WORKING PAPER SERIES

The scale of predictability

F.M. Bandi,B. Perron,A. Tamoni,C. Tebaldi

Working Paper n. 509

This Version: December 19, 2013

The scale of predictability[*]

F.M. Bandi,[†]B. Perron,[‡]A. Tamoni,[§]C. Tebaldi[¶]

First version: November 2012. This version: December 19, 2013

Abstract

We view economic time series as the result of a cascade of shocks occurring at different times and different frequencies (*scales*). We suggest that economic relations that are found to be elusive when using raw data may hold true for different layers (*details*) in the cascade of economic shocks. This observation leads to a notion of a *scale-specific predictability*. Using direct extraction of the details and two-way aggregation, we provide strong evidence of risk compensations in market returns, as well as of an unusually clear link between macroeconomic uncertainty and uncertainty in financial markets, at frequencies lower than the business cycle.

JEL classification: C22, E32, E44, G12, G17
Keywords: long run, predictability, aggregation, risk-return trade-off, Fisher hypothesis.

[†]Johns Hopkins University and Edhec-Risk Institute. Correspondence to: 100 International Drive, Baltimore MD, 21202, USA. E-mail: fbandi1@jhu.edu

[‡]Université de Montrèal, Department of Economics and CIRANO: Correspondence to: 3150 Jean-Brillant Street, Montreal QC H3T 1N8, Canada. E-mail: benoit.perron@umontreal.ca.

[§]London School of Economics, Department of Finance. Correspondence to: Houghton Street, WC2A 2AE London, UK. E-mail: a.g.tamoni@lse.ac.uk.

[¶]Universita' Bocconi. Correspondence to: Via Roentgen 1, Milano, Italy. E-mail: claudio.tebaldi@unibocconi.it.

1 Introduction

Low-frequency economic shocks may not just be long-run averages of high-frequency shocks. Similarly, low-frequency economic relations may not imply analogous relations at higher frequencies.

In agreement with this premise, we argue that every frequency of observation may, in general, be impacted by specific shocks and can, in consequence, carry unique information about the validity of economic relations. While such relations are typically tested - and often rejected - at a specific, high frequency, they may be valid at lower frequencies without requiring the researcher to impose, coherently with data, tight constraints at the highest frequency of observation for internal consistency.

To capture these ideas parsimoniously, we introduce an novel way to model economic time series leading to a new notion of *scale-specific predictability*. In essence, we view economic time series as a sum of scale-specific components, or *details*. We define the details as elements of the observed time series with a specific level of resolution. Higher scales are associated with lower resolution, lower frequencies, and higher calendar-time persistence. Higher scales are, also, affected by shocks which are relatively smaller in size but persist in the system relatively longer, as is typical of long-run shocks.

Often-studied economic relations - like the presumed dependence between market risk-premia and expected volatility (risk-return trade-offs) or between nominal rates and expected inflation (Fisher's effects) - may be hard to detect when using the raw series themselves. In the framework we propose, however, otherwise-elusive economic relations are found to apply to specific frequencies, levels of resolutions, or - in our jargon - *scales*.

We show that direct extraction of the details from observed regressands and regressors can shed important light on the validity of the assumed economic restrictions. Since economic data are viewed as aggregates of a cascade of scale-specific shocks with different sizes and different half-lives, focusing on individual elements of a time series with specific levels of resolution provides us with a suitable way to disaggregate information occurring at different frequencies. This, in turn, gives us a methodology to zoom in on to specific layers of the cascade of shocks affecting the system at different frequencies, isolate each layer, and identify those layers over which economic restrictions are more likely to be satisfied. To this extent, we find that the pattern of predictability in the details often reaches a peak corresponding to scales associated with business-cycle frequencies or

lower. At these scales - but only at these scales - the corresponding slope estimates have signs and magnitudes which are consistent with classical economic logic.

The paper shows that lagged values of the market return variance's detail with decade-long periodicity (or lower) forecast future values of the corresponding variance detail as well as future values of the excess market return's detail with the same periodicity. In essence, we provide evidence for an extremely slow-moving component in market variance which predicts itself and predicts a similarly slow-moving component of the market's excess returns, i.e., a scale-specific risk-return trade-off. Interestingly, the same finding applies to consumption variance. The consumption variance's detail with decade-long periodicity is positively autocorrelated and predicts future values of the excess market return's detail with the same periodicity. Said differently, higher past values of a slow-moving component of consumption variance predicts higher future values of the corresponding slow-moving component in market returns because it predicts higher future values of itself (and this higher variance component should, in agreement with classical economic theory, be compensated). This is, again, a low-frequency risk-return trade-off, i.e., a risk-return trade-off on selected details. While market variance and consumption variance are hardly correlated in the raw data, the low-frequency details delivering predictability have the same periodicity and, importantly, a correlation close to 90%. This finding establishes an extremely close link between macroeconomic uncertainty and uncertainty in financial markets, thereby lending support to sensible economic logic often contradicted by elusive empirical findings on the subject. Importantly, both in the case of market variance and in the case of consumption variance, when evaluating risk-return trade-offs by running predictive regressions on the 10-year details, rather than on the original series, we find R^2 values of 75% (for market variance) and 84% (for consumption variance). These are figures hardly seen in classical assessments of predictability. We deduce that short-run shocks hide equilibrium relations which careful signal extraction can bring to light.

What are the implications of scale-wise predictability for conducting long-run predictive analysis? We show formally that *two-way* (*forward* for the regressand, *backward* for the regressor) adaptive aggregation of the series, as suggested by Bandi and Perron (2008), leads to increased predictive ability precisely at frequencies corresponding to a scale, or level of resolution, over which the economic relation is more likely to apply. We demonstrate that aggregation works as a low-pass filter capable of eliminating high-frequency shocks, or short-term noise, while highlighting the low-frequency details to which economic restrictions apply. In this sense, finding increasing pre-

dictability upon forward/backward aggregation, as in the case of the long-run risk-return trade-offs illustrated by Bandi and Perron (2008), is symptomatic of risk compensations which apply to highly persistent, low-frequency *details* of returns and variance. When testing the restrictions on disaggregated raw data, such components are hidden by noisier high-frequency details. Their signal, however, dominates short-term noise when two-way aggregation is brought to data. Conversely, aggregation provides a way to make scale-specific predictability operational. We find that two-way aggregation of *both* the regressor and the regressand, rather than simple aggregation of the regressand, yields predictability at horizons corresponding with the scale(s) over which scale-specific predictability occurs. Because there is a close one-to-one map between predictability on the details and predictability upon two-way aggregation, the latter provides an operational way to make predictability on the details implementable in practical contexts (asset allocation being a classical example).

Figure 1 and 2 provide illustrations of these ideas. The left panels in Figure 1 present scatter plots of *forward* aggregates of excess market returns on *backward* aggregates of market variance for different levels of aggregation. The right panels present scatter plots of components (details) of the same series corresponding to analogous frequencies between one and two years ($j = 1$), two and four years ($j = 2$), and 8 and 16 years ($j = 4$), respectively. Figure 2 provides the same information for consumption variance. Predictability on the details at scale $j = 4$ (bottom right panels) translates into predictability upon two-way aggregation provided aggregation is conducted over analogous horizons (bottom left panels). The former (predictability on the details) amounts to a spectral feature of the two series of interest, one that carries important economic content in that it directly relates frequency, or scale, to predictable variation in returns. The latter (predictability upon forward/backward aggregation) is a detection tool based on raw data and, importantly for our purposes, it may be viewed a way to translate scale-specific predictability into return predictability for the long run, with all of its applied implications.

[Insert Figure 1 about here]

[Insert Figure 2 about here]

The separation of a time series in terms of details with different levels of resolution is conducted using wavelet methods as in Multiresolution Analysis (see, e.g., Mallat (1989), Dijkerman and Mazumdar (1994), Yazici and Kashyap (1997)).

Our use of wavelets is, however, solely intended to facilitate extraction of scale-specific information. Differently from the literature on wavelets and its reliance on traditional time series representations,[1] once extracted, the components are thought to be driven by time and scale-specific shocks which are not necessarily aggregates of short-term shocks. Hence, importantly, our proposed data generating process departs from classical Wold specifications in which low-frequency shocks are linear combinations of high-frequency shocks by allowing for specificities in the system's shocks across both time and scale. This new data generating process has proven successful in the context of structural consumption models to explain the market risk premium (Ortu, Tamoni, and Tebaldi (2013) and Tamoni (2011)) and provide an alternative view of cross-sectional asset pricing by virtue of a novel notion of scale-specific beta (Bandi and Tamoni, 2013). Here, we employ a scale-time process to broaden the scope and nature of tests of economic restrictions (with an emphasis on predictive relations) and introduce a new approach for modelling and testing these restrictions. To this extent, we study formally the dual role of suitable aggregation in detecting scale-specific economic restrictions and translating them into operational long-run features of the data.

The evaluation of low-frequency contributions to economic and financial time series has a long history, one which we can not attempt to review here. Barring fundamental methodological and conceptual differences having to do with our assumed data generating process, the approach adopted in this paper shares features with successful existing approaches. Consistent with band spectral methods (Hannan, 1963), the frequency dimension is important for our purposes. In light of the predictive nature of this study, time adaptation and localization in both frequency and time are, however, crucial. Band spectral methods only guarantee the latter and, by applying to traditional time series formulations, they are silent about the role of time and scale in the evaluation of the impact of economic shocks, which is something that we emphasize. As in Beveridge and Nelson (1981), who popularized time-series decompositions into stochastic trends and transitory components, we can view the details as components (more than two, in this paper) with different levels of (calendar-time) persistence operating at different frequencies. In our framework, the components' shocks are, again, functions of both time and scale. Comin and Gertler (2006) argue that the common practice, in business-cycle research, of including longer than 8-year oscillations into the trend (see e.g., Baxter and King, 1999), thereby effectively removing them from the analysis, may be associated with

---

[1]For stimulating treatments of wavelet methods for time series analysis, we refer the reader to Percival and Walden (2000) and Gençay, Selçuk, and Whitcher (2001).

significant loss of information. We aim at capturing analogous effects. While Comin and Gertler (2006) decompose a series into a "high-frequency" component between 2 and 32 quarters and a "medium-frequency" component between 32 and 200 quarters, our detail extraction allows us to disentangle multiple driving forces associated with different persistence levels within their assumed frequencies. Moreover, Comin and Gertler (2006) use a band-pass filter which amounts to a two-sided moving average and appears, therefore, not suitable for some of our purposes, i.e., to study predictability. In contrast, the method we propose is one-sided into the past, adapted to current information, and implementable in real time. As in Hansen and Sheinkman (2009), we employ operators to extract low-frequency information (in our case, the low-frequency information embedded in the details). Finally, *essential* scale-wise information in the extracted details can be summarized by a finite number of "typical" points, the result of an econometric process called "decimation". Figure 1 and 2, right panels, are constructed using them. These typical points can be viewed as being akin to "the small number of data averages" used by Müller and Watson (2008) to identify low-frequency information in the raw data. In our case, however, they are scale-specific and, as such, particularly useful to formalize our notion of frequency-specific, or scale-specific, predictability.

The work on stock-return prediction is broad[2] and has led to some controversy (e.g., Cochrane, 2008, for a well-known defense of predictability and references). While it is generally accepted that long-run prediction is more successful than short-run prediction, both are viewed as "reflections of a single underlying phenomenon" (Cochrane, 2001). This paper decouples short-run shocks from long-run shocks and offers an alternative mechanism through which predictability may arise, at certain frequencies *alone*, due to the presence of interconnected layers in the flow of economic shocks. It is this mechanism that justifies our notion of scale-wise predictability.

We proceed as follows. Section 2 provides intuition for the analysis of time series with different levels of resolution. We show how the data can be viewed as a collection of time-specific and frequency-specific shocks, as in a generalized Wold decomposition, or - equivalently - as the sum of details operating at different frequencies. Section 3 introduces scale-specific predictability, the idea that economic relations may hold true for individual layers in the cascade of shocks affecting the economy and, hence, for individual details, but may be hidden by high-frequency perturbations.

---

[2]The literature documents predictability induced by financial ratios, see e.g. Campbell and Shiller (1988), Lamont (1998), Kelly and Pruitt (2013), interest rate variables, see e.g. Fama and Schwert (1977), Fama and French (1989) and macroeconomic variables, see e.g. Lettau and Ludvigson (2001), Menzly, Santos, and Veronesi (2004), Nelson (1976), Campbell and Vuolteenaho (2004).

This section discusses the double role of low-pass filters based on two-way aggregation: detection of the level of resolution over which scale-wise predictability plays a role and operationalization of scale-wise predictability to forecast long-run returns. Section 4 employs direct extraction of the details, as well as aggregation, to provide strong evidence of long-run risk-return trade-offs in market returns. We focus on both low-frequency risk consumptions due to market variance and low-frequency risk compensations due to consumption variance. Their close link, and overall coherence, is established by providing significant evidence about the scale-wise co-movement between macroeconomic uncertainty, as captured by consumption variance, and uncertainty in financial markets. In Section 5 we verify the assumptions of theory using simulations. Specifically, we impose predictability on the "typical" points of certain details, re-construct the original series from these "typical" points, and show effectiveness of two-way aggregation in the identification of low-frequency economic relations. Section 6 turns to another well-know economic relation, namely Fisher hypothesis. Our interest in Fisher hypothesis is two-fold. First, we wish to show applicability of the methods to broad classes of economic relations. Second, contrary to risk-return trade-offs, the case of Fisher effects is one for which maximum predictability is not obtained over the very long haul. It is, instead, achieved over an horizon of about 8 years and is associated with tent-shaped behavior in the predictive slopes and R^2s. We show, using simulations and formal derivations, that this tent-like behavior represents an important implication of our assumed data generating process. Section 7 concludes. Technical details are in the Appendices.

2 Time-series modelling with multiple scales

Consider a weakly-stationary time series $\{x_{t-i}\}_{i \in \mathbb{Z}}$. We may write

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

where the $x^{(j)}$s are components (or *details*) associated with time (t) and *scale* (j) and $\pi_t^{(J)}$ is a long-run trend. The collection of values $x_t^{(j)}$ with j fixed corresponds to the representation of a time-series (viewed as a function of time t) at the jth scale.

While alternative choices are possible, in this paper we use Haar wavelets to decompose $\{x_{t-i}\}_{i \in \mathbb{Z}}$ into details. The use of Haar wavelets provides a clear connection between aggregation (and long-

7

run dynamics) and time-series components with low levels of resolution and high calendar-time persistence. We formally explore the link between Haar wavelets and aggregation in the next section. In general, however, any wavelet can be viewed as an aggregation scheme (e.g., Abry, Veitch, and Flandrin (1998)).

Consider the case $J = 1$. We have

$$x_t = \underbrace{\frac{x_t - x_{t-1}}{2}}_{x_t^{(1)}} + \underbrace{\left[\frac{x_t + x_{t-1}}{2}\right]}_{\pi_t^{(1)}},$$

which amounts to breaking the time series into a transitory and a persistent component. Set, now, $J = 2$. We obtain

$$x_t = \underbrace{\frac{x_t - x_{t-1}}{2}}_{x_t^{(1)}} + \underbrace{\left[\frac{x_t + x_{t-1} - x_{t-2} - x_{t-3}}{4}\right]}_{x_t^{(2)}} + \underbrace{\left[\frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}\right]}_{\pi_t^{(2)}},$$

which further separates the persistent component $\pi_t^{(1)}$ into an additional transitory and an additional persistent component.

The procedure can, of course, be iterated yielding a general expression for the detail $x_t^{(j)}$, i.e.,

$$x_t^{(j)} = \underbrace{\frac{\sum_{i=0}^{2^{(j-1)}-1} x_{t-i}}{2^{(j-1)}}}_{\pi_t^{(j-1)}} - \underbrace{\frac{\sum_{i=0}^{2^j - 1} x_{t-i}}{2^j}}_{\pi_t^{(j)}}$$

where the element $\pi_t^{(j)}$ satisfies the recursion

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

In essence, the time series can be written as a collection of details $x_t^{(j)}$ with different degrees of resolution (i.e., calendar-time persistence) along with a low-resolution approximation $\pi_t^{(J)}$. Equivalently, it can be written as a telescopic sum

$$x_t = \sum_{j=1}^{J} \underbrace{\left\{ \pi_t^{(j-1)} - \pi_t^{(j)} \right\}}_{x_t^{(j)}} + \pi_t^{(J)} = \pi_t^{(0)}, \tag{2}$$

in which the details are naturally viewed as changes in information between scale 2^{j-1} and scale 2^j. The scales are dyadic and, therefore, enlarge with j. The higher j, the lower the level of resolution, and the larger the scale. In particular, the innovations $x_t^{(j)} = \pi_t^{(j-1)} - \pi_t^{(j)}$ become smoother, and more persistent in calendar time, as j increases. The representation in Eq. (2) will be especially useful when discussing aggregation.

2.1 Decimation

Decimation is the process of defining *non-redundant* information, as contained in a suitable number of "typical" points, in the observed details. Returning to Figure 1 and Figure 2, the panels on the right-hand side are constructed from these "typical" points and, therefore, only contain *essential* information about the corresponding scale.

Let us return to the case $J = 2$, as in the example above, but similar considerations apply more generally. Define the vector

$$X_t = [x_t, x_{t-1}, x_{t-2}, x_{t-3}]^\mathsf{T}$$

and consider the orthogonal transform matrix

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

It is easy to see that $\mathcal{T}^{(2)}(\mathcal{T}^{(2)})^\top$ is diagonal and

$$\mathcal{T}^{(2)} X_t = \left[\pi_t^{(2)}, \; x_t^{(2)}, \; x_t^{(1)}, \; x_{t-2}^{(1)} \right]^\mathsf{T}.$$

By letting time t vary in the set $\{ t = k2^2 \text{ with } k \in \mathbb{Z} \}$ one can now define (from $\mathcal{T}^{(2)} X_t$) the

decimated counterparts of the calendar-time details, namely

$$\left\{ x_t^{(j)}, \ t = k2^j \ \text{with} \ k \in \mathbb{Z} \right\} \ \text{for} \ j = 1, 2$$

and

$$\left\{ \pi_t^{(2)}, \ t = k2^2 \ \text{with} \ k \in \mathbb{Z} \right\}.$$

Mallat (1989) provides a recursive algorithm for the general case with J not necessarily equal to 2.[3]

2.2 Scale-specific shocks

Assume, without loss of generality and for convenience, that the time series $\{x_{t-i}\}_{i \in \mathbb{Z}}$ is mean zero. The details $x^{(j)}$ are also mean zero and weakly-stationary for any fixed j (Wong, 1993). One can write the representation (understood in the mean-squared sense):

$$x_t = \sum_{j=1}^{J} \sum_{k=0}^{\infty} a_{j,k} \varepsilon_{t-k2^j}^{(j)}, \tag{3}$$

where $\varepsilon_t^{(j)} = x_t^{(j)} - \mathcal{P}_{\mathcal{M}_{j,t-2^j}} x_t^{(j)}$ and $\mathcal{P}_{\mathcal{M}_{j,t-2^j}}$ is a projection mapping onto the closed subspace $\mathcal{M}_{j,t-2^j}$ spanned by $\left\{ x_{t-k2^j}^{(j)} \right\}_{k \in \mathbb{Z}}$. This is a Wold representation which applies to every scale. Specifically, it is a decomposition which explicitly represents the time series of interest as a linear combination of shocks classified on the basis of their arrival time, as is typical in the analysis of linear stationary time-series, as well as their scale. The decomposition will reduce to a classical Wold representation for linear stationary processes *if*

$$\underbrace{\varepsilon_t^{(j)}}_{x_t^{(j)} - \mathcal{P}_{\mathcal{M}_{j,t-2^j}} x_t^{(j)}} = \underbrace{\left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right) / \sqrt{2^j}}_{\sum_{i=0}^{2^j-1} \left(x_{t-i} - \mathcal{P}_{\mathcal{M}_{t-i-1}} x_{t-i} \right)}.$$

Appendix A provides a proof of this result. Intuitively, if the scale-specific innovations are sums of high-frequency innovations, then the information contained by the series at every scale is an

---

[3]In general, we can use the components $x_t^{(j)}$, $j = 1, .., J$, and $\pi_t^{(J)}$ in their entirety to reconstruct the time series using (1). This is the *redundant* decomposition of a time series proposed in Renaud, Starck, and Murtagh (2005). Alternatively, one can reconstruct the time series signal from the decimated components using the (inverse of the) Haar unitary matrix, see Appendix B.

aggregate of that contained at higher frequencies. If, on the other hand, this is not the case, then there is a separation between scales - in terms of their informational content - which still preserves their consistency. This would translate into shocks which are specific to individual scales, thereby giving meaning to economic relations which, again, may only be satisfied at certain frequencies. The logic behind this approach to time series modelling, and its implications for detecting short and long-run dynamics in economic time series, is further described in Appendix B.

Consistent with Eq. (1) and Eq. (3), one may now represent the details as linear autoregressive processes. A convenient (2-parameter) way to do so is to write

$$x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \varepsilon_t^{(j)} \tag{4}$$

or, in terms of decimated observations (for an integer k),

$$x_{k2^j}^{(j)} = \rho_j x_{k2^j-2^j}^{(j)} + \varepsilon_{k2^j}^{(j)} \tag{5}$$

where the shocks $\varepsilon^{(j)}$ are uncorrelated across scales, white noise, and with a scale-specific variance σ_j. The model is autoregressive of order 1 in the dilated time of the scale being considered. The parameter ρ_j captures scale-specific persistence. We note that dependence in scale time can be considerably lower than dependence in calendar time, the later being an increasing function of the scale (Appendix C, Subsection C.2.1 for a formal proof). We also note that the assumption of uncorrelatedness across scales is due to the ability of a filter like $\mathcal{T}^{(2)}$ to "decorrelate" the original observations (see, e.g., Dijkerman and Mazumdar (1994) and Gençay, Selçuk, and Whitcher (2001)). Both properties, i.e., low correlation in scale time and uncorrelatedness across scales, will be verified in the data. While richer autoregressive specifications may be employed in the usual way, a parsimonious structure in scale time, consistent to the one in Eq. (5), is bound to lead, upon re-construction of the raw series, to rich dynamics in calendar time. Motivated by issues of signal processing akin to the economic issues of interest to us, Dijkerman and Mazumdar (1994) propose an analogous representation.

Eq. (3) and Eq. (5) are important. The former describes the idea that economic time series can be represented as aggregates of shocks that are both time and scale specific. For any scale indexed by j, the latter defines a parsimonious way in which scale-specific shocks impact the corresponding

layer of the time series of interest. We note that, given Eq. (5), by choosing the pair $\{\rho_j, \sigma_j\}$ for each j, one can readily view the specification on the scales as a novel data generating process for the original time series, resulting in (a form of) Eq. (3). The original time series can, in fact, be re-constructed from the details by using the properties of the matrix $\mathcal{T}^{(J)}$ (Mallat, 1989). Appendix B and simulations (below) provide further details. From an inferential standpoint, $\{\rho_j, \sigma_j\}$ can be readily identified once the details have been extracted.

In what follows, we begin with the extraction of the details and the analysis of their dynamics. We will then investigate the relation between details of suitable predictors and regressands in order to establish the presence, or lack thereof, of our notion of scale-wise predictability. We argue that scale-wise predictability is a channel through which economic restrictions may be satisfied at particular levels of resolution without having to be satisfied at *all* levels of resolution. By imposing Eq. (5), we will then simulate a model with multiple scales and show its effectiveness in replicating important stylized findings in the data. Before doing so, however, we turn to the role of aggregation in revealing predictability on the scales and, importantly, in providing a way to exploit it.

3   Two-way aggregation and scale-specific predictability

Consider a predictive variable y_t and a predictor x_t. It is standard in macroeconomics and finance to verify predictability by computing linear, or nonlinear, projections at the highest frequency of observation. It is also common to aggregate the regressand. A recent approach proposed by Bandi and Perron (2008) aggregates *both* the regressand (forward) and the regressor (backwards). The aggregate regressor is adapted to time t information and is, therefore, non anticipative. The logic for aggregating both the regressand and the regressor resides in the intuition according to which equilibrium implications of economic models may impact highly persistent components of the variables $\{y_t, x_t\}$ while being hidden by short-term noise (Bandi and Perron, 2008). Aggregation provides a natural way to filter out noise, thereby yielding a cleaner signal. We now formalize this intuition.

Assume the following predictive model postulated in terms of details:

$$y_{t+2^j}^{(j)} = \alpha + \beta x_t^{(j)} \qquad \text{for } j > s.$$

In this specification, the assumed relation applies to each jth scale (with $j > s$), but may be confounded by the presence of uncorrelated shocks at other scales. Said differently, the (scale) time $t^* + 1$ value of the jth scale of the variable y is linearly related to the time t^* value of the jth scale of the variable x. Since time is scale specific and dilated, the time $t^* + 1$ value of the jth scale is expressed as $t + 2^j$ in calendar time units.

Let us begin with a preliminary observation. Aggregation of the time series $\{x_{t-i}\}_{i \in \mathbb{Z}}$ uncovers information at different scales or, more precisely, for scales that are higher than the one corresponding to the aggregation level. To see this, using Eq. (2), write

$$x_{t-2^s+1,t} = \left(\sum_{i=0}^{2^s-1} x_{t-i} \right) / 2^s = \pi_t^{(s)} = \sum_{j=s+1}^{J} \underbrace{\left\{ \pi_t^{(j-1)} - \pi_t^{(j)} \right\}}_{x_t^{(j)}} + \pi_t^{(J)}. \tag{6}$$

where $s = 0, 1, \ldots, J$. The implication of this simple derivation is that economic relations which emerge from aggregation, and may not appear at higher frequencies, can be viewed as scale-specific.

Using Eq. (6), forward/backward aggregation yields

$$
\begin{aligned}
y_{t+1,t+2^s} &= \left(\sum_{i=0}^{2^s-1} y_{(t+2^s)-i} \right) / 2^s = \pi_{t+2^s}^{(s)} = \\
&= \sum_{j=s+1}^{J} y_{t+2^s}^{(j)} + \pi_{t+2^s}^{(J)} = \underbrace{\sum_{j=s+1}^{J} \left\{ \alpha + \beta x_t^{(j)} \right\} + \pi_{t+2^s}^{(J)}}_{=k+\beta x_{t-2^s+1,t}}.
\end{aligned}
\tag{7}
$$

Thus, predictability on the details implies predictability upon suitable aggregation of both the regressand and the regressor. In essence, economic relations which apply to highly persistent components will be revealed by two-way averaging since higher frequency dynamics will not affect inference.

Eq. (7) provides a clear way to understand the role of suitable (two-way) aggregation. As stated, however, the result hinges on two assumptions: predictability applies to scales $j > s$, where s is the level of aggregation, and the same slope β characterizes all scale-wise predictive regressions. In this sense, the relation is particularly useful to understand the implications of scale-wise predictability over the very long run (i.e., for a large j). In this scenario, in fact, Eq. (7) predicts that long-run forward/backward averaging of the regressand/regressor will lead to predictability upon aggregation.

13

Interestingly, it should also lead to slope estimates which are very closely related, in terms of their numerical value, with the true slopes of the relevant scale-wise predictive regression(s). This case will be very pertinent to understand the long-run risk-return trade-offs reported below.

An alternative scenario is one in which scale-wise predictability applies to a moderate scale, rather than to the very long haul. Under mild assumptions, Appendix C-C.2.2 shows that, unsurprisingly, the optimal amount of averaging should be conducted for time lengths corresponding to the scale over which predictability applies. More specifically, if predictability applies to a specific detail with fluctuations between 2^{j-1} and 2^j periods, the largest R^2 is achieved for a level of forward/backward aggregation corresponding to 2^j periods. Before and after, the R^2s should display a tent-like behavior.

The data generating process that we propose has an additional empirical implication worth mentioning. Under predictability at the same j^{th} scale, should forward predictors ($y_{t+1,t+2^s}$) be regressed on *differences* of aggregated regressors ($x_{t-2^s+1,t} - x_{t-2\times2^s+1,t-2^s}$), rather than on aggregated regressors ($x_{t-2^s+1,t}$), the maximum R^2 would be achieved for a level of aggregation corresponding to 2^{j-1} periods, rather than 2^j periods. Additionally, the sign of the slope estimate would be the opposite of the sign of the true slope linking details at the jth scale (see Appendix C-C.2.4). We will use this additional implication of theory to further validate the consistency between assumed data generating process and empirical findings in Section 6.

Next, we broaden the scope of classical predictability relations in the literature. We focus on risk-return trade-offs. Because of their different features, we later discuss Fisher effects. We first show the outcome of two-way aggregation and predictive regressions run on aggregated raw series. We then turn to regressions on the extracted details and illustrate the consistency of their findings with those obtained from two-way aggregation. This consistency is further confirmed by simulation as well as in the context of the illustrative treatment in Appendix C. From an applied standpoint, one could proceed in the opposite way: detect predictability on the scales and then utilize predictability on the scales by suitably aggregating regressands and regressors. The latter method is a way in which one could exploit the presence of a scale-specific risk-return trade-off to perform return predictability and, among other applications, asset allocation over suitable horizons.

4 Risk-return trade-offs

4.1 Equity returns on *market* variance

The basic observation driving our understanding of the analysis of risk-return trade-offs at different levels of aggregation performed by Bandi and Perron (2008) is the following: the "basis"of independent shocks yielding return time series must be classified along two dimensions: their time of arrival and their scale or level of resolution/persistence.

As discussed above, we propose an adapted linear decomposition which represents a time series as a sum of decorrelated details whose spectrum is concentrated on an interval of characteristic time scales (inverse of frequencies) ranging from 2^{j-1} to 2^j.[4] Each element of the time series x_t is decomposed into a sum of detail components $x_t^{(j)}$ classified by their degree of (scale-wise) persistence ρ_j plus a permanent component $\pi_t^{(\infty)}$. By construction each detail $x_t^{(j)}$ is stationary and mean zero. The permanent component may be viewed as the sum of deterministic and stochastic trend components with infinite persistence.

We apply the decomposition to logarithmic excess returns and realized variance series, r_t and v_t^2.[5] The details are shown in Figure 3. The hypothesis of uncorrelatedness among detail components with different degrees of persistence is not in contradiction with data. Table 2 presents pair-wise correlations between the individual details of market variance and excess market returns. Virtually all correlations are small and very statistically insignificant. Not surprisingly, the largest one (0.39) corresponds to the adjacent pair of variance's scales $j = 3$ and $j = 4$.[6]

[Insert Figure 3 about here]

Next, we consider the forward/backward regressions

$$r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t}^2 + u_{t,t+h}, \tag{8}$$

---

[4]For a clear interpretation of the j-th scale in terms of the corresponding time spans, we refer to Table 1.

[5]Appendix F describes the data and construction of variables.

[6]It is worth emphasizing that these pair-wise correlations are obtained by using *redundant* data on the details rather than the *decimated* counterparts described in Subsection 2.1. This is, of course, due to the need of having the same number of observations for each scale. Hence, even though they are small, we expect these correlations to be slightly upward biased.

There could also be leakage between adjacent time scales. It is possible to reduce the impact of leakage by replacing the Haar filter with alternative filters with superior robustness properties (the Daubechies filter is one example). The investigation of which filter is the most suitable for the purpose of studying predictability on the scales is beyond the scopes of the present paper. As pointed out earlier, also, the use of the Haar filter is particularly helpful to relate scale-wise predictability to aggregation, a core aspect of our treatment.

15

where $r_{t+1,t+h}$ and $v^2_{t-h+1,t}$ are aggregates of excess market returns and return variances over an horizon of length h. Empirical results are reported in Table 6-Panel A. We confirm the findings in Bandi and Perron (2008), namely *future* excess market returns are correlated with *past* market variance. Dependence increases with the horizon, and is strong in the long run, with R^2 values between 7 and 10 years ranging between about 16% and 51%.

A crucial observation in Bandi and Perron (2008) is that "*the long-run results are not compatible with classical short-term risk return trade-offs.*" Proposition 3 in that paper discusses the asymptotic properties of the long-run regressions' slope estimates and R^2s and shows that disaggregated asset pricing models which solely imply dependence between excess market returns and (autoregressive) conditional variance at the highest resolution *cannot* deliver the reported findings upon aggregation. This is easily seen. Consider a classical one-period ($h = 1$) predictive system:

$$
\begin{aligned}
r_{t,t+1} &= \alpha + \beta v^2_{t-1,t} + u_{t,t+1}, \\
v^2_{t,t+1} &= \rho v^2_{t-1,t} + \varepsilon_{t,t+1}.
\end{aligned}
\tag{9}
$$

By invoking a simple recursion, it can be verified that the parameter estimate from Eq. (8) would converge to $\rho^h \beta$ and, hence, to zero as $h \to \infty$. When using return and variance data, instead, we find that the slopes tend to increase with the horizon while becoming more statistically significant (Table 6-Panel A).

In essence, the empirical findings point to an alternative data generating process, one in which low-frequency shocks are not necessarily linear combinations of high frequency shocks and low-frequency dynamics are not simply successive iterations of high-frequency dynamics. To this extent, this paper argues that the relation between risk and return may be viewed as being *scale-specific*. In agreement with the implications of Eq. (7), aggregation is helpful to reveal these low-frequency risk compensations.

To corroborate this logic, we run detail-wise predictive regressions analogous to the classical predictive regressions in Eq. (9), namely

$$
r^{(j)}_{k2^j+2^j} = \beta_j v^{2(j)}_{k2^j} + u^{(j)}_{k2^j+2^j}
\tag{10}
$$

with $j = 1, ..., 4$ and k an integer. Similarly, we run

$$v^{2(j)}_{k2^j+2^j} = \rho_j v^{2(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j}. \tag{11}$$

Note that the elements at time-scale j are defined on a time-grid whose time unit is 2^j times the unit scale of the original time series of observations. These are the *decimated* elements of the details described in Subsection 2.1. They can be viewed as containing *essential* information about each scale. In this sense, they have an interpretation that is similar to the "small number of data averages" in Müller and Watson (2008), one that, however, applies to each individual detail rather than to the overall time series as in Müller and Watson (2008).

The results are based on yearly data and are reported in Table 6-Panel B. For a clear interpretation of the corresponding levels of resolution, we refer to Table 1. The strongest predictability is for $j = 4$, which corresponds to economic fluctuations of 8 to 16 years. The importance of this scale relates back to the increased significance of backward-aggregated variance as a predictor of forward-aggregated excess market returns at similar low frequencies (Table 6-Panel A). In fact, consistent with the derivation in Eq. (7), aggregation begins to reveal scale-wise predictability over an horizon (s) of about 7 years and, as documented, scale-wise predictability applies to a scale j satisfying $j > s$. For an explicit graphical representation based on scatter plots, we refer the reader back to Figure 1.

For $j = 4$, the R^2 on the detail-wise predictive regression is a considerable 74%. The R^2 on the detail-wise variance autoregression is 16%. For both regressions, the slope is positive. In the case of the detail-wise predictive regression, its value is - coherently, again, with theory in Eq. (7) - similar to that obtained from two-way aggregation (about 1.5). As for the autoregressive variance coefficient, while its numerical value appears small, we recall that it is a measure of correlation on the dilated time of a scale designed to capture economic fluctuations with 8 to 16 year cycles.

In essence, we find that, at scale $j = 4$, a very slow-moving component of the variance process predicts itself as well as the corresponding component in future excess market returns. Said differently, higher past values of a variance detail predict higher future values of the same variance detail and, consequently, higher future values of the corresponding detail in excess market returns, as required by conventional logic behind risk compensations. While this logic applies to a specific level of resolution in our framework, it translates - upon aggregation - into predictability in long-run

17

returns as shown formally (in Section 3) and in the data.

We now turn to consumption variance.

4.2 Equity returns on *consumption* variance

Replacing market variance with consumption variance, as justified structurally by Tamoni (2011), does not modify the previous results. If anything, it reinforces previous findings.[7]

We start off with two-way aggregation (Table 7-Panel A). The R^2 values between 7 and 10 years range between about 22% and 53%. Running detail-wise predictive regressions leads to maximum predictability (and an R^2 of 83%) associated with low-frequency cycles of about a decade on average, i.e., $j = 4$ (Table 7-Panel B). Similarly, for $j = 4$, a detail-wise autoregression of future consumption variance on past consumption variance yields a positive autocorrelation of 0.18 and an R^2 value of about 50%.

Again, consistent with theory, aggregation begins to uncover detail-wise predictability at horizons (7 years, in this case) just below the time length over which predictability on the details operates (between 8 and 16 years). In addition, aggregation leads to slope estimates which are very closely related, in terms of their numerical value, with the slope estimates of the corresponding scale-wise predictive regressions (about 3.5), c.f. Table 7-Panels A and B. Figure 2 provides a graphical representation.

In sum, because it predicts itself, a slow-moving component of consumption variance has forecasting ability for the corresponding slow-moving component of excess market returns. This finding points, once more, to a low-frequency risk compensation in market returns, one that - however - now operates through the economically-appealing channel of consumption risk.

4.3 The relation between market variance and consumption variance

These observations raise an important issue having to do with the relation between uncertainty in financial markets and macroeconomic uncertainty, as captured by consumption variance. Barring small differences, when exploring suitable scales, both variance notions have predictive power for excess market returns on the details. Similarly, they both have predictive power for long-run returns upon adaptive (two-way) aggregation.

---

[7]We note that the "decorrelation" property of the details applies to consumption variance very strongly (see Table 2).

While this result appears theoretical justifiable since there should be, in equilibrium, a close relation between consumption variance and market variance (see, e.g., Eq. (12) in Bollerslev, Tauchen, and Zhou, 2009, for a recent treatment), the empirical relation between these two notions of uncertainty is well-known to be extremely mild, at best. In an influential paper on the subject, Schwert (1989) finds a rather limited link between macroeconomic uncertainty and financial market variance. This work has spurred a number of contributions which, also, have provided evidence that the relation between variance in financial markets and a more "fundamental" notion of variance is extremely weak in US data (see, e.g., the discussion in Diebold and Yilmaz, 2008).

We argue that this statistical outcome may not be as counter-intuitive as generally believed. Specifically, it may be due to variance comparisons which focus on high frequencies. Schwert (1989), for instance, uses monthly data from 1857 to 1987. We conjecture that, being the result of equilibrium conditions, the presumed relation between macroeconomic variance and financial market variance may not occur at high frequencies and may, therefore, be irreparably confounded in the raw data. Using our jargon, the relation could, however, hold true for suitable lower frequency details of both variance processes. Figure 4-Panels A and B provides graphical representations supporting this logic. The upper panel relates market variance to the variance of consumption growth using yearly data. The lower panel looks at the link between the details of the two series with scale $j = 4$, i.e., the details capturing economic fluctuations between 8 and 16 years. The relation between the raw series is extremely mild, the correlation being about 0.05. The details are, instead, very strongly co-moving. Their estimated correlation is around 90%.

[**Insert Figure 4 about here**]

A large, successful literature has examined the validity of classical risk-return relations by refining the way in which conditional means and conditional variances are identified. Similarly, a large, equally successful literature has studied the properties of financial market volatility and, in some instances, looked for significant associations, dictated by theory, between macroeconomic uncertainty and uncertainty in financial markets. This paper addresses both issues by taking a unified view of the problem, one which emphasizes the role played by low-frequency shocks. We argue that equilibrium relations, the one between future excess market returns and past consumption/market variance or the one between contemporaneous market variance and contemporaneous consumption variance, may be satisfied at the level of individual layers of the raw series while being drastically

clouded by high-frequency variation in the data.

5   Simulating scale-specific predictability

One important observation about two-way aggregation is in order. One may argue that, by gener-
ating stochastic trends, aggregation could lead to spurious (in the sense of Granger and Newbold,
1974, and Phillips, 1986) predictability. If this were the case, *contemporaneous* aggregation should
also lead to patterns that are similar to those found with forward/backward aggregation. In all
cases above, one could show that this is not the case.[8] In other words, contemporaneous aggrega-
tion does not lead to any of the effects illustrated above (including consistency between the slope
estimates obtained from the aggregated series and from the details). Following a similar logic, one
could also argue that spurious behavior would prevent a tent-shape pattern from arising in the slope
estimates and R^2 from predictive regressions on the aggregated series because it would simply lead
to (approximate, at least) upward trending behavior in both. Again, this is not the case. Tent-shape
patterns may readily arise as shown formally in Appendix C and in the simulations below. The
study of Fisher effects (in Section 6) provides empirical evidence confirming the latter result.

In this section we establish, by simulation, that scale-wise predictability translates into pre-
dictability upon two-way aggregation. Supporting the implications of theory in Appendix C, we
show that tent-shaped patterns are possible provided, of course, predictability occurs at the appro-
priate scale. We also show that, if predictability on the details applies, contemporaneous aggregation
leads to insignificant outcomes. Similarly, if no predictability on the details applies, two-way ag-
gregation leads to insignificant outcomes. In sum, the findings discussed in this section provide
support for a genuine (close to) 10-year cycle in the predictable variation of the market's risk-return
trade-offs, as reported previously.

We begin by postulating processes for the (possibly related) details of the variance and return
series:

$$v^{2(j)}_{k2^j+2^j} = \rho_j v^{2(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j} \tag{12}$$

$$r^{(j)}_{k2^j+2^j} = v^{2(j)}_{k2^j}$$

---

[8]The corresponding tables are not reported for conciseness but can be provided by the authors upon request. See,
also, Bandi and Perron (2008) for further evidence.

for $j = j^*$ and

$$v^{2(j)}_{k2^j+2^j} = \varepsilon^{(j)}_{k2^j+2^j}$$
$$r^{(j)}_{k2^j+2^j} = u^{(j)}_{k2^j,k2^j+2^j},$$

for $j \neq j^*$, where k is defined as above and $j = 1, \ldots, J = 9$. The shocks $\varepsilon^{(j)}_t$ and $u^{(j)}_t$ satisfy $corr(u^{(j)}_t, \varepsilon^{(j)}_t) = 0 \ \forall t, j$. The model implies a predictive system on the scale j^* and unrelated details for all other scales. In other words, predictability only occurs at the level of the j^{*th} detail.

Here, the scales are defined at the monthly level. Due to the dyadic nature of the scales, this is simply done to gain granularity in the analysis. The data generating process is formulated for "deconstructed" or "decimated" data. One, then, has to recover the raw time series. To do so, we simulate the process at scale j every 2^j steps and multiply it by the inverse Haar Matrix. Appendix C illustrates within a tractable example the simulation procedure in the time-scale domain and the reconstruction steps in the time domain.

In agreement with the discussion in Section 3, we will now show that a predictive relation localized around the j^{*th} scale will produce a pattern of R^2s which has a peak for aggregation levels corresponding to the horizon 2^{j^*} (rather than 2^{j^*-1} or in-between).

5.1   Running the predictive regression

Table 3-Panel A shows the results obtained by running the regression in Eq. (8) on simulated data generated from Eq. (12). We compare these results to those in Table 4, where no scale-wise predictability is assumed.

When imposing the relation at scale $j^* = 6$, i.e., for a time span of 32 to 64 months (c.f., Table 1), we reach a peak in the R^2s of the two-way regressions at 5 years. The 5-year R^2 is about 25 times as large as the one obtained in the case of a spurious regression at the same horizon. Moreover, the slope estimates increase reaching their maximum value at 5 years and approaching the slope's true value on the 6th details of 1 (with some attenuation due to the impact of other scales). After the 5-year mark, the slope estimates decrease almost monotonically. This is a rough tent-shaped pattern which readily derives solely from imposing scale-wise predictability at a frequency lower than business-cycle frequencies but not as low as, say, the 10-year or 120-month frequency (c.f., Appendix C).

If we now impose the relation at scale $j^* = 7$, i.e., for a time span of 64 to 128 months, we would expect the peak in the R^2s from the two-way regressions to shift to about 128 months, again the upper bound of the range of possible horizons given scale $j^* = 7$. Should this horizon also be the upper bound of the horizons of aggregation, we would expect upward trending behavior in the estimated slopes, t-statistics, and R^2. This logic is consistent with the simulations in Table 5 confirming the ability of suitable aggregation to detect scale-wise predictability over the relevant scale.

As emphasized earlier, should aggregation lead, somewhat mechanically, to statistically significant, larger slopes and higher R^2 by virtue of the creation of stochastic trends, tent-shaped behaviors would be unlikely and contemporaneous aggregation would also lead to spurious predictability. We have shown that tent-shaped structures naturally arise from predictability at the corresponding scale. We now turn to contemporaneous aggregation. Again, we consider the cases $j^* = 6$ and $j^* = 7$ (in Table 3 and Table 5-Panel B). When both the regressor and the regressand are aggregated over the same time interval, no predictability is detected. Appendix C-C.2.3 provides a theoretical justification. Appendix D contains additional simulations and diagnostics.

6   Fisher hypothesis

In its simplest form, Fisher hypothesis postulates that the nominal rate of return on assets (interest rates as well as nominal returns on equities, for example) should move one-to-one with expected inflation (Fisher, 1930). The empirical work on the subject is broad and very mixed in terms of findings. Mishkin (1992) and Fisher and Seater (1993), for instance, run regressions of k-period continuously-compounded nominal interest rates (and h-period GDP growth) on contemporaneous h-period *expected* inflation (and the growth of nominal money supply). Boudoukh and Richardson (1993) run regressions of k-period nominal stock returns also on contemporaneous h-period expected inflation. In all cases, it is natural to test whether the slope of the predictive regression is equal to 1, as implied by theory.

Here, we study the relation between nominal rates of returns and inflation by exploring predictability using backward/forward aggregation. We find that, for a suitable horizon h, h-period continuously-compounded nominal returns are strongly correlated with *past h-period* realized inflation. The same logic as that employed for long-run risk-return trade-offs may be applied to explain

these findings. If low-frequency details of the nominal rates are linked to low-frequency details of realized inflation, two-way aggregation will uncover this dependence. Direct extraction of the details would also allow us to zoom in onto individual layers of information. We are now more specific.

The results are reported in Tables 8-Panels A, B, and C. For nominal stock returns, we find a tent-shaped predictability pattern (as aggregation increases) with a peak between 7 and 9 years. Importantly, as predictability increases, the corresponding beta estimates approach the value of 1. The 1-year beta is 0.43 with a t-statistic of 0.57 and an R^2 of 0.81%. The 8-year beta, instead, is equal to 1.03 with a t-statistic of 3.30 and an R^2 of about 28%. Turning to the details, at scale $j = 3$, i.e., for frequencies between 4 and 8 years, the R^2 of the predictive regression has a value of about 30%. Its associated slope estimate is positive (2.47). So, is the autocorrelation coefficient (0.13) associated with the autoregression on the 3rd detail of the inflation process. We recall that the estimated autocorrelation is on "decimated" data. The corresponding calendar-time autocorrelation would be considerably higher.

Analogous findings apply to nominal interest rates (Table 10). With two-way aggregation, the 1-year beta is 0.37 with a t-statistic of 2.92 and an R^2 of about 21%. The 8-year beta is twice as large and equal to 0.70 with a t-statistic of 3.14 and an R^2 of about 35%. The tent-shaped pattern is even more marked than in the previous case. The corresponding details ($j = 3$) yield a predictive regression with a positive slope and an R^2 value of 13.25%.

In sum, a predictable slow-moving component of the inflation process (operating between 4 and 8 years) appears to correlate with slow-moving components of nominal stock returns and interest rates. Higher past values of the $j = 3$ inflation detail predict higher future values of the same detail, as well as higher values of the nominal rates' details, thereby yielding compensation for inflation risk at a low level of resolution. Such a compensation is revealed by aggregation.

Gathering essential information about low-frequency dynamics is inevitably hard. Yet, even though the predictive and autoregressive slopes on the decimated details may not be accurately estimated, the results are striking. Differently from the risk-return trade-offs analyzed in Section 4, which operate at scale $j = 4$, the economic logic underlying Fisher hypothesis appears to be satisfied at scale $j = 3$. In agreement with our formal discussion in Section 3, two-way aggregation should yield maximum predictability over an horizon close to 2^3, i.e., close to 8 years. This is, in fact, consistent with data. It is also consistent with a new set of simulation, calibrated on the data, which assume scale-wise predictability at scale $j = 3$ and find a peak of predictability upon two-way

aggregation precisely at 8 years (Table 8-Panel A2).

As discussed in Section 3 (and shown in Appendix C-C.2.2) predictable variation induced by a detail, like $j = 3$, which is not at the upper bound of those that can be reliably handled given the available span of data and is not capturing the very long run ($j = 4$ in our case), would induce explicit tent-shaped behavior upon aggregation. Another interesting implication of the assumed data generating process is that, if we were to run regressions of forward aggregated nominal returns on differences of backward aggregated inflation (rather than on levels), the maximum level of predictability would occur at the horizon 2^{3-1} rather than at the horizon 2^3. In addition, the slope estimate would be negative, rather than positive (Appendix C-C.2.4). We confirm both implications of theory with data. Table 9-Panel B provides the corresponding results. Leaving the very long run aside (more on this later), the maximum R^2 is obtained for $h = 4$. The estimated slope at this horizon is negative and equal to -1.59. Simulations support these findings. As in Table 8, we assume a data generating process, calibrated on the data, with predictable variation corresponding to the 3rd scale. Again, leaving aside the very long run, the largest R^2 is obtained for $h = 4$. The corresponding estimated slope is also negative and rather close to what is found in the data (-1.22 rather than -1.59). It is interesting to notice that, not only does this model diagnostic provide support for the expected behavior of the assumed scale-based data generating process at $h = 4$, it also delivers long-run outcomes which are consistent with data. Both in the data and in simulation the long-run slopes (horizons between 8 and 10 years) are positive and rather significant. The corresponding R^2s are, also, somewhat larger.

7   Further discussions and conclusions

Shocks to economic time series can be time-specific and, importantly for our purposes, frequency-specific. We suggest that economic relations may apply to individual layers in the cascade of shocks affecting the economy and be hidden by effects at alternative, higher frequencies. These layers, and the frequency at which they operate, can be identified. In particular, the nature and the magnitude of the existing, low-frequency, economic relations can be studied. To do so, this paper proposes *direct* extraction of the time-series details - and regressions on the details - as well as *indirect* extraction by means of two-way aggregation of the raw series - and regressions on forward/backward aggregates of the raw series. The mapping between the two methods is established and their close relation

is exploited empirically. While the direct method allows one to identify, up to the customary estimation error, the data generating process (i.e., the details and, upon reconstruction, the original series), the indirect method provides one with a rather immediate way to evaluate the frequency at which layers in the information flow are connected across economic variables and employ this information for prediction. By providing an alternative way in which one may implement long-run predictability (aggregated regressand on past aggregated regressor, rather than on past regressor over one period), two-way aggregation provides a natural way to exploit scale-specific predictability (in asset allocation for the long run, for example). Using both direct extraction of the details and aggregation, we provide evidence of the long-run validity of certain economic relations (risk-return trade-offs and Fisher's hypothesis) typically found to be elusive when working with raw data at the highest frequency of observation.

The use of variance and inflation as predictors of asset returns is particularly appealing in our framework because the corresponding backward-aggregated measures do not lose their economic interpretation. Backward-aggregated variance and backward-aggregated inflation can readily be interpreted as long-run past variance and long-run past inflation. Having made this point, alternative popular predictors, like the dividend-yield and other financial ratios, may also be employed. While their long-run past averages are not as easily interpretable, the role played by aggregation in the extraction of low-frequency information contained in the details applies generally. So does the proposed approach to predictability. To the extent that market return data and the dividend-yield - for instance - contain relevant information about long-run cash-flow risk, regressions on their details and on properly-aggregated data appear very well-suited to uncover this information. We leave this issue for future work.

References

ABRY, P., D. VEITCH, AND P. FLANDRIN (1998): "Long-Range Dependence: revisiting Aggregation with Wavelets.," *Journal of Time Series Analysis*, 19, 253–266.

BANDI, F., AND A. TAMONI (2013): "Scale-specific risk in the consumption CAPM," *SSRN eLibrary*.

BANDI, F. M., AND B. PERRON (2008): "Long-run risk-return trade-offs," *Journal of Econometrics*, 143(2), 349–374.

BANSAL, R., V. KHATCHATRIAN, AND A. YARON (2005): "Interpretable asset markets?," *European Economic Review*, 49(3), 531 – 560.

BAXTER, M., AND R. G. KING (1999): "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *The Review of Economics and Statistics*, 81(4), pp. 575–593.

BEVERIDGE, S., AND C. R. NELSON (1981): "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*, 7(2), 151–174.

BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22(11), 4463–4492.

BOUDOUKH, J., AND M. RICHARDSON (1993): "Stock Returns and Inflation: A Long-Horizon Perspective," *American Economic Review*, 83(5), 1346–55.

CAMPBELL, J., AND R. SHILLER (1988): "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1(3), 195–228.

CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): "Inflation Illusion and Stock Prices," *American Economic Review*, 94(2), 19–23.

COCHRANE, J. H. (2001): *Asset Pricing.* Princeton University Press.

——— (2008): "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21(4), 1533–1575.

COMIN, D., AND M. GERTLER (2006): "Medium-Term Business Cycles," *American Economic Review*, 96(3), 523–551.

DIEBOLD, F. X., AND K. YILMAZ (2008): "Macroeconomic Volatility and Stock Market Volatility, Worldwide," NBER Working Papers 14269, National Bureau of Economic Research, Inc.

DIJKERMAN, R., AND R. MAZUMDAR (1994): "Wavelet representations of stochastic processes and multiresolution stochastic models," *Signal Processing, IEEE Transactions on*, 42(7), 1640 –1652.

Fama, E. F., and K. R. French (1989): "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25(1), 23–49.

Fama, E. F., and G. W. Schwert (1977): "Asset returns and inflation," *Journal of Financial Economics*, 5(2), 115–146.

Fisher, I. (1930): *"The Theory of Interest"*. New York: The Macmillan Co.

Fisher, M. E., and J. J. Seater (1993): "Long-Run Neutrality and Superneutrality in an ARIMA Framework," *American Economic Review*, 83(3), 402–15.

Gençay, R., F. Selçuk, and B. Whitcher (2001): *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, New York, first edn.

Granger, C. W. J., and P. Newbold (1974): "Spurious regressions in econometrics," *Journal of Econometrics*, 2(2), 111–120.

Hannan, E. J. (1963): "Regression for Time Series with Errors of Measurement," *Biometrika*, 50(3/4), pp. 293–302.

Hansen, L. P., and J. A. Sheinkman (2009): "Long-term Risk: An Operator Approach," *Econometrica*, 77(1), 177–234.

Kelly, B., and S. Pruitt (2013): "Market Expectations in the Cross-Section of Present Values," *The Journal of Finance*, 68(5), 1721–1756.

Lamont, O. (1998): "Earnings and Expected Returns," *The Journal of Finance*, 53(5), pp. 1563–1587.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Mallat, S. G. (1989): "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 11, 674–693.

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47.

MISHKIN, F. S. (1992): "Is the Fisher effect for real? : A reexamination of the relationship between inflation and interest rates," *Journal of Monetary Economics*, 30(2), 195–215.

MÜLLER, U. K., AND M. W. WATSON (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76(5), 979–1016.

NELSON, C. R. (1976): "Inflation and Rates of Return on Common Stocks," *The Journal of Finance*, 31(2), 471–483.

ORTU, F., A. TAMONI, AND C. TEBALDI (2013): "Long-Run Risk and the Persistence of Consumption Shocks," *Review of Financial Studies*, 26(11), 2876–2915.

PERCIVAL, D. B., AND A. T. WALDEN (2000): *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.

PHILLIPS, P. (1986): "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33(3), 311–340.

RENAUD, O., J.-L. STARCK, AND F. MURTAGH (2005): "Wavelet-Based Combined Signal Filtering and Prediction," *IEEE Transactions SMC, Part B*, 35, 1241 – 1251.

SCHWERT, G. W. (1989): "Why Does Stock Market Volatility Change over Time?," *Journal of Finance*, 44(5), 1115–53.

TAMONI, A. (2011): "The multi-horizon dynamics of risk and returns," *SSRN eLibrary*.

WONG, P. W. (1993): "Wavelet decomposition of harmonizable random processes," *IEEE Transactions on Information Theory*, 39(1), 7–18.

YAZICI, B., AND R. KASHYAP (1997): "A class of second order self-similar processes for 1/f phenomena," *IEEE Transactions on Signal Processing*, 45(2), 396–410.
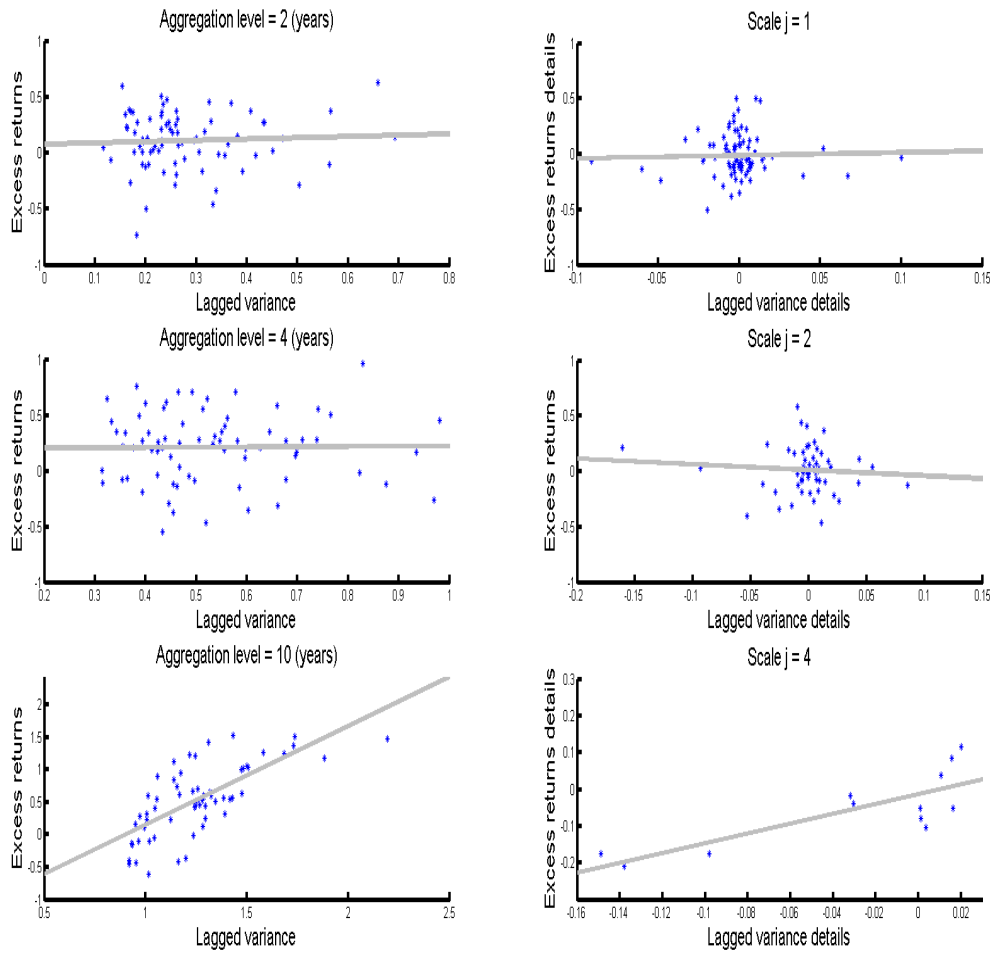
Figure 1: **Market Volatility.** The left panels present scatter plots of *forward* aggregates of excess market returns on *backward* aggregates of market variance for different levels of aggregation. The right panels present scatter plots of components (details) of the same series corresponding to analogous frequencies between one and two years ($j = 1$), two and four years ($j = 2$), and 8 and 16 years ($j = 4$), respectively. For a clear interpretation of the scales $j = 1, 2, \ldots$ into appropriate time horizons, please refer to Table 1.
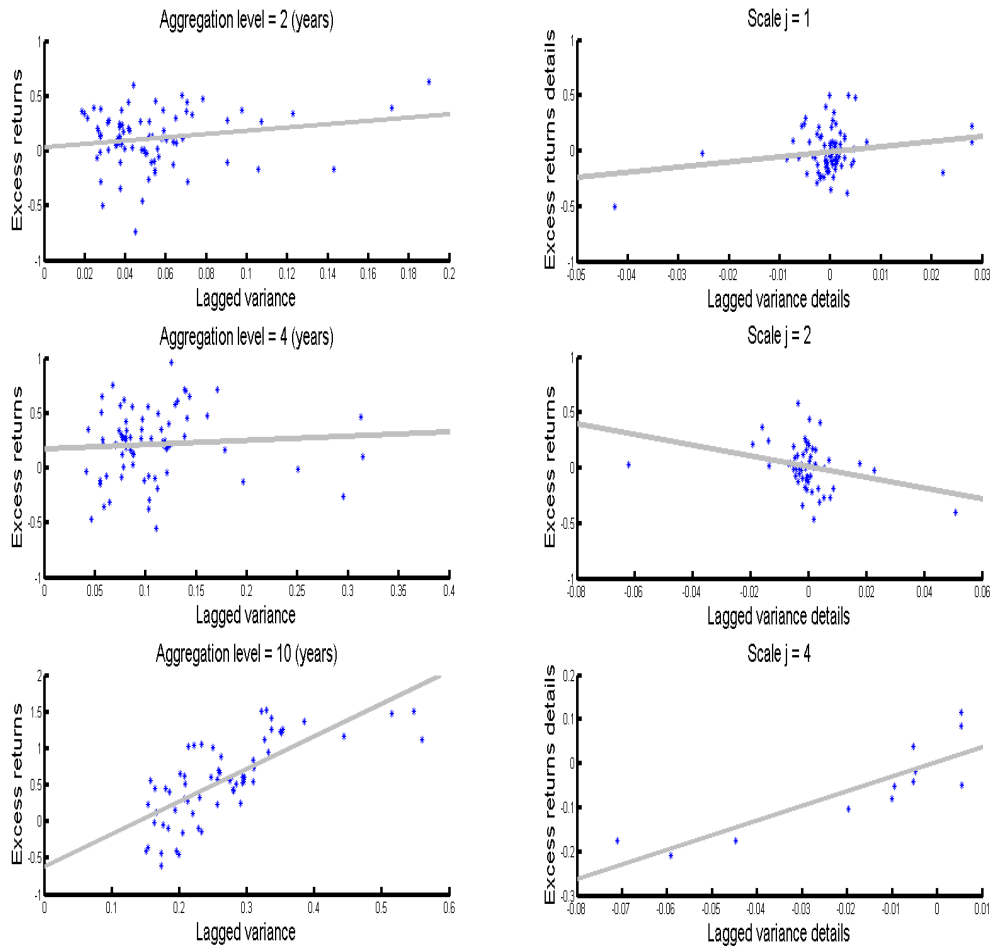
Figure 2: **Consumption Volatility.** The left panels present scatter plots of *forward* aggregates of excess market returns on *backward* aggregates of consumption variance for different levels of aggregation. The right panels present scatter plots of components (details) of the same series corresponding to analogous frequencies between one and two years ($j = 1$), two and four years ($j = 2$), and 8 and 16 years ($j = 4$), respectively. For a clear interpretation of the scales $j = 1, 2, \ldots$ into appropriate time horizons, please refer to Table 1.
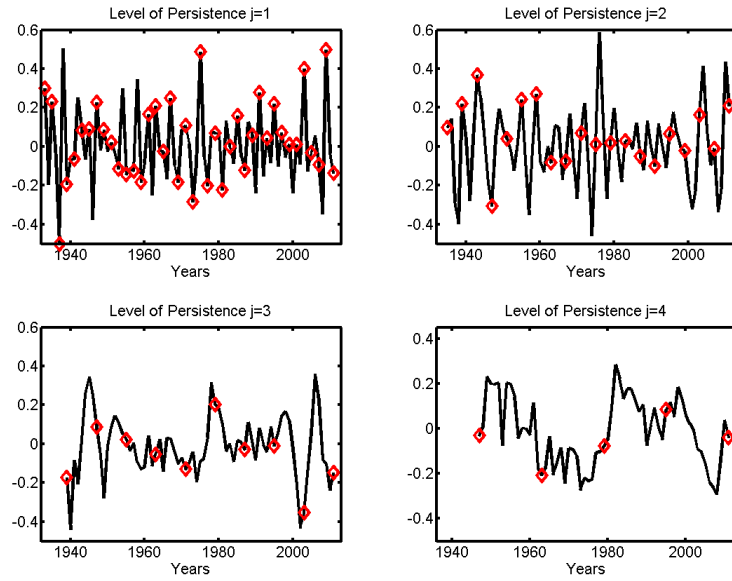
(a) Time-scale decomposition for market returns
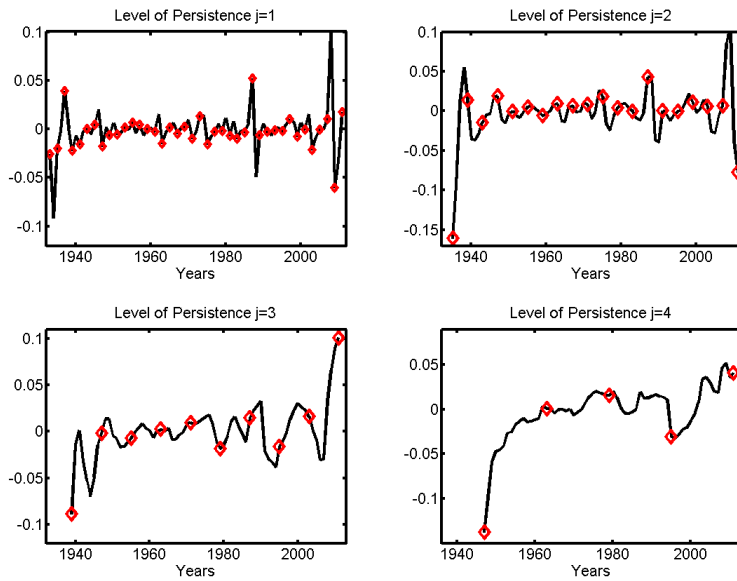
(b) Time-scale decomposition for market realized variance

Figure 3: Panel A displays the time-scale decomposition for the excess stock market returns. Panel B displays the time-scale decomposition for market realized variance. Solid lines represent the details, diamonds represent the decimated counterparts of the calendar-time details. For a clear interpretation of the scales $j = 1, 2, \ldots$ into appropriate time horizons, please refer to Table 1.
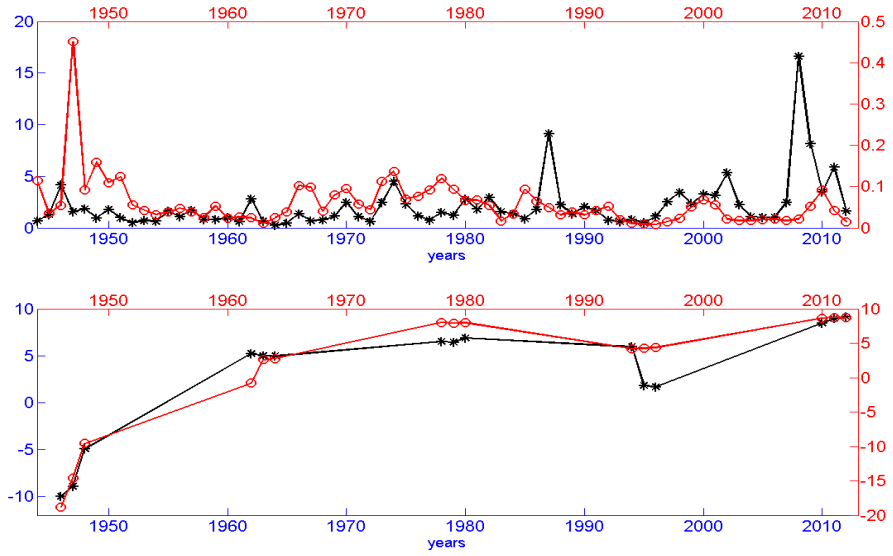
Figure 4: The upper panel in the figure displays the raw yearly series of market volatility (black-asterisk line) and consumption volatility (red-circle line); the lower panel displays the details of the two series with scale $j = 4$.

| Time-scale	Panel A: Monthly calendar time Frequency resolution	Panel B: Quarterly calendar time Frequency resolution	Panel C: Annual calendar time Frequency resolution
$j = 1$	$1 - 2$ months	$1 - 2$ quarters	$1 - 2$ years
$j = 2$	$2 - 4$ months	$2 - 4$ quarters	$2 - 4$ years
$j = 3$	$4 - 8$ months	$1 - 2$ years	$4 - 8$ years
$j = 4$	$8 - 16$ months	$2 - 4$ years	$8 - 16$ years
$j = 5$	$16 - 32$ months	$4 - 8$ years	$16 - 32$ years
$j = 6$	$32 - 64$ months	$8 - 16$ years	> 32 years
$j = 7$	$64 - 128$ months	$16 - 32$ years	
$\pi_t^{(7)}$	> 128	> 32 years	

Table 1: Interpretation of the time-scale (or persistence level) j in terms of time spans in the case of monthly (Panel A), quarterly (Panel B) and annual (Panel C) time series. Each scale corresponds to a frequency interval, or conversely an interval of periods, and thus each scale is associated with a range of time horizons.

	Panel A: Market volatility				Panel B: Consumption volatility				Panel C: Market excess returns			
Scales j =	1	2	3	4	1	2	3	4	1	2	3	4
1		0.28	-0.12	-0.03		0.29	0.10	-0.09		-0.04	-0.03	0.07
		(0.13)	(0.09)	(0.06)		(0.16)	(0.09)	(0.08)		(0.09)	(0.07)	(0.06)
2			-0.05	0.01			0.06	-0.04			-0.14	0.13
			(0.15)	(0.10)			(0.20)	(0.12)			(0.10)	(0.10)
3				0.39				0.10				0.15
				(0.11)				(0.13)				(0.13)

Table 2: **Pairwise correlations.** We report the pair-wise correlations between the individual details of market variance (Panel A), Consumption variance (Panel B) and excess market returns (Panel C). The pair-wise correlations are obtained by using redundant data on the details rather than the decimated counterparts. Standard errors for the correlation between $x_t^{(j)}$ and $x_t^{(j')}$, $j \neq j'$, are Newey-West with $2^{max(j,j')}$ lags.

Panel A: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	-0.01	-0.03	-0.09	-0.28	-0.10	0.43	0.71	0.67	0.44	0.18	0.04	-0.01
	(-0.13)	(-0.38)	(-0.96)	(-2.42)	(-0.80)	(4.40)	(8.19)	(7.61)	(4.47)	(1.54)	(0.36)	(-0.05)
$Adj.R^2$	[0.18]	[0.48]	[1.64]	[7.19]	[2.18]	[16.86]	[43.06]	[38.83]	[17.59]	[4.44]	[1.74]	[2.22]

Panel B: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t+1,t+h} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t+1,t+h}$	-0.00	-0.01	-0.02	-0.06	-0.14	-0.26	-0.34	-0.33	-0.21	-0.09	-0.03	-0.01
	(0.00)	(-0.07)	(-0.21)	(-0.57)	(-1.16)	(-1.85)	(-2.32)	(-2.26)	(-1.64)	(-0.78)	(-0.27)	(-0.07)
$Adj.R^2$	[0.21]	[0.51]	[1.08]	[2.45]	[4.57]	[8.36]	[12.55]	[11.57]	[6.63]	[3.46]	[2.56]	[2.47]

Table 3: **Simulation under the null of scale-dependent predictability. The relation is at scale** $j^* = 6$**.** We simulate excess market returns (y) and market variance (x) under the assumption of predictability at scale $j^* = 6$. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 6$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 500 replications. We set $T = 1024$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets. **Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period past realized market variances. **Panel B: contemporaneous aggregation.** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period realized market variances.

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00
	(0.01)	(-0.03)	(0.04)	(0.06)	(0.03)	(0.01)	(0.01)	(0.02)	(0.00)	(-0.06)	(-0.08)	(-0.07)
$Adj.R^2$	[0.14]	[0.29]	[0.74]	[1.48]	[1.70]	[1.96]	[1.78]	[1.94]	[2.26]	[2.08]	[2.34]	[2.24]

Table 4: **Simulation under the null of ABSENCE of scale-dependent predictability**. We simulate excess market returns (y) and market variance (x) under the assumption of no predictability. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_{t,j}$ for $j = 6$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 500 replications. We set $T = 1024$. We then run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period past realized market variances. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets.

Panel A: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	-0.00	-0.01	-0.02	-0.07	-0.15	-0.22	-0.21	-0.08	0.14	0.35	0.49	0.58
	(-0.01)	(-0.09)	(-0.21)	(-0.61)	(-1.21)	(-1.63)	(-1.48)	(-0.54)	(1.05)	(2.92)	(4.43)	(5.39)
$Adj.R^2$	[0.20]	[0.45]	[1.01]	[2.17]	[4.51]	[7.57]	[7.38]	[3.78]	[5.40]	[15.39]	[27.94]	[37.25]

Panel B: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t+1,t+h} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t+1,t+h}$	0.00	0.00	-0.00	-0.01	-0.02	-0.04	-0.06	-0.10	-0.15	-0.19	-0.24	-0.27
	(0.04)	(0.01)	(-0.03)	(-0.08)	(-0.14)	(-0.28)	(-0.48)	(-0.74)	(-0.97)	(-1.20)	(-1.44)	(-1.64)
$Adj.R^2$	[0.24]	[0.58]	[1.20]	[2.30]	[3.30]	[4.40]	[5.55]	[6.87]	[8.21]	[9.66]	[11.46]	[13.33]

Table 5: **Simulation under the null of scale-dependent predictability. The relation is at scale** $j^* = 7$**.** We simulate excess market returns (y) and market variance (x) under the assumption of predictability at scale $j^* = 7$. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 7$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 500 replications. We set $T = 1024$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets. **Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period past realized market variances. **Panel B: contemporaneous aggregation.** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period realized market variances.

Panel A: $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	1	2	3	4	5	6	7	8	9	10
$v_{t-h+1,t}$	0.22	0.12	-0.14	0.01	0.37	0.57	0.72	1.01	1.36	1.52
	(1.02)	(0.41)	(-0.50)	(0.04)	(1.19)	(1.68)	(2.18)	(3.38)	(5.72)	(7.46)
$R^2(\%)$	[0.68]	[0.31]	[0.48]	[0.01]	[3.86]	[10.00]	[15.69]	[27.14]	[43.16]	[50.43]

Panel B: $r_{t+2^j}^{(j)} - rf_{t+2^j}^{(j)} = \beta_j v_t^{(j)} + \epsilon_{t+2^j}$

	Time-scale j			
	1	2	3	4
$v_t^{(j)}$	0.56	-0.20	0.40	1.50
	[-1.21 3.12]	[-1.89 2.96]	[-1.90 4.05]	[0.18 2.96]
$R^2(\%)$	[0.85]	[2.64]	[2.83]	[74.35]

Panel C: $v_{t+2^j}^{(j)} = \rho_j v_t^{(j)} + \epsilon_{t+2^j}$

	Time-scale j			
	1	2	3	4
$v_t^{(j)}$	-0.16	-0.07	0.06	0.05
	[-0.37 0.06]	[-0.32 0.18]	[-0.30 0.39]	[-0.25 0.36]
$R^2(\%)$	[8.20]	[7.81]	[20.63]	[15.97]

Table 6: **Market Volatility. Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded market returns on the CRSP value-weighted index in excess of a 1-year Treasury bill rate on h-period past market variance. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses. **Panel B:** results of componentwise predictive regressions of the components of excess stock market returns on the components of market variance. **Panel C:** estimation results of the multiscale autoregressive system. For each level of persistence $j \in \{1, \ldots, 4\}$, we run a regression of the market variance component $v_{t+2^j}^{(j)}$ on its own lagged component $v_t^{(j)}$. For each regression, the table reports OLS estimates of the regressors, highest posterior density region with probability .95 (under a mildly informative prior) in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2012. For the translation of time-scales into appropriate range of time horizons refer to Table 1.

Panel A: $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	1	2	3	4	5	6	7	8	9	10
$v_{t-h+1,t}$	1.38	1.50	0.51	0.38	0.48	1.46	2.55	2.85	3.64	4.46
	(1.03)	(1.63)	(0.61)	(0.40)	(0.38)	(1.26)	(2.59)	(2.67)	(3.77)	(5.09)
$R^2(\%)$	[1.46]	[3.41]	[0.64]	[0.45]	[0.72]	[7.19]	[22.34]	[25.84]	[38.81]	[53.39]

Panel B: $r^{(j)}_{t+2^j} - rf^{(j)}_{t+2^j} = \beta_j v^{(j)}_t + \epsilon^{(j)}_{t+2^j}$

	Time-scale j			
	1	2	3	4
$v^{(j)}_t$	1.84	-6.47	-1.07	3.51
	[-1.16 5.57]	[-3.06 4.15]	[-2.10 4.44]	[0.33 4.95]
$R^2(\%)$	[6.31]	[15.13]	[3.98]	[82.96]

Panel C: $v^{(j)}_{t+2^j} = \rho_j v^{(j)}_t + \epsilon^{(j)}_{t+2^j}$

	Time-scale j			
	1	2	3	4
$v^{(j)}_t$	-0.12	0.11	-0.03	0.18
	[-0.27 0.16]	[-0.05 0.33]	[-0.15 0.12]	[0.03 0.35]
$R^2(\%)$	[4.90]	[10.17]	[22.51]	[49.85]

Table 7: **Consumption Volatility. Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded market returns on the CRSP value-weighted index in excess of a 1-year Treasury bill rate on h-period past consumption variance $v_{t-h,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses. **Panel B:** results of componentwise predictive regressions of the components of excess stock market returns on the components of consumption variance $v^2_{j,t}$. **Panel C:** estimation results of the multiscale autoregressive system. For each level of persistence $j \in \{1,\ldots,4\}$, we run a regression of the consumption variance component $v^{(j)}_{t+2^j}$ on its own lagged component $v^{(j)}_t$. For each regression, the table reports OLS estimates of the regressors, highest posterior density region with probability .95 (under a mildly informative prior) in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2012. For the translation of time-scales into appropriate range of time horizons refer to Table 1.

Panel A1 - Data: $r_{t+1,t+h} = \alpha_h + \beta_h \pi_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	1	2	3	4	5	6	7	8	9	10
$\pi_{t-h+1,t}$	0.43	0.02	0.11	0.65	0.99	1.23	1.14	1.08	1.15	1.11
	(0.57)	(0.05)	(0.22)	(3.41)	(5.34)	(5.82)	(4.73)	(3.30)	(3.60)	(2.68)
$R^2(\%)$	[0.81]	[0.00]	[0.15]	[6.53]	[15.43]	[27.54]	[28.42]	[27.62]	[31.40]	[28.24]

Panel A2 - Simulation: $r_{t+1,t+h} = \alpha_h + \beta_h \pi_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$\pi_{t-h+1,t}$	-0.04	-0.23	-0.52	-0.50	0.07	0.77	1.19	1.36	1.26	0.98
	(-0.18)	(-1.04)	(-2.08)	(-1.91)	(0.30)	(3.15)	(4.83)	(5.30)	(4.96)	(3.87)
$Adj.R^2$	[0.13]	[1.33]	[5.50]	[5.21]	[1.17]	[10.98]	[24.66]	[31.35]	[26.50]	[16.64]

Panel B: $r_{t+2^j}^{(j)} = \beta_j \pi_t^{(j)} + \epsilon_{t+2^j}^{(j)}$

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	0.48	-0.09	2.47	-0.21
	[-1.13 2.99]	[-1.47 2.60]	[0.04 4.42]	[-1.26 1.13]
$R^2(\%)$	[0.99]	[8.19]	[30.63]	[39.96]

Panel C: $\pi_{t+2^j}^{(j)} = \rho_j \pi_t^{(j)} + \epsilon_{t+2^j}^{(j)}$

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.22	0.11	0.13	-0.28
	[-0.35 0.10]	[-0.15 0.45]	[-0.19 0.55]	[-0.39 0.31]
$R^2(\%)$	[7.30]	[1.66]	[2.68]	[28.63]

Table 8: **Stock market return and inflation. Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded nominal stock market returns $r_{t+1,t+h}$ on h-period past realized inflation. We consider values of h equal to $1-10$ years. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. **Panel B:** scale-wise predictive regressions of the components of nominal excess stock market returns on the components of inflation $\pi_t^{(j)}$. **Panel C:** For each scale $j \in \{1, \ldots, 4\}$, we run a regression of the inflation rate $\pi_{t+2^j}^{(j)}$ on its own lagged component $\pi_t^{(j)}$. For each regression, the table reports OLS coefficient estimates of the regressors, highest posterior density region with probability .95 (under a mildly informative prior) in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2012. For the translation of time-scales into appropriate range of time horizons refer to Table 1 Panel A.

Panel A: $r_{t+1,t+h} = \alpha_h + \beta_h \Delta_h \pi_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$\Delta_h \pi_{t-h+1,t}$	0.11	0.15	-0.67	-1.22	-0.39	0.65	1.16	1.32	1.23	0.99
	(0.31)	(0.58)	(-2.49)	(-4.93)	(-1.63)	(1.80)	(4.30)	(4.43)	(4.33)	(3.85)
$Adj.R^2$	[0.27]	[0.48]	[7.27]	[23.79]	[2.98]	[3.90]	[21.36]	[28.09]	[24.40]	[16.16]

Panel B: $r_{t+1,t+h} = \alpha_h + \beta_h \Delta_h \pi_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$\Delta_h \pi_{t-h+1,t}$	-0.66	0.37	-0.32	-1.59	-0.25	0.25	0.50	0.80	1.04	1.45
	(-0.83)	(0.45)	(-0.45)	(-2.73)	(-0.41)	(0.44)	(0.84)	(1.43)	(2.47)	(5.53)
$Adj.R^2$	[0.39]	[0.33]	[0.37]	[8.84]	[0.26]	[0.40]	[2.07]	[6.52]	[12.82]	[27.53]

Table 9: **Forward on past differenced backward**. **Panel A: simulation** We simulate market returns and inflation under the assumption of predictability at scale $j^* = 3$. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_{t,j}$ for $j = 3$ and $x_t^{(j)} = \epsilon_{t,j}$ otherwise. We implement 500 replications. We set $T = 128$. **Panel B: data.** We run linear regressions (with an intercept) of h-period continuously compounded nominal market returns on h-period past differenced inflation. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets.

Panel A: $rf_{t+1,t+h} = \alpha_h + \beta_h \pi_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	1	2	3	4	5	6	7	8	9	10
$\pi_{t-h+1,t}$	0.37	0.44	0.50	0.61	0.70	0.73	0.73	0.70	0.66	0.61
	(2.92)	(2.49)	(2.57)	(2.83)	(2.97)	(3.04)	(3.06)	(3.14)	(3.20)	(3.12)
$R^2(\%)$	[20.76]	[24.07]	[26.36]	[32.36]	[37.31]	[39.72]	[38.64]	[35.29]	[30.95]	[25.53]

Panel B: $rf_{t+2^j}^{(j)} = \beta_j \pi_t^{(j)} + \epsilon_{t+2^j}$

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.01	-0.09	0.18	-0.05
	[-0.12 0.11]	[-0.30 0.13]	[-0.36 0.82]	[-0.76 0.79]
$R^2(\%)$	[0.04]	[10.84]	[13.24]	[1.38]

Panel C: $\pi_{t+2^j}^{(j)} = \rho_j \pi_t^{(j)} + \epsilon_{t+2^j}$

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.22	0.11	0.13	-0.28
	[-0.35 0.10]	[-0.15 0.45]	[-0.19 0.55]	[-0.39 0.31]
$R^2(\%)$	[7.30]	[1.66]	[2.68]	[28.63]

Table 10: **Risk-free rate and inflation. Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded nominal risk-free rate $rf_{t+1,t+h}$ on h-period past realized inflation. We consider values of h equal to $1-10$ years. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. **Panel B:** results of componentwise predictive regressions of the components of nominal risk-free rate on the components of inflation $\pi_{j,t}$. **Panel C:** estimation results of the multiscale autoregressive system. For each level of persistence $j \in \{1, \dots, 4\}$, we run a regression of the inflation rate $\pi_{t+2^j}^{(j)}$ on its own lagged component $\pi_t^{(j)}$. For each regression, the table reports OLS coefficient estimates of the regressors, highest posterior density region with probability .95 (under a mildly informative prior) in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2012. For the translation of time-scales into appropriate range of time horizons refer to Table 1 Panel A.

A Multiscale vs. standard Wold decomposition

We begin with

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

where we let

$$a_{j,k} = \left\langle x_t, \varepsilon_{t-k2^j}^{(j)} \right\rangle.$$

Consider, now, the time window $[t-6, t]$ and $J = 2$. We have

$$x_t = a_{1,0} \varepsilon_t^{(1)} + a_{1,1} \varepsilon_{t-2}^{(1)} + a_{1,2} \varepsilon_{t-4}^{(1)} + a_{1,3} \varepsilon_{t-6}^{(1)}$$
$$a_{2,0} \varepsilon_t^{(2)} + a_{2,1} \varepsilon_{t-4}^{(2)} + a_{J,0} \varepsilon_t^{(J)} + \dots$$

If

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

then

$$a_{1,0} = \left\langle x_t, \varepsilon_t^{(1)} \right\rangle = \frac{\psi_0}{\sqrt{2}} - \frac{\psi_1}{\sqrt{2}}$$

$$a_{1,1} = \left\langle x_t, \varepsilon_{t-2}^{(1)} \right\rangle = \frac{\psi_2}{\sqrt{2}} - \frac{\psi_3}{\sqrt{2}}$$

$$a_{1,2} = \left\langle x_t, \varepsilon_{t-4}^{(1)} \right\rangle = \frac{\psi_4}{\sqrt{2}} - \frac{\psi_5}{\sqrt{2}}$$

$$a_{1,3} = \left\langle x_t, \varepsilon_{t-6}^{(1)} \right\rangle = \frac{\psi_6}{\sqrt{2}} - \frac{\psi_7}{\sqrt{2}}$$

$$a_{2,0} = \left\langle x_t, \varepsilon_t^{(2)} \right\rangle = \frac{\psi_0}{2} + \frac{\psi_1}{2} - \frac{\psi_2}{2} - \frac{\psi_3}{2}$$

$$a_{2,1} = \left\langle x_t, \varepsilon_{t-4}^{(2)} \right\rangle = \frac{\psi_4}{2} + \frac{\psi_5}{2} - \frac{\psi_6}{2} - \frac{\psi_7}{2}$$

$$a_{J,0} = \left\langle x_t, \varepsilon_t^{(J)} \right\rangle = \frac{\psi_0}{2} + \frac{\psi_1}{2} + \frac{\psi_2}{2} + \frac{\psi_3}{2} + \frac{\psi_4}{2} + \frac{\psi_5}{2} + \frac{\psi_6}{2} + \frac{\psi_7}{2},$$

where we let

$$\psi_j = \langle x_t, \varepsilon_{t-j} \rangle.$$

Finally, note that

$$\psi_0 \left(\frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_0 \varepsilon_t$$

$$\psi_1 \left(-\frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_1 \varepsilon_{t-1}$$

$$\psi_2 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_2 \varepsilon_{t-2}$$

$$\psi_3 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_3 \varepsilon_{t-3}$$

$$\psi_4 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_4 \varepsilon_{t-4}$$

$$\psi_5 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_5 \varepsilon_{t-5}$$

$$\psi_6 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_6 \varepsilon_{t-6}$$

$$\psi_7 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \varepsilon_t^{(J)} \right) = \psi_7 \varepsilon_{t-7},$$

which yields the standard Wold Decomposition:

$$x_t = \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

B A primer on multiresolution analysis

In this section we provide a primer on multiresolution analysis (MRA). The fundamental idea behind MRA is to analyze data at different scales or resolutions. Here we assume for convenience that the original data lie in the space V_0.[9] For each vector space, there is another vector space of higher scale (or lower resolution) V_j. Each vector space contains all vector spaces that are of lower resolution. Also, for each vector space V_j, there is an orthogonal complement called $W_{j+1} = V_j - V_{j+1}$ and the basis function for this new space is the wavelet. Hence, we can define the coarse and detail spaces as $(V_1, W_1), (V_2, W_2), \dots$, i.e., increasing the index in the V-spaces is equivalent to coarsening the approximation to the data.

Now, we can represent a data series which lies in V_0 by projecting it onto the detail spaces W_j. In particular, we define the multiresolution decomposition of a series using

1. $\pi^{(J)}$, i.e., the coarsest scale

2. $\pi^{(J-1)} = \pi^{(J)} + x_t^{(J)}$

3. and, in general, $\pi^{(j-1)} = \pi^{(j)} + x_t^{(j)}$,

where $\pi^{(j)} \in V_j$ and $x_t^{(j)} \in W_j$. The decomposition takes the form:

$$\{\pi^{(J)}, x_t^{(J)}, x_t^{(J-1)}, \dots, x_t^{(j)}, \dots, x_t^{(1)}\}.$$

---

[9]For example if we have a deterministic signal of length 8 we can assume V_0 to be $\mathbb{R}^8$

At high frequencies, the wavelet is able to focus on short-lived phenomena, e.g., singularity points, whereas at low frequencies the wavelet has a large time-support allowing it to identify long, periodic fluctuations. This is a useful property for economic and financial systems in which variables may operate on a variety of time-scales simultaneously and in which, as discussed in the main text, the relation between variables may be different across time-scales. In the next subsection we look at how wavelet transforms can be formulated in terms of matrices and operators.

B.1 Expansion of a random process in a wavelet basis

This section explains the structure of wavelet algorithms using linear algebra. We work in discrete time and assume the following random process[10] $\omega(t) = \{\omega_0, \omega_1, \omega_2, \omega_3\}$, where - for simplicity - we can assume:

$$\omega_i = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } 1 - p \end{cases}.$$

The goal is to expand the process into $V^2 \oplus W^2 \oplus W^1$. The coefficients of the original signal are the coefficients of the process $\omega(t)$ expanded in V^0, i.e. $\langle \omega, \phi_0^0 \rangle = \omega_0$, $\langle \omega, \phi_1^0 \rangle = \omega_1$, $\langle \omega, \phi_2^0 \rangle = \omega_2$, $\langle \omega, \phi_3^0 \rangle = \omega_3$. The basis vectors of V^0 are translations of the Haar mother scaling function which has been dilated so that each basis function has a support equal to $1/4$.

The first step is to expand $\omega(t)$ into $V^1 \oplus W^1$. We carry out the matrix multiplication

$$W_1 = \mathcal{T}_1 \omega(t),$$

where

$$\mathcal{T}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

The first two rows correspond to the basis vectors ϕ_0^1 and ϕ_1^1 spanning V^1. The last two rows correspond to the basis vectors ψ_0^1 and ψ_1^1 spanning W^1. The matrix multiplication yields two averages

$$\pi_1^{(1)} = \frac{(\omega_0 + \omega_1)}{\sqrt{2}},$$

$$\pi_3^{(1)} = \frac{(\omega_2 + \omega_3)}{\sqrt{2}},$$

and two wavelet coefficients

$$\delta_1^{(1)} = \frac{(\omega_0 - \omega_1)}{\sqrt{2}},$$

$$\delta_3^{(1)} = \frac{(\omega_2 - \omega_3)}{\sqrt{2}}.$$

---

[10]Note that we assume the length (dimension) of the signal to be 2^j for some positive integer j. This assumption simplifies the analysis. The procedure can be generalized to any finite data series.

Hence, the process expanded into $V^1 \oplus W^1$ becomes

$$\omega(t) = \underbrace{\pi_1^{(1)}\phi_0^1 + \pi_3^{(1)}\phi_1^1}_{V^1} + \underbrace{\delta_1^{(1)}\psi_0^1 + \delta_3^{(1)}\psi_1^1}_{W^1}.$$

In the next step, we expand the process into $V^2 \oplus W^2 \oplus W^1$. We carry out another matrix multiplication

$$W_2 = \mathcal{T}_2 W_1$$

where

$$\mathcal{T}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that the identity matrix is in the bottom left side of $\mathcal{T}_2$. This is because we already have the random coefficients for the basis vectors of W^1 and, therefore, only the averages $(\pi_1^{(1)}, \pi_3^{(1)})$ become inputs for the next step. Alternatively, combining with the previous step:

$$W_2 = \mathcal{T}_2 \mathcal{T}_1 \omega(t),$$

where

$$\mathcal{T}_2 \mathcal{T}_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Notice that the first row corresponds to the basis vector ϕ_0^2 spanning V^2, while the second row corresponds to the basis vector ψ_0^2 spanning W^2. Eventually, the process expanded into $V^2 \oplus W^2 \oplus W^1$ becomes:

$$\omega(t) = \underbrace{\pi_4^{(2)}\phi_0^2}_{V^2} + \underbrace{\delta_4^{(2)}\psi_0^2}_{W^2} + \underbrace{\delta_1^{(1)}\psi_0^1 + \delta_3^{(1)}\psi_1^1}_{W^1},$$

where the average and the wavelet coefficients are given by

$$\pi_4^{(2)} = \frac{\sum_{i=0}^3 \omega_i}{2}$$

$$\delta_4^{(2)} = \frac{(\omega_0 + \omega_1)/\sqrt{2} - (\omega_2 + \omega_3)/\sqrt{2}}{\sqrt{2}}.$$

More generally, if the data set $\mathbf{x}_0 = \{x_{0,k}\}_{k=0}^{T-1}$ contains T elements, then the recursive iterations would continue until a single average is calculated. Indeed, the matrix is first applied to the original, full-length data vector. Then, the vector is smoothed and the matrix applied again. This process would replace the original data set of N elements with an average $\pi_T^{(J)}$, which we define as the smooth component, followed by a set of wavelet coefficients $D_J \equiv \{\delta_k^{(J)}\}_k, \delta^{(J-1)} \equiv \{\delta_k^{(J-1)}\}_k, \ldots, \delta^{(1)} \equiv \{\delta_k^{(1)}\}_k$, which make up the detail components. In the end, the discrete wavelet transform of the

sequence $\mathbf{x}_0$ can be represented as:

$$\mathbf{x}_0 \to \{\pi^{(J)}, \delta^{(J)}, \ldots, \delta^{(j)}, \ldots, \delta^{(2)}, \delta^{(1)}\}. \tag{B.1}$$

C Understanding two-way aggregation

C.1 Dynamics of time-scale components

Assume the following component (or detail) dynamics for $j = j^*$, where $j^* \in \{1, \ldots, J\}$:

$$y_{t+2^j}^{(j)} = \beta_j x_t^{(j)} \tag{C.1}$$

$$x_{t+2^j}^{(j)} = \rho_j x_t^{(j)} + \sigma_j \epsilon_{t+2^j} \tag{C.2}$$

For $j = 1, \ldots, J$, with $j \neq j^*$, we have

$$y_t^{(j)} = 0, \tag{C.3}$$

$$x_t^{(j)} = 0. \tag{C.4}$$

Assume - for conciseness - that $T = 16$, $j^* = 2$, and $J = 3$. Arrange the details of x as follows:

$$\begin{pmatrix} \pi_8^{(3)} & \pi_{16}^{(3)} \\ x_8^{(3)} & x_{16}^{(3)} \\ x_8^{(2)} & x_{16}^{(2)} \\ x_4^{(2)} & x_{12}^{(2)} \\ x_8^{(1)} & x_{16}^{(1)} \\ x_6^{(1)} & x_{14}^{(1)} \\ x_4^{(1)} & x_{12}^{(1)} \\ x_2^{(1)} & x_{10}^{(1)} \end{pmatrix} \tag{C.5}$$

and, analogously, for the details of y. Consider the following isometric transform matrix:

$$\mathcal{T}^{(3)} = \begin{pmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{C.6}$$

To reconstruct the time series x_t, we run through each column of the matrix (C.5) and, for each column, we perform the following operation:

$$X_8^{(3)} = \begin{pmatrix} x_8 \\ x_7 \\ x_6 \\ x_5 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \end{pmatrix} = \left(\mathcal{T}^{(3)}\right)^{-1} \begin{pmatrix} \pi_8^{(3)} \\ x_8^{(3)} \\ x_8^{(2)} \\ x_4^{(2)} \\ x_8^{(1)} \\ x_6^{(1)} \\ x_4^{(1)} \\ x_2^{(1)} \end{pmatrix} \tag{C.7}$$

and

$$X_{16}^{(3)} = \begin{pmatrix} x_{16} \\ x_{15} \\ x_{14} \\ x_{13} \\ x_{12} \\ x_{11} \\ x_{10} \\ x_9 \end{pmatrix} = \left(\mathcal{T}^{(3)}\right)^{-1} \begin{pmatrix} \pi_{16}^{(3)} \\ x_{16}^{(3)} \\ x_{16}^{(2)} \\ x_{12}^{(2)} \\ x_{16}^{(1)} \\ x_{14}^{(1)} \\ x_{12}^{(1)} \\ x_{10}^{(1)} \end{pmatrix}. \tag{C.8}$$

We do the same for the details of y_t. The matrix $\left(\mathcal{T}^{(3)}\right)^{-1}$ takes the following form:

$$\left(\mathcal{T}^{(3)}\right)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{C.9}$$

Using the dynamics of the state (C.2), (C.7) and (C.8), we obtain

$$
X_{16}^{(3)} = \begin{pmatrix}
x_{16} = & x_{16}^{(2)}/2 \\
x_{15} = & x_{16}^{(2)}/2 \\
x_{14} = & -x_{16}^{(2)}/2 \\
x_{13} = & -x_{16}^{(2)}/2 \\
x_{12} = & x_{12}^{(2)}/2 \\
x_{11} = & x_{12}^{(2)}/2 \\
x_{10} = & -x_{12}^{(2)}/2 \\
x_{9} = & -x_{12}^{(2)}/2
\end{pmatrix}
\tag{C.10}
$$

and

$$
X_{8}^{(3)} = \begin{pmatrix}
x_{8} = & x_{8}^{(2)}/2 \\
x_{7} = & x_{8}^{(2)}/2 \\
x_{6} = & -x_{8}^{(2)}/2 \\
x_{5} = & -x_{8}^{(2)}/2 \\
x_{4} = & x_{4}^{(2)}/2 \\
x_{3} = & x_{4}^{(2)}/2 \\
x_{2} = & -x_{4}^{(2)}/2 \\
x_{1} = & -x_{4}^{(2)}/2
\end{pmatrix}.
\tag{C.11}
$$

C.2 Aggregation

C.2.1 Fitting an AR(1) process to the regressor

We fit an AR(1) process to x_t:

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

From (C.10) and (C.11), it is easy to see that, for $j^* = 2$:

$$
\widetilde{\rho} = \frac{1 - \rho_{j^*}}{4}.
$$

For a more general j^*, i.e., if the process for x_t is given by (C.2) and (C.4), then

$$
\widetilde{\rho} = \frac{\underbrace{1 + 1 + \dots}_{2^{j^*-1}-1} - 1 + \underbrace{1 + 1 + \dots}_{2^{j^*-1}-1} - \rho_{j^*}}{2^{j^*}}.
$$

This result clarifies the relation between scale-wise persistence (ρ_{j^*}) and persistence in calendar time ($\widetilde{\rho}$). If $\rho_{j^*} < \frac{1}{5}$, then $\widetilde{\rho} > \rho_{j^*}$ for all j^*. However, as j^* grows large, $\widetilde{\rho}$ approximates 1. In other words, the largest the driving scale, the largest the calendar-time correlation *irrespective* of the actual scale-wise correlation.

C.2.2 Two-way (forward/backward) regressions

Let us construct the temporally-aggregated series

$$y_{t+1,t+h} = \sum_{i=1}^{h} y_{t+i}$$

and run the forward/backward regression

$$y_{t+1,t+h} = \tilde{\beta} x_{t-h+1,t} + \epsilon_{t+1,t+h},$$

where $x_{t+1,t+h}$ is defined like $y_{t+1,t+h}$. For $h = 4$, and using (C.1) and (C.3) together with (C.10) and (C.11), we have

$$y_{13,16} = 0 \qquad\qquad x_{13,16} = 0$$
$$y_{12,15} = (-y_{16}^{(2)} + y_{12}^{(2)})/2 = \beta\left(-x_{12}^{(2)} + x_8^{(2)}\right)/2 \qquad\qquad x_{12,15} = (-x_{16}^{(2)} + x_{12}^{(2)})/2$$
$$y_{11,14} = -y_{16}^{(2)} + y_{12}^{(2)} = \beta\left(-x_{12}^{(2)} + x_8^{(2)}\right) \qquad\qquad x_{11,14} = -x_{16}^{(2)} + x_{12}^{(2)}$$
$$y_{10,13} = (-y_{16}^{(2)} + y_{12}^{(2)})/2 = \beta\left(-x_{12}^{(2)} + x_8^{(2)}\right)/2 \qquad\qquad x_{10,13} = (-x_{16}^{(2)} + x_{12}^{(2)})/2$$
$$y_{9,12} = 0 \qquad\qquad x_{9,12} = 0$$
$$y_{8,11} = (-y_{12}^{(2)} + y_8^{(2)})/2 = \beta\left(-x_8^{(2)} + x_4^{(2)}\right)/2 \qquad\qquad x_{8,11} = (-x_{12}^{(2)} + x_8^{(2)})/2$$
$$y_{7,10} = -y_{12}^{(2)} + y_8^{(2)} = \beta\left(-x_8^{(2)} + x_4^{(2)}\right) \qquad\qquad x_{7,10} = -x_{12}^{(2)} + x_8^{(2)}$$
$$y_{6,9} = (-y_{12}^{(2)} + y_8^{(2)})/2 = \beta\left(-x_8^{(2)} + x_4^{(2)}\right)/2 \qquad\qquad x_{6,9} = (-x_{12}^{(2)} + x_8^{(2)})/2$$
$$y_{5,8} = 0 \qquad\qquad x_{5,8} = 0$$
$$y_{4,7} = (-y_8^{(2)} + y_4^{(2)})/2 = \beta\left(-x_4^{(2)} + x_0^{(2)}\right)/2 \qquad\qquad x_{4,7} = (-x_8^{(2)} + x_4^{(2)})/2$$
$$y_{3,6} = -y_8^{(2)} + y_4^{(2)} = \beta\left(-x_4^{(2)} + x_0^{(2)}\right) \qquad\qquad x_{3,6} = -x_8^{(2)} + x_4^{(2)}$$
$$y_{2,5} = (-y_8^{(2)} + y_4^{(2)})/2 = \beta\left(-x_4^{(2)} + x_0^{(2)}\right)/2 \qquad\qquad x_{2,5} = (-x_8^{(2)} + x_4^{(2)})/2$$
$$y_{1,4} = 0 \qquad\qquad x_{1,4} = 0.$$

Thus, regressing $y_{t+1,t+4}$ on $x_{t-3,t}$ yields $\tilde{\beta} = \beta$ with $R^2 = 100\%$. Hence, when scale-wise predictability applies to a scale operating between 2^{j^*-1} and 2^{j^*}, maximum predictability upon two-way aggregation arises over an horizon $h = 2^{j^*}$. In our case, $j^* = 2$ and $h = 4$. Consider, for

example, an alternative aggregation level: $h = 2$. We have

$$y_{15,16} = y_{16}^{(2)} = \beta x_{12}^{(2)} \qquad\qquad x_{15,16} = x_{16}^{(2)}$$
$$y_{14,15} = 0 \qquad\qquad x_{14,15} = 0$$
$$y_{13,14} = -y_{16}^{(2)} = -\beta x_{12}^{(2)} \qquad\qquad x_{13,14} = -x_{16}^{(2)}$$
$$y_{12,13} = (-y_{16}^{(2)} + y_{12}^{(2)})/2 = \beta(-x_{12}^{(2)} + x_{8}^{(2)})/2 \qquad\qquad x_{12,13} = (-x_{16}^{(2)} + x_{12}^{(2)})/2$$
$$y_{11,12} = y_{12}^{(2)} = \beta x_{8}^{(2)} \qquad\qquad x_{11,12} = x_{12}^{(2)}$$
$$y_{10,11} = 0 \qquad\qquad x_{10,11} = 0$$
$$y_{9,10} = -y_{12}^{(2)} = -\beta x_{8}^{(2)} \qquad\qquad x_{9,10} = -x_{12}^{(2)}$$
$$y_{8,9} = (-y_{12}^{(2)} + y_{8}^{(2)})/2 = \beta(-x_{8}^{(2)} + x_{4}^{(2)})/2 \qquad\qquad x_{8,9} = (-x_{12}^{(2)} + x_{8}^{(2)})/2$$
$$y_{7,8} = y_{8}^{(2)} = \beta x_{4}^{(2)} \qquad\qquad x_{7,8} = x_{8}^{(2)}$$
$$y_{6,7} = 0 \qquad\qquad x_{6,7} = 0$$
$$y_{5,6} = -y_{8}^{(2)} = -\beta x_{4}^{(2)} \qquad\qquad x_{5,6} = -x_{8}^{(2)}$$
$$y_{4,5} = (-y_{8}^{(2)} + y_{4}^{(2)})/2 = \beta(-x_{4}^{(2)} + x_{0}^{(2)})/2 \qquad\qquad x_{4,5} = (-x_{8}^{(2)} + x_{4}^{(2)})/2$$
$$y_{3,4} = y_{4}^{(2)} = \beta x_{0}^{(2)} \qquad\qquad x_{3,4} = x_{4}^{(2)}$$
$$y_{2,3} = 0 \qquad\qquad x_{2,3} = 0$$
$$y_{1,2} = -y_{4}^{(2)} = -\beta x_{0}^{(2)} \qquad\qquad x_{1,2} = -x_{4}^{(2)},$$

where we use the implied dynamics for x, see equations (C.10) and (C.11), and the equivalent ones for y together with (C.1) and (C.2). The regression of $y_{t+1,t+2}$ on $x_{t-1,t}$ yields (based on a fundamental block of four elements):

$$\tilde{\beta} = \frac{Cov(y_{15,16}, x_{13,14}) + Cov(y_{13,14}, x_{11,12})}{Var(x_{10,11}) + Var(x_{11,12}) + Var(x_{12,13}) + Var(x_{13,14})}$$

$$= \frac{-\beta Var\left(x_{12}^{(2)}\right)\rho - \beta Var\left(x_{12}^{(2)}\right)}{Var\left(x_{12}^{(2)}\right) + Var\left(\frac{x_{16}^{(2)}}{2}\right) + Var\left(\frac{x_{12}^{(2)}}{2}\right) - \frac{Cov\left(x_{16}^{(2)}, x_{12}^{(2)}\right)}{2} + Var\left(x_{16}^{(2)}\right)}$$

$$= -2\beta \frac{(1 + \rho_j)}{5 - \rho_j}$$

and, hence, an inconsistent slope estimate. This estimate could have a changed sign (with respect to β) and be drastically attenuated. In fact, $\tilde{\beta} = 0$ if $\rho_j = -1$ and $\tilde{\beta} = -\beta$ if $\rho_j = 1$.

C.2.3 Contemporaneous aggregation

We now run the contemporaneous regression

$$y_{t+1,t+h} = \tilde{\beta} x_{t+1,t+h} + \epsilon_{t+1,t+h}.$$

For $h = 4$, the relevant 4-term block contains terms like:

$$y_{t+1,t+h} = \beta(-x_{12}^{(2)} + x_8^{(2)})/2$$

$$x_{t+1,t+h} = (-x_{16}^{(2)} + x_{12}^{(2)})/2$$

By taking covariances we obtain

$$\tilde{\beta} = \beta \frac{\left(- \operatorname{Var}(x_{12}^{(2)}) + \rho_j \operatorname{Var}(x_{12}^{(2)}) - \rho_j^2 \operatorname{Var}(x_{12}^{(2)}) + \rho_j \operatorname{Var}(x_{12}^{(2)})\right)}{\operatorname{Var}(x_{16}^{(2)}) + \operatorname{Var}(x_{12}^{(2)}) - 2cov(x_{16}^{(2)}, x_{12}^{(2)})}$$

$$= \beta \frac{\left(-1 + 2\rho_j - \rho_j^2\right)}{2(1 - \rho_j)}$$

$$= -\beta \frac{(1 - \rho_j)}{2}.$$

Again, $\tilde{\beta} \neq \beta$. Its sign is also incorrect. We note that, in this case, $\tilde{\beta} = 0$ if $\rho_j = 1$ and $\tilde{\beta} = -\beta$ if $\rho_j = -1$. The R^2 is equal to

$$R^2 = \frac{\tilde{\beta}^2 \operatorname{Var}\left(-x_{12}^{(2)} + x_8^{(2)}\right)}{\beta^2 \operatorname{Var}\left(-x_{12}^{(2)} + x_8^{(2)}\right)} = \left(\frac{1 - \rho_j}{2}\right)^2.$$

The larger ρ_j, the smaller the R^2, and the more attenuated towards zero $\tilde{\beta}$ is.

C.2.4 Two-way (forward/backward) regressions on differences

Consider the regression

$$y_{t+1,t+h} = \tilde{\beta}\left(x_{t-h+1,t} - x_{t-2h+1,t-h}\right) + \epsilon_{t+1,t+h}.$$

Using the same methods as before, one can show that, if $h = 2$:

$$\tilde{\beta}_{h=2} = -\frac{\beta}{2} \quad \text{and} \quad R_{h=2}^2 = \frac{1}{2} \frac{[7 + \rho_j]}{[5 - \rho_j]}.$$

If $h = 3$:

$$\tilde{\beta}_{h=3} = \beta \left[\frac{\left[\frac{9}{4} - \frac{5}{4}\rho_j\right]}{9 - \frac{15}{2}\rho_j + \frac{1}{2}\rho_j^2}\right] \quad \text{and} \quad R_{h=3}^2 = \left[\frac{\left[\frac{9}{4} - \frac{5}{4}\rho_j\right]}{9 - \frac{15}{2}\rho_j + \frac{1}{2}\rho_j^2}\right]^2 \frac{\left[9 - \frac{15}{2}\rho_j + \frac{1}{2}\rho_j^2\right]}{3 - 2\rho_j}.$$

If $h = 1$:

$$\tilde{\beta}_{h=1} = \frac{\beta}{2}\left[\frac{1 + 3\rho_j}{3 + \rho_j}\right] \quad \text{and}$$

$$R_{h=1}^2 = \frac{1}{4}\left[\frac{1 + 3\rho_j}{3 + \rho_j}\right]^2 \left[\frac{3 + \rho_j}{2}\right].$$

Hence, when scale-wise predictability applies to a scale operating between 2^{j^*-1} and 2^{j^*}, two-way aggregation on *differences*, rather than on *levels*, would yield - at $h = 2^{j^*-1}$ - a slope coefficient whose sign is the opposite of the true sign. In our case, $j^* = 2$ and $h = 2$. As shown, $\tilde{\beta}_{h=2} = -\frac{\beta}{2}$. We note that $R^2_{h=2} = 1$ if $\rho_j = 1$. Also, $R^2_{h=2}$ reaches a minimum value of 0.5 for $\rho_j = -1$. Hence, the magnitude of the coefficient of determination is sizeable in correspondence with $h = 2$. In addition, $R^2_{h=2} > R^2_{h=1}$ and $R^2_{h=2} > R^2_{h=3}$ for all ρ, thereby determining tent-shaped behavior of the R^2s around $h = 2$.

D   Predicting long-run variance

Bandi and Perron (2008) find that, in spite of its predictive ability for long-run (i.e., forward-aggregated) excess market returns, long-run past (backward-aggregated) variance does not predict its future values. As emphasized by Bandi and Perron (2008), if taken literally, this result would contradict classical economic logic behind predictability. Why is realized variance aggregated over a certain horizon not predictable? We provide a justification based on measurement error.

In Section 5.1 we simulated models in which realized variance was characterized by autoregressive components at scales $j^* = 6$ or $j^* = 7$, respectively. Let us now focus on the variance dynamics given these specifications. Table D.1–Panels A and B report simulation results from running linear regressions of h-period realized market variances. We regress h-period future variance $v^2_{t+1,t+h} = \sum_{i=1}^{h} v^2_{t+i-1,t+i}$ on h-period past variance $v^2_{t-h+1,t} = \sum_{i=1}^{h} v^2_{t-i+1,t-i}$, i.e.,

$$v^2_{t+1,t+h} = \rho_h v^2_{t-h+1,t} + \varepsilon_{t,t+h}, \tag{D.1}$$

when the underlying data generating processes have predictable scales at $j^* = 6$ or $j^* = 7$. It is readily verified that both of these scales generate spurious *negative* predictability upon aggregation with (approximate) peaks corresponding to the upper bound of the corresponding interval of frequencies.

Thus, empirical evidence of some (negative) long-run variance predictability appears to be an additional diagnostic for the proposed data generating process relying on details. Having made this point, we note that the findings in Table D.1–Panel A and Table D.1–Panel B are at odds with the findings in Table D.2.[11]

Next, we show that this apparent contradiction is easily re-solved by taking into account measurement error generated by heterogeneity in the levels of persistence. More explicitly, we evaluate the effect on the dynamics of aggregated variance of the presence of uncorrelated predictable components at different frequencies. To this extent, we simulate variance according to the following specifications:

$$v^{2(j)}_{k2^j+2^j} = \rho_j v^{2(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j}$$

with $j = j^* \in S = \{7, 9\}$ and

$$v^{2(j)}_{k2^j+2^j} = \varepsilon^{(j)}_{k2^j+2^j}$$

for $j \notin S$, $j = 1, \ldots, J = 9$, where $corr(\varepsilon^{(j')}_{t'}, \varepsilon^{(j)}_{t}) = 0 \ \forall t, j, t', j'$ with $t' \neq t$ and $j' \neq j$. Hence, we now consider a variance process generated by two autoregressive details, one at scale 7 and one at

---

[11]Table 9 of Bandi and Perron (2008) also report estimates from a linear regression of realized variance on itself h periods in the past. Bandi and Perron (2008) show that "the autocorrelations become quickly statistically insignificant with the level of aggregation" and conclude that "realized variance is virtually uncorrelated in the long run."

Panel A - case $j^* = 6$: $x_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	0.24	0.13	0.06	-0.26	-0.43	-0.47	-0.48	-0.48	-0.49	-0.49	-0.48	-0.48
	(5.30)	(2.10)	(0.66)	(-2.56)	(-4.04)	(-4.57)	(-3.68)	(-3.80)	(-4.34)	(-3.96)	(-4.04)	(-3.59)
$Adj.R^2$	[5.89]	[2.16]	[1.19]	[7.78]	[18.99]	[23.20]	[24.95]	[24.64]	[25.28]	[25.67]	[24.18]	[24.82]

Panel B - case $j^* = 7$: $x_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	0.31	0.24	0.22	-0.00	-0.19	-0.32	-0.40	-0.44	-0.45	-0.46	-0.46	-0.47
	(6.89)	(4.02)	(2.77)	(-0.06)	(-1.62)	(-2.52)	(-2.91)	(-3.12)	(-3.31)	(-3.36)	(-3.28)	(-3.16)
$Adj.R^2$	[9.77]	[6.37]	[5.80]	[1.38]	[5.09]	[12.03]	[18.10]	[20.59]	[21.67]	[22.67]	[24.06]	[25.61]

Table D.1: **Simulation under the null of scale-dependent predictability.** We run linear regressions of h-period realized market variances $x_{t+1,t+h}$ on h-period past realized market variances $x_{t-h+1,t}$. We implement 500 replications. We set $T = 1024$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets. **Panel A:** We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 6$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. **Panel B.** We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 7$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise.

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$v_{t-h,t}^2$	0.47	0.34	0.33	0.28	0.21	0.20	0.23	0.20	0.13	0.05
	(5.05)	(4.08)	(2.46)	(1.83)	(1.37)	(1.06)	(1.15)	(0.99)	(0.70)	(0.26)
$Adj.R^2$	[17.35]	[13.08]	[10.39]	[6.53]	[4.31]	[4.83]	[6.73]	[4.67]	[1.73]	[0.19]

Table D.2: **Market Volatility.** We run linear regressions of h-period realized market variances $v_{t+1,t+h}^2$ on h-period past realized market variances $v_{t-h+1,t}^2$. The table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses. The sample is annual and spans the period 1930-2012. For the translation of time-scales into appropriate range of time horizons refer to Table 1.

scale 9.

Table D.3–Panel C reports the results. Comparing these findings with the ones in Table D.1-Panel B, we note that the effect of adding an additional predictable detail (relatively close to the permanent one) is two-fold. First, the autoregressive coefficient at horizon $h = 120$ is no longer significant and aggregate "realized variance is virtually uncorrelated in the long run," even though two autoregressive details are truly present. Second, the t-statistics and the R^2 for very short horizons, namely 3 months and 6 months, are now higher and significant.

This said, the previous conclusions about long-run predictability continue to hold. Earlier, we postulated a predictable relation between returns and realized variance only at scale $j^* = 7$. In other words, the variance process had only one predictable component. We now consider a case in which there are two details of the variance process which are *not* white noise ($j = \{7, 9\}$). However, we continue to impose that the only relevant predictability relation between returns and variance occurs at $j^* = 7$. Thus, we assume

$$r^{(j)}_{k2^j, k2^j + 2^j} = \begin{cases} v^{2(j)}_{k2^j} & \text{if } j = 7, \\ u^{(j)}_t & \text{otherwise .} \end{cases}$$

Results are reported in Table D.3–Panel A. We confirm the findings of the previous section, namely the relation at scale $j^* = 7$ implies a spike in the slope estimates, t-statistics, and R^2s at the horizon $h = 120$. Compared to Table 5–Panel A, we note that, when the variance process is characterized by two autoregressive components, Table D.3–Panel A shows that the estimated (upon aggregation) predictive relation is softened but continues to remain significant.

E   Computation of the Highest Posterior Density Intervals

Consider, again, the linear regression model:

$$y^{(j)}_{k2^j + 2^j} = \beta_j x^{(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j + 2^j}.$$

Suppose that the coefficient β_j can lie anywhere in the interval $(-\infty, \infty)$. A 95% credible interval for β_j is any interval $[a; b]$ so that:

$$p(a \leq \beta_j \leq b \mid y^{(j)}) = \int_a^b p(\beta_j \mid y^{(j)})d\beta_j = 0.95.$$

To choose among the infinite number of credible intervals, it is common to select the one with the smallest area. In the standard Normal example, $[-1.96; 1.96]$ is the shortest credible interval. This is the Highest Posterior Density Interval (HPDI). Each table presents 95% HPDIs in addition to point estimates. In simple words, the researcher is 95% sure that β_j lies within the HPDI.

Given the structure of the above regression, we employ a Normal-Gamma prior (i.e., the product of a Gamma and a conditional Normal) for β_j and $h = \frac{1}{\sigma^2}$:

$$\beta_j, h \sim NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}).$$

We use under-bars (e.g., $\underline{\beta}$) to denote the parameters of the prior density and over-bars (e.g., $\overline{\beta}$) to denote the parameters of the posterior density. The marginal posterior distribution for β_j is a

Panel A: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	-0.00	-0.01	-0.02	-0.09	-0.19	-0.26	-0.23	-0.09	0.11	0.28	0.39	0.43
	(-0.01)	(-0.10)	(-0.25)	(-0.69)	(-1.34)	(-1.80)	(-1.59)	(-0.61)	(0.81)	(2.23)	(3.27)	(3.94)
$Adj.R^2$	[0.17]	[0.38]	[0.86]	[2.03]	[4.62]	[8.07]	[7.83]	[3.73]	[4.72]	[13.01]	[22.75]	[28.79]

Panel B: $y_{t+1,t+h} = \alpha_h + \beta_h x_{t+1,t+h} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t+1,t+h}$	0.00	0.00	-0.00	-0.01	-0.02	-0.04	-0.06	-0.10	-0.13	-0.16	-0.19	-0.21
	(0.05)	(0.02)	(-0.02)	(-0.08)	(-0.15)	(-0.28)	(-0.45)	(-0.69)	(-0.89)	(-1.10)	(-1.30)	(-1.48)
$Adj.R^2$	[0.21]	[0.51]	[1.06]	[2.09]	[3.09]	[4.02]	[4.95]	[6.07]	[7.26]	[8.56]	[10.12]	[11.86]

Panel C: $x_{t+1,t+h} = \alpha_h + \beta_h x_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in months)											
	3	6	12	24	36	48	60	72	84	96	108	120
$x_{t-h+1,t}$	0.24	0.15	0.17	0.20	0.15	0.03	-0.05	-0.07	-0.07	-0.07	-0.09	-0.12
	(5.27)	(2.46)	(2.21)	(1.84)	(1.16)	(0.19)	(-0.37)	(-0.54)	(-0.59)	(-0.60)	(-0.66)	(-0.78)
$Adj.R^2$	[7.35]	[3.78]	[3.20]	[3.40]	[6.79]	[7.71]	[8.52]	[8.61]	[8.08]	[7.32]	[6.39]	[5.31]

Table D.3: **Simulation under the null of scale-dependent predictability. The relation is at scale $j^* = 7$. Two persistent components in the regressor.** We simulate excess market returns (y) and market variance (x) under the assumption of predictability at scale $j^* = 7$. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 7, 9$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 500 replications. We set $T = 1024$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with 2*(horizon-1) lags in parentheses and adjusted R^2 statistics in square brackets. **Panel A:** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period past realized market variances. **Panel B: contemporaneous aggregation.** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period realized market variances. **Panel C:** . We run linear regressions of h-period realized market variances $x_{t+1,t+h}$ on h-period past realized market variances $x_{t-h+1,t}$.

t-distribution:

$$\beta_j \mid y^{(j)} \sim t(\overline{\beta}, \overline{s}^2 \overline{V}, \overline{\nu}).$$

The prior hyper-parameter values are $\underline{s}^{-2} = \frac{1}{var(y)}$, $\underline{\nu} = 3$, and $\underline{\beta} = 1$. Since

$$var(\beta_j) = \frac{\underline{\nu}\underline{s}^2}{\underline{\nu} - 2}\underline{V},$$

setting $var(\beta_j) = 10$ for both risk-return trade-offs and Fisher effects implies a choice of $\underline{V}$ as well. We use the same prior density for all scales $j = 1, \ldots, J$.

We wish to attach little weight on to the prior. We do so by setting $\underline{\nu}$ to a value which is considerably smaller than the number of observations N. Since, in the case of the risk-return trade-offs, $N = 64$, setting $\underline{\nu} = 3$ is relatively uninformative. We also attach a large prior variance to each of the regression coefficients. Because $var(\beta_j) = 10$, we are giving an approximately 95% prior probability to the region $[-2\sqrt{(10)}, 2\sqrt{(10)}]$ for β_j, which translates into a wide interval.

Overall, the selected prior is considerably less informative than the data. This can be seen from figure E.1 (i.e., the prior p.d.f. is more dispersed than the likelihood) - both for the case of stock returns and consumption risk - see Figure 1(a) - and for the case of nominal stock returns and inflation - see Figure 1(b).
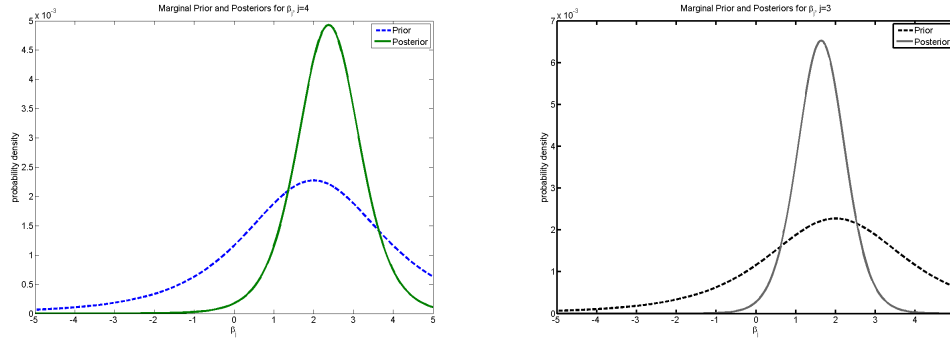
(a) Prior and posteriors for market returns and consumption volatility

(b) Prior and posteriors for nominal market returns and inflation

Figure E.1: The figure displays the marginal prior (dashed line) and posteriors for β_j. Panel A displays the case for the excess stock market returns and consumption risk at scale $j = 4$, and Panel B the relation between of nominal stock returns and inflation at scale $j = 3$.

In order to construct HPDIs for the Autoregressive AR(1) model for the variance, i.e.,

$$v^{2(j)}_{k2^j + 2^j} = \rho_j v^{2(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j + 2^j}$$

we assign prior distributions as follows:

- $\rho_j \sim 2\phi - 1$, where $\phi \sim Beta(a, b)$

- $h = 1/\sigma_j \sim Gamma(\underline{s}^{-2}, \underline{\nu})$

Again, we set uninformative priors. Naturally, the parameter ρ_j should be less than 1 in order for the process to be stationary. We use the beta distribution (which is, however, more informative than the uniform distribution)[12] and choose $a = 5$ and $b = 2$. We run a Random Walk Chain Metropolis-Hastings algorithm with 20,000 iterations to estimate the mean-reverting AR(1) model. We select independent normal proposal distributions to sample candidate points for the parameters. We choose the variance-covariance matrix of the multivariate Normal so that the acceptance probability tends to be neither too high nor too low. In particular, we make sure that the average acceptance probability is in the region 0.2 to 0.5.

F   Data

The empirical analysis in Sections 4, and 6 is conducted using annual data on consumption, inflation, stock returns, and short-term interest rates from 1930 to 2012, i.e., the longest available sample. We take the view that this sample is the most representative of the overall high/medium/low-frequency variation in asset prices and macroeconomic data.

Aggregate consumption is from the Bureau of Economic Analysis (BEA), series 7.1, and is defined as consumer expenditures on non-durables and services. Growth rates are constructed by taking the first difference of the corresponding logarithmic series. Our measure of consumption volatility is based on modeling consumption growth as following an AR(1)-GARCH(1,1), as in Bansal, Khatchatrian, and Yaron (2005).

For yearly inflation, we use the seasonally unadjusted CPI from the Bureau of Labor Statistics. Yearly inflation is the logarithmic (December to December) growth rate in the CPI.

We use the NYSE/Amex value-weighted index with dividends as our market proxy, R_{t+1}. Return data on the value-weighted market index are obtained from the Chicago Center for Research in Security Prices (CRSP). The nominal short-term rate ($R_{f,t+1}$) is the annualized yield on the 3-month Treasury bill taken from the CRSP treasury files.

The h-horizon continuously-compounded excess market return is calculated as $r_{t+1,t+h} = r_{t+1}^e + \ldots + r_{t+h}^e$, where $r_{t+j}^e = \ln(R_{t+j}) - \ln(R_{f,t+j})$ is the 1-year excess logarithmic market return between dates $t+j-1$ and $t+j$, R_{t+j} is a simple gross return, and $R_{f,t+j}$ is a gross risk-free rate (3-month Treasury bill).

The market's realized variance between the end of period t and the end of period $t+n$, a measure of integrated volatility, is obtained by computing

$$v_{t,t+n}^2 = \sum_{d=t_1}^{t_D} r_d^2,$$

where $[t_1, t_D]$ denotes the sample of available daily returns between the end of period t and the end of period $t+n$, and r_d is the market's logarithmic return on day d.

---

[12]It is known that U[0,1] is equivalent to Beta(1,1). The latter has values in the range of (0,1) with mean and standard deviation of 0.5 and 0.289, respectively.