



**cirano**

*Allier savoir et décision*

2015s-21

## **The scale of predictability**

*Federico M. Bandi, Benoit Perron, Andrea Tamoni,  
Claudio Tebaldi*

Série Scientifique/Scientific Series

2015s-21

## **The scale of predictability**

*Federico M. Bandi, Benoit Perron, Andrea Tamoni,  
Claudio Tebaldi*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal**  
**Mai 2015**

© 2015 *Federico M. Bandi, Benoit Perron, Andrea Tamoni, Claudio Tebaldi s.* Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère de l'Économie, de l'Innovation et des Exportations, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Économie, de l'Innovation et des Exportations, and grants and research mandates obtained by its research teams.*

### **Les partenaires du CIRANO**

#### **Partenaire majeur**

Ministère de l'Économie, de l'Innovation et des Exportations

#### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Intact  
Investissements PSP  
Ministère des Finances du Québec  
Power Corporation du Canada  
Rio Tinto Alcan  
Ville de Montréal

#### **Partenaires universitaires**

École Polytechnique de Montréal  
École de technologie supérieure (ÉTS)  
HEC Montréal  
Institut national de la recherche scientifique (INRS)  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

**ISSN 2292-0838 (en ligne)**

**Partenaire financier**

Économie,  
Innovation  
et Exportations

**Québec** 

# The scale of predictability<sup>\*</sup>

*Federico M. Bandi<sup>†</sup>, Benoit Perron<sup>‡</sup>  
Andrea Tamoni<sup>§</sup>, Claudio Tebaldi<sup>\*\*</sup>*

## Résumé/abstract

Stock return predictive relations found to be elusive when using raw data may hold true for different layers in the cascade of economic shocks. Consistent with this logic, we model stock market returns and their predictors as aggregates of uncorrelated components (details) operating over different scales and introduce a notion of scale-specific predictability, i.e., predictability on the details. We study and formalize the link between scale-specific predictability and aggregation. Using both direct extraction of the details and aggregation, we provide strong evidence of risk compensations in long-run stock 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. The reported tent-shaped behavior in long-run predictability is shown to be a theoretical implication of our proposed modelling approach.

**Mots clés/keywords** : long run, predictability, aggregation, risk-return trade-off, Fisher hypothesis.

**Codes JEL/JEL Codes** : C22, E32, E44, G12, G17

---

<sup>\*</sup>We are grateful to Yanqin Fan (the Toulouse discussant), Lars P. Hansen, C. Julliard, B. Kelly, L. Mancini (the EFA discussant), A. Neuberger and A. P. Taylor for their helpful comments. We thank seminar participants at the 2014 NBER-NSF Time Series Conference at the FRB of St. Louis, the 41st EFA in Lugano, the Workshop "Measuring and Modelling Financial Risk with High-Frequency Data" (Florence, 2014), the 2014 SoFiE Annual Conference in Toronto, the 2014 Financial Econometrics Conference at the Toulouse School of Economics, the 2014 Workshop on Uncertainty and Economic Forecasting at University College London, the 2013 NBER Summer Institute, the 2013 CIREQ Econometrics Conference on Time Series and Financial Econometrics, the Ninth CSEF-IGIER Symposium on Economics and Institutions, the 7th International Conference on Computational and Financial Econometrics, Aspect Capital, UPenn, Duke and the triangle, the London School of Economics, and Johns Hopkins University.

<sup>†</sup> Johns Hopkins University and Edhec-Risk Institute. E-mail: fbandi1@jhu.edu

<sup>‡</sup> Université de Montréal, Department of Economics and CIRANO. E-mail: benoit.perron@umontreal.ca.

<sup>§</sup> London School of Economics, Department of Finance. E-mail: a.g.tamoni@lse.ac.uk.

<sup>\*\*</sup> Università Bocconi. E-mail: claudio.tebaldi@unibocconi.it.

# 1 Introduction

It is generally accepted that the degree of stock-return predictability increases with the horizon. The introduction to the 2013 Nobel for Economic Sciences states: “*There is no way to predict whether the price of stocks and bonds will go up or down over the next few days or weeks. But it is quite possible to foresee the broad course of the prices of these assets over longer time periods, such as the next three to five years ...*”

The work on predictability assumes that all relevant economic information is embedded into shocks occurring at the highest frequency of observation. Mild short-term predictability and strong long-run predictability are, therefore, the “*reflection of a single underlying phenomenon*” (Cochrane, 2001).

An alternative view, put forward in this paper, is that low-frequency economic shocks may not just be long-run aggregates of high-frequency shocks. Consistent with this observation, low-frequency economic relations may not imply or require analogous relations at higher frequencies. In essence, every frequency of observation may be impacted by specific shocks and may, in consequence, carry unique information about the validity of economic relations.

To capture these ideas parsimoniously, we introduce the notion of *scale-specific predictability*. We view economic data as the result of a cascade of shocks with different magnitudes and different half lives. We provide a framework to de-couple the interconnection between short-run shocks as induced by, e.g., macroeconomic news announcements (Andersen, Bollerslev, Diebold and Vega, 2003) from those between medium and long-run shocks, as determined - for instance - by political phases (Santa-Clara and Valkanov, 2003), technological innovation (Hobijn and Jovanovic, 2001, Pastor and Veronesi, 2009, and Gârleanu, Panageas, and Yu, (2012)), and demographic changes (Abel, 2003, and Geanakoplos, Magill and Quinzii, 2004). In our approach, scale-specific predictability is a property of individual layers of the market returns and the predictors and is driven by scale-specific shocks. Importantly, we discuss how suitable aggregation can reveal scale-specific predictability and make it operational.

Intuitively, we model a generic economic time series  $x_t$  as a linear aggregate of  $J > 1$  scale-specific, uncorrelated, mean-zero components, or *details*,  $x_t^{(j)}$  with  $1 \leq j \leq J$ , in addition to a mean term  $\pi$ . We define the details as elements of the observed time series generated by *time-specific* (with  $t$  denoting, as always, time) and *scale-specific* (with  $j$  denoting scale) shocks. 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 longer, as is typical of long-run shocks.

Often-studied predictive relations - like the presumed dependence between market risk premia and expected volatility (risk-return trade-offs)<sup>1</sup> or between nominal returns and expected inflation (Fisher effects)<sup>2</sup> - are known to be hard to detect when using the raw series themselves. In the framework we propose, however, these elusive relations are found to apply to specific frequencies, levels of resolutions, or - in our jargon - *scales*. In essence, while the classical predictive regression  $y_{t+h} = \alpha + \beta x_t + u_{t+h}$  may lead to ambiguous outcomes for typical pairs  $\{y_{t+h}, x_t\}$  and an horizon  $h$ , the detail-wise regressions  $y_{t+h}^{(j)} = \beta x_t^{(j)} + u_{t+h}^{(j)}$ , with  $h = 2^j$  defining the  $j^{\text{th}}$ -scale's specific horizon, may imply a strong predictive link between hidden layers of the processes of interest.

These layers can be extracted using suitable filters, like the one-sided, linear Haar filter (e.g., Mallat, 1989). Extracting 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.

By virtue of direct identification of the pairs  $\{y_{t+h}^{(j)}, x_t^{(j)}\}$  for all  $1 \leq j \leq J$ , 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 may be interpreted as being consistent with classical economic logic.

To be specific, we show that lagged values of the market return variance's detail with a periodicity between 8 and 16 years 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

---

<sup>1</sup>See, *inter alia*, French, Schwert, and Stambaugh (1987), Campbell (1987), Campbell and Hentschel (1992), Glosten, Jagannathan, and Runkle (1993), Scruggs (1998), Harvey (2001), Ghysels, Santa-Clara, and Valkanov (2005), and Bali and Peng (2006).

<sup>2</sup>See, among others, Mishkin (1992) and Boudoukh and Richardson (1993).

variance's detail with periodicity between 8 and 16 years 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 8-to-16 year details, rather than on the original series, we find  $R^2$  values around 56%. 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 long-run prediction? We show formally that, should predictability on the scales occur, *long-run future* returns would be predicted by *long-run past* aggregates of the predictor. More explicitly, *two-way* (*forward* for the regressand, *backward* for the regressor) *adaptive aggregation*, as suggested by Bandi and Perron (2008), leads to increased predictive ability precisely at horizons corresponding to a scale, or level of resolution, over which the predictive relation is more likely to apply.

Since the work of Fama and French (1988) and Campbell and Shiller (1987, 1988), running predictive regressions of long-run returns on current values of the predictor is standard in the literature. Cochrane (2001) emphasizes the importance of multi-year regressions in illustrating the economic implications of forecastability, rather than as a tool to increase power in empirical assessments of predictability. The statistical pitfalls associated with long- (and short-)run prediction have been discussed by many (see, e.g., Stambaugh (1999), Valkanov (2003), Lewellen (2004), Campbell and Yogo (2006), Boudoukh, Richardson, and Whitelaw (2008), and the references therein).

In this paper, long-run returns are regressed on *past* long-run values of the predictor, rather than

on *current* values of the predictor. Not only are the regressions of forward aggregates on backward aggregates justified by our proposed data generating process, but also the scale-wise regressions are not affected by the statistical issues typically put forward to invalidate classical assessments of predictability (c.f., Section 5). In light of our reported empirical and theoretical correspondence between regressions on the details - and the spectral information they carry - and regressions on forward/backward averages, this observation is important.<sup>3</sup>

We discuss the *dual* role of aggregation. First, we illustrate how aggregation may work as a low-pass filter capable of eliminating high-frequency shocks, or short-term noise, while highlighting the low-frequency details to which predictive restrictions apply. In this sense, finding increasing predictability upon forward/backward aggregation is symptomatic of risk compensations which apply to highly persistent, low-frequency details of the return and variance process. 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. Second, aggregation provides a natural way to make scale-specific predictability operational. 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 translate predictability on the details into predictability for long-run market returns with its implication for, e.g., asset allocation for the long run.

We proceed as follows. Section 2 provides a graphical representation of our approach and findings. Section 3 reviews the literature. In Section 4 we model the data as a collection of details operating at different times and frequencies. The notion of scale autoregressive process (*scale-wise AR*) is discussed. Section 5 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. In Section 6, we discuss the dual 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 7 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 compensations due to market variance and low-frequency risk compensations due to

---

<sup>3</sup>For an analysis of the statistical properties of two-way regressions, we refer the interested reader to Bandi and Perron (2008) and the simulations in the present paper.

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 8 we verify the assumptions of theory using simulations. Section 9 turns to another well-known economic relation, namely the Fisher hypothesis. Our interest in the Fisher hypothesis is two-fold. First, we wish to show applicability of the methods to several, classical predictive problems. 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. Both in the case of the risk-return trade-offs and for Fisher effects, we report a marked *tent-shaped* behavior in the predictive slopes and  $R^2$ s. We show, using simulations and formal derivations, that this tent-like behavior represents an important implication of our assumed data generating process in Section 4. Section 10 returns to the notion of scale-specific shock and provides additional technical details. Section 11 concludes. Proofs are in the Appendices.

## 2 Scale-specific predictability: a visual representation and some intuition

Consider excess market returns and (market or consumption) variance. Assume the components (details) have been extracted. We will return to the extraction process in what follows. Figure 1 and 2 provide illustrations of the main ideas and findings.

The right panels present scatter plots of the details of excess market returns and market variance corresponding to frequencies between 1 and 2 years ( $j = 1$ , in our previous notation), 2 and 4 years ( $j = 2$ ), and 8 and 16 years ( $j = 4$ ), respectively. The left panels in Figure 1 present scatter plots of *forward* aggregates of returns on *backward* aggregates of market variance for different levels of aggregation. 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).<sup>4</sup> 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

---

<sup>4</sup>The wording “analogous” is sufficient to provide intuition here. In what follows, we provide a formal mapping.

aggregation) is a detection tool based on raw data and, importantly for our purposes, it may be viewed as a way to, again, 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 one-to-one mapping between scale-wise predictability (right panels) and long-run predictability upon two-way aggregation (left panels) is remarkable. The latter is justifiable by the former and, importantly, it would hardly be delivered by a classical predictive system. This is easy to see.

Assume  $y_{t+1}$  denotes future excess market returns and  $x_t$  denotes past market variance. A traditional predictive system (on a demeaned  $x_t$ ) would write:

$$y_{t+1} = \alpha + \beta x_t + u_{t+1}, \quad (1)$$

$$x_{t+1} = \rho x_t + \epsilon_{t+1}, \quad (2)$$

where  $u_{t+1}$  and  $\epsilon_{t+1}$  are possibly correlated shocks and  $0 < \rho < 1$ . When aggregating  $y_{t+1}$  forward and  $x_t$  backwards over an horizon  $h$ , the theoretical slope of the regression on forward/backward aggregates becomes  $\beta\rho^h$ , but  $\beta\rho^h \rightarrow 0$  as  $h \rightarrow \infty$ . The left panels in Figs. 1 and 2 show, instead, that the predictive slopes on aggregated backward/forward series increase strongly over a ten year horizon.

In stark contrast with the classical approach, we show formally that an analogous predictive system on suitable details of the raw series, i.e.,

$$y_{k2^j+2^j}^{(j)} = \beta_j x_{k2^j}^{(j)} + u_{k2^j+2^j}^{(j)}, \quad (3)$$

$$x_{k2^j+2^j}^{(j)} = \rho_j x_{k2^j}^{(j)} + \epsilon_{k2^j+2^j}^{(j)}, \quad (4)$$

with  $k \in \mathbb{Z}$  and  $1 \leq j \leq J$ , leads to the reported empirical pattern upon two-way aggregation of the raw series *provided* the dyadic frequency  $2^{j-1}$ ,  $2^j$  captures economic fluctuations over a 8-to-16 year horizon. The exact nature of this correspondence will be clarified below.

### 3 The literature

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.

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 our case) 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 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.

News arrivals at various frequencies, and the importance of shocks with alternative levels of persistence, also characterize the multifractal regime switching approach of Calvet and Fisher (2001, 2007). Multifrequency modelling and identification are conducted differently in the present paper. Importantly, our focus is on scale-specific economic relations and the role played by aggregation in their evaluation, rather than on inference, pricing and learning at high frequencies (daily, in Calvet and Fisher, 2007). To this extent, we explicitly spell out the conceptual links between the assumed scale-wise data generating process, its identification, and aggregation.

As in Hansen and Scheinkman (2009), we employ operators to extract low-frequency information. In our case, this is 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 non-overlapping, “fundamental” points (indexed, earlier, by  $k2^j$  with  $k \in \mathbb{Z}$ ), the result of an econometric process called “decimation”. Figure 1 and 2, right panels, are constructed using them. Similarly, the representation in Eqs. (3)-(4) hinges on decimation. These 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 framework, 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 predictability is broad. 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). The notion of predictability 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 the result of the same underlying phenomenon (Cochrane, 2001).

This paper offers an alternative mechanism through which predictability may arise, at certain frequencies *alone*, due to the presence of interconnected, across predictors and regressand, layers in the flow of economic shocks. It is this mechanism which justifies our concept of scale-wise predictability.

## 4 Time series modelling with scale-wise heterogeneity

We begin with an intuitive discussion of our proposed data generating process.

First, the informational flow at each scale  $j$  is represented by a mean-zero weakly-stationary (in scale time) component (or detail), denoted by  $x^{(j)}$ , localized with respect to time  $t$  and scale  $j$ . The evolution of the generic  $j^{\text{th}}$ -detail time series is specified on a dyadic time grid  $k2^j$  with  $k \in \mathbb{Z}$ . In particular, each detail is represented as an autoregressive process for which the conditional mean and the conditional variance are left unspecified, namely,

$$x_{k2^j+2^j}^{(j)} = \mu \left( x_{k2^j}^{(j)}, x_{k2^j-2^j}^{(j)}, \dots \right) + \sigma \left( x_{k2^j}^{(j)}, x_{k2^j-2^j}^{(j)}, \dots \right) \varepsilon_{k2^j+2^j}^{(j)} \quad \text{for all } j = 1, 2, \dots, J, \quad (5)$$

where the *scale-specific shocks*  $\varepsilon_{k2^j}^{(j)}$  are assumed to be uncorrelated across scales, white noise, mean zero and variance one. We note that every observation  $x_{k2^j}^{(j)}$  may be viewed as representing the outcome of informational flows between time  $k2^j - 2^j$  and time  $k2^j$ . The informational flows are driven by scale-specific shocks  $\varepsilon_{k2^j}^{(j)}$ . Importantly, these shocks may not be aggregates of higher

frequency shocks.

Second, we model an observed time series ( $x$ ) as the unique outcome of a *suitable* aggregation scheme applied to the details ( $x^{(j)}$ ). A sensible proposal would be to assume that the aggregation scheme is linear and each observation  $x$  can be expressed as an additive collection of scale-specific processes plus a mean term  $\pi$ :

$$x_t = \sum_{j=1}^J x_t^{(j)} + \pi. \quad (6)$$

However, Eq. (5) specifies the dynamics of details on a dyadic grid, i.e.,  $x_{k2^j}^{(j)}$ , while reconstruction of  $x_t$  using Eq. (6) would require that  $x_t^{(j)}$  is known for any time  $t$ , and hence on the finest grid, something which would be inconsistent with Eq. (5).

Two natural questions arise, the former having to do with *aggregation*, the latter having to do with *filtering*: Is the specification of the details on a coarse (scale-specific) grid sufficient to uniquely identify the observations  $x_t$  on the basis of a suitable aggregation scheme? Can the "inversion" of the adopted aggregation scheme provide a unique determination of the detail-specific observations  $\left\{ x_{t-2^j k_j}^{(j)} \right\}_{j=1, \dots, J, k_j \in \mathbb{N}}$  from observations on the aggregate process  $\{x_{t-n}\}_{n \in \mathbb{N}}$ ? These questions have a known, positive answer. Below, in order to *aggregate* the observations  $\left\{ x_{t-2^j k_j}^{(j)} \right\}_{j=1, \dots, J, k_j \in \mathbb{N}}$  into  $\{x_{t-n}\}_{n \in \mathbb{N}}$  (and *filter*  $\left\{ x_{t-2^j k_j}^{(j)} \right\}_{j=1, \dots, J, k_j \in \mathbb{N}}$  from  $\{x_{t-n}\}_{n \in \mathbb{N}}$ ) we employ a recursive algorithm introduced by Mallat (1993) and discussed in Appendix B. An intuitive application of the Mallat's algorithm produces the Haar matrix, the aggregation scheme used in this paper.

Eq. (6) clarifies the basic intuition behind our approach: we model a time series as an *additive* collection of scale-specific processes and, therefore, as a cascade of scale-specific shocks (Section 10 for more details).<sup>5</sup> At a more technical level, however, our adopted data generating process defines each observation  $x_t$  as a linear combination of scale-specific detail observations  $\left\{ x_{t-2^j k_j}^{(j)} \right\}_{j=1, \dots, J, k_j \in \mathbb{N}}$ . The coefficients of the linear combination are uniquely determined by the rows of the inverse Haar matrix. We call this data-generating process *scale-wise autoregressive (AR)*.

Econometrically, the technical approach that we propose (hinging on the Haar matrix) for

---

<sup>5</sup>One can show that, in the 2-component case, i.e.,  $x_t = \sum_{j=1}^2 x_t^{(j)} + \pi$ , if the details are AR(1) processes, the resulting process is an ARMA(6, 4). The presence of  $J > 2$  components would simply increase the order of the ARMA structure. Fuster, Laibson, and Mendel (2010) make a strong economic case for using high order structures *in spite of* the outcome of standard model-selection criteria, like AIC and BIC. These criteria are known to often recommend excessively low-order dynamics.

collecting the details and filtering them is only a possible one. The structure of the Haar filter is particularly helpful for us to formalize, below, the connection between scale-wise predictability and predictability upon two-way aggregation without introducing unnecessary complications. This said, any alternative multiresolution filter can be viewed as an aggregation scheme (e.g., Abry, Veitch, and Flandrin (1998)) and could be employed instead.

#### 4.1 Linear details

For our data, the *filtered* (using the Haar matrix) components are well described by a *linear* autoregressive process of order 1. We, therefore, adopt a convenient (2-parameter) linear autoregressive structure for them. Such a structure is not only justified empirically, but it will also allow us to draw a helpful comparison between classical predictive systems (in which the predictor is routinely modeled as an autoregressive process of order 1) and our proposed notion of scale-wise predictive system (Section 5). To this extent, we write

$$x_{k2^j+2^j}^{(j)} = \rho_j x_{k2^j}^{(j)} + \varepsilon_{k2^j+2^j}^{(j)}. \quad (7)$$

In essence, the details are autoregressive of order 1 in the dilated time of the scale being considered. The parameter  $\rho_j$  captures scale-specific persistence, the value  $\sigma_j$  represents the magnitude of economic shocks over different scales. From an inferential standpoint,  $\{\rho_j, \sigma_j\}$  can be readily estimated once the details have been extracted. Motivated by issues of signal processing akin to the economic issues of interest to us, Dijkerman and Mazumdar (1994) suggests an analogous linear dynamic structure for the details.

We note that dependence in scale time  $\rho_j$  may 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 of the details across scales is supported by the ability of the extraction filters we employ to “de-correlate” the original observations (see, e.g., Dijkerman and Mazumdar (1994) and Gençay, Selçuk, and Whitcher (2001)). Both features, i.e., low correlation in scale time and uncorrelatedness across scales, are verified in the data.

A data generating process with similar features to the one we proposed here 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 to provide an alternative view of cross-sectional asset

pricing by virtue of a novel notion of scale-specific beta (Bandi and Tamoni, 2013). In this paper we show that a scale-wise specification is key to explaining the long-run predictive features of the data. In essence, 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. In addition, we study formally the dual role of suitable (two-way) aggregation in detecting scale-specific economic restrictions and translating them into operational long-run features of the data.

Having made this point, we emphasize that our focus is *not* on *inference* at different frequencies, something which has routinely been tackled using spectral methods since the early work of Hannan (1963a,b). Our interest is on *modelling* and on the theoretical and empirical implications of the assumed specification for predictability at different horizons. Explicit localization in *both* time and scale is important for our purposes and is readily delivered by our proposed modelling approach. Scale-wise analogues to the traditional predictive systems, to which we now turn, are a central feature of the approach.

## 5 Scale-wise predictive systems

Consider a regressand  $y$  and a predictor  $x$ . Assume  $y$  and  $x$  are, as discussed above, additive (Haar-based) collections of details modelled as in Eq. (7). Assume, also, that for some  $1 \leq j \leq J$ ,

$$y_{k2^j+2^j}^{(j)} = \beta_j x_{k2^j}^{(j)} + u_{k2^j+2^j}^{(j)} \quad (8)$$

$$x_{k2^j+2^j}^{(j)} = \rho_j x_{k2^j}^{(j)} + \varepsilon_{k2^j+2^j}^{(j)}, \quad (9)$$

where  $u_{k2^j+2^j}^{(j)}$  is a scale-specific forecast error.

Eqs. (8)-(9) define a predictive system on individual layers of the  $\{y, x\}$  process to be contrasted with the traditional system written on the raw series. We will show that, for our data, the system is not affected by the inferential issues generally associated with predictability.

High first-order autocorrelation of the predictor, in particular, has been put forward as a leading cause of inaccurate inference in predictive systems (e.g., Stambaugh (1999), Valkanov (2003), Lewellen (2004), Campbell and Yogo (2006), and Boudoukh, Richardson, and Whitelaw (2008)). In our framework, however, we expect the magnitude of  $\rho_j$  to be smaller, the higher the scale and,

hence, the lower the frequency at which it operates. Consistent with this logic, at the low frequencies over which we identify scale-wise forecastability,  $\rho_j$  will be estimated to be small.

Before turning to empirical evaluations of Eqs. (8)-(9), we discuss the implications of the proposed data generating process for aggregation.

## 6 Two-way aggregation

It is standard in macroeconomics and finance to verify predictability by computing linear, or non-linear, 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  $\{y, x\}$  while being hidden by short-term noise. Aggregation provides a natural way to extract these components, filter out the noise, and generate a cleaner signal. Using the assumed data generating process, we now formalize this logic.

**Proposition.** *Assume that, for some  $j = j^*$ , we have*

$$\begin{aligned} y_{k2^{j^*}+2^{j^*}}^{(j^*)} &= \beta_{j^*} x_{k2^{j^*}}^{(j^*)}, \\ x_{k2^{j^*}+2^{j^*}}^{(j^*)} &= \rho_{j^*} x_{k2^{j^*}}^{(j^*)} + \varepsilon_{k2^{j^*}+2^{j^*}}^{(j)}, \end{aligned}$$

whereas  $\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\} = 0$  for  $j \neq j^*$ . We map decimated-time observations into calendar-time observations using the inverse Haar transform. Then, the forward-backward regressions

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

reach the maximum  $R^2$  of 1 over the horizon  $h = 2^{j^*}$  and, at that horizon,  $b_h = \beta_{j^*}$ .

**Proof.** See Appendix C.2.2.

For simplicity, we dispense with forecasting shocks  $u_t$  in the Proposition. Predictability applies to a specific  $j^*$  detail. All other details are set equal to zero.

The Proposition shows that predictability on the details implies predictability upon suitable

aggregation of both the regressand and the regressor. More explicitly, economic relations which apply to specific, low-frequency components will be revealed by two-way averaging. Adding short-term or long-term shocks in the form of uncorrelated details  $\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\}$ , for  $j < j^*$  or for  $j > j^*$ , or forecasting errors different from 0, would solely lead to a blurring of the resulting relation upon two-way estimation. We add uncorrelated details  $\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\}$  for  $j \neq j^*$  in the simulations in Section 8.

The Proposition also makes explicit the fact that the optimal amount of averaging should be conducted for time lengths corresponding to the scale over which predictability applies. More specifically, under the above assumptions, if predictability applies to a specific detail with fluctuations between  $2^{j^*-1}$  and  $2^{j^*}$  periods, an  $R^2$  of 1 would be achieved for a level of (forward/backward) aggregation corresponding to  $2^{j^*}$  periods. Before and after, the  $R^2$ s should display a *tent-like* behavior. At this horizon, the theoretical slope ( $b_{2^{j^*}}$ ) of the forward/backward regressions would coincide with the theoretical slope ( $\beta_{2^{j^*}}$ ) of the detail-wise regressions. This is an implication of our approach which will be verified in the data.

The proposed data generating process has, under the above assumptions, an additional implication worth mentioning. If predictability at the same  $j^{*th}$  scale applies, should forward predictors ( $y_{t+1,t+h}$ ) be regressed on *differences* of aggregated regressors ( $x_{t-h+1,t} - x_{t-2h+1,t-h}$ ), rather than on aggregated regressors ( $x_{t-h+1,t}$ ), a large  $R^2$  would be achieved for a level of aggregation corresponding to  $2^{j^*-1}$  periods, rather than  $2^{j^*}$  periods. The sign of the slope estimate upon two-way aggregation would also be the opposite of the sign of the theoretical slope linking details at the  $j^{*th}$  scale (see Appendix C-C.2.4). We will use this additional implication of theory to further validate the consistency between the assumed data generating process and the empirical findings in Section 9.

Next, we broaden the scope of classical predictability relations in the literature. We focus on risk-return trade-offs, and the low-frequency link between financial market variance and macro (consumption) variance. 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 theoretical 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.

## 7 Risk-return trade-offs

### 7.1 Equity returns on *market* variance

The basic observation driving our understanding of the analysis of risk-return trade-offs at different levels of aggregation 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 linear combination of decorrelated details whose spectrum is concentrated on an interval of characteristic time scales (inverse of frequencies) ranging from  $2^{j-1}$  to  $2^j$ .<sup>6</sup>

We apply the decomposition to logarithmic excess returns and realized variance series, i.e.,  $y_t = r_t$  and  $x_t = v_t^2$ .<sup>7</sup> 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 for both series. Virtually all correlations are small and very statistically insignificant. Not surprisingly, the largest one (0.39) corresponds to the adjacent pair of variance scales  $j = 3$  and  $j = 4$ .<sup>8</sup>

[Insert Figure 3 about here]

[Insert Table 2 about here]

---

<sup>6</sup>For a clear interpretation of the  $j$ -th scale in terms of the corresponding time span, we refer to Table 1.

<sup>7</sup>Appendix F describes the data and the construction of the variables.

<sup>8</sup>It is worth emphasizing that these pair-wise correlations are obtained by using overlapping or *redundant* data on the details rather than the non-overlapping or *decimated* counterparts described in Subsection A.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 overstate dependence.

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.

Next, we consider the forward/backward regressions

$$r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t}^2 + u_{t,t+h}, \quad (10)$$

where  $r_{t+1,t+h}$  and  $v_{t-h+1,t}^2$  are aggregates of excess market returns and return variances over an horizon of length  $h$ . Empirical results are displayed in Table 3-Panel A1 (horizons 1 to 10 years) and Table 3-Panel A2 (horizons 11 to 20 years). In Table 3-Panel A1, we confirm the findings in Bandi and Perron (2008) who report horizons of aggregations up to 10 years: *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 15.7% and 45.8%.

A crucial observation, for our purposes, is that the long-run results are not compatible with classical short-term risk return trade-offs. Disaggregated asset pricing models which solely imply dependence between excess market returns and (autoregressive) conditional variance at the *highest* resolution can hardly deliver the reported findings upon aggregation. As discussed in the Introduction, in fact, forward/backward aggregation of the system in Eqs. (1)-(2) would give rise to a theoretical slope equal to  $\beta\rho^h$ , but  $\beta\rho^h \rightarrow 0$  as  $h \rightarrow \infty$ , an outcome which is in stark contrast with the reported empirical evidence.

We argue that this argument points 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, we view the relation between risk and return as being *scale-specific*. In agreement with the implications of the discussion in Section 6, aggregation is helpful to reveal low-frequency, scale-wise risk compensations and make them operational.

To corroborate this logic, we run detail-wise predictive regressions as in Eqs. (8)-(9). The results are based on yearly data and are reported in Table 3-Panel B.<sup>9</sup> 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 between 8 and 16 years. For  $j = 4$ , the  $R^2$  of the detail-wise predictive regression is a considerable 56%.

---

<sup>9</sup>Given a scale  $j$ , we work with an effective sample size of  $[T/2^j]$  observations, where  $[\cdot]$  denotes the smallest integer near  $T/2^j$ . In this empirical analysis, we consider a sample spanning the period 1930 to 2013 and  $J = 4$ . We estimate scale-wise predictive systems using annual observations with five possible starts: 1930-2009, 1931-2010, ..., 1934-2013. Mean values of the estimates are reported. In practice, any choice of year start would yield very similar figures.

An important implication of theory is that, should predictability apply to a specific detail with fluctuations between  $2^{j-1}$  and  $2^j$  periods, the maximum  $R^2$  would be achieved for a level of (forward/backward) aggregation corresponding to  $2^j$  periods. Before and after, the  $R^2$  is expected to display a tent-shaped behavior. In our case,  $2^j = 16$  years. In Table 3-Panel A2 we extend the two-way regressions to horizons between 11 years and 20 years. Consistent with theory, the  $R^2$  values reach their peak (around 64%) between 14 and 16 years. Remarkably, the  $R^2$  structure, before and after, is roughly tent-shaped. The estimated beta values obtained by aggregating over an horizon of 14 to 16 years (1.55 to 1.38) are also numerically close to the estimated beta on the detail-wise regressions (1.16).

The study of low frequency relations is made difficult by the limited availability of observations over certain, long horizons. We do not believe that this difficulty detracts from the importance of inference at low frequency, provided it is conducted carefully. Importantly for the purposes of this paper, however, here we do not solely focus on low frequency dynamics. The reported tent-shaped behavior (a crucial by-product of the assumed data generating process) requires the dynamics at *all* frequencies to cooperate effectively, i.e., even at those high frequencies for which data availability would not be put forward as a statistical concern.

Standard economic theory views the presence of a market risk-return trade-off as compensation for variance risk. Given this logic, for past variance to affect future expected returns, higher past variance should predict higher future variance. Over the very long run, this is not the case in the raw data (see Appendix D). Remarkably, however, at the scale over which we report predictability (i.e., the 8 to 16 year scale), we find a positive dependence between past values of the variance detail and future values of the same detail (Table 3-Panel C). The  $R^2$  of the detail-wise variance autoregression is a rather substantial 15.3% with a positive slope of 0.04. As explained earlier, it is unsurprising to find a low *scale-wise* (for  $j = 4$ ) autocorrelation. While the autocorrelation 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. As shown in Appendix C.2.1, the corresponding autocorrelation in calendar-time would naturally be higher. In light of the examined long horizon, it is also unsurprising to find a large estimation error associated with the reported autocorrelation.

Importantly for our purposes, the documented low dependence between past and future variance dynamics at frequencies over which predictability applies differentiates our inferential problem from classical assessments of predictability. It is typically the case that high persistence of the

predictor makes classical inference incorrect (e.g., Valkanov (2003)). This issue does not arise in our framework.

[Insert Table 3 about here]

To summarize, 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 compensations for variance risk. While this logic applies to a specific level of resolution in our framework, it translates - upon aggregation - into predictability for long-run returns as shown formally in Section 6 and in the data.

We now turn to consumption variance.

## 7.2 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.<sup>10</sup>

Running detail-wise predictive regressions leads again to maximum predictability (and an  $R^2$  of 56%) associated with low-frequency cycles between 8 and 16 years, i.e.,  $j = 4$  (Table 4-Panel B). Similarly, for  $j = 4$ , a detail-wise autoregression of future consumption variance on past consumption variance yields a positive (and larger than in the market variance case) autocorrelation of 0.15 and an  $R^2$  value of about 56%.

Coherently with the implications of theory, two-way aggregation generates the largest  $R^2$  values over horizons close to 16 years (Table 4-Panel A1 and Panel A2). The largest  $R^2$  is obtained at the 14 year horizon (61.87%) but the values between 12 years and 16 years are all consistently between 60% and 57%. For shorter and longer horizons, the  $R^2$ s decline with a tent-shaped, almost monotonic, structure. They are between 0 and 3% over time periods between 1 and 5 years and close to 7% over the 20-year horizon.

Again, two-way aggregation leads to slope estimates which relate, in terms of their numerical value, with the slope estimates of the corresponding scale-wise predictive regressions. Over horizons

---

<sup>10</sup>We note that the “de-correlation” property of the details strongly applies to consumption variance (see Table 2).

between 8 and 16 years, we find estimated slopes ranging between 3.72 and 5.76. The slope estimate on the detail-wise predictive regression is 4.76.

**[Insert Table 4 about here]**

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 - operates now through the economically-appealing channel of consumption risk.

### **7.3 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.

While this result appears theoretically 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 (see Harvey (2001), Brandt and Kang (2004), Ludvigson and Ng (2007)). 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.

## 8 Simulating scale-specific predictability

One important observation about two-way aggregation is in order. One may argue that, by generating 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.<sup>11</sup> In other words, contemporaneous aggregation does not lead to any of the effects illustrated earlier (including consistency between the slope estimates obtained from the aggregated series and from the details). In addition, spurious behavior would prevent a tent-shape pattern from arising in the t-statistics and  $R^2$  from predictive regressions on the aggregated series because it would simply lead to (approximate, at least) upward trending

---

<sup>11</sup>The corresponding tables are not reported for conciseness but can be provided by the authors upon request.

behavior in both. As shown earlier in the data, as well as formally in Appendix C, Subsection C.2.2 and in the simulations below, however, tent-shaped patterns can easily arise.

In this section we establish, by simulation, that scale-wise predictability translates into predictability upon two-way aggregation. Supporting the implications of theory in Appendix C, we show that tent-shaped patterns are readily generated. 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 aggregation leads to insignificant outcomes. In sum, the findings discussed in this section provide support for a genuine 8-to-16 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:

$$\begin{aligned} r_{k2^j+2^j}^{(j)} &= v_{k2^j}^{2(j)} \\ v_{k2^j+2^j}^{2(j)} &= \rho_j v_{k2^j}^{2(j)} + \varepsilon_{k2^j+2^j}^{(j)} \end{aligned} \tag{11}$$

for  $j = j^*$  and

$$\begin{aligned} r_{k2^j+2^j}^{(j)} &= u_{k2^j+2^j}^{(j)} \\ v_{k2^j+2^j}^{2(j)} &= \varepsilon_{k2^j+2^j}^{(j)} \end{aligned}$$

for  $j \neq j^*$ , where  $k$  is defined as above and  $j = 1, \dots, J = 9$ . The shocks  $\varepsilon_t^{(j)}$  and  $u_t^{(j)}$  satisfy  $\text{corr}(u_t^{(j)}, \varepsilon_t^{(j)}) = 0 \forall t, j$ . 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  $j^{\text{th}}$  detail.

We note that the only conceptual difference between this simulation set-up and the assumptions in the Proposition is the addition of noise  $\{u_{k2^j+2^j}^{(j)}, \varepsilon_{k2^j+2^j}^{(j)}\}$  for scales  $j \neq j^*$ . As discussed, uncorrelated shocks will only lead to a blurring of the relation.

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 again formulated for non-overlapping or "decimated" data. As described earlier, we simulate the process at scale  $j$  every  $2^j$  steps and multiply it by the inverse Haar transformation in Appendix A.1 to obtain calendar-time observations. Appendix C, Subsection C.1 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 6, we will now show that a predictive relation localized around the  $j^{*th}$  scale will produce a pattern of  $R^2$ s which has a peak for aggregation levels corresponding to the horizon  $2^{j^*}$ .

### 8.1 Running the predictive regression

Table 5-Panel A shows the results obtained by running the regression in Eq. (10) on simulated data generated from Eq. (11). We compare these results to those in Table 6, 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  $R^2$ s 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 6<sup>th</sup> 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, Subsection C.2.2).

**[Insert Tables 5 and 6 about here]**

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  $R^2$ s 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 7 confirming the ability of suitable aggregation to detect scale-wise predictability over the relevant scale.

**[Insert Table 7 about here]**

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 5- and Table 7-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.

## 9 Fisher hypothesis

In its simplest form, the 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 somewhat mixed in terms of findings. For example, Mishkin (1992) and Fisher and Seater (1993) run regressions of  $h$ -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  $h$ -period nominal stock returns also on contemporaneous  $h$ -period expected inflation. In both cases, it is natural to test whether the slope of the predictive regression is equal to 1, as implied by theory. Both Mishkin (1992) and Boudoukh and Richardson (1993) find evidence supporting this null hypothesis. Using new asymptotic arguments, Valkanov (2003) re-examines their evidence: while in the pre-1979 period the effect is strong, after 1979 the effect is still present, but is not as convincing.

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, again, 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 0.42 with a t-statistic of 0.56 and an  $R^2$  of 0.74%. The 16-year beta is 0.75 with a t-statistic of 2.42 and an  $R^2$  of 16.6%. The 8-year beta, instead, is equal to 1.09 with a t-statistic of 3.34 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 25%. Its associated slope estimate is positive (2.10). So, is the autocorrelation coefficient (0.16) associated with the autoregression on the 3<sup>rd</sup> detail of the inflation process. We recall that the estimated autocorrelation is on non-overlapping low-frequency observations. The corresponding calendar-time autocorrelation would be considerably higher.

**[Insert Table 8 about here]**

Analogous findings apply to nominal interest rates (Table 9). With two-way aggregation, the 1-year beta is 0.37 with a t-statistic of 2.94 and an  $R^2$  of about 21%. The 8-year beta is, instead, twice as large and equal to 0.71 with a t-statistic of 3.13 and an  $R^2$  of about 35.53%. The tent-shaped pattern is even more marked than in the previous case. The corresponding detail ( $j = 3$ ) yields a predictive regression with a positive slope and an  $R^2$  value of 13.24%.

**[Insert Table 9 about here]**

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 two-way 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 7, 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 6, two-way aggregation should yield maximum predictability over an horizon close to  $2^3$ , i.e., close to 8 years. This is, in fact, largely consistent with data. It is also consistent with an additional set of simulations, 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 6 (and shown in Appendix C, Subsection C.2.2) predictable variation induced by a specific detail, like  $j = 3$  (here) or  $j = 4$  (in the previous section), would induce tent-shaped behavior upon aggregation. Another interesting implication of our 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), a large amount 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, Subsection C.2.4). We confirm both implications of theory with data.

Table 10-Panel A provides the corresponding results for nominal market returns. Leaving the 8 to 10 year horizon aside, 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 3<sup>rd</sup> scale. Again, leaving the 8 to 10 year horizon aside, the largest  $R^2$  is obtained for  $h = 4$ . The corresponding estimated slope is also negative and 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 lower frequency outcomes (over 8 to 10 years) 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  $R^2$ s are, also, somewhat larger.

**[Insert Table 10 about here]**

Before concluding, we turn to a more technical discussion about the nature - and interpretation - of scale-specific shocks.

## 10 Scale-specific shocks

Consider, using the terminology employed earlier, a *scale-wise AR process* with mean  $\pi$ . Since the details are autoregressive with uncorrelated (across time and scale) scale-specific shocks  $\varepsilon_t^{(j)}$ , the conventional Wold theorem, applied to the details, implies that each observation  $x_t$  can be

decomposed into a cascade of shocks, i.e.,

$$x_t = \sum_{j=1}^J \sum_{k=0}^{\infty} a_{j,k} \varepsilon_{t-k2^j}^{(j)} + \pi, \quad (12)$$

with  $a_{j,k} = E\left(x_t, \varepsilon_{t-k2^j}^{(j)}\right)$ . Hence, our assumed data generating process represents the idea that full information updates require the realization of the economic shocks affecting *all* frequencies.

It is interesting to observe that one can write an analogous decomposition (understood in the mean-squared sense) for *any* weakly stationary time series  $\{x_{t-i}\}_{i \in \mathbb{Z}}$ :

$$x_t = \sum_{j=1}^J \sum_{k=0}^{\infty} a_{j,k} \varepsilon_{t-k2^j}^{(j)} + \sum_{k=0}^{\infty} b_{J,k} \pi_{\varepsilon, t-k2^J, t-(k+1)2^{J+1}}^{(J)}, \quad (13)$$

where

$$\varepsilon_t^{(j)} = x_t^{(j)} - \mathcal{P}_{\mathcal{M}_{j,t-2^j}} x_t^{(j)} = \sqrt{2^j} \left( \frac{\sum_{i=0}^{2^j-1} \varepsilon_{t-i}}{2^{j-1}} - \frac{\sum_{i=0}^{2^j-1} \varepsilon_{t-i}}{2^j} \right) \quad (14)$$

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  $\left\{x_{t-k2^j}^{(j)}\right\}_{k \in \mathbb{Z}}$ ,

$$\pi_{\varepsilon, t-k2^J, t-(k+1)2^{J+1}}^{(J)} = \sqrt{2^J} \left( \frac{\sum_{i=t-(k+1)2^{J+1}}^{t-k2^J} \varepsilon_i}{2^J} \right),$$

with  $\varepsilon_t = x_t - \mathcal{P}_{\mathcal{M}_{t-1}} x_t$  satisfying  $Var(\varepsilon_t) = 1$ ,

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

and

$$b_{j,k} = E\left(x_t, \pi_{\varepsilon, t-k2^J, t-(k+1)2^{J+1}}^{(J)}\right).$$

Again, Eq. (13) is a (form of) Wold representation which applies to each scale and, hence, to the full process. Specifically, it explicitly describes any weakly-stationary time series of interest as a linear combination of shocks classified on the basis of their arrival time as well as their scale. We refer the reader to Wong (1993) for a similar decomposition. Appendix B shows formally that Eq. (13) reduces to the classical Wold representation for weakly-stationary time series.<sup>12</sup>

---

<sup>12</sup>Ortu, Severino, Tamoni, and Tebaldi (2015) use the Abstract Wold Theorem to provide a set of characterization theorems that classify the class of linear aggregation-decompositions schemes that can be achieved along these lines.

It is important to remark that the expression in Eq. (13) is neither *unique*, nor *economically-motivated*. Specifically, the expression hinges on the Haar filter. A different filter would give rise to an alternative expression for  $\pi_t^{(J)}$  as well as for the low-frequency shocks  $\varepsilon_t^{(j)}$  as *aggregates of high-frequency shocks* (c.f., Eq. (14)).

A crucial innovation of the approach advocated in this paper is to highlight that Eq. (13) (and, given their equivalency, the classical Wold representation for weakly-stationary process) can, in fact, be viewed as the result of *restrictions* on the shocks across scales. Eq. (14) is, in effect, a restriction. Our preferred approach in Eq. (12), instead, frees up the shocks in order to generate what we consider to be an economically-important separation between innovations - and information - at different scales.

## 11 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, predictive 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 offers 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 supply strong evidence about the long-run validity of certain predictive 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 and discount rate risks, 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

- ABEL, A. B. (2003): “The Effects of a Baby Boom on Stock Prices and Capital Accumulation in the Presence of Social Security,” *Econometrica*, 71(2), pp. 551–578.
- ABRY, P., D. VEITCH, AND P. FLANDRIN (1998): “Long-Range Dependence: revisiting Aggregation with Wavelets,” *Journal of Time Series Analysis*, 19, 253–266.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND C. VEGA (2003): “Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange,” *American Economic Review*, 93(1), 38–62.
- BANDI, F. M., AND B. PERRON (2008): “Long-run risk-return trade-offs,” *Journal of Econometrics*, 143(2), 349–374.
- BANDI, F. M., AND A. TAMONI (2013): “Scale-specific risk in the consumption CAPM,” *SSRN eLibrary*.
- 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.
- BOUDOUGH, J., AND M. RICHARDSON (1993): “Stock Returns and Inflation: A Long-Horizon Perspective,” *American Economic Review*, 83(5), 1346–55.
- BOUDOUGH, J., M. RICHARDSON, AND R. F. WHITELAW (2008): “The Myth of Long-Horizon Predictability,” *Review of Financial Studies*, 21(4), 1577–1605.

- BRANDT, M. W., AND Q. KANG (2004): “On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach,” *Journal of Financial Economics*, 72(2), 217–257.
- CALVET, L. E., AND A. J. FISHER (2001): “Forecasting multifractal volatility,” *Journal of Econometrics*, 105(1), 27–58.
- (2007): “Multifrequency news and stock returns,” *Journal of Financial Economics*, 86(1), 178–212.
- CAMPBELL, J. Y. (1987): “Stock returns and the term structure,” *Journal of Financial Economics*, 18(2), 373–399.
- CAMPBELL, J. Y., AND L. HENTSCHEL (1992): “No news is good news: An asymmetric model of changing volatility in stock returns,” *Journal of Financial Economics*, 31(3), 281–318.
- CAMPBELL, J. Y., 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 R. J. SHILLER (1987): “Cointegration and Tests of Present Value Models,” *Journal of Political Economy*, 95(5), 1062–88.
- CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): “Inflation Illusion and Stock Prices,” *American Economic Review*, 94(2), 19–23.
- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient tests of stock return predictability,” *Journal of Financial Economics*, 81(1), 27–60.
- 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.
- CROWLEY, P. M. (2007): “A Guide To Wavelets For Economists,” *Journal of Economic Surveys*, 21(2), 207–267.

- 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 (1988): “Dividend yields and expected stock returns,” *Journal of Financial Economics*, 22(1), 3–25.
- (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.
- FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): “Expected stock returns and volatility,” *Journal of Financial Economics*, 19(1), 3–29.
- FUSTER, A., D. LAIBSON, AND B. MENDEL (2010): “Natural Expectations and Macroeconomic Fluctuations,” *Journal of Economic Perspectives*, 24(4), 67–84.
- GÂRLEANU, N., S. PANAGEAS, AND J. YU (2012): “Technological Growth and Asset Pricing,” *Journal of Finance*, 67(4), 1265–1292.
- GEANAKOPOLOS, J., M. MAGILL, AND M. QUINZII (2004): “Demography and the Long-run Predictability of the Stock Market,” *Brookings Papers on Economic Activity*.
- 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.
- GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2005): “There is a Risk-Return Tradeoff After All,” *Journal of Financial Economics*, 76(3), 509 – 548.

- GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *The Journal of Finance*, 48(5), pp. 1779–1801.
- GRANGER, C. W. J., AND P. NEWBOLD (1974): “Spurious regressions in econometrics,” *Journal of Econometrics*, 2(2), 111–120.
- HANSEN, L. P., AND J. A. SCHEINKMAN (2009): “Long-term Risk: An Operator Approach,” *Econometrica*, 77(1), 177–234.
- HARVEY, C. R. (2001): “The specification of conditional expectations,” *Journal of Empirical Finance*, 8(5), 573–637.
- HOBIIJN, B., AND B. JOVANOVIĆ (2001): “The Information-Technology Revolution and the Stock Market: Evidence,” *American Economic Review*, 91(5), 1203–1220.
- 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.
- LEWELLEN, J. (2004): “Predicting returns with financial ratios,” *Journal of Financial Economics*, 74(2), 209–235.
- LUDVIGSON, S. C., AND S. NG (2007): “The empirical risk-return relation: A factor analysis approach,” *Journal of Financial Economics*, 83(1), 171–222.
- 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., F. SEVERINO, A. TAMONI, AND C. TEBALDI (2015): “A persistence-based Wold-type decomposition for stationary time series,” Working paper series.
- 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.
- PASTOR, L., AND P. VERONESI (2009): “Technological Revolutions and Stock Prices,” *The American Economic Review*, 99(4), pp. 1451–1483.
- 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. C. B. (1986): “Understanding spurious regressions in econometrics,” *Journal of Econometrics*, 33(3), 311–340.
- RAMSEY, J. B. (1999): “The contribution of wavelets to the analysis of economic and financial data,” *Phil. Trans. R. Soc. Lond. A*, 357(1760), 2593–2606.
- RENAUD, O., J.-L. STARCK, AND F. MURTAGH (2005): “Wavelet-Based Combined Signal Filtering and Prediction,” *IEEE Transactions SMC, Part B*, 35, 1241 – 1251.
- SANTA-CLARA, P., AND R. VALKANOV (2003): “The Presidential Puzzle: Political Cycles and the Stock Market,” *The Journal of Finance*, 58(5), pp. 1841–1872.
- SCHWERT, G. W. (1989): “Why Does Stock Market Volatility Change over Time?,” *Journal of Finance*, 44(5), 1115–53.

- SCRUGGS, J. T. (1998): “Resolving the Puzzling Intertemporal Relation between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach,” *The Journal of Finance*, 53(2), 575–603.
- STAMBAUGH, R. F. (1999): “Predictive regressions,” *Journal of Financial Economics*, 54(3), 375–421.
- TAMONI, A. (2011): “The multi-horizon dynamics of risk and returns,” *SSRN eLibrary*.
- VALKANOV, R. (2003): “Long Horizon Regressions: Theoretical Results and Applications,” *Journal of Financial Economics*, 68, 201–232.
- 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.

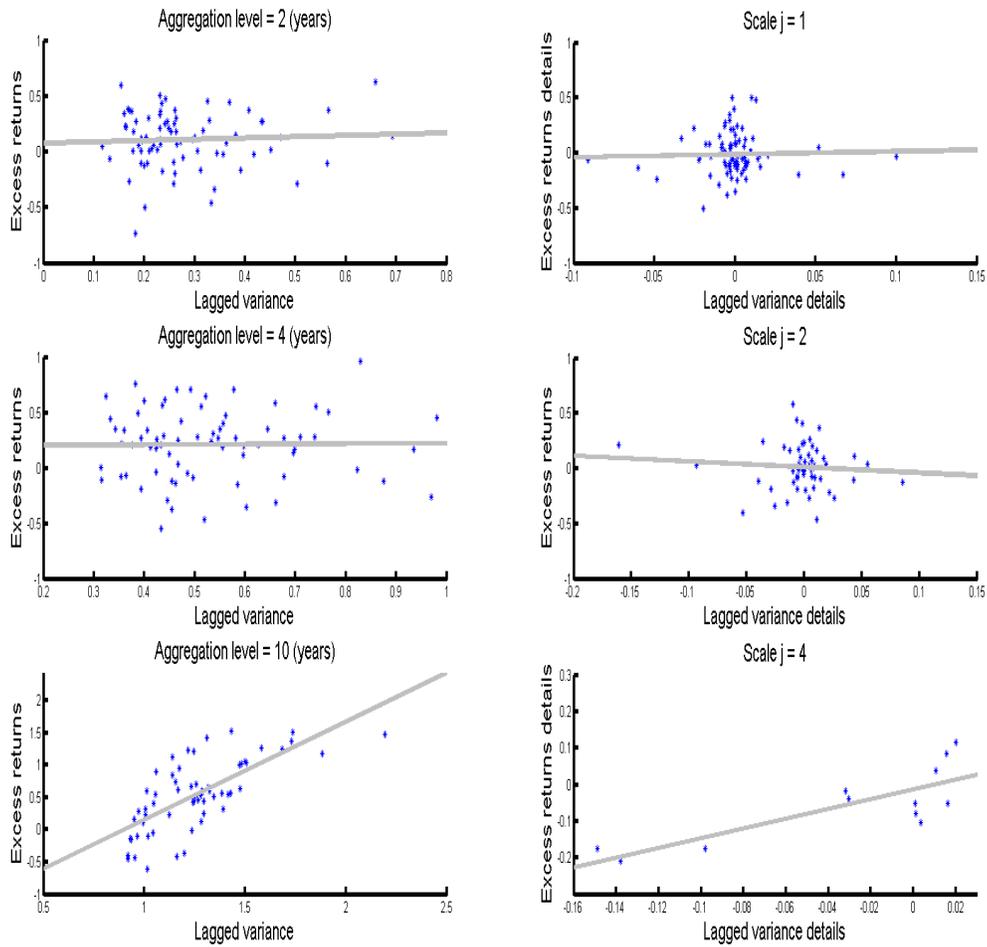


Figure 1: **Market Volatility.** The left panels present scatter plots of *forward* aggregates of excess market returns on *backward* aggregates of market variance for different levels of aggregation. The right panels present scatter plots of components (details) of the same series corresponding to analogous 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  $j = 1, 2, \dots$  into appropriate time horizons, please refer to Table 1.

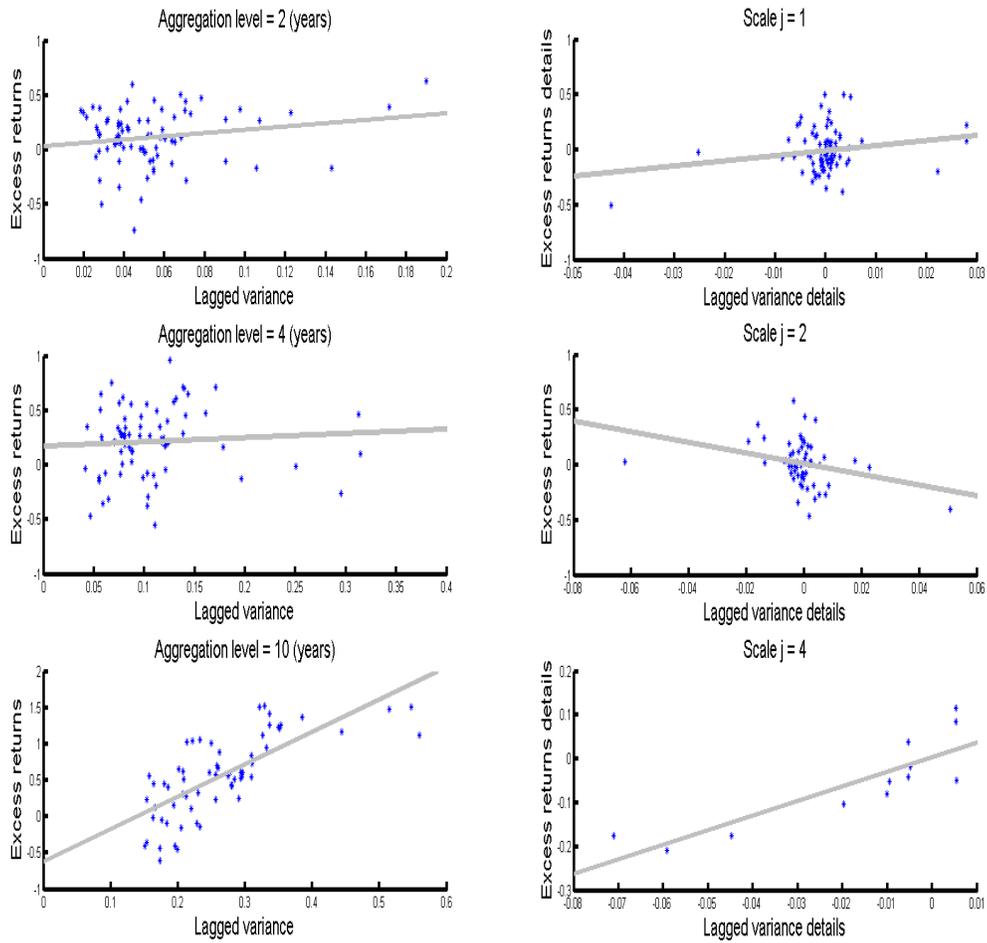
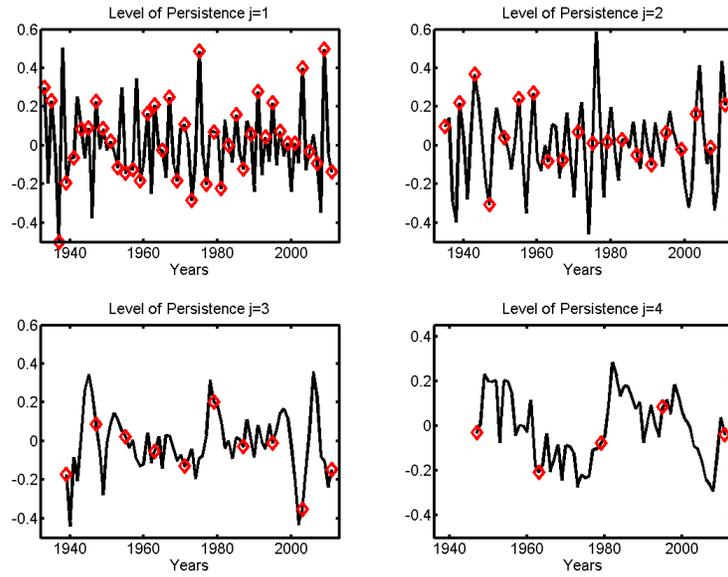
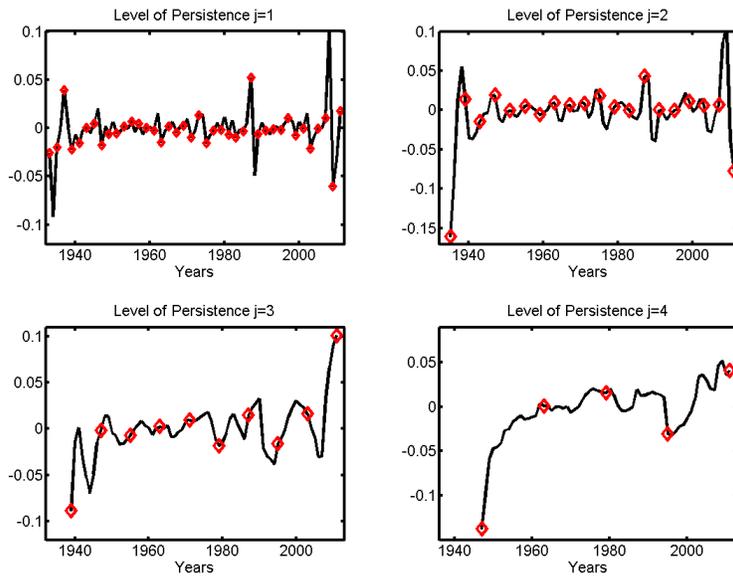


Figure 2: **Consumption Volatility.** The left panels present scatter plots of *forward* aggregates of excess market returns on *backward* aggregates of consumption variance for different levels of aggregation. The right panels present scatter plots of components (details) of the same series corresponding 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  $j = 1, 2, \dots$  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  $j = 1, 2, \dots$  into appropriate time horizons, please refer to Table 1.

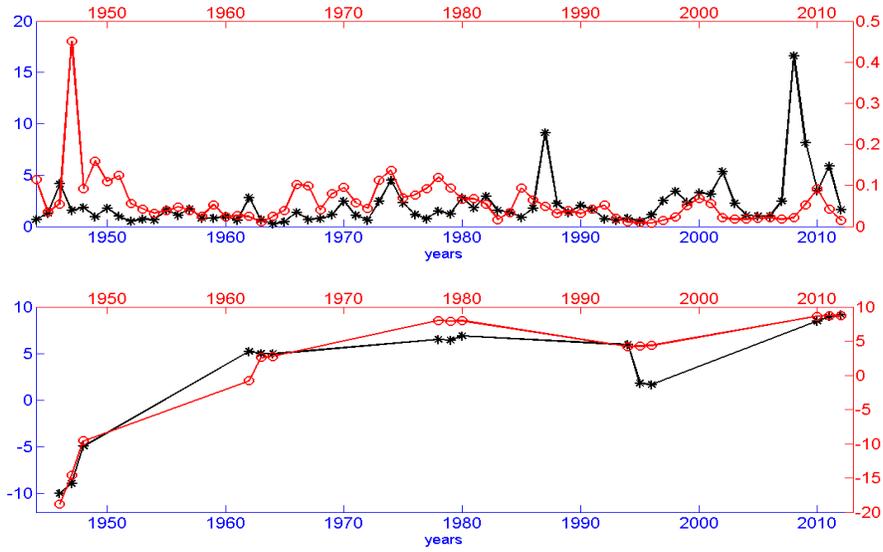


Figure 4: The upper panel in the figure displays the raw yearly series of market volatility (black-asterisk line) and consumption volatility (red-circle line); the lower panel displays the details of the two series with scale  $j = 4$ .

	<b>Panel A:</b>	<b>Panel B:</b>	<b>Panel C:</b>
	<b>Monthly calendar time</b>	<b>Quarterly calendar time</b>	<b>Annual calendar time</b>
Time-scale	Frequency resolution	Frequency resolution	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.

	<b>Panel A:</b>				<b>Panel B:</b>				<b>Panel C:</b>			
	<b>Market volatility</b>				<b>Consumption volatility</b>				<b>Market excess returns</b>			
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 A1:**  $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	1	2	3	4	5	6	7	8	9	10
$v_{t-h+1,t}$	0.13 (0.57)	0.10 (0.35)	-0.07 (-0.26)	0.05 (0.17)	0.29 (0.95)	0.41 (1.68)	0.54 (1.88)	0.81 (2.30)	1.17 (3.97)	1.40 (6.51)
$R^2(\%)$	[0.24]	[0.24]	[0.15]	[0.07]	[2.73]	[10.00]	[15.69]	[19.80]	[35.08]	[45.88]

**Panel A2:**  $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)									
	11	12	13	14	15	16	17	18	19	20
$v_{t-h+1,t}$	1.53 (8.86)	1.55 (8.53)	1.57 (9.17)	1.55 (10.35)	1.46 (9.77)	1.38 (8.94)	1.17 (7.68)	0.92 (5.83)	0.59 (4.61)	0.34 (2.77)
$R^2(\%)$	[50.39]	[53.75]	[59.58]	[64.75]	[63.57]	[63.78]	[53.68]	[37.02]	[16.05]	[5.85]

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

	Time-scale $j$			
	1	2	3	4
$v_t^{(j)}$	0.87 [-1.11 3.48]	0.25 [-1.53 3.18]	0.48 [-2.10 4.14]	1.16 [-0.19 2.89]
$R^2(\%)$	[1.36]	[2.42]	[3.02]	[56.13]

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

	Time-scale $j$			
	1	2	3	4
$v_t^{(j)}$	-0.15 [-0.37 0.07]	-0.12 [-0.37 0.14]	0.19 [-0.18 0.53]	0.04 [-0.24 0.32]
$R^2(\%)$	[7.50]	[11.09]	[21.31]	[15.32]

Table 3: **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  $j \in \{1, \dots, 4\}$ , we run a regression of the market variance component  $v_{t+2j}^{(j)}$  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 1930-2013. For the translation of time-scales into appropriate range of time horizons refer to Table 1.

**Panel A1:**  $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon $h$ (in years)									
	1	2	3	4	5	6	7	8	9	10
$v_{t-h+1,t}$	2.05 (1.06)	2.22 (1.69)	0.38 (0.32)	-0.04 (-0.04)	0.04 (0.03)	1.70 (1.16)	3.34 (2.63)	3.72 (2.70)	4.89 (3.94)	6.01 (5.19)
$R^2(\%)$	[1.54]	[3.37]	[0.16]	[0.00]	[0.00]	[4.46]	[17.71]	[22.51]	[33.59]	[46.59]

**Panel A2:**  $(r_{t+1,t+h} - rf_t) = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon $h$ (in years)									
	11	12	13	14	15	16	17	18	19	20
$v_{t-h+1,t}$	6.59 (5.32)	6.90 (5.75)	6.68 (5.54)	6.57 (6.56)	6.10 (6.92)	5.76 (7.63)	5.10 (8.11)	4.20 (7.25)	3.02 (5.04)	1.71 (2.69)
$R^2(\%)$	[53.95]	[60.40]	[60.89]	[61.87]	[57.74]	[56.93]	[50.78]	[38.21]	[21.17]	[7.72]

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

	Time-scale $j$			
	1	2	3	4
$v_t^{(j)}$	5.01 [-1.20 5.83]	-5.05 [-2.40 5.17]	1.12 [-1.70 5.56]	4.76 [-0.93 6.06]
$R^2(\%)$	[5.79]	[14.35]	[10.91]	[56.00]

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

	Time-scale $j$			
	1	2	3	4
$v_t^{(j)}$	-0.11 [-0.26 0.17]	0.19 [-0.02 0.47]	-0.03 [-0.14 0.11]	0.15 [0.04 0.29]
$R^2(\%)$	[3.87]	[9.97]	[35.10]	[55.87]

Table 4: **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  $v_{j,t}^2$ . **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 consumption variance component  $v_{t+2j}^{(j)}$  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 1930-2013. For the translation of time-scales into appropriate range of time horizons refer to Table 1.

**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 (-0.13)	-0.03 (-0.38)	-0.09 (-0.96)	-0.28 (-2.42)	-0.10 (-0.80)	0.43 ( 4.40)	0.71 ( 8.19)	0.67 ( 7.61)	0.44 ( 4.47)	0.18 ( 1.54)	0.04 (0.36)	-0.01 (-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 (0.00)	-0.01 (-0.07)	-0.02 (-0.21)	-0.06 (-0.57)	-0.14 (-1.16)	-0.26 (-1.85)	-0.34 (-2.32)	-0.33 (-2.26)	-0.21 (-1.64)	-0.09 (-0.78)	-0.03 (-0.27)	-0.01 (-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 5: **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  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  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  $2^*(\text{horizon}-1)$  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 (0.01)	-0.00 (-0.03)	0.00 (0.04)	0.00 (0.06)	0.00 (0.03)	0.00 (0.01)	0.00 (0.01)	0.00 (0.02)	0.00 (0.00)	-0.00 (-0.06)	-0.00 (-0.08)	-0.00 (-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 6: **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  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_{t,j}$  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  $2^*(\text{horizon}-1)$  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)	-0.01 (-0.09)	-0.02 (-0.21)	-0.07 (-0.61)	-0.15 (-1.21)	-0.22 (-1.63)	-0.21 (-1.48)	-0.08 (-0.54)	0.14 ( 1.05)	0.35 ( 2.92)	0.49 (4.43)	0.58 ( 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 (0.04)	0.00 (0.01)	-0.00 (-0.03)	-0.01 (-0.08)	-0.02 (-0.14)	-0.04 (-0.28)	-0.06 (-0.48)	-0.10 (-0.74)	-0.15 (-0.97)	-0.19 (-1.20)	-0.24 (-1.44)	-0.27 (-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 7: **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  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  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  $2^*(\text{horizon}-1)$  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 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	11	12	13	14	15	16
$\pi_{t-h+1,t}$	0.42 (0.56)	0.02 (0.05)	0.11 (0.22)	0.64 (3.41)	0.98 (5.39)	1.24 (5.84)	1.15 (4.74)	1.09 (3.34)	1.17 (3.68)	1.12 (2.75)	1.07 (2.18)	1.09 (2.22)	1.08 (2.31)	1.06 (2.85)	0.89 (2.58)	0.75 (2.42)
$R^2(\%)$	[0.74]	[0.00]	[0.14]	[6.29]	[14.71]	[27.64]	[28.66]	[27.95]	[31.85]	[28.62]	[25.68]	[26.33]	[26.52]	[26.75]	[21.54]	[16.60]

**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	11	12	13	14	15	16
$\pi_{t-h+1,t}$	-0.04 (-0.18)	-0.23 (-1.04)	-0.52 (-2.08)	-0.50 (-1.91)	0.07 (0.30)	0.77 (3.15)	1.19 (4.83)	1.36 (5.30)	1.26 (4.96)	0.98 (3.87)	0.61 (2.20)	0.27 (0.94)	0.10 (0.39)	0.04 (0.14)	0.02 (0.06)	0.04 (0.10)
$R^2(\%)$	[0.13]	[1.33]	[5.50]	[5.21]	[1.17]	[10.98]	[24.66]	[31.35]	[26.50]	[16.64]	[7.41]	[2.92]	[1.62]	[1.36]	[1.87]	[2.37]

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

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	0.99 [-1.13 3.11]	0.70 [-1.28 2.67]	2.10 [-0.26 4.46]	-0.10 [-1.35 1.14]
$R^2(\%)$	[0.92]	[7.71]	[25.43]	[30.83]

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

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.10 [-0.33 0.13]	0.17 [-0.13 0.48]	0.16 [-0.19 0.54]	-0.01 [-0.37 0.36]
$R^2(\%)$	[5.97]	[2.28]	[2.96]	[25.80]

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  $j \in \{1, \dots, 4\}$ , we run a regression of the inflation rate  $\pi_{t+2j}^{(j)}$  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 1930-2013. For the translation of time-scales into appropriate range of time horizons refer to Table 1 Panel A.

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

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\pi_{t-h+1,t}$	0.37 (2.94)	0.44 (2.50)	0.50 (2.59)	0.62 (2.83)	0.70 (2.97)	0.74 (3.03)	0.73 (3.06)	0.71 (3.13)	0.67 (3.19)	0.62 (3.13)	0.55 (2.80)	0.46 (2.24)	0.36 (1.69)	0.24 (1.13)	0.11 (0.53)	-0.04 (-0.18)
$R^2(\%)$	[20.85]	[24.05]	[26.69]	[32.37]	[37.12]	[39.41]	[38.71]	[35.53]	[31.32]	[26.03]	[19.74]	[13.29]	[7.95]	[3.72]	[0.83]	[0.10]

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

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.01 [-0.12 0.11]	-0.09 [-0.30 0.12]	0.20 [-0.42 0.88]	-0.02 [-0.87 0.82]
$R^2(\%)$	[0.06]	[10.96]	[13.24]	[4.26]

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

	Time-scale j			
	1	2	3	4
$\pi_t^{(j)}$	-0.10 [-0.33 0.13]	0.17 [-0.14 0.48]	0.16 [-0.22 0.54]	-0.01 [-0.37 0.35]
$R^2(\%)$	[5.97]	[2.28]	[2.96]	[25.80]

Table 9: **Risk-free rate and inflation.** **Panel A:** We run linear regressions (with an intercept) of  $h$ -period continuously compounded nominal risk-free rate  $r f_{t+1,t+h}$  on  $h$ -period past realized inflation. We consider values of  $h$  equal to 1 – 10 years. **Panel B:** results of componentwise predictive regressions of the components of nominal risk-free rate on the components of inflation  $\pi_t^{(j)}$ . **Panel C:** estimation results of the multiscale autoregressive system. The sample is annual and spans the period 1930-2013.

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

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$\Delta_h \pi_{t-h+1,t}$	-0.66 (-0.83)	0.37 (0.45)	-0.32 (-0.45)	-1.59 (-2.73)	-0.25 (-0.41)	0.25 (0.44)	0.50 (0.84)	0.80 (1.43)	1.04 (2.47)	1.45 (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]

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

	Horizon h (in months)									
	1	2	3	4	5	6	7	8	9	10
$\Delta_h \pi_{t-h+1,t}$	0.11 ( 0.31)	0.15 ( 0.58)	-0.67 (-2.49)	-1.22 (-4.93)	-0.39 (-1.63)	0.65 ( 1.80)	1.16 ( 4.30)	1.32 ( 4.43)	1.23 ( 4.33)	0.99 ( 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]

Table 10: **Forward on past differenced backward. Panel A: data.** We run linear regressions (with an intercept) of  $h$ -period continuously compounded nominal market returns on  $h$ -period past differenced inflation. **Panel B: simulation** We simulate market returns and inflation under the assumption of predictability at scale  $j^* = 3$ . We simulate  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_{t,j}$  for  $j = 3$  and  $x_t^{(j)} = \epsilon_{t,j}$  otherwise. We implement 500 replications. We set  $T = 128$ . For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with  $2*(\text{horizon}-1)$  lags in parentheses and adjusted  $R^2$  statistics in square brackets.

## Appendix

### A Filtering

This Appendix provides a concise introduction to the extraction procedure for the details.

Let  $\{x_{t-i}\}_{i \in \mathbb{Z}}$  be a time series of interest. 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 effectively amounts to breaking the series down 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 generic detail  $x_t^{(j)}$ , i.e.,

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

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, for every  $t$ ,  $\{x_{t-i}\}_{i \in \mathbb{Z}}$  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

$$x_t = \sum_{j=1}^J \underbrace{\left\{ \pi_t^{(j-1)} - \pi_t^{(j)} \right\}}_{x_t^{(j)}} + \pi_t^{(J)} = \pi_t^{(0)}, \quad (\text{A.1})$$

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. 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. As discussed in the main text, the representation in Eq. (A.1) is especially useful when discussing aggregation.

## A.1 Decimation

Decimation is the process of defining *non-redundant* information, as contained in a suitable number of *non-overlapping* “typical” points, in the observed details. Referring back 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 now return to the case  $J = 2$ , as in the example above, but similar considerations apply more generally. Define the vector

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

and consider the orthogonal Haar 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

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

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  $\{x_t^{(j)}, t = k2^j \text{ with } k \in \mathbb{Z}\}$  for  $j = 1, 2$  and  $\{\pi_t^{(2)}, t = k2^2 \text{ with } k \in \mathbb{Z}\}$ . Mallat (1989) provides a recursive algorithm for the general case with  $J$  not necessarily equal to 2.<sup>13</sup>

The separation of a time series in terms of details with different levels of resolution is conducted using wavelets as in Multiresolution Analysis (see, e.g., Mallat (1989), Dijkerman and Mazumdar (1994), Yazici and Kashyap (1997)). Wavelets analysis has been widely employed to study economic and financial time series (we refer to the comprehensive treatments in Ramsey (1999), Percival and Walden (2000), Gençay, Selçuk, and Whitcher (2001), and Crowley (2007) for discussions and references). Our use of multiresolution filters is, however, solely intended to facilitate extraction of scale-specific information. As emphasized, differently from the existing economic literature on wavelets, and its reliance on traditional time-series representations of the Wold type, once extracted, the components are thought to be driven by time and scale-specific shocks. In light of our discussion in the main text, the proposed data generating process is key to justify the reported results and the novel notion of scale-specific predictability introduced in this paper.

---

<sup>13</sup>In general, we can use the components  $x_t^{(j)}$ ,  $j = 1, \dots, J$ , and  $\pi_t^{(J)}$  in their entirety to reconstruct the time series using (A.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 C.

## B The classical Wold representation as a restriction on a generalized (time/scale) Wold representation

We begin with

$$x_t = \sum_{j=1}^J \sum_{k=0}^{\infty} a_{j,k} \varepsilon_{t-k2^j}^{(j)} + \sum_{k=0}^{\infty} b_{J,k} \pi_{\varepsilon, t-k2^J, t-(k+1)2^J}^{(J)},$$

where all the terms were defined in Section 1.

Consider the case  $J = 2$  and the interval  $[t, t - 7]$ . We have

$$\begin{aligned} x_t = & a_{1,0} \varepsilon_t^{(1)} + a_{1,1} \varepsilon_{t-2}^{(1)} + a_{1,2} \varepsilon_{t-4}^{(1)} + a_{1,3} \varepsilon_{t-6}^{(1)} + \dots + \\ & + a_{2,0} \varepsilon_t^{(2)} + a_{2,1} \varepsilon_{t-4}^{(2)} + \dots + \\ & + b_{2,0} \pi_{\varepsilon, t, t-3}^{(2)} + b_{2,1} \pi_{\varepsilon, t-4, t-7}^{(2)} + \dots \end{aligned}$$

Now, recall that

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

Then,

$$\begin{aligned} a_{1,0} &= E \left( x_t, \varepsilon_t^{(1)} \right) = E \left( x_t, \frac{\varepsilon_t}{\sqrt{2}} - \frac{\varepsilon_{t-1}}{\sqrt{2}} \right) = \frac{\psi_0}{\sqrt{2}} - \frac{\psi_1}{\sqrt{2}} \\ a_{1,1} &= E \left( x_t, \varepsilon_{t-2}^{(1)} \right) = E \left( x_t, \frac{\varepsilon_{t-2}}{\sqrt{2}} - \frac{\varepsilon_{t-3}}{\sqrt{2}} \right) = \frac{\psi_2}{\sqrt{2}} - \frac{\psi_3}{\sqrt{2}} \\ a_{1,2} &= E \left( x_t, \varepsilon_{t-4}^{(1)} \right) = E \left( x_t, \frac{\varepsilon_{t-4}}{\sqrt{2}} - \frac{\varepsilon_{t-5}}{\sqrt{2}} \right) = \frac{\psi_4}{\sqrt{2}} - \frac{\psi_5}{\sqrt{2}} \\ a_{1,3} &= E \left( x_t, \varepsilon_{t-6}^{(1)} \right) = E \left( x_t, \frac{\varepsilon_{t-6}}{\sqrt{2}} - \frac{\varepsilon_{t-7}}{\sqrt{2}} \right) = \frac{\psi_6}{\sqrt{2}} - \frac{\psi_7}{\sqrt{2}} \\ &\dots \\ a_{2,0} &= E \left( x_t, \varepsilon_t^{(2)} \right) = E \left( x_t, \frac{\varepsilon_t + \varepsilon_{t-1}}{2} - \frac{\varepsilon_{t-2} + \varepsilon_{t-3}}{2} \right) = \frac{\psi_0}{2} + \frac{\psi_1}{2} - \frac{\psi_2}{2} - \frac{\psi_3}{2} \\ a_{2,1} &= E \left( x_t, \varepsilon_{t-4}^{(2)} \right) = E \left( x_t, \frac{\varepsilon_{t-4} + \varepsilon_{t-5}}{2} - \frac{\varepsilon_{t-6} + \varepsilon_{t-7}}{2} \right) = \frac{\psi_4}{2} + \frac{\psi_5}{2} - \frac{\psi_6}{2} - \frac{\psi_7}{2} \\ &\dots \\ b_{2,0} &= E \left( x_t, \pi_{\varepsilon, t, t-3}^{(2)} \right) = \frac{\psi_0}{2} + \frac{\psi_1}{2} + \frac{\psi_2}{2} + \frac{\psi_3}{2} \\ b_{2,1} &= E \left( x_t, \pi_{\varepsilon, t-4, t-7}^{(2)} \right) = \frac{\psi_4}{2} + \frac{\psi_5}{2} + \frac{\psi_6}{2} + \frac{\psi_7}{2} \\ &\dots \end{aligned}$$

with

$$\psi_j = E(x_t, \varepsilon_{t-j}).$$

Finally, notice that

$$\begin{aligned}
\psi_0 \left( \frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t,t-3}^{(2)} \right) &= \psi_0 \varepsilon_t \\
\psi_1 \left( -\frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t,t-3}^{(2)} \right) &= \psi_1 \varepsilon_{t-1} \\
\psi_2 \left( \frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t,t-3}^{(2)} \right) &= \psi_2 \varepsilon_{t-2} \\
\psi_3 \left( -\frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t,t-3}^{(2)} \right) &= \psi_3 \varepsilon_{t-3} \\
\psi_4 \left( \frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4,t-7}^{(2)} \right) &= \psi_4 \varepsilon_{t-4} \\
\psi_5 \left( -\frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4,t-7}^{(2)} \right) &= \psi_5 \varepsilon_{t-5} \\
\psi_6 \left( \frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4,t-7}^{(2)} \right) &= \psi_6 \varepsilon_{t-6} \\
\psi_7 \left( -\frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4,t-7}^{(2)} \right) &= \psi_7 \varepsilon_{t-7},
\end{aligned}$$

which yields the standard Wold representation:

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

## C Understanding two-way aggregation

### C.1 Dynamics of time-scale components

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

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

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

For  $j = 1, \dots, J$ , 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:

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

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

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

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:

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

and

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

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

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

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

$$X_{16}^{(3)} = \begin{pmatrix} x_{16} = x_{16}^{(2)}/2 \\ x_{15} = x_{16}^{(2)}/2 \\ x_{14} = -x_{16}^{(2)}/2 \\ x_{13} = -x_{16}^{(2)}/2 \\ x_{12} = x_{12}^{(2)}/2 \\ x_{11} = x_{12}^{(2)}/2 \\ x_{10} = -x_{12}^{(2)}/2 \\ x_9 = -x_{12}^{(2)}/2 \end{pmatrix} \quad (\text{C.10})$$

and

$$X_8^{(3)} = \begin{pmatrix} x_8 = x_8^{(2)}/2 \\ x_7 = x_8^{(2)}/2 \\ x_6 = -x_8^{(2)}/2 \\ x_5 = -x_8^{(2)}/2 \\ x_4 = x_4^{(2)}/2 \\ x_3 = x_4^{(2)}/2 \\ x_2 = -x_4^{(2)}/2 \\ x_1 = -x_4^{(2)}/2 \end{pmatrix}. \quad (\text{C.11})$$

## C.2 Aggregation

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

Given the assumed data-generating process in scale time, we fit an AR(1) process in calendar time to  $x_t$ :

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

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

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

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

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

This result clarifies the relation between scale-wise persistence ( $\rho_{j^*}$ ) and persistence in calendar time ( $\tilde{\rho}$ ). If  $\rho_{j^*} < 1 - \frac{4}{2^{j^*+1}}$ , then  $\tilde{\rho} > \rho_{j^*}$  for all  $j^*$ . However, as  $j^*$  grows large,  $\tilde{\rho}$  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

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

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

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

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):

$$\begin{aligned}
\tilde{\beta} &= \frac{Cov(y_{15,16}, x_{13,14}) + Cov(y_{13,14}, x_{11,12})}{Var(x_{10,11}) + Var(x_{11,12}) + Var(x_{12,13}) + Var(x_{13,14})} \\
&= \frac{-\beta Var(x_{12}^{(2)}) \rho - \beta Var(x_{12}^{(2)})}{Var(x_{12}^{(2)}) + Var\left(\frac{x_{16}^{(2)}}{2}\right) + Var\left(\frac{x_{12}^{(2)}}{2}\right) - \frac{Cov(x_{16}^{(2)}, x_{12}^{(2)})}{2} + Var(x_{16}^{(2)})} \\
&= -2\beta \frac{(1 + \rho_j)}{5 - \rho_j}
\end{aligned}$$

and, hence, an inconsistent slope estimate. This estimate could have a changed sign (with respect to  $\beta$ ) 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:

$$\begin{aligned} y_{t+1,t+h} &= \beta(-x_{12}^{(2)} + x_8^{(2)})/2 \\ x_{t+1,t+h} &= (-x_{16}^{(2)} + x_{12}^{(2)})/2 \end{aligned}$$

By taking covariances we obtain

$$\begin{aligned} \tilde{\beta} &= \beta \frac{\left(-\text{Var}(x_{12}^{(2)}) + \rho_j \text{Var}(x_{12}^{(2)}) - \rho_j^2 \text{Var}(x_{12}^{(2)}) + \rho_j \text{Var}(x_{12}^{(2)})\right)}{\text{Var}(x_{16}^{(2)}) + \text{Var}(x_{12}^{(2)}) - 2\text{cov}(x_{16}^{(2)}, x_{12}^{(2)})} \\ &= \beta \frac{\left(-1 + 2\rho_j - \rho_j^2\right)}{2(1 - \rho_j)} \\ &= -\beta \frac{(1 - \rho_j)}{2}. \end{aligned}$$

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

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

The larger  $\rho_j$ , 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

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

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

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

If  $h = 3$ :

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

If  $h = 1$ :

$$\begin{aligned} \tilde{\beta}_{h=1} &= \frac{\beta}{2} \left[ \frac{1 + 3\rho_j}{3 + \rho_j} \right] \text{ and} \\ R_{h=1}^2 &= \frac{1}{4} \left[ \frac{1 + 3\rho_j}{3 + \rho_j} \right]^2 \left[ \frac{3 + \rho_j}{2} \right]. \end{aligned}$$

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  $R_{h=2}^2 = 1$  if  $\rho_j = 1$ . Also,  $R_{h=2}^2$  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,  $R_{h=2}^2 > R_{h=1}^2$  and  $R_{h=2}^2 > R_{h=3}^2$  for all  $\rho$ , thereby leading to a tent-shaped behavior of the  $R^2$ s 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 8.1 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  $v_{t+1,t+h}^2 = \sum_{i=1}^h v_{t+i-1,t+i}^2$  on  $h$ -period past variance  $v_{t-h+1,t}^2 = \sum_{i=1}^h v_{t-i+1,t-i}^2$ , i.e.,

$$v_{t+1,t+h}^2 = \rho_h v_{t-h+1,t}^2 + \varepsilon_{t,t+h}, \quad (\text{D.1})$$

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.<sup>14</sup>

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:

$$v_{k2^j+2^j}^{2(j)} = \rho_j v_{k2^j}^{2(j)} + \varepsilon_{k2^j+2^j}^{(j)}$$

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

$$v_{k2^j+2^j}^{2(j)} = \varepsilon_{k2^j+2^j}^{(j)}$$

for  $j \notin S$ ,  $j = 1, \dots, J = 9$ , where  $\text{corr}(\varepsilon_{t'}^{(j')}, \varepsilon_t^{(j)}) = 0 \forall t, j, t', j'$  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

<sup>14</sup>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 (5.30)	0.13 (2.10)	0.06 (0.66)	-0.26 (-2.56)	-0.43 (-4.04)	-0.47 (-4.57)	-0.48 (-3.68)	-0.48 (-3.80)	-0.49 (-4.34)	-0.49 (-3.96)	-0.48 (-4.04)	-0.48 (-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 (6.89)	0.24 (4.02)	0.22 (2.77)	-0.00 (-0.06)	-0.19 (-1.62)	-0.32 (-2.52)	-0.40 (-2.91)	-0.44 (-3.12)	-0.45 (-3.31)	-0.46 (-3.36)	-0.46 (-3.28)	-0.47 (-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  $2^*(\text{horizon}-1)$  lags in parentheses and adjusted  $R^2$  statistics in square brackets. **Panel A:** We simulate  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  for  $j = 6$  and  $x_t^{(j)} = \epsilon_t^{(j)}$  otherwise. **Panel B.** We simulate  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  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 (5.05)	0.34 (4.08)	0.33 (2.46)	0.28 (1.83)	0.21 (1.37)	0.20 (1.06)	0.23 (1.15)	0.20 (0.99)	0.13 (0.70)	0.05 (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

$$r_{k2^j, k2^j+2^j}^{(j)} = \begin{cases} v_{k2^j}^{2(j)} & \text{if } j = 7, \\ u_t^{(j)} & \text{otherwise .} \end{cases}$$

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  $R^2$ s at the horizon  $h = 120$ . Compared to Table 7–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:

$$y_{k2^j+2^j}^{(j)} = \beta_j x_{k2^j}^{(j)} + \varepsilon_{k2^j+2^j}^{(j)}.$$

Suppose that the coefficient  $\beta_j$  can lie anywhere in the interval  $(-\infty, \infty)$ . A 95% credible interval for  $\beta_j$  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  $\beta_j$  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  $\beta_j$  and  $z = \frac{1}{\sigma^2}$ :

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

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  $\beta_j$  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)	-0.01 (-0.10)	-0.02 (-0.25)	-0.09 (-0.69)	-0.19 (-1.34)	-0.26 (-1.80)	-0.23 (-1.59)	-0.09 (-0.61)	0.11 ( 0.81)	0.28 ( 2.23)	0.39 ( 3.27)	0.43 ( 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 (0.05)	0.00 (0.02)	-0.00 (-0.02)	-0.01 (-0.08)	-0.02 (-0.15)	-0.04 (-0.28)	-0.06 (-0.45)	-0.10 (-0.69)	-0.13 (-0.89)	-0.16 (-1.10)	-0.19 (-1.30)	-0.21 (-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 (5.27)	0.15 (2.46)	0.17 (2.21)	0.20 (1.84)	0.15 (1.16)	0.03 (0.19)	-0.05 (-0.37)	-0.07 (-0.54)	-0.07 (-0.59)	-0.07 (-0.60)	-0.09 (-0.66)	-0.12 (-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  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  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  $2*(\text{horizon}-1)$  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:

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

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

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

setting  $\text{Var}(\beta_j) = 10$  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  $j = 1, \dots, J$ .

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  $\text{Var}(\beta_j) = 10$ , we are giving an approximately 95% prior probability to the region  $[-2\sqrt{(10)}, 2\sqrt{(10)}]$  for  $\beta_j$ , 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).

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

$$v_{k2^j+2^j}^{2(j)} = \rho_j v_{k2^j}^{2(j)} + \varepsilon_{k2^j+2^j}^{(j)}$$

we assign prior distributions as follows:

- $\rho_j \sim 2\phi - 1$ , where  $\phi \sim \text{Beta}(a, b)$
- $z = 1/\sigma_j \sim \text{Gamma}(\underline{s}^{-2}, \underline{\nu})$

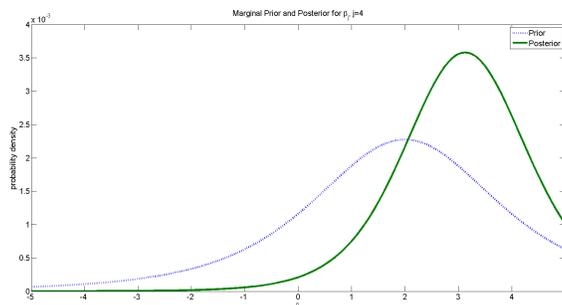
Again, we set uninformative priors. Naturally, the parameter  $\rho_j$  should be less than 1 in order for the process to be stationary. We use the beta distribution (which is, however, more informative than the uniform distribution)<sup>15</sup> and choose  $a = 5$  and  $b = 2$ . We run a Random Walk 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

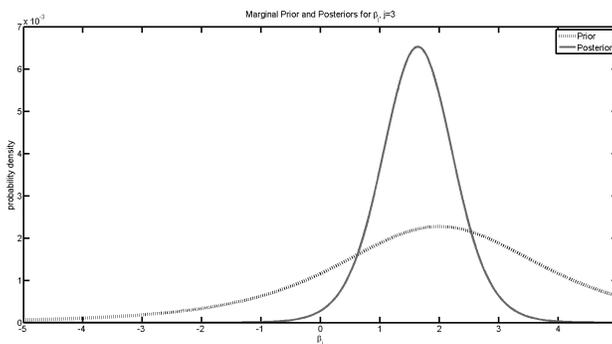
The empirical analysis in Sections 7, and 9 is conducted using annual data on consumption, inflation, stock returns, and short-term interest rates from 1930 to 2013, i.e., the longest available sample. We take the view that this sample is the most representative of the overall high/medium/low-frequency 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

<sup>15</sup>It is known that  $U[0,1]$  is equivalent to  $\text{Beta}(1,1)$ . The latter has values in the range of  $(0,1)$  with mean and standard deviation of 0.5 and 0.289, respectively.



(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  $\beta_j$ . 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  $j = 3$ .

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  $r_{t+1,t+h} = r_{t+1}^e + \dots + r_{t+h}^e$ , 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  $d$ .



1130, rue Sherbrooke Ouest, bureau 1400, Montréal (Québec) H3A 2M8

Tél. : 514-985-4000 • Téléc. : 514-985-4039

[www.cirano.qc.ca](http://www.cirano.qc.ca) • [info@cirano.qc.ca](mailto:info@cirano.qc.ca)

Centre interuniversitaire de recherche en analyse des organisations  
Center for Interuniversity Research and Analysis on Organizations