

Online Supplement for  
“Scale of Predictability”

F.M. Bandi, B. Perron, A. Tamoni and C. Tebaldi  
May 8, 2017

# Contents

<b>A</b>	<b>Low-frequency humps and classical predictive systems: simulations</b>	<b>1</b>
<b>B</b>	<b>Low-frequency humps and <i>scale-wise predictive systems</i>: simulations</b>	<b>5</b>
B.1	Running two-way regressions on data from a component model . . . . .	6
<b>C</b>	<b>Further results</b>	<b>10</b>
C.1	Fitting an AR(1) process to the regressor . . . . .	10
C.2	Contemporaneous aggregation . . . . .	10
<b>D</b>	<b>The Long-run Risk Model with Persistence Heterogeneity in Variance</b>	<b>11</b>
D.1	The Valuation Approach: the details of the derivation . . . . .	15
D.2	The Financial Ratios . . . . .	15
D.3	The Risk Premium and Return Volatility . . . . .	20

## A Low-frequency humps and classical predictive systems: simulations

Leaving theoretical considerations aside, how likely are classical predictive systems to generate (in any finite sample) the tent-shaped dynamics detected in the data? We run forward/backward regressions using simulations from a traditional predictive model for excess market returns and *consumption* variance, namely Eqs. (2) and (3) in the main text. The parameters are estimated from the data employed previously:  $\beta = 1.80$ ,  $\rho = 0.734$ ,  $\sigma_u = 0.180$ ,  $\sigma_\epsilon = 0.0095$ , and  $\rho_{u,\epsilon} = -0.045$ .

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

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

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

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

[Insert Table A.1 about here]

**Panel A: Distribution of coefficient estimates.**

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

**Panel B: Distribution of  $R^2$ s.**

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

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

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

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

Simulations are performed using the parameters  $\beta = 1.7990$ ,  $\rho = 0.7335$ ,  $\sigma_\varepsilon = 0.1802$ ,  $\sigma_u = 0.0095$  and  $corr_{\varepsilon,u} = -0.0454$ . There are 85 observations for each simulation. **Panel A: Distribution of coefficient estimates.** The Table reports the mean, standard deviation, and median of the coefficient estimates from the predictive regression across 100,000 simulations. **Panel B: Distribution of  $R^2$ s.** The Table reports the mean, standard deviation, and median of the  $R^2$ s from the predictive regression across 100,000 simulations. “ $\beta$  ( $R^2$  increasing 6-12” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are monotonic in the horizons 6 to 12 years, i.e.  $\beta_{12} > \beta_{11} > \dots > \beta_6$  ( $R_{12}^2 > R_{11}^2 > \dots > R_6^2$ , respectively). “ $\beta$  ( $R^2$  decreasing 16-20” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are decreasing in the horizons 16 to 20 years, i.e.  $\beta_{16} > \beta_{17} > \dots > \beta_{20}$  ( $R_{16}^2 > R_{17}^2 > \dots > R_{20}^2$ , respectively). “ $\beta$  ( $R^2$ ) hump-shape” is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. “ $R^2$  hump-shape &  $> 50\%$ ” is the percentage of the simulations that produce  $R^2$ s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an  $R_h^2 > 50\%$  in the range  $12 \leq h \leq 16$ .

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

Should the data generating process not allow for predictability, the estimated slopes would be slightly increasing across the board. This is, as discussed above, akin to a typical spurious regression problem: the generation of unit root behavior by virtue of aggregation leads to the appearance of dependence. Such appearance is also reflected in increasing mean and median  $R^2$ s across simulations. This increasing behavior is pervasive across frequencies and, importantly, inconsistent with the marked tent-shaped structures reported in the data. We find, for instance, that the percentages of hump-shaped slopes and  $R^2$ s (again, monotonically increasing between 6 and 12 years *and* monotonically decreasing between 16 and 20 years) are 7.83% and 7.56%, respectively. When including the requirement of an  $R^2$  larger than 50%, the latter percentage drops to 3.41%. Should we also add the requirement that the drop in the  $R^2$  value between the 16-year horizon and the 20-year horizon is larger than 30%, something which is consistent with data, the percentage would decrease further to 1.40%.

**[Insert Table A.2 about here]**

Additional metrics are reported in Table A.2. All of them provide a consistent picture. In agreement with the arguments in Boudoukh, Richardson, and Whitelaw (2008), lack of predictability in the usual sense ( $\beta=0$  in the system in Eqs. (2)-(3)) could generate mild upward trending behavior in the slopes and coefficients of determination. It will however find it difficult to replicate both the reported humps and their magnitudes at the peaks.

**Panel A: Distribution of coefficient estimates.**

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

**Panel B: Distribution of  $R^2$ s.**

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

**Table A.2: Classical predictive system.** We simulate under the assumption of no predictability and an AR(1) process on  $x_t$ ,

$$R_{t+1} = \varepsilon_{t+1},$$

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

Simulations are performed using the parameters  $\rho = 0.7335$ ,  $\sigma_\varepsilon = 0.1951$ ,  $\sigma_u = 0.0095$  and  $corr_{\varepsilon, u} = -0.0454$ . There are 85 observations for each simulation. **Panel A: Distribution of coefficient estimates.** The Table reports the mean, standard deviation, and median of the coefficient estimates from the predictive regression across 100,000 simulations. **Panel B: Distribution of  $R^2$ s.** The Table reports the mean, standard deviation, and median of the  $R^2$ s from the predictive regression across 100,000 simulations. “ $\beta$  ( $R^2$ ) increasing 6-12” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are monotonic in the horizons 6 to 12 years, i.e.  $\beta_{12} > \beta_{11} > \dots > \beta_6$  ( $R_{12}^2 > R_{11}^2 > \dots > R_6^2$ , respectively). “ $\beta$  ( $R^2$ ) decreasing 16-20” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are decreasing in the horizons 16 to 20 years, i.e.  $\beta_{16} > \beta_{17} > \dots > \beta_{20}$  ( $R_{16}^2 > R_{17}^2 > \dots > R_{20}^2$ , respectively). “ $\beta$  ( $R^2$ ) hump-shape” is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. “ $R^2$  hump-shape &  $> 50\%$ ” is the percentage of the simulations that produce  $R^2$ s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an  $R_h^2 > 50\%$  in the range  $12 \leq h \leq 16$ .

## B Low-frequency humps and *scale-wise predictive systems*: simulations

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

We begin by postulating processes for the decimated details of the consumption variance and return series:

$$\begin{aligned} r_{k2^j+2^j}^{(j)} &= \beta_j v_{k2^j}^{(j)} \\ v_{k2^j+2^j}^{(j)} &= \rho_j v_{k2^j}^{(j)} + \varepsilon_{k2^j+2^j}^{(j)} \end{aligned} \tag{B.1}$$

for  $j = j^*$  and

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

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

The model implies a predictive system on the scale  $j^*$  and unrelated details for all other scales. In other words, predictability only occurs at the level of the  $j^*$ -th detail. Consistent with data, we set  $j^* = 4$ , i.e., only the fourth component of the return and variance process correlate with each other, their relation being based on the previously-reported betas in Table 2 in the main text (we set  $b_j$  equal to 2.8 for  $j = 4$  and zero otherwise). Moreover  $u_{k2^j}^{(j)}$  is  $N(0, \sigma_u^{(j)})$ , where  $\sigma_u^{(j)}$  is chosen so as to match the variance of the component  $r_{k2^j}^{(j)}$  at scale  $j$ .<sup>1</sup>

The fourth component of the market variance follows an autoregressive process of order one in scale time, with a scale-specific autoregressive parameter  $\rho_j$  calibrated to the data (we set  $\rho_j$  equal to 0.2 for  $j = 4$  and zero otherwise).<sup>2</sup> All other variance components

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

<sup>2</sup>The small autoregressive parameter in scale time is enough to generate a persistent process in calendar time. The autocorrelation of our simulated variance at lags 1, 2 and 3 is  $\text{ACF}(1)=0.71$ ,  $\text{ACF}(2)=0.50$ , and  $\text{ACF}(3)=0.33$ ; in the data, the first 3 lags of the sample ACF for consumption variance are 0.73, 0.48, and 0.33 respectively.

are assumed to be white noise shocks  $\varepsilon_{k2^j}^{(j)} \sim N(0, \sigma_\varepsilon^{(j)})$  with a variance chosen so as to match the variance of the component  $v_{k2^j}^{(j)}$  at scale  $j$ .

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

The data generating process is again formulated for (decimated) observations defined in scale time. We, therefore, simulate the process at scale  $j$  every  $2^j$  steps. To obtain the (calendar-time observations of) aggregate return and variance series from scale-time details, we aggregate the simulated details via the inverse Haar matrix (defined in the main text – Section 5 (Subsection 5.1 illustrates within a tractable example with  $J = 2$  the reconstruction steps from decimated observations to observations in the time domain).

In agreement with the discussion in the main text – Section 6, we now show that a predictive relation localized around the  $j^*$ -th scale will produce a pattern of  $R^2$ s which has a peak for aggregation levels corresponding to the horizon  $2^{j^*}$ .

## B.1 Running two-way regressions on data from a component model

Table B.1-Panel A shows the results obtained by running the regression

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

(see also Eq. (1) in the main text) on simulated data generated from Eq. (B.1). We compare these results to those in Table B.2, where no scale-wise predictability is assumed.<sup>3</sup> The tables report standard errors rather than  $t$ -statistics since, in the case of overlapping observations, the rate at which the standard errors grow may be informative.

When imposing the relation at scale  $j^* = 4$ , i.e., for a time span of 8 to 16 years (c.f., Table 1), we reach a peak in the  $R^2$ s of the two-way regressions exactly at 16 years. The 16-year  $R^2$  is 53.65% and is very comparable to its empirical counterpart from Table 2 in the main text. The 16-year  $R^2$  is also about 25 times as large as the one obtained in the case of no-predictability at the same horizon (see Table B.2). We notice that the slope estimates increase, reaching their maximum value of 4.48 at 16 years, a value identical to the the slope's estimated value on data of 4.48 (see Table 2 in the main text). After the 16-year mark, the slope estimates decrease almost monotonically. The coefficients are strongly significant between 12 and 18 years. They are insignificant before and after

---

<sup>3</sup>Under the null of no scale-wise predictability we set  $\beta_4$  equal to zero. In other words under the null we have  $r_{k2^j+2^j}^{(j)} = u_{k2^j+2^j}^{(j)}$  for all  $j$ , whereas under the alternative we had  $r_{k2^j+2^j}^{(j)} = \beta_j v_{k2^j}^{(j)}$  for  $j = j^*$ . Once again the variance of the shock is chosen so as to match the variance of the component  $r_{k2^j}^{(j)}$  at scale  $j$ . All the other calibrated parameters are untouched so that the first 2 lags of the ACF for the simulated consumption variance are the same as before.



those horizons. Hence, the simulations generate an hump-shaped pattern in the estimated slopes,  $t$ -statistics, and  $R^2$  which derives solely from imposing scale-specific predictability at a frequency lower than business-cycle frequencies.

**[Insert Tables B.1 and B.2 about here]**

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

Should we strengthen the metric a bit and consider now the percentage of  $R^2$  values which are hump-shaped, larger than 50% and such that the drop in  $R^2$  between the 16-year horizon and the 20-year horizon is larger than 30%, we would find that about 19% of the sample paths would reproduce exactly the pattern in the data. This observation translates into a likelihood of observing paths with those characteristics which is more than 13 times larger than the corresponding likelihood in the case of a traditional predictive model. Additional metrics, providing analogous information, are reported in Table B.1.

As emphasized earlier, if aggregation were to lead, somewhat mechanically, to statistically significant, larger slopes and higher  $R^2$  values by virtue of the creation of stochastic trends, hump-shaped behaviors would be unlikely and contemporaneous (forward/forward) aggregation would also lead to spurious predictability. We have shown, instead, that hump-shaped structures may naturally arise from predictability at the corresponding scale. We now turn to forward/backward aggregation. Again, we simulate under  $j^* = 4$  (in Table B.1-Panel B). When both the regressor and the regressand are aggregated over the same time interval, no statistical significant predictability is detected. The Online Supplement (see Section “Contemporaneous aggregation”) provides a theoretical justification.

These effects are similar to what one would obtain if, instead of aggregating forward/backward, one were to aggregate forward/backward while simulating from a component model like Eqs. (12)-(13) in the main text under the assumption of *absence* of scale-specific predictability, i.e.,  $\beta_j = 0$ . Absence of predictability would lead to statistically insignificant slopes numerically very close to zero for all horizons. The resulting  $R^2$  values would also be very small and rather far from magnitudes seen in the data.

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

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

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

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

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

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

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

Table B.1: **Simulation under the null of scale-dependent predictability. The relation is at scale  $j^* = 4$ .** We simulate excess market returns ( $y$ ) and consumption volatility ( $x$ ) under the assumption of predictability at scale  $j^* = 4$ . We simulate  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_t^{(j)}$  for  $j = 4$  and  $x_t^{(j)} = \epsilon_t^{(j)}$  otherwise. We implement 100000 replications. We set  $T = 128$ . For each regression, the table reports the median and the standard deviation (in parentheses) of the coefficient estimates from the predictive regression as well as the median of the adjusted  $R^2$  statistics (in square brackets). **Panel A: two-way (forward/backward) regressions.** We run linear regressions (with an intercept) of  $h$ -period continuously compounded excess market returns on  $h$ -period *past* consumption volatility. **Panel B: contemporaneous aggregation.** We run linear regressions (with an intercept) of  $h$ -period continuously compounded excess market returns on  $h$ -period *contemporaneous* consumption volatility. **Panel C: Distribution of coefficient estimates and of  $R^2$ s.** “ $\beta$  ( $R^2$ ) increasing 6-12” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are monotonic in the horizons 6 to 12 years, i.e.  $\beta_{12} > \beta_{11} > \dots > \beta_6$  ( $R_{12}^2 > R_{11}^2 > \dots > R_6^2$ , respectively). “ $\beta$  ( $R^2$ ) decreasing 16-20” is the percentage of the simulations that produce coefficients ( $R^2$ s) that are decreasing in the horizons 16 to 20 years, i.e.  $\beta_{16} > \beta_{17} > \dots > \beta_{20}$  ( $R_{16}^2 > R_{17}^2 > \dots > R_{20}^2$ , respectively). “ $\beta$  ( $R^2$ ) hump-shape” is the percentage of the simulations that produce coefficients that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years. “ $R^2$  hump-shape &  $> 50\%$ ” is the percentage of the simulations that produce  $R^2$ s that are increasing in the horizons 6 to 12 years, and decreasing in the horizons 16 to 20 years, and with an  $R_h^2 > 50\%$  in the range  $12 \leq h \leq 16$ .

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

Table B.2: **Simulation under the null of ABSENCE of scale-dependent predictability.** We simulate excess market returns ( $y$ ) and market variance ( $x$ ) under the assumption of no predictability. We simulate  $x_t^{(j)} = \rho_j x_{t-2j}^{(j)} + \epsilon_{t,j}$  for  $j = 4$  and  $x_t^{(j)} = \epsilon_t^{(j)}$  otherwise. We implement 500 replications. We set  $T = 128$ . We then run linear regressions (with an intercept) of  $h$ -period continuously compounded excess market returns on  $h$ -period past realized market variances. For each regression, the table reports the median and the standard deviation (in parentheses) of the coefficient estimates from the predictive regression as well as the median of the adjusted  $R^2$  statistics (in square brackets).

## C Further results

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

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

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

From (A.9) in the main text, it is easy to see that, for  $j^* = 2$ :

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

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

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

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

### C.2 Contemporaneous aggregation

We now run the contemporaneous regression

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

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

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

By taking covariances we obtain

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

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

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

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

## D The Long-run Risk Model with Persistence Heterogeneity in Variance

We incorporate in the standard long-run risk model the decomposition of time series into components realized over different time horizons so that the log consumption growth,  $g_t$ , and the log dividend growth,  $gd_t$  take the following form:

$$g_t = \sum_{j=1}^J g_t^{(j)}$$

$$gd_t = \sum_{j=1}^J gd_t^{(j)}$$

where  $g_t^{(j)}$  and  $gd_t^{(j)}$  are meant to capture the behavior of the original time series over time-scale of length  $2^{j-1}$  periods. Now, we closely follow the approach of Bollerslev, Tauchen, and Zhou (2009) and Zhou (2010) and suppose that the growth rate of consumption in the economy is not predictable. The novelty is to assume that each component of consumption growth,  $g_t^{(j)}$  and of dividend growth,  $gd_t^{(j)}$  is driven by its own stochastic volatility,  $\sigma_{j,t}$ , i.e.:

$$g_{t+2j}^{(j)} = \sigma_t^{(j)} e_{g,t+2j}^{(j)}, \quad \text{where } e_{g,t+2j}^{(j)} \sim N(0, 1) \quad (\text{D.1})$$

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

with the shocks  $e_{g,t+2j}^{(j)}$  and  $e_{gd,t+2j}^{(j)}$  being correlated for  $j = j'$  and uncorrelated otherwise. For parsimony, as in Bansal and Yaron (2004), volatility of consumption and dividends is driven by a common time-varying component. To close the dynamics of the model we assume that each of the stochastic variance components  $\{\sigma_{j,t}^2\}_{j=1}^J$  follows its own autoregressive process, i.e.

$$\sigma_{t+2j}^{(j)2} = \rho_j \sigma_t^{(j)2} + \varepsilon_{t+2j}^{(j)}, \quad \text{with } \varepsilon_{t+2j}^{(j)} \sim N(0, \sigma^{(j)2}). \quad (\text{D.3})$$

This statistical description of volatility is called multi-scale autoregressive process, for it has an autoregressive process occurring on all scales  $j = 1, \dots, J$ . This is the key innovation of our model. We consider variance's components observed over time intervals of different sizes. The heterogeneity of the model derives from the fact that different autoregressive structures are present at each time scale allowing volatility components to decay at different rates. With this regard we sometimes refer to  $j$  as the level of persistence of the  $j$ -th components. Importantly, equations (D.1) to (D.3) represent a natural way to incorporate persistence heterogeneity in the macroeconomic uncertainty framework while retaining its pedagogical simplicity. Note that in this simple specification of the model variance can go negative. To ensure the positivity of the variance process, one can pursue the approach of Barndorff-Nielsen and Shephard (2001) and assume that the innovations in the components of variance processes follow a Gamma distribution. This would just slightly complicate the algebra without adding much to the intuition to the model. In the following we therefore assume that consumption variance shocks are Gaussian to explain the major implications of the model.

To give economic and structural meaning to the parameters we assume, as in BY04, a pure exchange economy with a representative agent with Epstein-Zin recursive preferences. The well known Euler condition for such an agent is:

$$E_t \left[ e^{m_{t+1} + r_{t+1}^i} \right] = 1 \quad (\text{D.4})$$

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

$$m_{t+1} = \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{t+1}^a, \quad (\text{D.5})$$

$r_{t+1}^a$  is the log return of the claim which distributes a dividend equals to aggregate consumption and  $r_{t+1}^i$  is the log return on any asset  $i$ . The parameter  $\beta$  is the preference discount factor. The preference parameter  $\psi$  measures the intertemporal elasticity of substitution,  $\gamma$  measures the risk aversion and  $\theta = (1 - \gamma) / (1 - 1/\psi)$ .

In what follows we provide the basic steps to determine the pricing kernel and risk premia on the market portfolio in our long-run risk model with persistence heterogeneity.<sup>4</sup> Recall first that by the standard Campbell and Shiller (1988) log-linear approximation for returns one obtains:

$$\begin{aligned} r_{a,t+1} &= \kappa_0 + \kappa_1 z_{t+1}^a - z_t^a + g_{t+1} \\ r_{m,t+1} &= \kappa_{0,m} + \kappa_{1,m} z_{t+1}^m - z_t^m + g d_{t+1} \end{aligned} \quad (\text{D.6})$$

where  $z_t^a$ ,  $z_t^m$ , denote the log price-consumption and the log price-dividend ratio respec-

---

<sup>4</sup>All details behind our calculations are given in the Appendix D.1.

tively. Recalling our decomposition of consumption and dividends into components with different levels of persistence, and denoting with  $z_{a,t}^{(j)}$ ,  $z_{m,t}^{(j)}$ , the components with persistence  $j$  of the (log) price-consumption ratio and price-dividend ratio respectively, it is natural to conjecture that there exists *component by component* a linear relation between the financial ratios and our state variables  $\sigma_{j,t}^2$ , i.e.

$$\begin{aligned} z_{a,t}^{(j)} &= A_{0,j} + A_j \sigma_{j,t}^2 \\ z_{m,t}^{(j)} &= A_{0,j}^m + A_j^m \sigma_{j,t}^2 \end{aligned} \tag{D.7}$$

As long as  $A_j$  and  $A_j^m$  are not vanishing, these relations tells us that the variation in valuation ratios can be attributed to fluctuations in economic uncertainty. These relations together with equation (D.3) imply that measures of economic uncertainty (conditional variance of consumption) are predicted by components of valuation ratios.

The values of  $A_{0,j}$ ,  $A_j$ ,  $A_{0,j}^m$ ,  $A_j^m$  in terms of the parameters of the model are obtained from the Euler condition (D.4) after the log stochastic discount factors and the returns are all expressed in terms of the factors  $\{\sigma_{j,t}^2\}_{j=1}^J$  and of the innovations  $\{e_{g,t+2j}^{(j)}\}_j$  and  $\{\varepsilon_{t+2j}^{(j)}\}_j$ . In Appendix D.1 we show that plugging these expressions for the stochastic discount factor and for the returns into the Euler equation and using the method of undetermined coefficients one obtains a set of equations for the coefficients  $A_{0,j}$ ,  $A_j$ ,  $A_{0,j}^m$ ,  $A_j^m$ , the solution of which are given by the following vectors of sensitivities:

$$\begin{aligned} \underline{A} &= 0.5 \frac{\left(\theta - \frac{\theta}{\psi}\right)^2}{\theta} (\mathbb{I}_J - \kappa_1 M)^{-1} \underline{1} \\ \underline{A}_m &= (\mathbb{I}_J - \kappa_{1,m} M)^{-1} \left( \frac{H_m}{2} - .5(1 - \gamma) \left( \frac{1}{\psi} - \gamma \right) \right) \end{aligned}$$

where

$$M = \text{diag}(\rho_1, \dots, \rho_J)$$

and  $\underline{A}$  and  $\underline{A}_m$  denote the column vectors with entries,  $A_1, \dots, A_J$ ,  $A_1^m, \dots, A_J^m$ , respectively. Two features of this model specification are noteworthy. First, if the IES and risk aversion are larger than 1, then  $\theta$  is negative, and a rise in volatility lowers the price-consumption ratio. Similarly, an increase in economic uncertainty will make consumption more volatile, which lowers asset valuations and increases the risk premia on all assets. This highlights that an IES larger than 1 is critical for capturing the negative correlation between price-dividend ratios and consumption variance. Second, an increase in the permanence of variance shocks, that is  $M$ , magnifies the effects of volatility shocks on valuation ratios, as changes in economic uncertainty are perceived as being long-lasting.

To study the scale-dependent consequences of our model for the equity premium it is important to recall that the innovation in the returns' component at level of persistence

$j$  are given by:

$$\begin{aligned} r_{a,t+2j}^{(j)} - E_t[r_{a,t+2j}^{(j)}] &= \sigma_{j,t} e_{g,t+2j}^{(j)} + \kappa_1 \left( A_j \varepsilon_{t+2j}^{(j)} \right) \\ r_{m,t+2j}^{(j)} - E_t[r_{m,t+2j}^{(j)}] &= \varphi_{d,j} \sigma_{j,t} e_{d,t+2j}^{(j)} + \underbrace{\kappa_{1,m} \underline{A}_m}_{\beta_{m,\varepsilon}} \varepsilon_{t+2j}^{(j)} \end{aligned}$$

and that the innovations of the stochastic discount factor's components are given by

$$m_{t+2j}^{(j)} - E_t[m_{t+2j}^{(j)}] = -\lambda_g \sigma_{j,t} e_{g,t+2j}^{(j)} - \lambda_j \varepsilon_{t+2j}^{(j)} \quad j = 1, \dots, J \quad (\text{D.8})$$

Recall that the risk premium on any asset  $i$  satisfy, in this set-up,  $E_t[r_{i,t,t+h} - r_{f,t+h}] + 0.5\sigma_{r_{i,t,t+h}}^2 = -\text{cov}_t(m_{t,t+h}, r_{i,t,t+h})$  where  $r_{i,t,t+h}$  and  $m_{t,t+h}$  are the stock return and stochastic discount factor aggregated over  $h$ -period. With the innovations to the equilibrium returns at hand and using (D.8) one finally can compute the risk premia for the consumption claim asset,  $r_{a,t+2h-1}$  and for the market portfolio,  $r_{m,t+2h-1}$  for any horizon  $2h-1$ :

$$\begin{aligned} E_t[r_{a,t,t+2h-1} - r_{f,t+2h-1}] + 0.5\sigma_{r_{a,t,t+2h-1}}^2 &= -\text{cov}_t \left( \sum_{j=h}^J m_{t+2j}^{(j)}, \sum_{j=h}^J r_{a,t+2j}^{(j)} \right) \\ &= \lambda_g \sum_{j=h}^J \sigma_{j,t}^2 + \kappa_1 \underline{\lambda}_\varepsilon \mathbf{Q} \underline{A}' \end{aligned} \quad (\text{D.9})$$

$$\begin{aligned} E_t[r_{m,t,t+2h-1} - r_{f,t+2h-1}] + 0.5\sigma_{r_{m,t,t+2h-1}}^2 &= -\text{cov}_t \left( \sum_{j=h}^J m_{t+2j}^{(j)}, \sum_{j=h}^J r_{m,t+2j}^{(j)} \right) \\ &= \lambda_g \varphi_{d,j} \sum_{j=h}^J \sigma_{j,t}^2 \rho_j + \kappa_{1,m} \underline{\lambda}_\varepsilon \mathbf{Q} \underline{A}'_m \end{aligned} \quad (\text{D.10})$$

where  $\lambda_g \equiv \left( \frac{\theta}{\psi} - \theta + 1 \right)$ ,  $\underline{\lambda}_\varepsilon \equiv \kappa_1 (1 - \theta) \underline{A}$  and  $\mathbf{Q} = \mathbf{E}_t [\varepsilon_{t+1} \varepsilon'_{t+1}]$ .



## D.1 The Valuation Approach: the details of the derivation

In this Section we show the steps to obtain the values of the financial ratios coefficients  $A_{0,j}$ ,  $A_j$ ,  $A_{0,j}^m$ ,  $A_j^m$  in terms of the parameters of the model. We then compute the equity premia on both the consumption claim asset and the market return. In what follows, we make use of the decomposition of time series into layers with different levels of persistence<sup>5</sup>

$$x_t = \sum_{j=1}^J x_t^{(j)} + \pi_t^{(J)} \quad (\text{D.12})$$

We solve first for the price-consumption coefficients  $A_{0,j}$ ,  $A_j$  and hence for the consumption return  $r_{a,t+1}$ . This determines the pricing kernel. Then we solve for the price-dividend ones  $A_{0,j}^m$ ,  $A_j^m$  and consequently for the risk premia on the market portfolio,  $r_{m,t+1}$ .

To obtain the values of the coefficients  $A_{0,j}$ ,  $A_j$  we exploit the Euler condition

$$\begin{aligned} & E_t \left[ \exp \left( \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1} + r_{i,t+1} \right) \right] \\ &= E_t \left[ \exp \left( \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + \theta r_{a,t+1} \right) \right] = 1 \end{aligned}$$

which is derived from (D.4) for the special case where the asset being priced is the aggregate consumption claim, i.e.  $r_{i,t+1} = r_{a,t+1}$ . We then express the log consumption growth  $g_{t+1}$  and the return  $r_{a,t+1}$  in terms of the factors  $\{x_t^{(j)}\}_j$  and of the innovations  $\{e_{g,t+2j}^{(j)}\}_j$  and  $\{\varepsilon_{t+2j}^{(j)}\}_j$ . To do so we plug first the Campbell and Shiller (1988) approximation for log returns, see equation (D.6), into the above expression to obtain:

$$E_t \left[ \exp \left( \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + \theta \underbrace{(\kappa_0 + \kappa_1 z_{t+1}^a - z_t^a + g_{t+1})}_{r_{a,t+1}} \right) \right] = 1$$

By the backward decomposition (D.12) applied to the (demeaned) price-consumption ratio at time  $t$  and by the forward decomposition (D.11) applied to the (demeaned)

---

<sup>5</sup>Alternatively, to reconstruct the realization of  $x_{t+1}$  from the effect that this realization will have at different horizons we use

$$x_{t+1} = - \sum_{j=1}^J x_{t+2j}^{(j)} + \pi_{t+2j}^{(J)} \quad (\text{D.11})$$

## D.2 The Financial Ratios

consumption growth and price-consumption processes at time  $t + 1$  we have:

$$z_{a,t} \simeq \sum_{j=1}^J z_{a,t}^{(j)} + \pi_a \quad (\text{D.13})$$

$$z_{a,t+1} \simeq - \sum_{j=1}^J z_{a,t+2j}^{(j)} + \pi_a \quad (\text{D.14})$$

$$g_{t+1} \simeq - \sum_{j=1}^J g_{t+2j}^{(j)} + \pi_g \quad (\text{D.15})$$

where we denote by  $\pi_g$ ,  $\pi_a$  the mean consumption growth and price-consumption ratio (with  $\pi_a$  to be determined in equilibrium) respectively.<sup>6</sup>

We now show that the variance of the consumption component at level  $j$ ,  $\sigma_{j,t}^2$ , coincides with the component at level  $j$  of the consumption variance. In fact we have that:

$$\sigma_t^2 = \text{Var}_t(g_{t+1}) = \text{Var}_t\left(\sum_{j=1}^J g_{t+2j}^{(j)}\right) = \text{Var}_t\left(\sum_{j=1}^J \sigma_{j,t} e_{g,t+2j}^{(j)}\right) = \sum_{j=1}^J \sigma_{j,t}^2$$

where we use the fact that the shocks  $e_{j,t}$  and  $e_{j',t}$  are uncorrelated for all  $j \neq j'$ .

Plugging the above expressions into the Euler condition yields:

$$E_t \left[ \exp \left( \theta \log \beta - \frac{\theta}{\psi} \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) + \theta \left( \kappa_0 + \kappa_1 \left( \sum_{j=1}^J (-z_{t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) + \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) \right) \right) \right]$$

Finally using the dynamics for the components of log consumption growth given in equation (D.1) together with our guess for the components of price-consumption ratio solution given in equation (D.7), rearranging terms and using the log normal properties of the

---

<sup>6</sup>Recalling that by definition  $\pi_t^{(J)}$ ,  $\pi_{a,t}^{(J)}$  are sample means of past realizations of consumption growth and of the price-consumption ratio, we assume that there exists  $J$  large enough so that  $\pi_t^{(J)} \simeq \pi_g$  and  $\pi_{a,t}^{(J)} \simeq \pi_a$  for all  $t$ .

shocks we obtain:

$$\begin{aligned}
& E_t \left[ \exp \left( \theta \log \beta + \theta \left( 1 - \frac{1}{\psi} \right) \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) + \theta \left( \kappa_0 + \kappa_1 \left( \sum_{j=1}^J (-z_{t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) \right) \right) \right] \\
&= E_t \left[ \exp \left( \theta \log \beta + \theta \left( 1 - \frac{1}{\psi} \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{g,t+2j}^{(j)} \right) \right) + \right. \right. \\
&\quad \left. \left. \theta \left( \kappa_0 + \kappa_1 \left( - \sum_{j=1}^J A_{0,j} - \sum_{j=1}^J A_j \sigma_{j,t+2j}^2 \right) - \left( \sum_{j=1}^J A_{0,j} + \sum_{j=1}^J A_j \sigma_{j,t}^2 \right) \right) \right) \right] \\
&= E_t \left[ \exp \left( \theta (\log \beta + \kappa_0 - (\kappa_1 + 1) \sum_{j=1}^J A_{0,j}) + \dots \right. \right. \\
&\quad \left. \left. \theta \left( 1 - \frac{1}{\psi} \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{g,t+2j}^{(j)} \right) \right) + \theta \left( \kappa_1 \sum_{j=1}^J -A_j \sigma_{j,t+2j}^2 - \sum_{j=1}^J A_j \sigma_{j,t}^2 \right) \right) \right] \\
&= E_t \left[ \exp \left( \theta (\log \beta + \kappa_0 - (\kappa_1 + 1) \sum_{j=1}^J A_{0,j}) + \dots \right. \right. \\
&\quad \left. \left. \theta \left( 1 - \frac{1}{\psi} \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{g,t+2j}^{(j)} \right) \right) + \theta \left( \kappa_1 \sum_{j=1}^J A_j \underbrace{(e_j M \tilde{\Sigma}_t + e_j \varepsilon_{t+2j})}_{\sigma_{j,t+2j}^2} - \sum_{j=1}^J A_j \sigma_{j,t}^2 \right) \right) \right] = 1
\end{aligned}$$

where we defined  $\tilde{\Sigma}_t \equiv [\sigma_{1,t}^2, \dots, \sigma_{J,t}^2]^\top$ ,  $e_j$  is the vector of the canonical basis,  $M$  is the  $J$ -dimensional diagonal matrix with the opposite of the persistence parameters  $\rho_j$  on the diagonal. Collecting terms in  $\tilde{\Sigma}_t$  yields eventually a system of equations

$$e_j \left( 0.5 \left( \theta - \frac{\theta}{\psi} \right)^2 + \theta A_j (\kappa_1 M - \mathbb{I}_J) \right) = 0$$

for all  $j = 1, \dots, J$ . If we introduce the following column vectors

$$\underline{A} \equiv [A_1, \dots, A_J]^\top$$

the solution to these equations is given by the following vectors of sensitivities:

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

To derive the expression for  $A_j^m$  we exploit once again the Euler condition

$$E_t \left[ \exp \left( \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1} + r_{m,t+1} \right) \right] = 1$$

where now the asset being priced is the market return  $r_{m,t+1}$ . Following the same steps as

above, and additionally using the Campbell and Shiller (1988) log-linear approximation for  $r_{m,t+1}$ , see equation (D.6) we can then rewrite the Euler equation as:

$$\begin{aligned}
& E_t \left[ \exp \left( \theta \log \beta + \theta \left( 1 - \frac{1}{\psi} \right) \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) + \right. \right. \\
& \quad (\theta - 1) \left( \kappa_0 + \kappa_1 \left( \sum_{j=1}^J (-z_{t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) \right) - \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) + \\
& \quad \left. \left. \underbrace{\kappa_{0,m} + \kappa_{1,m} \left( \sum_{j=1}^J (-z_{m,t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_{m,t}^{(j)} \right) + \left( \sum_{j=1}^J (-gd_{t+2j}^{(j)}) \right)}_{r_{m,t+1}} \right) \right] \\
& = E_t \left[ \exp \left( \theta \log \beta + \left( \theta - \frac{\theta}{\psi} - 1 \right) \left( \sum_{j=1}^J (-g_{t+2j}^{(j)}) \right) + \right. \right. \\
& \quad (\theta - 1) \left( \kappa_0 + \kappa_1 \left( \sum_{j=1}^J (-z_{t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) \right) + \\
& \quad \left. \left. \underbrace{\kappa_{0,m} + \kappa_{1,m} \left( \sum_{j=1}^J (-z_{m,t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_{m,t}^{(j)} \right) + \left( \sum_{j=1}^J (-gd_{t+2j}^{(j)}) \right)}_{r_{m,t+1,t+h}} \right) \right] = 1
\end{aligned}$$

Let's focus on the term

$$(\theta - 1) \left( \kappa_0 + \kappa_1 \left( \sum_{j=1}^J (-z_{t+2j}^{(j)}) \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) \right)$$

This can be written

$$\begin{aligned}
& = (\theta - 1) \left( \kappa_1 \left( - \sum_{j=1}^J A_j \sigma_{j,t+2j}^2 \right) - \left( \sum_{j=1}^J A_j \sigma_{j,t}^2 \right) \right) \\
& = (\theta - 1) \left( \underline{A} (\kappa_1 M - \mathbb{I}_J) \tilde{\Sigma}_t \right)
\end{aligned}$$

and plugging the solution for  $\underline{A}$  we eventually obtain

$$\begin{aligned}
& = -(\theta - 1) 0.5 \theta \left( 1 - \frac{1}{\psi} \right)^2 \left( \tilde{\Sigma}_t \right) \\
& = -0.5 (1 - \gamma) \left( \frac{1}{\psi} - \gamma \right) \left( \tilde{\Sigma}_t \right)
\end{aligned}$$

Plugging into the Euler equation the above simplifying expression, using the dynamics for the components of the log consumption growth and the log dividend growth given in

formula (D.1) and (D.2), respectively and rearranging terms we have:

$$\begin{aligned}
& E_t \left[ \exp \left( \theta \log \beta + \left( \theta - \frac{\theta}{\psi} - 1 \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{j,t+2j}^g \right) \right) - 0.5(1-\gamma) \left( \frac{1}{\psi} - \gamma \right) \left( \tilde{\Sigma}_t \right) + \dots \right. \right. \\
& \quad \left. \left. + \kappa_{0,m} + \kappa_{1,m} \left( \sum_{j=1}^J \left( -z_{m,t+2j}^{(j)} \right) \right) - \left( \sum_{j=1}^J z_{m,t}^{(j)} \right) + \left( \sum_{j=1}^J \left( -gd_{t+2j}^{(j)} \right) \right) \right) \right] \\
&= E_t \left[ \exp \left( \theta \log \beta + \left( \theta - \frac{\theta}{\psi} - 1 \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{g,t+2j}^{(j)} \right) \right) - 0.5(1-\gamma) \left( \frac{1}{\psi} - \gamma \right) \left( \tilde{\Sigma}_t \right) + \dots \right. \right. \\
& \quad \left. \left. + \kappa_{0,m} + \kappa_{1,m} \left( \sum_{j=1}^J \left( -z_{m,t+2j}^{(j)} \right) \right) - \left( \sum_{j=1}^J z_{m,t}^{(j)} \right) + \left( \sum_{j=1}^J \varphi_{d,j} \sigma_{j,t} \left( -e_{d,t+2j}^{(j)} \right) \right) \right) \right] \\
&= E_t \left[ \exp \left( \theta \log \beta + \left( \theta - \frac{\theta}{\psi} - 1 \right) \left( \sum_{j=1}^J \sigma_{j,t} \left( -e_{g,t+2j}^{(j)} \right) \right) - 0.5(1-\gamma) \left( \frac{1}{\psi} - \gamma \right) \left( \tilde{\Sigma}_t \right) + \dots \right. \right. \\
& \quad \left. \left. + \kappa_{0,m} + \left( \kappa_{1,m} \left( \sum_{j=1}^J A_j^m \underbrace{\left( e_j M \tilde{\Sigma}_t + e_j \varepsilon_{t+2j} \right)}_{\sigma_{j,t+2j}^2} \right) - \sum_{j=1}^J A_m^{(j)} \sigma_{j,t} \right) + \left( \sum_{j=1}^J \varphi_{d,j} \sigma_{j,t} \left( -e_{d,t+2j}^{(j)} \right) \right) \right) \right] =
\end{aligned}$$

where in the last line we substitute for our guess for the components of log price-dividend ratio given in equation (D.7). Finally using the log normal property of the shocks, collecting all the  $\sigma_{j,t}$  terms and defining the vectors

$$H_m \equiv [\lambda_g^2 + \varphi_{d,1}^2, \dots, \lambda_g^2 + \varphi_{d,J}^2]$$

where  $\lambda_g = \theta - \frac{\theta}{\psi} - 1$ , and

$$\begin{aligned}
\underline{A}_m &\equiv [A_1^m, \dots, A_J^m] \\
\underline{\phi} &\equiv [\phi_1, \dots, \phi_J]
\end{aligned} \tag{D.16}$$

we obtain the following restriction in vector notation,

$$\begin{aligned}
\underline{A}^m (\kappa_{1,m} M - \mathbb{I}_J) &= -\frac{H_m}{2} + 0.5(1-\gamma) \left( \frac{1}{\psi} - \gamma \right) \\
\underline{A}^m &= (\mathbb{I}_J - \kappa_{1,m} M)^{-1} \left( \frac{H_m}{2} - .5(1-\gamma) \left( \frac{1}{\psi} - \gamma \right) \right)
\end{aligned}$$

For reasonable parameter values  $\underline{A}^m$  are negative resulting in the well-known leverage effect, i.e., shocks to return is negatively correlated with shocks to variance process.

### D.3 The Risk Premium and Return Volatility

The risk premium for any asset is determined by the conditional covariance between the return and the SDF. For instance we can compute the risk premium on any asset  $i$  as

$$E_t[r_{i,t+1} - r_{f,t}] + 0.5\sigma_{r_{i,t}}^2 = -cov_t(m_{t+1}, r_{i,t+1})$$

We therefore need to compute first the innovations in the stochastic discount factor and in the returns.

Given the solution above for  $z_{a,t}^{(j)}$  it is possible to derive the innovation to the return  $r_{a,t+1}$  as a function of the evolution of the state variables and the parameters of the model. In particular the equilibrium return innovations can be found by plugging the expressions (D.13), (D.14) and (D.15) into the Campbell and Shiller (1988) approximation for log returns, see equation (D.6) to obtain

$$\begin{aligned} r_{a,t+1} - E_t[r_{a,t+1}] &= \left( \sum_{j=1}^J g_{t+2j}^{(j)} \right) + \kappa_0 + \kappa_1 \left( \sum_{j=1}^J z_{t+2j}^{(j)} \right) - \left( \sum_{j=1}^J z_t^{(j)} \right) - E_t[r_{a,t+1}] \\ &= \sum_{j=1}^J \sigma_{j,t} e_{j,t+2j}^g + \kappa_1 \left( \sum_{j=1}^J A_j (e_j \varepsilon_{t+2j}) \right) \\ &= \sigma_{j,t} \odot e_{j,t+1}^g + \kappa_1 \underline{A} \boldsymbol{\varepsilon}_{t+1} \end{aligned} \quad (\text{D.17})$$

where we define

$$\begin{aligned} \boldsymbol{\varepsilon}_{t+1}^\top &\equiv [\varepsilon_{1,t+2^1}, \dots, \varepsilon_{J,t+2^J}] \\ \sigma_{j,t} \odot e_{j,t+1}^g &\equiv \sum_{j=1}^J \sigma_{j,t} e_{j,t+2j}^g \end{aligned}$$

The innovation in the return component at level of persistence  $j$  is given by:

$$r_{a,t+2j}^{(j)} - E_t[r_{a,t+2j}^{(j)}] = \sigma_{j,t} e_{j,t+2j}^g + \kappa_1 (A_j \varepsilon_{j,t+2j}) \quad (\text{D.18})$$

It is trivial to show that

$$r_{a,t+1} - E_t[r_{a,t+1}] = \sum_{j=1}^J r_{a,t+2j}^{(j)} - E_t[r_{a,t+2j}^{(j)}]$$

which allows us to decompose the innovation in aggregate return into the sum of the innovation in the market return components. Further, it follows that the conditional

variance of  $r_{a,t+1}$  is:

$$Var_t(r_{a,t+1}) = \sum_{j=1}^J \sigma_{j,t}^2 + \kappa_1^2 \underline{A} \underline{Q} \underline{A}' \quad (\text{D.19})$$

where we define

$$\underline{Q} \equiv \mathbf{E}_t [\boldsymbol{\varepsilon}_{t+1} \boldsymbol{\varepsilon}'_{t+1}]$$

Analogous steps yield the following expression for the market return innovations

$$r_{m,t+1} - E_t[r_{m,t+1}] = \varphi_{d,j} \sigma_{j,t} \odot e_{j,t+1}^d + \underbrace{\kappa_{1,m} \underline{A}_m}_{\beta_{m,\varepsilon}} \boldsymbol{\varepsilon}_{t+1} \quad (\text{D.20})$$

where we define

$$\sigma_{j,t} \odot e_{j,t+1}^d \equiv \sum_{j=1}^J \sigma_{j,t} e_{j,t+2j}^d$$

The innovation in the market return component at level of persistence  $j$  is given by:

$$r_{m,t+2j}^{(j)} - E_t[r_{m,t+2j}^{(j)}] = \varphi_{d,j} \sigma_{j,t} e_{j,t+2j}^d + \underbrace{\kappa_{1,m} \underline{A}_m}_{\beta_{m,\varepsilon}} \boldsymbol{\varepsilon}_{j,t+2j} \quad (\text{D.21})$$

Using the expression (D.20) we can compute the conditional variance of  $r_{m,t+1}$  as follows

$$Var_t(r_{m,t+1}) = \sum_{j=1}^J \varphi_{d,j}^2 \sigma_{j,t}^2 + \kappa_1^2 \underline{A}_m \underline{Q} \underline{A}_m' \quad (\text{D.22})$$

To find the innovations in the stochastic discount factor, we plug the expressions (D.13), (D.14) and (D.15), together with the dynamics for the components of log consumption growth given in equation (D.1) and our guess for the components of price-

consumption ratio solution given in equation (D.7) into equation (D.5) to obtain:

$$\begin{aligned}
m_{t+1} &= \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1} \\
&= \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) (\kappa_0 + \kappa_1 z_{t+1}^a - z_t^a + g_{t+1}) \\
&= \theta \log \beta - \frac{\theta}{\psi} \sum_{j=1}^J g_{t+2j}^{(j)} + (\theta - 1) \left( \kappa_0 + \kappa_1 \sum_{j=1}^J z_{t+2j}^{(j)} - \sum_{j=1}^J z_t^{(j)} + \sum_{j=1}^J g_{t+2j}^{(j)} \right) \\
&= \theta \log(\beta) - \left( 1 - \theta + \frac{\theta}{\psi} \right) \sum_{j=1}^J g_{t+2j}^{(j)} + (\theta - 1) \left( \kappa_0 + \kappa_1 \sum_{j=1}^J z_{t+2j}^{(j)} - \sum_{j=1}^J z_t^{(j)} \right) \\
&= \theta \log(\beta) - \left( 1 - \theta + \frac{\theta}{\psi} \right) \sum_{j=1}^J g_{t+2j}^{(j)} + (\theta - 1) \left( \kappa_0 + \kappa_1 \sum_{j=1}^J z_{t+2j}^{(j)} - \sum_{j=1}^J z_t^{(j)} \right) \\
&= \theta \log(\beta) - \left( 1 - \theta + \frac{\theta}{\psi} \right) \sum_{j=1}^J g_{t+2j}^{(j)} + (\theta - 1) \left( \kappa_0 + \kappa_1 \sum_{j=1}^J A_{0,j} + A_j \sigma_{j,t+2j}^2 - \sum_{j=1}^J A_{0,j} - A_j \sigma_{j,t}^2 \right)
\end{aligned}$$

Finally using the dynamics for our latent factors (D.3) we obtain

$$\begin{aligned}
m_{t+1} &= \theta \log(\beta) + (\theta - 1) \left( \kappa_0 + \kappa_1 \sum_{j=1}^J A_{0,j} + A_j \rho_j \sigma_{j,t}^2 - \sum_{j=1}^J A_{0,j} - A_j \sigma_{j,t}^2 \right) \\
&\quad - \left( 1 - \theta + \frac{\theta}{\psi} \right) \sum_{j=1}^J \sigma_{j,t} e_{g,t+2j}^{(j)} + (\theta - 1) \kappa_1 \left( \sum_{j=1}^J A_j \varepsilon_{t+2j}^{(j)} \right)
\end{aligned}$$

which implies

$$\begin{aligned}
m_{t+1} - E_t[m_{t+1}] &= - \left( 1 - \theta + \frac{\theta}{\psi} \right) \sum_{j=1}^J \sigma_{j,t} e_{j,t+2j}^g + (\theta - 1) \kappa_1 \left( \sum_{j=1}^J A_j \varepsilon_{j,t+2j} \right) \\
&= -\lambda_g \sum_{j=1}^J \sigma_{j,t} e_{j,t+2j}^g - \sum_{j=1}^J \lambda_j \varepsilon_{j,t+2j} \\
&= -\lambda_g \sigma_{j,t} \odot e_{j,t+1}^g - \underline{\lambda}_\varepsilon \varepsilon_{t+1} \tag{D.23}
\end{aligned}$$

where

$$\begin{aligned}
\lambda_g &\equiv \left( \frac{\theta}{\psi} - \theta + 1 \right) = \gamma \\
\underline{\lambda}_\varepsilon &\equiv \kappa_1 (1 - \theta) \underline{A}
\end{aligned}$$

Using the formula (D.18) and the innovation in the SDF (D.23) we obtain the risk premium for the components of consumption claim asset

$$E_t[r_{a,t+2j} - r_{f,t+2j}] + 0.5 \sigma_{r_{a,t,j}}^2 = \lambda_g \sigma_{j,t}^2 + \kappa_1 [\underline{\lambda}_\varepsilon \mathbf{Q}]_j A_j$$



and using the formula for the return on aggregate wealth (D.17) and the innovation in the SDF (D.23) we obtain the risk premium for the consumption claim asset,

$$E_t[r_{a,t+1} - r_{f,t}] + 0.5\sigma_{r_{a,t}}^2 = \lambda_g \sum_{j=1}^J \sigma_{j,t}^2 + \kappa_1 \lambda_\varepsilon \mathbf{Q} \mathbf{A}'$$

where  $\sigma_{r_{a,t}}^2$  is defined in equation (D.19).

Similarly to what we have just done, using the formula (D.21) and the innovation in the SDF (D.23) the premia to the market return components become:

$$E_t[r_{m,t+2j} - r_{f,t+2j}] + 0.5\sigma_{r_{m,t,j}}^2 = \lambda_g \varphi_{d,j} \sigma_{j,t}^2 \rho_j + \kappa_{1,m} [\lambda_n \mathbf{Q}]_j A_j^m$$

and using the formula for the innovations in the market return (D.20) and in the SDF (D.23) the premium to the market return becomes:

$$E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma_{r_{m,t}}^2 = \lambda_g \varphi_{d,j} \sigma_{j,t}^2 \odot \rho_j + \kappa_{1,m} \lambda_n \mathbf{Q} \mathbf{A}'_m$$

where  $\sigma_{r_{m,t}}^2$  is defined in equation (D.19) and  $\rho_j$  is the correlation at level of persistence  $j$  between the stochastic discount factor shocks  $e_{j,t+2j}^g$  and the dividend growth shocks  $e_{j,t+2j}^d$ .

## References

- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2001): “Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics,” *Journal Of The Royal Statistical Society Series B*, 63(2), 167–241.
- BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU (2009): “Expected Stock Returns and Variance Risk Premia,” *Review of Financial Studies*, 22(11), 4463–4492.
- BOUDOUKH, J., M. RICHARDSON, AND R. F. WHITELAW (2008): “The Myth of Long-Horizon Predictability,” *Review of Financial Studies*, 21(4), 1577–1605.
- ORTU, F., A. TAMONI, AND C. TEBALDI (2013): “Long-Run Risk and the Persistence of Consumption Shocks,” *Review of Financial Studies*, 26(11), 2876–2915.
- ZHOU, H. (2010): “Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty,” *SSRN eLibrary*.