The scale of predictability*

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

October 4, 2016

Abstract

We introduce a new stylized fact: the hump-shaped behavior of slopes and coefficients of determination as a function of the aggregation horizon when running (forward/backward) predictive regressions of *future* excess market returns onto *past* economic uncertainty (as proxied by market variance, consumption variance, or economic policy uncertainty). To justify this finding formally, we propose a novel framework in which predictability is a property of low-frequency components of both excess market returns and economic uncertainty. We show that predictability on these low-frequency components (i.e., *scale-specific predictability*) translates theoretically into hump-shaped patterns of slopes and coefficients of determination upon forward/backward regressions on the raw series. If past long-run uncertainty predicts future long-run returns, it also has to predict future long-run dividend growth. We report that it does so strongly.

JEL classification: C22, E32, E44, G12, G17

Keywords: long run, predictability, aggregation, risk-return trade-off

*This paper was briefly circulated under the title "Economic uncertainty and predictability." We are grateful to Yanqin Fan (the Toulouse discussant), Lars P. Hansen, L. Mancini (the EFA discussant), A. Neuberger, A. Patton and A. P. Taylor for their helpful comments. We thank seminar participants at the 69th European Meeting of the Econometric Society in Geneva, the Australian Conference for Economists 2015 (Brisbane, 2015), 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 of 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, U Penn Economics, Duke Economics, LSE Finance, Johns Hopkins Economics, Tilberg University, Erasmus University, Tinbergen Institute, Rutgers Business School, CK Graduate School of Business and Tsinghua University.

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

‡Université de Montrèal, Department of Economics, CIRANO and CIREQ. E-mail: benoit.perron@umontreal.ca.

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

¶Universita' Bocconi. E-mail: claudio.tebaldi@unibocconi.it.

1 Introduction

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...*"

Hard-to-detect predictability over short horizons is generally viewed as the result of a low signal-to-noise problem. The magnitude of shocks to returns swamps predictable variation in expected stock returns. The aggregation of stock returns over longer horizons, however, operates as a signal extraction process uncovering predictability.

Existing work (Bandi and Perron (2008)) has highlighted the empirical usefulness of aggregating *both* the regressand (excess market returns) and the regressor (the market return predictor). Specifically, Bandi and Perron (2008) have suggested running adapted (to time t information) regressions of *forward* aggregated returns (i.e., long-run *future* returns) on *backward* aggregated predictors (long-run *past* market variance, in their case), rather than on raw predictors, as common in the literature. The use of forward/backward aggregation was shown to lead to a strengthening of variance-induced predictability over the long-run, the 10-year horizon being the longest prediction horizon considered in the paper. The long-run predictability of past variance was reported to be robust to the use of alternative variance notions (Tamoni (2011) uses consumption variance) and the dynamics of the variance process (Sizova, 2013, assumes long memory in variance). Among other stylized facts regarding stock returns, such predictability was justified in the context of an asset pricing model with loss aversion (see Bonomo, Garcia, Meddahi, and Tedongap (2015)).

We make four contributions. First, we show that the relation between *future* excess market returns and *past* uncertainty, as proxied by market variance (Bandi and Perron (2008)), consumption variance (Tamoni (2011)) and economic policy uncertainty (EPU, henceforth; see Baker, Bloom, and Davis (2013)), is *hump-shaped*. The forward/backward regressions are conducted over horizons of aggregation reaching 20 years, thereby doubling the 10-year horizon spanned in the existing work. The peak of predictability is around 16 years. Estimated slopes and R^2s feature increasing (resp. decreasing) dynamics before (resp. after) the 16-year mark. Around 16 years, the reported R^2s may reach a value of about 55%.

Second, we show that a traditional predictive system in which excess market returns are predicted

by a persistent uncertainty process would find it hard to replicate the structure and magnitude of the reported hump-shaped behavior upon two-way aggregation. Theory and simulations reported in the paper lead to this conclusion. If a traditional predictive system is an unlikely data generating process for the reported result, what would a more likely data generating process look like?

In his third contribution, this paper proposes a component-wise model in which returns and uncertainty can be viewed as linear aggregates of components operating over different frequencies. In the context of the model, predictability is not viewed as a property of the raw series. Rather, it is a property of individual return and uncertainty components. We dub it *scale-specific predictability*. Importantly, we show theoretically that should components with cycles of suitable lengths be linked by a predictability relation, then two-way aggregation would yield hump-shaped patterns in estimated slopes and R^2s. Empirically, we filter excess returns and uncertainty components and find predictability between components with cycles between 8 and 16 years. In agreement with the implications of theory, this scale-specific predictability should yield a hump-shaped pattern with a peak at 16 years upon two-way aggregation, as found in the data.

Fourth, we conceptualize the predictive ability of economic uncertainty within the classical Campbell and Shiller's present value identity (Campbell and Shiller (1988)).

What the present value identity does is, by construction, attributing a fundamental predictive role to the dividend-to-price ratio.[1] When seen through the lens of the identity, other successful predictors are, in fact, often viewed as proxies for it. As an example, alternative financial ratios, like earnings-to-price and book-to-market, perform reasonably well. Like in the case of dividend-to-price, they all have price in the denominator. Low prices are thought to predict high expected returns, thereby justifying their predictive performance as well as that of the more celebrated dividend-to-price ratio.

What the present value identity does *not* do is excluding the predictive ability of variables other than the dividend-to-price ratio. Yet, identifying variables capable to add to the predictive ability of the dividend-to-price ratio, particularly over the long run, is known not to be an easy task. The successful consumption-to-wealth ratio (see Lettau and Ludvigson (2001)), for instance, appears to change the term structures of short and medium-term return predictability, but does not lead to

[1] As emphasized by Cochrane (2008), ignoring possible bubble terms, the dividend-to-price ratio should predict returns, dividend growth or both. If it does not predict the latter, it ought to predict the former, and vice versa. Cochrane (2008), in particular, stresses that the predictive ability of the dividend-to-price ratio for long-run returns is economically compelling, long-run dividend growth predictability being not so.

significant long-run return forecasts (Cochrane (2011)).

We find that the *orthogonal* (with respect to the dividend-to-price ratio) component of past long-run uncertainty, whether given by market variance, consumption variance, or EPU, leads to superior return predictability over the long run.

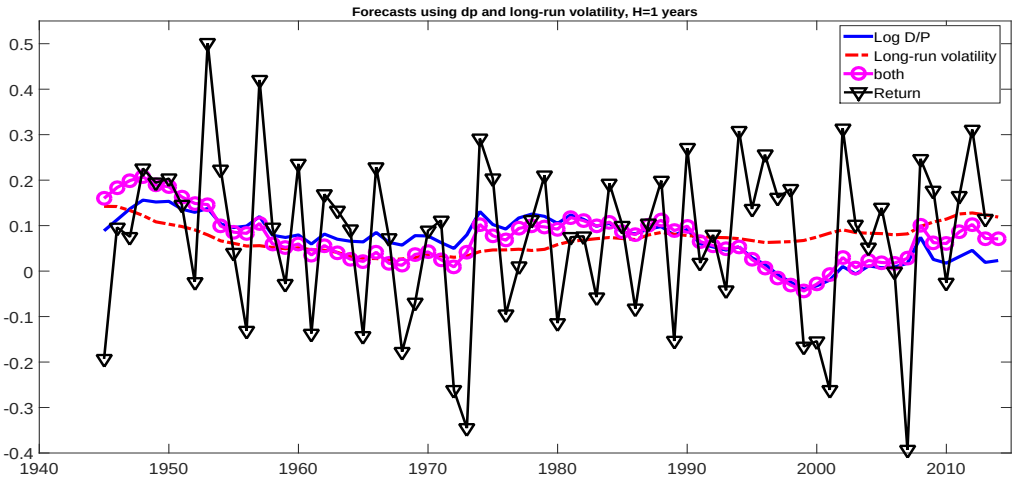
In order to immediately visualize matters, we report a graphical representation analogous to that provided by Cochrane (2011) in his Figs. 3 and 4. Figs. 1(a), 1(b) and 1(c), below, plot returns over three different horizons (1 year, 7 years, and 10 years) as well as return forecasts based on the dividend-to-price ratio *alone*, on a proxy for past long-run uncertainty (past long-run market variance, in this case) *alone*, and on the dividend-to-price and past long-run market variance *jointly*.

The effect is apparent. As we transition to longer horizons, past long-run market variance captures more and more of the slow-moving adjustments in long-run returns failed to be captured by the dynamics of the dividend-to-price ratio. The numbers are remarkable. At 1 year, the dividend-to-price ratio captures 10% of the variability in returns, long-run variance less than 3%. At 7 years, the R^2 associated with the dividend-to-price ratio reaches 35% and that associated with long-run variance 30%, at 10 years the corresponding numbers are near 42% for both. Said differently, the joint R^2 from a regression of 10-year returns on *both* variables is close to 85%.

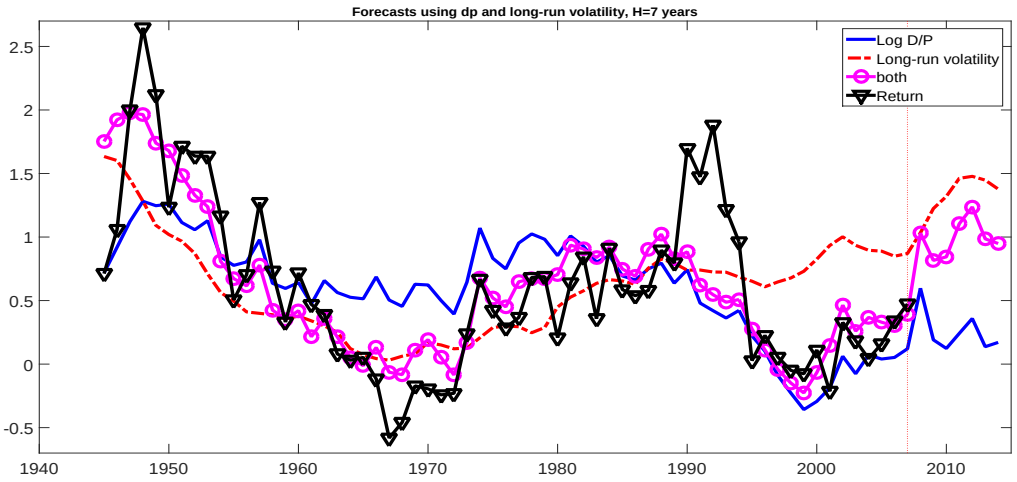
Past long-run uncertainty serves as a powerful long-run predictor, but improves predictability at all horizons. To highlight only a few more numbers using market variance again, the joint use of dividend-to-price and the orthogonal component of past long-run market variance leads to R^2 values higher than 40% at the 4 year horizon, higher than 50% at the 6-year horizon, higher than 70% at the 8-year horizon and higher than 80% at the 9-year horizon.

We show that the orthogonal component of past long-run uncertainty also predicts dividend growth strongly. Some numbers: at 5 years, 10 years, and 15 years, the R^2s from regressions of long-run dividend growth (without continuous compounding) onto past long-run market variance are 36%, 51%, and 60%, respectively. The corresponding numbers for the same regressions with the dividend-to-price ratio as a regressor are 0.9%, 3%, and 0.27%.

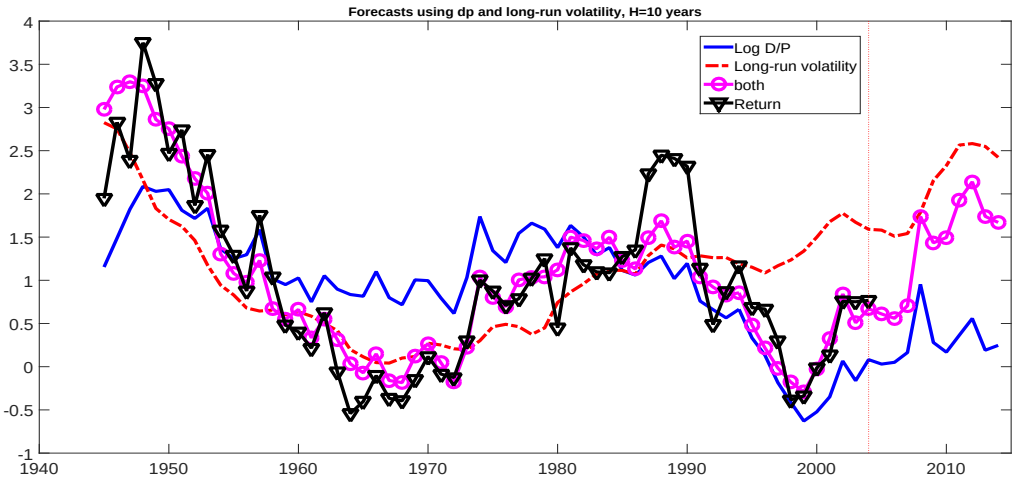
Because Campbell and Shiller's identity implies that all information about changes in long-run returns minus long-run dividend growth is contained in the dividend-to-price ratio, anything which is orthogonal to the dividend to price ratio, and predicts returns, should lead to an exactly offsetting prediction for dividend growth. Hence, one should expect to find, as we do, long-run dividend growth predictability in the presence of long-run return predictability.



(a) Forecast and actual 1-year returns



(b) Forecast and actual 7-year returns



(c) Forecast and actual 10-year returns

Figure 1: Actual vs. forecast returns at the $h = 1, 7$, and 10 year horizons. The forecasts are fitted values of regressions of returns on log dividend-price only (solid line), log market variance only (dashed line), and on log dividend-price and log market variance (line with circles). We line up the forecasts with the actual returns, i.e. we plot $\alpha_h + \beta_{x,h} x_t$ together with $R_{t+1,t+h}$.

Interestingly, we show in the paper that *indirect* long-run dividend growth regressions are a more informative test of long-run return predictability than the *direct* long-run return regressions themselves. This is true because of (1) the Campbell and Shiller identity, (2) the orthogonality of the regressor(s) and (3) the long-run predictability of the dividend-to-price ratio. These indirect dividend growth regressions confirm the long-run predictive ability of *past* economic uncertainty strongly.

2 Hump-shaped dynamics and *scale-specific predictability*: a preview

Fig. 2 plots R^2 values from regressions of forward aggregated excess market returns onto backward aggregated uncertainty (as proxied by either market variance, consumption variance, or (squared) EPU). The data is detailed in Appendix C.

The x axis is the common horizon over which forward and backward aggregation are conducted. We observe a marked hump-shaped behavior with a peak in the R^2 values reaching 55% around the 16-year horizon. Before and after, the R^2s have a roughly monotonic upward and downward trend, respectively.

Classical predictive system would find it hard to replicate the observed hump-shaped behavior. This is easy to see in theory. We will return to it using the simulations in Section 10. Assume y_{t+1} denotes future excess market returns and x_t denotes past uncertainty. A traditional predictive system (on a demeaned x_t) would write:

$$y_{t+1} = \alpha + \beta x_t + u_{t+1}, \tag{1}$$

$$x_{t+1} = \rho x_t + \epsilon_{t+1}, \tag{2}$$

where u_{t+1} and ϵ_{t+1} are possibly correlated, white noise shocks and $0 < \rho < 1$.

When aggregating y_{t+1} forward and x_t backward over an horizon h, the theoretical slope of the regression on forward/backward aggregates becomes $\beta\rho^h$, but $\beta\rho^h \to 0$ as $h \to \infty$.[2] Similarly, the

[2]The reported "slope" should be intended as the resulting slope from direct forward/backward iterations of the model. In light of the dependence between the regression residuals obtained by iteration and the backward-aggregated regressors, this slope does not coincide with that associated with the true conditional mean of forward-aggregated regressands onto backward-aggregated regressors. Such slope, for a large aggregation horizon h, would be approximately $\frac{\beta}{1+\rho}\frac{1}{h}$. Hence, it would also vanish as $h \to \infty$. We thank Nour Meddahi for discussions about this point.

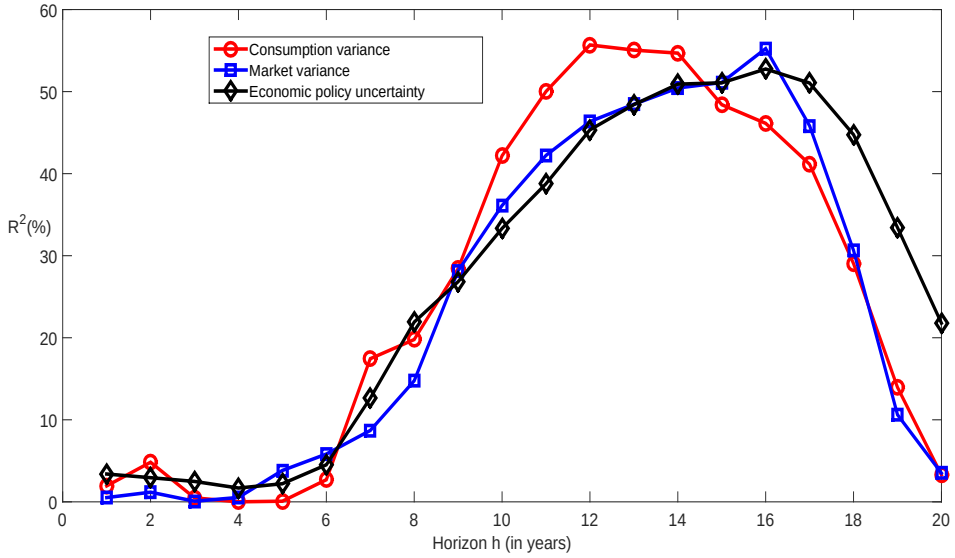


Figure 2: R^2 values obtained by regressing forward-aggregated excess market returns on backward aggregated market variance (blue line, with square), consumption variance (red line, with circles), and (squared) economic policy uncertainty (black line, with diamonds) for different levels of aggregation (on the horizontal axis).

R^2 should go to zero with the horizon of aggregation. Fig. 2 shows, instead, that the R^2 increases steeply to about 55% before decreasing equally sharply.

Assume now excess market returns and economic uncertainty, y_t and x_t, are linear aggregates of $J > 1$ uncorrelated, mean-zero components, or *details*, $y^{(j)}_{k2^j}$ and $x^{(j)}_{k2^j}$ with $k \in \mathbb{Z}$ and $1 \leq j \leq J$. The details are elements of the specific series generated by *scale-specific* (with j denoting scale) and *time-specific* (with $k2^j$ denoting "dilated" scale time for a running $k \in \mathbb{Z}$ and a fixed j) shocks. More explicitly, the detail associated with the j-th scale captures informational flow between time $k2^j - 2^j$ and time $k2^j$ $\forall k$. Lower scales are associated with higher frequencies and lower calendar-time persistence in the impact of sudden economic shocks (like, e.g., macroeconomic news announcements, Andersen, Bollerslev, Diebold and Vega, 2003). Higher scales, on the other hand, are affected by shocks which are relatively smaller in size but persist in the system longer, as is typical of medium and long-run shocks, such as those induced 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 contrast with the system above, below we demonstrate formally that an otherwise analogous *scale-wise predictive system* on details of the raw series, i.e.,

$$y^{(j)}_{k2^j+2^j} = \beta_j x^{(j)}_{k2^j} + u^{(j)}_{k2^j+2^j}, \tag{3}$$

$$x^{(j)}_{k2^j+2^j} = \rho_j x^{(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j}, \tag{4}$$

with $k \in \mathbb{Z}$ and $1 \leq j \leq J$, leads to the reported empirical pattern in R^2 values upon two-way aggregation of the raw series *provided* the dyadic frequency 2^{j-1}, 2^j over which scale-specific predictability applies captures economic fluctuations over a 8-to-16 year horizon.

We note that 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 to predictable variation in the return process. Predictability upon forward/backward aggregation is, instead, a way to translate scale-specific predictability into return predictability for the long haul, with all of its applied implications, including long-run asset allocation. Importantly, the mapping between the two is a natural outcome of scale-wise predictive systems, as our treatment below shows.

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 modeling 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 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 with its low-frequency implications.

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.

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"(see Section 6). 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.

In recent work, Bollerslev, Osterrieder, Sizova, and Tauchen (2013) also evaluate the relation between alternative notions of variance (and the variance risk premium) and market returns across frequencies. Their work focuses on dependencies over frequencies in excess of 3.5 months, between a day and 3.5 months, and intra-daily. Their emphasis is, therefore, on higher frequencies than in this paper. Over their assumed frequencies, their study finds mild compensations for variance risk, but a statistically significant compensation associated with the variance risk premium (Bollerslev, Tauchen and Zhou, 2009). We provide a framework to *jointly* explain mild risk-return trade-offs at high frequencies (as in Bollerslev et al., 2013, among others) and strong risk-return trade-offs at low frequencies, as in this paper. We show (in Proposition II below) that consistency between the two can be the outcome of a novel data generating process allowing for scale-specific predictability over suitable, low-frequency, scales.

Scale-wise specifications have proven successful in consumption models to explain the market risk premium (Ortu, Tamoni, and Tebaldi (2013) and Tamoni (2011)) and define granular notions of systematic risk through scale-wise betas (Bandi and Tamoni, 2013). This paper formalizes the

role of scale-wise specifications in the context of a new methodological approach to predictability.

The work on stock return predictability is broad. The literature documents return forecastability 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 return predictability has led to controversy (e.g., Welch and Goyal (2008), for a critique, and Cochrane (2008), for a well-known defense).

The literature on dividend growth predictability is less voluminous, the consensus being that growth rates of fundamentals, such as dividends or earnings, are less forecastable than returns when using financial ratios as predictors. Nonetheless, recent work shows that dividend growth is predictable by the dividend yield (see, e.g., van Binsbergen and Koijen (2010) and Rangvid, Schmeling, and Schrimpf (2014)) or by a consumption-(labor) income ratio, see Lettau and Ludvigson (2005). Koijen and Nieuwerburgh (2011) survey the return and dividend growth predictability debate through the lens of the present-value relation.

4 Low-frequency humps

4.1 Equity returns on *market* variance

The analysis is based on yearly data from 1930 to 2014. As indicated previously, Appendix C describes the data and the construction of the variables.

We begin with forward/backward regressions of long-run future excess market returns $y_t = r_t$ on long-run past variance $x_t = v_t$:

$$r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + u_{t,t+h}, \tag{5}$$

where $r_{t+1,t+h}$ and $v_{t-h+1,t}$ are aggregates of excess market returns and return variances over an horizon of length h. Empirical results are displayed in Table 1-Panel A1 (horizons 1 to 10 years) and Table 1-Panel A2 (horizons 11 to 20 years). Panel A1 and A2 report the estimated regression coefficient, the adjusted R^2 statistic in square brackets, and an heteroskedasticity and autocorrelation-consistent t-statistic for the hypothesis that the regression coefficient is zero (in parentheses). The

table also reports, in curly brackets, the rescaled t-statistic recommended by Valkanov (2003) to test the hypothesis that the regression coefficient is zero. Valkanov's methods have become standard tools in the predictability literature. We use them here to evaluate robustness. We recall, however, that they are justifiable under a classical data generating process (as in Eq. (1) and (2)), regressors near unity, and aggregation of the regressand, of the regressor, or both.[3]

[Insert Table 1 about here]

In Table 1-Panel A1, we report horizons of aggregations up to 10 years only: *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 8 and 10 years ranging between 14.7% and 36.1%.

In Table 1-Panel A2 we extend the two-way regressions to horizons between 11 years and 20 years. The R^2 values reach their peak (around 55%) at 16 years. The structure of the R^2s, before and after, is roughly tent-shaped (see Fig. 2). Using Valkanov's rescaled t-statistics as a metric, past market variance is a powerful predictor of future excess returns (leading to statistically significant slope estimates at the 2.5% level) for horizons ranging between 11 and 16 years.

4.2 Equity returns on *consumption* variance

Replacing market variance with consumption variance does not modify the previous results in a meaningful way. Two-way aggregation generates the largest R^2 values over horizons longer than 10 years (Table 2-Panel A1 and Panel A2). The largest R^2 is obtained at the 12 year horizon (55.7%) but the values between 11 years and 16 years are all consistently between 46% and 55%, with minimal differences between them. For shorter and longer horizons, the R^2s decline with a tent-shaped, almost monotonic, structure. They are between 0 and 5% over time periods between 1 and 5 years and close to 3% over the 20-year horizon.

[Insert Table 2 about here]

4.3 Equity returns on (squared) EPU

Results are provided in Table 3. They are analogous to the corresponding findings for market and consumption variance. We will not comment on them in detail.

[3] We discuss construction in Appendix D.

[Insert Table 3 about here]

Two observations are in order.

One may argue that, by generating stochastic trends, forward/backward aggregation could lead to spurious (in the sense of Granger and Newbold, 1974, and Phillips, 1986) predictability. If spuriousness were induced (somewhat mechanically) by aggregation, however, *contemporaneous* (i.e., forward/forward) 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.[4] 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 components). In addition, spurious behavior would prevent a tent-shape pattern from arising in the slopes, t-statistics and R^2 from predictive regressions on the aggregated series because it would simply lead to upward trending behavior in both. We will show that tent-shaped patterns are a natural by-product of an alternative data generating process. We will also return to these considerations in the simulation section.

Disaggregated asset pricing models which solely imply dependence between excess market returns and (autoregressive) conditional variance at the *highest* resolution can not easily deliver the reported findings upon aggregation. As discussed in Section 2, 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 $\beta\rho^h \to 0$ as $h \to \infty$, an outcome which is in contrast with the reported empirical evidence. As shown by Sizova (2013), a "large" ρ, captured by long memory in her framework, would help over horizons over which the statistics are reported as being monotonically increasing (1 to about 16 years). A long-memory variance process would, however, find it hard to capture the hump-shaped dynamics illustrated above and further discussed below.

Again, we argue that this argument points to an alternative data generating process, one to which we now turn.

5 Scale-specific shocks

The new data generating process will emphasize the differential role played by different scales. It is, therefore, important to introduce the notion of scale-specific shock. We do so by making explicit

[4]The corresponding tables are not reported for conciseness but can be provided by the authors upon request.

the connection between the traditional Wold representation of a covariance-stationary process and an alternative, *extended* Wold representation in which the series is broken down into scale-specific components and the shocks are, indeed, *scale-specific*.

Predictability will be a property of individual components and, therefore, will be induced by scale-specific shocks, which we now define.

5.1 A scale-wise representation of a covariance-stationary process

We begin with the classical Wold representation of a generic covariance-stationary time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$.

Formally, let $\mathbf{x}$ be a covariance-stationary process defined onto the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$. For simplicity, assume the process is mean zero. Adding the mean would, of course, simply add a constant to its Wold representation(s). The classical Wold representation of $\mathbf{x}$ states that there exists a unit variance white noise process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that, for any t in $\mathbb{Z}$,

$$x_t = \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k}, \tag{6}$$

where the equality is in the L^2-norm and $\{\alpha_k\}_{k\in\mathbb{N}_0}$ is a square-summable sequence of real coefficients with $\alpha_k = \mathrm{E}(x_t, \varepsilon_{t-k})$.

Let us now define the innovation process at scale j with $j \in \mathbb{N}$. If $j = 1$, the innovation process at scale 1, denoted by $\boldsymbol{\varepsilon^{(1)}} = \left\{\varepsilon_t^{(1)}\right\}_{t\in\mathbb{Z}}$, is the process whose terms are

$$\varepsilon_t^{(1)} = \frac{\varepsilon_t - \varepsilon_{t-1}}{\sqrt{2}}.$$

We observe that $\varepsilon_t^{(1)}$ has zero mean and its variance is equal to 1 for all t. More generally, we define the innovation process at scale j as the process $\boldsymbol{\varepsilon^{(j)}} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ with

$$\varepsilon_t^{(j)} = x_t^{(j)} - \mathcal{P}_{\mathcal{M}_{j,t-2^j}} x_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right) \tag{7}$$

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

The structure of the shocks $\boldsymbol{\varepsilon^{(j)}}$ is that of a Discrete Haar Transform (DHT). Once a scale

level j is set, in order not to have any overlaps between the shocks $\varepsilon_t^{(j)}$, one may consider the sub-series of $\boldsymbol{\varepsilon^{(j)}}$ defined on the support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$. The process $\boldsymbol{\varepsilon^{(j)}}$ is a moving average of order $2^j - 1$ with respect to the classical Wold innovations of $\mathbf{x}$. In calendar time, there is correlation between the variables $\varepsilon_{t-k2^j}^{(j)}$ and $\varepsilon_{\tau-k2^j}^{(j)}$ with $|t - \tau| \leq 2^j - 1$. Nonetheless, each sub-series $\left\{\varepsilon_{t-k2^j}^{(j)}\right\}_{k\in\mathbb{Z}}$ is a unit variance white noise process on the support $S_t^{(j)}$.

Technically, we are decomposing the space $\mathcal{H}_t(\boldsymbol{\varepsilon})$ of infinite moving averages whose underlying white noise process is $\boldsymbol{\varepsilon}$, i.e.,

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

into orthogonal subspaces:[5]

$$\mathcal{W}_t^{(j)} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in \mathbb{R} \right\}$$

with $j = 1, ..., +\infty$.

We now turn to the Wold coefficients at scale j. The process $\mathbf{x}$ is contained in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. Therefore, the decomposition of the space $\mathcal{H}_t(\boldsymbol{\varepsilon})$ induces the following decomposition of $\mathbf{x}$:

$$x_t = \sum_{j=1}^{+\infty} x_t^{(j)}, \tag{8}$$

where each random variable $x_t^{(j)}$ is the projection of x_t on the subspace $\mathcal{W}_t^{(j)}$, with $j \in \mathbb{N}$, and the equality is in the L^2-norm. By construction, for a fixed t, the components $x_t^{(j)}$ are orthogonal to each other. Since each $x_t^{(j)}$ belongs to $\mathcal{W}_t^{(j)}$, we have that

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

for some square-summable sequence of real coefficients $\left\{\psi_k^{(j)}\right\}_{k\in\mathbb{Z}}$.

Each coefficient $\psi_k^{(j)}$ is obtained by projecting $\mathbf{x}$ on the linear subspace generated by the (scale-

[5]See Theorem 1 in Ortu, Severino, Tamoni, and Tebaldi (2015), OSTT henceforth.

specific) innovation $\varepsilon^{(j)}_{t-k2^j}$:

$$\psi^{(j)}_k = \mathrm{E}\left[x_t \varepsilon^{(j)}_{t-k2^j}\right]. \tag{9}$$

Substituting the expression of $x^{(j)}_t$ into Eq. (8), we derive the *extended* Wold representation of x_t, that is[6]

$$x_t = \sum_{j=1}^{+\infty}\sum_{k=0}^{+\infty} \psi^{(j)}_k \varepsilon^{(j)}_{t-k2^j}. \tag{10}$$

Eq. (10) is a (type 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.

It is now interesting to discuss the connection between the coefficients $\psi^{(j)}_k$ of the *extended* Wold representation and the coefficients α_k of the *classical* Wold representation. Exploiting Eq. (9) after expressing $\varepsilon^{(j)}_t$ as a finite linear combination of variables ε_t as in Eq. (7) and using Eq. (6), we obtain

$$\psi^{(j)}_k = \frac{1}{\sqrt{2^j}}\left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}\right). \tag{11}$$

Just like the shocks $\varepsilon^{(j)}$s are Discrete Haar Transforms (DHTs) of the high-frequency shocks in a classical Wold (see Eq. (7)), the coefficients $\psi^{(j)}$s are DHTs of the coefficients in a classical Wold. Since the Wold coefficients α_k are unique, it follows that the coefficients $\psi^{(j)}_k$ are unique, time-invariant, functions of the coefficients α_h given the Haar structure.

We now turn to two examples.

Example 1: Unfolding the extended Wold (with $J = 2$).

We truncate J and set it equal to 2. Consider

$$x_t = \sum_{j=1}^{J}\sum_{k=0}^{\infty} \psi^{(j)}_k \varepsilon^{(j)}_{t-k2^j} + \sum_{k=0}^{\infty} \varphi^{(J)}_k \pi^{(J)}_{\varepsilon,t-k2^J},$$

where $\pi^{(J)}_{\varepsilon,t} \equiv \frac{1}{\sqrt{2^J}}\left(\sum_{i=0}^{2^J-1} \varepsilon_{t-i}\right)$ and $\varphi^{(J)}_k = \mathrm{E}\left(x_t, \pi^{(J)}_{\varepsilon,t-k\cdot 2^J}\right)$. This is an *extended* Wold which we are going to unfold to obtain a *classical* Wold. We work, for simplicity again, on the interval $[t, t-7]$.

[6] See Theorem 2 in Ortu, Severino, Tamoni, and Tebaldi (2015).

We have

$$\begin{aligned} x_t = \quad & \psi_0^{(1)}\varepsilon_t^{(1)} + \psi_1^{(1)}\varepsilon_{t-2}^{(1)} + \psi_2^{(1)}\varepsilon_{t-4}^{(1)} + \psi_3^{(1)}\varepsilon_{t-6}^{(1)} + ... + \\ & + \psi_0^{(2)}\varepsilon_t^{(2)} + \psi_1^{(2)}\varepsilon_{t-4}^{(2)} + \ldots + \\ & + \varphi_0^{(2)}\pi_{\varepsilon,t}^{(2)} + \varphi_1^{(2)}\pi_{\varepsilon,t-4}^{(2)} + \ldots \end{aligned}$$

Now, recall that

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

Thus,

$$\begin{aligned} \psi_0^{(1)} &= \mathrm{E}\left(x_t, \varepsilon_t^{(1)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_t}{\sqrt{2}} - \frac{\varepsilon_{t-1}}{\sqrt{2}}\right) = \frac{\alpha_0}{\sqrt{2}} - \frac{\alpha_1}{\sqrt{2}} \\ \psi_1^{(1)} &= \mathrm{E}\left(x_t, \varepsilon_{t-2}^{(1)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_{t-2}}{\sqrt{2}} - \frac{\varepsilon_{t-3}}{\sqrt{2}}\right) = \frac{\alpha_2}{\sqrt{2}} - \frac{\alpha_3}{\sqrt{2}} \\ \psi_2^{(1)} &= \mathrm{E}\left(x_t, \varepsilon_{t-4}^{(1)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_{t-4}}{\sqrt{2}} - \frac{\varepsilon_{t-5}}{\sqrt{2}}\right) = \frac{\alpha_4}{\sqrt{2}} - \frac{\alpha_5}{\sqrt{2}} \\ \psi_3^{(1)} &= \mathrm{E}\left(x_t, \varepsilon_{t-6}^{(1)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_{t-6}}{\sqrt{2}} - \frac{\varepsilon_{t-7}}{\sqrt{2}}\right) = \frac{\alpha_6}{\sqrt{2}} - \frac{\alpha_7}{\sqrt{2}} \\ & ... \\ \psi_0^{(2)} &= \mathrm{E}\left(x_t, \varepsilon_t^{(2)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_t + \varepsilon_{t-1}}{2} - \frac{\varepsilon_{t-2} + \varepsilon_{t-3}}{2}\right) = \frac{\alpha_0}{2} + \frac{\alpha_1}{2} - \frac{\alpha_2}{2} - \frac{\alpha_3}{2} \\ \psi_1^{(2)} &= \mathrm{E}\left(x_t, \varepsilon_{t-4}^{(2)}\right) = \mathrm{E}\left(x_t, \frac{\varepsilon_{t-4} + \varepsilon_{t-5}}{2} - \frac{\varepsilon_{t-6} + \varepsilon_{t-7}}{2}\right) = \frac{\alpha_4}{2} + \frac{\alpha_5}{2} - \frac{\alpha_6}{2} - \frac{\alpha_7}{2} \\ & ... \\ \varphi_0^{(2)} &= \mathrm{E}\left(x_t, \pi_{\varepsilon,t}^{(2)}\right) = \frac{\alpha_0}{2} + \frac{\alpha_1}{2} + \frac{\alpha_2}{2} + \frac{\alpha_3}{2} \\ \varphi_1^{(2)} &= \mathrm{E}\left(x_t, \pi_{\varepsilon,t-4}^{(2)}\right) = \frac{\alpha_4}{2} + \frac{\alpha_5}{2} + \frac{\alpha_6}{2} + \frac{\alpha_7}{2} \\ & ... \end{aligned}$$

with $\alpha_j = \mathrm{E}(x_t, \varepsilon_{t-j})$. Finally, notice that

$$\begin{aligned}
\alpha_0 \left(\frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t}^{(2)} \right) &= \alpha_0 \varepsilon_t \\
\alpha_1 \left(-\frac{1}{\sqrt{2}} \varepsilon_t^{(1)} + \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t}^{(2)} \right) &= \alpha_1 \varepsilon_{t-1} \\
\alpha_2 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t}^{(2)} \right) &= \alpha_2 \varepsilon_{t-2} \\
\alpha_3 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-2}^{(1)} - \frac{1}{2} \varepsilon_t^{(2)} + \frac{1}{2} \pi_{\varepsilon,t}^{(2)} \right) &= \alpha_3 \varepsilon_{t-3} \\
\alpha_4 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4}^{(2)} \right) &= \alpha_4 \varepsilon_{t-4} \\
\alpha_5 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-4}^{(1)} + \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4}^{(2)} \right) &= \alpha_5 \varepsilon_{t-5} \\
\alpha_6 \left(\frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4}^{(2)} \right) &= \alpha_6 \varepsilon_{t-6} \\
\alpha_7 \left(-\frac{1}{\sqrt{2}} \varepsilon_{t-6}^{(1)} - \frac{1}{2} \varepsilon_{t-4}^{(2)} + \frac{1}{2} \pi_{\varepsilon,t-4}^{(2)} \right) &= \alpha_7 \varepsilon_{t-7},
\end{aligned}$$

which yields the classical Wold representation:

$$x_t = \alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots$$

Example 2: the AR(1) case. Here, we express explicitly the coefficients $\psi_k^{(j)}$ for a weakly stationary $AR(1)$ process $\mathbf{x}$, namely

$$x_t = \rho x_{t-1} + \varepsilon_t,$$

where $|\rho| < 1$ and $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a unit variance white noise. By using the lag operator $\mathbf{L}$ and the identity map I, we can rewrite the previous equation as

$$(I - \rho \mathbf{L}) x_t = \varepsilon_t.$$

As is well-known, the process $\mathbf{x}$ admits a moving average representation that we can find by exploiting the invertibility of the operator $I - \rho \mathbf{L}$. In fact,

$$x_t = (I - \rho \mathbf{L})^{-1} \varepsilon_t = \varepsilon_t + \sum_{l=1}^{+\infty} \rho^l \varepsilon_{t-l} = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h},$$

where $\alpha_h = \rho^h$.

Let us, now, fix a scale level $j \in \mathbb{N}$. The expressions of the coefficients $\psi_k^{(j)}$ can be easily obtained by virtue of Eq. (11):

$$\psi_k^{(j)} = \frac{\rho^{k2^j}\left(1-\rho^{2^{j-1}}\right)^2}{\sqrt{2^j}(1-\rho)}$$

for any $k \in \mathbb{N}_0$. The processes $x_t^{(j)}$ are

$$x_t^{(j)} = \frac{\left(1-\rho^{2^{j-1}}\right)^2}{\sqrt{2^j}(1-\rho)} \sum_{k=0}^{+\infty} \rho^{k2^j} \varepsilon_{t-k2^j}^{(j)}.$$

Barring the proportionality term $\frac{\left(1-\rho^{2^{j-1}}\right)^2}{\sqrt{2^j}(1-\rho)}$, the structure of the $\psi_k^{(j)}$ coefficients is that of an AR(1) defined on the support $S_t^{(j)} = \{t - k2^j : \quad k \in \mathbb{Z}\}$. In essence, then, we can rewrite the original AR(1) as an infinite sum of AR(1)s with time steps 2^j and autoregressive coefficients given by ρ^{2j}, i.e.,

$$\begin{aligned} x_t &= \sum_{j=1}^{\infty} \frac{\left(1-\rho^{2^{j-1}}\right)^2}{\sqrt{2^j}(1-\rho)} \times x_t^{(j)}, \\ x_t^{(j)} &= \rho^{2^j} x_{t-2^j}^{(j)} + \varepsilon_t^{(j)}. \end{aligned}$$

It is interesting to note that the magnitude of the autoregressive parameter declines with j. Hence, lower frequency scales entail less persistence in scale time.

6 Filtering the scale-specific shocks

As said, predictability will be a property of individual components and, therefore, will be induced by scale-specific shocks. To this extent, this section provides a discussion of the extraction procedure for the components.

Let $\{x_{t-i}\}_{i\in\mathbb{Z}}$ be the time series of interest. Consider the case $J = 1$. We have

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

which effectively amounts to breaking the series down into a "transitory" and a "persistent" com-

ponent. Set, now, $J = 2$. We obtain

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

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 component (or detail) $\widehat{x}_t^{(j)}$, i.e.,

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

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 $\widehat{x}_t^{(j)}$ with different calendar-time persistence along with a low-frequency trend $\pi_t^{(J)}$. Equivalently, it can be written as a telescopic sum

$$x_t = \sum_{j=1}^{J} \underbrace{\left\{\pi_t^{(j-1)} - \pi_t^{(j)}\right\}}_{\widehat{x}_t^{(j)}} + \pi_t^{(J)} = \pi_t^{(0)}, \tag{12}$$

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 frequency. In particular, the innovations $\widehat{x}_t^{(j)} = \pi_t^{(j-1)} - \pi_t^{(j)}$ become smoother, and more persistent in calendar time, as j increases. In essence, Eq. (12) is naturally viewed as an estimate, based on the proposed nonparametric filter, of Eq. (8).

One may also write a convenient representation of the filter using a suitable orthonormal operator. To illustrate how the operator works, set $J = 2$. In this case,

$$\pi_t^{(2)} = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}. \tag{13}$$

and

$$\widehat{x}_t^{(2)} = \frac{\pi_t^{(1)} - \pi_{t-1}^{(1)}}{2} = \frac{1}{2}\left(\frac{x_t + x_{t-1}}{2} - \frac{x_{t-2} + x_{t-3}}{2}\right)$$
$$\widehat{x}_t^{(1)} = \frac{\pi_t^{(0)} - \pi_{t-1}^{(0)}}{2} = \left(\frac{x_t - x_{t-1}}{2}\right) \tag{14}$$
$$\widehat{x}_{t-2}^{(1)} = \frac{\pi_{t-2}^{(0)} - \pi_{t-3}^{(0)}}{2} = \left(\frac{x_{t-2} - x_{t-3}}{2}\right).$$

Stacking now Eq. (13) on top of Eq. (14) we obtain:

$$\begin{pmatrix} \pi_t^{(2)} \\ \widehat{x}_t^{(2)} \\ \widehat{x}_t^{(1)} \\ \widehat{x}_{t-2}^{(1)} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix}. \tag{15}$$

Denoting by $\mathcal{T}^{(2)}$ the (4×4) matrix in Eq. (15), we notice that $\mathcal{T}^{(2)}$ is orthogonal, that is $\Lambda^{(2)} \equiv \mathcal{T}^{(2)} \left(\mathcal{T}^{(2)}\right)^\top$ is diagonal. Moreover, the diagonal elements of $\Lambda^{(2)}$ are non-vanishing so that $\left(\mathcal{T}^{(2)}\right)^{-1} = \left(\mathcal{T}^{(2)}\right)^\top \left(\Lambda^{(2)}\right)^{-1}$ is well-defined. Importantly, the matrix $\mathcal{T}^{(2)}$ permits to construct the filtered components by virtue of a simple matrix multiplication.

Similarly, by matrix inversion, one could reconstruct the original process given the filtered components. This is, again, easy to see:

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix} = \left(\mathcal{T}^{(2)}\right)^{-1} \begin{pmatrix} \pi_t^{(2)} \\ \widehat{x}_t^{(2)} \\ \widehat{x}_t^{(1)} \\ \widehat{x}_{t-2}^{(1)} \end{pmatrix}. \tag{16}$$

For an extension of this procedure to any $J \geq 2$ and a recursive algorithm for the construction of the matrix $\mathcal{T}^{(J)}$, associated to an arbitrary level of persistence J, we refer the reader to Mallat (1989). The matrix $\mathcal{T}^{(J)}$ will be referred to in what follows as the Haar transform (or the Haar matrix).

As in Beveridge and Nelson (1981), who initiated the filtering of a time series into a stochastic trend and a transitory component, we can view the adopted filter as a nonparametric method to separate the time series into 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 functions of both time and scale.

6.1 Decimation

Decimation is the process of defining *non-redundant* information, as contained in a suitable number of *non-overlapping* "typical" points, in the observed components.

For clarity, let us now return to the case $J = 2$, as above, with the understanding that identical considerations apply more generally. Define the vector

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

By letting time t vary in the set $\left\{t = k2^2 \text{ with } k \in \mathbb{Z}\right\}$ one can now define (from $\mathcal{T}^{(2)}X_t$) the *decimated* counterparts of the calendar-time details, namely $\left\{\widehat{x}_t^{(j)},\ t = k2^j \text{ with } k \in \mathbb{Z}\right\}$ for $j = 1, 2$ and $\left\{\pi_t^{(2)},\ t = k2^2 \text{ with } k \in \mathbb{Z}\right\}$

The decimated points have a similar intuition as the "the small number of data averages" advocated by Müller and Watson (2008) to identify low-frequency information in the raw data. In our framework, however, these points are scale-specific and useful to formalize our notion of frequency-specific predictability.

6.2 Further discussion

In multiresolution analysis,[7] the Haar matrix is routinely used to *filter* $\left\{x_{t-k2^j}^{(j)}\right\}_{j=1,..,J,\ k\in\mathbb{Z}}$ from $\{x_{t-n}\}_{n\in\mathbb{Z}}$. While we also filter the details using the Haar matrix, the methodological novelty of this work is to operate in the opposite direction, i.e., to propose a data generating process which specifies the law of motion of the details $\left\{x_{t-k2^j}^{(j)}\right\}_{j=1,..,J,\ k\in\mathbb{Z}}$ and, only then, obtains each observation x_t as a linear combination of the details themselves. In our specification, the coefficients of the linear combination are uniquely determined by the rows of the inverse Haar matrix.

Some observations are now in order. First, why Haar? There are two reasons, the first has a methodological nature, the second is economically motivated.

[7] See, e.g., Mallat (1989), Dijkerman and Mazumdar (1994), and Yazici and Kashyap (1997). For applications to 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).

1. The structure of the Haar matrix is particularly helpful for us to understand and formalize, below, the connection between scale-wise predictability and predictability upon two-way aggregation without introducing unnecessary complications. Proposition II will, in fact, be derived assuming an Haar mapping between the details $\left\{x^{(j)}_{t-k2^j}\right\}_{j=1,..,J,\ k\in\mathbb{Z}}$ and $\{x_{t-n}\}_{n\in\mathbb{Z}}$. This said, all multiresolution matrices could be viewed as aggregation schemes (e.g., Abry, Veitch, and Flandrin (1998)) and could have been employed instead.

2. The dyadic nature of the Haar matrix is consistent with details having cycles between 2^{j-1} and 2^j periods (years, in our case). The cycles of the details are economically meaningful and help interpretation of the results: 1 to 2 years for the first, 2 to 4 for the second, 4 to 8 for the third, and 8 to 16 for the fourth. In essence, the first detail spans periods shorter than the business cycle, the second and the third detail jointly cover the business cycle (its short and its long end, respectively), and the fourth detail spans cycles longer than the business cycle.

Second, in light of our emphasis on frequencies, why not using more classical methods in the frequency domain, like band spectrum regression (Hannan 1963a, 1963b)? Again, there are, at least, two reasons. Once more, the first has a methodological flavor, the second has to do with the economics of the problem.

1. This is a paper about modeling. We are not excluding the possibility that a non-anticipative (i.e., adapted to time t information) filter in the frequency domain would capture some, or all, of the effects which two-way aggregation (in the time domain) would deliver (like a tent-shaped behavior in R^2, as evidenced by Fig. 2). What we are after is a more fundamental justification for the reported dynamics across frequencies. To use standard econometric jargon, we are interested in a data generating process conforming with economic logic and capable of yielding the reported aggregation result as a very natural by-product of its structure. Our emphasis on scale-specific shocks (as driving shocks in the context of a component-wise model) and component-wise predictability is designed to achieve this goal.

2. The economic literature on predictability is broad and successful. Its reliance on linear predictive systems, like that in Eq. (1) and Eq. (2), is ubiquitous. There is, therefore, much value in casting our problem in the same framework (see Eq. (3)) and Eq. (4)) as the existing literature. Our adopted component-wise framework allows us to do so rather naturally. We do it, in fact,

with one critical conceptual difference: the introduction of the superscript j symbolizing that our exploration into predictability views it as a scale-specific phenomenon with interesting implications for the raw series, rather than as a feature of the raw series themselves, something which has been the exclusive focus of the successful work on the subject.

We now turn to scale-specific predictability.

7 Scale-wise predictive systems

Consider a regressand y and a predictor x. Assume y and x are covariance-stationary and, therefore, admit an extended Wold representation (see Section 5). Assume, also, that for some scale $1 \leq j^* \leq J$, the scale-specific components of the processes y and x are such that

$$y^{(j^*)}_{k2^{j^*}+2^{j^*}} = \beta_{j^*} x^{(j^*)}_{k2^{j^*}} + u^{(j^*)}_{k2^{j^*}+2^{j^*}} \tag{17}$$

$$x^{(j^*)}_{k2^{j^*}+2^{j^*}} = \rho_{j^*} x^{(j^*)}_{k2^{j^*}} + \varepsilon^{(j^*)}_{k2^{j^*}+2^{j^*}}, \tag{18}$$

where $\varepsilon^{(j^*)}_{k2^{j^*}+2^{j^*}}$ is a mean zero, variance $\sigma^2_{\varepsilon^{j^*}}$, white noise process in scale time and $u^{(j^*)}_{k2^{j^*}+2^{j^*}}$ is a mean zero, variance $\sigma^2_{u^{j^*}}$, white noise (again, in scale time) forecast error possibly correlated with $\varepsilon^{(j^*)}_{k2^{j^*}+2^{j^*}}$. Also, assume all other components $\{y^{(j)}_{k2^j+2^j}, x^{(j)}_{k2^j+2^j}\}$ with $j \neq j^*$ are mean-zero, finite variance, white noise processes, uncorrelated with each other.

Eqs. (17)-(18) define a predictive system on individual layers of the $\{y, x\}$ process to be contrasted with the traditional system written on the raw series.

It is interesting to notice that the scale-wise predictive systems are not expected to be affected by the inferential issues that one typically associates with predictability problems. High first-order autocorrelation of the predictor, in particular, has been put forward as a leading cause of inaccurate inference in predictability (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 ρ_j to be smaller, the higher the scale. Consistent with this logic, at the low frequencies over which we identify scale-wise forecastability, ρ_j will be estimated to be small.

There is an understanding in the predictability literature that slow-moving predictors should drive slow-moving conditional means. This relation is, however, hidden by short-term noise. The noise leads to the appearance of low predictability in the short-run, large long-run predictability be-

ing the outcome of noise reduction through return aggregation. There is also an understanding that predictors may be imperfect (Pastor and Stambaugh, 2009). In the context of a different conceptual framework, Eqs. (17) and (18) account for both effects: the link between slow-moving components (for a large j) and, by being defined on components rather than on noisier (i.e., "imperfect") raw series, the "imperfection" of predictors (and regressands).

As an aside, we began this section by stating that y and x are covariance-stationary. Because y is an excess market return, this property is clearly desirable. However, we defined the data generating process of y as an implication of the assumed scale-wise predictive system. The following proposition states that y is, indeed, covariance-stationary provided mild assumptions are satisfied.

Proposition I. *Assume there is a basis of shocks ε_t so that*

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

and the scale-specific shocks $\varepsilon_{t-k2^j}^{(j)}$ are uncorrelated on $S_t^{(j)} = \{t - k2^j : \quad k \in \mathbb{Z}\}$. Then, the process y_t, which can be written as

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

is covariance-stationary provided

$$\sum_{j=1}^{+\infty} \sum_{h=0}^{+\infty} \left(\psi_{\left\lfloor \frac{h}{2^j} \right\rfloor}^{(j)} \delta_{h-2^j \left\lfloor \frac{h}{2^j} \right\rfloor}^{(j)} \right)^2 < +\infty.$$

Proof. *See Appendix A.*

Before turning to empirical evaluations of Eqs. (17)-(18), we discuss the implications of the proposed data generating process for two-way aggregation.

8 The mapping between two-way aggregation and scale-wise predictive systems

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 re-

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

$$\begin{aligned} y^{(j^*)}_{k2^{j^*}+2^{j^*}} &= \beta_{j^*} x^{(j^*)}_{k2^{j^*}}, \\ x^{(j^*)}_{k2^{j^*}+2^{j^*}} &= \rho_{j^*} x^{(j^*)}_{k2^{j^*}} + \varepsilon^{(j)}_{k2^{j^*}+2^{j^*}}, \end{aligned}$$

whereas $\left\{y^{(j)}_{k2^j}, x^{(j)}_{k2^j}\right\} = 0$ for $j \neq j^*$. We map scale-time observations into calendar-time observations by using the inverse Haar transform. Then, the forward-backward regressions

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

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

Proof. *See Appendix A.*

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

Proposition II 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 $\left\{y^{(j)}_{k2^j}, x^{(j)}_{k2^j}\right\}$, for $j < j^*$ or for $j > j^*$, or forecasting errors different from zero, would solely lead to a blurring of the resulting relation upon two-way aggregation. We add uncorrelated details $\left\{y^{(j)}_{k2^j}, x^{(j)}_{k2^j}\right\}$ for $j \neq j^*$ in the simulations in Section 11.

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^2s should display a *tent-like* behavior. At the same horizon, the theoretical slope ($\beta_{h=2^{j^*}}$) of the forward/backward regressions would coincide with the theoretical slope (β_{j^*}) of the detail-wise regressions. These are implications of our approach which will be supported by the data.

Importantly, the proposition has implications for *all* frequencies, not only for the very long run. The conclusions we draw will, therefore, not depend solely on regions of the data space - at very low frequencies - with inevitably high statistical uncertainty. We will show that all frequencies provide support for the suggested approach, thereby reducing classical concerns in the literature about statistical significance over long horizons.

Next, we broaden the scope of classical predictability relations in the literature. We turn to regressions on the extracted details and illustrate the consistency of their findings with those obtained from two-way aggregation. This consistency, which is an implication of Proposition II, is further confirmed by simulation.

From an applied standpoint, one could proceed in the opposite way: detect predictability on the scales and then exploit predictability on the scales by suitably aggregating regressand and regressors, by way of forward/backward aggregation. The latter method is a way in which one could exploit the presence of a scale-specific predictability to perform return predictability and, among other applications, asset allocation over suitable horizons.

9 Predictability on the components

We report predictive regressions on the components. We preserve the same structure as for Section 4 and begin with the component-wise relation between equity returns and market variance.

9.1 Equity returns on *market* variance

The filtered details are shown in Figs. 3 and 4. For an explicit interpretation of the j-th scale in terms of yearly time spans, we refer to Table 4.

[Insert Table 4 about here]

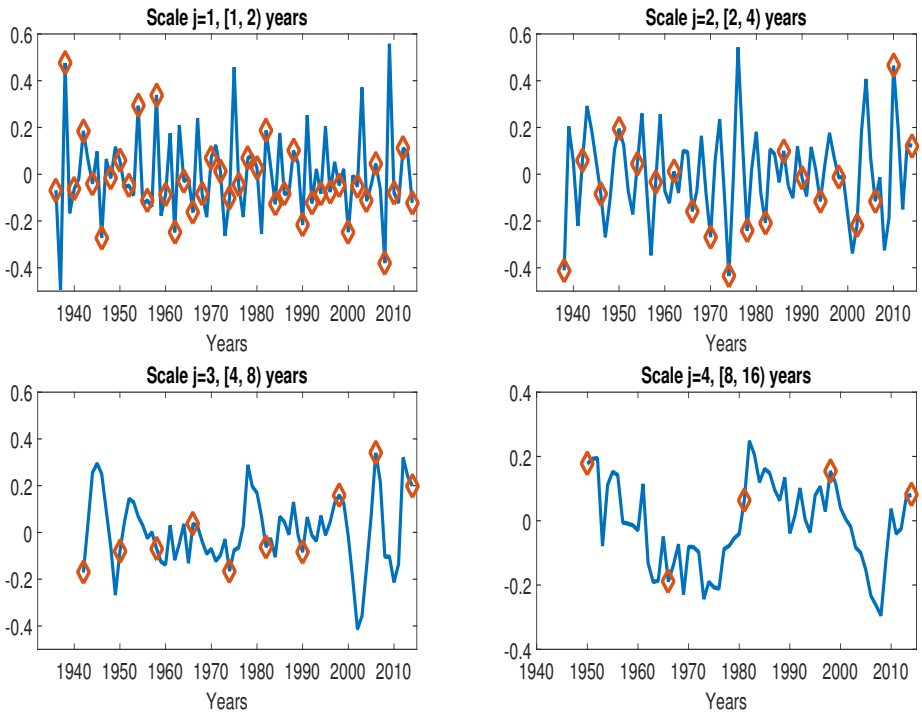


Figure 3: Detail decomposition for market returns: calendar-time observations in solid blue, scale-time observations in red diamonds.

The hypothesis of uncorrelatedness among detail components with different degrees of persistence is not in contradiction with data. Table 5 presents pair-wise correlations between the individual details for both series. Virtually all correlations are small and statistically insignificant. Not surprisingly, the largest one (0.33) corresponds to the adjacent pair of variance scales $j = 3$ and $j = 4$.[8]

[Insert Table 5 about here]

We run detail-wise predictive regressions as in Eqs. (17)-(18). The results are reported in Table

[8]It is worth emphasizing that these pair-wise correlations are obtained by using overlapping, calendar-time, or *redundant* data on the details rather than the non-overlapping, scale-time or *decimated* counterparts described in Subsection 6.1. 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, and yield components with cycles whose length is easily interpretable.

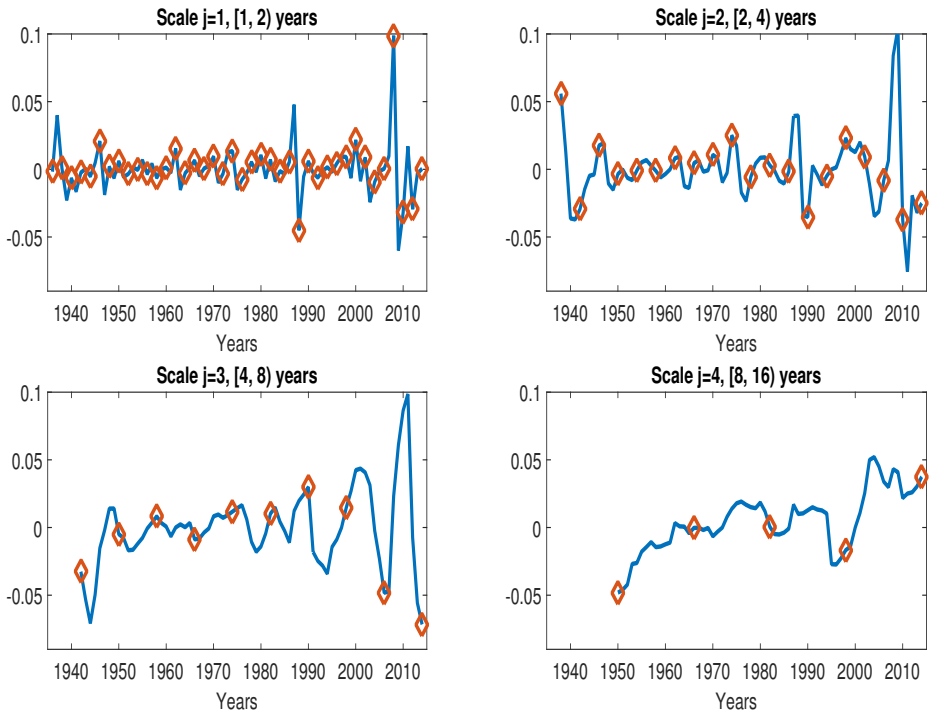


Figure 4: Detail decomposition for market realized variance: calendar-time observations in blue, scale-time observations in red diamonds.

1-Panel B.[9] 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 58.2%.

An important implication of Proposition II 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. Consistent with theory, the R^2 values upon two-way aggregation reach their peak (around 55%) between 14 and 16 years (see Section 4). Remarkably, the structure of the R^2s, before and after, is roughly tent-shaped (see Fig. 2).

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 such inference is conducted carefully. Importantly for the

[9]Given a scale j, we work with an effective sample size of $\left[T/2^j\right]$ observations, where [.] denotes the smallest integer near $T/2^j$. In this empirical analysis, we consider a sample spanning the period 1930 to 2014 and $J = 4$.

purposes of this paper, however, here we do not solely focus on low-frequency dynamics. A crucial implication of our conceptual framework is, in fact, the existence of a tent-shaped behavior in R^2 values as a by-product of scale-wise predictability. The reported tent-shaped behavior 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. In sum, we are not just drawing conclusions from frequencies associated with high statistical uncertainty, we are relying on *all* frequencies. We now turn to consumption variance.

9.2 Equity returns on *consumption* variance

We note that the "de-correlation" property of the details strongly applies to consumption variance (see Table 5). In agreement with theory, running detail-wise predictive regressions leads again to maximum predictability (and an R^2 of 61.62%) associated with low-frequency cycles between 8 and 16 years, i.e., $j = 4$ (Table 2-Panel B).

9.3 Equity returns on (squared) EPU

Results are provided in Table 3. They are analogous to the corresponding findings for market and consumption variance. We will not comment on them further.

9.4 An interpretation

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. Importantly, this is not the case when simply aggregating the raw data over the long run.[10]

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. In other words, consistent with an autoregressive (of order 1) specification of the details, the $j = 4$ variance detail has a positive autocorrelation with uncorrelated residuals. The R^2 of the detail-wise autoregression on market variance is a rather substantial 43.28% with a positive slope of 0.12. As explained earlier, it is unsurprising to find a low *scale-wise* (for $j = 4$) autocorrelation. While the

[10]The corresponding results are available upon request.

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 B.1, the corresponding autocorrelation in calendar-time would naturally be higher. 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 0.24 and an R^2 value of about 58.8%. Again, analogous findings apply to the (squared) EPU's details which, also, behave as an AR(1), with an estimated autocorrelation equal to 0.53, for $j = 4$.

Importantly for our purposes, the documented low dependence between past and future uncertainty 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.

In essence, we find that, at scale $j = 4$, a very slow-moving component of the uncertainty process predicts itself as well as the corresponding component in future excess market returns. Said differently, higher past values of the uncertainty detail appear to predict higher future values of the same uncertainty detail and, consequently, higher future values of the corresponding detail in excess market returns, as required by conventional logic behind compensations for uncertainty (i.e., variance) risk. While this logic applies to a specific level of resolution in our framework, it translates - upon forward/backward aggregation - into predictability for long-run returns as shown formally in Section 8 and in the data.

10 Hump-shaped patterns and classical predictive systems: simulations

How likely are classical predictive systems to generate the tent-shaped dynamics detected in the data? We run forward/backward regressions using simulations from a traditional predictive model for excess market returns and consumption variance, namely:

$$\begin{aligned} r_{t+1} &= \beta v_t + u_{t+1}, \\ v_{t+1} &= \rho v_t + \epsilon_{t+1}. \end{aligned}$$

The parameters are estimated from the data employed previously: $\beta = 1.80$, $\rho = 0.734$, $\sigma_u = 0.180$, $\sigma_\epsilon = 0.0095$, and $\rho_{u,\epsilon}$ = -0.045.

The mean and the median of the slope estimates decline monotonically over time (Table 6-Panel A). This finding is consistent with the observations that, in theory, forward/backward aggregation should lead to decreasing slopes (when $|\rho| < 1$) in the presence of classical predictive systems (Section 2). It is, however, inconsistent with data. In the data, we find monotonically increasing slopes between 6 and 12 years and monotonically decreasing slopes between 16 and 20 years (see, e.g., Table 2). Instead, the percentages of simulations delivering β estimates which are monotonically increasing between 6 and 12 years, monotonically decreasing between 16 and 20 years and hump-shaped (monotonically increasing between 6 and 12 *and* monotonically decreasing between 16 and 20 years) are 14.64%, 35.21%, and only 6.14%, respectively.

The coefficients of determination are monotonically increasing over time, in terms of both their mean and median across simulations. If we employ the same metrics used for the slope estimates (increasingly monotonic between 6 and 12 years, decreasingly monotonic between 16 and 20 years, and hump-shaped), we find percentages equal to 16.81%, 18.29% and 6.54%, respectively. Yet, similarly to what is found for the slope estimates, these shapes are strongly in the data.

Sheer magnitude of the R^2 is an additional, important metric. The reported maximum R^2 value from data is sizable and around 55%. The percentage of simulations yielding hump-shaped R^2 values as well as R^2 magnitudes larger than 50% is only 3.05%. The percentage of simulations delivering *both* hump-shaped R^2 values and hump-shaped slope values, as well as R^2s in excess of 50%, is instead 1.15%.

In essence, it is hard to argue that the empirical findings yielded by forward/backward aggregation can be explained by a traditional predictive system.

[Insert Table 6 about here]

In the spirit of Boudoukh, Richardson, and Whitelaw (2008), we now set $\beta = 0$ and ask whether *absence* of predictability can lead to the effects in the data. When setting $\beta = 0$, the estimated values from data are $\rho = 0.688$, $\sigma_u = 0.1872$, $\sigma_\epsilon = 0.008$, and $\rho_{u,\epsilon}$ = -0.142.

Should the data generating process not allow for predictability, the estimated slopes would be slightly increasing across the board. This is akin to a typical spurious regression problem: the generation of unit root behavior by virtue of aggregation leads to the appearance of dependence. Such

appearance is also reflected in increasing mean and median R^2s across simulations. This increasing behavior is pervasive across frequencies and, importantly, inconsistent with the marked tent-shaped structures reported in the data. We find, for instance, that the percentages of hump-shaped slopes and R^2s (again, monotonically increasing between 6 and 12 years *and* monotonically decreasing between 16 and 20 years) are 7.83% and 7.56%, respectively. When including the requirement of an R^2 larger than 50%, the latter percentage drops to 3.41%. Should we also add the requirement that the drop in the R^2 value between the 16-year horizon and the 20-year horizon is larger than 30%, something which is consistent with data, the percentage would decrease further to 1.40%.

[Insert Table 7 about here]

Additional metrics are reported in Table 7. All of them provide a consistent picture. In agreement with the arguments in Boudoukh, Richardson, and Whitelaw (2008), lack of predictability in the usual sense (β=0 in the system in Eqs. (1)-(2)) could generate mild upward trending behavior in the slopes and coefficients of determination. It will however find it difficult to replicate both the reported humps and their magnitudes at the peaks. We now turn to simulations under a component model.

11 Hump-shaped patterns and *scale-wise predictive systems*: simulations

In this section we confirm, once more by simulation, that scale-specific predictability translates into predictability upon two-way aggregation. We do so by carefully calibrating the simulation parameters to the data. Supporting the implications of Proposition II, we show that hump-shaped patterns are readily generated. We also show that, if predictability on the components applies, contemporaneous (i.e., forward/forward) aggregation leads to insignificant outcomes. Similarly, if no predictability on the components applies, forward/backward aggregation leads to insignificant outcomes.

We begin by postulating processes for the (possibly related) details of the consumption variance

and return series:

$$\begin{aligned} r^{(j)}_{k2^j+2^j} &= \beta_j v^{(j)}_{k2^j} \\ v^{(j)}_{k2^j+2^j} &= \rho_j v^{(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j} \end{aligned} \tag{19}$$

for $j = j^*$ and

$$\begin{aligned} r^{(j)}_{k2^j+2^j} &= u^{(j)}_{k2^j+2^j} \\ v^{(j)}_{k2^j+2^j} &= \varepsilon^{(j)}_{k2^j+2^j} \end{aligned}$$

for $j \neq j^*$, where k is defined as above and $j = 1, \ldots, J = 4$. As in the data, the scales are defined at the annual level. The shocks $\varepsilon^{(j)}_t$ and $u^{(j)}_t$ satisfy $corr(u^{(j)}_t, \varepsilon^{(j)}_t) = 0 \ \forall t, j$.

The model implies a predictive system on the scale j^* and unrelated details for all other scales. In other words, predictability only occurs at the level of the j^*-th detail. Consistent with data, we set $j^* = 4$, i.e., only the fourth component of the return and variance process correlate with each other, their relation being based on the previously-reported betas in Table 2 (we set b_j equal to 2.8 for $j = 4$ and zero otherwise). Moreover $u^{(j)}_{k2^j}$ is $N(0, \sigma^{(j)}_u)$, where $\sigma^{(j)}_u$ is chosen so as to match the variance of the component $r^{(j)}_{k2^j}$ at scale j.[11]

The fourth component of the market variance follows an autoregressive process of order one in scale time, with a scale-specific autoregressive parameter ρ_j calibrated to the data (we set ρ_j equal to 0.2 for $j = 4$ and zero otherwise).[12] All other variance components are assumed to be white noise shocks $\varepsilon^{(j)}_{k2^j} \sim N(0, \sigma^{(j)}_\varepsilon)$ with a variance chosen so as to match the variance of the component $v^{(j)}_{k2^j}$ at scale j.

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

[11] Ortu, Tamoni, and Tebaldi (2013) show that the variance of a stationary time series equals the sum of the variances of its decimated components. For the return series, the variance of the components is $Var(r^{(1)}_{k2^1}) = 0.02$, $Var(r^{(2)}_{k2^2}) = 0.012$, $Var(r^{(3)}_{k2^3}) = 0.005$ and $Var(r^{(4)}_{k2^4}) = 0.002$. Indeed, $\sqrt{\sum_j Var(r^{(j)}_{k2^j})} = 19.75\%$, which equals the stock market volatility over our sample.

[12] The small autoregressive parameter in scale time is enough to generate a persistent process in calendar time. The autocorrelation of our simulated variance at lags 1, 2 and 3 is ACF(1)=0.71, ACF(2)=0.50, and ACF(3)=0.33; in the data, the first 3 lags of the sample ACF for consumption variance are 0.73, 0.48, and 0.33 respectively.

The data generating process is again formulated for observations defined in scale time. We therefore simulate the process at scale j every 2^j steps. To obtain the (calendar-time observations of) aggregate return and variance series from scale-time details, we aggregate the simulated details via the inverse Haar matrix (defined in Section 6). Appendix A, Subsection A.2 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 8, we now show that a predictive relation localized around the j^*-th scale will produce a pattern of R^2s which has a peak for aggregation levels corresponding to the horizon 2^{j^*}.

11.1 Running two-way regressions on data from a component model

Table 8-Panel A shows the results obtained by running the regression in Eq. (5) on simulated data generated from Eq. (19). We compare these results to those in Table 9, where no scale-wise predictability is assumed.[13] The tables report standard errors rather than t-statistics since, in the case of overlapping observations, the rate at which the standard errors grow may be informative.

When imposing the relation at scale $j^* = 4$, i.e., for a time span of 8 to 16 years (c.f., Table 1), we reach a peak in the R^2s of the two-way regressions exactly at 16 years. The 16-year R^2 is 53.65% and is very comparable to its empirical counterpart from Table 2. The 16-year R^2 is also about 25 times as large as the one obtained in the case of no-predictability at the same horizon (see Table 9). We notice that the slope estimates increase, reaching their maximum value of 4.48 at 16 years, a value identical to the the slope's estimated value on data of 4.48 (Table 2). After the 16-year mark, the slope estimates decrease almost monotonically. The coefficients are strongly significant between 12 and 18 years. They are insignificant before and after those horizons. Hence, the simulations generate an hump-shaped pattern in the estimated slopes, t-statistics, and R^2 which derives solely from imposing scale-specific predictability at a frequency lower than business-cycle frequencies.

[Insert Tables 8 and 9 about here]

[13] Under the null of no scale-wise predictability we set b_4 equal to zero. In other words under the null we have $r^{(j)}_{k2^j+2^j} = u^{(j)}_{k2^j+2^j}$ for all j, whereas under the alternative we had $r^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j}$ for $j = j^*$. Once again the variance of the shock is chosen so as to match the variance of the component $r^{(j)}_{k2^j}$ at scale j. All the other calibrated parameters are untouched so that the first 2 lags of the ACF for the simulated consumption variance are the same as before.

Returning to the distribution of the slope estimates and the R^2s, the percentages of βs (R^2s) which are monotonically increasing between 6 and 12 years and monotonically decreasing between 16 and 20 years are 52.2% (39.2%) and 66.8% (63.4%), respectively. The percentage of βs (R^2s) which satisfy both condition is 37.2% (31.2%). Hence, more than 1 in 3 simulated paths deliver what we observe in the data. Since the corresponding numbers in the case of a classical predictive system are around 6.5%, this metric suggests that a component model appears to have a probability 5 times higher than a traditional predictive model to feature the effects observed in the data.

Should we strengthen the metric a bit and consider now the percentage of R^2 values which are hump-shaped, larger than 50% and such that the drop in R^2 between the 16-year horizon and the 20-year horizon is larger than 30%, we would find that about 19% of the sample paths would reproduce exactly the pattern in the data. This observation translates into a likelihood of observing paths with those characteristics which is more than 13 times larger than the corresponding likelihood in the case of a traditional predictive model. Additional metrics, providing analogous information, are reported in Table 8.

As emphasized earlier, if aggregation were to lead, somewhat mechanically, to statistically significant, larger slopes and higher R^2 values by virtue of the creation of stochastic trends, hump-shaped behaviors would be unlikely and contemporaneous (forward/forward) aggregation would also lead to spurious predictability. We have shown, instead, that hump-shaped structures may naturally arise from predictability at the corresponding scale. We now turn to forward/forward aggregation. Again, we simulate under $j^* = 4$ (in Table 8-Panel B). When both the regressor and the regressand are aggregated over the same time interval, no statistical significant predictability is detected. Appendix A-B.2 provides a theoretical justification.

These effects are similar to what one would obtain if, instead of aggregating forward/forward, one where to aggregate forward/backward while simulating from a component model like Eqs. (17)-(18) above under the assumption of *absence* of scale-specific predictability, i.e., $\beta_j = 0$. Absence of predictability would lead to statistically insignificant slopes numerically very close to zero for all horizons. The resulting R^2 values would also be very small and rather far from magnitudes seen in the data.

In sum, we view these simulations as giving an important role to predictability on the components and confirming the ability of suitable (forward/backward) aggregation to detect it.

12 Past uncertainty within Campbell and Shiller's framework

In this section, we use the well-known Campbell and Shiller's present value identity (Campbell and Shiller (1988)) in order to provide a conceptual framework to understand the long-run predictive ability of past economic uncertainty. Within the Campbell and Shiller's framework, variables that are orthogonal to the dividend-to-price ratio should predict long-run returns only if they have equal predictive ability for future long-run cash flows. We show that the orthogonal (to the dividend-to-price ratio) component of past economic uncertainty satisfies this property.

We recall that the classical Campbell and Shiller's log-linearization implies the following

$$d_t - p_t = r_{t+1} - \Delta d_{t+1} - k + \rho \left(d_{t+1} - p_{t+1} \right), \tag{20}$$

where d_t is log-dividend, p_t is log-price, $k = \log\left(1 + \exp\left(\mathrm{E}\left[p - d\right]\right)\right) - \rho \mathrm{E}\left[p - d\right]$ and $\rho = \frac{e^{\mathrm{E}[p-d]}}{1+e^{\mathrm{E}[p-d]}}$. Iterating Eq. (20) forward, and taking conditional expectations, one obtains

$$d_t - p_t = \text{const.} + \mathrm{E}_t \left[\sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right) \right]. \tag{21}$$

Eq. (21) is derived by ruling out the explosive behavior of stock prices, i.e., $\lim_{j\to\infty} \rho^j \left(d_{t+j} - p_{t+j} \right) = 0$. Eq. (20), and its forward iteration, are identities. They hold ex-post as well as in expectation.[14] In fact,

$$\begin{aligned} d_t - p_t &= \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right) \\ &= \mathrm{E} \left[\sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right) \mid d_t - p_t \right]. \end{aligned}$$

In words, the quantity $d_t - p_t$ is, somewhat mechanically, informative about investor's expectations regarding either long-run dividend growth or long-run returns, or *both*. This observation justifies

[14] We ignore the constant, and interpret all variables from now on as being de-meaned. Recent work by Lettau and Nieuwerburgh (2008) and Favero, Gozluklu, and Tamoni (2011) reports evidence for structural shifts in the long-run mean of the dividend-price ratio and advocates the use of a de-trended dividend-price series as the return predictor. Although in what follows we assume that the long-term mean is constant, our results would be even stronger had we used a de-trended dividend price ratio. This is so because our orthogonalization procedure attributes part of the movement in past uncertainty to low-frequency changes in the dividend-price ratio, these changes being potentially related to a time-varying long-run mean. Absent these changes, the additional contribution of past uncertainty to return and dividend predictability based solely on dividend-price ratio would increase.

the attention the price-to-dividend ratio has received (see, e.g., Cochrane (2008)).

When taking the identity to data, the infinite sums ought to be truncated. We use the notation

$$r_t^k \equiv \sum_{j=1}^{k} \rho^{j-1} r_{t+j}, \quad \Delta d_t^k \equiv \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}, \quad (d/p_t)^k \equiv \rho^k \left(d_{t+k} - p_{t+k}\right), \tag{22}$$

where r_t^k, Δd_t^k, and $(d/p_t)^k$ define k-period (discounted) log returns, k-period (discounted) log dividend growth and the k-step ahead (discounted) log dividend-to-price ratio. Write $r_t^k = r_t^\infty$, $\Delta d_t^k = \Delta d_t^\infty$, and $(d/p_t)^k \to 0$ (ruling out bubbles) when $k \to \infty$. Interpret r_t^∞ and Δd_t^∞ as notions of long-run (weighted, by ρ) returns and long-run (weighted) dividend growth.

When truncating, i.e., for a k small enough, the bubble term may be empirically (and conceptually) important. Iterating Eq. (20) forward, write:

$$\begin{aligned} d_t - p_t =& \sum_{j=1}^{k} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j}\right) + \rho^k \left(d_{t+k} - p_{t+k}\right), & (23) \\ =& r_t^k - \Delta d_t^k + (d/p_t)^k. & (24) \end{aligned}$$

This expression readily implies that

$$\begin{aligned} \operatorname{cov}(d_t - p_t, d_t - p_t) =& \operatorname{cov}(d_t - p_t, r_t^k - \Delta d_t^k + d/p_t^k) \\ =& \operatorname{cov}(d_t - p_t, r_t^k) - \operatorname{cov}(d_t - p_t, \Delta d_t^k) + \operatorname{cov}(d_t - p_t, d/p_t^k). \end{aligned}$$

In terms of regression coefficients, the restriction on the covariances becomes

$$1 = \beta_{r,dp}^k - \beta_{\Delta d,dp}^k + \beta_{d/p,dp}^k, \tag{25}$$

where the three slopes derive from the following univariate regressions:

$$\begin{aligned} r_t^k &= \beta_{r,dp}^k (d_t - p_t) + \varepsilon_{r,t+k}, \\ \Delta d_t^k &= \beta_{\Delta d,dp}^k (d_t - p_t) + \varepsilon_{d,t+k}, \\ d/p_t^k &= \beta_{d/p,dp}^k (d_t - p_t) + \varepsilon_{d/p,t+k}. \end{aligned}$$

In the long run (i.e., for $k \to \infty$), the third equation can be ignored. As in Cochrane (2008), the

restriction becomes

$$1 = \beta^{\infty}_{r,dp} - \beta^{\infty}_{\Delta d,dp}. \tag{26}$$

The dividend-to-price ratio should, therefore, predict either long-run returns or long-run dividend growth, or both. Because $\beta^{\infty}_{r,dp}$ and $\beta^{k}_{\Delta d,dp}$ are found to be economically close to 1 and 0, Cochrane (2008) emphasizes that it predicts the former, rather than the latter.

12.1 Other (than dividend-to-price) forecasting variables

Eq. (23) and its long-run implications continue to hold. Hence, if a variable x_t, *orthogonal* to the dividend-to-price ratio, were to forecast r^{∞}_t in addition to $d_t - p_t$, the same variable would also have to forecast dividend growth Δd^{∞}_t. These forecasts would "offset" each other so that, given $d_t - p_t$, the forecast of the entire right hand side of the present value identity is not altered. Write

$$\begin{aligned}\text{cov}(x_t, d_t - p_t) =& \text{cov}(x_t, r^k_t - \Delta d^k_t + (d/p_t)^k) \\ =& \text{cov}(x_t, r^k_t) - \text{cov}(x_t, \Delta d^k_t) + \text{cov}\left(x_t, (d/p_t)^k\right),\end{aligned}$$

where $\text{cov}(x_t, d_t - p_t) = 0$, due to the assumed *orthogonality* of x_t and $d_t - p_t$. In terms of regression coefficients, the restriction on the covariances becomes:

$$0 = \beta^k_{r,x} - \beta^k_{\Delta d,x} + \beta^k_{d/p,x}, \tag{27}$$

where the three slopes derive from the following univariate regressions:

$$\begin{aligned} r^k_t &= \beta^k_{r,x} x_t + \epsilon_{r,t+k}, \\ \Delta d^k_t &= \beta^k_{\Delta d,x} x_t + \epsilon_{d,t+k}, \\ d/p^k_t &= \beta^k_{d/p,x} x_t + \epsilon_{d/p,t+k}. \end{aligned}$$

In the long run ($k \to \infty$), the third equation can be ignored and the restriction becomes

$$\beta^{\infty}_{r,x} = \beta^{\infty}_{\Delta d,x}. \tag{28}$$

Importantly, this restriction is completely mechanical and is solely a by-product of (i) Campbell

and Shiller's identity and (2) the orthogonality of the predictor. *In other words, for a large enough k any orthogonal variable would satisfy it, irrespective of its predictive ability for long-run returns and dividend growth.*

This said, any orthogonal variable which predicts long-run returns (resp. long-run dividend growth) in a statistically significant manner should also predict long-run dividend growth (resp. long-run returns). Whether, econometrically, one obtains more signal from either a regression of long-run returns on the assumed predictor or from a regression of long-run dividend growth depends on testable restrictions. This is easy to see. Write

$$r_t^{\infty} = \beta_{r,x}^{\infty} x_t + \epsilon_{r,t+\infty}. \tag{29}$$

Eq. (29) and Eq. (24), with $k \to \infty$, imply that

$$\Delta d_t^{\infty} = \beta_{r,x}^{\infty} x_t + \epsilon_{r,t+\infty} - (d_t - p_t).$$

Hence, we have $\mathrm{E}[\Delta d_t^{\infty}|x_t] = \mathrm{E}[r_t^{\infty}|x_t] = \beta_{r,x}^{\infty} x_t$ since $\mathrm{E}[d_t - p_t|x_t] = \mathrm{E}[d_t - p_t] = 0$. The first equality derives from the orthogonality of $d_t - p_t$ and x_t and the second equality derives from the fact that $d_t - p_t$ is de-meaned. *For a large enough k, the true beta can be estimated consistently both from long-run returns and from long-run dividend growth. Both regressions are correctly specified.*

Importantly, however, the relative signal of a regression of long-run dividend growth on x_t or long-run returns on x_t depends on the relation between $\mathrm{var}(\epsilon_{r,t+\infty} - (d_t - p_t))$ and $\mathrm{var}(\epsilon_{r,t+\infty})$.

Immediately, the signal from the former regression is stronger than that from the latter when

$$\rho_{\epsilon_r,pd} \geq \frac{\sigma_{pd}}{2\sigma_{\epsilon_r}},$$

where $\rho_{\epsilon_r,pd}$ is the correlation between shocks to long-run returns and $d_t - p_t$, σ_{ϵ_r} is the standard deviation of shocks to long-run returns, and σ_{pd} is the standard deviation of the price-to-dividend ratio.

Now, notice that $\epsilon_{r,t+\infty}$ is positively correlated with $d_t - p_t$ since the latter has predictive ability for long-run returns and is orthogonal to the new variable x_t. In fact, by orthogonality, the correlation $\rho_{\epsilon_r,pd}$ is proportional to $\beta_{r,dp}^{\infty}$ from the univariate regression of long-run returns (or, equivalently, the errors $\epsilon_{r,t+\infty}$) onto the dividend-to-price ratio:

$$\epsilon_{r,t+\infty} = \beta^{\infty}_{r,dp}(d_t - p_t) + \epsilon^{*}_{r,t+\infty}.$$

Hence,

$$\rho_{\epsilon_r,pd} = \beta^{\infty}_{r,dp}\frac{\sigma_{pd}}{\sigma_{\epsilon_r}},$$

which implies

$$\beta^{\infty}_{r,dp} \geq 1/2.$$

In other words, should the slope associated with the classical long-run predictive regression on the dividend-to-price ratio be larger than 1/2, an *indirect* long-run dividend-growth regression provides more signal to detect the predictability of alternative predictors than a *direct* long-run return regression. *This is like saying that, should the dividend-to-price ratio display more predictability, economically, for long-run returns than for long-run dividend growth ($\beta^{\infty}_{r,dp} \geq 1/2$), then long-run dividend growth should be, statistically, more informative about the predictive ability of an orthogonal regressor than long-run returns.* Since the condition $\beta^{\infty}_{r,dp} \geq 1/2$ is easily satisfied in the data, we expect the standard errors of the slope estimates derived from long-run dividend growth regressions on x_t to be smaller. Indirect long-run dividend growth regressions are, therefore, more powerful to detect the predictability of orthogonal regressor(s) than direct long-run return regressions.

More can be said. Should one agree with Cochrane's view that $\beta^{\infty}_{r,dp} = 1$, the signal about additional predictability from a regression of long-run dividend growth onto the orthogonal predictor would be as large as the signal about additional predictability from a full model in which long-run returns are regressed onto *both* the orthogonal regressor and the dividend-to-price ratio. This is again easy to see.

Using the same notation as before, the *full* specification would read:

$$r^{\infty}_t = \beta^{\infty}_{r,x}x_t + \underbrace{\beta^{\infty}_{r,dp}(d_t - p_t) + \epsilon^{*}_{r,t+\infty}}_{\epsilon_{r,t+\infty}}.$$

Given the definition of $\epsilon_{r,t+\infty}$ with $\beta^{\infty}_{r,dp} = 1$, the long-run dividend growth regression would now

be:

$$\begin{aligned}\Delta d_t^{\infty} =& \beta_{r,x}^{\infty} x_t + \epsilon_{r,t+\infty} - (p_t - d_t) \\ =& \beta_{r,x}^{\infty} x_t + \epsilon_{r,t+\infty}^{*}.\end{aligned}$$

By the orthogonality of $d_t - p_t$ and x_t and the fact that the two regressions have the same error terms, running a long-run dividend growth regression on the assumed predictor yields the same standard error (for the quantity of interest, $\hat{\beta}_{r,x}^{\infty}$) as working with the completely-specified, full model. In sum:

- *An orthogonal variable x_t which helps $d_t - p_t$ forecast long-run returns should also forecast long-run dividend growth (and, for a finite horizon, future dividend to price).* This result is rather explicit in the discussion in Cochrane (2008).

- *For a long enough horizon k, the prediction slopes should cancel each other out exactly, i.e., $\beta_{r_t^{\infty},x_t} = \beta_{\Delta d_t^{\infty},x_t}$.* This result is mechanical. Any orthogonal (to the dividend-to-price ratio) variable would lead to it, irrespective of its predictive ability, for k large enough.

- *While uninformative about actual predictability, the "time of closure" of the estimated slopes, k^* say, does provide information about the horizon k^* over which "the long run" begins to show up in the data.*

- *How to then assess the long-run predictive ability of an orthogonal variable at this specific horizon k^*? In light of the theoretical equality of their slopes, both a direct regression of long-run returns on the variable and an indirect regression of long-run dividend growth on the variable could be run. Statistical significance of the corresponding slopes should, of course, be evaluated.*

- *This said, because the restriction $\beta_{r,dp}^{\infty} \geq 1/2$ is satisfied in the data, regressions of long-run dividend growth on the presumed predictor have more signal.*

- *Thus, strong statistical significance as derived from long-run dividend growth regressions may be viewed as a necessary (and sufficient) condition for the long-run predictive ability of an orthogonal (to the dividend-to-price ratio) regressor.*

In what follows, we identify the horizon k^* over which the data contains information about long-run dynamics. At this horizon k^*, we report substantial statistical significance associated to the predictive ability of past uncertainty for *both* future long-run returns and future long-run dividend growth. The former represents a direct test. The latter constitutes an indirect, albeit more powerful, test.

The above discussion hinges on implications drawn from a standard log-linearization. Below, we evaluate those implications as well as predictability for long-run returns and long-run dividend growth without continuous compounding. Modifying the compounding frequency does not affect matters.

13 Past uncertainty, future returns and dividend growth

We showed that components of the uncertainty process with cycles between 8 and 16 years predict low-frequency return components, 16 years into the future. A theoretical implication of this *scale-specific predictability* is that backward aggregated variance over 16 years has maximum predictive ability for future returns when reaching the 16-year horizon.

Consistent with the results in Section 4, in order to extract a slow-moving signal about future conditional mean returns, we use past uncertainty aggregated over a 16-year horizon. We make use of Campbell and Shiller's identity (and its implications, as discussed above) as the relevant conceptual framework. For reasons of interpretation and technical clarity (again, discussed above), we now focus on the strictly *orthogonal* (to the dividend-to-price ratio) component of past uncertainty. It is, then, within the Campbell and Shiller's framework that we justify the strong long-run cash-flow predictability associated with this orthogonal component as the flip side of the same component's predictive ability for long-run discount rates. As is the case for other quantities, we work with logarithmic transformations of uncertainty measures. Hence, by Proposition II, the regressor (i.e., past uncertainty) should be viewed as a low-pass filter for a slow-moving component in logarithmic uncertainty. We discuss robustness to transformations in what follows.

13.1 (Log) Market variance

Table 11 contains univariate regressions of k-period (discounted, by ρ) log returns, r_t^k, k-period (discounted, by ρ) log dividend growth, Δd_t^k, and k-period ahead (discounted, by ρ) log dividend-

to-price ratio, d/p_t^k, onto the dividend-to-price ratio and past long-run (log) market variance. The notation was defined in Eq. (22). The value of ρ is estimated to be equal to 0.9677.[15,16]

In theory, the long run yields clean implications. We use an empirical definition of "long run" as the horizon k^* over which we witness "closure" or "convergence" of the slope estimates associated with (i) a regression of r_t^k on variance and (ii) a regression of Δd_t^k on variance. When "closure" is not achieved, the difference between the two slopes should (mechanically) be the slope associated with a regression of d/p_t^k on variance.

Table 11 reports two panels. Panel A refers to an horizon of k=16 years, Panel B refers to an horizon of k=18 years. At k=16, there is still a slight mismatch between estimated slopes on cash flow regressions and estimated slopes on discount rate regressions. The mismatch is erased at 18 years. While at this horizon there is still some variation left in discounted k-period ahead dividend-to-price ratios, such variation is statistically and economically insignificant.

Should k=18 be a true "long-run" proxy, the discussion in Section 12 implies that the 18-year dividend-growth regression will provide a stronger signal about predictability, or lack thereof, then the 18-year return regression. The latter delivers an estimated slope of 0.03 with an associated t-statistic of 2.01 and an R^2 value of 13.73%. The former, instead, yields the same estimated slope but with a t-statistic of 4.96 and an R^2 value of 58.21%. As emphasized, it is the predictive ability of the dividend-to-price ratio which enhances the signal-to-noise ratio of dividend growth regressions on predictors. One may, therefore, view the considerably smaller standard errors from cash flow regressions (relative to those from discount rate regressions) as an indirect test of the predictive ability of the dividend-to-price ratio. Leaving *relative* magnitude of the standard errors aside, and focusing on their *absolute* magnitude instead, statistical significance from cash flow regressions is critical for long-run predictability, i.e., more so than statistical significance from discount rate regressions.

We now turn to returns and dividend growth without continuous compounding. We do not

[15] $\rho = \frac{\exp^{\mathrm{E}(p-d)}}{1+\exp^{\mathrm{E}(p-d)}}$, where $\mathrm{E}(p-d)$ is estimated using the full sample 1930-2014. We also run the regressions when ρ is estimated using the effective sample 1945-2014. Results (available upon request) do not change.

[16] CRSP computes annual return series under the stock market reinvestment assumption. This approach is problematic because it imparts some of the properties of returns to cash flows, see Chen (2009) and Koijen and Nieuwerburgh (2011). Thus, we construct the annual return series from monthly data under the assumption of dividend reinvestment at a zero-rate. It is also interesting to note that the reinvestment rate of monthly dividends leads to contamination mainly in the pre-1945 era, see Koijen and Nieuwerburgh (2011). Since we use the first 16 years of data (1930-1945) to construct our measure of past uncertainty, our effective sample spans the post-1945 period and our results are less affected by the choice of the dividend reinvestment strategy.

expect the implications derived earlier from log-linearization to hold but, of course, continue to be interested in predictability for the long run (and at higher frequencies). Both the dividend-to-price ratio (Panel A, Table 10) and the orthogonal component of variance (Panel B, Table 10) have considerable predictive ability for returns, at all reported frequencies. The joint predictions (Panel C, Table 10) are remarkable. At 2 years, dividend-to-price and past long-run variance explain 20% of the variability in excess returns. The number becomes 48% at 5 years, 84% at 10 years, and 78% at 15 years. Dividend growth is different. The dividend-to-price ratio is insignificant across the board (Panel A, Table 12). Past long-run variance, instead, has predictive ability for dividend growth at all horizons (Panel B, Table 12): the R^2 is 10% at 1 year, 35% at 5 years, 51% at 10 years, and 60% at 15 years.

[Insert Tables 10, 11 and 12 about here.]

13.2 (Log) Consumption variance

Past long-run consumption variance operates in a similar fashion (Table 14). As earlier, the horizon k over which we observe "closure" of the estimated slopes on cash flow and discount rate regressions is 18 years. At 18 years, the estimated slope from cash flow regressions on past long-run consumption variance (0.03) is very statistically significant (t-statistics of 5.30). The statistical significance of the estimated slope associated with the discount rate regressions is lower but, as explained, this is a by-product of the lower signal from long-run return regressions.

Turning to simple returns and predictability over different frequencies, past long-run consumption has somewhat lower predictive ability than past long-run market variance at all horizons (Table 13, Panel B). Statistical significance begins at 7 years. As earlier, past long-run consumption variance predicts dividend growth strongly (Table 15, Panel B).

[Insert Tables 13, 14 and 15 about here.]

13.3 (Log) Economic Policy Uncertainty

At k=18, the slopes from cash flow and discount rate regressions onto past EPU have not converged yet (Table 17). Specifically, there is residual predictable variation associated with the 18-period ahead (discounted) log dividend-to-price ratio. The estimated slope is -0.01 versus estimated slopes

on r_t^{18} and Δd_t^{18} equal to 0.02 and 0.01, respectively. The value is, therefore, economically meaningful. The corresponding t-statistic is a substantial -5.44. Closure of the slopes would, in this case, occur over a slightly longer 20-year horizon.

Predictability for returns (Table 16) and dividend growth (Table 18) over different horizons much resembles the case of past market variance. Some numbers: at 5, 10, and 15 years, dividend-to-price and past EPU predict returns with an R^2 of about 50%, 85%, and 82%, respectively.

[Insert Tables 16, 17 and 18 about here.]

13.4 Robustness

In this subsection, we discuss various robustness checks. When not reported in enclosed tables, the numerical results are available upon request.

First, we investigated robustness with respect to the aggregation horizon. Past uncertainty was defined by aggregating single period proxies over a 16-year time period. Tables 1, 2, and 3, in fact, justify this choice. Changing the aggregation horizon to a number of years between 10 and 16, however, does not modify the reported findings in any relevant fashion. In particular, aggregation over a 10-year horizon is sufficient to extract economically-relevant, slow-moving, uncertainty components.

Next, we verified robustness with respect to transformations. Previously, we used logarithmic transformations (of variance). Using variance, or volatility, does not affect the reported results meaningfully.

Finally, we performed robustness checks with respect to alternative measures of macroeconomic uncertainty. Employing, e.g., the proxy for macroeconomic uncertainty suggested in Jurado, Ludvigson, and Ng (2015), would not modify our findings. We report the corresponding results in Table 19. For this series, the sample is shorter and spans 1960-2014. Due to data limitations, macroeconomic uncertainty is now aggregated over 8 years. Analogous results, however, are obtained for levels of aggregations between 8 and 12 years. The table confirms our previous conclusions from Table 13, 14 and 15.

[Insert Table 19 about here.]

14 Conclusions and further discussion

Economic relations may apply to individual low-frequency components and be hidden by transient effects at higher frequencies.

To capture this idea parsimoniously, this paper models stock market returns and their predictors as aggregates of uncorrelated components operating over different scales and introduce a notion of *scale-specific predictability,* i.e., predictability on the components, layers, or details.

We propose *direct* extraction of the time-series details - and predictive regressions on the details - as well as *indirect* extraction by means of two-way aggregation of the raw series - and predictive regressions on forward/backward aggregates of the raw series. The mapping between the two methods is established *theoretically* and in *simulation* and their close relation is exploited *empirically.* The direct method allows one to identify the data generating process (i.e., the details and, upon reconstruction, the original series) as well as their predictive link. 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. Two-way aggregation, in particular, offers a natural way to exploit scale-specific predictability (in asset allocation over alternative low frequencies, including the very long run, for instance).

Using both direct extraction of the details and two-way aggregation, we offer unusually strong evidence about the existence of risk-return trade-offs on slow-moving components of the return and variance process with cycles between 8 and 16 years. These *scale-wise trade-offs* translate into equally strong dependence between long-run future returns and long-run past variance with a peak of predictability around 16 years.

The reported tent-shaped pattern - an implication of the theory we propose - requires all frequencies to cooperate. In this sense, we jointly exploit the informational content of the data at all frequencies, i.e., in regions of the data space with inevitably high statistical uncertainty as well as in largely-populated regions of the data space.

Our predictive results hold regardless of the notion of variance used, whether it is consumption variance, market variance, or a proxy for economic policy uncertainty (EPU). This robustness speaks to the link between macro uncertainty, as represented by consumption variance and EPU, and uncertainty in financial markets, as represented by market variance, when focusing on their components with 8-to-16 year cycles. While the existing literature has found it hard to relate,

empirically, consumption variance and market variance, we find a remarkable 80%/90% correlation between their components with fluctuations lower than the business cycle.

Successful return predictors (other than the dividend-to-price ratio) generally modify the term structures of short and medium-term return predictability. However, they do not lead to significant long-run return forecasts (Cochrane (2011)). These predictors do not forecast dividend growth either. In contrast, this paper shows that, in agreement with its long-run predictive ability, past long-run uncertainty yields substantial cash-flow predictability as well.

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., and T. Bollerslev (1997): "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52(3), 975–1005.

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.

Baker, S. R., N. Bloom, and S. J. Davis (2013): "Measuring economic policy uncertainty," *Chicago Booth research paper*, 02(13-).

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., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and return: Long-run relations, fractional cointegration, and return predictability," *Journal of Financial Economics*, 108(2), 409–424.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22(11), 4463–4492.

Bonomo, M., R. Garcia, N. Meddahi, and R. Tedongap (2015): "The long and the short of the risk-return trade-off," *Journal of Econometrics*, 187(2), 580–592.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (2008): "The Myth of Long-Horizon Predictability," *Review of Financial Studies*, 21(4), 1577–1605.

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

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

Campbell, J. Y., and M. Yogo (2006): "Efficient tests of stock return predictability," *Journal of Financial Economics*, 81(1), 27–60.

Chen, L. (2009): "On the reversal of return and dividend growth predictability: A tale of two periods," *Journal of Financial Economics*, 92(1), 128–151.

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

Cochrane, J. H. (2011): "Presidential address: Discount rates," *The Journal of Finance*, 66(4), 1047–1108.

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.

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

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

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

Favero, C. A., A. E. Gozluklu, and A. Tamoni (2011): "Demographic Trends, the Dividend-Price Ratio, and the Predictability of Long-Run Stock Market Returns," *JFQA*, 46(4), 1493–1520.

Gârleanu, N., S. Panageas, and J. Yu (2012): "Technological Growth and Asset Pricing," *Journal of Finance*, 67(4), 1265–1292.

Geanakoplos, 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.

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

Hannan, E. J. (1963a): *Regression for Time Series*. in Time Series Analysis, ed. by M. Rosenblatt. New York: Wiley.

——— (1963b): "Regression for Time Series with Errors of Measurement," *Biometrika*, 50, pp. 293–302.

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

Hobijn, B., and B. Jovanovic (2001): "The Information-Technology Revolution and the Stock Market: Evidence," *American Economic Review*, 91(5), 1203–1220.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *American Economic Review*, 105(3), 1177–1216.

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

Koijen, R. S., and S. V. Nieuwerburgh (2011): "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3(1), 467–491.

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.

Lettau, M., and S. C. Ludvigson (2005): "Expected returns and expected dividend growth," *Journal of Financial Economics*, 76(3), 583–626.

Lettau, M., and S. V. Nieuwerburgh (2008): "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21(4), 1607–1652.

Lewellen, J. (2004): "Predicting returns with financial ratios," *Journal of Financial Economics*, 74(2), 209–235.

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.

Muller, U. A., M. M. Dacorogna, R. D. Dave, R. B. Olsen, O. V. Pictet, and J. E. von Weizsacker (1997): "Volatilities of different time resolutions – Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4(2), 213–239.

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," *SSRN eLibrary*.

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 R. F. Stambaugh (2009): "Predictive Systems: Living with Imperfect Predictors," *Journal of Finance*, 64(4), 1583–1628.

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.

Rangvid, J., M. Schmeling, and A. Schrimpf (2014): "Dividend Predictability Around the World," *Journal of Financial and Quantitative Analysis*, 49, 1255–1277.

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.

Sizova, N. (2013): "Long-Horizon Return Regressions With Historical Volatility and Other Long-Memory Variables," *Journal of Business and Economic Statistics*, 31(4), 546–559.

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 Regressioms: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 201–232.

van Binsbergen, J. H., and R. S. J. Koijen (2010): "Predictive Regressions: A Present-Value Approach," *Journal of Finance*, 65(4), 1439–1471.

Welch, I., and A. Goyal (2008): "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21(4), 1455–1508.

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

Yazici, B., and R. Kashyap (1997): "A class of second order self-similar processes for 1/f phenomena," *IEEE Transactions on Signal Processing*, 45(2), 396–410.

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
$v_{t-h+1,t}$	0.42	0.55	0.12	0.32	0.89	1.10	1.35	1.86	2.80	3.36
	(0.86)	(0.83)	(0.19)	(0.44)	(1.28)	(1.36)	(1.60)	(2.24)	(3.68)	(5.02)
	{0.07}	{0.11}	{0.03}	{0.07}	{0.20}	{0.25}	{0.03}	{0.41}	{0.62*}	{0.74*}
$R^2(\%)$	[0.52]	[1.18]	[0.08]	[0.54]	[3.82]	[5.82]	[8.70]	[14.75]	[28.17]	[36.12]
$[5^{th},\ 95^{th}]$	[0.00, 3.74]	[0.01, 7.10]	[0.01, 8.17]	[0.01, 9.64]	[0.03, 20.92]	[0.11, 32.60]	[0.15, 38.30]	[2.27, 50.55]	[11.24, 61.52]	[20.81, 65.43]

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
$v_{t-h+1,t}$	3.79	3.94	3.93	3.79	3.71	3.70	3.11	2.41	1.38	0.77
	(6.29)	(6.34)	(6.34)	(6.56)	(7.44)	(8.61)	(7.57)	(5.73)	(3.69)	(2.08)
	{0.84**}	{0.91**}	{0.95**}	{0.99**}	{1.00*}	{1.09**}	{0.90}	{0.65}	{0.34}	{0.19}
$R^2(\%)$	[42.22]	[46.35]	[48.46]	[50.44]	[51.11]	[55.24]	[45.83]	[30.64]	[10.62]	[3.50]
$[5^{th},\ 95^{th}]$	[24.59, 67.43]	[26.70, 73.01]	[31.12, 77.42]	[36.46, 78.03]	[35.45, 76.56]	[33.64, 83.37]	[21.67, 86.63]	[7.48, 76.41]	[0.65, 48.81]	[0.11, 36.72]

Panel B: $r^{(j)}_{k2^j+2^j} - rf^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + \epsilon_{k2^j+2^j}$

	Time-scale j			
	1	2	3	4
$v^{(j)}_t$	-0.68	0.87	-0.34	1.89
	(-0.69)	(0.43)	(-0.26)	(2.36)
$R^2(\%)$	[1.16]	[0.96]	[0.77]	[58.29]

Table 1: **Market risk. Panel A:** We run linear regressions 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, $v_{t-h,t}$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with $2 * (\text{horizon} - 1)$ lags (in parentheses), the $t/\sqrt{T}$ test suggested in Valkanov (2003) (in curly brackets), and R^2 (in square brackets). Ninety percent confidence intervals for the true R^2s are reported in brackets below the sample values. Significance at the 5%, 2.5%, and 1% level of the $t/\sqrt{T}$ test using Valkanov's (2003) critical values is indicated by *, **, and ***, respectively. **Panel B:** Component-wise predictive regressions of the components of excess stock market returns on the components of market variance. For each regression (in Panel B), the table reports OLS estimates of the regressors, t-stats in parentheses and R^2 statistics (and bootstrapped confidence intervals) in square brackets. The sample is annual and spans the period 1930-2014.

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
$v_{t-h+1,t}$	1.80	2.04	0.50	-0.04	-0.19	1.07	2.71	3.05	3.89	5.00
	(1.12)	(2.31)	(0.60)	(-0.05)	(-0.17)	(0.89)	(2.28)	(2.31)	(2.94)	(4.11)
	{0.14}	{0.22}	{0.07}	{-0.01}	{-0.03}	{0.16}	{0.45}	{0.49}	{0.62*}	{0.84**}
$R^2(\%)$	[1.94]	[4.83]	[0.48]	[0.00]	[0.08]	[2.70]	[17.44]	[19.78]	[28.43]	[42.18]
$[5^{th},\ 95^{th}]$	[0.03, 8.54]	[0.09, 15.12]	[0.01, 10.74]	[0.01, 15.78]	[0.01, 26.83]	[0.03, 32.83]	[6.55, 44.24]	[8.05, 49.43]	[12.97, 56.32]	[25.21, 61.04]

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
$v_{t-h+1,t}$	5.54	5.84	5.61	5.43	4.87	4.48	3.97	3.14	2.08	0.93
	(4.44)	(5.20)	(5.32)	(6.80)	(8.01)	(9.34)	(9.29)	(6.14)	(3.38)	(1.34)
	{0.98***}	{1.10***}	{1.09***}	{1.08**}	{0.95*}	{0.91*}	{0.82}	{0.63}	{0.39}	{0.18}
$R^2(\%)$	[50.01]	[55.68]	[55.06]	[54.69]	[48.41]	[46.13]	[41.15]	[29.03]	[13.95]	[3.29]
$[5^{th},\ 95^{th}]$	[32.65, 69.62]	[38.15, 71.14]	[38.18, 71.03]	[31.30, 72.59]	[19.49, 69.39]	[11.88, 71.55]	[4.85, 71.52]	[0.76, 62.01]	[0.42, 42.73]	[0.12, 26.83]

Panel B: $r^{(j)}_{k2^j+2^j} - rf^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + \epsilon^{(j)}_{k2^j+2^j}$

	Time-scale j			
	1	2	3	4
$v^{(j)}_t$	-6.01	-9.64	-3.24	2.81
	(-1.24)	(-1.55)	(-0.85)	(2.53)
$R^2(\%)$	[3.70]	[25.83]	[7.41]	[61.62]

Table 2: **Consumption risk. Panel A:** We run linear regressions 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, Newey-West t-statistics with $2 * (\text{horizon} - 1)$ lags (in parentheses), the $t/\sqrt{T}$ test suggested in Valkanov (2003) (in curly brackets), and R^2 (in square brackets). Ninety percent confidence intervals for the true R^2s are reported in brackets below the sample values. Significance at the 5%, 2.5%, and 1% level of the $t/\sqrt{T}$ test using Valkanov's (2003) critical values is indicated by *, **, and ***, respectively. **Panel B:** Component-wise predictive regressions of the components of excess stock market returns on the components of consumption variance $v^2_{j,t}$. For each regression (in Panel B), the table reports OLS estimates of the regressors, t-stats in parentheses and R^2 statistics (and bootstrapped confidence intervals) in square brackets. The sample is annual and spans the period 1930-2014.

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
$v_{t-h+1,t}$	0.03	0.02	0.01	0.01	0.01	0.01	0.02	0.03	0.03	0.03
	(1.94)	(1.56)	(1.05)	(0.72)	(0.75)	(0.99)	(1.64)	(2.42)	(3.07)	(3.51)
	{0.19}	{0.17}	{0.16}	{0.13}	{0.15}	{0.21}	{0.37}	{0.52}	{0.60*}	{0.70*}
$R^2(\%)$	[3.39]	[2.94]	[2.49]	[1.68]	[2.21]	[4.52]	[12.63]	[21.91]	[26.83]	[33.32]
$[5^{th},\ 95^{th}]$	[0.08, 11.81]	[0.04, 12.91]	[0.03 , 15.86]	[0.02, 18.85]	[0.02, 21.09]	[0.05, 29.56]	[0.29, 43.24]	[2.65, 53.49]	[4.05, 56.10]	[5.81, 63.18]

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
$v_{t-h+1,t}$	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02
	(3.70)	(3.87)	(3.54)	(3.27)	(3.07)	(3.00)	(2.92)	(2.92)	(3.00)	(3.35)
	{0.78*}	{0.90**}	{0.95**}	{1.00**}	{1.00**}	{1.04*}	{1.00*}	{0.88}	{0.69}	{0.52}
$R^2(\%)$	[38.77]	[45.34]	[48.41]	[50.91]	[51.07]	[52.76]	[51.04]	[44.75]	[33.43]	[21.78]
$[5^{th},\ 95^{th}]$	[7.85, 70.50]	[13.59, 77.32]	[15.11, 82.42]	[15.50, 84.17]	[13.68, 81.64]	[15.51, 79.83]	[12.93, 76.72]	[9.15, 70.75]	[4.27, 61.60]	[1.12, 58.55]

Panel B: $r^{(j)}_{k2^j+2^j} - rf^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + \epsilon_{k2^j+2^j}$

	Time-scale j			
	1	2	3	4
$v^{(j)}_t$	-0.02	0.04	0.05	0.06
	(-0.74)	(1.11)	(1.65)	(3.21)
$R^2(\%)$	[1.39]	[6.04]	[23.32]	[72.00]

Table 3: **Economic policy uncertainty (Baker, Bloom and Davis, 2015). Panel A:** We run linear regressions 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 (squared) economic policy uncertainty (EPU), $v_{t-h,t}$. For each regression, the table reports OLS estimates of the regressors, Newey-West t-statistics with $2*(\text{horizon}-1)$ lags (in parentheses), the $t/\sqrt{T}$ test suggested in Valkanov (2003) (in curly brackets), and R^2 (in square brackets). Ninety percent confidence intervals for the true R^2s are reported in brackets below the sample values. Significance at the 5%, 2.5%, and 1% level of the $t/\sqrt{T}$ test using Valkanov's (2003) critical values is indicated by *, **, and ***, respectively. **Panel B:** Component-wise predictive regressions of the components of excess stock market returns on the components of EPU. For each regression (in Panel B), the table reports OLS estimates of the regressors, t-stats in parentheses and R^2 statistics (and bootstrapped confidence intervals) in square brackets. The sample is annual and spans the period 1930-2014.

	Annual calendar time
Time-scale	Frequency resolution
$j = 1$	$1 - 2$ years
$j = 2$	$2 - 4$ years
$j = 3$	$4 - 8$ years
$j = 4$	$8 - 16$ years
$\pi_t^{(4)}$	> 16 years

Table 4: Interpretation of the time-scale (or persistence level) j in terms of time spans in the case of annual time series. Each scale corresponds to a frequency interval, or conversely an interval of periods, and thus each scale is associated with a range of time horizons.

	Panel A: Market excess returns				**Panel B: Consumption risk**			
Scales j =	1	2	3	4	1	2	3	4
1		-0.03	-0.04	0.09		0.33	0.17	-0.09
		(0.09)	(0.07)	(0.05)		(0.18)	(0.10)	(0.09)
2			-0.13	0.15			-0.09	-0.12
			(0.09)	(0.10)			(0.26)	(0.15)
3				0.14				-0.06
				(0.13)				(0.13)

	Panel C: Market risk				**Panel D: Economic policy uncertainty**			
Scales j =	1	2	3	4	1	2	3	4
1		0.27	-0.13	0.00		0.08	0.09	0.06
		(0.12)	(0.09)	(0.06)		(0.09)	(0.11)	(0.05)
2			-0.03	-0.00			0.22	0.09
			(0.15)	(0.09)			(0.09)	(0.08)
3				0.33				0.32
				(0.12)				(0.15)

Table 5: **Pairwise correlations.** We report the pair-wise correlations between the individual details of excess market returns (Panel A), consumption variance (Panel B), market variance (Panel C), and (squared) economic policy uncertainty (Panel D). The pair-wise correlations are obtained by using redundant data on the details rather than the decimated counterparts. Standard errors for the correlation between $x_t^{(j)}$ and $x_t^{(j')}$, $j \neq j'$, are Newey-West with $2^{max(j,j')}$ lags.

Panel A: Distribution of coefficient estimates.

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Mean	1.83	1.56	1.03	0.39	0.17	0.00	-0.14	-0.27	-0.39	-0.50
Median	1.82	1.55	1.05	0.42	0.20	0.03	-0.12	-0.27	-0.39	-0.50
SD	1.53	1.61	1.82	2.29	2.56	2.85	3.16	3.50	3.86	4.23
$[5^{th},\ 95^{th}]$	[-0.66, 4.35]	[-1.10, 4.19]	[-1.98, 3.98]	[-3.39, 4.08]	[-4.01, 4.29]	[-4.66, 4.58]	[-5.26, 4.94]	[-5.90, 5.37]	[-6.59, 5.84]	[-7.25, 6.30]
β increasing 6-12 (%)	14.64									
β decreasing 16-20 (%)	35.21									
β hump-shape (%)	6.14									

Panel B: Distribution of R^2s.

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Mean	2.98	4.39	5.88	9.25	11.46	13.93	16.58	19.37	22.19	24.97
Median	1.89	2.55	3.08	4.90	6.33	8.09	10.15	12.43	15.15	17.93
SD	3.26	5.07	7.24	11.09	13.28	15.50	17.66	19.73	21.61	23.25
$[5^{th},\ 95^{th}]$	[0.02, 9.62]	[0.02, 14.92]	[0.03, 21.24]	[0.05, 33.02]	[0.05, 39.93]	[0.07, 47.25]	[0.09, 54.08]	[0.12, 60.63]	[0.15, 66.11]	[0.19, 71.00]
R^2 increasing 6-12 (%)	16.81									
R^2 decreasing 16-20 (%)	18.29									
R^2 hump-shape (%)	6.54									
R^2 hump-shape & $R^2 > 50\%$ (%)	3.05									
R^2 hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	1.28									
R^2 and β hump-shape (%)	2.28									
R^2 and β hump-shape & $R^2 > 50\%$ (%)	1.15									
R^2 and β hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	0.55									

Table 6: **Classical predictive system.** We simulate under the assumption of predictability and an AR(1) process on x_t,

$$R_{t+1} = \beta x_t + \varepsilon_{t+1} \ ,$$
$$x_{t+1} = \rho x_t + u_{t+1}$$

Simulations are performed using the parameters $\beta = 1.7990$, $\rho = 0.7335$, $\sigma_\epsilon = 0.1802$, $\sigma_u = 0.0095$ and $corr_{\epsilon,u} = -0.0454$. There are 85 observations for each simulation. **Panel A: Distribution of coefficient estimates.** The Table reports the mean, standard deviation, and median of the coefficient estimates from the predictive regression across 100,000 simulations. **Panel B: Distribution of R^2s.** The Table reports the mean, standard deviation, and median of the R^2s from the predictive regression across 100,000 simulations. "β (R^2) increasing 6-12" is the percentage of the simulations that produce coefficients (R^2s) that are monotonic in the horizons 6 to 12 years, i.e. $\beta_{12} > \beta_{11} > \ldots > \beta_6$ ($R^2_{12} > R^2_{11} > \ldots > R^2_6$, respectively). "$\beta$ (R^2) decreasing 16-20" is the percentage of the simulations that produce coefficients (R^2s) that are decreasing in the horizons 16 to 20 years, i.e. $\beta_{16} > \beta_{17} > \ldots > \beta_{20}$ ($R^2_{16} > R^2_{17} > \ldots > R^2_{20}$, respectively). "$\beta$ (R^2) hump-shape" is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. "R^2 hump-shape & $> 50\%$" is the percentage of the simulations that produce R^2s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an $R^2_h > 50\%$ in the range $12 \leq h \leq 16$.

Panel A: Distribution of coefficient estimates.

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Mean	0.04	0.04	0.05	0.07	0.08	0.09	0.10	0.11	0.12	0.13
Median	0.03	0.03	0.04	0.07	0.09	0.10	0.11	0.11	0.13	0.14
SD	1.66	1.75	1.96	2.45	2.72	3.03	3.36	3.72	4.10	4.49
$[5^{th},\ 95^{th}]$	[-2.67, 2.76]	[-2.81, 2.91]	[-3.15, 3.27]	[-3.90, 4.06]	[-4.33, 4.52]	[-4.81, 4.98]	[-5.31, 5.51]	[-5.86, 6.12]	[-6.45, 6.73]	[-7.06, 7.31]
β increasing 6-12 (%)	21.02									
β decreasing 16-20 (%)	32.35									
β hump-shape (%)	7.83									

Panel B: Distribution of R^2s.

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Mean	1.22	2.28	4.45	9.06	11.53	14.11	16.77	19.48	22.19	24.81
Median	0.56	1.07	2.20	4.81	6.40	8.25	10.33	12.60	15.12	17.88
SD	1.70	3.09	5.78	10.92	13.33	15.62	17.78	19.78	21.54	23.10
$[5^{th},\ 95^{th}]$	[0.00, 4.65]	[0.01, 8.61]	[0.02, 16.59]	[0.04, 32.56]	[0.06, 40.16]	[0.08, 47.63]	[0.10, 54.31]	[0.12, 60.69]	[0.15, 65.98]	[0.18, 70.57]
R^2 increasing 6-12 (%)	19.05									
R^2 decreasing 16-20 (%)	18.80									
R^2 hump-shape (%)	7.56									
R^2 hump-shape & $R^2 > 50\%$ (%)	3.41									
R^2 hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	1.40									
R^2 and β hump-shape (%)	2.66									
R^2 and β hump-shape & $R^2 > 50\%$ (%)	1.29									
R^2 and β hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	0.64									

Table 7: **Classical predictive system.** We simulate under the assumption of no predictability and an AR(1) process on x_t,

$$R_{t+1} = \varepsilon_{t+1}\ ,$$
$$x_{t+1} = \rho x_t + u_{t+1}$$

Simulations are performed using the parameters $\rho = 0.7335$, $\sigma_\epsilon = 0.1951$, $\sigma_u = 0.0095$ and $corr_{\epsilon,u} = -0.0454$. There are 85 observations for each simulation. **Panel A: Distribution of coefficient estimates.** The Table reports the mean, standard deviation, and median of the coefficient estimates from the predictive regression from the predictive regression across 100,000 simulations. **Panel B: Distribution of R^2s.** The Table reports the mean, standard deviation, and median of the R^2s from the predictive regression across 100,000 simulations. "β (R^2) increasing 6-12" is the percentage of the simulations that produce coefficients (R^2s) that are monotonic in the horizons 6 to 12 years, i.e. $\beta_{12} > \beta_{11} > \ldots > \beta_6$ ($R^2_{12} > R^2_{11} > \ldots > R^2_6$, respectively). "$\beta$ (R^2) decreasing 16-20" is the percentage of the simulations that produce coefficients (R^2s) that are decreasing in the horizons 16 to 20 years, i.e. $\beta_{16} > \beta_{17} > \ldots > \beta_{20}$ ($R^2_{16} > R^2_{17} > \ldots > R^2_{20}$, respectively). "$\beta$ (R^2) hump-shape" is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. "R^2 hump-shape & $> 50\%$" is the percentage of the simulations that produce R^2s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an $R^2_h > 50\%$ in the range $12 \leq h \leq 16$.

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

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Median of β_h	0.07	-0.26	-1.20	-1.70	0.28	2.58	4.12	4.48	4.05	2.14
SD of β_h	(1.15)	(1.11)	(0.90)	(1.18)	(1.21)	(1.22)	(1.43)	(1.70)	(1.77)	(1.82)
Median of $Adj.R^2$	[0.04]	[0.09]	[2.91]	[7.55]	[10.65]	[17.75]	[43.82]	[53.65]	[44.69]	[26.67]

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

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
Median of β_h	0.26	0.24	0.16	-0.28	-0.69	-1.25	-1.81	-2.10	-1.84	-1.45
SD of β_h	(1.34)	(1.43)	(1.51)	(1.62)	(1.72)	(1.77)	(1.91)	(2.10)	(2.07)	(2.06)
Median of $Adj.R^2$	[0.42]	[1.12]	[2.11]	[3.03]	[4.34]	[5.24]	[8.54]	[12.13]	[10.57]	[8.25]

Panel C: Distribution of coefficient estimates and of R^2s

β increasing 6-12 (%)	52.20
β decreasing 16-20 (%)	66.80
β hump-shape (%)	37.20
R^2 increasing 6-12 (%)	39.20
R^2 decreasing 16-20 (%)	63.40
R^2 hump-shape (%)	31.20
R^2 hump-shape & $R^2 > 50\%$ (%)	23.40
R^2 hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	18.80
R^2 and β hump-shape (%)	26.20
R^2 and β hump-shape & $R^2 > 50\%$ (%)	21.20
R^2 and β hump-shape & $R^2 > 50\%$ & $R^2_{16} - R^2_{20} > 30\%$(%)	18.80

Table 8: **Simulation under the null of scale-dependent predictability. The relation is at scale $j^* = 4$.** We simulate excess market returns (y) and consumption volatility (x) under the assumption of predictability at scale $j^* = 4$. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_t^{(j)}$ for $j = 4$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 100000 replications. We set $T = 128$. For each regression, the table reports the median and the standard deviation (in parentheses) of the coefficient estimates from the predictive regression as well as the median of the adjusted R^2 statistics (in square brackets). **Panel A: two-way (forward/backward) regressions.** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period *past* consumption volatility. **Panel B: contemporaneous aggregation.** We run linear regressions (with an intercept) of h-period continuously compounded excess market returns on h-period *contemporaneous* consumption volatility. **Panel C: Distribution of coefficient estimates and of R^2s.** "β (R^2) increasing 6-12" is the percentage of the simulations that produce coefficients (R^2s) that are monotonic in the horizons 6 to 12 years, i.e. $\beta_{12} > \beta_{11} > \ldots > \beta_6$ ($R^2_{12} > R^2_{11} > \ldots > R^2_6$, respectively). "$\beta$ (R^2) decreasing 16-20" is the percentage of the simulations that produce coefficients (R^2s) that are decreasing in the horizons 16 to 20 years, i.e. $\beta_{16} > \beta_{17} > \ldots > \beta_{20}$ ($R^2_{16} > R^2_{17} > \ldots > R^2_{20}$, respectively). "$\beta$ (R^2) hump-shape" is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. "R^2 hump-shape & $> 50\%$" is the percentage of the simulations that produce R^2s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an $R^2_h > 50\%$ in the range $12 \leq h \leq 16$.

	Horizon h (in years)									
	1	2	4	8	10	12	14	16	18	20
$x_{t-h+1,t}$	0.00	0.00	-0.00	-0.00	0.00	-0.01	0.00	-0.01	0.00	-0.00
	(0.10)	(0.12)	(0.14)	(0.10)	(0.12)	(0.14)	(0.16)	(0.19)	(0.20)	(0.21)
$Adj.R^2$	[0.01]	[0.09]	[0.58]	[0.33]	[1.11]	[1.42]	[1.67]	[2.20]	[2.54]	[2.24]

Table 9: **Simulation under the null of ABSENCE of scale-dependent predictability**. We simulate excess market returns (y) and market variance (x) under the assumption of no predictability. We simulate $x_t^{(j)} = \rho_j x_{t-2^j}^{(j)} + \epsilon_{t,j}$ for $j = 4$ and $x_t^{(j)} = \epsilon_t^{(j)}$ otherwise. We implement 500 replications. We set $T = 128$. 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 the median and the standard deviation (in parentheses) of the coefficient estimates from the predictive regression as well as the median of the adjusted R^2 statistics (in square brackets).

Panel A: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.39)	(2.57)	(2.78)	(3.19)	(3.44)	(3.20)	(3.07)	(3.05)	(3.24)	(3.13)	(2.97)	(2.85)	(2.86)	(2.83)	(2.89)	(3.22)
$R^2(\%)$	[7.30]	[13.04]	[15.47]	[18.98]	[24.85]	[32.08]	[35.48]	[38.00]	[39.73]	[42.04]	[42.99]	[41.84]	[42.32]	[44.56]	[47.73]	[49.35]

Panel B: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.11	0.13	0.14	0.16	0.17	0.17	0.18
	(1.44)	(2.48)	(3.18)	(3.63)	(3.86)	(3.59)	(3.58)	(4.02)	(4.36)	(4.35)	(4.23)	(3.84)	(3.33)	(2.86)	(2.48)	(2.15)
$R^2(\%)$	[2.82]	[8.36]	[15.72]	[21.48]	[23.34]	[25.95]	[30.46]	[35.69]	[42.47]	[42.75]	[40.46]	[39.59]	[40.53]	[35.58]	[30.42]	[27.72]

Panel C: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.38)	(2.64)	(3.27)	(4.56)	(5.81)	(6.54)	(8.73)	(11.79)	(22.01)	(23.33)	(17.96)	(16.92)	(19.36)	(11.89)	(9.39)	(10.06)
$v_{t-15,t}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.11	0.13	0.14	0.16	0.17	0.17	0.18
	(1.42)	(2.17)	(2.79)	(3.34)	(3.50)	(3.92)	(4.66)	(5.57)	(6.87)	(6.66)	(6.22)	(7.30)	(10.87)	(11.20)	(11.38)	(12.87)
$R^2(\%)$	[10.11]	[21.40]	[31.19]	[40.46]	[48.20]	[58.02]	[65.93]	[73.69]	[82.20]	[84.79]	[83.45]	[81.44]	[82.85]	[80.14]	[78.14]	[77.08]

Table 10: **Excess returns and Market risk. Panel A: log DP-only.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio. **Panel B: log market variance only.** We run linear regressions (with an intercept) of excess returns on log market variance. **Panel C: Multiple regressions of excess returns on log dividend-price and log market variance.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio and log market variance. $R_{t+1,t+h}$ represents the total return from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Market variance is aggregated over $H = 16$ years. The sample is annual and spans the period 1945-2014.

Panel A: Direct regression, $k = 16$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.52	0.35	-0.17
	(4.96)	(1.90)	(-0.98)
$R^2(\%)$	[60.67]	[25.79]	[3.48]
$v_{t-h,t}$	0.04	0.02	-0.01
	(2.50)	(4.19)	(-1.66)
$R^2(\%)$	[16.07]	[45.83]	[7.55]

Panel B: Direct regression, $k = 18$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.50	0.31	-0.19
	(5.74)	(1.69)	(-1.19)
$R^2(\%)$	[61.08]	[20.09]	[4.96]
$v_{t-h,t}$	0.03	0.03	-0.00
	(2.01)	(4.96)	(-1.39)
$R^2(\%)$	[13.73]	[58.21]	[1.89]

Table 11: **Long-Run Regression Coefficients: Market risk.** Direct regressions on log DP and log market variance aggregated over $H = 16$ years. **Panel A: Direct regression, $k = 16$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. **Panel B: Direct regression , $k = 18$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. Table entries are long-run regression coefficients, for example, $\beta_{r,dp}^{(k)}$ in $\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \alpha + \beta_{r,dp}^{(k)} dp_t + \varepsilon_{t+k}^r$. Annual CRSP data, 1945-2014.

Panel A: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.50)	(-0.14)	(-0.28)	(-0.37)	(-0.53)	(-0.49)	(-0.59)	(-0.59)	(-0.61)	(-0.61)	(-0.65)	(-0.55)	(-0.47)	(-0.28)	(-0.12)	(0.01)
$R^2(\%)$	[0.50]	[0.06]	[0.25]	[0.36]	[0.90]	[1.19]	[2.49]	[2.95]	[3.19]	[2.93]	[3.73]	[3.05]	[2.80]	[1.18]	[0.27]	[0.00]

Panel B: $D_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04
	(2.36)	(3.68)	(4.49)	(5.06)	(5.37)	(5.08)	(5.36)	(5.74)	(6.01)	(6.08)	(5.54)	(5.51)	(6.22)	(7.89)	(9.80)	(10.73)
	[9.69]	[20.45]	[28.78]	[33.26]	[35.85]	[36.73]	[41.81]	[47.85]	[50.62]	[51.09]	[49.84]	[49.08]	[50.31]	[54.02]	[60.09]	[64.89]

Panel C: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.54)	(-0.17)	(-0.37)	(-0.58)	(-1.07)	(-1.12)	(-1.38)	(-1.42)	(-1.55)	(-1.63)	(-1.65)	(-1.33)	(-1.09)	(-0.64)	(-0.28)	(0.03)
$v_{t-15,t}$	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04
	(2.37)	(3.70)	(4.61)	(5.25)	(5.78)	(5.57)	(6.12)	(6.47)	(6.71)	(6.84)	(6.60)	(6.48)	(6.92)	(7.99)	(9.59)	(10.77)
$R^2(\%)$	[10.19]	[20.51]	[29.03]	[33.62]	[36.75]	[37.91]	[44.30]	[50.80]	[53.81]	[54.02]	[53.57]	[52.12]	[53.11]	[55.20]	[60.36]	[64.89]

Table 12: **Dividend growth and Market risk. Panel A: log DP-only.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio. **Panel B: log market variance only.** We run linear regressions (with an intercept) of dividend growth on log market variance. **Panel C: Multiple regressions of dividend growth on log dividend-price and log market variance.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio and log market variance. $D_{t+1,t+h}$ represents the growth rate of dividends from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Market variance is aggregated over $H = 16$ years. The sample is annual and spans the period 1945-2014.

Panel A: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.39)	(2.57)	(2.78)	(3.19)	(3.44)	(3.20)	(3.07)	(3.05)	(3.24)	(3.13)	(2.97)	(2.85)	(2.86)	(2.83)	(2.89)	(3.22)
$R^2(\%)$	[7.30]	[13.04]	[15.47]	[18.98]	[24.85]	[32.08]	[35.48]	[38.00]	[39.73]	[42.04]	[42.99]	[41.84]	[42.32]	[44.56]	[47.73]	[49.35]

Panel B: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	-0.01	-0.01	0.00	0.01	0.02	0.03	0.06	0.07	0.09	0.11	0.12	0.14	0.17	0.17	0.16	0.17
	(-2.16)	(-1.30)	(0.20)	(0.87)	(0.88)	(1.41)	(2.67)	(2.60)	(2.54)	(2.71)	(2.77)	(3.16)	(3.65)	(3.35)	(2.72)	(2.40)
$R^2(\%)$	[4.89]	[2.42]	[0.09]	[2.18]	[2.40]	[6.00]	[15.01]	[16.04]	[20.30]	[20.40]	[18.69]	[20.90]	[26.18]	[22.58]	[16.06]	[14.26]

Panel C: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.45)	(2.55)	(2.80)	(3.59)	(4.13)	(4.29)	(4.94)	(4.97)	(5.88)	(5.56)	(5.00)	(5.04)	(5.78)	(5.11)	(4.48)	(4.73)
$v_{t-15,t}$	-0.01	-0.01	0.00	0.01	0.02	0.03	0.06	0.07	0.09	0.11	0.12	0.14	0.17	0.17	0.16	0.17
	(-2.40)	(-1.85)	(0.27)	(1.00)	(0.94)	(1.52)	(3.03)	(3.31)	(4.25)	(4.73)	(5.00)	(6.36)	(9.81)	(9.49)	(9.28)	(12.60)
$R^2(\%)$	[12.18]	[15.46]	[15.56]	[21.16]	[27.25]	[38.08]	[50.48]	[54.04]	[60.03]	[62.45]	[61.68]	[62.74]	[68.50]	[67.14]	[63.78]	[63.61]

Table 13: **Excess returns and Consumption risk. Panel A: log DP-only.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio. **Panel B: log consumption variance only.** We run linear regressions (with an intercept) of excess returns on log consumption variance. **Panel C: Multiple regressions of excess returns on log dividend-price and log consumption variance.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio and log consumption variance. $R_{t+1,t+h}$ represents the total return from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Consumption variance is aggregated over $H = 16$ years. The sample is annual and spans the period 1945-2014.

Panel A: Direct regression, $k = 16$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.52	0.35	-0.17
	(4.96)	(1.90)	(-0.98)
$R^2(\%)$	[60.67]	[25.79]	[3.48]
$v_{t-h,t}$	0.03	0.02	-0.01
	(2.36)	(4.74)	(-0.76)
$R^2(\%)$	[6.96]	[34.09]	[0.75]

Panel B: Direct regression, $k = 18$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.50	0.31	-0.19
	(5.74)	(1.69)	(-1.19)
$R^2(\%)$	[61.08]	[20.09]	[4.96]
$v_{t-h,t}$	0.03	0.03	0.00
	(1.92)	(5.30)	(0.29)
$R^2(\%)$	[5.60]	[40.24]	[0.05]

Table 14: **Long-Run Regression Coefficients: Consumption risk.** Direct regressions on log DP and log consumption variance aggregated over $H = 16$ years. **Panel A: Direct regression, $k = 16$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. **Panel B: Direct regression , $k = 18$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. Table entries are long-run regression coefficients, for example, $\beta_{r,dp}^{(k)}$ in $\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \alpha + \beta_{r,dp}^{(k)} dp_t + \varepsilon_{t+k}^r$. Annual CRSP data, 1945-2014.

Panel A: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.50)	(-0.14)	(-0.28)	(-0.37)	(-0.53)	(-0.49)	(-0.59)	(-0.59)	(-0.61)	(-0.61)	(-0.65)	(-0.55)	(-0.47)	(-0.28)	(-0.12)	(0.01)
$R^2(\%)$	[0.50]	[0.06]	[0.25]	[0.36]	[0.90]	[1.19]	[2.49]	[2.95]	[3.19]	[2.93]	[3.73]	[3.05]	[2.80]	[1.18]	[0.27]	[0.00]

Panel B: $D_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
	(1.01)	(1.39)	(1.85)	(2.02)	(2.47)	(3.00)	(4.79)	(5.20)	(5.29)	(4.80)	(6.11)	(7.35)	(6.28)	(6.39)	(7.27)	(7.75)
$R^2(\%)$	[2.82]	[3.78]	[6.43]	[10.55]	[16.96]	[23.83]	[38.28]	[40.30]	[36.88]	[32.70]	[29.93]	[28.94]	[35.23]	[35.72]	[37.32]	[35.34]

Panel C: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.49)	(-0.13)	(-0.27)	(-0.40)	(-0.64)	(-0.69)	(-0.85)	(-0.86)	(-0.89)	(-0.89)	(-0.95)	(-0.80)	(-0.71)	(-0.41)	(-0.17)	(0.01)
$v_{t-15,t}$	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
	(1.00)	(1.36)	(1.77)	(1.97)	(2.43)	(3.02)	(5.11)	(5.15)	(4.60)	(3.92)	(4.51)	(5.49)	(6.16)	(6.15)	(7.12)	(7.79)
$R^2(\%)$	[3.33]	[3.84]	[6.68]	[10.92]	[17.86]	[25.02]	[40.77]	[43.26]	[40.07]	[35.63]	[33.65]	[31.99]	[38.02]	[36.90]	[37.58]	[35.34]

Table 15: **Dividend growth and Consumption risk. Panel A: log DP-only.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio. **Panel B: log consumption variance only.** We run linear regressions (with an intercept) of dividend growth on log consumption variance. **Panel C: Multiple regressions of dividend growth on log dividend-price and log consumption variance.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio and log consumption variance. $D_{t+1,t+h}$ represents the growth rate of dividends from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Consumption variance is aggregated over $H = 16$ years. The sample is annual and spans the period 1945-2014.

Panel A: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.39)	(2.57)	(2.78)	(3.19)	(3.44)	(3.20)	(3.07)	(3.05)	(3.24)	(3.13)	(2.97)	(2.85)	(2.86)	(2.83)	(2.89)	(3.22)
$R^2(\%)$	[7.30]	[13.04]	[15.47]	[18.98]	[24.85]	[32.08]	[35.48]	[38.00]	[39.73]	[42.04]	[42.99]	[41.84]	[42.32]	[44.56]	[47.73]	[49.35]

Panel B: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.09	0.09
	(1.63)	(2.48)	(2.67)	(2.84)	(3.03)	(3.07)	(3.23)	(3.50)	(3.65)	(3.58)	(3.45)	(3.13)	(2.76)	(2.42)	(2.13)	(1.86)
$R^2(\%)$	[3.01]	[6.88]	[12.76]	[18.45]	[21.28]	[24.57]	[29.25]	[33.34]	[38.61]	[39.69]	[40.16]	[39.98]	[40.21]	[36.73]	[32.43]	[29.57]

Panel C: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.11	0.22	0.29	0.40	0.59	0.78	0.94	1.15	1.32	1.56	1.82	2.06	2.32	2.75	3.38	4.07
	(2.41)	(2.69)	(3.22)	(4.36)	(5.43)	(5.75)	(6.48)	(7.33)	(9.01)	(10.14)	(11.55)	(12.53)	(13.91)	(15.81)	(15.71)	(17.58)
$v_{t-15,t}$	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.09	0.09
	(1.66)	(2.43)	(2.75)	(3.14)	(3.63)	(4.51)	(6.00)	(8.57)	(11.69)	(16.13)	(19.06)	(17.94)	(13.15)	(11.98)	(11.15)	(9.06)
$R^2(\%)$	[10.08]	[19.33]	[28.03]	[37.66]	[46.77]	[57.58]	[66.04]	[73.29]	[80.70]	[83.88]	[85.06]	[83.51]	[84.16]	[82.75]	[81.21]	[79.54]

Table 16: **Excess returns and economic policy uncertainty (Baker, Bloom and Davis, 2015) aggregated over $H = 16$ years. Panel A: log DP-only.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio. **Panel B: economic policy uncertainty only.** We run linear regressions (with an intercept) of excess returns on economic policy uncertainty. **Panel C: Multiple regressions of excess returns on log dividend-price and economic policy uncertainty.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio and (log of squared) economic policy uncertainty. $R_{t+1,t+h}$ represents the total return from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Economic policy uncertainty is aggregated over $H = 16$ years. The sample is annual and spans the period 1930 - 2014.

Panel A: Direct regression, $k = 16$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.52	0.35	-0.17
	(4.96)	(1.90)	(-0.98)
$R^2(\%)$	[60.67]	[25.79]	[3.48]
$v_{t-h,t}$	0.03	0.01	-0.02
	(2.27)	(2.60)	(-5.61)
$R^2(\%)$	[32.42]	[37.26]	[45.71]

Panel B: Direct regression, $k = 18$

	Coefficients		
Right-Hand Variable	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
dp_t	1.50	0.31	-0.19
	(5.74)	(1.69)	(-1.19)
$R^2(\%)$	[61.08]	[20.09]	[4.96]
$v_{t-h,t}$	0.02	0.01	-0.01
	(2.17)	(3.05)	(-5.44)
$R^2(\%)$	[29.81]	[47.39]	[33.34]

Table 17: **Long-Run Regression Coefficients on log DP and (log of squared) economic policy uncertainty (Baker, Bloom and Davis, 2015) aggregated over $H = 16$ years. Panel A: Direct regression, $k = 16$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. **Panel B: Direct regression , $k = 18$.** "Direct" regression estimates are calculated using k-year ex post returns, dividend growth, and dividend yields as left-hand variables. Table entries are long-run regression coefficients, for example, $\beta_{r,dp}^{(k)}$ in $\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \alpha + \beta_{r,dp}^{(k)} dp_t + \varepsilon_{t+k}^r$. Annual CRSP data, 1945-2014.

Panel A: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.50)	(-0.14)	(-0.28)	(-0.37)	(-0.53)	(-0.49)	(-0.59)	(-0.59)	(-0.61)	(-0.61)	(-0.65)	(-0.55)	(-0.47)	(-0.28)	(-0.12)	(0.01)
$R^2(\%)$	[0.50]	[0.06]	[0.25]	[0.36]	[0.90]	[1.19]	[2.49]	[2.95]	[3.19]	[2.93]	[3.73]	[3.05]	[2.80]	[1.18]	[0.27]	[0.00]

Panel B: $D_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-15,t}$	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
	(1.85)	(1.75)	(1.73)	(1.77)	(1.86)	(2.02)	(2.38)	(2.84)	(3.38)	(3.96)	(4.46)	(5.04)	(5.44)	(5.46)	(5.34)	(5.27)
$R^2(\%)$	[3.79]	[7.65]	[9.93]	[11.78]	[13.71]	[17.39]	[23.18]	[30.58]	[35.33]	[38.28]	[39.93]	[42.65]	[46.28]	[50.87]	[56.47]	[59.96]

Panel C: $D_{t+1,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.10	-0.11	-0.13	-0.13	-0.12	-0.08	-0.04	0.00
	(-0.52)	(-0.15)	(-0.31)	(-0.47)	(-0.80)	(-0.83)	(-1.06)	(-1.17)	(-1.36)	(-1.57)	(-1.67)	(-1.42)	(-1.16)	(-0.71)	(-0.32)	(0.03)
$v_{t-15,t}$	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
	(1.84)	(1.75)	(1.72)	(1.75)	(1.83)	(1.99)	(2.31)	(2.76)	(3.25)	(3.77)	(4.15)	(4.58)	(4.70)	(4.87)	(5.05)	(5.30)
$R^2(\%)$	[4.30]	[7.72]	[10.18]	[12.14]	[14.61]	[18.58]	[25.68]	[33.54]	[38.52]	[41.21]	[43.65]	[45.70]	[49.08]	[52.05]	[56.74]	[59.97]

Table 18: **Dividend growth and economic policy uncertainty (Baker, Bloom and Davis, 2015) aggregated over $H = 16$ years. Panel A: log DP-only.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio. **Panel B: economic policy uncertainty only.** We run linear regressions (with an intercept) of dividend growth on economic policy uncertainty. **Panel C: Multiple regressions of dividend growth on log dividend-price and economic policy uncertainty.** We run linear regressions (with an intercept) of dividend growth on log dividend-price ratio and (log of squared) economic policy uncertainty. $D_{t+1,t+h}$ represents the growth rate of dividends from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Economic policy uncertainty is aggregated over $H = 16$ years. The sample is annual and spans the period 1930 - 2014.

Panel A: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.07	0.12	0.13	0.17	0.30	0.39	0.47	0.60	0.75	0.90	1.09	1.24	1.43	1.76	2.27	2.91
	(1.17)	(1.22)	(1.24)	(1.90)	(3.30)	(4.87)	(4.57)	(4.28)	(3.98)	(4.47)	(5.33)	(6.80)	(7.33)	(6.25)	(5.05)	(4.55)
$R^2(\%)$	[2.45]	[3.68]	[2.74]	[3.49]	[6.87]	[10.64]	[13.59]	[18.11]	[22.05]	[25.74]	[27.88]	[28.36]	[31.02]	[33.42]	[37.19]	[40.77]

Panel B: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$v_{t-7,t}$	0.06	0.15	0.25	0.37	0.48	0.61	0.70	0.80	0.95	1.21	1.52	1.91	2.22	2.45	2.49	2.49
	(0.96)	(1.57)	(1.84)	(2.15)	(2.27)	(2.64)	(2.84)	(3.46)	(4.36)	(6.88)	(8.55)	(7.09)	(6.56)	(7.30)	(7.02)	(5.38)
$R^2(\%)$	[1.28]	[3.62]	[6.71]	[9.88]	[10.93]	[15.01]	[17.38]	[18.80]	[20.49]	[27.40]	[33.39]	[43.46]	[49.99]	[47.19]	[38.30]	[32.37]

Panel C: $R_{t+1,t+h} - R^f_{t,t+h} = \alpha_h + \beta_h dp_t + \beta_{h,v} v_{t-h+1,t} + \epsilon_{t+h}$

	Horizon h (in years)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dp_t	0.07	0.12	0.13	0.17	0.30	0.39	0.47	0.60	0.75	0.90	1.09	1.24	1.43	1.76	2.27	2.91
	(1.15)	(1.17)	(1.16)	(1.68)	(2.41)	(2.84)	(3.98)	(4.23)	(4.15)	(4.87)	(6.55)	(8.37)	(11.08)	(31.27)	(13.68)	(9.21)
$v_{t-7,t}$	0.06	0.15	0.25	0.37	0.48	0.61	0.70	0.80	0.95	1.21	1.52	1.91	2.22	2.45	2.49	2.49
	(0.95)	(1.43)	(1.65)	(1.86)	(1.84)	(2.06)	(2.18)	(2.52)	(3.12)	(5.17)	(9.98)	(11.68)	(10.92)	(13.53)	(16.95)	(10.47)
$R^2(\%)$	[3.73]	[7.30]	[9.45]	[13.36]	[17.79]	[25.65]	[30.97]	[36.91]	[42.54]	[53.14]	[61.27]	[71.82]	[81.01]	[80.61]	[75.49]	[73.14]

Table 19: **Excess returns and Macroeconomic Uncertainty (Kurado, Ludvigson and Ng, 2014) aggregated over $H = 8$ years. Panel A: log DP-only.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio. **Panel B: macroeconomic uncertainty only.** We run linear regressions (with an intercept) of excess returns on macroeconomic uncertainty. **Panel C: Multiple regressions of excess returns on log dividend-price and macroeconomic uncertainty.** We run linear regressions (with an intercept) of excess returns on log dividend-price ratio and macroeconomic uncertainty. $R_{t+1,t+h}$ represents the total return from time t to time $t+h$. For each regression, the table reports OLS estimates of the regressors, corrected t-stats in parentheses and R^2 statistics in square brackets at the different horizons. Standard errors are Newey–West with $2 \times (\text{horizon} - 1)$ lags. We use overlapping observations for the full sample. Macroeconomic Uncertainty is aggregated over $H = 8$ years. The sample is annual and spans the period 1967 - 2014.

Appendix

A Proofs

A.1 Proof of Proposition I.

The logic of the proof is as follows.

1. Postulate the dynamics of the decimated components for each level $j = 1, \ldots, \infty$. For example, we may have, for some scale j,

$$x^{(j)}_{k2^j+2^j} \quad = \quad \rho_j x^{(j)}_{k2^j} + e^{(j)}_{k2^j+2^j}.$$

2. For each scale j, rewrite the dynamics in the form

$$x^{(j)}_t = \sum_{k=0}^{+\infty} \psi^{(j)}_k \varepsilon^{(j)}_{t-k2^j}.$$

   If the dynamics are autoregressive, as above, we immediately see that $\psi^{(j)}_k = \rho^k_j$ and $\varepsilon^{(j)}_{t-k2^j} = e^{(j)}_{t-k2^j}$.

3. For the series x_t, defined as

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \psi^{(j)}_k \varepsilon^{(j)}_{t-k2^j},$$

   to be stationary, we need to verify that

   - the shocks $\varepsilon^{(j)}_{t-k2^j}$ are uncorrelated on $S^{(j)}_t = \{t - k2^j : \quad k \in \mathbb{Z}\}$;
   - and the weighted series of coefficients $\psi^{(j)}_k$ is square summable, i.e.,

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

     where the weights $\delta^{(j)}_i$ are so that we can write

$$\varepsilon^{(j)}_t = \sum_{i=0}^{2^j-1} \delta^{(j)}_i \varepsilon_{t-i}.$$

Square summability guarantees finite variance. Why, however, do we need to weight the $\psi^{(j)}_k$s? Because x_t ought to be defined for all t. Thus, $\varepsilon^{(j)}_t$ has to be defined for all t and not only on the set $S^{(j)}_t = \{t - k2^j : k \in \mathbb{Z}\}$. We have already imposed a condition on $\varepsilon^{(j)}_t$ on the grid $S^{(j)}_t$. Theorem 3 in Ortu, Severino, Tamoni, and Tebaldi (2015) requires that the shocks $\varepsilon^{(j)}_t$ are moving averages on the points within the grid:

$$\varepsilon^{(j)}_t = \sum_{i=0}^{2^j-1} \delta^{(j)}_i \varepsilon_{t-i}.$$

If this is the case, then x_t is defined for all t and covariance-stationary.

Let us, now, return to our setting.

We assume that

$$\begin{aligned} y^{(j)}_{k2^j+2^j} &= \beta_j x^{(j)}_{k2^j} + u^{(j)}_{k2^j+2^j} \\ x^{(j)}_{k2^j+2^j} &= \rho_{x,j} x^{(j)}_{k2^j} + e^{(j)}_{k2^j+2^j}, \end{aligned} \tag{A.1}$$

for $j = j^*$ and

$$y^{(j)}_{k2^j+2^j} = \rho_{y,j} y^{(j)}_{k2^j} + \varepsilon^{(j)}_{k2^j+2^j}$$

for $j \neq j^*$, where k is defined as above and $j = 1, \ldots, J$. This specification is more general than what we consider in the main text (where $y^{(j)}_{k2^j+2^j} = \varepsilon^{(j)}_{k2^j+2^j}$). Note that, at scale $j = j^*$,

$$y^{(j)}_{k2^j+2^j} = u^{(j)}_{k2^j+2^j} + \beta_j \times \left(\sum_{i=0}^{\infty} \rho^i_{x,j} e^{(j)}_{k2^j-i2^j} \right).$$

Naturally, $y^{(j)}_{k2^j+2^j}$ inherits the shocks of x. At scale $j = j^*$ the sequence of shocks is

$$\underbrace{u^{(j)}_{k2^j+2^j}}_{\varepsilon^{(j)}_{k2^j+2^j}}, \underbrace{e^{(j)}_{k2^j}}_{\varepsilon^{(j)}_{k2^j}}, \underbrace{e^{(j)}_{k2^j-2^j}}_{\varepsilon^{(j)}_{k2^j-2^j}}, \underbrace{e^{(j)}_{k2^j-2\times 2^j}}_{\varepsilon^{(j)}_{k2^j-2\times 2^j}}, \ldots$$

Being constructive, we show that it is possible to find a series of one-period high-frequency shocks so that

$$\varepsilon^{(j)}_t = \sum_{i=0}^{2^j-1} \delta^{(j)}_i \varepsilon_{t-i} \tag{A.2}$$

holds true. Assume, for conciseness, that $T = 16$, $j^* = 2$, and $J = 3$. Arrange the decimated shocks (coming from y or x depending on the scale) as follows:

$$\begin{pmatrix} \varepsilon^{(3)}_{\pi,8} & \varepsilon^{(3)}_{\pi,16} \\ \varepsilon^{(3)}_{r,8} & \varepsilon^{(3)}_{r,16} \\ e^{(2)}_{v,8} & u^{(2)}_{16} \\ e^{(2)}_{v,4} & e^{(2)}_{v,12} \\ \varepsilon^{(1)}_{r,8} & \varepsilon^{(1)}_{r,16} \\ \varepsilon^{(1)}_{r,6} & \varepsilon^{(1)}_{r,14} \\ \varepsilon^{(1)}_{r,4} & \varepsilon^{(1)}_{r,12} \\ \varepsilon^{(1)}_{r,2} & \varepsilon^{(1)}_{r,10} \end{pmatrix}. \tag{A.3}$$

Consider the following isometric transform matrix (i.e., the Haar matrix for the case J=3):

$$\mathcal{T}^{(3)} = \begin{pmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{A.4}$$

To find the high-frequency shocks, the counterpart of the scale-specific shocks, we run through each column of the matrix (A.3) and, for each column, we perform the following operations:

$$\begin{pmatrix} \varepsilon_8 \\ \varepsilon_7 \\ \varepsilon_6 \\ \varepsilon_5 \\ \varepsilon_4 \\ \varepsilon_3 \\ \varepsilon_2 \\ \varepsilon_1 \end{pmatrix} = \left(\mathcal{T}^{(3)}\right)^{-1} \begin{pmatrix} \varepsilon^{(3)}_{\pi,8} \\ \varepsilon^{(3)}_{r,8} \\ e^{(2)}_{v,8} \\ e^{(2)}_{v,4} \\ \varepsilon^{(1)}_{r,8} \\ \varepsilon^{(1)}_{r,6} \\ \varepsilon^{(1)}_{r,4} \\ \varepsilon^{(1)}_{r,2} \end{pmatrix} \tag{A.5}$$

and

$$\begin{pmatrix} \varepsilon_{16} \\ \varepsilon_{15} \\ \varepsilon_{14} \\ \varepsilon_{13} \\ \varepsilon_{12} \\ \varepsilon_{11} \\ \varepsilon_{10} \\ \varepsilon_9 \end{pmatrix} = \left(\mathcal{T}^{(3)}\right)^{-1} \begin{pmatrix} \varepsilon^{(3)}_{\pi,16} \\ \varepsilon^{(3)}_{r,16} \\ u^{(2)}_{16} \\ e^{(2)}_{v,12} \\ \varepsilon^{(1)}_{r,16} \\ \varepsilon^{(1)}_{r,14} \\ \varepsilon^{(1)}_{r,12} \\ \varepsilon^{(1)}_{r,10} \end{pmatrix}. \tag{A.6}$$

Now, since the matrix in Eq. (A.4) is invertible, we can rewrite the shocks $\varepsilon_t^{(j)}$ as weighted averages of our newly-defined one-period shocks. The coefficient $\delta_i^{(j)}$ in Eq. (A.2) will be of the form $\pm 1/\sqrt{2}$ raised to some power. We can also define the scale-specific shocks within the grid starting from $\varepsilon_1, \ldots, \varepsilon_{16}$ and using Eq. (A.2). Finally, the square summability condition, i.e.,

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

is satisfied since the coefficients $\psi^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor}$ are of the form $\rho_j^{\left\lfloor \frac{h}{2^j} \right\rfloor}$ and the coefficients δs dampen their effect further.

A.2 Proof of Proposition 2

We begin with some matrix manipulations analogous to what was done to prove Proposition I. Consider the following component (or detail) dynamics for $j = j^*$, where $j^* \in \{1, \ldots, J\}$:

$$y^{(j)}_{t+2^j} = \beta_j x^{(j)}_t \tag{A.7}$$

$$x^{(j)}_{t+2^j} = \rho_j x^{(j)}_t + \sigma_j \epsilon_{t+2^j} \tag{A.8}$$

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

$$y_t^{(j)} = 0, \tag{A.9}$$

$$x_t^{(j)} = 0. \tag{A.10}$$

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

$$\begin{pmatrix} \pi_8^{(3)} & \pi_{16}^{(3)} \\ x_8^{(3)} & x_{16}^{(3)} \\ x_8^{(2)} & x_{16}^{(2)} \\ x_4^{(2)} & x_{12}^{(2)} \\ x_8^{(1)} & x_{16}^{(1)} \\ x_6^{(1)} & x_{14}^{(1)} \\ x_4^{(1)} & x_{12}^{(1)} \\ x_2^{(1)} & x_{10}^{(1)} \end{pmatrix} \tag{A.11}$$

and, analogously, for the details of y. To reconstruct the time series x_t, we run through each column of the matrix (A.11) and, for each column, we perform the following operation:

$$X_8^{(3)} = \begin{pmatrix} x_8 \\ x_7 \\ x_6 \\ x_5 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \end{pmatrix} = \left(\mathcal{T}^{(3)}\right)^{-1} \begin{pmatrix} \pi_8^{(3)} \\ x_8^{(3)} \\ x_8^{(2)} \\ x_4^{(2)} \\ x_8^{(1)} \\ x_6^{(1)} \\ x_4^{(1)} \\ x_2^{(1)} \end{pmatrix} \tag{A.12}$$

and

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

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

$$\left(\mathcal{T}^{(3)}\right)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{A.14}$$

Using the dynamics of the state (A.8), (A.12) and (A.13), we obtain

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

and

$$X_{8}^{(3)} = \begin{pmatrix} x_{8} = x_{8}^{(2)}/2 \\ x_{7} = x_{8}^{(2)}/2 \\ x_{6} = -x_{8}^{(2)}/2 \\ x_{5} = -x_{8}^{(2)}/2 \\ x_{4} = x_{4}^{(2)}/2 \\ x_{3} = x_{4}^{(2)}/2 \\ x_{2} = -x_{4}^{(2)}/2 \\ x_{1} = -x_{4}^{(2)}/2 \end{pmatrix}. \tag{A.16}$$

We now turn to forward/backward aggregation. 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 (A.7) and (A.9) together with (A.15)

and (A.16), 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 (A.15) and (A.16), and the equivalent ones for y together with (A.7) and (A.8). The regression of $y_{t+1,t+2}$ on $x_{t-1,t}$ yields (based on a fundamental block of four elements):

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

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

B Further results

B.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:

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

From (A.15) and (A.16), it is easy to see that, for $j^* = 2$:

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

For a more general j^*, i.e., if the process for x_t is given by (A.8) and (A.10), then

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

This result clarifies the relation between scale-wise persistence (ρ_{j^*}) and persistence in calendar time ($\widetilde{\rho}$). If $\rho_{j^*} < 1 - \frac{4}{2^{j^*}+1}$, then $\widetilde{\rho} > \rho_{j^*}$ for all j^*. However, as j^* grows large, $\widetilde{\rho}$ approximates 1. In other words, the largest the driving scale, the largest the calendar-time correlation *irrespective* of the actual scale-wise correlation.

B.2 Contemporaneous aggregation

We now run the contemporaneous regression

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

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

$$y_{t+1,t+h} = \beta(-x_{12}^{(2)} + x_{8}^{(2)})/2$$
$$x_{t+1,t+h} = (-x_{16}^{(2)} + x_{12}^{(2)})/2$$

By taking covariances we obtain

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

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

$$R^2 = \frac{\tilde{\beta}^2 \operatorname{Var}\left(-x_{12}^{(2)} + x_{8}^{(2)}\right)}{\beta^2 \operatorname{Var}\left(-x_{12}^{(2)} + x_{8}^{(2)}\right)} = \left(\frac{1-\rho_j}{2}\right)^2.$$

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

C Data

The empirical analysis in Sections 4, is conducted using annual data on consumption and stock returns from 1930 to 2014, 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. Our measure of consumption volatility is based on modeling consumption growth as following an AR(1) with an error variance evolving as an heterogeneous ARCH model (see Muller, Dacorogna, Dave, Olsen, Pictet, and von Weizsacker (1997)). The HARCH dynamics accommodates numerous heterogeneous information arrival processes, see, e.g., Andersen and Bollerslev (1997). Similar results are obtained by modeling consumption growth as following an AR(1)-GARCH(1,1), as in Bansal, Khatchatrian, and Yaron (2005).

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 annual return series is constructed from monthly data under the assumption of reinvestment at a zero-rate. The nominal short-term rate ($R_{f,t+1}$) is the yield on the 1-year Treasury bill.

The h-horizon continuously-compounded excess market return is calculated as $r_{t+1,t+h} = r_{t+1}^e + \ldots + r_{t+h}^e$, where $r_{t+j}^e = \ln(R_{t+j}) - \ln(R_{f,t+j})$ is the 1-year excess logarithmic market return between

dates $t+j-1$ and $t+j$, R_{t+j} is a simple gross return, and $R_{f,t+j}$ is a gross risk-free rate (3-month Treasury bill).

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

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

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

The measure of economic policy uncertainty (EPU) is based on Baker, Bloom, and Davis (2013).

D Rescaled *t*-statistics

Panels A1 and A2 in tables 1, 2 and 3 report, in curly brackets, the rescaled t-statistic recommended by Valkanov (2003) for the hypothesis that the regression coefficient is zero.

Under Valkanov's assumptions, when the forecasting horizon is a nontrivial fraction of the sample size, the t-statistic to test whether the predictive variable is statistically different from zero diverges at rate $\sqrt{T}$. Thus, to address this potential inferential problem in the context of a classical data generating process (different from the one we advocate), we compute the rescaled $t/\sqrt{T}$ statistic (where T is the sample size), recommended by Valkanov (2003).

Valkanov (2003) shows that the rescaled $t/\sqrt{T}$ statistic has a well-defined limiting distribution. This distribution is, however, nonstandard and depends on two nuisance parameters, δ and c. The parameter δ measures the covariance between innovations in the variable to be forecast and innovations in some forecasting variable, call it x_t. The parameter c measures deviations from unity in the highest autoregressive root for x_t, in a decreasing (at rate T) neighborhood of 1. Both of these parameters can be consistently estimated using the methodology described in Valkanov (2003). With these estimates in hand, the rescaled t-statistic, $t/\sqrt{T}$, can be compared against critical values computed as in Valkanov (2003).