider the PBD of a stationary time series $\{x_{t-k}\}_{k\in\mathbb{N}}$ with Wold decomposition $x_t = \mu + \psi(L)\,\varepsilon_t$ and denote by P_H the orthogonal projection onto the linear subspace H, then the PBD decomposition is determined by the decomposition induced by the isometric operator R through the following equalities:*

$$\begin{aligned} P_{R^j \mathcal{W}_t^R} x_t &= \delta_t^{(j)} \\ P_{\mathcal{H}_{t,R}^{(\infty)}} x_t &= \mu \end{aligned}$$

Proof. The proof is provided in Appendix B. ■

The stationary hypothesis plays a central role in the characterization of $\mathcal{H}_{t,R}^{(\infty)}$. By definition

$$\mathcal{H}_{t,R}^{(\infty)} = \cap_{j=0,..,+\infty} R^j \mathcal{H}_t(\mathbf{x})$$

Note that the fixed point property pertains to the vector space and not necessarily to its elements. Consider for example the linear space of constant processes, this one dimensional space is invariant as a whole with respect to rescaling, in fact the image under a rescaling R of a constant process $\mu = (\mu, \mu, \mu, ..)$ is still a constant process:

$$R\mu = \sqrt{2}\mu$$

but each constant process is not a fixed point for the rescaling operator, in fact $R\mu \neq \mu$. This proves that the space of constant processes is a subspace of $\mathcal{H}_{t,R}^{(\infty)}$. Then additional independent generators of $\mathcal{H}_{t,R}^{(\infty)}$ are to be found in the set of zero mean stationary processes. Let $\mathbf{x}_t$ a zero mean stationary process with summable covariances and consider the image of $\mathbf{x}_t$ under the application J times of the rescaling operator $R^J\mathbf{x}_t$, in the limit of $J \to +\infty$. Then application of the CLT for stationary time series with summable covariances, see Hall and Heyde (1980), yields immediately the result:

$$\left(\mathbf{s}_t^{(\infty)}\right)_k \equiv \lim_{J\to+\infty} \left(R^J \mathbf{x}_t\right)_k = \lim_{J\to\infty} \sqrt{2^J}\pi_{t-2^J k}^{(J)} = \lim_{J\to\infty} \left(\frac{\sum_{n=0}^{2^J} x_{t-k2^J-n}}{\sqrt{2^J}} \right) \stackrel{d}{=} \psi(1)\,\varepsilon_{t-k} \tag{10}$$

where $\psi(1)$ is the long run variance of $\mathbf{x}_t$ which is finite given the summability of the covariances and ε_{t-k} is a normally distributed standard innovation. Hence the limit process $\mathbf{s}_t^{(\infty)}$ is a fixed point process for the rescaling R:

$$R\mathbf{s}_t^{(\infty)} \overset{d}{=} \mathbf{s}_t^{(\infty)} \tag{11}$$

The corresponding linear subspace belongs to $\mathcal{H}_{t,R}^{(\infty)}$ and exhausts all the possible limiting expressions that can be obtained from a a zero mean stationary time series with summable covariances $\mathbf{x}$ and thus the zero mean elements of $\mathcal{H}_{t,R}^{(\infty)}$.

While the contribution of $\mathbf{s}_t^{(\infty)}$ to the PBD of a stationary time series becomes negligible as $J \to +\infty$, it plays a key role when the decomposition is applied to a nonstationary unit root process. In fact it is possible to prove that this component is the stochastic trend as defined in Beveridge and Nelson (1981) (BN hereafter).

Definition 11 *(Hayashi pg. 562) The permanent component (stochastic trend) p_t^{BN} of a unit root non stationary process y_t process is that component whose effect is not expected to decay but "persists" at any horizon. A transitory component of a process is that component which becomes negligible when looked at a sufficiently large horizon. Hence for a time series y_t the BN permanent component is defined by:*

$$\begin{aligned} p_t^{BN} &= \lim_{h\to+\infty} y_t + E\left[\sum_{k=1}^{h} \Delta y_{t+k} - \mu h \mid \Omega_t\right] \\ \mu &= E\left[\Delta y_t\right] \end{aligned} \tag{12}$$

where Ω_t is the information set available at time t.

This purely formal definition requires the specification of a probabilistic time series model to produce an effective identification and estimation procedure. The simplest computational scheme of the permanent component p_t^{BN} assumes that the first difference process $x_t = \Delta y_t$ admits a (standard) Wold representation with summable coefficients:

$$\begin{aligned} \Delta y_t &= x_t = \mu + \psi(L)\epsilon_t \\ \psi(L) &= \sum_{j=0}^{+\infty} c_j L^j \end{aligned} \tag{13}$$

then:

$$\begin{aligned} \Delta y_t &= x_t = \mu + \psi(L)\epsilon_t \\ &= \mu + \psi(1)\epsilon_t + \left[\psi(L) - \psi(1)\right]\epsilon_t \end{aligned} \tag{14}$$

and using the telescopic relation $y_t - y_0 = \sum_{t'=0}^{t} \Delta y_{t'}$, the process y_t can be written as:

$$\begin{aligned}
y_t - y_0 &= \mu \cdot t + \psi(1) z_t + \psi^*(L)\epsilon_t \\
z_t &= z_{t-1} + \epsilon_t \\
\psi^*(L) &= [\psi(L) - \psi(1)][1-L]^{-1} = -\sum_{k=0}^{\infty} \psi_k^* L^k \\
\psi_k^* &= \left(\sum_{j>k} \psi_j\right)
\end{aligned}$$

hence y_t is decomposed as a sum of a linear deterministic trend μt, a stochastic trend (martingale component) $\psi(1) z_t$ and a transitory (stationary) component represented by $\psi^*(L)\epsilon_t$. Recall that $\psi(1)$ determines the Long Run Variance of the $I(0)$ process x_t. In this framework the permanent component coincides with the stochastic trend, hence:

$$\begin{aligned}
p_t^{BN} &\equiv \mu t + \psi(1) z_t \\
p_t^{CY} &= y_t - p_t^{BN} = \psi^*(L) = -[\psi(1) - \psi(L)][1-L]^{-1}\epsilon_t
\end{aligned} \tag{15}$$

In order to relate BN analysis with the PBD, the natural starting point is the observation that the first differences of the BN permanent component are fixed points of the rescaling operator R:

Theorem 12 *Let $\mathbf{y}_t = \{y_{t-k}\}_{k\in\mathbb{N}}$ a zero mean unit root process such that its increments $x_{t-k} = \Delta y_{t-k}$ admit a representation $x_t = \mu + \psi(L)\varepsilon_t$. Let $\mathbf{p}_t^{BN}$ and $\mathbf{p}_t^{CY}$ the BN permanent and transitory components of $\mathbf{y}_t$, let $\Delta\mathbf{p}_t^{BN} := \{p_{t-k}^{BN} - p_{t-1-k}^{BN}\}_{k\in\mathbb{N}}$ the time series of first differences of the stochastic trend component and $\mathbf{s}_t^{(\infty)}$ defined by (11). Then the following equalities hold:*

$$\Delta\mathbf{p}_t^{BN} = \mathbf{s}_t^{(\infty)}$$

thus time series of the persistent component increments satisfies the fixed point equation:

$$R\Delta\mathbf{p}_t^{BN} \overset{d}{=} \Delta\mathbf{p}_t^{BN} \tag{16}$$

where R is the rescaling operator defined in eq.(8).

Proof. Apply the rescaling transformation to eq.(14), then it is immediate to verify that $\Delta p_t^{BN} = \psi(1)\varepsilon_t$. It is directly implied by the stability of the Gaussian distribution with respect to the scaling operator. ■

Hence in the space $\mathcal{H}_t(\mathbf{x})$, the shocks driving the BN "permanent" component belong to $\mathcal{H}_{R,t}^{(\infty)}(\mathbf{x})$, i.e. to the scale component of the PBD, and the theorem formalizes the intuition that the permanent component is driven by shocks which "do not show a decay over any timescale".

The substantial improvement achieved by making use of the abstract Wold theorem is determined by the action of the PBD on the transitory component. The detail components $\delta_t^{(j)} \in R^j \mathcal{W}_t^R$ appearing in the PBD of $\mathbf{x}_t$ are by definition orthogonal to the permanent component which belongs to $\mathcal{H}_{R,t}^{(\infty)}(\mathbf{x})$ and by Theorem (6) each detail time series is the first difference of a stationary process. Hence the detail components selected by the PBD of $\mathbf{x}_t$ contribute to the transitory component $\mathbf{p}_t^{CY}$ of the integrated process $\mathbf{y}_t$.

The above results critically hinge on the possibility to have an infinite number of observations. A sharper characterization of the finite sample version of the PBD is analyzed in the next section.

5 The asymptotic limit theory for $T \to +\infty$ of the PBD.

In this section a sample of observations of size T is thought to be the $T-th$ row of a triangular double array random process $\left\{\left\{x_t^{(T)}\right\}_{t=1}^{T}\right\}_{T=1}^{+\infty}$, in order to analyze the asymptotic properties of the PBD as $T \to +\infty$.

Within this framework, following Stock (1994), Davidson (1999,2002), White (2001, p.179) and Breitung (2002) it is natural to introduce:

Definition 13 *The process* $\left\{y_{t-k}^T\right\}_{k=1}^T$ *is a zero mean* $I(1)$ *process if:*

$$\exists \sigma > 0 : T^{-1/2}\sigma^{-1} y_{\lfloor rT \rfloor}^{(T)} \overset{T \to +\infty}{\Longrightarrow} W(r)$$

where $\lfloor \cdot \rfloor$ *is the largest smaller integer function,* $W(\cdot)$ *is a standard Wiener process in the unit interval and* $\overset{T\to+\infty}{\Longrightarrow}$ *denotes weak convergence*[6] *as* $T \to +\infty$.

The process $\left\{x_{t-k}^T\right\}_{k=1}^T$ *is a zero mean* $I(0)$ *process if it is the first difference of an* $I(1)$ *process, hence:*

$$\forall r \in [0,1], \exists \sigma > 0 : T^{-1/2}\sigma^{-1} \sum_{t=1}^{\lfloor rT \rfloor} x_t^{(T)} \overset{T \to +\infty}{\Longrightarrow} W(r)$$

[6] Convergence has to be understood as the weak convergence in the space of cadlag processes in the unit interval equipped with the Skorohod topology.

The weak limit procedure fixes r in the unit interval and imposes that the fraction of observations with (discrete) time index $t < rT$ as $T \to +\infty$ converge to a Wiener process on the unit interval, the unique process with continuous paths having increments that are stationary, normally distributed and uncorrelated.

This asymptotic limit has important implications for the asymptotic expression of the scale component in the PBD. The limit $T \to +\infty$ and $J \to +\infty$ are not independent and can be simply related thanks to the following observation:

Proposition 14 *Consider the time series of observations* $\{x_{t-k}^{T}\}_{k=1}^{T}$ *with sample size* T. *The maximum scale component corresponding to a level of persistence* $J_{\max}(T) = \lfloor \log_2(T) \rfloor$ *is determined by the time series* $\left\{\pi_t^{(J_{\max})}(\omega)\right\}_{0 \le t \le T}$, *with elements:*

$$\pi_t^{(J_{\max})} = \frac{1}{2^{J_{\max}(T)}} \sum_{k=0}^{t} x_k^{(T)} \quad 0 \le t \le T$$

and cannot be further decomposed on the basis of the finite set of observations. Hence the detail components $\delta_t^{(j)}$ *with level of persistence* $j > J_{\max} = \lfloor \log_2(T) \rfloor$ *and the scale component* $\pi_t^{(\infty)}$ *cannot be disentangled on the basis of the finite sample of* T *observations.*

On the basis of Definition 13 and of the previous observation, it is immediate to state the following:

Corollary 15 *Apply the PBD to a zero mean* $I(0)$ *process, then the triangular array of random variables* $\{\left\{\sqrt{2^J}\pi_t^{(J)}(\omega)\right\}_{0 \le t \le 2^J}\}_{J=1}^{+\infty}$ *weakly converges as* $J \to +\infty$ *to the limit:*

$$\sqrt{2^J}\pi_{\lfloor 2^J r \rfloor}^{(J)} \Rightarrow \psi(1) W(r)$$

where $\psi(1)$ *is the long run variance of the* $I(0)$ *process.*

Apply the PBD to a zero mean $I(1)$ *process, then the triangular array of random variables* $\{\left\{\pi_t^{(J)}(\omega)/\sqrt{2^J}\right\}_{0 \le t \le 2^J}\}_{J=1}^{+\infty}$ *weakly converges as* $J \to +\infty$ *to the limit:*

$$\pi_{\lfloor 2^J r \rfloor}^{(J)}/\sqrt{2^J} \Rightarrow \psi(1) \int_0^1 W(r)dr$$

where $\psi(1)$ *is the long run variance of the first difference process.*

Proof. Fix $T^{(J)} = 2^J$, then the triangular array generated by $\sqrt{2^J}\pi^{(J)}_{\lfloor 2^J r \rfloor}$, $J \to \infty$, is a subsequence of the triangular array with elements $y^{(T)}_{\lfloor rT \rfloor} = \frac{1}{\sqrt{T}} \sum_{k=0}^{\lfloor rT \rfloor} x^{(T)}_k$ where we recall that $x^{(T)}_k$ are observations extracted from a $I(0)$ processes. Hence the two triangular arrays have the same limiting process and by the functional central limit theorem for $I(1)$ processes $T^{-1/2} y^{(T)}_{\lfloor rT \rfloor} \overset{T \to +\infty}{\Longrightarrow} W(r)$ and the thesis is proved.

The second part of the theorem is also derived through a simple application of the functional limit theorem. For any driftless random walk y_t, the functional $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} y_{t-1}$ converges to:

$$\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} y_{t-1} \overset{T \to +\infty}{\Rightarrow} \sigma \int_0^1 W(r)dr$$

■

The theorem is very important for practical applications when the order of integration is unknown. Then the computation of the scale component at different levels of persistence offers a non parametric procedure to discriminate between the $I(0)$ and $I(1)$ alternatives based on the growth rate of volatility of the scale component $\pi^{(J)}_t$ is $\left(2^J\right)^{1/2}$ for an $I(0)$ process and $\left(2^J\right)^{3/2}$ for a $I(1)$ process and offers the opportunity to design accurate tests.

5.1 Simulation analysis

5.1.1 The scale component and the long run variance in a AR(1) process

In this section we analyze the convergence of the scale coefficient under the nonisometric PBD and under the isometric Haar transform.[7] We simulate a series of length $T = 1024$[8] for an AR(1)

$$x_{t+1} = \rho x_t + k + \epsilon_{t+1}$$

with $\epsilon_{t+1} \sim i.i.d.N(0,1)$ and for each process we compute the scaling coefficient obtained both under the isometric transform and the not-isometric transform. We choose $k = 0.3$, $\phi = 0.8$[9] and we repeat the experiment on a sample of 1000 paths. The theory suggests that the scaling coefficient π_t obtained under the non-isometric transform should converge to the unconditional mean. This is in fact the case as it is shown in Figure 3 where we plot the scaling coefficient obtained using the Non Isometric Transform. We see that the coefficients

[7] See Appendix B.

[8] We also simulate sample of length $T = 512$ and $T = 2048$ and results are virtually the same.

[9] We also check results using $\phi = 0, 0.7, 0.9, 0.95$.

wanders around the true unconditional mean $\mu = \frac{k}{1-\phi} = 1.5$ of our AR(1) process with a standard deviation equal to 0.073.

We now study the behavior of the Isometric Haar transform. Results are shown in Figure 4 where we plot the isometric scaling coefficient at time T for the autoregressive process. As Theorem ... suggests, the variance of the scale coefficient for the PBD for the AR(1) case converges to the long run variance. In fact we compute the analytical long run variance for our AR(1) process and we obtain $Avar(x_t) = 25.00$ whereas the variance of our scaling coefficient is equal to $Var(\pi_x(T)) = 23.72$.[10] Figure 6 illustrates the result of the central limit theorem for stationary stochastic processes, i.e.

$$\sqrt{T}\left(\bar{X}_T - \mu\right) \to N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right) \tag{17}$$

In particular the top panel of Figure 6 shows the $Var(\sqrt{T}\pi_x(T))$ for $T = 64, 128, 256, 512, 1024$. We see that $Var(\sqrt{T}\pi_x(T))$ converges to the long run variance of our AR(1) process.

[Insert Figures 3 and 4 about here.]

5.1.2 The stochastic trend of the random walk process.

In this section we analyze the properties of the scaling coefficients under the Isometric Haar transform for a non stationary pure Random Walk process. For an $I(1)$ this component corresponds to the scale component selected by the $I(1)$ version of the PBD. We simulate a series of length $T = 1024$[11] for a random walk

$$y_{t+1} = y_t + \epsilon_{t+1}$$

Figure 5 determines the isometric scaling coefficient at time T,i.e. $\pi_y(T)$. In the first panel we see that the variance is huge. In the second panel we show the coefficient $\pi_y(T)/T^{\frac{3}{2}}$ whose variance instead is equal to 0.3.

In the second panel of Figure 6 we show that controlling for the rate of convergence $2^{\frac{3}{2}J}$ then the long-run variance of our driftless random walk goes to a constant value. In conclusion the presence of a stochastic trend generates a growth of order $2^{\frac{3}{2}J}$ variance as the sample of observations is increased.

[10] We also compute the long run variance using the Newey-West estimator with 300 lags for the bandwidth kernel and we obtain $Avar(x_t) = 18.1661$

[11] We also simulate sample of length $T = 512$ and $T = 2048$ and results are virtually the same.

[Insert Figures 5 and 6 about here.]

5.1.3 Persistence based decomposition and the cointegration of dividends and prices.

To illustrate the performance of the filter we apply the decomposition to the time series of log prices p_t and log dividends d_t[12]. The empirical exercise is based upon quarterly data series of the return on the value portfolio of all US stocks — NYSE, AMEX and NASDAQ — with and without dividends[13]. From these two series we are then able to obtain the price and dividend ones. The dataset is collected from CRSP and the sample spans from the first quarter of 1926 to the fourth quarter of 2007. We report the log dividend and the log price series in Figure 7 and their complete decomposition in Figure 8.

[Insert Figure 7 and Figure 8 about here.]

Each inset numbered from 1 to 9 shows the corresponding detail components ordered by an increasing level of persistence. The last inset shows the permanent component. By definition each of these components has an increasing level of persistence that determines the decrease of the speed of mean reversion which can be observed in the figure.

Two important observations are in order: Figure 9 and 10 compares the first detail $\delta_t^{(1)}$of the Persistence Based Decomposition (PBD) with the transitory component of a standard univariate Beveridge Nelson (BN) decomposition, both time series selected through the standard first order differencing procedure. Visual inspection reveals immediately that they almost overlap. Since the BN permanent component is defined as the difference between the original time series and the transitory part, it is immediate to verify that all the detail components from 2 to $J = 9$ of the PBD are included in the BN permanent component.[14]

[Insert Figure 9, 10 and 11 about here.]

On the contrary the permanent components selected by the PBD for the log dividend and log prices corresponds only to the component which persists beyond scale $J = 9$. According

[12] Lower-case letters will be used to denote variables in logs.

[13] In particular we have $vwretd = (P_t + D_t)/P_{t-1} - 1$, $vwretx = P_t/P_{t-1} - 1$.

[14] See Figure 11 for the log price series; analogous results are obtained for the dividends series and are available upon request from the corresponding author

to the the Campbell and Shiller log-linear approximation the permanent components of p_t and d_t should be cointegrated with cointegration vector $\beta = [1 \ -1]$.

We consider the estimation of the cointegrating vector β considering two separate procedures to analyze the deterministic trend and the stochastic trend.

5.1.4 Deterministic trend

Using a simple OLS regression of the finite sample permanent component of the log-dividend π_t^d on the finite sample permanent component of the price π_t^p:

$$\pi_t^d = \beta_0 + \beta_1 \pi_t^p + u_t$$

selects a cointegrating vector [15] $\beta = [1 \ -1.01]$ very close to the theoretical expectations.

5.1.5 Stochastic trend

We now investigate whether the scaling coefficients for the series of log prices and log dividends grow at rate $T^{\frac{3}{2}}$ as it is the case for an I(1) process. In Figure 12 we plot for each series the scaling coefficient properly rescaled with the rate of convergence of $T^{1.5}$ together with the best linear fit. This linear hypothesis is verified by the behavior of the scaling coefficients surprisingly well and supports the thesis of a unit root behavior for prices and dividends and identifies a cointegration relation for the stochastic component given by $\beta = [1 \ -1.01]$ very close to the theoretical expectations and identical to the one obtained for the deterministic trend.

[Insert Figure 12 about here.]

6 Empirical test (DRAFT VERSION): forecasting GDP and Inflation using the PBD decomposition

One important finding, also remarked in Nelson (2008), is that most of the variation in macroeconomic time series can be ascribed to permanent shocks, which are largely unpre-

[15] Note that the estimated coefficients β_0 and β_1 from the cointegrated regression are superconsistent according to Stock (1987) and dynamic misspecification is not a problem in this regression. Moreover we obtain an R^2 of almost 98% suggesting that if there is any bias due to small sample then this is going to be small.

dictable, whereas the transitory component has very little amplitude. This section will re-examine such finding using our suggested decomposition.

In particular the section presents two illustrations dealing with two relevant macroeconomic indicators of the U.S. economy: quarterly real gross domestic product (GDP, 100 x logarithms, sample period: 1947.q1-2010.q4) and quarterly inflation obtained as $\pi_t = \ln(CPI_t/CPI_{t-1})$, where CPI is the consumer price index for all urban consumers released by the U.S. Bureau of Labor Statistics (seasonally adjusted, January 1947 - December 2008). The series were downloaded from the FRED (Federal Reserve Economic Data) database.

In both cases we apply the predictive validity test proposed by Cogley (2002) and Nelson (2008), which aims at evaluating whether the transitory components extracted through our proposed decomposition contain information that is useful for predicting the future growth of both GDP and inflation.

6.1 U.S. Gross Domestic Product

To validate the predictive content of the transitory components we run the OLS regression of

$$\Delta x_{t+1} = \alpha + \sum_{j=1}^{J} \beta_j x_{t,j} + \epsilon_{t+1} \tag{18}$$

where $\Delta x_{t+1} = x_{t+1} - x_t$ is the next period change in US GDP and $x_{t,j}$ is the transitory component at level of persistence j. Importantly note that the forecast comparisons presented here are limited to univariate methods so the information set for predicting the growth of GDP is only past GDP.

It is important to stress that our measures of cycle are obtained for any given historical quarter only using the observations up to the historical date. Therefore unlike many other approximations to low-pass filters filters, such as those advocated by Baxter and King (1999) or Hodrick and Prescott (1997), this filter is one sided into the past and can be implemented in real time.

The first column in Table 1 presents the estimated regression coefficient and the associated t-value, and finally the coefficient of determination, R^2 of the regression.

[Insert Tables 1 about here.]

The first thing to note is that the components that are significant in the regression and that contribute to the R^2 of 14% are the ones at short horizon.[16]

Next we try to understand how successful is our decomposition compared to competing decompositions in predicting quarterly growth in GDP one quarter ahead. Among the most popular method of trend-cycle decomposition is the filter of Hodrick and Prescott (1997) which seeks to balance smoothness of the cycle against variance of the measured cycle. Further analysis of the 'HP filter' is given by Harvey and Jaeger (1993). Also highly influential have been the Unobserved Component models of Harvey (1985), Watson (1986) and Clark (1987) which both model the trend as a random walk and the cycle as an AR process; here we use the latter which allows the growth rate of the trend to evolve as a random walk as well.

In what follows we will estimates the competing cycles with the benefit of all the subsequent data in revised form. Similar to Nelson (2008) we find that the BN estimates is inherently a one-sided estimate of cycle, so future data only influences estimation of the ARMA parameters. However in the case of GDP growth, the model is AR(1) as suggested by lag selection based on SIC, and the AR parameter is very stable over the sample period. However future data matters the most for the HP filter we will use both a two-sided and one-sided filter.

The next columns in Table 1 presents the results from the regression in (18) augmented with cycle estimates using Beveridge-Nelson, Clark, one-sided and two-sided HP filters. This exercise is in the spirit of Granger's composite prediction and is intended to suggest the marginal information content of cycle estimates. Explanatory power is low, R-square is only .14 for our decomposition alone and none of the other cycle estimates are able to raise this number. None of the other cycle estimates is significant in the presence of our transient components. Note that the difference between R-squared in the last two columns is a measure of the value of hindsight for the HP filter. However as noted by Bryan and Cecchetti (1993), two-sided filters are less useful to monetary policy makers because they reduce the timeliness of incoming inflation data. In contrast, our filter is one-sided into the past, so its output would be available to policymakers as soon as new inflation data became available.

The figure 14 shows the results from a regression of h-period change in GDP, $\Delta_h x_{t+h} = x_{t+h} - x_t$ on the transitory component(s). Panel (a) shows for each forecast lead time h

[16] Exclusion of the components with $j > 3$ would not change the R^2 of 14%. Results are available upon requests.

on the horizontal axis, the adjusted R^2, whereas panel (b) plots the efficiency of the OTTb decomposition measured by

$$G(h) = 100 \times \left(1 - \frac{\widehat{MSFE}(h)_{OTT}}{\widehat{MSFE}(h)_{TC}} \right)$$

i.e. percent gain in forecast accuracy arising from our suggested decomposition compared to alternative TC (trend-cycle) measures.

The R^2 for our suggested decomposition increases up to 45% for horizons around 8 years (32 quarters).

[Insert Figures 13 and 14 about here.]

Our decomposition is able to capture at short horizons, namely $h = 1, 2$ quarters, as much predictability as the Beveridge Nelson decomposition. The results confirm Nelsons overall conclusion that when $h = 1$ only a small fraction of future GDP growth is predictable using the BN transitory component. However our decompositon plays differential role in explaining the GDP fluctuations at longer horizons. In fact, contrary to a body of recent empirical work that finds that fluctuations in GDP are permanent, our evidence shows that GDP does, in fact, revert toward a "trend" following a shock. However, confirming the results in Cochrane (1988), that reversion occurs over a time horizon characteristic of business cycles-several years at least. Therefore, quoting from Cochrane (1988) "the short-run properties of GDP are consistent with a model with very persistent shocks, and one can incorrectly infer a great deal of long-horizon persistence by fitting a time-series model to this short-run behavior".

6.2 U.S. Inflation

Our second exercise is motivated by a series of papers, Cecchetti (1997), Bryan and Cecchetti (1993, 1994, 1995), and Bryan, Cecchetti and II (1997) that stress that as the focus of central banks shifts toward inflation targeting, it becomes increasingly important to have accurate measures of inflation. In what follows, we show that our proposed method can significantly improve inflation forecasts over horizons of up to eight years.

Therefore it makes sense to investigates the univariate properties of the proposed measures by studying the predictive content of the transitory component(s) with respect to the h-period change in inflation, i.e. the regression of $\Delta_h x_{t+h} = x_{t+h} - x_t$ on the transitory component(s) extracted using our suggested decomposition and competing ones proposed in the literature.

$$\Delta_h x_{t+h} = \alpha_h + \sum_{j=1}^{J} \beta_j x_{t,j} + \epsilon_{t+h}$$

where $x_t = \pi_t$.

Among the alternative candidate measures we consider the Beveridge-Nelson component at time t, $x_t - \hat{m}_t$, where m_t is the Beveridge-Nelson trend, and the exponential smoother of inflation proposed in Cogley (2002).[17] This last measure computes first a measure of core inflation[18] using a "constant-gain" update measure, i.e.

$$\mu_t = \mu_{t-1} + g_0(\pi_t - \mu_{t-1})$$

where μ_t is the period t estimate of mean inflation and g_0 is the gain parameter, which is assumed to lie between 0 and 1. Then the core deviation, $\pi_t - \mu_t$ is used as a predictor. This is reasonable since a successful measure of core inflation should purge the transients from actual inflation and the difference between actual and core inflation should predict subsequent changes in inflation. That is, when actual inflation is above its core value, inflation should fall as the transients accounting for the high current level die out. Importantly both our measure and the one in Cogley (2002) can be seen as the output of a one-sided low-pass filter applied to current and past inflation.

The Figure 15 Panel (a) plots the R^2 statistics from these regressions against each forecast lead time h on the horizontal axis, whereas panel (b) plots the efficiency of the PBD decomposition measured by

$$G(h) = 100 \times \left(1 - \frac{\widehat{MSFE}(h)_{OTT}}{\widehat{MSFE}(h)_{TC}}\right)$$

i.e. percent gain in forecast accuracy arising from our suggested decomposition compared to alternative TC (trend-cycle) measures. Candidates that account for a greater percentage of subsequent changes in inflation filter out more transient variation and are preferred to those that account for less.

[Insert Figures 15 and 16 about here.]

[17] We do not consider in our analysis measures such as the median and trimmed mean price change among CPI components suggested by Bryan and Cecchetti, since, similar to Cogley (2002), we found that in regression-based combinations these candidate measures were dominated by the exponentially smoothed measure.

[18] According to Bryan and Cecchetti (1994) "core inflation" is defined as "the component of price changes that is expected to persist over medium-run horizons of several years".

Whereas at short horizon we confirm the evidence that the transitory component has a large predictive power for the next change in quarterly inflation, with an $R^2 = 0.22$, our decomposition is a superior predictor also of subsequent changes in inflation thanks to the marginal predictive power of each of the locally mean-reverting components of inflation.

Note that for forecast horizons of one to four quarters, the fit of our decomposition is almost the same as the one delivered by the measure suggested in Cogley (2002). The reason is explained in Figure 17 where we plot quarterly data on CPI inflation (shown as a solid black line), along with the candidate measure of core inflation, π_{mt}, suggested in Cogley (2002) and the sum of the components with level of persistence $j \geq 5$, $\sum_{j=5}^{J} x_{t,j}$. The correlation between the measure of core inflation and our sum of cyclical components is about 85%. When we forecast future changes over horizon of $1-4$ quarters using our cyclical components, we find that only the first 4 components are significant at standard levels. Hence our explanatory regressor is

$$\sum_{j=1}^{4} \hat{\beta}_j x_{t,j}$$

On the other hand when we forecast future changes using the core deviation, our regressor is

$$\pi_t - \pi_{mt} \approx \beta_h \sum_{j=1}^{4} x_{t,j}$$

Overall the Cogley (2002) is a restricted version of our regression where the restriction amounts to

$$\beta_j = \beta_h \;\; \forall j = 1, \ldots, 4$$

However at longer horizons, our decomposition is the most informative with R^2 statistics between 30 and 50 percent.

[Insert Figure 17 about here.]

7 Conclusions

A The abstract formulation of the Wold Decomposition.

This section provides an orthogonal decomposition of a time series in terms of uncorrelated innovations which are classified by two indices. The first one accounts for its level of persistence, the second one accounts for the time at which the shock impinges the economy.

The introduction of this decomposition requires the formulation of Wold theorem in an abstract Hilbert space, as provided in Sz. Nagy and Foias (Sz.-Nagy and Foias (1970)). This theorem shows that the Wold decomposition can be formulated in terms of properties which hold in a generic Hilbert space and its validity extends beyond the traditional probabilistic model adopted in conventional time series analysis.

For this reason let us consider an abstract Hilbert space H and let $\mathcal{L}(H)$ denote the set of bounded linear operators from H to itself.

Definition 16 *The vector space H is said to be the direct sum of its subspaces H_1 and H_2, and we write*

$$H = H_1 \oplus H_2$$

if $H = H_1 + H_2$ and $H_1 \cap H_2 = 0$

As a consequence, let H_1 and H_2 be subspaces of H. Then $H = H_1 \oplus H_2$ if and only if for every $h \in H$ there exist unique vectors $h_1 \in H_1$ and $h_2 \in H_2$ such that $h = h_1 + h_2$.

Definition 17 *Let $F \in \mathcal{L}(H)$. An* ***invariant*** subspace of F is a subspace V of H such that $F(V)$ is contained in V. An invariant subspace of F is also said to be F-invariant. If V is F-invariant, we can restrict F to V to arrive at a new linear mapping

$$F|_V : V \to V$$

Definition 18 *Let $H_1 \subset H$ and $H_2 \subset H$ be two Hilbert spaces. Let $\langle\ ,\ \rangle_1$ be the scalar product in H_1 and $\langle\ ,\ \rangle_2$ the scalar product in H_2. Then the operator $F : H_1 \to H_2$ with domain H_1 and image H_2 is isometric if*

$$\langle Fx, Fy \rangle_2 = \langle x, y \rangle_1$$

for any $x, y \in H_1$.

Definition 19 *Given a densely defined linear operator F on H we define its adjoint F^* as follows:*

- *The domain of F^* consists of elements $x \in H$ such that*

$$y \mapsto \langle x, Fy \rangle$$

  is a continuous linear functional. By continuity and density of the domain of F, it extends to a unique continuous linear functional on all of H.

- *if x is in the domain of F^* then we have*[19]

$$\langle F^* x, y \rangle = \langle x, Fy \rangle$$

  for all y in the domain of F.

Definition 20 *Let $F \in \mathcal{L}(H)$ and denote F its adjoint. Then F is said to be unitary if it satisfies*

$$F^* F = F F^* = \mathbb{I}$$

and $\mathbb{I} : H \to H$ is the identity operator.

Note that a unitary operator in H is a special case of an isometric operator for which $H_1 = H_2 = H$. In fact an isometric operator can be non invertible.

Let $F \in \mathcal{L}(H)$ be an isometry and define

$$H^{(\infty)} \equiv \cap_n F^n H$$

This is invariant under both F and F^*. Now we can define

$$H_S \equiv H \ominus H^{(\infty)}$$

This is invariant under F.

Definition 21 *Define the **detail subspace*** [20] with respect to F as

$$\mathcal{W} \equiv H_S \ominus F H_S$$

[19] This follows thanks to the Riesz representation theorem for linear functionals.
[20] Sometimes it is also known as **wandering subspace** .

Note that

$$\mathcal{W} = (im\ F|_{H_S})^{\perp} = ker\ F^*$$

where the second equality comes from the relationship between the image of F and the kernel of its adjoint[21]. We then obtain the following decomposition for H_S:

$$H_S = \mathcal{W} \oplus F\mathcal{W} \oplus F^2\mathcal{W} \oplus \ldots = \bigoplus_{n=0}^{\infty} F^n\mathcal{W}$$

We are now ready to present the following general decomposition:

Theorem 22 (Wold decomposition, see Sz.-Nagy and Foias (1970) [Theorem 1.1, page 3]) *Let F be an isometry on an Hilbert space H. Then H can be decomposed into an orthogonal sum*

$$H = H_S \oplus H^{(\infty)}$$

of F–invariant subspaces such that the restriction of F on $H^{(\infty)}$ is unitary and the restriction of F on H_S is a unilateral shift. More precisely, for $\mathcal{W} = (im\ F|_{H_S})^{\perp} \subset H_S$ one has[22]

$$F^n\mathcal{W} \perp F^m\mathcal{W}$$

for all $n \neq m$ with $n, m \in \mathbb{N}$ and $H_S = \bigoplus_{n=0}^{\infty} F^n\mathcal{W}$ and we can rewrite

$$H = \bigoplus_{n=0}^{\infty} F^n\mathcal{W} \oplus H^{(\infty)}$$

where $\bigoplus_{n=0}^{\infty} F^n\mathcal{W}$ is called the shift part.

Note that the shift part is present if and only if $ker\ F^* \neq \{0\}$. Moreover if $H^{(\infty)} = \{0\}$ then $H = \bigoplus_{n=0}^{\infty} F^n\mathcal{W}$ and this is why sometimes $\mathcal{W}$ is called the **generating** wandering subspace for the operator F.

[21] Indeed take $x \in ker\ F^*$. Then

$$F^*x = 0 \Leftrightarrow \langle F^*x, y\rangle = 0\ \forall y \Leftrightarrow \langle x, Fy\rangle = 0\ \forall y \Leftrightarrow x \perp im\ F$$

[22] This property gives the reason for calling $\mathcal{W}$ the **wandering** subspace.

A.1 The traditional realization of the Wold Decomposition

The formulation of the Wold decomposition introduced in the previous section applies in a generic Hilbert space. Following standard mathematical convention [23], it has been called "abstract" Wold decomposition to make clear that its validity is not restricted to the standard framework which is usually considered in the econometrics of time series analysis. In this section we illustrate the relation between the abstract Wold decomposition and the classical one.

Repeating the considerations of Brockwell and Davis (2002) we can introduce a "Universal Probability Space" assuming that the series of observations possibly starts at $t \to -\infty$ and is endless. Hence the state space becomes $\Omega = l_2(\mathbb{Z})$ the filtration becomes an increasing sequence of σ algebras $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. Standard Wold decomposition applies to wide sense stationary sequences, i.e. the probability measure $\mathbb{P}$ on the sequence $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}}$ is such that:

$$\begin{aligned} E^{\mathbb{P}}[X_{t-k+1}] &= 0 \\ E^{\mathbb{P}}\left[X_t^2\right] &= \gamma(0) < +\infty, \\ E^{\mathbb{P}}[X_t X_{t-k+1}] &= \gamma(k-1) < +\infty, \qquad k \in \mathbb{N} \end{aligned}$$

Now we associate an Hilbert space $\mathcal{H}_\infty^\gamma(\mathbf{X})$ to each wide sense stationary sequence, it is the Hilbert space of the square integrable linear combinations of the elements of the sequence $\mathbf{X}$ as follows:

$$\mathcal{H}_\infty^\gamma(\mathbf{X}) = \left\{ Y = \sum_{n\in\mathbb{Z}} \alpha_n X_n \mid \sum_{n,m\in\mathbb{Z}} \alpha_n \gamma(n-m)\alpha_m < +\infty, \quad \left\langle Y^1, Y^2 \right\rangle = \sum_{n,m\in\mathbb{Z}} \alpha_n^1 \gamma(n-m)\alpha_m^2 \right\}$$

We will denote with $\mathcal{H}_t^\gamma(\mathbf{X})$ the subspace of sequences observed up to time t, i.e. such that $\alpha_\tau = 0$ if $\tau > t$. We can define

Definition 23 *The linear prediction $LP(Y \mid \mathcal{K})$ of a generic element $Y \in \mathcal{H}_\infty(\mathbf{X})$ with respect to the subspace $\mathcal{K} \subset \mathcal{H}_\infty(\mathbf{X})$ is given by the orthogonal projection, of Y on $\mathcal{K}$:*

$$LP_\gamma(Y \mid \mathcal{K}) \equiv \arg\min_{Z\in\mathcal{K}} \langle Z - Y, Z - Y \rangle_{\mathcal{H}_\infty^\gamma(\mathbf{X})}^{1/2}$$

[23] The word "abstract" refers to those properties and statements whose validity does not rely in the specific choice of the Hilbert space basis (the so called Hilbert space realization) but on intrinsic (defining) properties of the Hilbert space itself.

We remark that for a given sequence of observations truncated at n, $\mathbf{X}_n \equiv \{X_k\}_{k\leq n}$ then the best linear predictor of Y is a linear combination of the elements of $\mathbf{X}_n$.[24]

Within the Hilbert space framework introduced above, the traditional Wold decomposition for time series can be obtained as a straightforward application of the abstract Wold theorem. Consider the linear span of all the observations up to time t and associate to it the Hilbert space $\mathcal{H}_t^\gamma(\mathbf{X}) = \left\{Y = \sum_{k\in\mathbb{N}} \alpha_{t-k} X_{t-k},\, Y \in \mathcal{H}_\infty^\gamma(\mathbf{X})\right\}$ and define the lag operator L as the linear operator on the sequence of observations up to time t, $\mathcal{H}_t^\gamma(\mathbf{X})$:

$$
\begin{aligned}
L \quad &: \quad \mathcal{H}_t^\gamma(\mathbf{X}) \to \mathcal{H}_t^\gamma(\mathbf{X}) \\
\mathbf{X}_t \quad &= \quad \{X_{t-k+1}\}_{k\in\mathbb{N}} \to L\mathbf{X} = \{X_{t-k}\}_{k\in\mathbb{N}}
\end{aligned}
$$

The lag operator L is an isometry in $\mathcal{H}_t(\mathbf{X})$ since $\langle LZ, LZ\rangle_{\mathcal{H}_t^\gamma(\mathbf{X})} = \langle Z, Z\rangle_{\mathcal{H}_t^\gamma(\mathbf{X})}$ and the sequence is wide sense stationary and infinite. Hence L satisfies all the hypotheses required by the abstract Wold theorem and we can conclude that:

$$
\mathcal{H}_t^\gamma(\mathbf{X}) = \bigoplus_{n=0}^{\infty} L^n \mathcal{W} \oplus \mathcal{H}^{\mathrm{det}}
$$

In particular the wandering subspace is defined by $\mathcal{W} = (Im\ L)^\perp$ hence: $\mathcal{W} = \mathcal{H}_t(\mathbf{X}) \ominus L\mathcal{H}_t(\mathbf{X})$ and corresponds to the space spanned by the t-th innovation: $\epsilon_t = X_t - LP^\gamma\left(X_t\,|L\mathcal{H}_t^\gamma(\mathbf{X})\right)$. Hence the iterated application of the lag operator defines the linear projection of the sequence $\mathbf{X}$ on the subspaces of orthogonal innovations:

$$
\begin{aligned}
P_{\bigoplus_{k=0}^{\infty} L^k\mathcal{W}} X_t \quad &= \quad \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \\
\psi_k \quad &= \quad \frac{1}{\sigma^2}\langle X_t, \epsilon_{t-k}\rangle_{\mathcal{H}_t^\gamma(\mathbf{X})} \\
\sigma^2 \quad &= \quad \langle \epsilon_t, \epsilon_t\rangle_{\mathcal{H}_t^\gamma(\mathbf{X})}
\end{aligned}
$$

Note that wide sense stationarity grants that ψ_k do not depend on index t and by definition of the Hilbert space the sequence $\{\psi_k\}_{k\in\mathbb{N}}$ is square integrable. The space $\mathcal{H}^{\mathrm{det}}$ is the deterministic component which is orthogonal to all past innovations defined by

$$
\nu_t = X_t - \sum_{k=0}^{+\infty} \psi_k \epsilon_{t-k}
$$

hence we obtain the well known Wold decomposition for weakly stationary time series, i.e.

[24]Hence in general it will differ from the conditional expectation.

Theorem 24 (Wold decomposition for Weakly Stationary Time Series) *If* $\{X_{t-k}\}_{k=0,..+\infty}$ *is a weakly stationary and mean zero process, then it can be expressed as*

$$X_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} + \nu_t$$

with

- $\epsilon_k \sim WN(0, \sigma^2)$
- $\nu_t \in \mathcal{H}^{\text{det}} \equiv \cap_{n \in \mathbb{N}} L^n \mathcal{H}_t^{\gamma} (\mathbf{X})$ *and it is deterministic,*
- $\sum_{k=0}^{\infty} \psi_k^2 < \infty$

Throughout this paper the deterministic component is assumed to equal a constant $\nu_t = \mu$.

B The Wold Decompostion of the scaling operator isometry and Proof of Theorem 6

In this section Theorem 6 is proved. The proof is based on the application of the abstract Wold theorem. Consider the Hilbert spaces

$$
\begin{aligned}
\mathcal{H}_{\infty}(\mathbf{x}) &= \left\{ Z = \sum_{n\in\mathbb{Z}} \alpha_n x_n \mid \sum_{n\in\mathbb{Z}} |\alpha_n|^2 < +\infty, \quad \left\langle Z^1, Z^2 \right\rangle = \sum_{n\in\mathbb{Z}} \alpha_n^1 \alpha_n^2 \right\} \\
\mathcal{H}_{t}(\mathbf{x}) &= \left\{ Z = \sum_{k\in\mathbb{N}} \alpha_{t-k} x_{t-k}, \; \in \mathcal{H}^2_{\infty}(\mathbf{x}) \right\}
\end{aligned}
$$

where $\mathcal{H}_{\infty}(\mathbf{x})$ is the Hilbert space of square integrable linear combinations of elements of the time series $\mathbf{x}_t$ obtained by considering the observations of the time series $\mathbf{x}$ truncated at time t, with the standard scalar product inherited from $l^2(\mathbb{Z})$. Without loss of generality we can assume $\mu = 0$. The persistence based decomposition of x_t in terms of a linear combination of uncorrelated innovations is produced applying the abstract Wold theorem to a different isometric operator. The new isometry operator is given by the composition of the following operators:

Definition 25 *The mean operator M centered at time t is defined by:*

$$
\begin{aligned}
&M : \mathcal{H}_t(\mathbf{x}) \to \mathcal{H}_t(\mathbf{x}) \\
&\mathbf{x} = \{x_{t-k}\}_{k\in\mathbb{N}} \to M\mathbf{x} = \{\widetilde{x}_{t-k}\}_{k\in\mathbb{N}}, \\
&\widetilde{x}_{t-k} = \frac{x_{t-k} + x_{t-k-1}}{2}
\end{aligned}
$$

the dyadic dilation operator centered at time t is defined by:

$$
\begin{aligned}
&D : \mathcal{H}_t(\mathbf{x}) \to \mathcal{H}_t(\mathbf{x}) \\
&\widetilde{\mathbf{x}} = \{\widetilde{x}_{t-k}\}_{k\in\mathbb{N}} \to \mathbf{x}^{(1)} = \left\{ x^{(1)}_{t-k} \right\}_{k\in\mathbb{N}} = D\widetilde{\mathbf{x}} \\
&x^{(1)}_{t-k} = \sqrt{2}\,\widetilde{x}_{t-2k}
\end{aligned}
$$

The rescaling operator R centered in t is defined by:

$$R = D \circ M$$

i.e. as the composition of the dyadic dilation operator D with the averaging one M.

The rescaling operator R defines a Wold decomposition of the Hilbert space $\mathcal{H}_t(\mathbf{x})$ in fact the following lemma holds:

Lemma 1 (Isometry Lemma) *The rescaling operator is isometric in $\mathcal{H}_t(\mathbf{x})$.*

Proof. The rescaling operator R can be represented as the composition of two operators: a rotation matrix *Rot* obtained as the direct sum of two by two rotation matrices:

$$Rot : \bigoplus_{k=0}^{+\infty} \begin{bmatrix} \frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and a projection on the odd components

$$P^{Odd} : \{\alpha_{-n}\}_{n\in\mathbb{N}} \to \{\alpha_{-2n-1}\}_{n\in\mathbb{N}}$$

Their joint action produces

$$P^{Odd} \circ Rot : (\alpha_t, \alpha_{t-1}, \alpha_{t-2}....) \to \left(\frac{\alpha_t + \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_{t-2} + \alpha_{t-3}}{\sqrt{2}}.....\right)$$

$$Rot : (\alpha_t, \alpha_{t-1}, \alpha_{t-2}....) \to \left(\frac{\alpha_t + \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_t - \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_{t-2} + \alpha_{t-3}}{\sqrt{2}}, ...\right)$$

$$P^{Odd} : \left(\frac{\alpha_t + \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_t - \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_{t-2} + \alpha_{t-3}}{\sqrt{2}}.....\right) \to \left(\frac{\alpha_t + \alpha_{t-1}}{\sqrt{2}}, \frac{\alpha_{t-2} + \alpha_{t-3}}{\sqrt{2}}.....\right)$$

and this implies that:

$$R = P^{Odd} \circ Rot$$

Direct verification proves that:

$$\langle R\mathbf{x}, R\mathbf{x} \rangle_{\mathcal{H}_t(\mathbf{x})} = \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}_t(\mathbf{x})},$$

in fact by definition:

$$\left\| \frac{x_{t-2k} + x_{t-2k-1}}{\sqrt{2}} \right\|_{\mathcal{H}_t(\alpha)} = \|x_{t-k}\|_{\mathcal{H}_t(\mathbf{x})} \quad \forall k \in \mathbb{N}$$

As a consequence the abstract Wold theorem admits a realization in $\mathcal{H}_t(\mathbf{x})$ where the rescaling operator R acts isometrically.

Then application of the abstract Wold theorem grants the existence of the following decomposition:

$$\mathcal{H}_t(\mathbf{x}) = \bigoplus_{j=1}^{+\infty} R^j \mathcal{W}^R \oplus \mathcal{H}_R^{(\infty)}$$

where

$$\mathcal{W}^R = \mathcal{H}_t(\mathbf{x}) \ominus R\mathcal{H}_t(\mathbf{x})$$

The generic element of the "innovation " subspace in the new decomposition is given by:

$$\delta_t^{(1)} \propto x_t - LP_{Rx_t}(x_t)$$

it corresponds to the detail which is removed by dyadic averaging. The detail $\delta_t^{(1)}$ has to verify the following conditions:

$$\begin{aligned} x_t &= LP_{\left\{Rx_t, \delta_t^{(1)}\right\}}(x_t) \\ 0 &= \left\langle \delta_t^{(1)}, Rx_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \end{aligned}$$

and for an arbitrary k all the constraints are satisfied by:

$$\delta_{t-k}^{(1)} = \frac{x_{t-k} - x_{t-k-1}}{2} \tag{B.1}$$

The detail subspace defines the information of the time series which is "removed" by aggregation or equivalently the "innovation" which is discovered when the resolution is increased. Note that the "normalized innovation" would be $\sqrt{2}\delta_{t-2k}^{(1)}$. The detail subspace for $\delta_{t-4k}^{(2)}$ is determined by the conditions

$$\begin{aligned} x_t &= LP_{\left\{\delta_t^{(1)}, \delta_t^{(2)}, R^2 x_t\right\}}(x_t) \\ 0 &= \left\langle \delta_t^{(1)}, Rx_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \\ 0 &= \left\langle \delta_t^{(2)}, R^2 x_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \end{aligned}$$

hence

$$\delta_t^{(2)} \propto \pi_t^{(1)} - LP_{(R^2 x)_t}(x_t)$$

and a possible choice is given by:

$$\delta_t^{(2)} = \pi_t^{(1)} - \frac{\pi_t^{(1)} + \pi_{t-2}^{(1)}}{2} = \frac{\pi_t^{(1)} - \pi_{t-2}^{(1)}}{2}$$

The $j-th$ iteration of the above construction implies that at resolution j the (unnormalized) detail, $\delta_t^{(j)}$ is given by:

$$\begin{aligned} \delta_{t-2^j k}^{(j)} &= \left(\frac{\left(R^j x\right)_{t-2^j k}}{\sqrt{2^j}} - \frac{\left(R^{j+1} x\right)_{t-2^j k}}{\sqrt{2^{j+1}}} \right) \\ \delta_{t-2^j k}^{(j)} &= \pi_{t-2^j k}^{(j-1)} - \frac{\pi_{t-2^j k}^{(j-1)} + \pi_{t-2^j (k+1)}^{(j-1)}}{2} = \frac{\pi_{t-2^j k}^{(j-1)} - \pi_{t-2^j (k+1)}^{(j-1)}}{2} \end{aligned}$$

By the assertion of the abstract Wold theorem the space $\mathcal{H}_R^{(\infty)}$ is determined by

$$\mathcal{H}_R^{(\infty)} = \cap_{j=0,..,+\infty} R^j \mathcal{H}_t(\mathbf{x})$$

and can be identified by that component which is not removed by an arbitrary number of averaging operations. In particular applying the law of large numebrs we can identify this component as the constant time series μ.

In addition the CLT central limit theorem for stationary processes presented in Hall and Heyde (1980) Corollary 5.2 pg.135 implies that:

$$\lim_{J\to\infty} \sqrt{Var\left[\sqrt{2^J}\pi_t^{(J)}\right]} = \psi(1)$$

hence point 1 and 2 are proved.

Point 3 is easily verified observing that each detail time series is obtained by a recursive application of a finite difference filter to the original time series which is stationary and the polynomial which defines the $j-th$ detail component $2^{-j}\left(1-L^{2^j}\right)\left(\sum_{i=0}^{2^j-1} L^i\right)$ is the first difference of a stationary process and the corresponding spectral density at frequency 0. ■

C Proof of Theorem 7

Observe that the isometry property for R does not hold in $\mathcal{H}_t^{\gamma}(\mathbf{X})$. In fact the space $\mathcal{H}_t^{\gamma}(\mathbf{x})$, where the standard Wold decomposition is applied, is different from $\mathcal{H}_t(\mathbf{x})$ because these two spaces have different definitions of the metric form. The two metric definitions coincide only when $\mathbf{X}$ is a white noise sequence, in which case:

$$\begin{aligned} E^{\mathbb{P}}[x_{t-k+1}] &= 0 \\ E^{\mathbb{P}}\left[x_t^2\right] &= \gamma(0) < +\infty, \\ E^{\mathbb{P}}[x_t x_{t-k}] &= 0, \qquad k \neq 0 \end{aligned}$$

On the other hand, given any stationary purely nondeterministic sequence, standard Wold decomposition states that it can be represented as an infinite moving average acting on a sequence of white noise innovations, i.e. in the space $\mathcal{H}_t(\varepsilon)$. Hence it is quite natural to formulate the Persistence Based Decomposition for the innovations process ε_t and then analyze the time series $\mathbf{x}_t \in \mathcal{H}_t(\varepsilon)$. In this space there exists a natural basis of uncorrelated shocks $\left\{\varepsilon_{j,t-2^j k_j}\right\}_{j\geq 1, k_j\geq 0}$ which are defined by an application of the isometric Haar transform to the original sequence of innnovations: let $\{\varepsilon_{t-k}\}_{k\in\mathbb{N}}$, the original sequence of uncorrelated innovations. We are now ready to state the Proof of Theorem 7.

Proof. Consider the decomposition of the Hilbert space $\mathcal{H}_t^{\gamma}(\mathbf{x})$

$$\begin{aligned}
\mathcal{H}_t^{\gamma}(\mathbf{x}) &= \bigoplus_{n=0}^{\infty} L^n \mathcal{W} \\
\mathcal{W}^L &= \mathcal{H}_t(\mathbf{x}) \ominus L\mathcal{H}_t(\mathbf{x})
\end{aligned}$$

Define now the sequence of white noise innovations $\{\varepsilon_{t-k}\}$ such that:

$$\begin{aligned}
\mathcal{W}^L &= Sp\{\varepsilon_t\} \\
L^k\mathcal{W}^L &= Sp\{\varepsilon_{t-k}\}
\end{aligned}$$

and we observe that $\mathcal{H}_t^{\gamma}(\mathbf{x})$ is unitary equivalent to $\mathcal{H}_t(\varepsilon)$. Then, without loss of generality, we can show that an application of the PBD in $\mathcal{H}_t(\varepsilon)$ determines the decomposition:

$$\begin{aligned}
\mathcal{W}^L &: = \mathcal{H}_t(\varepsilon) \ominus L\mathcal{H}_t(\varepsilon) \\
\mathcal{H}_t(\varepsilon) &= \bigoplus_{k=0,..,+\infty} L^k\mathcal{W}^L \\
\mathcal{W}^R &: = \mathcal{H}_t(\varepsilon) \ominus R\mathcal{H}_t(\varepsilon) \\
\mathcal{H}_t(\varepsilon) &= \bigoplus_{j=1,..,+\infty} R^j\mathcal{W}^R \oplus \mathcal{H}_R^{(\infty)} \\
\mathcal{W}^{L,R} &: = \mathcal{W}^L \ominus R\left(\mathcal{W}^L\right) \\
\mathcal{H}_t(\varepsilon) &= \bigoplus_{j=1,..,+\infty} R^j\left[\mathcal{H}_t(\varepsilon) \ominus R\mathcal{H}_t(\varepsilon)\right] \oplus \mathcal{H}_R^{(\infty)} \qquad \text{(C.1)} \\
&= \bigoplus_{j=1,..,+\infty} R^j\left[\left(\bigoplus_{k=0,..,+\infty} L^k\mathcal{W}^L\right) \ominus R\left(\bigoplus_{k=0,..,+\infty} L^k\mathcal{W}^L\right)\right] \oplus \mathcal{H}_R^{(\infty)} \\
&= \bigoplus_{j=1,..,+\infty} R^j \bigoplus_{k=0,..,+\infty} \left[\left(L^k\mathcal{W}^L\right) \ominus R\left(L^k\mathcal{W}^L\right)\right] \oplus \mathcal{H}_R^{(\infty)} \\
&= \bigoplus_{\substack{j=1,..,+\infty \\ k=0,..,+\infty}} R^jL^k\left[\mathcal{W}^L \ominus R\left(\mathcal{W}^L\right)\right] \oplus \mathcal{H}_R^{(\infty)} \\
&= \bigoplus_{\substack{j=1,..,+\infty \\ k=0,..,+\infty}} R^jL^k\mathcal{W}^{L,R} \oplus \mathcal{H}_R^{(\infty)}
\end{aligned}$$

where $P^{j,k}$ is the orthogonal projection of x_t along $\mathcal{W}_{j,k} = R^jL^k\mathcal{W}^{L,R}$ is the stochastic trend component of x_t. Then if we define $\varepsilon_{j,t-2^jk}$ as the normalized innovation such that:

$$LP\left(\varepsilon_{j,t-2^jk}\right) = \mathcal{W}_{j,k}$$

Consider the standard Wold decomposition $x_t = \sum_{k=0,..,+\infty} \psi_k \varepsilon_{t-k}$ and define:

$$\varepsilon_{j,t-2^j k_j} := \sum_{k=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k} \varepsilon_{t-k}$$

then application of the PBD to the white noise innovation implies the extended Wold decomposition:

$$\begin{aligned} x_t &= \sum_{k,k'=0,..,+\infty} \psi_{k'} \left(\left(\mathcal{T}_{Haar}^{(\infty)}\right)^{-1} \left(\mathcal{T}_{Haar}^{(\infty)}\right)\right)_{k',k} \varepsilon_{t-k} \\ &= \sum_{j\geq 1} \sum_{k_j=0}^{+\infty} \left(\sum_{k'=1,..,+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k'} \psi_{k'} \right) \varepsilon_{j,t-2^j k_j} \end{aligned}$$

Recall that the BNdecomposition is determined by

$$\begin{aligned} \Delta y_t &= \mu + x_t = \mu + \psi(L)\varepsilon_t \\ &= \mu + \psi(1)\varepsilon_t + \left[\psi(L) - \psi(1)\right]\varepsilon_t \end{aligned} \tag{C.2}$$

Now in order to conclude it is enough to define

$$\begin{aligned} \widetilde{\delta}_t^{(j)} &= -\sum_{k_j=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)} \widetilde{\psi}\right)_{j,k_j} \varepsilon_{j,t-2^j k_j} \quad \varepsilon_{j,t-2^j k_j} = \sum_{k=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k} \varepsilon_{t-k} \\ \left(\mathcal{T}_{Haar}^{(\infty)} \widetilde{\psi}\right)_{j,k_j} &= \sum_{k=0}^{+\infty} \left(\mathcal{T}_{Haar}^{(\infty)}\right)_{j,k_j;k} \widetilde{\psi}_k, \qquad \widetilde{\psi}_k \equiv \left(\sum_{j>k} \psi_j\right) \\ z_t &= z_{t-1} + \varepsilon_t \\ \tilde{\pi}_t^{(\infty)} &= \mu \cdot t + \psi(1)\, z_t \end{aligned}$$

and conclude that

$$y_t - y_0 = \tilde{\pi}_t^{(\infty)} + \sum_{j=1}^{+\infty} \widetilde{\delta}_t^{(j)}$$

■

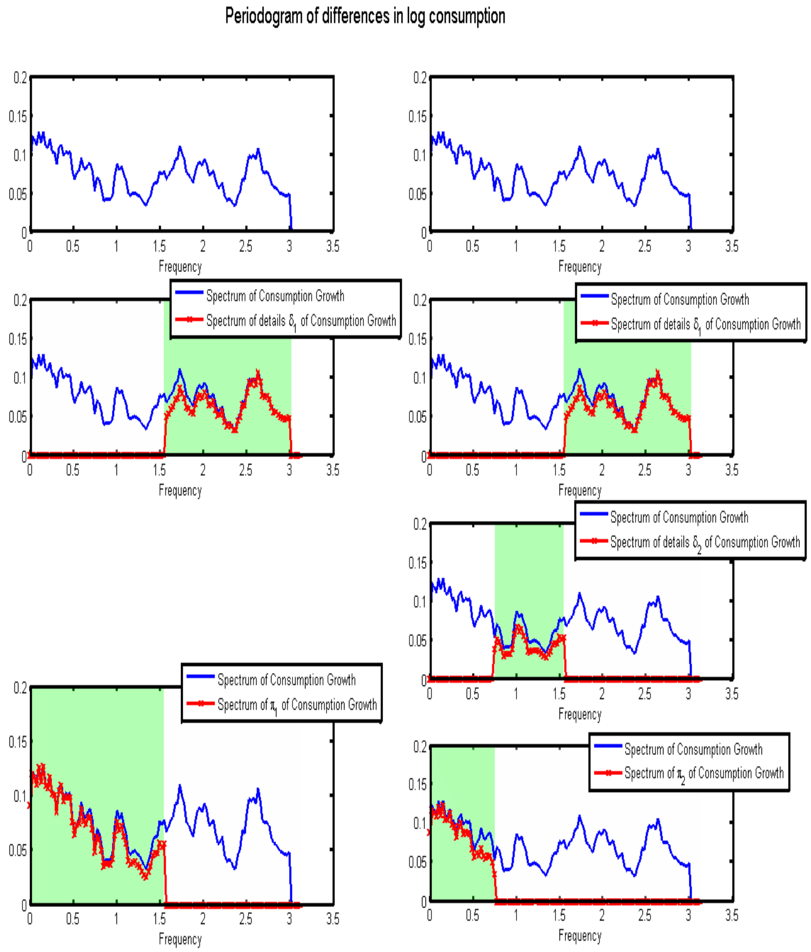


Figure 1: The figure displays the effects of the persistence based decomposition of the consumption time series applied up to level $J = 1$ (left panels) and $J = 2$ (right panels). In particular the top panels displays the smoothed periodogram the consumption process for the data. An equally weighted "nearest neighbor" kernel was used to perform the smoothing, equally weighting the 4 nearest frequencies. The bottom right panel displays the Fourier spectrum of the time series $\pi_t^{(2)}$ whereas the bottom left panel displays the Fourier spectrum of the time series $\pi_t^{(1)}$.

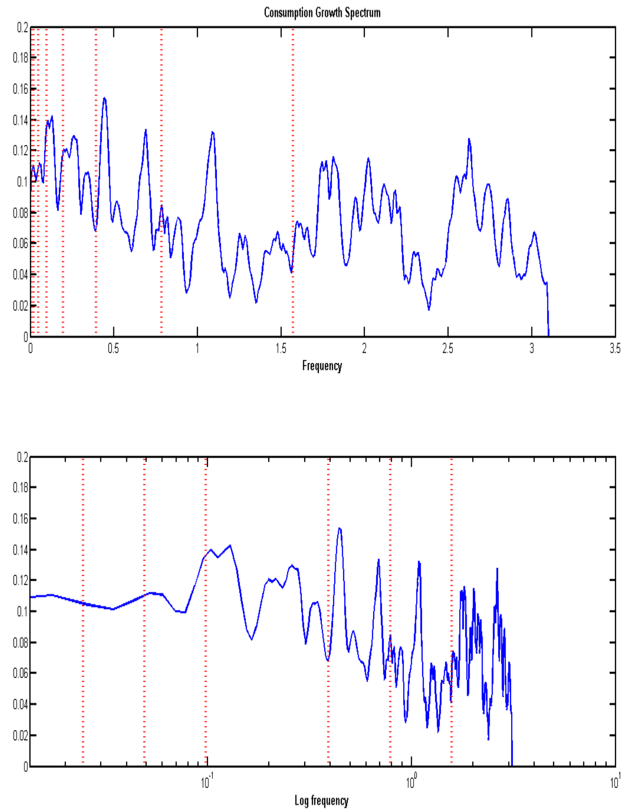


Figure 2: The figure displays the smoothed periodogram the consumption process for the data together with the intervals $\left[\frac{f_{max}}{2^j}\ \frac{f_{max}}{2^{j-1}}\right)$ $j = 1, \ldots, 8$. An equally weighted "nearest neighbor" kernel was used to perform the smoothing, equally weighting the 4 nearest frequencies. In the top panel linear scale is used for frequencies whereas in the bottom panel logarithmic scale is used for the X-axis.

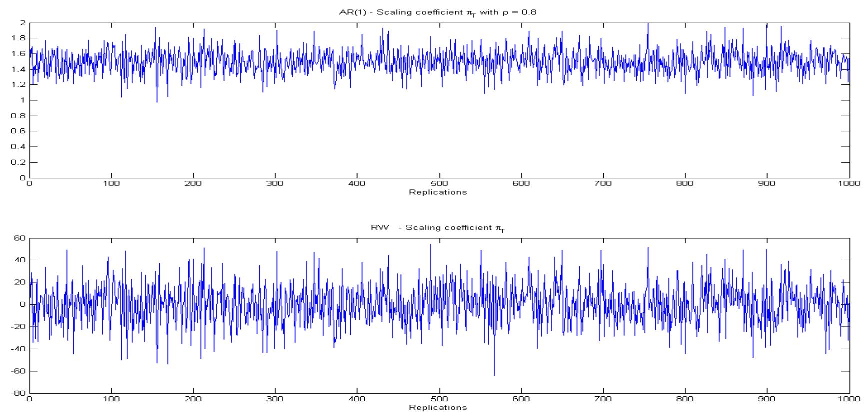


Figure 3: We plot the scaling coefficient for an AR(1) process with unconditional mean 1.5 and autoregressive coefficient equal to 0.8.

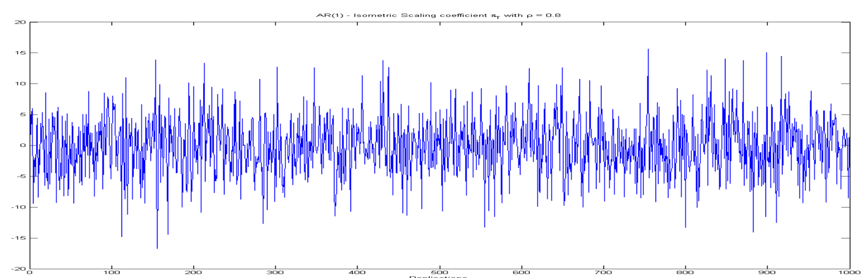


Figure 4: We plot the isometric scaling coefficient at time T for the AR(1) process described in the text.

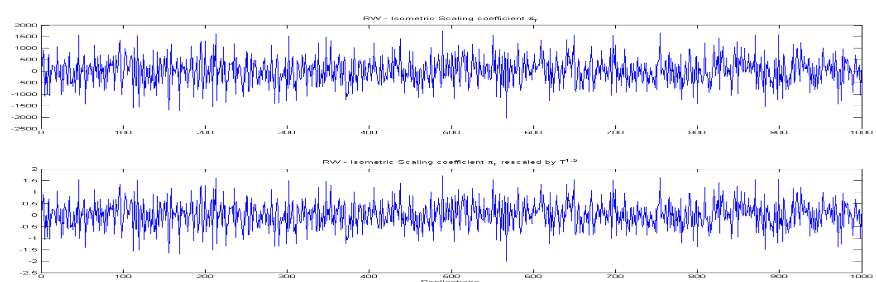
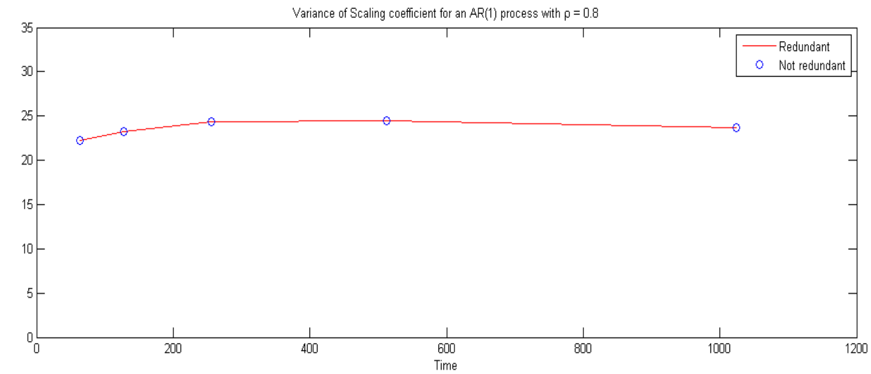


Figure 5: We plot the isometric scaling coefficient at time T for the Random Walk process described in the text.



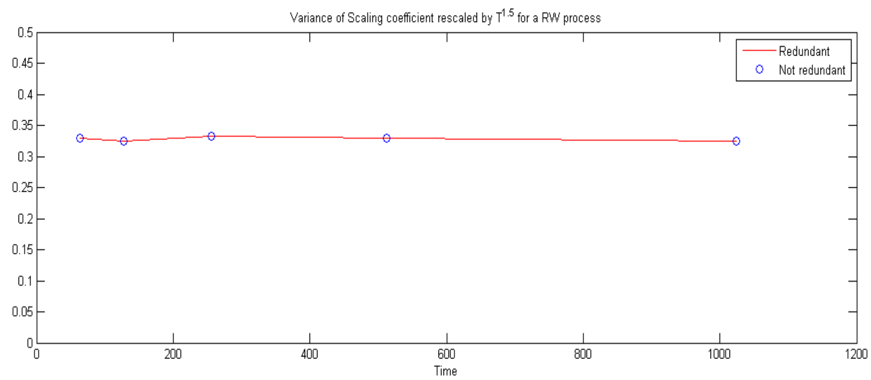


Figure 6: We plot the variance of the isometric scaling π_t for each time instant t for the AR(1) and the Random Walk process.

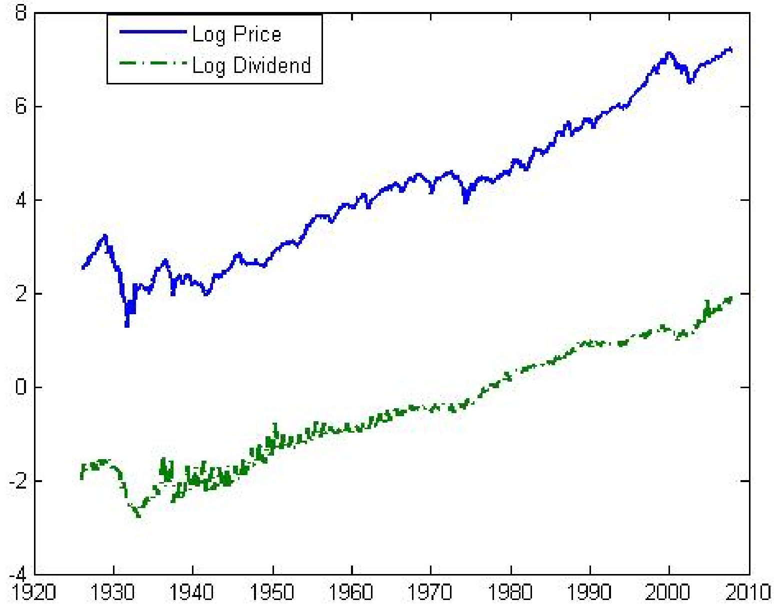


Figure 7: Log dividend and log price, 1926Q3-2007Q4.

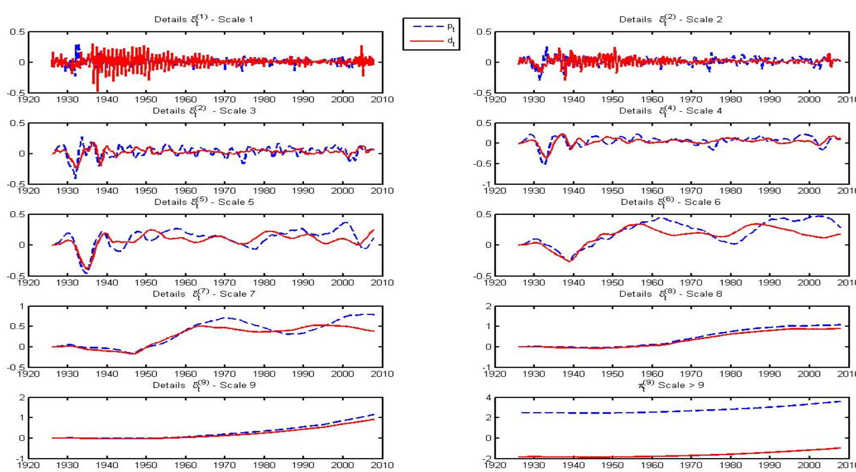


Figure 8: Time-scale decomposition for the log price p_t and log dividend d_t.

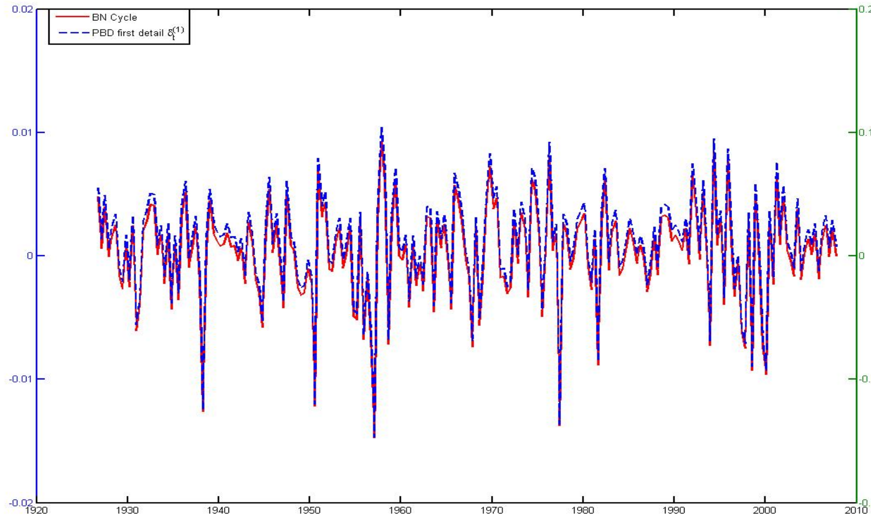


Figure 9: Comparison between the transitory component obtained through Beveridge–Nelson (BN) decomposition and through the Persistence Based Decomposition (PBD) for the log price p_t, i.e. $\delta_t^{(1)}$.

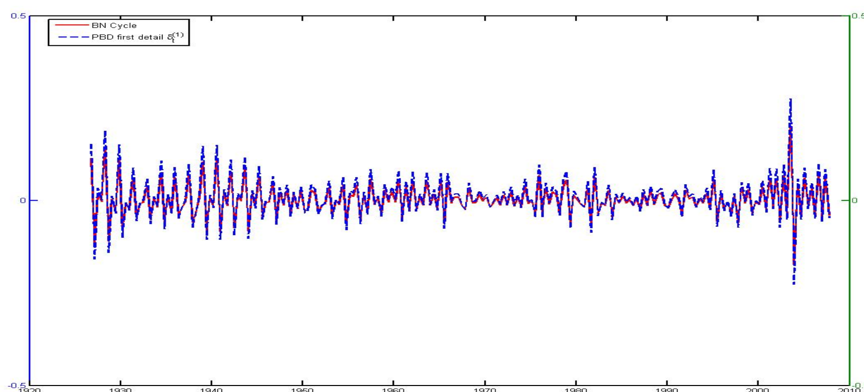


Figure 10: Comparison between the transitory component obtained through Beveridge–Nelson (BN) decomposition and through the Persistence Based Decomposition (PBD) for the log dividend d_t, i.e. $\delta_t^{(1)}$.

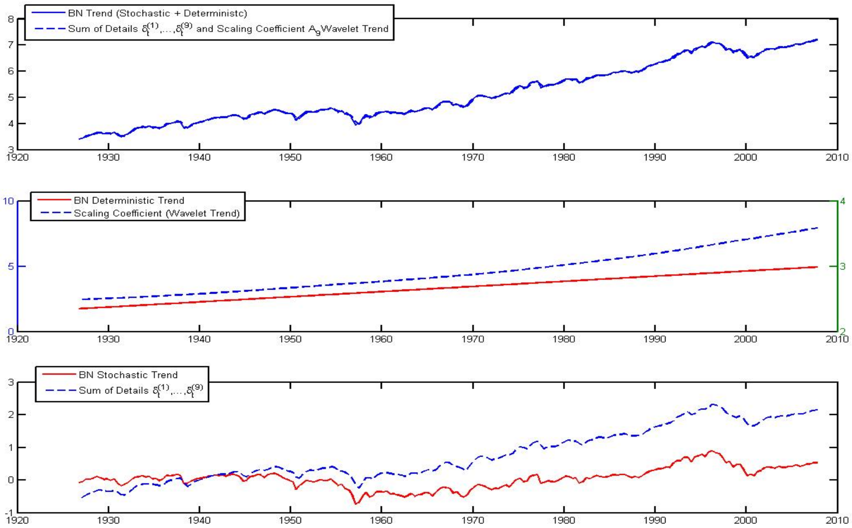


Figure 11: Comparison between the permanent component obtained through Beveridge–Nelson (BN) decomposition and through the Persistence Based Decomposition (PBD) for the log price p_t.

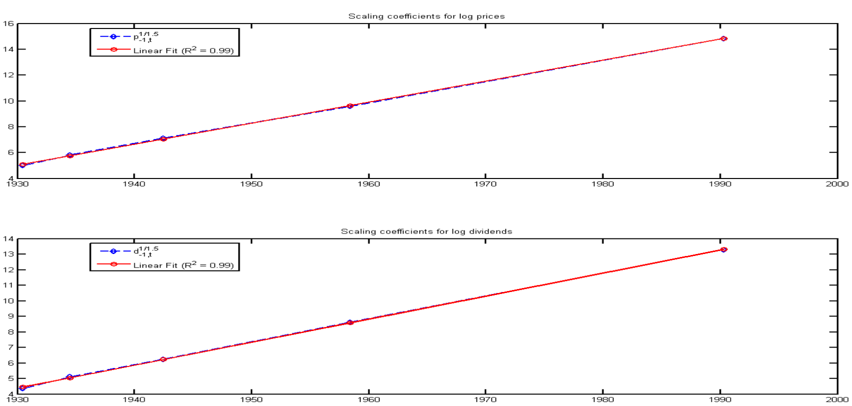
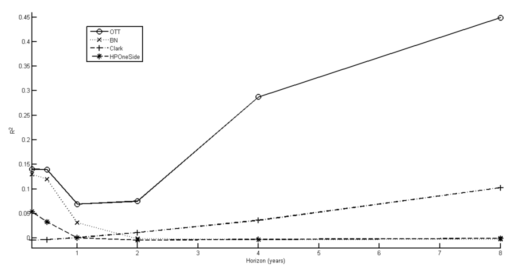
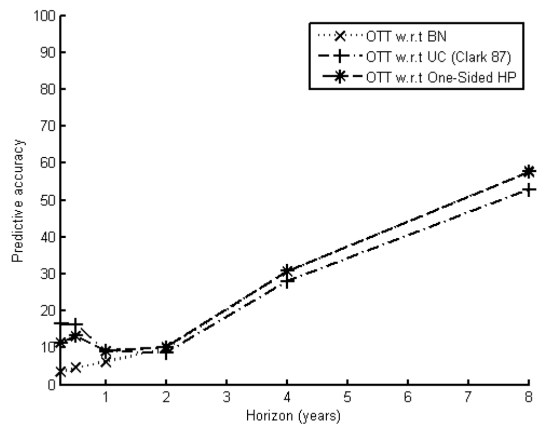


Figure 12: We plot the rescaled isometric scaling π_t for each time instant $t = 16, 32, 64, 128, 256$ quarters starting from 1927Q1 for the log prices (top panel) and the log dividends (bottom panel) process together with the best linear fit. The best linear fit for the dividends is given by $d_{-1,t} = 0.037x + 3.86$ whereas for the prices we have $p_{-1,t} = 0.041x + 4.44$.



(a)



(b)

Figure 13: U.S. real gross domestic product. Panel (a) shows the $\bar{R}^2$ statistics from univariate regressions versus forecast horizon h. Panel (b) shows the percent efficiency gain versus forecast horizon h.

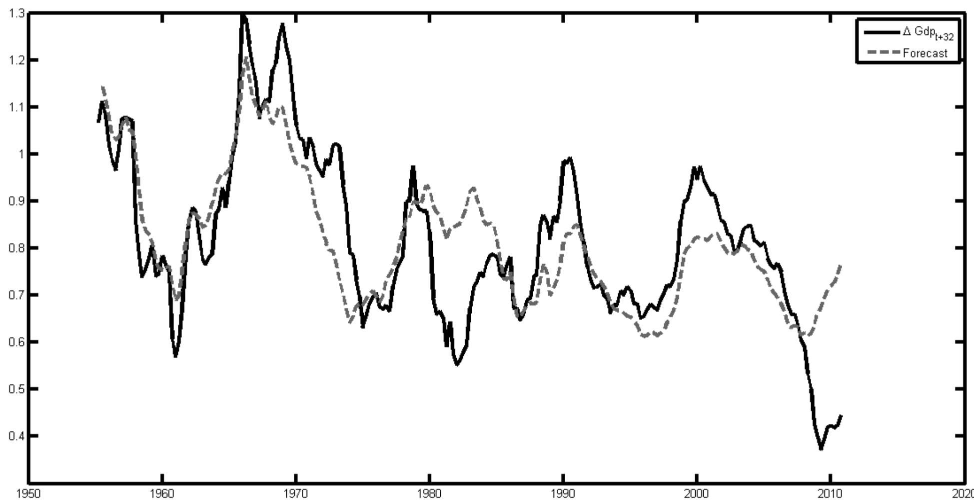
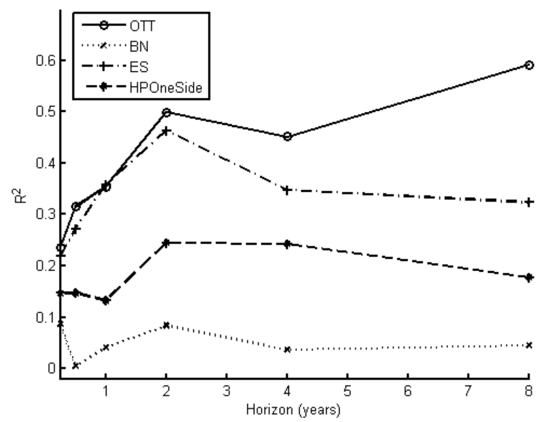
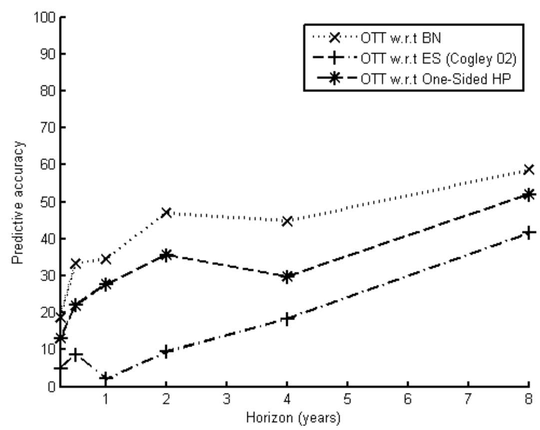


Figure 14: Realized h-changes in U.S. real gross domestic product along with (in-sample) fitted values of the forecast-based regression. The forecast horizon is $h = 32$ quarters.



(a)



(b)

Figure 15: U.S. inflation. Panel (a) shows the $\bar{R}^2$ statistics from univariate regressions versus forecast horizon h. Panel (b) shows the percent efficiency gain versus forecast horizon h.

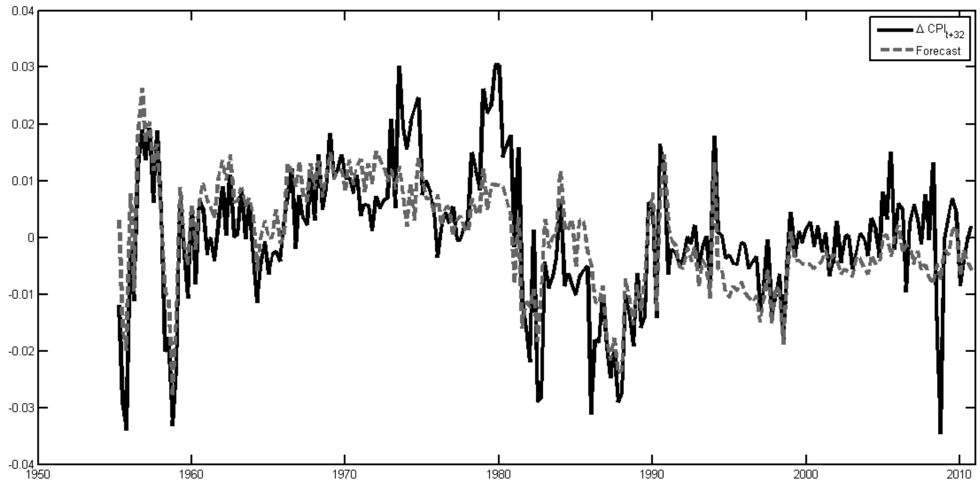


Figure 16: Realized h-changes in U.S. inflation along with (in-sample) fitted values obtained from the forecast-based regression. The forecast horizon is $h = 32$ quarters.

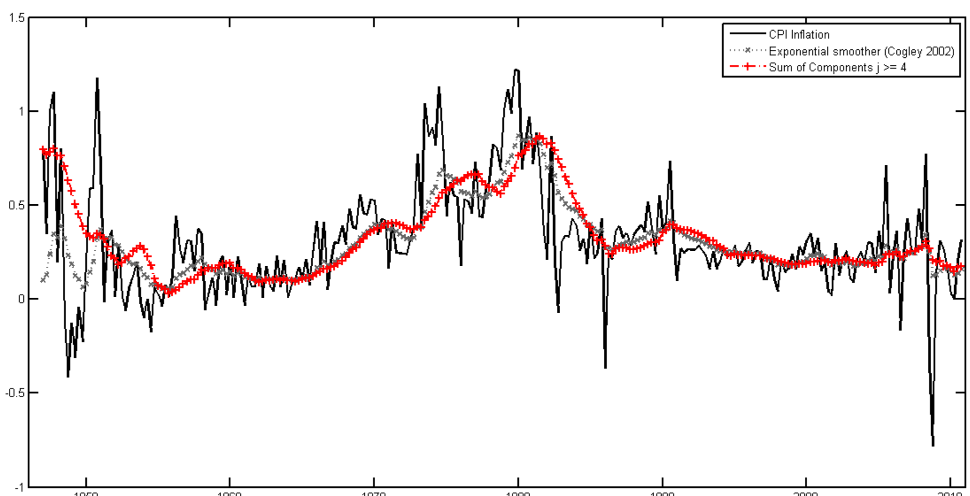


Figure 17: U.S. inflation.

Panel A: Final cycle estimates vs. Extended BN

C_1	0.568	2.564	0.574	0.186	0.591
	(3.62)	(1.29)	(3.64)	(1.758)	(4.34)
C_2	0.228	0.227	0.218	0.630	0.462
	(1.85)	(1.84)	(1.74)	(4.79)	(4.19)
C_3	-0.172	-0.174	-0.191	0.325	0.050
	(-2.32)	(-2.34)	(-2.22)	(1.382)	(0.73)
C_4	0.031	0.029	0.017	0.273	0.127
	(0.60)	(0.56)	(0.28)	(0.66)	(2.79)
C_5	-0.028	-0.031	-0.031	0.255	-0.020
	(-0.78)	(-0.85)	(-0.84)	(0.826)	(-0.664)
C_6	0.004	0.002	0.006	-0.154	-0.013
	(0.16)	(0.09)	(0.24)	(-0.500)	(-0.637)
C_7	-0.008	-0.008	-0.008	0.557	0.004
	(-0.87)	(-0.90)	(-0.895)	(1.83)	(0.489)
C_8					
$Beveridge - Nelson$		1.723			
		(1.01)			
$Clark$			0.341		
			(0.43)		
Hodrick-Prescott 1-sided				0.006	
				(0.062)	
Hodrick-Prescott 2-sided					-0.388
					(-9.06)
$\bar{R}^2$	0.15	0.15	0.15	0.15	0.35

Table 1: Predictive regressions for real GDP growth using lag of cycle estimates.

References

Bandi, Federico M. and Benot Perron, “Long-run risk-return trade-offs,” *Journal of Econometrics*, April 2008, *143* (2), 349–374.

Bansal, Ravi and Amir Yaron, “Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles,” *Journal of Finance*, 08 2004, *59* (4), 1481–1509.

Beveridge, Stephen and Charles R. Nelson, “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’,” *Journal of Monetary Economics*, 1981, *7* (2), 151–174.

Brockwell, P. J. and R. A. Davis, *Introduction to time series and forecasting*, second ed., Springer, 2002.

Bryan, Michael F. and Stephen G. Cecchetti, “The consumer price index as a measure of inflation,” *Economic Review*, 1993, (Q IV), 15–24.

—— **and** —— , “Measuring Core Inflation,” in “Monetary Policy” NBER Chapters, National Bureau of Economic Research, Inc, 1994, pp. 195–219.

—— **and** —— , “The seasonality of consumer prices,” *Economic Review*, 1995, (Q II), 12–23.

—— **,** —— **, and Rodney L. Wiggins II**, “Efficient Inflation Estimation,” NBER Working Papers 6183, National Bureau of Economic Research, Inc September 1997.

Campbell, John Y and Luis M Viceira, “The Term Structure of the Risk-Return Tradeoff,” CEPR Discussion Papers 4914, C.E.P.R. Discussion Papers February 2005.

Clark, Peter K, “The Cyclical Component of U.S. Economic Activity,” *The Quarterly Journal of Economics*, November 1987, *102* (4), 797–814.

Cochrane, John H., “How Big Is the Random Walk in GNP?,” *The Journal of Political Economy*, 1988, *96* (5), pp. 893–920.

Cogley, Timothy, “A Simple Adaptive Measure of Core Inflation,” *Journal of Money, Credit and Banking*, February 2002, *34* (1), 94–113.

Daubechies, I., “The wavelet transform, time–frequency localization and signal analysis,” *IEEE Trans. Info. Theory*, Sep. 1990, *36* (5), 961–1005.

Daubechies, Ingrid, *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*, Soc for Industrial & Applied Math, December 1992.

Gencay, Ramazan and Nikola Gradojevic, “Errors in Variables Estimation with No Instruments,” Working Paper Series wp30 09, Rimini Centre for Economic Analysis January 2009.

——— **and Yanqin Fan**, "Unit Root Tests with Wavelets," *Econometric Theory, Forthcoming*, 2008.

———**, Faruk Selcuk, and Brandon Whitcher**, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, September 2001.

Hall, P. and C.C. Heyde, *Martingale limit theory and its application*, Academic Press, June 1980.

Hansen, Lars Peter and Jose A Sheinkman, "Long-term Risk: An Operator Approach," *Econometrica*, November 2009, *77* (1), 177–234.

——— **and** ——— , "Pricing growth-rate risk," *Finance and Stochastics*, September 2009.

Harvey, A C, "Trends and Cycles in Macroeconomic Time Series," *Journal of Business & Economic Statistics*, June 1985, *3* (3), 216–27.

——— **and A Jaeger**, "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, July-Sept 1993, *8* (3), 231–47.

Hayashi, Fumio, *Econometrics*, Princeton University Press, December 2000.

Hodrick, Robert J and Edward C Prescott, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, February 1997, *29* (1), 1–16.

Mallat, Stephane G., "Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R)," September 1989, *315* (1).

——— , "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, *11*, 674–693.

Morley, James C., "The Two Interpretations Of The BeveridgeNelson Decomposition," *Macroeconomic Dynamics*, June 2011, *15* (03), 419–439.

———**, Charles R. Nelson, and Eric Zivot**, "Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different?," *The Review of Economics and Statistics*, 03 2003, *85* (2), 235–243.

Muller, Ulrich and Mark W. Watson, "Low-Frequency Robust Cointegration Testing," NBER Working Papers 15292, National Bureau of Economic Research, Inc August 2009.

Muller, Ulrich K. and Mark W. Watson, "Testing Models of Low-Frequency Variability," *Econometrica*, 09 2008, *76* (5), 979–1016.

Nelson, Charles R., "The Beveridge-Nelson decomposition in retrospect and prospect," *Journal of Econometrics*, October 2008, *146* (2), 202–206.

Oh, Kum Hwa, Eric Zivot, and Drew D. Creal, "The Relationship between the Beveridge-Nelson Decomposition andUnobserved Component Models with Correlated Shocks," Working Papers UWEC-2006-16-FC, University of Washington, Department of Economics 2006.

Ortu, Fulvio, Andrea Tamoni, and Claudio Tebaldi, "Long Run Risk and the Persistence of Consumption Shocks," November 2011.

Proietti, Tommaso, "Trend-Cycle Decompositions with Correlated Components," *Econometric Reviews*, 2006, *25* (1), 61–84.

Ramsey, J.B. and C. Lampart, "The Decomposition of Economic Relationships by Time Scale Using Wavelets," *Macroeconomic Dynamics*, September 1998, *2* (01).

Renaud, Olivier, Jean-Luc Starck, and Fionn Murtagh, "Wavelet-Based Combined Signal Filtering and Prediction," *IEEE Transactions SMC, Part B*, 2005, *35*, 1241 – 1251.

Sz.-Nagy, Bela and Ciprian Foias, *Harmonic analysis of operators on Hilbert space*, North-Holland Pub. Co., December 1970.

Watson, Mark W., "Univariate detrending methods with stochastic trends," *Journal of Monetary Economics*, July 1986, *18* (1), 49–75.

Wold, Herman, *A Study in the Analysis of Stationary Time Series*, Uppsala, Sweden: Almqvist and Wiksell, 1938.