Journal of Econometrics 208 (2019) 120–140

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The scale of predictability☆

F.M. Bandi[a,b,*], B. Perron[c], A. Tamoni[d], C. Tebaldi[e]

[a] Johns Hopkins University, United States
[b] Edhec-Risk Institute, France
[c] Université de Montrèal, CIRANO and CIREQ, Canada
[d] London School of Economics, United Kingdom
[e] Department of Finance, Universitá Bocconi, Baffi-Carefin and IGIER, Italy

ARTICLE INFO

Article history:
Available online 10 October 2018

JEL classification:
C22
E32
E44
G12
G17

Keywords:
Predictability
Frequency
Aggregation
Risk-return trade-off

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 modeling framework in which predictability is specified as a property of components of both excess market returns and economic uncertainty. We dub this property *scale-specific predictability*. We show that classical predictive systems imply restricted forms of scale-specific predictability. We conclude that for certain predictors, like economic uncertainty, the restrictions imposed by classical predictive systems may be excessively strong.

© 2018 Elsevier B.V. All rights reserved.

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.

☆ We are grateful to the Editor, Zhengjun Zhang, two anonymous reviewers, two discussants (Yanqin Fan and Loriano Mancini), Lars Peter Hansen, Nour Meddahi, Anthony Neuberger, Fulvio Ortu and Andrew Patton for their helpful comments. We thank seminar participants at the 2016 European Meetings of the Econometric Society in Geneva, the 2015 Australian Conference for Economists in Brisbane, the 2014 NBER-NSF Time Series Conference at the FRB of St. Louis, the 2014 European Finance Association Meetings in Lugano, the 2014 Workshop "Measuring and Modeling Financial Risk with High-Frequency Data" in Florence, 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 in Boston, the 2013 CIREQ Econometrics Conference on Time Series and Financial Econometrics in Montrèal, the Ninth CSEF-IGIER Symposium on Economics and Institutions in Capri, the 7th International Conference on Computational and Financial Econometrics in London, 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.

* Correspondence to: Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, United States.
E-mail addresses: fbandi1@jhu.edu (F.M. Bandi), benoit.perron@umontreal.ca (B. Perron), a.g.tamoni@lse.ac.uk (A. Tamoni), claudio.tebaldi@unibocconi.it (C. Tebaldi).

https://doi.org/10.1016/j.jeconom.2018.09.008
0304-4076/© 2018 Elsevier B.V. All rights reserved.

Existing work has highlighted the empirical usefulness of aggregating *both* the regressand (excess market returns) and the predictor. Specifically, Bandi and Perron (2008) have suggested running adapted (to time t information) regressions of *forward* aggregated excess market returns (i.e., long-run *future* excess market 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 that paper.[1]

We make three 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) or economic policy uncertainty (EPU, henceforth; Baker et al., 2016), is *hump-shaped*. The forward/backward regressions are conducted in this paper over horizons of aggregation reaching 20 years, thereby doubling the 10-year horizon reported 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 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 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 its third contribution, the paper introduces a data generating process in which excess returns and uncertainty are interpreted as linear aggregates of components operating over different frequencies or *scales*. Predictability is not modeled directly on the raw series. Rather, it is modeled as a property of individual (excess) return and uncertainty components. We dub this property *scale-specific predictability*. Theoretically, we show that, should components with cycles of suitable lengths be linked by a predictability relation, then two-way (forward/backward) aggregation would yield hump-shaped patterns in estimated slopes and R^2s. Empirically, after filtering excess returns and uncertainty components, we find predictability between components with cycles between 8 and 16 years. In agreement with theory, this type of scale-specific predictability should yield, upon two-way aggregation, a hump-shaped pattern with a peak at 16 years. This outcome is, as discussed above, consistent with data.

We show that the notion of scale-specific predictability is general in the sense that it would also apply to classical predictive systems. However, in the case of classical predictive systems, scale-specific predictability would be tightly parametrized across components and, therefore, frequencies. Our proposed approach can, therefore, be viewed as freeing up the nature of predictability across components and, as a result, for the raw series.

We conclude with an asset pricing model with Epstein–Zin recursive preferences which begins with component-wise structures for consumption growth, dividend growth and their conditional variance processes. Consistent with our empirical findings, the model yields scale-specific risk-return trade-offs, i.e., linear dependencies between future values of specific components of the (excess) market return process and past values of the *same* components of the variance process, as a feature of its equilibrium. While the model provides an economic channel which justifies our empirical results regarding the existence of scale-wise dependencies between excess returns and uncertainty, we view the paper's contribution as being methodological and, as such, broadly applicable to predictors other than uncertainty. In our final remarks in Section 9, we comment further on this issue.

The evaluation of low-frequency contributions to economic and financial time series has a long history, one which we cannot 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.

Beveridge and Nelson (1981) popularized time series decompositions into stochastic trends and transitory components. We too operate with component-wise decompositions in which the components (more than two, in our framework) feature different levels of (calendar-time) persistence. 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) may be associated with significant loss of information. Comin and Gertler (2006) decompose a series into a "high-frequency" component with cycles between 2 and 32 quarters and a "medium-frequency" component with cycles between 32 and 200 quarters. As emphasized, we focus on a number of components larger than two, thereby capturing cycles of different length (32 quarters, or 8 years, being the upper bound of one of these cycles and, of course, the lower bound of the subsequent one). News arrivals at various frequencies also characterize the multifractal regime switching approach of Calvet and Fisher (2001, 2007). As we will show below, modeling (through Wold representations, c.f. Section 3) and identification (through multiresolution filters, c.f. Section 5) are conducted differently in the present paper. As in Hansen and Scheinkman (2009), we employ operators to extract low-frequency information (c.f. Section 5). In our case, it is the low-frequency information embedded in the components. Finally, we show that essential frequency-wise information in the extracted components can be summarized by a finite number of non-overlapping, "fundamental" points, the result of an econometric process called "decimation"(c.f., again, Section 5). 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, these "fundamental" points are scale-specific and their locations correspond to the support of uncorrelated innovations over a specific scale. As such, they are particularly useful to formalize our notion of frequency-specific, or scale-specific, predictability, as shown in Section 6 and in Section 7.

[1] The long-run predictability of past variance has been documented to be robust to the use of alternative variance notions (Tamoni, 2011, employs consumption variance) and to the assumed dynamics of the variance process (Sizova, 2013, imposes long memory in variance). Among other stylized facts regarding stock returns, such predictability has also been justified in the context of an asset pricing model with loss aversion (Bonomo et al., 2015).

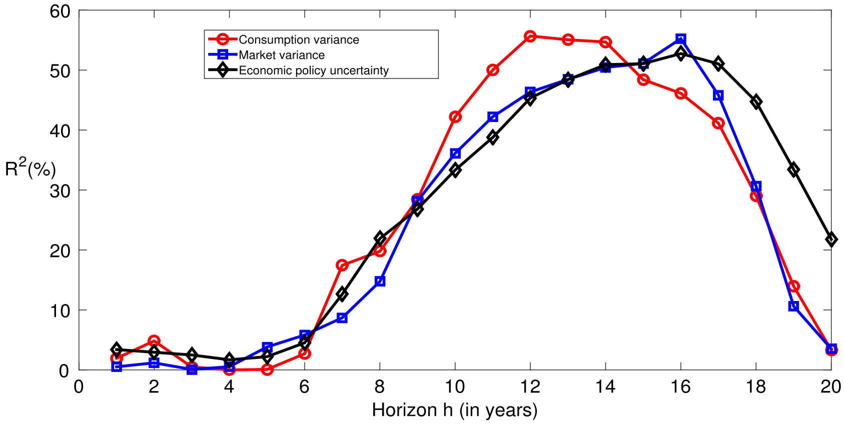

Fig. 1. R^2 values obtained by regressing forward-aggregated excess market returns on backward-aggregated market variance (blue line, with squares), consumption variance (red line, with circles), and (squared) economic policy uncertainty (black line, with diamonds) for different levels of aggregation (on the horizontal axis).

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) and Kelly and Pruitt (2013), interest rate variables, see e.g. Fama and Schwert (1977) and Fama and French (1989), and macroeconomic variables, see e.g. Lettau and Ludvigson (2001), Menzly et al. (2004), Nelson (1976) and 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).

We contribute to this extensive literature by emphasizing that traditional predictive systems may be viewed as restricted versions of the data generating process we propose. We free up *implicit* (given standard modeling approaches) links across frequencies and provide a parsimonious framework to capture predictive relations at different frequencies.

Before continuing, we stress that the paper has an Online Supplement containing extensive simulations, supporting proofs for the pricing model in Section 8, and additional results. In what follows, when we refer to the simulations, it should always be understood that they can be found in the Online Supplement. Should other results (mentioned in the main text) also be in the Online Supplement, we will state it explicitly.

2. Low-frequency humps

The empirical analysis in this paper is based on yearly data from 1930 to 2014. Appendix B describes the data and the construction of the variables.

We begin by running forward/backward regressions of long-run future excess market returns on long-run past *market* variance:

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

where $r_{t+1,t+h}$ and $v_{t-h+1,t}$ are non-overlapping sums of logarithmic excess market returns and return variances over an horizon of length h years.

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 estimated regression coefficients, adjusted R^2 statistics (in square brackets) and heteroskedasticity and autocorrelation-consistent t-statistics for the hypothesis that the regression coefficients are zero (in parentheses). The table also reports, in curly brackets, the rescaled t-statistics recommended by Valkanov (2003).[2]

In Table 1-Panel A1, we report horizons of aggregations up to 10 years: *future* excess market returns are correlated with *past* market variance. Dependence increases with the horizon, and is strong in the long run, with R^2 values between 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 (c.f., Fig. 1). 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.[3]

2 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 Eqs. (2) and (3)), regressors near unity, and aggregation of the regressand, of the regressor, or both.

3 By generating stochastic trends, forward/backward aggregation could lead to spurious (in the sense of Granger and Newbold, 1974, and Phillips, 1986) predictability. If the reported predictability were induced mechanically by aggregation, however, *contemporaneous* (i.e., forward/forward) aggregation

Table 1
Market variance. **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. For each regression, the table reports OLS estimates, Newey–West t-statistics with 2(horizon − 1) lags (in parentheses), the $t/\sqrt{T}$ tests suggested by Valkanov (2003) (in curly brackets) and R^2s (in square brackets). Ninety percent confidence intervals for the true R^2s are also reported in brackets below the R^2 estimates. Significance at the 5%, 2.5%, and 1% level of the $t/\sqrt{T}$ test using Valkanov's (2003) critical values is denoted 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, the table reports OLS estimates, t-statistics in parentheses and R^2 values in square brackets. The sample is annual and spans the period 1930–2014.

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

	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]
[5th, 95th]	[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: $r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+1,t+h}$

	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]
[5th, 95th]	[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: $r^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + u^{(j)}_{k2^j+2^j}$

	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]

Replacing market variance with *consumption* variance and EPU does not modify the previous results in any meaningful way (see Tables 2 and 3). We will therefore not comment on these measures separately.

We argue that classical predictive systems would find it hard to replicate the observed hump-shaped behavior. This is easy to see in theory. A traditional predictive system (on demeaned variables) would write:

$$r_{t+1} = \beta v_t + u_{t+1}, \tag{2}$$

$$v_{t+1} = \rho v_t + e_{t+1}, \tag{3}$$

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

When aggregating r_{t+1} forward and v_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 $h \to \infty$.[4] Similarly, the R^2 would also go to zero with the horizon of aggregation. Fig. 1 shows, instead, that the R^2 increases steeply to about 55% before decreasing equally sharply.[5]

Leaving theoretical considerations aside, in the simulations we conduct an experiment similar, in spirit, to that in Boudoukh et al. (2008) and show that classical predictive systems are, in any finite sample, unlikely to generate the tent-shaped dynamics detected in the data. Using data-driven parameter values for Eq. (2) and Eq. (3), the experiment finds that the percentage of simulated paths delivering *both* hump-shaped R^2 values and hump-shaped slope values, as well as R^2s in excess of 50%, is about 1.15.

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. The corresponding tables are not reported for conciseness but can be provided by the authors upon request. Simulations confirm the statement. In addition, spurious behavior would prevent a tent-shaped pattern from arising in the slopes, t-statistics and R^2s from predictive regressions on the aggregated series because it would likely lead to upward trending dynamics. We will later show that tent-shaped patterns are, instead, a natural by-product of a suitable component-based data generating process.

4 The reported "slope" (i.e., $\beta\rho^h$) 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 the one associated with the true conditional mean of forward-aggregated regressands onto backward-aggregated regressors. Such a 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 regarding this point.

5 As shown by Sizova (2013), a "large" ρ, captured by long memory in her framework, would help for 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 difficult to capture the hump-shaped dynamics illustrated above and further discussed below.

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

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

	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]
[5th, 95th]	[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: $r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+1,t+h}$

	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]
[5th, 95th]	[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: $r^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + u^{(j)}_{k2^j+2^j}$

	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]

In the following sections, we express excess market returns and economic uncertainty as linear combinations of $J > 1$ uncorrelated, mean-zero components $r^{(j)}$ and $v^{(j)}$ with $1 \le j \le J$. Each component will be shown to operate over a specific frequency. We will discuss how a predictive system analogous to Eqs. (2) and (3), but applied to specific components (i.e., a *scale-specific predictive system* as in Eqs. (12) and (13) of Section 4), may yield the reported hump-shaped pattern(s) upon two-way aggregation of the raw series. Proposition I in Section 6 will then link scale-specific predictability to two-way (forward/backward) aggregation.

We will view scale-wise predictability as a "spectral" feature of the series. Predictability upon two-way aggregation will, instead, be interpreted as 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.

3. Scale-specific components

In the next subsection, we discuss how to represent a covariance-stationary process in terms of orthogonal, scale-specific components. Importantly, since the procedure is applicable to any process for which a classical Wold representation applies, it can be regarded as general. A formalization based on Hilbert space theory of our proposed representation is given by Ortu et al. (2018) in a recent paper.

3.1. A scale-wise representation of a covariance-stationary scalar process

Let $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ be a covariance-stationary scalar process defined onto the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$. For simplicity, but without loss of generality, assume the process has a zero mean.[6]

The Wold representation of $\mathbf{x}$ states that there exists a unit variance, zero mean white noise process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that, for any t in $\mathbb{Z}$,

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

6 Adding a mean to the process would, of course, simply add a constant to its Wold representation.

Table 3
Economic policy uncertainty (Baker et al., 2016). **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) EPU. For each regression, the table reports OLS estimates, Newey–West t-statistics with 2(horizon − 1) lags (in parentheses), the $t/\sqrt{T}$ tests suggested by Valkanov (2003) (in curly brackets) and R^2s (in square brackets). Ninety percent confidence intervals for the true R^2s are reported in brackets below the R^2 estimates. Significance at the 5%, 2.5%, and 1% level of the $t/\sqrt{T}$ test using Valkanov's (2003) critical values is denoted 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, the table reports OLS estimates, t-statistics in parentheses and R^2 values in square brackets. The sample is annual and spans the period 1930–2014.

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

	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]
[5th, 95th]	[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: $r_{t+1,t+h} = \alpha_h + \beta_h v_{t-h+1,t} + \epsilon_{t+1,t+h}$

	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]
[5th, 95th]	[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: $r^{(j)}_{k2^j+2^j} = \beta_j v^{(j)}_{k2^j} + u^{(j)}_{k2^j+2^j}$

	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]

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 $\alpha_k = \mathbb{E}(x_t\varepsilon_{t-k})$.

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

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

We observe that $\varepsilon^{(1)}_t$ has a 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 $\boldsymbol{\varepsilon}^{(j)} = \left\{\varepsilon^{(j)}_t\right\}_{t\in\mathbb{Z}}$ with

$$\varepsilon^{(j)}_t = \frac{1}{\sqrt{2^j}}\left(\sum_{i=0}^{2^{j-1}-1}\varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1}\varepsilon_{t-2^{j-1}-i}\right). \tag{5}$$

The structure of the shocks $\boldsymbol{\varepsilon}^{(j)}$ is that of a Discrete Haar Transform (DHT). Once a scale level j is set, one may consider the sub-series of $\boldsymbol{\varepsilon}^{(j)}$ defined on the support $S^{(j)}_t = \{t - k2^j : k \in \mathbb{Z}\}$. The process $\boldsymbol{\varepsilon}^{(j)}$ is a moving average of order $2^j - 1$ with respect to the Wold innovations of $\mathbf{x}$. In calendar time, there is correlation between the variables $\varepsilon^{(j)}_{t-k2^j}$ and $\varepsilon^{(j)}_{\tau-k2^j}$ when $|t-\tau| \le 2^j - 1$. However, each sub-series $\left\{\varepsilon^{(j)}_{t-k2^j}\right\}_{k\in\mathbb{Z}}$ is a unit variance, zero mean white noise process on the support $S^{(j)}_t$.

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[7]

$$\mathcal{W}_t^{(j)} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad \sum_{k=0}^{+\infty} b_k^{(j)2} < +\infty \right\},$$

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

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 representation of $\mathbf{x}$:

$$x_t = \sum_{j=1}^{+\infty} x_t^{(j)}, \tag{6}$$

where each $x_t^{(j)}$ is the projection of x_t on the subspace $\mathcal{W}_t^{(j)}$, with $j \in \mathbb{N}$, and the equality is, again, 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 $\left\{\psi_k^{(j)}\right\}_{k\in\mathbb{N}_0}$.

Each coefficient $\psi_k^{(j)}$ is obtained by projecting $\mathbf{x}$ on the linear subspace generated by the (scale-specific) innovations $\varepsilon_{t-k2^j}^{(j)}$:

$$\psi_k^{(j)} = \mathbb{E}\left[x_t \varepsilon_{t-k2^j}^{(j)}\right]. \tag{7}$$

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

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \psi_k^{(j)} \varepsilon_{t-k2^j}^{(j)}. \tag{8}$$

Eq. (8) is a (type of) Wold representation describing 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.

It is now interesting to discuss the relation between the coefficients $\psi_k^{(j)}$ of the *extended* Wold representation and the coefficients α_k of the *classical* Wold representation. Exploiting Eq. (7) after expressing $\varepsilon_t^{(j)}$ as a finite linear combination of variables ε_t as in Eq. (5) and using Eq. (4), we obtain

$$\psi_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right). \tag{9}$$

Just like the shocks $\boldsymbol{\varepsilon}^{(j)}$ are DHTs of the high-frequency shocks in the classical Wold (see Eq. (5)), the coefficients in Eq. (9) are DHTs of the coefficients in the classical Wold. Because the Wold coefficients α_k are unique, it follows that the coefficients $\psi_k^{(j)}$ are also unique, and time-invariant, functions of the coefficients α_k, given the Haar structure.

3.2. The multivariate case

Consider the zero mean, covariance stationary, bi-variate process $\boldsymbol{X} = \{(y_t, x_t)^\intercal\}_{t\in\mathbb{Z}}$. Its Wold representation is, again,

$$\boldsymbol{X}_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \sum_{k=0}^{+\infty} \boldsymbol{\alpha}_k \boldsymbol{\varepsilon}_{t-k},$$

where $\boldsymbol{\varepsilon} = \{(u_t, e_t)^\intercal\}_{t\in\mathbb{Z}}$ is a bi-variate vector of (possibly) cross-correlated white noise shocks and the Wold coefficients $\boldsymbol{\alpha}_k$ are, for all $k \in \mathbb{N}_0$, 2×2 matrices.

By analogy with the scalar case, we can now write

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \boldsymbol{\Psi}_k^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)}. \tag{10}$$

7 The orthogonal decomposition into subspaces has been formalized by Ortu et al. (2018) in their Theorem 1.

8 Theorem 2 in Ortu et al. (2018) provides a formal justification.

For any $j \in \mathbb{N}$, the 2×2 matrices $\boldsymbol{\Psi}_k^{(j)}$ are unique DHTs of the original Wold coefficients:

$$\boldsymbol{\Psi}_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \boldsymbol{\alpha}_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \boldsymbol{\alpha}_{k2^j+2^{j-1}+i} \right).$$

Similarly,

$$\boldsymbol{\varepsilon}_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \boldsymbol{\varepsilon}_{t-i} - \sum_{i=0}^{2^{j-1}-1} \boldsymbol{\varepsilon}_{t-2^{j-1}-i} \right). \tag{11}$$

We emphasize that the components of the *extended* Wold in Eq. (10) are uncorrelated, for all t. As an example, take the first and the second scale and notice that the corresponding shocks are defined as follows:

$$\boldsymbol{\varepsilon}_t^{(1)} = \begin{pmatrix} \frac{u_t - u_{t-1}}{\sqrt{2}} \\ \frac{e_t - e_{t-1}}{\sqrt{2}} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{t-2}^{(1)} = \begin{pmatrix} \frac{u_{t-2} - u_{t-3}}{\sqrt{2}} \\ \frac{e_{t-2} - e_{t-3}}{\sqrt{2}} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{t-4}^{(1)} = \begin{pmatrix} \frac{u_{t-4} - u_{t-5}}{\sqrt{2}} \\ \frac{e_{t-4} - e_{t-5}}{\sqrt{2}} \end{pmatrix}, \ldots$$

and

$$\boldsymbol{\varepsilon}_t^{(2)} = \begin{pmatrix} \frac{(u_t + u_{t-1}) - (u_{t-2} + u_{t-3})}{\sqrt{4}} \\ \frac{(e_t + e_{t-1}) - (e_{t-2} + e_{t-3})}{\sqrt{4}} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{t-4}^{(2)} = \begin{pmatrix} \frac{(u_{t-4} + u_{t-5}) - (u_{t-6} + u_{t-7})}{\sqrt{4}} \\ \frac{(e_{t-4} + e_{t-5}) - (e_{t-6} + e_{t-7})}{\sqrt{4}} \end{pmatrix}, \ldots$$

It is immediate to verify that the first and the second components are orthogonal at all leads and lags on the supports $S_t^{(1)}$ and $S_t^{(2)}$ defined in Section 3.1. Formally, for generic components j and l, we have

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

where $\boldsymbol{X}^{(j)} = \left\{ \left(y_t^{(j)}, x_t^{(j)} \right)^\top \right\}_{t \in \mathbb{Z}}$.

We now turn to the focus of our analysis, i.e., scale-specific predictability.

4. Scale-specific predictive systems

Consider a regressand $\mathbf{y}$ and a predictor $\mathbf{x}$. Assume $\mathbf{y}$ and $\mathbf{x}$ are covariance-stationary and, therefore, admit an *extended* Wold representation as in Eq. (10). Assume, also, that for some scale j^*, the components of the processes $\mathbf{y}$ and $\mathbf{x}$ (written in scale-specific time) are such that

$$y_{k2^{j^*}+2^{j^*}}^{(j^*)} = \beta_{j^*} x_{k2^{j^*}}^{(j^*)} + u_{k2^{j^*}+2^{j^*}}^{(j^*)}, \tag{12}$$

$$x_{k2^{j^*}+2^{j^*}}^{(j^*)} = \rho_{j^*} x_{k2^{j^*}}^{(j^*)} + e_{k2^{j^*}+2^{j^*}}^{(j^*)}, \tag{13}$$

where $e_{k2^{j^*}+2^{j^*}}^{(j^*)}$ is a white noise process in scale time with a zero mean and a variance $\sigma_e^{(j^*)2}$ and $u_{k2^{j^*}+2^{j^*}}^{(j^*)}$ is a white noise (again, in scale time) forecast error with a zero mean and a variance $\sigma_u^{(j^*)2}$. The shocks $u_{k2^{j^*}+2^{j^*}}^{(j^*)}$ and $e_{k2^{j^*}+2^{j^*}}^{(j^*)}$ are possibly cross-correlated. Assume all other components $\{y_{k2^j+2^j}^{(j)}, x_{k2^j+2^j}^{(j)}\}$ with $j \neq j^*$ are mean-zero white noise processes, uncorrelated with each other.[9]

Eqs. (12)–(13) define a predictive system on individual layers of the bivariate process $\{(y_t, x_t)^\top\}_{t \in \mathbb{Z}}$ to be contrasted with the traditional system written on the raw series directly.

As mentioned in the Introduction, 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 being the outcome of noise reduction through (excess) 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. (12) and (13) account for both effects: the link between slow-moving components (for a large j^*) of $\mathbf{y}$ and $\mathbf{x}$ and, because predictability is defined on components rather than on noisier (i.e., "imperfect") raw series, the "imperfection" of traditional predictors (and regressands).

9 Needless to say, we are not excluding the possibility of multiple scale-wise predictive systems at different frequencies. Our emphasis on a single frequency is simply meant to illustrate the potential of a simple scale-specific predictive structure for yielding patterns found in the data.

4.1. Classical predictive systems and scale-specific predictability

Can a typical predictive system yield scale-specific predictability? The answer is positive. However, classical predictive systems impose tight restrictions on the nature of predictability across scales. To discuss this issue, consider a vector autoregressive process of order 1 (VAR(1)) for the bi-variate process $(y_t, x_t)^\intercal$, i.e.,

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = A \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ e_t \end{pmatrix}.$$

As in Section 2, the vector $(u_t, e_t)^\intercal = \boldsymbol{\varepsilon}_t$ is a vector of cross-correlated white noise shocks. The standard Wold representation yields

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \underbrace{I_2}_{\boldsymbol{\alpha}_0} \begin{pmatrix} u_t \\ e_t \end{pmatrix} + \underbrace{A}_{\boldsymbol{\alpha}_1} \begin{pmatrix} u_{t-1} \\ e_{t-1} \end{pmatrix} + \underbrace{A^2}_{\boldsymbol{\alpha}_2} \begin{pmatrix} u_{t-2} \\ e_{t-2} \end{pmatrix} + \cdots$$

with I_2 defining the 2×2 identity matrix. Hence, in terms of the notation in Section 3.2, we have $\boldsymbol{\alpha}_k = A^k$.

We can now obtain the same *extended* Wold representation as in Eq. (10) with the following coefficients $\boldsymbol{\Psi}_k^{(j)}$:

$$\boldsymbol{\Psi}_0^{(1)} = \frac{I_2 - A}{\sqrt{2}}$$

$$\boldsymbol{\Psi}_1^{(1)} = \frac{A^2 - A^3}{\sqrt{2}} = \frac{I_2 - A}{\sqrt{2}} A^2,$$

$$\ldots$$

$$\boldsymbol{\Psi}_k^{(1)} = \frac{A^{2k} - A^{2k+1}}{\sqrt{2}} = \frac{I_2 - A}{\sqrt{2}} A^{2k},$$

for $j = 1$. Similarly, for a generic $j = 1, 2, \ldots,$ and for $k = 0$:

$$\begin{aligned}
\boldsymbol{\Psi}_0^{(j)} &= \frac{\underbrace{I_2 + A + \cdots + A^{2^{(j-1)}-1}}_{2^{(j-1)} \text{ terms}} - \left(\underbrace{A^{2^{(j-1)}} + \cdots + A^{2^j-1}}_{2^{(j-1)} \text{ terms}} \right)}{\sqrt{2^j}} \\
&= \frac{\left(I_2 - A^{2^{(j-1)}}\right)(I_2 - A)^{-1} - \left(I_2 - A^{2^{(j-1)}}\right)(I_2 - A)^{-1} A^{2^{(j-1)}}}{\sqrt{2^j}} \\
&= \frac{\left(I_2 - A^{2^{(j-1)}}\right)^2 (I_2 - A)^{-1}}{\sqrt{2^j}}.
\end{aligned}$$

For $j = 1, 2, \ldots,$ and for $k > 0$:

$$\begin{aligned}
\boldsymbol{\Psi}_k^{(j)} &= \frac{\underbrace{I_2 + A + \cdots + A^{2^{(j-1)}-1}}_{2^{(j-1)} \text{ terms}} - \left(\underbrace{A^{2^{(j-1)}} + \cdots + A^{2^j-1}}_{2^{(j-1)} \text{ terms}} \right)}{\sqrt{2^j}} \times A^{k2^j} \\
&= \frac{\left(I_2 - A^{2^{(j-1)}}\right)^2 (I_2 - A)^{-1}}{\sqrt{2^j}} \times A^{k2^j}.
\end{aligned}$$

Let us now be explicit about the components. Since $\boldsymbol{\Psi}_k^{(j)} = \boldsymbol{\Psi}_0^{(j)} \times A^{k2^j}$, we obtain

$$\left(y_t^{(j)}, x_t^{(j)}\right)^\intercal = \sum_{k=0}^{+\infty} \boldsymbol{\Psi}_k^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)} = \widehat{\boldsymbol{\varepsilon}}_t^{(j)} + A^{2^j} \widehat{\boldsymbol{\varepsilon}}_{t-2^j}^{(j)} + A^{2 \times 2^j} \widehat{\boldsymbol{\varepsilon}}_{t-2\times 2^j}^{(j)} + \cdots,$$

where $\widehat{\boldsymbol{\varepsilon}}^{(j)} = \boldsymbol{\Psi}_0^{(j)} \boldsymbol{\varepsilon}^{(j)}$. Hence, the structure of the coefficients $\boldsymbol{\Psi}_k^{(j)}$ is that of a VAR(1) defined on the support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$. However, this implies

$$\begin{pmatrix} y_t^{(j)} \\ x_t^{(j)} \end{pmatrix} = A^{2^j} \begin{pmatrix} y_{t-2^j}^{(j)} \\ x_{t-2^j}^{(j)} \end{pmatrix} + \widehat{\boldsymbol{\varepsilon}}_t^{(j)},$$

which is, in our jargon, a *scale-wise predictive system*.

The above derivations lead to two observations. First, we can rewrite the VAR(1) as an infinite sum of VAR(1)s with time steps 2^j and autoregressive matrix given by A^{2^j}. We are, of course, interested in the case

$$A = \begin{pmatrix} 0 & \beta \\ 0 & \rho \end{pmatrix},$$

which yields

$$A^{2^j} = \begin{pmatrix} 0 & \beta \times \rho^{2^j-1} \\ 0 & \rho^{2^j} \end{pmatrix}.$$

Thus, a classical predictive system implies a tightly parametrized form of scale-specific predictability. For each scale j, the predictive slope should be $\beta \times \rho^{2^j-1}$. Our approach frees up this restriction and, therefore, allows for rich dependence structures in the raw series without dispensing with parsimony.

Second, the standard predictive system only allows for scale-specific predictability between identical scales of the regressand and the regressor, cross-predictability (i.e., predictability between two generic scales, say j and l) being excluded. Just like in the standard predictive system in this subsection, even in our *less* restricted framework, lack of cross-predictability is *not* an assumption. In light of the discussion in Section 3.2, the property readily rests on the Haar structure of the scale-specific shocks.

5. Filtering the scale-specific components

Because the components $x_t^{(j)}$ are needed to validate scale-specific predictability in the data (as we do in Section 7), their extraction – the topic of this section – is empirically important.

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

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

which effectively amounts to breaking the series down into a "transitory" component $\widehat{x}_t^{(1)}$ and a "persistent" component $\pi_t^{(1)}$. 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 another "transitory" component $\widehat{x}_t^{(2)}$ and an another "persistent" component $\pi_t^{(2)}$.

The procedure can, of course, be iterated yielding a general expression for the generic component $\widehat{x}_t^{(j)}$, 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

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

with $\pi_t^{(0)} \equiv x_t$.

In essence, for every t, $\{x_{t-i}\}_{i\in\mathbb{Z}}$ can be written as a collection of components $\widehat{x}_t^{(j)}$ with different calendar-time persistence along with a low-frequency trend $\pi_t^{(J)}$. Equivalently, it can be written as a telescopic sum

$$x_t = \sum_{j=1}^{J} \underbrace{\left\{\pi_t^{(j-1)} - \pi_t^{(j)}\right\}}_{\widehat{x}_t^{(j)}} + \pi_t^{(J)} = \pi_t^{(0)}, \tag{14}$$

in which the filtered components are naturally viewed as changes in information between scale 2^{j-1} and scale 2^j. The scales are dyadic[10] and, therefore, enlarge with j. The higher j, the lower the frequency. In particular, the innovations

[10] The term "dyadic" refers to the fact that the time horizon of the scales, i.e., 2^{j-1} to 2^j periods, increases like powers of 2.

$\widehat{x}_t^{(j)} = \pi_t^{(j-1)} - \pi_t^{(j)}$ become smoother, and more persistent in calendar time, as j increases. Eq. (14) is the filtered analogue of Eq. (6).

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

$$\pi_t^{(2)} = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}. \tag{15}$$

and

$$\begin{aligned}
\widehat{x}_t^{(2)} &= \frac{\pi_t^{(1)} - \pi_{t-2}^{(1)}}{2} = \frac{1}{2}\left(\frac{x_t + x_{t-1}}{2} - \frac{x_{t-2} + x_{t-3}}{2}\right) \\
\widehat{x}_t^{(1)} &= \frac{\pi_t^{(0)} - \pi_{t-1}^{(0)}}{2} = \left(\frac{x_t - x_{t-1}}{2}\right) \\
\widehat{x}_{t-2}^{(1)} &= \frac{\pi_{t-2}^{(0)} - \pi_{t-3}^{(0)}}{2} = \left(\frac{x_{t-2} - x_{t-3}}{2}\right).
\end{aligned} \tag{16}$$

Stacking now Eq. (15) on top of Eq. (16), we obtain

$$\begin{pmatrix} \pi_t^{(2)} \\ \widehat{x}_t^{(2)} \\ \widehat{x}_t^{(1)} \\ \widehat{x}_{t-2}^{(1)} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix}. \tag{17}$$

Denoting by $\mathcal{T}^{(2)}$ the (4×4) matrix in Eq. (17), we notice that $\mathcal{T}^{(2)}$ is orthogonal, that is $\Lambda^{(2)} \equiv \mathcal{T}^{(2)} \left(\mathcal{T}^{(2)}\right)^{\intercal}$ 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:

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix} = \left(\mathcal{T}^{(2)}\right)^{-1} \begin{pmatrix} \pi_t^{(2)} \\ \widehat{x}_t^{(2)} \\ \widehat{x}_t^{(1)} \\ \widehat{x}_{t-2}^{(1)} \end{pmatrix}. \tag{18}$$

For an extension of this procedure to any $J \geq 2$, and a recursive algorithm for the construction of the operator 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).[11]

As in Beveridge and Nelson (1981), who popularized 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.

5.1. 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. Equivalently, it is the process of sampling in scale time, the type of sampling used in Section 6 and in Section 7.

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 components, 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

[11] For any dilation index j, one can intepret the components $\widehat{x}_t^{(j)}$ as projections of the original series x_t onto the space spanned by a system of Haar basis with translation parameter $k \in \mathbb{Z}$ (see Mallat, 1989, and Dijkerman and Mazumdar, 1994).

useful to formalize our notion of frequency-specific predictability. Implicitly, they were used to define scale-wise predictive systems in Eq. (12) and in Eq. (13).

In multiresolution analysis[12] the Haar matrix is routinely used to *filter* $\left\{x_{t-k2^j}^{(j)}\right\}_{j=1,\ldots,J,\ k\in\mathbb{Z}}$ from $\{x_{t-i}\}_{i\in\mathbb{Z}}$. While we also filter the components using the Haar matrix, one methodological novelty of this work is to also operate in the opposite direction, i.e., to propose a data generating process which specifies the law of motion of the components $\left\{x_{t-k2^j}^{(j)}\right\}_{j=1,\ldots,J,\ k\in\mathbb{Z}}$ as in Eq. (13) (or as in Eqs. (12)–(13) in the bivariate case) and, only then, obtains each observation x_t as a linear combination of the components themselves (using Eq. (18)). This is, in fact, how we generate observations from a data generating process allowing for scale-specific predictability in the simulations.

Before turning to an empirical evaluation of Eqs. (12)–(13), we discuss the implications of scale-wise predictability for two-way aggregation.

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

We showed in Section 2 that forward–backward regressions uncover a strong risk-return trade-off beyond business-cycle frequencies. The logic for aggregating both the regressand (**y**, say) and the regressor (**x**, say) resides in the intuition according to which equilibrium implications of economic models may hold for specific, highly persistent components of the variables **y** and **x**, while being hidden by short-term noise. Aggregation provides a natural way to filter out the noise, generate a cleaner signal and extract these components. Using the assumed data generating process, we now formalize this logic.

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

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

whereas $\left\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\right\} = 0$ for $j \neq j^*$. We map scale-time (or decimated) 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, in Proposition I we dispense with the forecasting shocks $u_{k2^{j^*}+2^{j^*}}^{(j^*)}$. Predictability applies to a specific j^* component. All other components are set equal to zero.

Proposition I shows that predictability on the components 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.

The proposition also makes explicit the fact that the optimal level of (forward/backward) aggregation should coincide with the horizon over which frequency-specific predictability applies. More specifically, under the above assumptions, if predictability applies to a component 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.

Adding short-term or long-term shocks in the form of uncorrelated components $\left\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\right\}$, for $j < j^*$ or for $j > j^*$, or forecast errors different from zero, would solely lead to a blurring of the resulting relation (e.g., lower R^2 values) upon two-way aggregation. We add uncorrelated components $\left\{y_{k2^j}^{(j)}, x_{k2^j}^{(j)}\right\}$ for $j \neq j^*$ in the simulations.[13]

Next, we broaden the scope of classical predictability relations in the literature. We turn to regressions on the extracted components and illustrate the consistency of their findings with those obtained, in Section 2, from two-way aggregation. This consistency, which is an implication of Proposition I, is further confirmed by simulation.

7. Component-wise predictability

We focus on the component-wise relation between excess equity returns and market variance (Table 1). The case of consumption variance and EPU is completely analogous (Tables 2 and 3) and, as earlier, we will not comment on it independently.

[12] See, e.g., Mallat (1989). For applications to economic and financial time series, we refer to the comprehensive treatments in Ramsey (1999), Percival and Walden (2000), Gençay et al. (2001), and Crowley (2007).

[13] Future research will study the link between scale-wise predictability and predictability upon aggregation by working directly with the Wold components.

Table 4
Interpretation of the scale j in terms of time length for annual time series. Each scale corresponds to a frequency or, equivalently, to a time interval reported in the second column.

	Time (in years)
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

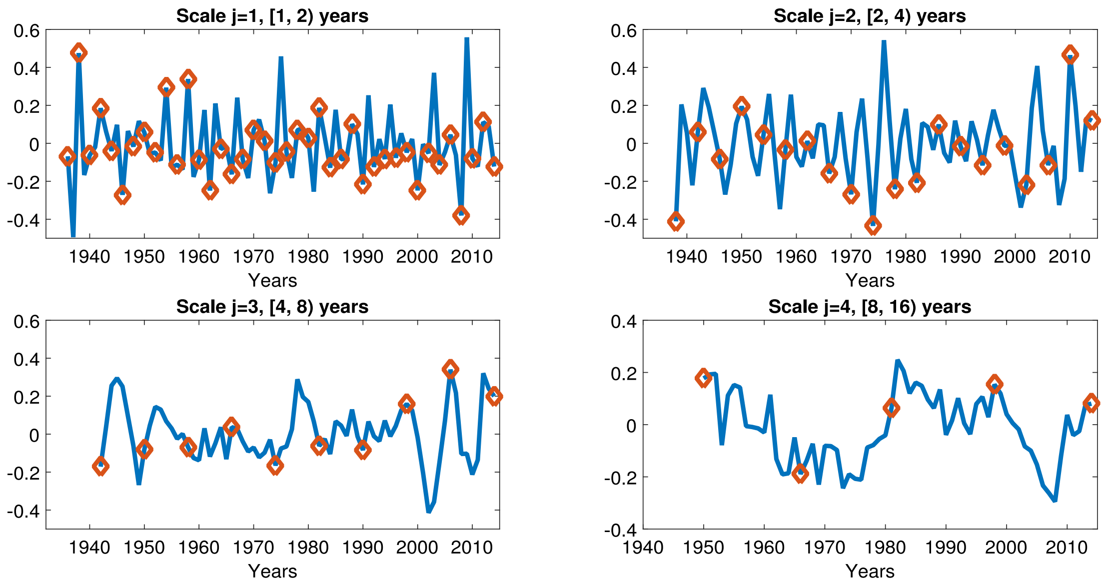

Fig. 2. Decomposition into components of excess market returns. The calendar-time observations are solid blue lines, the scale-time or "decimated" observations are red diamonds.

The filtered components are shown in Figs. 2 and 3. For an explicit interpretation of the jth scale in terms of yearly time spans, we refer to Table 4.

Table 5 presents pair-wise correlations between individual components, for both series. Virtually all correlations are small and statistically insignificant. This evidence is suggestive of the fact that the Haar-based filtering procedure described in Section 5 is effective in extracting the orthogonal Wold components.[14]

We run scale-wise predictive regressions as in Eqs. (12)–(13). The results are reported in Table 1-Panel B.[15] 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 scale-wise predictive regression is a considerable 58.3%. Consistent with Proposition I, the R^2 values upon two-way aggregation reach their peak (around 55%) at 16 years ($2^4 = 16$ years). Remarkably, the structure of the R^2s, before and after, is roughly tent-shaped (see Fig. 1).

The study of low-frequency relations is made difficult by the limited availability of observations over certain, long horizons. We, of course, do not believe that this difficulty detracts from the importance of inference at low frequency, provided such inference is conducted carefully. Importantly for the purposes of this paper, 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

14 The pair-wise correlations are obtained by using overlapping, calendar-time, or *redundant* data on the components rather than the non-overlapping, scale-time, or *decimated* counterparts described in Section 5.1. This is, of course, due to the need of having the same number of observations for each scale. 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 (1) relate scale-wise predictability to aggregation, a core aspect of our treatment, and to (2) yield components with dyadic cycles whose length is economically easily interpretable: because the length of the business cycle is 2 to 8 years, the 8 to 16-year horizon that we highlight in this section captures frequencies lower than the business cycle.

15 Given a scale j, we work with an effective sample size of $\lfloor T/2^j \rfloor$ observations, where $\lfloor . \rfloor$ denotes the floor function. In this empirical analysis, we set $J = 4$.

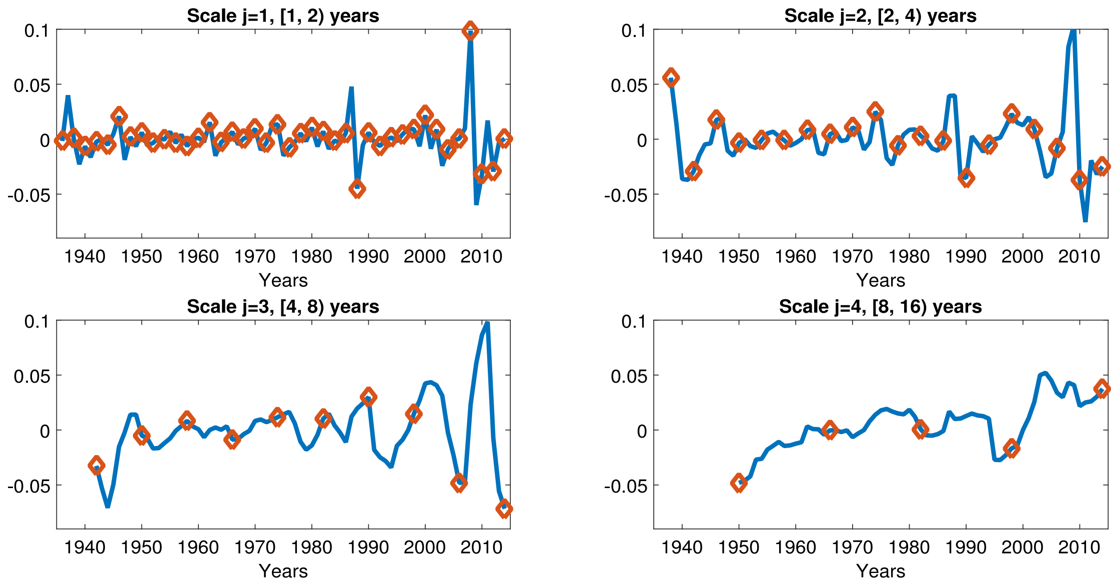

Fig. 3. Decomposition into components of market variance. The calendar-time observations are solid blue lines, the scale-time or "decimated" observations are red diamonds.

Table 5
Pairwise correlations. We report pairwise correlations between the components of excess market returns (Panel A), consumption variance (Panel B), market variance (Panel C), and (squared) EPU (Panel D). The pair-wise correlations are obtained by using *redundant* (i.e., calendar time) observations on the components rather than their decimated counterparts. Standard errors for the correlation between $x_t^{(j)}$ and $x_t^{(j')}$, with $j \neq j'$, are Newey–West with $2^{max(j,j')}$ lags.

	Panel A:				**Panel B:**			
	Market excess returns				Consumption variance			
Scale $j =$	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:				**Panel D:**			
	Market variance				Economic policy uncertainty			
Scale $j =$	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)

in R^2 values as a by-product of scale-wise predictability. Tent-shaped behavior, an implication of Proposition I, 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.

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 excess 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.[16] 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 component and future values of the same component. In other words, consistent with an autoregressive (of order 1) specification for the components, the $j = 4$ variance component has a positive autocorrelation with uncorrelated residuals.

16 The corresponding results are available upon request.

The R^2 of the scale-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 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 the Online Supplement (see Subsection C.1), the corresponding autocorrelation in calendar-time would naturally be higher.

Interestingly, 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. 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 et al. (2008)). In our framework, however, the autocorrelation ρ_{j*} is scale-specific. At the low frequencies over which we identify scale-wise forecastability, ρ_{j*} is estimated to be small.

In essence, we show 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 component appear to predict higher future values of the same uncertainty component and, consequently, higher future values of the corresponding component 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 excess returns, as shown formally in Proposition I and in the data. Extensive simulations from a data generating process allowing for scale-wise predictability provide further support.

8. Scale-wise risk-return trade-offs in equilibrium

We conclude with an asset pricing model which yields scale-specific risk-return trade-offs as a result of its equilibrium. While the model offers an economic channel which justifies our empirical findings, we view the paper's contribution as being methodological and, as such, applicable to predictors other than uncertainty.

Our assumed specification follows that of the long-run risk model of Bansal and Yaron (2004) duly modified to account for heterogeneity in persistence as in Ortu et al. (2013). We introduce component-wise representations for both consumption growth and dividend growth. However, we shut down long-run risk and set the expectation of both growth rates equal to zero. As in Bollerslev et al. (2009), who also make this assumption in their model specification, this is solely done for conciseness and in order to focus on the substantive core of our analysis, i.e., the role of scales and their equilibrium implications.

We begin with component-wise decompositions for log consumption growth, g_t, and log dividend growth, gd_t, taking the form:

$$g_t = \sum_{j=1}^{J} g_t^{(j)}, \quad gd_t = \sum_{j=1}^{J} gd_t^{(j)}.$$

We assume that each component of the consumption growth process, $g_t^{(j)}$, and of the dividend growth process, $gd_t^{(j)}$, is driven by its own scale-specific stochastic variance, $\sigma_t^{(j)2}$, i.e.,

$$g_{t+2^j}^{(j)} = \sigma_t^{(j)} e_{g,t+2^j}^{(j)} \quad \text{with } e_{g,t+2^j}^{(j)} \sim N(0,1), \tag{19}$$

$$gd_{t+2^j}^{(j)} = \varphi_d^{(j)} \sigma_t^{(j)} e_{gd,t+2^j}^{(j)} \quad \text{with } e_{gd,t+2^j}^{(j)} \sim N(0,1), \tag{20}$$

where the shocks $e_{g,t+2^j}^{(j)}$ and $e_{gd,t+2^{j'}}^{(j')}$ are correlated for $j = j'$ (with scale-specific correlation denoted by $\rho_e^{(j)}$) and are uncorrelated otherwise.[17] For parsimony, as in Bansal and Yaron (2004), the variance of consumption growth and the variance of dividend growth are driven by a common time-varying component (the terms $\varphi_d^{(j)}$ are, therefore, just scaling factors).

To complete the dynamics of the model, we assume that each of the stochastic variance components $\left\{\sigma_t^{(j)2}\right\}_{j=1}^{J}$ follows an autoregressive process, i.e.,

$$\sigma_{t+2^j}^{(j)2} = \rho_j \sigma_t^{(j)2} + \varepsilon_{t+2^j}^{(j)}, \quad \text{with } \varepsilon_{t+2^j}^{(j)} \sim N\left(0, \sigma^{(j)2}\right), \tag{21}$$

where the innovations $\varepsilon_{t+2^j}^{(j)}$ are uncorrelated with all other shocks and, of course, across scales.[18]

[17] If, as in Section 3, we were to assume that the scale-j shocks are DHTs of white noise shocks, then $e_{g,t}^{(j)}$ and $e_{g,s}^{(j)}$ would be uncorrelated for all t and s so that $|t - s| > 2^j - 1$. The same would, of course, apply to $e_{gd,t}^{(j)}$ and $e_{gd,s}^{(j)}$. In addition, assuming that the scale-j shocks are DHTs of white noise shocks implies that the components of the variance of consumption growth coincide with the variances of the components of the consumption growth process (see the Online Supplement for a proof). Since we use the former (i.e., the components of the variance of the consumption growth process) in Section 7 and we directly model the latter (i.e., the variances of the components of the consumption growth process) in this section, this equivalence is interesting.

[18] The zero-mean property of the variance components is coherent with the construction in Section 3.1. Fig. 3 contains a graphical representation based on data.

We operate in a pure exchange economy with a representative agent endowed with Epstein–Zin recursive preferences. The representative agent's Euler equation is well known:

$$\mathbb{E}_t \left[e^{m_{t+1}+r^i_{t+1}} \right] = 1, \tag{22}$$

where m_{t+1} is the log stochastic discount factor given by

$$m_{t+1} = \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r^a_{t+1} \ , \tag{23}$$

r^a_{t+1} is the log return on an asset yielding a dividend equal to aggregate consumption and r^i_{t+1} is the log return on any asset i. The parameter β is a discount factor. The preference parameter ψ measures the intertemporal elasticity of substitution (IES), γ measures risk aversion and $\theta \equiv (1-\gamma)/(1-1/\psi)$.

We focus on the market risk premium.[19] Recall, first, that the standard Campbell and Shiller (1988) log-linear approximation gives

$$r^a_{t+1} = \kappa_0 + \kappa_1 z^a_{t+1} - z^a_t + g_{t+1}, \tag{24}$$

$$r^m_{t+1} = \kappa_{0,m} + \kappa_{1,m} z^m_{t+1} - z^m_t + gd_{t+1}, \tag{25}$$

where z^a_t and z^m_t denote the log price–consumption ratio and the log price–dividend ratio, respectively.

Because of our decompositions of consumption growth and dividend growth into components with different levels of persistence, representing by $z^{a,(j)}_t$ and $z^{m,(j)}_t$ the components of the log price–consumption ratio and the log price–dividend ratio, respectively, it is natural to conjecture that there exists a *component-by-component* linear relation between the financial ratios and the state variables $\sigma^{(j)2}_t$, i.e.,

$$z^{a,(j)}_t = A_{0,j} + A_j \sigma^{(j)2}_t, \tag{26}$$

$$z^{m,(j)}_t = A^m_{0,j} + A^m_j \sigma^{(j)2}_t. \tag{27}$$

As long as A_j and A^m_j are not zero, these relations imply that the variation in the valuation ratios can be attributed to fluctuations in economic uncertainty.

The values of A_j and A^m_j in terms of the parameters of the model are obtained from the Euler condition in Eq. (22) after expressing the log stochastic discount factor and the returns in Eqs. (24) and (25) in terms of the state variables $\{\sigma^{(j)2}_t\}^J_{j=1}$ and the innovations $\{e^{(j)}_{g,t+2^j}\}^J_{j=1}$, $\{e^{(j)}_{gd,t+2^j}\}^J_{j=1}$ and $\{\varepsilon^{(j)}_{t+2^j}\}^J_{j=1}$. In particular, using the method of undetermined coefficients, one obtains the following J-column vectors of sensitivities:

$$\underline{A} = 0.5 \frac{\left(\theta - \frac{\theta}{\psi}\right)^2}{\theta} \left(I_J - \kappa_1 M\right)^{-1} \underline{1},$$

$$\underline{A}^m = 0.5 \left(I_J - \kappa_{1,m} M\right)^{-1} \left(H_m - (1-\gamma) \left(\frac{1}{\psi} - \gamma \right) \underline{1} \right),$$

where I_J is the J-identity matrix, $M \equiv -\mathrm{diag}\left(\rho_1, \ldots, \rho_J\right)$, $\underline{1}$ is a J-column vector of ones, $\lambda_g \equiv \left(\frac{\theta}{\psi} - \theta + 1\right) = \gamma$ and $H_m \equiv \left[\lambda_g^2 + \varphi_d^{(1)2} + 2\lambda_g \varphi^{(1)} \rho_e^{(1)}, \ldots, \lambda_g^2 + \varphi_d^{(J)2} + 2\lambda_g \varphi^{(J)} \rho_e^{(J)} \right]^\intercal$. In essence, $\underline{A}$ and $\underline{A}^m$ are J-column vectors with entries $A_1, \ldots, A_J$ and $A^m_1, \ldots, A^m_J$, respectively.

Next, one may show that the innovations to the return components are given by

$$r^{a,(j)}_{t+2^j} - \mathbb{E}_t[r^{a,(j)}_{t+2^j}] = \sigma^{(j)}_t e^{(j)}_{g,t+2^j} + \kappa_1 A_j \varepsilon^{(j)}_{t+2^j}, \tag{28}$$

$$r^{m,(j)}_{t+2^j} - \mathbb{E}_t[r^{m,(j)}_{t+2^j}] = \varphi^{(j)}_d \sigma^{(j)}_t e^{(j)}_{gd,t+2^j} + \kappa_{1,m} A^m_j \varepsilon^{(j)}_{t+2^j}. \tag{29}$$

Similarly, the innovations to the components of the stochastic discount factor are given by

$$m^{(j)}_{t+2^j} - \mathbb{E}_t[m^{(j)}_{t+2^j}] = -\lambda_g \sigma^{(j)}_t e^{(j)}_{g,t+2^j} - \underline{\lambda}^{(j)}_\epsilon \varepsilon^{(j)}_{t+2^j}, \tag{30}$$

with $\underline{\lambda}_\epsilon \equiv \kappa_1 (1-\theta) \underline{A}$. Given frequency-specific risk premia equal to

$$\mathbb{E}_t[r^{m,(j)}_{t+2^j} - r^{f,(j)}_{t+2^j}] + 0.5 \mathbb{V}_t(r^{m,(j)}_{t+2^j}) = -\mathbb{C}_t(m^{(j)}_{t+2^j}, r^{m,(j)}_{t+2^j}), \tag{31}$$

Eq. (29) and Eq. (30) lead to

$$\mathbb{E}_t[r^{m,(j)}_{t+2^j} - r^{f,(j)}_{t+2^j}] + 0.5 \mathbb{V}_t(r^{m,(j)}_{t+2^j}) = \gamma \varphi^{(j)}_d \sigma^{(j)2}_t \rho^{(j)}_e + \kappa_{1,m} [\underline{\lambda}^\intercal_\epsilon \mathbf{Q}]_j A^m_j, \tag{32}$$

where $\boldsymbol{\varepsilon}^\intercal_{t+1} \equiv [\varepsilon^{(1)}_{t+1}, \ldots, \varepsilon^{(J)}_{t+1}]$ and $\mathbf{Q} \equiv \mathbb{E}_t \left[\boldsymbol{\varepsilon}_{t+1} \boldsymbol{\varepsilon}^\intercal_{t+1} \right]$.

[19] Derivations and technical details are contained in the Online Supplement.

Eq. (32) represents a risk-return trade-off on the generic scale j obtained as the result of a simple equilibrium argument based on a classical change of measure (relying on Eq. (23)) and scale-wise decompositions for consumption growth and dividend growth. Sections 2 and 7 provide empirical content to Eq. (32).

9. Conclusions and further discussion

Economic relations may apply to individual low-frequency components of the series of interest and be hidden by transient effects at high frequencies. To capture this idea parsimoniously, this paper models market excess returns and their predictors (economic uncertainty, in our case) as aggregates of uncorrelated components operating over different frequencies and introduces a notion of *scale-specific predictability*, i.e., predictability on the components.

We conclude with two observations. First, the use of variance as a predictor of asset returns is appealing in our framework because the corresponding backward-aggregated measure does not lose its economic interpretation. Backward-aggregated variance has, in fact, a natural interpretation in terms of long-run past variance (under conditions). Having made this point, all alternative predictors, like the dividend-yield and other financial ratios, may also be employed. While their long-run past averages are not as easily interpretable, the role played by aggregation in the extraction of low-frequency information contained in the components and its translation into long-run properties of the data, as laid out in Proposition I, apply generally. To the extent that market return data and the dividend yield – for instance – contain relevant information about long-run cash-flow and discount rate risks, regressions on their components and on properly-aggregated data appear very well-suited to uncover this information.

Some have argued that the dividend yield may be nonstationary (e.g., Cai and Wang, 2014, and the references therein).[20] Our decompositions apply to all processes for which a classical Wold representation apply and, therefore, to all covariance-stationary process. Thus, even though extreme forms of persistence can be accommodated, unit root type behavior would fall outside of our framework. However, given a Beveridge–Nelson decomposition of the dividend yield (dp) into a transitory (covariance-stationary) component and a stochastic trend, i.e.,

$$dp_t = x_{dp,t} + \psi(1) \sum_{k=0}^{+\infty} \varepsilon_{t-k},$$

the extended Wold representation will still be valid for the transitory component $x_{dp,t}$. In particular, when interpreting the stochastic trend ($\psi(1) \sum_{k=0}^{+\infty} \varepsilon_{t-k}$) as one of the components of the raw series dp_t, scale-wise predictability may apply to it as well as to individual components of the transitory component $x_{dp,t}$. We leave this issue for future work.

Second, we have shown that a classical predictive system is not incompatible with scale-specific predictability. On the contrary, it imposes predictability at all scales with tight parametrizations on the predictive slopes. Our data generating process, instead, frees up predictability at different scales and reconstructs the original series from unrestricted, yet parsimonious, scale-wise relations. An interesting question to ask is, then: what would be the *implied* nature of predictability on the raw series? Our conjecture is that one could obtain interesting predictive relations which go well beyond the typical autoregressive structure of order 1 of classical predictive systems and give an important role to *long lags*, something that appears to be a recurrent theme in recent macro-finance work (see, e.g., Fuster et al., 2010). A formalization of predictability on the raw series given scale-specific predictability represents the next step in this research program.

Appendix A. Proof of Proposition I

Consider the following component dynamics for $j = j^*$, where $j^* \in \{1, \ldots, J\}$:

$$y^{(j^*)}_{k2^{j^*}+2^{j^*}} = \beta_{j^*} x^{(j^*)}_{k2^{j^*}}, \tag{A.1}$$

$$x^{(j^*)}_{k2^{j^*}+2^{j^*}} = \rho_{j^*} x^{(j^*)}_{k2^{j^*}} + e^{(j^*)}_{k2^{j^*}+2^{j^*}}. \tag{A.2}$$

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

$$y^{(j)}_{k2^j} = 0, \tag{A.3}$$

$$x^{(j)}_{k2^j} = 0. \tag{A.4}$$

In agreement with $y^{(j)}_{k2^j} = x^{(j)}_{k2^j} = 0$, we also have $\pi^{(J)}_{k2^j} = 0$ for both series.

[20] Nonstationarity in the dividend yield is *economically* questionable. Upon predictability, it would also induce nonstationarity in risk premia and, therefore, raise further economic issues. This said, well-accepted tests would generally fail to reject nonstationarity.

Solely for clarity, we provide details for $T = 16, j^* = 2$, and $J = 3$. The general case is completely analogous. Arrange the components of $\mathbf{x}$ as follows:

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

The components of $\mathbf{y}$ are arranged similarly. In order to reconstruct x_t, we run through each column of the matrix in Eq. (A.5) and, for each column, we perform the following operation:

$$X_8^{(3)} = \begin{pmatrix} x_8 \\ x_7 \\ x_6 \\ x_5 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \end{pmatrix} = \left(\widetilde{\mathcal{T}}^{(3)}\right)^{-1} \begin{pmatrix} \pi_8^{(3)} \\ x_8^{(3)} \\ x_8^{(2)} \\ x_4^{(2)} \\ x_8^{(1)} \\ x_6^{(1)} \\ x_4^{(1)} \\ x_2^{(1)} \end{pmatrix} \tag{A.6}$$

and

$$X_{16}^{(3)} = \begin{pmatrix} x_{16} \\ x_{15} \\ x_{14} \\ x_{13} \\ x_{12} \\ x_{11} \\ x_{10} \\ x_9 \end{pmatrix} = \left(\widetilde{\mathcal{T}}^{(3)}\right)^{-1} \begin{pmatrix} \pi_{16}^{(3)} \\ x_{16}^{(3)} \\ x_{16}^{(2)} \\ x_{12}^{(2)} \\ x_{16}^{(1)} \\ x_{14}^{(1)} \\ x_{12}^{(1)} \\ x_{10}^{(1)} \end{pmatrix}. \tag{A.7}$$

We do the same to reconstruct y_t. The matrix $\left(\widetilde{\mathcal{T}}^{(3)}\right)^{-1}$ takes the form:

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

We emphasize that $\widetilde{\mathcal{T}}^{(3)}$ is a suitably "re-scaled" version of the matrix $\mathcal{T}^{(3)}$ discussed in the main text. The "re-scaling" is, for instance, such that the generic term $x_{k2^{j^*}}^{(2)}$ of the second component of the $\mathbf{x}$ process is defined as

$$x_{k2^{j^*}}^{(2)} = \frac{x_{k2^{j^*}} + x_{k2^{j^*}-1} - x_{k2^{j^*}-2} - x_{k2^{j^*}-3}}{2}$$

rather than as

$$x^{(2)}_{k2^{j^*}} = \frac{x_{k2^{j^*}} + x_{k2^{j^*}-1} - x_{k2^{j^*}-2} - x_{k2^{j^*}-3}}{4},$$

which is the expression in the main text (c.f. Section 5). Similarly, given the expressions in the main text, the first component is multiplied by $\sqrt{2}$ and the third component, as well as the term $\pi^{(3)}$, are multiplied by $2\sqrt{2}$.

Using Eq. (A.4), together with Eqs. (A.6) and (A.7), we obtain

$$X^{(3)}_{16} = \begin{pmatrix} x_{16} = & x^{(2)}_{16}/2 \\ x_{15} = & x^{(2)}_{16}/2 \\ x_{14} = & -x^{(2)}_{16}/2 \\ x_{13} = & -x^{(2)}_{16}/2 \\ x_{12} = & x^{(2)}_{12}/2 \\ x_{11} = & x^{(2)}_{12}/2 \\ x_{10} = & -x^{(2)}_{12}/2 \\ x_{9} = & -x^{(2)}_{12}/2 \end{pmatrix} \quad \text{and} \quad X^{(3)}_{8} = \begin{pmatrix} x_{8} = & x^{(2)}_{8}/2 \\ x_{7} = & x^{(2)}_{8}/2 \\ x_{6} = & -x^{(2)}_{8}/2 \\ x_{5} = & -x^{(2)}_{8}/2 \\ x_{4} = & x^{(2)}_{4}/2 \\ x_{3} = & x^{(2)}_{4}/2 \\ x_{2} = & -x^{(2)}_{4}/2 \\ x_{1} = & -x^{(2)}_{4}/2 \end{pmatrix}. \tag{A.9}$$

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 $x_{t-h+1,t}$ is defined consistently with $y_{t+1,t+h}$. For $h = 4$, and using Eqs. (A.1) and (A.3) together with Eq. (A.9), we have

$$\begin{aligned}
y_{13,16} &= 0 & x_{13,16} &= 0 \\
y_{12,15} &= (-y^{(2)}_{16} + y^{(2)}_{12})/2 = \beta_2\left(-x^{(2)}_{12} + x^{(2)}_{8}\right)/2 & x_{12,15} &= (-x^{(2)}_{16} + x^{(2)}_{12})/2 \\
y_{11,14} &= -y^{(2)}_{16} + y^{(2)}_{12} = \beta_2\left(-x^{(2)}_{12} + x^{(2)}_{8}\right) & x_{11,14} &= -x^{(2)}_{16} + x^{(2)}_{12} \\
y_{10,13} &= (-y^{(2)}_{16} + y^{(2)}_{12})/2 = \beta_2\left(-x^{(2)}_{12} + x^{(2)}_{8}\right)/2 & x_{10,13} &= (-x^{(2)}_{16} + x^{(2)}_{12})/2 \\
y_{9,12} &= 0 & x_{9,12} &= 0 \\
y_{8,11} &= (-y^{(2)}_{12} + y^{(2)}_{8})/2 = \beta_2\left(-x^{(2)}_{8} + x^{(2)}_{4}\right)/2 & x_{8,11} &= (-x^{(2)}_{12} + x^{(2)}_{8})/2 \\
y_{7,10} &= -y^{(2)}_{12} + y^{(2)}_{8} = \beta_2\left(-x^{(2)}_{8} + x^{(2)}_{4}\right) & x_{7,10} &= -x^{(2)}_{12} + x^{(2)}_{8} \\
y_{6,9} &= (-y^{(2)}_{12} + y^{(2)}_{8})/2 = \beta_2\left(-x^{(2)}_{8} + x^{(2)}_{4}\right)/2 & x_{6,9} &= (-x^{(2)}_{12} + x^{(2)}_{8})/2 \\
y_{5,8} &= 0 & x_{5,8} &= 0 \\
y_{4,7} &= (-y^{(2)}_{8} + y^{(2)}_{4})/2 = \beta_2\left(-x^{(2)}_{4} + x^{(2)}_{0}\right)/2 & x_{4,7} &= (-x^{(2)}_{8} + x^{(2)}_{4})/2 \\
y_{3,6} &= -y^{(2)}_{8} + y^{(2)}_{4} = \beta_2\left(-x^{(2)}_{4} + x^{(2)}_{0}\right) & x_{3,6} &= -x^{(2)}_{8} + x^{(2)}_{4} \\
y_{2,5} &= (-y^{(2)}_{8} + y^{(2)}_{4})/2 = \beta_2\left(-x^{(2)}_{4} + x^{(2)}_{0}\right)/2 & x_{2,5} &= (-x^{(2)}_{8} + x^{(2)}_{4})/2 \\
y_{1,4} &= 0 & x_{1,4} &= 0.
\end{aligned}$$

Thus, regressing $y_{t+1,t+4}$ on $x_{t-3,t}$ yields $\tilde{\beta} = \beta_2$ with $R^2 = 100\%$. In general, when scale-wise predictability applies to a scale operating between 2^{j^*-1} and 2^{j^*} periods, maximum predictability upon two-way (forward/backward) aggregation arises over a horizon $h = 2^{j^*}$. In our case, $j^* = 2$ and $h = 4$.

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

$$\begin{aligned}
y_{15,16} &= y^{(2)}_{16} = \beta_2 x^{(2)}_{12} & x_{15,16} &= x^{(2)}_{16} \\
y_{14,15} &= 0 & x_{14,15} &= 0 \\
y_{13,14} &= -y^{(2)}_{16} = -\beta_2 x^{(2)}_{12} & x_{13,14} &= -x^{(2)}_{16} \\
y_{12,13} &= (-y^{(2)}_{16} + y^{(2)}_{12})/2 = \beta_2(-x^{(2)}_{12} + x^{(2)}_{8})/2 & x_{12,13} &= (-x^{(2)}_{16} + x^{(2)}_{12})/2
\end{aligned}$$

$$
\begin{aligned}
y_{11,12} &= y_{12}^{(2)} = \beta_2 x_8^{(2)} & x_{11,12} &= x_{12}^{(2)} \\
y_{10,11} &= 0 & x_{10,11} &= 0 \\
y_{9,10} &= -y_{12}^{(2)} = -\beta_2 x_8^{(2)} & x_{9,10} &= -x_{12}^{(2)} \\
y_{8,9} &= (-y_{12}^{(2)} + y_8^{(2)})/2 = \beta_2(-x_8^{(2)} + x_4^{(2)})/2 & x_{8,9} &= (-x_{12}^{(2)} + x_8^{(2)})/2 \\
y_{7,8} &= y_8^{(2)} = \beta_2 x_4^{(2)} & x_{7,8} &= x_8^{(2)} \\
y_{6,7} &= 0 & x_{6,7} &= 0 \\
y_{5,6} &= -y_8^{(2)} = -\beta_2 x_4^{(2)} & x_{5,6} &= -x_8^{(2)} \\
y_{4,5} &= (-y_8^{(2)} + y_4^{(2)})/2 = \beta_2(-x_4^{(2)} + x_0^{(2)})/2 & x_{4,5} &= (-x_8^{(2)} + x_4^{(2)})/2 \\
y_{3,4} &= y_4^{(2)} = \beta_2 x_0^{(2)} & x_{3,4} &= x_4^{(2)} \\
y_{2,3} &= 0 & x_{2,3} &= 0 \\
y_{1,2} &= -y_4^{(2)} = -\beta_2 x_0^{(2)} & x_{1,2} &= -x_4^{(2)},
\end{aligned}
$$

where, again, we used Eqs. (A.1) and (A.3) along with Eq. (A.9).

Given the dynamics in Eq. (A.2), using a fundamental block of four elements, the regression of $y_{t+1,t+2}$ on $x_{t-1,t}$ yields:

$$
\begin{aligned}
\tilde{\beta} &= \frac{\mathbb{C}(y_{t+1,t+2}, x_{t-1,t}) + \mathbb{C}(y_{t-1,t}, x_{t-3,t-2})}{\mathbb{V}\left(x_{t-1,t}\right) + \mathbb{V}\left(x_{t-2,t-1}\right) + \mathbb{V}\left(x_{t-3,t-2}\right) + \mathbb{V}\left(x_{t-4,t-3}\right)} \\
&= \frac{-\beta_2 \mathbb{V}\left(x_{t-2}^{(2)}\right)\rho_2 - \beta_2 \mathbb{V}\left(x_{t-2}^{(2)}\right)}{\mathbb{V}\left(x_{t+2}^{(2)}\right) + \mathbb{V}\left(\frac{x_{t+2}^{(2)}}{2}\right) + \mathbb{V}\left(\frac{x_{t-2}^{(2)}}{2}\right) - \frac{\mathbb{C}\left(x_{t+2}^{(2)}, x_{t-2}^{(2)}\right)}{2} + \mathbb{V}\left(x_{t-2}^{(2)}\right)} \\
&= -2\beta_2 \frac{(1+\rho_2)}{5-\rho_2}
\end{aligned}
$$

and, hence, an inconsistent slope estimate, upon estimation. The forward/backward beta could have a changed sign (with respect to β_2) or be drastically attenuated. In fact, $\tilde{\beta} = -\beta_2$ if $\rho_2 = 1$ and $\tilde{\beta} = 0$ if $\rho_2 = -1$.

Appendix B. Data

The empirical analysis is conducted using annual observations (and aggregates of annual observations) from 1930 to 2014.

Annual consumption is from the Bureau of Economic Analysis, series 7.1, and is defined as consumer expenditures on non-durables and services. Our measure of consumption variance is based on modeling consumption growth as an AR(1) with an error variance evolving as a heterogeneous ARCH model (c.f. Muller et al. (1997)). The HARCH dynamics accommodate heterogeneous information arrival, 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 et al. (2005).

We use the NYSE/Amex value-weighted index with dividends as our market proxy. 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 risk-free rate is the yield on the 1-year Treasury bill.

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

The market's realized variance over a year (i.e., between t and $t+1$), a measure of integrated variance, is obtained by computing

$$
v_{t,t+1} = \sum_{d=t_{first}}^{t_{last}} r_d^2,
$$

where r_d is the market's logarithmic return on day d and $[t_{first}, \ldots, t_{last}]$ denotes the days in the year.

The measure of economic policy uncertainty (EPU) is based on Baker et al. (2016).

Appendix C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.09.008.

References

Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. J. Finance 52 (3), 975–1005.
Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131 (4), 1593–1636.
Bandi, F.M., Perron, B., 2008. Long-run risk-return trade-offs. J. Econometrics 143 (2), 349–374.
Bansal, R., Khatchatrian, V., Yaron, A., 2005. Interpretable asset markets? Eur. Econ. Rev. 49 (3), 531–560.
Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. J. Financ. 59 (4), 1481–1509.
Baxter, M., King, R.G., 1999. Measuring business cycles: Approximate band-pass filters for economic time series. Rev. Econ. Stat. 81 (4), 575–593.
Beveridge, S., Nelson, C.R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. J. Monet. Econ. 7 (2), 151–174.
Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financ. Stud. 22 (11), 4463–4492.
Bonomo, M., Garcia, R., Meddahi, N., Tedongap, R., 2015. The long and the short of the risk-return trade-off. J. Econometrics 187 (2), 580–592.
Boudoukh, J., Richardson, M., Whitelaw, R.F., 2008. The myth of long-horizon predictability. Rev. Financ. Stud. 21 (4), 1577–1605.
Cai, Z., Wang, Y., 2014. Testing predictive regression models with nonstationary regressors. J. Econometrics 178 (P1), 4–14.
Calvet, L.E., Fisher, A.J., 2001. Forecasting multifractal volatility. J. Econometrics 105 (1), 27–58.
Calvet, L.E., Fisher, A.J., 2007. Multifrequency news and stock returns. J. Financ. Econ. 86 (1), 178–212.
Campbell, J.Y., Shiller, R., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1 (3), 195–228.
Campbell, J.Y., Vuolteenaho, T., 2004. Inflation illusion and stock prices. Amer. Econ. Rev. 94 (2), 19–23.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81 (1), 27–60.
Cochrane, J.H., 2008. The dog that did not bark: A defense of return predictability. Rev. Financ. Stud. 21 (4), 1533–1575.
Comin, D., Gertler, M., 2006. Medium-term business cycles. Amer. Econ. Rev. 96 (3), 523–551.
Crowley, P.M., 2007. A guide to wavelets for economists. J. Econ. Surv. 21 (2), 207–267.
Dijkerman, R., Mazumdar, R., 1994. Wavelet representations of stochastic processes and multiresolution stochastic models. IEEE Trans. Signal Process. 42 (7), 1640–1652.
Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. J. Financ. Econ. 25 (1), 23–49.
Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. J. Financ. Econ. 5 (2), 115–146.
Fuster, A., Hebert, B., Laibson, D., 2010. Natural expectations and macroeconomic fluctuations. J. Econ. Perspect. 24 (4), 67–84.
Gençay, R., Selçuk, F., Whitcher, B., 2001. An Introduction to Wavelets and Other Filtering Methods in Finance and Economics, first ed. Academic Press, New York.
Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. J. Econometrics 2 (2), 111–120.
Hansen, L.P., Scheinkman, J.A., 2009. Long-term risk: An operator approach. Econometrica 77 (1), 177–234.
Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. J. Finance 68 (5), 1721–1756.
Lamont, O., 1998. Earnings and expected returns. J. Finance 53 (5), 1563–1587.
Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. J. Finance 56 (3), 815–849.
Lewellen, J., 2004. Predicting returns with financial ratios. J. Financ. Econ. 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., Santos, T., Veronesi, P., 2004. Understanding predictability. J. Polit. Econ. 112 (1), 1–47.
Muller, U.A., Dacorogna, M.M., Dave, R.D., Olsen, R.B., Pictet, O.V., von Weizsacker, J.E., 1997. Volatilities of different time resolutions – Analyzing the dynamics of market components. J. Empir. Finance 4 (2), 213–239.
Müller, U.K., Watson, M.W., 2008. Testing models of low-frequency variability. Econometrica 76 (5), 979–1016.
Nelson, C.R., 1976. Inflation and rates of return on common stocks. J. Finance 31 (2), 471–483.
Ortu, F., Severino, F., Tamoni, A., Tebaldi, C., 2018. A persistence-based Wold-type decomposition for stationary time series, SSRN eLibrary, SSRN.
Ortu, F., Tamoni, A., Tebaldi, C., 2013. Long-run risk and the persistence of consumption shocks. Rev. Financ. Stud. 26 (11), 2876–2915.
Pastor, L., Stambaugh, R.F., 2009. Predictive systems: Living with imperfect predictors. J. Finance 64 (4), 1583–1628.
Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. J. Econometrics 33 (3), 311–340.
Ramsey, J.B., 1999. The contribution of wavelets to the analysis of economic and financial data. Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci. 357 (1760), 2593–2606.
Sizova, N., 2013. Long-Horizon return regressions with historical volatility and other long-memory variables. J. Bus. Econom. Statist. 31 (4), 546–559.
Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54 (3), 375–421.
Tamoni, A., 2011. The multi-horizon dynamics of risk and returns, SSRN eLibrary, SSRN.
Valkanov, R., 2003. Long horizon regressions: Theoretical results and applications. J. Financ. Econ. 68, 201–232.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.