A persistence-based Wold-type decomposition for stationary time series*

Fulvio Ortu† Federico Severino‡
Università Bocconi & IGIER Università Bocconi & USI

Andrea Tamoni§ Claudio Tebaldi¶
LSE Università Bocconi, IGIER & CAREFIN

July 30, 2019

*We thank T. Andersen, S. Cerreia-Vioglio, M. Chernov, F. Corsi, R. Gallant, L.P. Hansen, M. Henry, M. Marcellino, G. Primiceri, P. Reggiani, E. Renault, L. Sala, E. Sentana, P. Veronesi, and M. Watson, and participants at the 2017 NBER/NSF Time Series Conference at Kellogg, University of Chicago, 11th World Congress of the Econometric Society in Montréal, 8th International Conference on CFE in Pisa, Università degli Studi di Udine, 11th CSEF-IGIER Symposium on Economics and Institutions in Capri, XLI AMASES Meeting in Cagliari and 11^{th} Annual SoFiE Conference in Lugano for helpful suggestions.

†Università Bocconi, Department of Finance, Milan, 20136, Italy, e-mail: fulvio.ortu@unibocconi.it

‡Università Bocconi, Department of Finance, Milan, 20136, Italy, and USI Lugano, Department of Economics, Lugano, 6900, CH, e-mail: federico.severino@usi.ch

§London School of Economics, Department of Finance, London, WC2A 2AE, UK, e-mail: a.g.tamoni@lse.ac.uk

¶Università Bocconi, Department of Finance, Milan, 20136, Italy, e-mail: claudio.tebaldi@.unibocconi.it

A persistence-based Wold-type decomposition for stationary time series

Abstract

This paper shows how to decompose weakly stationary time series into the sum, across time scales, of uncorrelated components associated with different degrees of persistence. In particular, we provide an *Extended Wold Decomposition* based on an isometric *scaling operator* that makes averages of process innovations. Thanks to the uncorrelatedness of components, our representation of a time series naturally induces a persistence-based variance decomposition of any weakly stationary process. We provide two applications to show how the tools developed in this paper can shed new light on the determinants of the variability of economic and financial time series.

JEL classification: C18, C22, C50

Keywords: Wold decomposition, temporal aggregation, persistence heterogeneity, forecasting.

1 Introduction

This paper formalizes the idea that a stationary time series decomposes into the sum of orthogonal layers with heterogeneous levels of durations. Notably, our representation is obtained in the time-domain. Thus, although it shares many of the insights of frequency domain methods, our approach is well suited for predictive analysis where time adaptation, and localization in both time and frequency, are crucial.

Our main result shows how to represent a stationary time series as a sum of orthogonal components each one characterized by a specific level of persistence (or scale). Our novel representation produces an *Extended Wold Decomposition* in which each component at scale j is an infinite moving average driven by a sequence of uncorrelated innovations that are located on a time grid whose unit spacing is 2^j the original unit grid.

As a by-product of the orthogonality of components, our approach naturally provides a variance decomposition of the time series. Differently from the traditional principal component analysis (PCA) approach, each component in our analysis is associated with a specific scale, a fact that comes to aid when one has to provide an economic interpretation of the key driving factors. Indeed, different economic phenomena most likely operate at different frequencies: e.g., price movements can be related to announcements of macroeconomic conditions and corporate activity at the short-end of spectrum, to political cycles at the medium-end, to technological and demographic changes at the long-end, and to uncertainty regarding exhaustible energy resources and climate change at the very long-end (see e.g. the literature review in Ortu et al. 2013). Therefore, by being informative about the duration of the fluctuation most relevant for the time series variability, our approach sheds light on the potential economic mechanism driving the time series.

Our Extended Wold Decomposition also provides an interpretation of the notion of

persistence for stationary processes. Indeed, our construction exploits the isometry of a *scaling operator* that smooths the original time series by making successive averages of the process innovations. Such procedure implicitly defines an index j that we dub *scale* or *degree of persistence*. In fact, a shock at scale j has, by construction, a mean reversion time ranging between 2^{j-1} and 2^j, and its spectrum is localized in the corresponding band of frequencies.

Whereas our first main result (Theorem 1) shows how to decompose and analyze a time series in terms of its persistent components, our second main result (Theorem 2) shows how to build a stationary process starting from uncorrelated components with heterogeneous degrees of persistence. This second result is important as it offers a new family of (data generating) processes suitable to capture phenomena evolving over time scales of different length (or frequency). Interestingly, Robinson (1978) and Granger (1980) propose contemporaneous aggregation of individual, *random* coefficient AR(1) processes as a mechanism that leads to long memory. Relying on Theorem 2, the simulations in Section 2.2.1 show that a *finite* (in fact as small as two) sum of autoregressive processes with *fixed* autoregressive coefficients is able to generate hyperbolically decaying autocorrelation functions as long as the processes evolve over different time scales (e.g., monthly and annual). On the technical side, the proof of Theorem 2 highlights that the reconstructed time series is stationary because each scale-specific shock is assumed to be a weighted linear combination of underlying innovations.

We provide two applications to show the relevance of our decomposition.

In the first empirical application, we investigate the drivers of realized volatility in currency markets. We show that three components operating over semiannual, annual and biannual frequencies account for a substantial fraction of variability of daily realized variance. A forecasting model using only these three components performs well relative to the Heterogeneous Autoregressive model (HAR) model of Corsi (2009).

Interestingly, also the HAR model relies on three regressors - namely daily, weekly, and monthly realized volatility series - to forecast daily volatility. Both our model based on Wold components and the HAR model provide evidence that volatility reflects the aggregate impact of processes evolving over different time scales. However, our analysis further shows that the daily, weekly, and monthly components of Corsi (2009) indeed proxy for phenomena occurring over longer time scales in between half-year and two years. Moreover, if one adopts the "Heterogeneous Market Hypothesis" of Müller et al. (1993) where each volatility component is generated by the actions of specific types of market participants, then our analysis is informative about the horizons of such investors.

In the second application, we employ our Extended Wold Decomposition to investigate the relation between yields and bond returns. Our analysis shows that the Extended Wold Decomposition is able to detect yield cycles useful for bond predictability. We show that a model based on a high frequency component of the level and a low frequency component of the slope attains a fit similar to the Cieslak and Povala (2015) model. The latter detrends the level of the yield curve using an inflation trend. Despite the same fit, the two models call for different economic interpretation. In our analysis, the factors driving bond returns are orthogonal and hence economic theory needs to explain the high- and low-frequency cycles in the yield curve only. In the Cieslak and Povala (2015) framework, the factors are negatively correlated so that economic theory needs to explain not only the (detrended) level and slope, but also their relative value.

In order to clarify positioning, we discuss how our results relate to some existing approaches. Ortu et al. (2013) propose to use multiresolution analysis to study the asset prices reaction in dynamic economies hit by shocks of heterogeneous durations. The authors show that exposure to specific consumption components (rather than raw consumption) is a source of priced risk and it can explain the level of the equity

premium. Bandi, Perron, Tamoni and Tebaldi (2019) show that the empirical relation between future excess market returns and past economic uncertainty is hump-shaped, i.e. the R-squared (R^2) values reach their peak at 16 years and the structure of the R^2s, before and after, is roughly tent-shaped. These authors show that classical predictive systems imply restrictions across scales which are inconsistent with the hump-shaped relation between uncertainty and returns. To justify their findings formally, Bandi et al. (2019) introduce a novel modeling framework in which predictability is specified as a property of the components of excess market returns and economic uncertainty. The Extended Wold Decomposition formalized in this paper is used, in their framework, to model predictability across scales and justify extraction of the components. Finally, the Extended Wold Decomposition proves also useful in the context of cross-sectional asset pricing. Bandi, Lo, Chaudhuri and Tamoni (2019) employ the Extended Wold Decomposition to introduce novel spectral factor models in which systematic risk is allowed (without being forced) to vary across different frequencies. The authors show that spectral factors models lead to portfolios with lower out-of-sample variance relative to portfolios constructed under the assumption that risk is constant across frequencies.

The present paper shows that a stationary time series can be described as the sum of uncorrelated components with different levels of persistence. A component at level of persistence j admits a classical Wold decomposition in terms of innovations that are moving averages of the original innovations and are localized on a time grid whose unit is 2^j units of the original observation grid. The coarsening of the grid where low-frequency shocks are located is closely related to the frequency-domain procedure adopted by Müller and Watson (2008) and Müller and Watson (2015), where long-run sample information is isolated using a small number of low-frequency trigonometric weighted averages which, in turn, can be used to conduct inference about long-run variability and covariability. Differently from them, our resulting representation uses

information only up to time t and, thus, it does not suffer from the look-ahead bias typical of a two-sided frequency representation.

Our persistence-based Wold-type representation lays the foundation for processes that have been proposed in the literature to capture the long-range dependence exhibited by many financial time series, such the Heterogeneous Information Arrivals process suggested by Andersen and Bollerslev (1997), the Markov-Switching Multifractal work by Calvet and Fisher (2007), the Heterogeneous Autoregressive Model of Realized Volatility (HAR-RV) developed by Corsi (2009), the component-GARCH models of Engle and Lee (1999) and the Spline-Garch framework of Engle and Rangel (2008).

Our decomposition into components with different persistence levels is additive. This is in contrast with the MIDAS (Mi(xed) Da(ta) S(ampling)) regressions developed in Ghysels et al. (2004) and Ghysels et al. (2006), where the smoothing coefficients interact multiplicatively in a way that makes it difficult to isolate the effects of each components.

The rest of the paper is organized as follows. Section 2 presents the theoretical results of the paper. We start with a finite-dimensional example to illustrate the key steps needed to decompose a zero-mean, purely non-deterministic process into the sum of uncorrelated components associated with different levels of persistence. In subsection 2.1 we introduce our isometric scaling operator, we derive the Extended Wold Decomposition, and we analyze its main properties. We also look at our decomposition from the standpoint of spectral analysis and show that our operator acts as a low-pass filter. We devote Subsection 2.2 to establish the converse result, i.e. how to compute the moving average representation of a weakly stationary time series that is obtained by summing heterogeneous components at different scales. In that section we also illustrate the reconstruction theorem by means of an example. Subsection 2.3 describes the relation between the Extended Wold Decomposition and the

multiresolution-based decomposition proposed in Ortu et al. (2013). In particular, we contrast, for a fixed level of persistence, the spectrum of the components obtained from our scaling operator with that generated by multiresolution analysis. Section 3 describes two empirical applications of the Extended Wold Decomposition to realized volatility and bond yields. Section 4 concludes. All proofs are in the Online Supplement.

2 The Extended Wold Decomposition

To provide the intuition behind our Extended Wold Decomposition, we begin with a finite-dimensional example. Specifically, we consider the $MA(7)$ process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ defined by

$$x_t = \sum_{h=0}^{7} \alpha_h \varepsilon_{t-h}, \tag{1}$$

where $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a unit variance white noise and each α_h quantifies the impact on x_t of a unitary shock occurred h periods before.

An alternative representation of x_t obtains by averaging the innovations ε_t and taking the differences between subsequent shocks:

$$x_t = \sum_{k=0}^{3} \beta_k^{(1)}\ \varepsilon_{t-2k}^{(1)} + \sum_{k=0}^{3} \gamma_k^{(1)}\ \bar{\varepsilon}_{t-2k}^{(1)}, \tag{2}$$

where, for $k = 0, \ldots, 3$,

$$\varepsilon_{t-2k}^{(1)} = \frac{\varepsilon_{t-2k} - \varepsilon_{t-2k-1}}{\sqrt{2}}, \qquad \bar{\varepsilon}_{t-2k}^{(1)} = \frac{\varepsilon_{t-2k} + \varepsilon_{t-2k-1}}{\sqrt{2}}.$$

Each $\bar{\varepsilon}_{t-2k}^{(1)}$ can be interpreted as a low-frequency shock while each $\varepsilon_{t-2k}^{(1)}$ can be interpreted as a higher-frequency innovation. The coefficient $\gamma_k^{(1)}$ quantifies the sensitivity of x_t with respect to $\bar{\varepsilon}_{t-2k}^{(1)}$, while the coefficient $\beta_k^{(1)}$ captures the exposure to $\varepsilon_{t-2k}^{(1)}$. Since x_t features a unique representation with respect to the initial innovations ε_t,

the coefficients $\beta_k^{(1)}$ and $\gamma_k^{(1)}$ are determined from the coefficients α_h. Specifically, by substituting the definitions of $\varepsilon_{t-2k}^{(1)}$ and $\bar{\varepsilon}_{t-2k}^{(1)}$ into eq. (2) and comparing with eq. (1), $\beta_k^{(1)}$ and $\gamma_k^{(1)}$ are uniquely obtained from the coeffcients α_h via the linear system

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \beta_k^{(1)} \\ \gamma_k^{(1)} \end{bmatrix} = \begin{bmatrix} \alpha_{2k} \\ \alpha_{2k+1} \end{bmatrix}.$$

We focus now on the low-frequency component $\pi_t^{(1)} = \sum_{k=0}^{3} \gamma_k^{(1)} \bar{\varepsilon}_{t-2k}^{(1)}$ and we consider the innovations $\bar{\varepsilon}_t$ as drivers. Specifically, for $k = 0, 1$, we let

$$\varepsilon_{t-4k}^{(2)} = \frac{\bar{\varepsilon}_{t-4k}^{(1)} - \bar{\varepsilon}_{t-4k-2}^{(1)}}{\sqrt{2}}, \qquad \bar{\varepsilon}_{t-4k}^{(2)} = \frac{\bar{\varepsilon}_{t-4k}^{(1)} + \bar{\varepsilon}_{t-4k-2}^{(1)}}{\sqrt{2}}$$

with the aim of determining the coefficients $\beta_k^{(2)}$ and $\gamma_k^{(2)}$ such that

$$\pi_t^{(1)} = \sum_{k=0}^{1} \beta_k^{(2)} \ \varepsilon_{t-4k}^{(2)} + \sum_{k=0}^{1} \gamma_k^{(2)} \ \bar{\varepsilon}_{t-4k}^{(2)}.$$

We do so by solving, for $k = 0, 1$, the linear system

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \beta_k^{(2)} \\ \gamma_k^{(2)} \end{bmatrix} = \begin{bmatrix} \gamma_{2k}^{(1)} \\ \gamma_{2k+1}^{(1)} \end{bmatrix}.$$

As a final step, we concentrate on the term $\pi_t^{(2)} = \sum_{k=0}^{1} \gamma_k^{(2)} \bar{\varepsilon}_{t-4k}^{(2)}$ generated by the shocks $\bar{\varepsilon}_t^{(2)}$. Similarly to the previous construction, we obtain

$$\pi_t^{(2)} = \beta_0^{(3)} \ \varepsilon_t^{(3)} + \gamma_0^{(3)} \ \bar{\varepsilon}_t^{(3)}$$

with

$$\varepsilon_t^{(3)} = \frac{\bar{\varepsilon}_t^{(2)} - \bar{\varepsilon}_{t-4}^{(2)}}{\sqrt{2}}, \qquad \bar{\varepsilon}_t^{(3)} = \frac{\bar{\varepsilon}_t^{(2)} + \bar{\varepsilon}_{t-4}^{(2)}}{\sqrt{2}}.$$

Accordingly, $\beta_0^{(3)}$ and $\gamma_0^{(3)}$ constitute the unique solution of the system

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \beta_0^{(3)} \\ \gamma_0^{(3)} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(2)} \\ \gamma_1^{(2)} \end{bmatrix}.$$

By reassembling the results of this recursive procedure, we obtain the following representation of x_t:

$$x_t = \sum_{j=1}^{3} \sum_{k=0}^{2^{3-j}-1} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)} + \gamma_0^{(3)} \bar{\varepsilon}_t^{(3)}, \tag{3}$$

where each $\varepsilon_t^{(j)}$ and $\bar{\varepsilon}_t^{(3)}$ can be expressed in terms of the initial innovations ε_t as

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right), \qquad \bar{\varepsilon}_t^{(3)} = \frac{1}{\sqrt{2^3}} \sum_{i=0}^{2^3-1} \varepsilon_{t-i}.$$

Moreover, by solving iteratively the linear systems defined above, one sees that the coefficients $\beta_k^{(j)}$ and $\gamma_0^{(3)}$ obtain uniquely from the initial α_h as follows:

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right), \qquad \gamma_k^{(3)} = \frac{1}{\sqrt{2^3}} \sum_{i=0}^{2^3-1} \alpha_{k2^3+i}.$$

We use the term *scale* for the index j and we call *persistent component at scale* j the sum $g_t^{(j)} = \sum_{k=0}^{2^{3-j}-1} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}$. The scale determines the moving average order of the rescaled block-of-2^{j-1} differences $\varepsilon_t^{(j)}$ in terms of the original shocks. Moreover, the components $g_t^{(j)}$ at higher scales are associated with higher persistence levels.

Our *nested* construction delivers a representation of x_t based on innovations that involve sparser and sparser time grids. In fact, at any scale j the processes $\{\varepsilon_{t-k2^j}^{(j)}\}_{k\in\mathbb{Z}}$ and $\{\bar{\varepsilon}_{t-k2^j}^{(3)}\}_{k\in\mathbb{Z}}$ are white noise because there is no overlap of innovations ε_t when a different k is considered on a grid of 2^j time steps. In addition, at any instant t the correlation between any $\varepsilon_t^{(j)}$ and $\varepsilon_t^{(l)}$ with $j \neq l$ is null. Each $\varepsilon_t^{(j)}$ is also uncorrelated with $\bar{\varepsilon}_t^{(3)}$. Therefore, eq. (3) provides a decomposition of x_t in terms of a linear

combination of increasingly persistent components that are orthogonal.

Interestingly, the scale-specific shocks $\varepsilon_t^{(j)}$ and $\bar{\varepsilon}_t^{(3)}$ can be seen as the application of the Discrete Haar Transform to the sequence of the original innovations ε_t (see, for instance, Addison 2002). The same property holds for the coefficients $\beta_k^{(j)}$ and $\gamma_0^{(3)}$, which are obtained from the Discrete Haar Transform of the sequence of coefficients α_h. Hence, we can readily compute $\beta_k^{(j)}$ and $\gamma_0^{(3)}$ from the original MA representation of x_t.

In light of this example, we ask whether it is possible to generalize the decomposition of eq. (3) in two main directions. First, we want to deal with MA processes of any order. By the Classical Wold Decomposition Theorem, this permits to analyze any weakly stationary time series. Second, we want to allow for a potentially infinite number of scales in order to reduce as much as possible the magnitude of the term $\gamma_0^{(3)}\bar{\varepsilon}_t^{(3)}$ so to obtain a univocal persistence-based decomposition. We do this in the next section.

2.1 The Extended Wold Decomposition of x_t

We consider, throughout, a zero-mean, weakly stationary purely non-deterministic process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$, in the sense that there exists a unit variance white noise $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h}, \tag{4}$$

where the coefficients α_h are square-summable, independent of t, and $\alpha_h = \mathbb{E}\left[x_t \varepsilon_{t-h}\right]$. The process $\boldsymbol{\varepsilon}$ is commonly called the *sequence of fundamental innovations* of $\mathbf{x}$. This is the Classical Wold Decomposition for time series (Wold 1938).

By applying the iterative procedure of the introductory example on the infinite MA representation of x_t, we obtain a Wold-type decomposition of x_t based on increasingly persistent white noise processes.

Theorem 1. *Let* $\mathbf{x}$ *be a zero-mean, weakly stationary purely non-deterministic stochastic process. Then* x_t *decomposes as*

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}, \tag{5}$$

where the equality is in the L^2*-norm and*

i) for any fixed $j \in \mathbb{N}$*, the process* $\boldsymbol{\varepsilon^{(j)}} = \{\varepsilon_t^{(j)}\}_{t\in\mathbb{Z}}$ *with*

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right) \tag{6}$$

is an $MA(2^j - 1)$ *with respect to the classical Wold innovations of* $\mathbf{x}$ *and* $\{\varepsilon_{t-k2^j}^{(j)}\}_{k\in\mathbb{Z}}$ *is a unit variance white noise;*

ii) for any $j \in \mathbb{N}$*,* $k \in \mathbb{N}_0$*, the coefficients* $\beta_k^{(j)}$ *are uniquely determined via*

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right), \tag{7}$$

hence, they are independent of t *and* $\sum_{k=0}^{\infty} (\beta_k^{(j)})^2 < +\infty \ \forall j \in \mathbb{N}$*;*

iii) letting

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}, \tag{8}$$

for any $j, l \in \mathbb{N}, p, q, t \in \mathbb{Z}$*,* $\mathbb{E}[g_{t-p}^{(j)} g_{t-q}^{(l)}]$ *depends at most on* $j, l, p-q$*. Moreover,* $\mathbb{E}[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)}] = 0$ *for all* $j \neq l$*,* $m, n \in \mathbb{N}_0$ *and* $t \in \mathbb{Z}$*.*

We call *Extended Wold Decomposition of* x_t the linear representation in eq. (5). Moreover, we call $\beta_k^{(j)}$ the scale-specific moving average (MA) coefficient associated to scale j and time translation $k2^j$. Finally, we call *detail process at scale* j and *persistent component at scale* j the processes defined by $\varepsilon_t^{(j)}$ and $g_t^{(j)}$, respectively.

Putting together eqs. (5) and (8) in Theorem 1, x_t can be expressed as the infinite

orthogonal sum

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}. \tag{9}$$

Note that, as it is apparent from the iterative procedure described in the introductory example, there is also a finite-scale version of the Extended Wold Decomposition, in which a maximum scale J is fixed. In this case,

$$x_t = \sum_{j=1}^{J} g_t^{(j)} + \pi_t^{(J)},$$

where the *residual component at scale* J is defined by

$$\pi_t^{(J)} = \sum_{k=0}^{+\infty} \gamma_k^{(J)} \bar{\varepsilon}_{t-k2^J}^{(J)} \tag{10}$$

and $\bar{\varepsilon}_t^{(J)}$ and $\gamma_k^{(J)}$ satisfy, respectively,

$$\bar{\varepsilon}_t^{(J)} = \frac{1}{\sqrt{2^J}} \sum_{i=0}^{2^J-1} \varepsilon_{t-i}, \qquad \gamma_k^{(J)} = \frac{1}{\sqrt{2^J}} \sum_{i=0}^{2^J-1} \alpha_{k2^J+i}. \tag{11}$$

Theorem 1 follows from the application of the Abstract Wold Theorem (see Theorem 1.1 in Nagy et al. 2010) to the Hilbert space $\mathcal{H}_t(\boldsymbol{\varepsilon})$ spanned by the sequence of fundamental innovations $\{\varepsilon_{t-k}\}_{k\in\mathbb{N}_0}$:

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad \sum_{k=0}^{+\infty} a_k^2 < +\infty \right\}.$$

In words, $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is the space of infinite moving averages generated by $\boldsymbol{\varepsilon}$. The Abstract Wold Theorem provides an orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ by iteratively applying an isometric linear operator that generalizes the recursive procedure of the introductory example. Specifically, in our case we define the *scaling operator*

$\mathbf{R} : \mathcal{H}_t(\boldsymbol{\varepsilon}) \longrightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ by

$$\mathbf{R}: \quad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right) = \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k}, \tag{12}$$

where $\lfloor \cdot \rfloor$ associates any real number c with the integer $\lfloor c \rfloor = \max\{n \in \mathbb{Z} : n \leqslant c\}$. The scaling operator is well-defined, linear and isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$. As a result, $\mathcal{H}_t(\boldsymbol{\varepsilon})$ decomposes into the direct sum $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=0}^{\infty} \mathbf{R}^j \mathcal{L}_t^{\mathbf{R}}$, where $\mathcal{L}_t^{\mathbf{R}}$ is the orthogonal complement of $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$. From the Classical Wold Decomposition, x_t belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$ and so the decomposition into uncorrelated components of eq. (9) follows from projecting x_t on the orthogonal subspaces.[1]

Property iii) in Theorem 1 concerns the relation among the components $g_t^{(j)}$ at different layers of persistence. When t is fixed, the variables $g_t^{(j)}$ and $g_t^{(l)}$, with $j \neq l$, are uncorrelated. Besides, the uncorrelation involves also shifted variables $g_{t-m2^j}^{(j)}$ and $g_{t-n2^l}^{(l)}$, for any $m, n \in \mathbb{Z}$, where the time translation is proportional to 2^j and 2^l respectively. In general, the covariance between $g_{t-p}^{(j)}$ and $g_{t-q}^{(l)}$ depends at most on the scales j, l and on the difference $p - q$.

The orthogonality of the persistent components is a key property since it allows for a decomposition of the total variance of x_t across persistence levels:

$$var\left(x_t\right) = \sum_{j=1}^{+\infty} var\left(x_t^{(j)}\right) = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \left(\beta_k^{(j)}\right)^2.$$

Thus, the Extended Wold Decomposition induces a persistence-based variance decomposition of any weakly stationary process, thanks to the uncorrelatedness of components ensured by the Abstract Wold Theorem. The key novelty is that our persistence-based variance decomposition speaks about the role of both *time and frequency* in the evaluation of the impact of economic shocks, whereas classical time-

[1] Interestingly, the same logic underlies the Classical Wold Decomposition, that can be derived by employing the lag operator as isometry. See Severino (2016).

series methods, like principal component analysis, find it often onerous to disentangle the two dimensions.

We remark that it is also possible to build an Extended Wold Decomposition based on time grids of N^j steps instead of 2^j. This orthogonal decomposition exploits an isometric operator that generalizes $\mathbf{R}$ by averaging N subsequent innovations instead of just 2. Details can be found in the Addendum available on the website of the authors.

Importantly, the approach that exploits the Abstract Wold Theorem and the isometric scaling operator to identify the persistent components generalizes to stationary multivariate processes. In a nutshell, Theorem 1 continues to hold even when $\mathbf{x}$, $\boldsymbol{\varepsilon}$ and all $\boldsymbol{\varepsilon}^{(j)}$ are m-dimensional processes, so that α_h and $\beta_k^{(j)}$ are $m \times m$ matrices of coefficients. Therefore, any multivariate stationary process admits a decomposition of the type

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

where square-summability is ensured by the convergence of $\sum_{k=0}^{\infty} \mathrm{Tr}(\beta_k^{(j)} \beta_k^{(j)\prime})$, the trace of the matrix $\beta_k^{(j)} \beta_k^{(j)\prime}$. We refer the reader interested in the technical details to Cerreia-Vioglio et al. (2017).

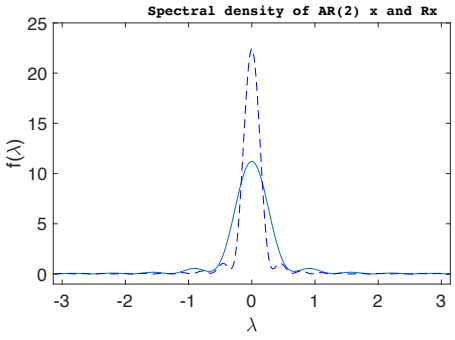


Figure 1 Comparison between the spectral density functions of x_t (solid line) and $\mathbf{R}x_t$ (dashed line) when x_t is an $AR(2)$ process with $\phi_1 = 1.16$ and $\phi_2 = -0.24$.

Finally, we look at our decomposition from the standpoint of spectral analysis. We assume that the classical Wold coefficients of x_t are absolutely summable, so that

x_t has a well-defined spectral density function $f_x : (-\pi, \pi] \longrightarrow \mathbb{C}$, namely $f_x(\lambda) = \sum_{n\in\mathbb{Z}} e^{-in\lambda}\gamma(n)/2\pi$. The scaling operator is an approximate low-pass filter and it is this feature that makes it valuable to capture persistence. Specifically, the spectral density functions of $\mathbf{R}x_t$ is

$$f_{\mathbf{R}}(\lambda) = 2\cos^2\left(\frac{\lambda}{2}\right) f_x(2\lambda),$$

as computed in Proposition A.1 in the Online Supplement. Hence, $f_{\mathbf{R}}$ associates the largest weight with frequencies around zero, while frequencies near to $|\pi|$ are negligible. An example of the impact of $\mathbf{R}$ on spectral density functions is described by Figure 1, where x_t is assumed to be a stationary $AR(2)$ process.

The next subsection provides the theory for the aggregation of given persistent components to a single stationary process.

2.2 Reconstructing a time series from its components

This subsection investigates the converse of our Extended Wold Decomposition: once the dynamics are given at various time scales, what can we say about the resulting process $\mathbf{x}$, built by summing up such components?

Given the dynamics at different time scales, we want to recover the whole process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$. To make the sum of components feasible, we assume a common innovation process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$. In fact, while MA processes may have multiple representations that involve different innovations, considering MA processes driven by the same source of randomness has the advantage that the sum of such time series is still based on the same random shocks.

At each scale $j \in \mathbb{N}$, we define the detail process $\boldsymbol{\varepsilon}^{(j)} = \{\varepsilon_t^{(j)}\}_{t\in\mathbb{Z}}$ as an $MA(2^j - 1)$

driven by the underlying innovations $\boldsymbol{\varepsilon}$. More precisely, we assume that

$$\varepsilon_t^{(j)} = \sum_{i=0}^{2^j-1} \delta_i^{(j)} \varepsilon_{t-i}$$

for some real coefficients $\delta_i^{(j)}$. Next, we consider the processes $\mathbf{g^{(j)}} = \{g_t^{(j)}\}_{t\in\mathbb{Z}}$ associated with level of persistence j. We build these processes in a way that on the time grids $S_t^{(j)} = \{t - k2^j : \ k \in \mathbb{N}_0\}$ they are well-defined $MA(\infty)$ with respect to $\boldsymbol{\varepsilon^{(j)}}$. In particular, for any given j, we assume that there exists a sequence of coefficients $\{\beta_k^{(j)}\}_{k\in\mathbb{N}_0}$ such that

$$\sum_{j=1}^{+\infty}\sum_{h=0}^{+\infty}\left(\beta^{(j)}_{\lfloor \frac{h}{2^j} \rfloor}\delta^{(j)}_{h-2^j\lfloor \frac{h}{2^j} \rfloor}\right)^2 < +\infty. \tag{13}$$

Then, for any t, we define the variables

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon^{(j)}_{t-k2^j}$$

and, in addition, we assume

$$\mathbb{E}\left[g^{(j)}_{t-m2^j} g^{(l)}_{t-n2^l}\right] = 0 \qquad \forall j \neq l, \quad \forall m, n \in \mathbb{N}_0, \quad \forall t \in \mathbb{Z}. \tag{14}$$

While Condition (13) involves the interaction of the coefficients $\beta_k^{(j)}$ and $\delta_i^{(j)}$, it is interesting to discuss two special cases in which Condition (13) holds by imposing restrictions separately on $\beta_k^{(j)}$ and $\delta_i^{(j)}$.

First, consider the special case of Haar coefficients, i.e.

$$\delta_i^{(j)} = \begin{cases} \frac{1}{\sqrt{2^j}} & \text{if} \quad i \in \{0, \ldots, 2^{j-1}-1\} \\ -\frac{1}{\sqrt{2^j}} & \text{if} \quad i \in \{2^{j-1}, \ldots, 2^j-1\}. \end{cases} \tag{15}$$

In this case, Condition (13) translates into the standard square-summability of the

coefficients $\beta_k^{(j)}$, i.e.

$$\sum_{j=1}^{+\infty}\sum_{k=0}^{+\infty}\left(\beta_k^{(j)}\right)^2 < +\infty.$$

In addition, Condition (14) is also satisfied as a consequence of the orthogonal structure discussed in Subsection 2.1.

Second, one can assume that the coefficients $\delta_i^{(j)}$ are uniformly bounded over $j \in \mathbb{N}$ and $i = 1, \dots 2^j$, and require that

$$\sum_{j=1}^{+\infty} 2^j \sum_{k=0}^{+\infty}\left(\beta_k^{(j)}\right)^2 < +\infty.$$

so that Condition (13) holds once again. In this case the condition on the coefficients $\delta_i^{(j)}$ is weaker, while the square-summability of the $\beta_k^{(j)}$ is more restrictive.

The following theorem states the properties of the process $\mathbf{x}$ obtained by summing the components $\mathbf{g^{(j)}}$ over all the scales $j \in \mathbb{N}$ when Conditions (13) and (14) hold.

Theorem 2. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a unit variance white noise process. For any $j \in \mathbb{N}$, define the detail process $\boldsymbol{\varepsilon^{(j)}} = \{\varepsilon_t^{(j)}\}_{t\in\mathbb{Z}}$ as*

$$\varepsilon_t^{(j)} = \sum_{i=0}^{2^j-1} \delta_i^{(j)} \varepsilon_{t-i},$$

with $\delta_i^{(j)} \in \mathbb{R}$ for $i = 0, \dots, 2^j - 1$, and consider a stochastic process $\mathbf{g^{(j)}} = \{g_t^{(j)}\}_{t\in\mathbb{Z}}$ such that there exists a sequence of real coefficients $\{\beta_k^{(j)}\}_{k\in\mathbb{N}_0}$ so that Conditions (13) and (14) are satisfied. Then, the process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ defined by

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}$$

is zero-mean, weakly stationary purely non-deterministic and

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h} \qquad with \qquad \alpha_h = \sum_{j=1}^{+\infty} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \delta^{(j)}_{h-2^j\left\lfloor \frac{h}{2^j} \right\rfloor} \quad \forall h \in \mathbb{N}_0. \tag{16}$$

The theorem provides the moving average representation, with respect to the innovations ε_t, of the process $\mathbf{x}$ built via the aggregation of persistent components. In case the shocks ε_t are the classical Wold innovations of $\mathbf{x}$, Theorem 2 supplies the Classical Wold Decomposition of x_t, starting from its persistence-based representation. As a consequence, it constitutes a converse result of Theorem 1.

Note that, when the Haar coefficients are employed, the classical impulse responses can be retrieved via

$$\alpha_h = \sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\lfloor \frac{h}{2^j} \rfloor} \chi^{(j)}(h),$$

where

$$\chi^{(j)}(h) = \begin{cases} -1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \{h - 2^j + 1, \ldots, h - 2^{j-1}\}, \\ 1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \{h - 2^{j-1} + 1, \ldots, h\}. \end{cases}$$

2.2.1 An illustrative example of the reconstruction theorems

Next we provide a simple example of a multiscale construction of a stationary process. We consider a two-scale process given by the sum of a fast evolving autoregressive process and a second, more persistent autoregressive process evolving on a coarser time scale. We show by means of a numerical example that different choices of parameters have different impacts on the short and long lags of the ACF of the resulting process.

Given a common innovation process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$, we let the coefficients $\delta_i^{(j)}$ in Theorem 2 be the Haar coefficients of eq. (15). We start from two weakly stationary purely non-deterministic time series, $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ and $\mathbf{y} = \{y_t\}_{t\in\mathbb{Z}}$, defined by two families of scale-specific moving average coefficients: $\{\beta^{(j)}_{x,k}\}_{j,k}$ and $\{\beta^{(j)}_{y,k}\}_{j,k}$ with $j \in \mathbb{N}$ and $k \in \mathbb{N}_0$.

We assume the scale-specific moving average coefficients of x_t to be

$$\beta_{x,k}^{(j)} = \frac{\rho_x^{k2^j}}{\sqrt{2^j}} \frac{\left(1-\rho_x^{2^{j-1}}\right)^2}{1-\rho_x}, \qquad j \in \mathbb{N}, k \in \mathbb{N}_0,$$

i.e. the extended Wold coefficients of a weakly stationary $AR(1)$ process with parameter ρ_x, with $|\rho_x| < 1$.

We now fix a scale J and we define the scale-specific coefficients of y_t by setting $\beta_{y,k}^{(j)} = 0$ if $j = 0, \ldots, J$ and

$$\beta_{y,k}^{(j)} = \frac{\rho_y^{k2^{j-J}}}{\sqrt{2^{j-J}}} \frac{\left(1-\rho_y^{2^{j-J-1}}\right)^2}{1-\rho_y}$$

if $j \geqslant J+1$, with $|\rho_y| < 1$. A simple comparison with $\beta_{x,k}^{(j)}$ shows that the coefficients $\beta_{y,k}^{(j)}$ identify the autoregressive process $\mathbf{y}$ defined on the grid $S_t^{(J)} = \{t - k2^J : \ k \in \mathbb{N}_0\}$ by

$$y_t = \sum_{k=0}^{+\infty} \rho_y^k \tilde{\varepsilon}_{t-k2^J}^{(J)},$$

where

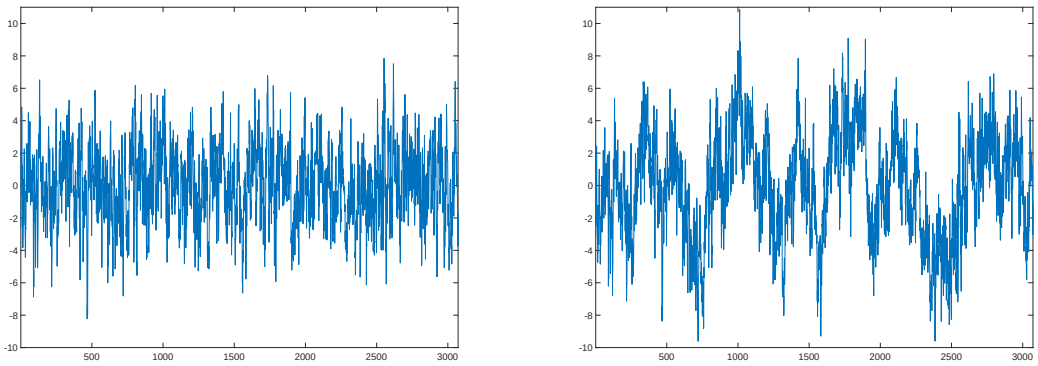
$$\tilde{\varepsilon}_t^{(J)} = \frac{1}{\sqrt{2^J}} \sum_{i=0}^{2^J-1} \varepsilon_{t-i}$$

denotes the unit variance white noise $\tilde{\boldsymbol{\varepsilon}}^{(\boldsymbol{J})} = \{\tilde{\varepsilon}_{t-k2^J}^{(J)}\}_{k\in\mathbb{Z}}$. In contrast with $\mathbf{x}$, which is a standard $AR(1)$ process, we call $\mathbf{y}$ an $AR(1)$ process *with horizon* 2^J. To enhance interpretation, one can think of $\mathbf{x}$ as a daily process, while $\mathbf{y}$ is a time series acting on longer lags (e.g. monthly, yearly, ...), depending on the choice of J.

By construction, the scale-specific moving average coefficients of $\mathbf{z} = \{z_t\}_{t\in\mathbb{Z}}$ are the sum of the extended Wold coefficients of $\mathbf{x}$ and $\mathbf{y}$, that is $\beta_{z,k}^{(j)} = \beta_{x,k}^{(j)} + \beta_{y,k}^{(j)}$ for any $j \in \mathbb{N}$, $k \in \mathbb{N}_0$. It is readily checked that all the conditions of Theorem 2 are satisfied. Therefore, $\mathbf{z}$ is weakly stationary and purely non-deterministic. Note that while the sum of two autoregressive processes with different horizons is not necessarily

stationary, the structure of the shocks at different scales required by Theorem 2 ensures that this holds true in this case. Hereafter we plot the ACF of $\mathbf{z}$ for some choices of the parameters ρ_x, ρ_y and of the scale J.

We start by setting $\rho_x = 0.7$, $\rho_y = 0.9$ and $J = 3$. Figure 2 compares a simulated path of the basic autoregressive process $\mathbf{x}$ with a path of the process $\mathbf{z}$ that is obtained after the addition of the persistent component. In the resulting process $\mathbf{z}$ the fast process corresponding to the process $\mathbf{x}$ mean reverts toward a trend which is itself a slowly mean reverting process, as determined by the low-frequency component.

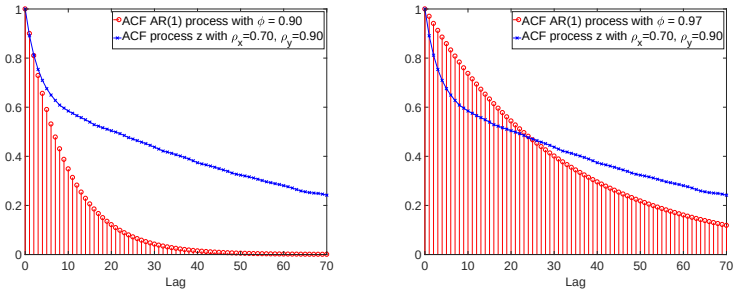


(a) (b)

Figure 2 Simulated path for the process $\mathbf{x}$ with $\rho_x = 0.7$ and for the process $\mathbf{z}$ with $\rho_x = 0.7$, $\rho_y = 0.9$, $J = 3$.

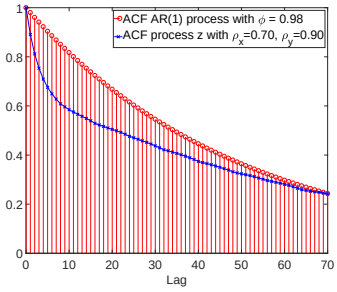
In Figure 3 we see that the ACF of $\mathbf{z}$ is piecewise approximated by the ACFs of $AR(1)$ processes with different parameters. Moreover, the persistence of $\mathbf{z}$ increases with the scale J.

Alternative choices of ρ_y allow us to modify the short lags of the ACF of $\mathbf{x}$, while keeping the long lags unchanged. For instance, this is possible by setting $\rho_y = \rho_x^{2^J}$, as we show in Addendum Z.3. Summing up, the scale-by-scale construction of $\mathbf{z}$ and the choices of the parameters ρ_x, ρ_y and of the scale J allow us to obtain ACFs with predetermined features. Further numerical examples are in the Addendum.



(a)

(b)



(c)

Figure 3 Comparison between the ACF of the process $\mathbf{z}$ with $\rho_x = 0.7$, $\rho_y = 0.9$, $J = 3$ and the ACF of an $AR(1)$ process with parameter $\phi = 0.9, 0.97, 0.98$ in panel (a), (b) and (c), respectively.

2.3 The Extended Wold Decomposition vs the multiresolution approach

In this section we compare our Extended Wold Decomposition with the multiresolution approach to time series. Formal statements and proofs are in the Addendum.

A standard way to isolate phenomena with heterogeneous persistence in economic time series comes from the application of the Discrete Haar Transform to the realizations of a given process. As explained, for instance, in Ortu et al. (2013) a *multiresolution approach* helps to disentangle low-frequency disturbances from high-frequency fluctuations. To compare this approach with our Extended Wold Decomposition, given a weakly stationary process $\mathbf{x}$, we follow Ortu et al. (2013) and we

build moving averages of size 2^j of past realizations of x

$$\breve{\pi}_t^{(j)} = \frac{1}{2^j} \sum_{p=0}^{2^j-1} x_{t-p}$$

that include fluctuations whose half-life exceeds 2^j periods. Accordingly, the differences between moving averages of sizes 2^{j-1} and 2^j, i.e.

$$\breve{g}_t^{(j)} = \breve{\pi}_t^{(j-1)} - \breve{\pi}_t^{(j)} = \frac{1}{2^j} \left(\sum_{i=0}^{2^{j-1}-1} x_{t-i} - \sum_{i=0}^{2^{j-1}-1} x_{t-2^{j-1}-i} \right),$$

capture fluctuations with half-life in the interval $[2^{j-1}, 2^j)$. Since $\breve{\pi}_t^{(0)} = x_t$, it follows readily that, given a maximum scale $J \in \mathbb{N}$,

$$x_t = \sum_{j=1}^{J} \breve{g}_t^{(j)} + \breve{\pi}_t^{(J)}.$$

Thus, x_t is decomposed into a finite sum of variables $\breve{g}_t^{(j)}$ related to different persistence levels, plus a residual long-run average term. If $\breve{\pi}_t^{(J)}$ converges to zero in norm as J goes to infinity, then x_t tends to the infinite sum of $\breve{g}_t^{(j)}$'s. We remark that the convergence of $\breve{\pi}_t^{(J)}$ is ensured for those processes $\mathbf{x}$ whose autocovariance function γ is vanishing, namely $\lim_{n\to\infty} \gamma(n) = 0$. As a result, we obtain the *multiresolution-based decomposition*

$$x_t = \sum_{j=1}^{+\infty} \breve{g}_t^{(j)}.$$

To make this approach comparable to the Extended Wold Decomposition, we consider the closed space spanned by the sequence $\{x_{t-k}\}_{k\in\mathbb{N}_0}$

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl} \left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty \right\}.$$

The multiresolution approach exploits the operator $\mathbf{R}_{\mathbf{x}} : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ that acts

on the generators of $\mathcal{H}_t(\mathbf{x})$ as follows

$$\mathbf{R}_{\mathbf{x}}: \qquad \sum_{k=0}^{+\infty} a_k x_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(x_{t-2k} + x_{t-2k-1}\right) = \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} x_{t-k}.$$

While weak stationarity is not enough to guarantee that $\mathbf{R}_{\mathbf{x}}$ is well-defined on $\mathcal{H}_t(\mathbf{x})$, $\mathbf{R}_{\mathbf{x}}$ is indeed well-defined upon restricting it to the space of *finite* linear combinations of past realizations x_{t-k}. The difference between $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{R}$ is an outcome of the peculiar interaction between the scaling operator and the lag operator, which do not commute. This fact makes $\mathbf{R}$ different from $\mathbf{R}_{\mathbf{x}}$ and so the Extended Wold Decomposition turns out to be different from the multiresolution-based decomposition, which is induced by $\mathbf{R}_{\mathbf{x}}$.

Differently from the components $g_t^{(j)}$, a certain amount of correlation may be present between $\breve{g}_t^{(j)}$ and $\breve{g}_t^{(l)}$ associated with different scales. For example, if $\mathbf{x}$ is a weakly stationary $AR(1)$ process with parameter ρ, we have $\mathbb{E}[\breve{g}_t^{(1)} \breve{g}_t^{(2)}] = \rho/8$.

In general, each $\breve{g}_t^{(j)}$ can be expressed in terms of details $\varepsilon_t^{(j)}$:

$$\breve{g}_t^{(j)} = \sum_{h=0}^{+\infty} \frac{\alpha_h}{\sqrt{2^j}} \varepsilon_{t-h}^{(j)}. \tag{17}$$

This expression makes the comparison between $\breve{g}_t^{(j)}$ and $g_t^{(j)}$ clear. The variables $\breve{g}_t^{(j)}$ involve all the lags of the details $\varepsilon_t^{(j)}$ and they exploit directly the classical Wold coefficients α_h. On the other hand, the components $g_t^{(j)}$ concern just a selection of details $\varepsilon_t^{(j)}$, namely those on the grid $S_t^{(j)}$, and they concentrate the relevant information in the coefficients $\beta_k^{(j)}$. Moreover, the persistent components $g_t^{(j)}$ at different scales are uncorrelated and so the Extended Wold Decomposition can be interpreted as an orthogonalization of the multiresolution-based decomposition.

A further insight in the differences between the operator $\mathbf{R}$ defined in eq. (12) and the operator $\mathbf{R}_{\mathbf{x}}$ is provided by spectral analysis in case the classical Wold innovations of x_t are absolutely summable. Both $\mathbf{R}$ and $\mathbf{R}_{\mathbf{x}}$ are, indeed, approximate low-pass

filters: the spectral density functions of $\mathbf{R}x_t$ and $\mathbf{R}_{\mathbf{x}}x_t$ are respectively

$$f_{\mathbf{R}}(\lambda) = 2\cos^2\left(\frac{\lambda}{2}\right) f_x(2\lambda) \qquad \text{and} \qquad f_{\mathbf{R}_{\mathbf{x}}}(\lambda) = 2\cos^2\left(\frac{\lambda}{2}\right) f_x(\lambda).$$

While the two densities are different, they both associate the largest weight with frequencies around zero, while frequencies near to π or $-\pi$ are negligible.

3 Empirical Analysis

We present now two applications of our Extended Wold Decomposition. The first regards the analysis of Realized Volatility, while the second discusses the relation between yields and Treasury bond returns. We show that our decomposition is able to detect the components of the process under scrutiny that are most relevant for forming future expectations, and separate these from the noisy components. The orthogonality of the components, and their association with a specific level of persistence, makes the economic interpretation of the results easier compared to well established forecasting benchmark in the respective literature.

3.1 Persistence of Realized Volatility

One of the most intriguing features revealed by empirical work on volatility is its high persistence. Among many alternative explanations for the source of such persistence, Andersen and Bollerslev (1997) argue that the long-memory feature may arise naturally if the volatility process reflects the aggregate impact of several distinct information arrival processes, each one characterized by its own degree of persistence. Similarly, the Heterogeneous Market Hypothesis of Müller et al. (1997) suggests that the long memory of volatility can be the result of an additive cascade of different volatility components, each driven by the actions of different (in terms of investment horizon) types of market participants. In this subsection we use the insights of our

Extended Wold Decomposition to provide supporting evidence for these hypotheses. Additional details and figures are collected in the Addendum.

3.1.1 Data and component extraction

We consider the time series of daily USD/CHF exchange rate Realized Volatility built by Corsi (2009). Starting from tick-by-tick series of USD/CHF exchange rates from December 1989 to December 2003, the daily Realized Volatility $\mathbf{d} = \{d_t\}_t$ is constructed by computing spot logarithmic middle prices (averages of log bid and ask quotes) and the related returns as in Andersen et al. (2003): $d_t^2 = \sum_{j=0}^{M-1} r_{t-j/M}^2$, where $r_{t-j/M} = p(t - j/M) - p(t - (j+1)/M)$ and $M = 12$ refers to time intervals of two hours in a 24 hours trading day.

We assume that $\mathbf{d}$ is a weakly stationary time series and we estimate an autoregressive process with 25 lags, as suggested by the BIC. We then retrieve the related MA representation of d_t in terms of its own shocks ε_t, and we compute the innovations $\varepsilon_t^{(j)}$ and the scale-specific coefficients $\beta_k^{(j)}$ as prescribed by eq. (6) and eq. (7). Finally, each component $d_t^{(j)}$ at level of persistence j is obtained from eq. (8).

The sample of daily Realized Volatility contains 3599 data and we estimate persistent components up to scale $J = 10$.

3.1.2 Results

The left panel of Figure 4 provides a variance decomposition of Realized Volatility. We observe that most of the variance is explained by scales $j = 7, 8$ and 9, which involve shocks that last 128, 256 and 512 working days. This preliminary evidence is potentially consistent with the existence of market participants with different time horizons as postulated by Müller et al. (1997). Importantly, our analysis shows that most of the action is in the long run and hence it is more likely associated with long-term investors rather than, e.g., short-term FX dealers.

We next introduce a forecasting method that exploits the components of d_t. In particular, we first estimate the OLS coefficients $a^{(0)}$ and $a^{(j)}$ for $j = 1, \ldots, 10$ in the regression

$$d_t = a^{(0)} + \sum_{j=1}^{J} a^{(j)} d_t^{(j)} + \xi_t. \tag{18}$$

Then, we consider eq. (18) at time $t+1$, and we apply the conditional expectation at t to obtain the one-step ahead forecast

$$\mathbb{E}_t \left[d_{t+1} \right] = a^{(0)} + \sum_{j=1}^{J} a^{(j)} \mathbb{E}_t \left[d_{t+1}^{(j)} \right]. \tag{19}$$

To estimate the conditional expectations $\mathbb{E}_t[d_{t+1}^{(j)}]$ at scales j we use the forecasting formulas that we provide in the Online Supplement A.2. By employing the estimates of $a^{(0)}$ and $a^{(j)}$ from eq. (18), we obtain a one-day ahead forecast for d_{t+1} based on the components of Realized Volatility.

Rather than considering all the components at once, we study the performance of models with increased complexity as captured by the number of regressors. In particular, we add one component at a time based on the explained variance reported in the leftmost panel of Figure 4. Hence, we start with a model that includes only the component $j = 8$, then we move to a model with two components $j = 7, 8$, and so on. Interestingly, the addition of a new component leaves (largely) unaffected the coefficients on the components included by the previous model. This is so because the Wold components are uncorrelated by construction. Indeed, one could even impose the restriction $a^{(j)} = 1$ with $j \in \mathbb{N}$ as suggested by the Extended Wold Decomposition.

The right panel of Figure 4 shows the relation between in-sample forecasting performance and the number of persistent components added according to their explained variance. After including the three scales $j = 7, 8, 9$, we do not see a meaningful im-

provement in forecasting ability. Hence, our preferred model is given by

$$d_t = a^{(0)} + a^{(7)} d_t^{(7)} + a^{(8)} d_t^{(8)} + a^{(9)} d_t^{(9)} + \xi_t, \tag{20}$$

which delivers the forecast

$$\mathbb{E}_t \left[d_{t+1} \right] = a^{(0)} + a^{(7)} \mathbb{E}_t \left[d_{t+1}^{(7)} \right] + a^{(8)} \mathbb{E}_t \left[d_{t+1}^{(8)} \right] + a^{(9)} \mathbb{E}_t \left[d_{t+1}^{(9)} \right]. \tag{21}$$

In other words, we are making predictions of daily volatility d_t from its long cycles with persistence of about 128, 256 and 512 days.

Table 1 quantifies the performance of our persistence-based forecasting with three components and compares it with the Heterogeneous Autoregressive model (HAR) of Corsi (2009). We follow Corsi (2009) and report results for out-of-sample forecasts of the realized volatility in which the models are reestimated daily on a moving window of 2600 observations.[2] The forecasting performances are compared over two different time horizons: one day and three months. The multistep-ahead forecasts are evaluated considering the aggregated volatility realized and predicted over the multiperiod horizon.

Interestingly, the HAR model forecasts Realized Volatility through the equation

$$d_{t+1} = a_0 + a_d d_t + a_w w_t + a_m m_t + \nu_t, \tag{22}$$

where w_t and m_t denote weekly and monthly Realized Volatility obtained by averaging

[2]Large moving windows allow for a proper estimation of persistent components at high scales. By using 2600 observations we can build time series of 528 realizations for each component from scale 1 to 9. In particular, we want to have at least four coefficients $\beta_k^{(j)}$ for $j = 9$. Hence, 2048 observations are required at any t. Considering 528 realizations of $g_t^{(9)}$, we get 2575 observations. Additional 25 lags are needed in the preliminary AR estimate of realized volatility.

subsequent past realizations of d_t:

$$w_t = \frac{1}{5}\sum_{h=0}^{4} d_{t-h}, \qquad m_t = \frac{1}{22}\sum_{h=0}^{21} d_{t-h}.$$

Hence, also the HAR model exploits three components as our preferred model in eq. (20). For completeness, Table 1 also reports the performance measures from a model which exploits all available Wold components (except the component at scale $j = 10$ which would require a much longer estimation window and, hence, shrink the out-of-sample period).

We start with the one-day-ahead forecasts. Our preferred persistence-based forecasting model with three components fares well compared to the HAR model, while keeping the same parsimony of parameters. Adding more components improves the performance only marginally. Indeed, the persistence-based forecast that relies on three components only and the HAR forecast are highly correlated at 0.92. This correlation increases only slightly to 0.96 if we consider a model with all components. Thus Table 1-Panel A suggests that the persistent components at scales $j = 7, 8$ and 9 are the drivers behind the forecasting performance of the daily, weekly and monthly averages used in the HAR by Corsi (2009).

Looking at the longer forecasting horizon in Panel B reinforces our analysis. Our model continues to fare well, this time with performance slightly superior to the HAR model. Commenting on the performance of the HAR at long horizons, Corsi (2009, p. 192) states: "What is surprising is the ability of the HAR-RV model to attain these results with only a few parameters." At the light of our analysis, it is clear that the ability of the model to forecast well at short- and long-horizons is dictated by the use of components that converge slowly to their unconditional mean relative to the forecasting horizon.

Overall our analysis supports the fact that the forecasting ability of the HAR

model truly depends on high scales.

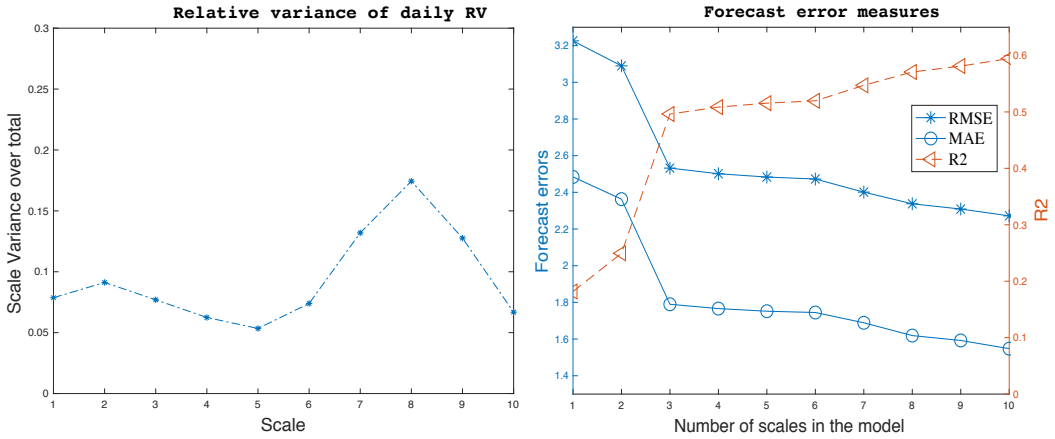


Figure 4 On the left, variance ratio explained by each scale for daily Realized Volatility. On the right, forecasting performance of persistence-based models with increasing number of components as regressors. Persistent components are added according to their explained variance. RMSE, MAE and R^2 denote Root Mean Square Error, Mean Absolute Error and R^2 of Mincer-Zarnowitz regression.

Table 1 Forecasting performance of HAR model and persistence-based models with 3 and 9 components. RMSE, MAE and R^2 denote Root Mean Square Error, Mean Absolute Error and R^2 of Mincer-Zarnowitz regression.

Panel A: 1-day-ahead out-of-sample forecasts			
	RMSE	MAE	R^2
HAR	2.144	1.548	0.658
Extended Wold (3)	2.449	1.873	0.561
Extended Wold (9)	2.300	1.705	0.627
Panel B: 66-day-ahead out-of-sample forecasts			
	RMSE	MAE	R^2
HAR	2.110	1.693	0.523
Extended Wold (3)	2.125	1.646	0.530
Extended Wold (9)	2.106	1.620	0.602

3.2 Persistence of yields to maturity

In this application we show that using the (multivariate) Wold components extracted from the term structure of interest rates sheds new light on the predictability of bond returns.

3.2.1 Data and component extraction

We fit a vector autoregressive process of order p (VAR(p)) to yields $y_t^{(n)}$ with maturities of $n = 1, 2, 3, 4, 5, 7$, and 10 years. The data are from Gürkaynak et al. (2007). We use end-of-month data from September 1971 to December 2016. Using the companion form of the VAR(p) process, we obtain the Wold moving average representation of yields. Using eq. (6) adapted to multivariate processes (see the remark at the end of Section 2.1), we obtain the frequency-specific innovations of the extended Wold representation. Finally, using eq. (7) we evaluate the coefficients of the extended Wold representation. Combining the extended Wold coefficients together with the frequency-specific innovations as in eq. (8) one obtains the Wold components of yields.

While we choose $p = 24$, we remark that the results are robust and the conclusions are unaffected if we increase the lag length to $p = 36, 48, 60$. We stress the importance of using long lags for uncovering cyclical fluctuations in yields. This fact is consistent with the evidence in Cochrane and Piazzesi (2005) that lags of forward rates add to the bond return predictability, in direct contrast with standard Markovian models. Such evidence has spurred interest in long lags model: Monfort and Pegoraro (2007) consider cases of Gaussian Dynamic Term Structure Models based on VAR with $0 < p < \infty$ lags; Feunou and Fontaine (2017) propose a model that allows to deal with the case $p = \infty$ parsimoniously, and generates a role of past yields for predictability.

Finally, we choose $J = 6$.[3] Thus, for each yields we obtain six components plus

[3]The choice is dictated by our sample length. We estimate our VAR(p) starting on 31-Jan-1962

the residual $\pi_t^{(6)}$. We proxy for the level of the yield curve with $L = \sum_{n\geq 2} y_t^{(n)}/6$, and we define the slope as $S = L - y_t^{(1)}$. From the components of yields it is immediate to obtain the components of the level and slope factors.

3.2.2 Results

Table 2 shows the results for bond returns forecasting regressions of the form:

$$\overline{rx}_{t+1} = \beta_0 + \beta_1 X_t + \varepsilon_{t+1}$$

where $\overline{rx}_{t+1}$ is the average (across maturities) excess returns and X_t denotes the vector of predictors. In Panel A, X_t refers to the level or its components; in Panel B, X_t refers to the slope or its components.

We start with the level. Specification (a) shows the standard result in the literature: the level of the term structure of interest rates is not a statistically significant predictor for bond returns. However, specification (b) shows that our Extended Wold Decomposition is able to detect *several* cycles in the level that are indeed useful to predict bond returns: the R^2 raises from 6% to 29%. By the orthogonality of the components, we can move from specification (b) down to (d), and drop the components that are insignificant. The procedure leaves the loadings on the (significant) components and the goodness of fit of the model almost unaffected. Importantly, specification (e) shows that one can bundle together the level components used in specification (d) into *a single* level cycle leaving the results essentially unaffected. Cieslak and Povala (2015) were the first to show that a cyclical component of the level contains information for expected bond returns. In addition, our analysis in Panel A shows that: (1) removing long term trends (like our sixth component and

VAR(p). This leaves about 10 years of data to initialize the components. Indeed we loose 2 years of data to construct lags of the regressors in the VAR. Moreover we need 2^6 observations (roughly 5 years) to obtain the first innovation at scale $j = 6$. We then restrict the effective sample of predictive regressions to September 1971 to December 2016.

the residual) works well because it uncovers high-frequency cycles that range from $1-2$ months ($j=1$) to $16-32$ months ($j=5$); (2) these cycles can be uncovered without using macro information like the inflation trend of Cieslak and Povala (2015). This latter point begs the question of whether the long-term inflation is an economically informative de-trender that bears some peculiar information.

We next turn to the slope. Panel B, specification (a) shows the well-known result in the literature that the slope of the term structure has considerable predictive power for excess returns. Specifications (b) to (e) show a result that is novel and interesting: the slope contains cycles as well, cycles that once properly extracted convey a cleaner signal of future excess returns. This is witnessed by an increase in R^2 of about 5%. Interestingly, the cyclical components of the slope evolve over a longer time scale than those of the level. We come back to this fact in the interpretation of our results below.

Panel C compares our level and slope factors extracted using the Extended Wold Decomposition with other factors employed in the literature. The first row is our benchmark model where we use the level and slope factors from specification (e) in Panel A and B together to predict bond returns. Our factors are (close to) orthogonal as testified by the fact that the R^2 of the multiple regression is close to the sum of the R^2 from simple regressions. Specification (b) is akin to the Cochrane and Piazzesi (2005) model, and it shows that a linear combination of forward rates has predictive ability above and beyond the slope (c.f. Panel B, specification (a)). Yet, the goodness of fit of the Cochrane and Piazzesi (2005) is lower than our benchmark specification by about 12%. Specification (c) is akin to the Cieslak and Povala (2015) level factor: in particular we remove from the level of the yield curve a secular inflation trend. Consistent with Cieslak and Povala (2015), the level cycle so constructed has forecasting ability for bond returns. However, specification (d) shows that the level cycle extracted using our Extended Wold Decomposition drives away a level cycle

obtained using an inflation trend.

Specification (e) is akin to the full specification by Cieslak and Povala (2015). Rebonato and Hatano (2018) shows that the Cieslak and Povala (2015) factor can be rewritten as a combination of the slope and the detrended level used in specification (c). Two remarks are in order. First, the Cieslak and Povala (2015) specification attains the same fit as our benchmark model with components of the level and slope extracted from the extended Wold. Both specifications are based on a high-frequency component related to the level and a low-frequency component related to the slope (recall that our level factor is composed of components $j = 1, 2, 3, 5$ as per specification (d) in Panel A, whereas our slope factor is composed of components $j = 5, 6$ as per specification (d) in Panel B). Despite these similarities, the two models call for potentially different economic explanations. This leads to our second remark. Our benchmark specification (a) provides a very simple attribution of the predictability: 25% comes from the high-frequency cycle in the level (c.f. specification (e) in Panel A), and 14% comes from the low-frequency cycle in the slope (c.f. specification (e) in Panel B), for a total of about 38% (c.f. specification (a) in Panel C). One can then try to link these *two* frequencies to economic factors evolving over different time scales. For instance, Rebonato and Hatano (2018) link the slope to business cycle fluctuations induced by time-varying risk aversion, and the high frequency cycle in the level to temporary deterioration in market liquidity by arbitrageurs. On the other hand, specification (e) shows that one needs to consider also a third factor. This is clearly seen by observing that the sum of the R^2 statistics from the two separate return predictive factors (c.f. specification (a) in Panel B and specification (c) in Panel C) falls short of adding up to the full explanatory power of the bivariate regression (c.f. specification (e) in Panel C). Together, the level cycle and the slope of the yield curve enhance the explanatory power because they are strongly negative correlated at -47%. Overall, specification (e) requires: (1) an economic story for the

high frequency component in the level; (2) an economic story for the low frequency component in the slope; *and* (3) a third factor to explain that the relative values of the slope and detrended level is important.

Comparing specification (a) to specification (e) we can sum up our main findings as follows: (1) it is possible to recover the same goodness of fit without using long term inflation trends: this fact raises concern whether the long-term inflation detrending is really special; (2) despite the same statistical fit, the two specifications call for a different number of factors: our approach relies on the orthogonality of the components and does not need to explain interaction terms in the attribution of predictability.

Panel A: Regression of monthly excess bond returns on Level factor and its components										
Forecasted Variable	Predictors									R^2
$\overline{rx}_{t+1}$	L	Components of Level, L							Wold Level	
		$L^{(1)}$	$L^{(2)}$	$L^{(3)}$	$L^{(4)}$	$L^{(5)}$	$L^{(6)}$	$\pi^{(6)}$		
(a)	0.40									0.06
	(1.46)									
(b)		3.40	3.57	4.21	1.66	5.91	-1.49	0.29		0.29
		(5.53)	(4.58)	(4.50)	(1.07)	(5.55)	(-1.03)	(1.10)		
(c)		3.38	3.61	4.27	1.64	5.92	-1.15			0.27
		(5.45)	(4.39)	(4.20)	(1.04)	(5.28)	(-0.75)			
(d)		3.37	3.54	4.18		5.86				0.26
		(5.33)	(4.13)	(4.06)		(5.36)				
(e)									5.16	0.25
									(6.49)	

Panel B: Regression of monthly excess bond returns on Slope factor and its components										
Forecasted Variable	Predictors									R^2
$\overline{rx}_{t+1}$	S	Components of Slope, S							Wold Slope	
		$S^{(1)}$	$S^{(2)}$	$S^{(3)}$	$S^{(4)}$	$S^{(5)}$	$S^{(6)}$	$\pi^{(6)}$		
(a)	2.09									0.10
	(2.67)									
(b)		-0.99	-1.12	-0.87	-0.53	5.04	2.49	0.10		0.15
		(-0.71)	(-0.69)	(-0.42)	(-0.33)	(3.00)	(2.65)	(0.05)		
(c)		-0.99	-1.12	-0.87	-0.53	5.04	2.49			0.15
		(-0.71)	(-0.69)	(-0.42)	(-0.33)	(3.00)	(2.63)			
(d)						5.01	2.48			0.16
						(2.96)	(2.64)			
(e)									3.02	0.14
									(3.11)	

Panel C: Horse race against factors										
Forecasted Variable	Predictors									R^2
$\overline{rx}_{t+1}$	Wold factors		CPo (2015)		CP (2005)					
	Wold Level	Wold Slope	Level Cycle	Slope	f_t^1	f_t^2	f_t^3	f_t^4	f_t^5	
(a)	5.04	2.89								0.38
	(6.25)	(3.35)								
(b)					-3.35	6.69	-6.15	-1.27	4.88	0.26
					(-3.44)	(0.90)	(-0.27)	(-0.04)	(0.37)	
(c)			1.41							0.10
			(3.45)							
(d)	4.61		0.56							0.27
	(5.56)		(1.56)							
(e)			2.72	4.06						0.38
			(6.50)	(6.45)						

Table 2 Forecasting Average Excess Returns to Treasury Bonds: 1971:M1 to 2016:M12 sample. Excess returns of Treasury bonds are regressed on the month-t values of the first three principal components of the Treasury term structure. The table reports parameter estimates, Newey and West (1987) corrected t-statistics with lag order 18 months for the individual coefficients (in parentheses), and R^2s for the forecasting regressions. A constant is always included in the regression even though its estimate is not reported in the table. Wold Level in Panel A denotes the single level cycle obtained by summing the level components in specification (d), i.e. Wold Level $= L^{(1)} + L^{(2)} + L^{(3)} + L^{(5)}$. Analogously, Wold Slope in Panel B denotes the single slope cycle obtained by summing the slope components in specification (d), i.e. Wold Slope $= S^{(5)} + S^{(6)}$.

4 Conclusions

We provide an orthogonal decomposition of x_t into uncorrelated components $g_t^{(j)}$ that are associated with different layers of persistence. This decomposition, which we dub *Extended Wold Decomposition*, results from the application of the Abstract Wold Theorem on the Hilbert space $\mathcal{H}_t(\boldsymbol{\varepsilon})$, spanned by the classical Wold innovations of $\mathbf{x}$, where the scaling operator $\mathbf{R}$ is isometric. We also show how to construct a stationary time series starting from the law of motion of the components $g_t^{(j)}$ defined over different time scales.

From a technical perspective, our Wold-type decomposition of weakly stationary stochastic processes is the outcome of a multiresolution analysis of an abstract Hilbert space and it is connected to the work by Baggett et al. (2010), where the authors study the wavelet decompositions of Hilbert spaces. Their construction and our approach rely on the isometry associated with the operator: this is key to permit the use of the Abstract Wold Theorem (see Nagy et al. 2010).

From an empirical perspective, we see our Extended Wold Decomposition as a useful tool for the analysis of macroeconomic and financial time series that are commonly affected by shocks with different frequencies. Using two applications, we have shown that the ability of our Extended Wold Decomposition to separate and zoom-in the different fluctuations which may co-exist in the original time series has important consequences for forecasting, as well as the economic interpretation of the results. In addition, Di Virgilio et al. (2019) provide an application of our theory to optimal asset allocation when security returns are subject to shocks with heterogeneous persistence.

Our Extended Wold Decomposition is based on the estimated moving average coefficients. Thus, uncertainty around the Wold coefficients affect inference on the Wold components. In this paper we fit a VAR of length p on the underlying process to recover its infinite moving average representation. Similar to what is done in

the (Structural) VAR literature to compute confidence intervals for impulse response functions, one can quantify the uncertainty around the Wold components by, e.g., bootstrapping the data using the estimated VAR parameters and the fitted residuals. Alternatively, one could conduct inference directly on the moving average representation of the data (without estimation and inversion of the VAR) along the line of Barnichon and Matthes (2018) and Plagborg-Moller (2019). We view this as an interesting avenue for future research.

References

Addison, P. S. (2002), *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*, CRC Press.

Andersen, T. G. and Bollerslev, T. (1997), 'Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns', *The Journal of Finance* **52**(3), 975–1005.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Baggett, L. W., Larsen, N. S., Packer, J. A., Raeburn, I. and Ramsay, A. (2010), 'Direct limits, multiresolution analyses, and wavelets', *Journal of Functional Analysis* **258**(8), 2714–2738.

Bandi, F., Lo, A. W., Chaudhuri, S. and Tamoni, A. (2019), Spectral factor models, LSE Working Papers, Financial Market Group.

Bandi, F., Perron, B., Tamoni, A. and Tebaldi, C. (2019), 'The scale of predictability', *Journal of Econometrics* **208**(1), 120 – 140.

Barnichon, R. and Matthes, C. (2018), 'Functional Approximation of Impulse Responses', *Journal of Monetary Economics* **99**(C), 41–55.

Calvet, L. E. and Fisher, A. J. (2007), 'Multifrequency news and stock returns', *Journal of Financial Economics* **86**(1), 178–212.

Cerreia-Vioglio, S., Ortu, F., Severino, F. and Tebaldi, C. (2017), Multivariate Wold Decompositions, IGIER Working Paper n. 606.

Cieslak, A. and Povala, P. (2015), 'Expected returns in treasury bonds', *The Review of Financial Studies* **28**(10), 2859–2901.

Cochrane, J. H. and Piazzesi, M. (2005), 'Bond risk premia', *American Economic Review* **95**(1), 138–160.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Di Virgilio, D., Ortu, F., Severino, F. and Tebaldi, C. (2019), Optimal asset allocation with heterogeneous persistent shocks and myopic and intertemporal hedging demand, *in* 'Behavioral Finance: The Coming Of Age', World Scientific, pp. 57–108.

Engle, R. F. and Lee, G. (1999), *A long-run and short-run component model of stock return volatility*, Cointegration, Causality, and Forecasting: Oxford University Press.

Engle, R. F. and Rangel, J. G. (2008), 'The spline-GARCH model for low-frequency volatility and its global macroeconomic causes', *Review of Financial Studies* **21**(3), 1187–1222.

Feunou, B. and Fontaine, J.-S. (2017), 'Bond risk premia and gaussian term structure models', *Management Science* **64**(3), 1413–1439.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004), The MIDAS touch: Mixed Data Sampling Regression Models, CIRANO Working Papers 2004s-20, CIRANO.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006), 'Predicting volatility: getting the most out of return data sampled at different frequencies', *Journal of Econometrics* **131**(1-2), 59–95.

Granger, C. W. J. (1980), 'Long memory relationships and the aggregation of dynamic models', *Journal of Econometrics* **14**(2), 227–238.

Gürkaynak, R. S., Sack, B. and Wright, J. H. (2007), 'The US Treasury yield curve: 1961 to the present', *Journal of Monetary Economics* **54**(8), 2291–2304.

Luenberger, D. G. (1968), *Optimization by vector space methods*, John Wiley and Sons.

Monfort, A. and Pegoraro, F. (2007), 'Switching VARMA term structure models-extended version', *CREST-Laboratoire de Finance et Assurance* .

Müller, U. A., Dacorogna, M., Dave, R. D., Pictet, O. V., Olsen, R. and Ward, J. (1993), Fractals and intrinsic time - a challenge to econometricians, Working Papers 1993-08-16, Olsen and Associates.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions - analyzing the dynamics of market components', *Journal of Empirical Finance* **4**(2), 213–239.

Müller, U. K. and Watson, M. W. (2008), 'Testing models of low-frequency variability', *Econometrica* **76**(5), 979–1016.

Müller, U. K. and Watson, M. W. (2015), Low-frequency econometrics, Technical report, NBER.

Nagy, B. S., Foias, C., Bercovici, H. and Kérchy, L. (2010), *Harmonic analysis of operators on Hilbert space*, Springer.

Newey, W. K. and West, K. D. (1987), 'Hypothesis testing with efficient method of moments estimation', *International Economic Review* pp. 777–787.

Ortu, F., Tamoni, A. and Tebaldi, C. (2013), 'Long-run risk and the persistence of consumption shocks', *Review of Financial Studies* **26**(11), 2876–2915.

Plagborg-Moller, M. (2019), 'Bayesian inference on structural impulse response functions', *Quantitative Economics* **10**(1), 145–184.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/QE926*

Rebonato, R. and Hatano, T. (2018), 'The economic origin of treasury excess returns: A cycles and trend explanation'.

Robinson, P. M. (1978), 'Statistical inference for a random coefficient autoregressive model', *Scandinavian Journal of Statistics* **5**(3), 163–168.

Severino, F. (2016), 'Isometric operators on Hilbert spaces and Wold decomposition of stationary time series', *Decisions in Economics and Finance* **39**(2), 203–234.

Wold, H. (1938), *A study in the analysis of stationary time series*, Almqvist and Wiksells Boktryckeri.