Long Run and the Temporal Aggregation of Risks

Ortu F., Tamoni A., Tebaldi C. [1]

March 15, 2010

[1]Fulvio Ortu is at the Department of Finance, Bocconi University. Andrea Tamoni is at the Department of Finance, Bocconi University. Claudio Tebaldi is at the Department of Finance, Bocconi University. The authors thanks C.A. Favero for valuable insights and the IGIER-Asset Pricing Group seminar participants. The usual disclaimer applies. **Correspondence Information:** Fulvio Ortu, `mailto:fulvio.ortu@unibocconi.it`, Andrea Tamoni `mailto:andrea.tamoni@unibocconi.it`, Claudio Tebaldi `mailto:claudio.tebaldi@unibocconi.it`, Bocconi University, Via Roentgen, 1, 20136 MILANO.

Abstract

The long-run risk model introduced by R.Bansal and A.Yaron (2004) assumes the existence of a small predictable component in consumption growth and an elasticity of intertemporal substitution of the representative agent larger than one for the substitution effect to dominate the income one. Previous tests fail to detect predictability in mean consumption growth fluctuations and the estimated value for the elasticity of intertemporal substitution is smaller than one. We argue that these apparent inconsistencies are due to a severe error-in-variables problem generated by the heterogeneity in the persistence levels of shocks to consumption growth. In this paper the original long run risk model is extended introducing a novel persistence based decomposition which classifies innovation shocks on the basis of their half life. Correspondingly the relations between equity return variations, cash flow risk and persistent fluctuations in the consumption mean are disaggregated across different levels of persistence and the complete term structure of risk return tradeoffs is computed. Quite remarkably, the empirical tests performed within this extended setup find evidence of consumption growth predictability, produce sizeable estimated for the equity premia and explain the value premium as an effect of the differential exposure of cash flows to consumption risk.

KEYWORDS: Persistence heterogeneity; Temporal aggregation; Permanent-transitory decomposition; Term-structure of risks; Long-run risks

J.E.L. CLASSIFICATION NUMBERS: C10, E44, G12

1 Introduction

Shocks that impinge an economy can be classified along two competing dimensions: their size as measured by their instantaneous volatility and their persistence as measured by their half life. Short run risk is determined by transitory risk components with large volatility which are expected to have a fast decay over a characteristic time horizon, the half life of the component. Long run risk is dominated by highly persistent components with small size; in the limiting situation of a diverging horizon, valuation is affected only by the permanent components whose effect does not show a decay.

There are many economic arguments which suggest that this classification is a critical issue for asset valuation. Consider for example slow moving structural changes like those induced by technological innovation, by demographic trends or human capital accumulation. In the short run the effects on prices of these economic variables is completely hidden by the myopic and volatile reaction of markets to the undifferentiated flow of incoming information. As the valuation horizon increases, the effect of transitory shocks is averaged away while persistent structural trends emerge as the driving forces of long run expectations and play a pivotal role in the rational valuation of assets. The accurate reconstruction of the entire term structure of risk return tradeoffs requires therefore to decompose aggregate shocks while controlling simultaneously for volatility and for persistence.

This paper proposes an extension of the long run valuation model of Bansal and Yaron (2004) explicitly designed to optimally account for the heterogeneity of shock persistence levels. For this reason we introduce a novel spectral decomposition of innovation shocks which extends the spectral approach of Hansen, Heaton, and Li (2008) and Beveridge and Nelson (1981). On the empirical side, our new decomposition offers a natural framework to discuss the impact of the long run consumption growth predictability on asset valuation, a central issue in the analysis of long run risks. Our findings prove that after controlling for persistence many puzzling results which have been obtained using traditional methods in the field of linear long horizon valuation models find a natural explanation. In particular we find new empirical evidence of predictability of specific persistent components of consumption growth. These consumption growth components are characterized by a negligible size in volatility but a high level of persistence and account for a sizable equity risk premia. These findings are not in contradiction with the empirical findings in Beeler and Campbell (2009): in fact the variance of very persistent shocks in themselves is so small that it does not generate any predictability on aggregate consumption due to an error-in-variables problem. In the specific analysis of the long run consumption risk valuation model we identify a predictable consumption growth component characterized by a volatility smaller than 5% and by a degree of persistence compatible with a half life longer than 8 years. Due to its negligible size, the identification of this component crucially hinges on the use of a pre-filtering procedure which is performed applying a new persistence based decomposition. As a robustness check, we verify that the detection of the predictable component emerges using different, alternative filtering procedures. We concentrate our analysis on the fluctuation in mean of consumption growth predictability, although the same analysis could be applied including also fluctuations in consumption growth volatility.

The new decomposition proposed in this paper can be considered as a refined permanent-transitory decomposition where the transitory shock is further split in orthogonal components. This new decomposition provides an effective method to represent a time series as a linear combination of uncorrelated innovations classified by their level of persistence and their time of arrival. Applying this decomposition to the long run valuation model allows to disentangle across time scales the interaction between stochastic discounting, cash flow risk and consumption growth fluctuations. As a result it is possible to define a whole term structure of consumption risks. Shocks with different levels of persistence which have the same arrival time define a term structure of innovation shocks. The knowledge of the term structure of innovations allows us to reconstruct the risk-return tradeoffs for different holding periods as in Campbell and Viceira (2005) or in Bandi and Perron (2008).

In the standard long run risk picture, see e.g. Bansal, Dittmar, and Lundblad (2005) and Bansal, Dittmar, and Kiku (2007), the cross sectional variability of risk premia finds an explanation in terms of the differential exposure of the cash flows to consumption risk. Within our extended valuation framework a test on the portfolios formed on book-to-market (BE/ME) portfolios provides an empirical confirmation of the picture. The duration profile of cash flow risk is defined as the vector of sensitivities which measures the reaction of cash flow to consumption growth innovations at different level of persistence. Empirical analysis proves that the duration profiles of the value and growth portfolios are significantly different, and the value premium finds an explanation in terms of the their differential exposure to consumption growth at different horizons.

Our empirical evidence sheds new light and opens new questions also on the agent preferences about intertemporal substitution of consumption which are elicited assuming the market equilibrium. A necessary condition for the long-run risk to be priced is that the income effect and substitution effects should not cancel each other. In particular, the long-run risk paradigm requires the substitution effect to dominate the income one and thus necessitates an intertemporal elasticity of substitution (IES) greater than one. In this way a positive shock to expected consumption growth provides a strong incentive to reduce consumption and increase demand for financial assets with the final effect that stock prices rise. In our empirical examination of the IES we find that if we focus only on low persistence consumption growth components then results are consistent with a representative agent for which the wealth effect dominates the intertemporal substitution one. On the contrary the determination of the IES based on high persistence consumption growth components detects an elasticity parameter larger than one, consistent with the long run risk picture where the substitution effect dominates the income effect. This evidence explains the contradictory results found by different authors[1], and seems to be consistent with a framing effect: the representative agent shows different preferences when facing intertemporal substitution of short or long term consumption risk.

The paper is organized as follows. The next subsection concludes the introduction with a review of the literature. The Section 2 discusses a simple model for consumption growth to explain the effects of persistence heterogeneity and the errors-in-variable problem. This example motivates the needs to find an optimal and viable method to separate a time series in components characterized by their level of persistence. Section 3 introduces such a method, which we call the persistence based decomposition, and provides the definition of the term structure of risk. Section 4 extends the Long-Run valuation framework using the persistence based decomposition. Section 5 explores the main empirical findings obtained applying the new persistence based decomposition.

1.1 Literature Review

Our research contributes to the fast growing stream of literature which looks at the long run regime to explain many of the inconsistencies which affect predictions of dynamic asset pricing models: Parker and Julliard (2005) prove that the consumption CAPM performs better at predicting the cross sectional differences if one uses long run consumption growth rates instead of short run ones. In their seminal contribution to long run risk valuation Bansal and Yaron (2004) (BY04 henceforth) explain stock price variations as a response to small persistent fluctuations in the mean and volatility of aggregate consumption growth by an agent with elasticity of substitution greater than one and recursive preferences a la Epstein Zin (Kreps and Porteus 1978, Epstein and Zin 1989, Epstein and Zin 1991). In a similar set-up Backus, Routledge, and Zin (2009) show a positive correlation between future economic growth and equity returns. Kaltenbrunner and Lochstoer (2007) explain a high price of risk and a low volatility of consumption growth with a low coefficient of risk aversion in a one-sector DSGE model in which long-run risk arises endogenously. Bansal, Dittmar, and Lundblad (2005) show that long-run risks in cash flows are an important risk source in accounting for asset returns and Bansal, Dittmar, and Kiku (2007) show that economic restrictions of cointegration between asset cash flows and aggregate consumption have important implications for cross-sectional variation in equity returns, particularly for long horizons. Hansen, Heaton, and Li (2008) analyze the pricing implications of the exposure of cash flows to macroeconomic risks in the limit of a diverging valuation horizon. Campbell

[1]Hall (1988) and Campbell and Mankiw (1989) estimated an extremely small value of IES. Campbell (2003) summarizes these results and finds similar patterns in international data. Attanasio and Weber (1993) and Beaudry and van Wincoop (1996) have found higher values for IES using disaggregated cohort-level and state-level consumption data. Vissing-Jorgensen (2002) points out that many consumers do not participate actively in asset markets; using household data she finds a higher value for among asset market participants.

and Viceira (2005) describe the term structure of the risk-return tradeoffs which is obtained varying the holding periods when returns are predictable. Bandi and Perron (2008) identify a remarkable relation between the past market variance aggregated over a time period h and the future excess returns. Similarly Lettau, Ludvigson, and Wachter (2007) find a strong correlation between low-frequency movements in macroeconomic volatility and low-frequency movements in the stock market. Garleanu, Panageas, and Yu (2009) focus on the impact of major technological innovations and real options on consumption and the cross-section of asset prices. These innovations are assumed to occur at a very low frequency (about 20 years), and are shown to carry over into a small, highly persistent component of aggregate consumption. Alvarez and Jermann (2005) analyze the persistence of the marginal utility of wealth applying a multiplicative permanent-transitory decomposition in line with the methodology proposed by Hansen and Sheinkman (2009).

The long-run risk theoretical framework has motivated many empirical tests on the presence of long run consumption risk (see Bansal, Dittmar, and Lundblad (2005)) and long run volatility risk (Bansal, Yaron, and Kiku (2007), Bandi and Perron (2008), Bollerslev, Tauchen, and Zhou (2009)) in the data. Recent findings, see e.g. Constantinides and Ghosh (2008) and Beeler and Campbell (2009) have shown that empirical evidence on long run predictable consumption growth is often controversial. Bansal, Dittmar, and Kiku (2007) and Hansen, Heaton, and Li (2008) suggest that cointegration between consumption and dividends can be a source of long run consumption risk, while Beeler and Campbell (2009) and Calvet and Fisher (2007) indicate consumption growth volatility as a possible channel of transmission of long run risk.

Structural weaknesses of econometrics of stationary time series when dealing with heterogeneity of persistence effects have certainly played a role in reducing the discriminatory power of these tests. The main methodological contribution of this paper is a new approach to the identification of low frequency structural relations. The relevant time series are decomposed in a sequence of uncorrelated shocks which are classified by the time of their arrival, as in the standard Wold decomposition, and by an additional index which measures their level of persistence. In this regard our work is close to Calvet and Fisher (2007) who investigate the role of heterogeneity in persistence of volatility in a partial equilibrium set-up by means of non linear regime switching multifractal models.

According to our classification of shocks based on both their level of persistence and their time of arrival, a specific component accounts for those shocks which do not show a decay within any time interval smaller than the observation sample. This component is the permanent component as defined by Beveridge and Nelson (1981), and the new persistence based decomposition can be interpreted as a refined permanent-transitory decomposition where the transitory shocks are further split in orthogonal components. The effective viability of the proposed technology as an empirical investigation tool is demonstrated by the improved quality of these statistical tests.

While traditional linear time series analysis is based on Fourier spectral analysis, our extended approach is based on spectral multiresolution analysis (see Daubechies (1992), Daubechies (1990), Mallat (1989a) and Mallat (1989b) for an introduction to the topic and more recently Gencay and Fan (2008), Gencay and Gradojevic (2009), Gencay, Selcuk, and Whitcher (2001) for a specific analysis of unit roots and filtering using multiresolution analysis). Far from being the first attempt to use multiresolution methods in economics and finance, it will be shown that the use of a multiresolution filter can be motivated by a mild variation of the spectral approach proposed in Hansen, Heaton, and Li (2008) and Hansen and Sheinkman (2009) and provides an optimal solution to analyze long-term risk return profiles and low frequency structural relations among their economic determinants.

2 A motivating example

In their seminal paper Bansal and Yaron propose the following DGP for consumption:

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma\eta_{t+1} \\ x_{t+1} &= \rho x_t + \phi_e \sigma e_{t+1} \\ e_{t+1}, \eta_{t+1} &\sim i.i.d.N(0,1) \end{aligned} \tag{1}$$

where g_{t+1} is the growth rate of consumption, and x_t is the state variable of the model. As suggested by Beeler and Campbell (2009), it is natural to test the model by evaluating the ability of the log price-dividend ratio to predict long-run consumption.

A key requirement of BY04 is that there exists a predictable component of expected consumption growth x_t which is highly persistent and small in volatility. This factor x_t turns to be the driving factor of stock price variability relative to dividends. In fact at equilibrium the log price-consumption, z_t, and price-dividend ratio z_t^m, are linear functions of the conditional mean of consumption growth namely

$$\begin{aligned} z_t^m &= A_{0,m} + A_{1,m} x_t \\ z_t &= A_0 + A_1 x_t \end{aligned} \tag{2}$$

where the sensitivity parameters $A_0, A_1, A_{0,m}, A_{1,m}$ can be explicitly computed inserting the Campbell-Shiller (1988) approximation and imposing first order Euler conditions. Since the focus of our analysis is on the predictability of consumption growth, volatility of consumption will be assumed to be constant.

A preliminary investigation of the consumption growth and the price-dividend ratio series highlights that the main difference between the consumption growth and price-dividend series lies in their persistence. Figure 1 shows the two data series. The price-dividend ratio is an extremely persistent series[2] and it is often modeled by a close-to-unity process whereas the consumption growth resembles more closely a white noise process.

[Insert Figure 1 about here.]

This simple observation casts some doubts on the effective ability of the data generating process proposed in equation (1) to describe the predictability pattern, if any, arising in two time series which show very different persistence levels.

In order to gain some intuition on the effects of for persistence heterogeneity it is sufficient to consider a simple modification of the data generating process: the price-dividend ratio

[2] In fact since 1872, the price-dividend ratio has crossed its mean value 29 times, with intervals between crossings ranging from one year to twenty years (the twenty-year interval being between 1955 and 1975). See also Campbell and Shiller (2001).

and the consumption growth are assumed to be obtained as a sum of two components. The first component, which we call the characteristic component, will be common across the two processes and generating covariation and predictability effect. This is similar to BY04 setup where a process x_t both drives the dividend price ratio and the consumption growth. In addition we add to each process a second component, which we call idiosyncratic. This component is not present in the BY04 set up and is necessary to account for those components of consumption growth which do not covary with the price-consumption ratios and have a different level of persistence with respect to the characteristic (predictable) component.

More formally we assume the following dynamics for g_t and z_t^m:

$$\begin{aligned} g_{t+1} &= x_t^{(char)} + \widetilde{x}_t^{(idyo)} \\ z_t^m &= x_t^{(char)} + x_t^{(idyo)} \end{aligned} \tag{3}$$

where $\widetilde{x}_t^{(idyo)}$ is the consumption growth idiosyncratic components, $x_t^{(idyo)}$ represent the idiosyncratic components of the price dividend and where $x_t^{(char)}$ is the common component across the two processes. This component is therefore responsible for predictability. The key aspect of our model is that each components $\widetilde{x}_t^{(idyo)}, x_t^{(idyo)}$ differs from the other due to its level of persistence.

According to BY04 we expect $x_t^{(char)}$ to be highly persistent and to explain a small fraction of the total variance of the consumption growth. Hence the predictability effect manifests itself on a characteristic time scale much longer than the interval of observation and for a component whose size as measured by instantaneous (single period) volatility can be very small if compared with the instantaneous total volatility of consumption growth and price dividend time series. For this reason the problem of the filtering of $x_t^{(char)}$ out of pd_t becomes a critical issue for empirical analysis.

In fact given that the covariation relation holds only for this small component of consumption growth $x_t^{(char)}$ then the predictive effect could be largely underestimated in the absence of a pre-filtering procedure which extracts $x_t^{(char)}$ out of pd_t. This can be easily understood considering the following simulation test. Simulate the trajectories of price-dividend ratio and consumption growth based on the DGP in equation (3) by assuming that the characteristic level of persistence of the predictable component $\tau\left(x_t^{(char)}\right) \in (8, 16)$[3] years while its variance explains less than 10% of the variability in the time series [4]. Notice that $\tau\left(x_t^{(char)}\right)$ is different from the standard half life usually estimated for the price-dividend ratio, which is

[3] $\tau\left(x_t^{(char)}\right) \equiv -\log(2)/\log(\rho)$

[4] These assumptions are motivated by a preliminary analysis of the US data. Note that in the quarterly time sample 1947Q2-2007Q4 the ratio of the unconditional variance of log consumption growth to log price-dividend ratio is $(1.09)^2/(0.38)^2 \approx 8$. We simulate a process x_t whose unconditional variance is $(v)^2/(1-\rho^2) = 0.1$ whereas the unconditional variance of the process g_t is around 1, thus yielding a ratio of 7.

$\tau(pd_t) \approx 5$ years [5]. This makes sense because the estimation of the half life estimated on the price-dividend time series will be determined by the components of large size (as measured by volatility) while, on the contrary, the effect of the characteristic component is small.

Figure 2 and Figure 3 show a typical realization of these processes. The top panel of Figure 2 displays a realization of the series z_t^m and of its components $x_t^{(char)}$ whereas the top panel of Figure 3 we give an example of the process g_t. We can first observe that our DGP is able to produce both a very slow mean reverting process thus resembling the price-dividend behavior and a close to white noise process similar to the consumption growth one.

[Insert Figure 2 and 3 about here.]

The BY04 model given by equation (1) and (2) would suggest to run the following regression:

$$g_{t+1} = \beta_0 + \beta_1 z_t^m + w_{t,t+1} \tag{4}$$

The results of applying this econometric model on our simulated data is shown in Table 1 where the predictive effect of z_t^m on g_{t+1} is not statistically significant in the simulated data. Hence this statistical regression method fails to detect the structural relation which, by construction, is in the data. Little thought can convince the reader that the explanatory power of $x_t^{(char)}$ is removed because it is dominated by the non predictive, large variance components $\widetilde{x}_t^{(spur)}$, $x_t^{(spur)}$. In the Appendix A we report the detailed explanation of the error-in-variables problem which explains the inability of standard linear statistical inference to detect the corresponding predictive relation between consumption growth and dividend-price ratio which holds on a specific set of frequencies. The top panel of Figure 3 gives an example of the process g_t before and after adding the predictable components $x_t^{(char)}$. We can note in the top panel of Figure 3 that the processes g_t with and without the *persistent* component $x_t^{(char)}$ look almost indistinguishable. The explanation can be found in the bottom panel of Figure 3 where we show (a realization of) the processes g_t together with $x_t^{(char)}$. Observe that the series $x_t^{(char)}$ is very small and it is overwhelmed by $\widetilde{x}_t^{(idyo)}$.

[Insert Table 1 about here.]

In fact, the necessity to pre-filter the data in the analysis of consumption-income relations can be traced back to Friedman as highlighted by the following observation made by Engle (1974): " If the error component [6] is assumed to be confined to a particular frequency band, then a natural procedure is to eliminate that frequency band from the regression. This is essentially the technique used by Friedman in defining permanent income as a moving

[5] In annual data we estimate for the log price-dividend ratio an autoregressive coefficient of $\hat{\rho} = 0.89$ whereas for quarterly data we estimate $\hat{\rho} = 0.97$.

[6] In our case $x^{(spur)}$.

average of measured income. This filter eliminated high frequency noise which he calls the transient component of income".

The first issue addressed by this paper is the selection of an optimal procedure to filter the predictable component $x_t^{(char)}$ out of pd_t and out of g_t.

In fact traditional filtering based on Fourier decomposition (see Christiano and Fitzgerald (2003), Canova (1998), Yu (2009) and Garleanu, Panageas, and Yu (2009)) could be a viable choice, since the persistence level of a component could be controlled selecting a narrow band of frequencies in the spectrum. However such an approach would imply a great loss of information in the time representation of the series. It is well known that a frequency representation of a time series generates a complete loss of information regarding the temporal localization of the shocks and of the price response.

If the set of characteristic frequencies were known then we could define the corresponding narrow band component of g_t as $g_t^{(char)}$ and run the regression:

$$g_{t+1}^{(char)} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_t^{(char)} + \widetilde{\epsilon}_{t,t+1}$$

The point estimate is reported in Table 2. We are now able to detect the predictive effect of $x_t^{(char)}$ on $g_{t+1}^{(char)}$ with a statistically significant $t - stat = 3.318$ in the simulated data.

[Insert Table 2 about here.]

According to BY04 the persistence of a small predictable component drives expectations on future long run consumption movements and therefore generates relevant effects on valuation. The second important question addressed in this paper is whether the same effect is produced when a specific consumption growth components is selected through a filtering procedure. To this aim the BY04 valuation model will be formally extended in order to account for a persistence based classification of consumption risk shocks.

Last but not least, an extensive empirical analysis will test whether such heterogeneous persistence long-run risk equilibrium finds an empirical confirmation.

The next section introduces a new filtering procedure which will allow to filter the predictable component $x_t^{(char)}$ out of pd_t and out of g_t.

3 The persistence based decomposition of the time series.

Recall that the presence of predictable components of consumption is a necessary condition to accept a long run risk picture. The motivating example discussed in the previous section highlights the importance of identifying and classifying components of consumption growth with different levels of persistence. In this section we introduce a procedure which we call the Persistence Based Decomposition (PBD) for a time series. The PBD extends the Beveridge Nelson (1981) permanent-transitory decomposition in the sense that the transitory part is represented as a linear sum of components classified by their level of persistence.

Quite remarkably its construction is based on a well known object in financial econometrics: the moving average filter. From the point of view of spectral methods, a moving average filter is the prototype of a multiresolution filter, a sophisticate class of filters whose use is widespread in signal analysis. In terms of econometrics of time series a multiresolution filter is an aggregation operator acting on the space of the time series of observations[7].

As already stated, the starting point of our analysis is the choice of the (dyadic) moving average as the aggregation operator:

Definition 1 *The dyadic (sample) mean aggregation operator acting on the time series of observations up to time t, $\mathbf{x}_t = \{x_{t-k}\}_{k \in 0,..,+\infty}$, is defined by:*

$$M : \mathbf{x}_t \to \boldsymbol{\pi}_t^{(1)} \triangleq M\mathbf{x}_t = \left\{\pi_{t-k}^{(1)}\right\}_{k \in 0,..,+\infty}$$
$$\pi_{t-k}^{(1)} = \frac{x_{t-k} + x_{t-k-1}}{2}$$

Using the operator M it is possible to define a sequence of time series whose elements can be computed recursively:

$$\pi_{t-2^J k}^{(J)} = \left(M\boldsymbol{\pi}_t^{(J-1)}\right)_{t-2^J k} = \frac{\pi_{t-2^J k}^{(J-1)} + \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2} \tag{5}$$

and the element $\pi_{t-2^J k}^{(J)}$ corresponds to the sample mean over a window of past observations with size 2^J:

$$\pi_{t-2^J k}^{(J)} = \frac{1}{2^J} \sum_{p=0}^{2^J-1} x_{t-2^J k-p}$$

In words, iterated applications of the dyadic aggregation operator maps the original time series into the new time series $\pi_t^{(J)}$ where each element is the $h-$period moving average with $h = 2^J$ and time is consistently scaled by a factor 2^J.

Note that the elements of $\boldsymbol{\pi}_t^{(J)}$ are 2^J-period moving averages of the original time series, thus the fluctuations generated by a shock whose characteristic time scale is smaller that 2^J leave the elements of $\boldsymbol{\pi}_t^{(J)}$ unaffected. Equivalently, in the frequency representation, a

[7]For the basic definitions and notational conventions employed we refer to Brockwell and Davis (1996).

2^j-period moving average operator works as a low band pass filter which removes all those components whose frequency is smaller than 2^j.

The effect of the moving average filter on a generic time series is easily visualized in terms of the Fourier spectrum of the time series. The top subplot of Figure 4 shows the Fourier spectrum of the aggregate consumption growth time series; the shadowed region in the bottom left panel identifies the part of the spectrum which survives after the first application of the moving average filter, namely the spectrum of $\pi_{t-k}^{(1)}$. The unshadowed, high frequency part of the spectrum is removed by the application of the filter. In fact a moving average is a low band pass filter. Note that the spectrum is cut at a frequency $f_{max}/2$ ($f_{max} = 2\pi/h$ where h is the time discretization step of the time series and is the highest frequency in the spectrum of $\mathbf{x}_t$) and is proportional to the inverse of the averaging window size.

[Insert Figure 4 about here.]

On the basis of the illustration it is intuitive to verify that, the difference between $\boldsymbol{\pi}_t^{(J-1)}$ and $\boldsymbol{\pi}_t^{(J)}$ identifies the component of the original time series with half life in the interval $\left[2^{J-1}, 2^J\right)$ and with a Fourier spectrum localized in the finite interval of frequencies $\left[\frac{f_{max}}{2^j} \; \frac{f_{max}}{2^{j-1}}\right)$. This intuition suggests the following:

Definition 2 *Assume the following notational convention:* $\boldsymbol{\pi}_t^{(0)} = \mathbf{x}_t$. *The time series* $\delta^{(J)} = \left\{\delta_{t-2^J k}^{(J)}\right\}_{k\in\mathbb{N}}$ *with elements*

$$\delta_{t-2^J k}^{(J)} = \pi_{t-2^J k}^{(J-1)} - \frac{\pi_{t-2^J k}^{(J-1)} + \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2} \quad (6)$$

$$= \frac{\pi_{t-2^J k}^{(J-1)} - \pi_{t-2^{J-1}(2k+1)}^{(J-1)}}{2} \quad (7)$$

is called the J-th detail component of the time series $\mathbf{x}_t$. *The index J is called the level of persistence of the component.* 2^J *is sometimes called the characteristic "scale" of the J-th, component.*

Figure 4 inset 3 shows the spectrum of the details δ_t^1 as extracted by the time series for consumption growth.

Thanks to the recursive nature of the definition of $\boldsymbol{\pi}_t^{(J)}$ and $\delta_t^{(J)}(\mathbf{x}_t)$ it is immediate to verify that:

Corollary 3 *The element x_t of the original time series stopped at time t, $\mathbf{x}_t$, can be decomposed as:*

$$x_t = \sum_{j=1}^{J} \delta_t^{(j)} + \pi_t^{(J)} \tag{8}$$

Continuing the exemplification in Figure 4 the insets in the second column show the Fourier spectra of each term (namely the details $\delta_t^{(1)}$ and $\delta_t^{(2)}$ and of the persistent component $\pi_t^{(2)}$) obtained by a decomposition truncated at level $J = 2$.

In the sequel of the paper we will refer to $\boldsymbol{\pi}_t^{(J)}$ as the component of x_t with level of persistence larger than J while the vector obtained joining the sequence of details $\left\{\delta_t^{(j)}\right\}_{j=1,..,J}$ with the persistent component $\boldsymbol{\pi}_t^{(J)}$ will be denominated the term structure of risk truncated at level J extracted at time t from $\mathbf{x}_t$.

The following theorem proves the convergence for the sequence obtained by the iteration of the dyadic averaging operation:

Theorem 4 *Consider a linear $I(0)$ time series $\mathbf{x}_t = \{x_{t-k}\}_{k\in 0,..,+\infty}$ as defined in Hayashi pg.563 such that $x_t = \psi(L)\varepsilon_t$. Let $E_t[x_{t+1}] = \mu$, $Var_t[x_{t+1}] = \sigma^2$ and consider the decomposition (8). Then:*

1. *the sequence of random variables $\{\pi_t^{(J)}\}_{J=0}^{+\infty}$ converges in mean square norm to a constant equal to the conditional mean. More formally*

$$\pi_t^{(\infty)} \equiv \lim_{J\to+\infty} \pi_t^{(J)} = \mu$$

with

$$\lim_{j\to\infty} \sqrt{Var\left[\sqrt{2^j}\pi_t^{(j)}\right]} = \psi(1)$$

2. *the time series formed by the details $\delta_t^{(j)}$ as defined by eq.(6) is stationary and has zero mean for all j.*

In conclusion the following decomposition holds for $\mathbf{x}_t$:

$$x_t = \sum_{j=1}^{+\infty} \delta_t^{(j)} + \pi^{(\infty)} \tag{9}$$

where the convergence has to be understood in mean square norm.

Proof. The proof of this theorem is provided in Appendix B. ■

The construction of the decomposition is obtained by an application of the abstract Wold theorem (Wold 1938). [8]

On the basis of the above discussion the detail $\delta_t^{(j)}$ is the component of the original time series which includes an interval of characteristic time scales ranging from 2^{j-1} to 2^j. The theorem states that it is possible to decompose x_t as a linear sum of detail components $\delta_t^{(j)}$ with increasing degree of persistence $j = 1, ..., +\infty$ and the Beveridge Nelson permanent component $\pi_t^{(\infty)}$, using only the set of observations up to time t. By construction each component $\delta_t^{(j)}$ is stationary and has zero mean. Observe also that the element $\pi_t^{(\infty)}$ is a deterministic constant time trend which corresponds to the "zero frequency" component of the spectrum, in fact the interval $\left[\frac{f_{max}}{2^j} \frac{f_{max}}{2^{j-1}}\right)$ shrinks to the zero point set $\{0\}$ as $j \to +\infty$ as shown in Figure 5. The above formulation of the Persistence Based Decomposition decomposes a stationary process x_t, while the standard Beveridge Nelson (1991) decomposition applies to processes integrated of order $1.y_t$. The connection is immediately obtained if the decomposition is applied to time series formed by the first differences $x_t = \Delta y_t$. In fact in the present framework, we will follow the procedure used by BY04 and consider as observed stationary variables first differences of consumption and dividends, i.e. consumption and dividend growth. It must be remarked that the above decomposition turns to be a practical instrument to identify and separate the (stochastic+deterministic) trend component from the stationary components. The formulation of the decomposition which applies to $I(1)$ processes and a statistical analysis of its performance in detecting unit root and cointegration relations are discussed in the on-line Appendix (Ortu, Tamoni, and Tebaldi 2010). In fact in this setup the "zero frequency" component $\pi^{(\infty)}$ is shown to correspond to the (first difference of the) Beveridge Nelson stochastic trend.

[Insert Figure 5 about here.]

Of course the infinite PBD must be estimated from the observation of a finite sample of observations. While we leave for future research a systematic investigation of the asymptotic inference theory for the PBD, in the next proposition we introduce the basic computational scheme which will be used to identify the layer of details with the maximum level of persistence which can be identified given T observations.

Proposition 5 *Consider the finite set of observations* $\{x_s\}_{s=0,..,T}$ *of the time series* $\mathbf{x}$ *and let* $J_{\max} = \lfloor \log_2(T) \rfloor$ *then*

1. *each detail component with a level of persistence* $j \geq J^{\max}$ *cannot be separated from the permanent component on the basis of the finite set of observations;*

[8] In the on-line Appendix (Ortu, Tamoni, and Tebaldi 2010) we review the theorem and its relation with the traditional Wold decomposition which is used in stationary time series analysis.

2. $\pi_T^{(J_{\max})}(\omega)$, *the component with persistence larger or equal to* $J_{\max}$ *is said to be the "scale component" and is identified by:*

$$\pi_T^{(J_{\max})} = \frac{1}{2^{J_{\max}}} \sum_{s=1}^{2^{J_{\max}}} x_{T+s-2^{J_{\max}}} \quad \forall j \geq J_{\max}$$

when $T = 2^{J_{\max}}$ $\pi_T^{(J_{\max})}$ *corresponds to the unconditional sample mean over the observation interval.*

Proof. By definition the computation of the details with level of persistence $j \geq J_{\max}$ requires at least $2^{J_{\max}+1}$ independent observations. ■

The use of the above definitions provides an explicit procedure to construct the vector with $J^{\max}+1$ components formed by the $J^{\max}$ details and the scale component. Details are numbered by the increasing persistence index j while we will conventionally denote the scale component with the index -1. This computational approach is known in multiresolution analysis under the name of "redundant transformation" (see Renaud, Starck, and Murtagh (2005)). In the Appendix C.1 we report the simple recursive procedure which is practically used. It has to be highlighted that the attribute "redundant" denotes the fact that the elements computed using this transformation are not independent. The use of the redundant transformation significantly increases the sample size of observations, in particular for the time series of details with high level of persistence, however the lack of independence between the observed sequences of details poses serious statistical concerns. These concerns are very similar to those generated by the use of overlapping windows in standard time series analysis. Their statistical effects will be carefully discussed during the analysis of the empirical results.

A computational method which allows to avoid any problem and generates only detail components which are linearly uncorrelated is based on the simple observation that the linearity of the truncated persistence based decomposition implies the existence of a linear transformation which transforms the original time series $\mathbf{x} = \{x_{t-k}\}_{k\in\mathbb{N}}$ to a sequence of independent term structure of risks. A key result of multiresolution spectral analysis (see e.g. Daubechies (1992)) states that this linear mapping can be explicitly represented for any filter using a matrix of size 2^J, $\mathcal{T}^{(J)}$. While we refer to the above references for the construction of $\mathcal{T}^{(J)}$ in the general case, to clarify our approach we exemplify its construction for $J = 2$. In this case the matrix $\mathcal{T}^{(2)}$ that maps the time series in the term structure of risks is given by:

$$\mathcal{T}^{(2)} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{10}$$

To see this we first group the variables $\{x_{t-k}\}_{k\in\mathbb{N}}$ in (disjoint)blocks of length 2^2. The first block is:

$$\underline{X}_t^{(2)} = \begin{bmatrix} x_{t-3} \\ x_{t-2} \\ x_{t-1} \\ x_t \end{bmatrix}$$

Then let $\widetilde{\underline{X}}_t^{(2)}$ be the vector obtained by stacking the sequence of details $\{\delta_t^{(2)}, \delta_{t-2}^{(1)}, \delta_t^{(1)}\}$ below the persistent component $\left\{\pi_t^{(2)}\right\}$. Using equation (5) and (6) we get

$$\widetilde{\underline{X}}_t^{(2)} = \begin{pmatrix} \pi_t^{(2)} \\ \delta_t^{(2)} \\ \delta_{t-2}^{(1)} \\ \delta_t^{(1)} \end{pmatrix} = \begin{pmatrix} \frac{x_t+x_{t-1}+x_{t-2}+x_{t-3}}{4} \\ \frac{x_t+x_{t-1}-x_{t-2}-x_{t-3}}{4} \\ \frac{x_{t-2}-x_{t-3}}{2} \\ \frac{x_t-x_{t-1}}{2} \end{pmatrix}$$

and given that the following must hold true:

$$\widetilde{\underline{X}}_t^{(2)} = T^{(2)}\underline{X}_t^{(2)}$$

the expression (10) is found.

Using this transformation it is now possible to obtain the law of motion of the details given the law of motion of the time series. In particular consider an AR(1) process:

$$x_{t+1} = \rho x_t + w_{t+1} \tag{11}$$

where $w_t \sim N(0, \sigma_w^2)$. From (11) it follows that we can rewrite the dynamics of x_t as:

$$\underline{X}_{t+4} = \bar{\rho}\underline{X}_t + \bar{w}_{t+4} \tag{12}$$

where

$$\bar{\rho} = \begin{pmatrix} \rho^4 & 0 & 0 & 0 \\ 0 & \rho^4 & 0 & 0 \\ 0 & 0 & \rho^4 & 0 \\ 0 & 0 & 0 & \rho^4 \end{pmatrix}$$

and the noise vector is given by

$$\begin{aligned} \bar{w}_t &= B\mathbf{w}_t \\ B &= \begin{pmatrix} \rho^3 & \rho^2 & \rho & 1 & 0 & 0 & 0 \\ 0 & \rho^3 & \rho^2 & \rho & 1 & 0 & 0 \\ 0 & 0 & \rho^3 & \rho^2 & \rho & 1 & 0 \\ 0 & 0 & 0 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \\ \mathbf{w}_t &= [w_{t-4}\ w_{t-3}\ \ldots\ w_t]^T \end{aligned}$$

satisfying

$$E[\bar{w}_t] = 0$$
$$E[\bar{w}_t \bar{w}_t^T] = \sigma_w^2 BB' \equiv \bar{Q}$$

If we now apply the matrix linear transformation to both right and left hand sides of (12) we can obtain the desired dynamics (and in this particular case a vector $AR(1)$) for the decomposed quantities. If we again focus on the $J = 2$ case we obtain

$$\underline{\widetilde{X}}_{t+4} = M_t^g \underline{\widetilde{X}}_t + e_{t+4}$$

where

$$M^g = T^{(2)} \bar{\rho} \left(T^{(2)}\right)^{-1}$$
$$E[e_t] = 0$$
$$E[e_t \cdot e_t'] = T^{(2)} \; \bar{Q} \; \left(T^{(2)}\right)^{-1}$$

The matrix linear transformation can be easily extended to an arbitrary index J. The state vector $\underline{\widetilde{X}}_t^{(J)} = T^{(J)} \underline{X}_t^{(J)}$ has 2^J elements, the first element is denoted by $\widetilde{X}_{-1,t}$ and selects the trend component (the scale function in the multiresolution terminology); the remaining elements of the vector contain the details necessary to reconstruct the vector $\overline{X}_t$ in decreasing order: the elements with index in the interval $2^{J-j} + 1 < i \leq 2^{J-j+1}$ identify the details with level of persistence j:[9]

$$\begin{aligned} \underline{\widetilde{X}}_t^{(J)} &= \left(\widetilde{X}_i\right)_{i=2^{J-j}+n-1} = \widetilde{X}_{j,t-2^j(n-1)} \\ j &= 1,..,J \quad n = 1,..,2^{J-j} \end{aligned}$$

where j denotes the level of persistence, while n determines the sequential ordering of innovation details with level of persistence equal to j. The details with highest order n are those which are necessary to reconstruct the details of the time series in the distant past.

The lower the level of persistence, the larger the number of details which are necessary to reconstruct the original time series. If the inverse transformation is applied to the vector of details excluding the elements with degree of persistence smaller than a given level, say j_0, the time series which is obtained corresponds to the original time series aggregated (averaged) over intervals of time of length 2^{j_0}.

The representation of the original time series is the so called non-redundant representation of a time series, in fact the complete sequence of $\left\{\underline{\widetilde{X}}_{t-k}^{(j)}\right\}_{k\in\mathbb{N}, j=1,..,J,-1}$ is necessary to reconstruct the original time series and the elements are linearly independent.

[9] In the wavelet literature it is customary to classify the details using double indexing: $g_{j,t-2^j(n-1)} = (g_t)_{j,n}$

3.1 Consumption and Dividend growth

This novel decomposition is useful in those cases where the original time series is characterized by a high dispersion of persistence levels i.e. the time series results from the aggregation of a large number of stationary components with different degrees of persistence.

In this section the persistence based decomposition is applied to the three time series which turn to be important for long run consumption based valuation. In particular we apply the decomposition to the time series of consumption growth g_t, dividend growth gd_t and price dividend pd_t. Following BY04 and Beeler and Campbell (2009)[10] we use data on US nondurables and services consumption from the Bureau of Economic Analysis. All variables are measured in real terms. We consider a postwar quarterly US series over the period 1947:Q2-2007:Q4 and for robustness a long-run annual series over the period 1930-2006.

After performing the decomposition we fit the time series of the vectors of details $\delta_t^g = [g_{j,t}]_{j=1,\ldots 8}$, $\delta_t^{gd} = [gd_{j,t}]_{j=,1,\ldots 8}$ and $\delta_t^{dp} = [dp_{j,t}]_{j=1,\ldots 8}$using a VAR(1) with 8 detail components:

$$
\begin{aligned}
\delta_{t+1}^g &= \rho^g \delta_t^g + \varepsilon_{t+1}^g & \varepsilon_t^g &\sim N\left(\mathbf{0}, \mathbf{Q}^g\right), \\
\delta_{t+1}^{gd} &= \rho^{gd} \delta_t^{gd} + \varepsilon_{t+1}^{gd} & \varepsilon_t^{gd} &\sim N\left(\mathbf{0}, \mathbf{Q}^{gd}\right), \\
\delta_{t+1}^{dp} &= \rho^{dp} \delta_t^{dp} + \varepsilon_{t+1}^{dp} & \varepsilon_t^{dp} &\sim N\left(\mathbf{0}, \mathbf{Q}^{dp}\right),
\end{aligned}
$$

Estimates of the autoregression matrices is shown in Table 4 and Table 5 for the consumption process. Similar results (not reported) are obtained for the dividend growth and price-dividend series 6. As expected we verify that components with different degree of persistence are strongly stationary and almost uncorrelated and that the half life extrapolated by the root of the AR(1) using the formula:

$$\rho_j \simeq \exp\left(-HL\left(j\right)\right)$$

is consistent with the level of persistence of the detail:

$$HL\left(j\right) \simeq 2^{j-1}$$

[Insert Table 4, 5 and 6 about here.]

The significance of this result is tight to the spectral approach: the classification of shocks on the basis of their level of persistence improves the statistical identification of dynamic correlations in the time series. With this respect our work can be compared to the one of Calvet and Fisher (2007) which develop an asset pricing equilibrium model with shocks of heterogeneous durations. Moreover the diagonal nature of the matrices ρ^g, ρ^{gd} and ρ^{pd} demonstrates

[10]We thank Jason Beeler for kindly providing us the data.

that testing the predictability of aggregate long run consumption and dividend growth without controlling for persistence can generate a severe error-in-variables problem.Notice that the application of the redundant transformation implies that the details are autocorrelated; in line of principle this autocorrelation effect could be removed by considering the time series of details at scale j $\delta_{t+2^j k}$, $k \in \mathbb{N}$, i.e. subsampling the time series over time intervals of length 2^j. Of course this procedure would largely reduce the number of observations. We checked in the data and more extensively on simulations that the results are robust and remain unaffected when passing from redundant to non redundant transformations.

In the next section we explore the possible applications of this different decomposition to the Long Run valuation paradigm.

4 Long run valuation and the term structure of consumption risks

In this section we revisit the model of Bansal Yaron (2004) proposing an extended long run valuation model. The novelty of our modeling approach lies in the possibility to generate a consumption growth dynamics that accounts for the persistence heterogeneity and is consistent at any level of aggregation while retaining analytical tractability and a linear specification. This is possible thanks to the use of the persistence based decomposition to represent all the relevant processes. Correspondingly the relations between equity return variations, cash flow risk and persistent fluctuations in consumption growth are disaggregated across different levels of persistence.

As in the original model we consider a pure exchange economy with a representative agent which has Epstein-Zin recursive preferences which allow for a separation between the intertemporal elasticity of substitution (IES) and risk aversion, and consequently permits both parameters to be simultaneously greater than 1. The Euler condition on an holding period h can be written as:

$$E_t\left[\beta^{\theta} G_{t+h}^{\frac{\theta}{\psi}} R_{t+h}^{a(1-\theta)} R_{t+h}^{i}\right] = 1 \quad \forall h > 0 \tag{13}$$

where G_{t+h} is the consumption growth between t and $t+h$, R^a_{t+h} is the return on the holding period between t and $t+h$ of the claim which releases an aggregate dividend equal to consumption, R^i_{t+h} is the holding period return on the asset i between t and $t+h$. The parameter β is the preference discount factor. The preference parameter ψ measures the intertemporal elasticity of substitution, γ measures the risk aversion and $\theta = (1-\gamma)\,/\,(1-1/\psi)$.

Note that the R^a_{t+h} is in general different from the market portfolio R^m_{t+h} corresponding to a claim to aggregate dividends, implicitly assuming that the agent has access to the labor income market. The log-linear expression of the intertemporal marginal rate of substitution on the holding period return h is given by:

$$m_{t+h} = \theta \log \beta - \frac{\theta}{\psi} g_{t+h} + (\theta - 1) r_{a,t+h} \tag{14}$$

To derive the solutions for the model, we use the standard log-linear approximation used for log returns (see Campbell and Shiller, 1988) extended to a general holding period equal to h:

$$r_{a,t+h} = \kappa_0 + \kappa_1 z_{t+h} - z_t + g_{t+h}$$

where we define z^a_{t+h} as the log price dividend at time $t+h$ considering the data aggregated over a time interval h. Thanks to the fact that the aggregation operator leaves the unconditional mean of the process unchanged, $\kappa_{1,h} = \kappa_1 = \exp(E[z])/(1+\exp(E[z]))\ \forall h$.

Analogously we define $r_{m,t+h}$ the market return and $z_{m,t+h}$ the price dividend ratio aggregated over a time interval h. Although there are no formal obstructions to extend the analysis to stochastic second moments, in this research volatilities of log-consumption growth and log-dividend growth are assumed to be constant in order to focus on the fluctuations in the mean whose effect in the data has been questioned and is a primary research question under debate.

If we assume that there exists a maximum level of persistence J such that shocks which persist beyond time scales larger than 2^J cannot be separated by the permanent shocks, then consumption growth and dividend growth can be decomposed as:

$$g_t = \sum_{j=1}^{J} g_{j,t} + g_{-1,t}$$

$$gd_t = \sum_{j=1}^{J} gd_{j,t} + gd_{-1,t}$$

The crucial extension with respect to BY04 is the specification of the dynamics of the predictable components, $\{x_{j,t}\}_{j=1,..,J}$, disaggregated over different persistence levels. As we note in the previous section it makes sense to assume that the J components $x_{j,t}$ are independent and follow an AR(1) process:

$$\begin{aligned} x_{j,t+2^j} &= \rho_j x_{j,t} + \varepsilon_{j,t+2^j} \\ \varepsilon_{j,t+2^j} &\sim N\left(0, \left(\sigma^{(j)}\right)^2\right) \end{aligned} \tag{15}$$

Notice that the assumption that the noises $\varepsilon_{j,t+2^j}$ are not autocorrelated forced us to assume that the process at scale j is innovated every 2^j periods.

The relation between (log) dividend growth, gd_t, (log) consumption growth, g_t and $\{x_{j,t}\}$ follows from the specification given in BY04 with the key difference that our relations holds scale by scale:

$$\begin{aligned} g_{j,t+2^j} &= x_{j,t} + e^g_{j,t+2^j} & \forall j \in S \qquad (16) \\ e^g_{j,t+2^j} &\sim N\left(0, \sigma^2_{g,j}\right) \\ gd_{j,t+2^j} &= \phi_j x_{j,t} + e^d_{j,t+2^j} & \forall j \in S \qquad (17) \\ e^d_{j,t+2^j} &\sim N\left(0, \sigma^2_{d,j}\right) \end{aligned}$$

where the set $S \subset \{1, .., J\}$ selects those components of consumption growth (dividend growth) which are lead by the price consumption (price dividend) ratio. Consistently at any level of persistence $j \in S$ we assume the following linear relations

$$\begin{aligned} z^a_{j,t} &= A_{0,j} + A_j x_{j,t} \ \forall j \in S \qquad (18) \\ z^m_{j,t} &= A^m_{0,j} + A^m_j x_{j,t}. \end{aligned}$$

These relations imply that backward components for price consumption $z^a_{j,t}$ and price dividends $z^m_{j,t}$ lead the forward components of consumption and dividends. Hence the levels of persistence belonging to S select the characteristic components of consumption growth which are predictable. By construction, shocks with different levels of persistence are orthogonal. As the prediction horizon is increased, the effect of the most persistent components of shocks becomes dominant, while those with a faster decay rate are averaged out and disappear.

The values of $A_{0,j}, A_j, A^m_{0,j}, A^m_j$ are determined by imposing that the Euler conditions hold on a decreasing sequence of aggregation intervals $h = 2^j$, $j = 1, ..., J$. This hierarchical procedure imposes the Euler equations on a sequence of time scales of decreasing length and correspondingly determines the values of the coefficient $A_{0,j}, A_j, A^m_{0,j}, A^m_j$ with decreasing index of persistence j. Let $m_{j,t+h}$ be the time $t + h$, $j - th$ component of the persistence based decomposition for the IMRS.

Using the above dynamics eqs. (16) and (17) together with eq. (14) it is possible to find the expression of $m_{j,t+2^j}$ as a function of the factors $x_{j,t}$ and of the innovations $e^g_{j,t+2^j}$ and $\varepsilon_{j,t+2^j}$:

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

and using the linear relation (18) for the price-consumption ratio we obtain

$$
\begin{aligned}
&= \theta log(\beta) - \left(1-\theta+\frac{\theta}{\psi}\right) g_{j,t+2^j} \\
&= \theta log(\beta) + (\theta-1)\left(\kappa_0+\kappa_1 A_{0,j}-A_{0,j}\right) \\
&- \left(1-\theta+\frac{\theta}{\psi}\right) x_{j,t} + (\theta-1)\left((\kappa_1\rho_j-1)A_j x_{j,t}\right) \\
&- \left(1-\theta+\frac{\theta}{\psi}\right) e^g_{j,t+2^j} + (\theta-1)\kappa_1 A_j \varepsilon_{j,t+2^j}
\end{aligned}
$$

The innovation details are then given by:

$$
m_{j,t+2^j} - E_t\left[m_{j,t+2^j}\right] = -\left(1-\theta+\frac{\theta}{\psi}\right) e^g_{j,t+2^j} + (\theta-1)\kappa_1 A_j \varepsilon_{j,t+2^j} \tag{19}
$$

Hence all the quantities $z_{j,t+2^j}, z_{j,t}$, $m_{j,t+2^j}$,$g_{j,t+2^j}$,and $gd_{j,t+2^j}$ have been expressed as linear function of the factors $x_{j,t}$. Using either the redundant or the non.redundant transformations is is immediate to verify that:

$$
m_{t+1} = \sum_{j=1}^{J} m_{j,t+2^J} \tag{20}
$$

It is then trivial to show using (19) and (20) that:

$$
\begin{aligned}
m_{t+1} - E_t[m_{t+1}] &= -\left(\frac{\theta}{\psi}-\theta+1\right) e^g_{t+1} - \kappa_1(1-\theta)\underline{A}\cdot(\boldsymbol{\varepsilon}_{t+1}) \\
&= -\lambda_\eta e^g_{t+1} - \underline{\lambda}_n \cdot (\boldsymbol{\varepsilon}_{t+1})
\end{aligned} \tag{21}
$$

where

$$
\begin{aligned}
\lambda_\eta &\equiv \left(\frac{\theta}{\psi}-\theta+1\right) \\
\underline{\lambda}_n &\equiv \kappa_1(1-\theta)\underline{A} \\
\boldsymbol{\varepsilon}_{t+1} &= \begin{cases} \varepsilon_{j,t+2^j} & j \in S \\ 0 & otherwise \end{cases} \\
e^g_{t+1} &= \sum_j e^g_{j,t+2^j}
\end{aligned}
$$

It is a central result the fact that the sensitivities to long run shocks $\underline{\lambda}_n$ are a vector and the model can account for a different clsses.

Imposing that the log expression of the Euler equation (13) holds for an arbitrary value of $x_{j,t}$ we get a set of recursive linear equations for the the coefficient $A_{0,j}, A_j, A^m_{0,j}, A^m_j$. If we introduce the following line vectors of dimension 2^J[11]

$$\begin{aligned}
\underline{A} &\equiv [A_1 e^{(1)} + \ldots + A_J e^{(J)} + A_{-1} e^{(-1)}] \\
\underline{A_m} &\equiv [A^m_1 e^{(1)} + \ldots + A^m_J e^{(J)} + A^m_{-1} e^{(-1)}] \\
\underline{1} &\equiv [1 e^{(1)} + \ldots + 1 e^{(J)} + 1 e^{(-1)}] \\
\underline{\phi} &\equiv [\phi_1 e^{(1)} + \ldots + \phi_J e^{(J)} + \phi_{-1} e^{(-1)}]
\end{aligned}$$

the solution to these equations is given by the following vectors of sensitivities[12]:

$$\begin{aligned}
\underline{A} &= \left(1 - \frac{1}{\psi}\right) \underline{1} \left(\mathbb{I}_{2^J} - \kappa_1 M\right)^{-1} \\
\underline{A_m} &= \left(\underline{\phi} - \frac{1}{\psi}\underline{1}\right) \left(\mathbb{I}_{2^J} - \kappa_1 M\right)^{-1} \\
M &= diag\left(\overline{\rho}\right) \quad \left(\overline{\rho}\right)_k = \sum_{j \in S} \rho_j e^{(j)}
\end{aligned}$$

Relying on Abel (1999) the vector $\underline{\phi}$ reflects the exposures of the market dividends to the components of the consumption growth.

The new innovations and $\varepsilon_{j,t+1}$ is the $j-th$ component of the (forward) persistence based decomposition of the innovation between time t and time $t+1$. The parameters λ_η and $\underline{\lambda}_n$ determine the risk compensation for the independent consumption shock e^g_{t+1} while $\varepsilon_{j,t+1}$ for expected growth rate shock at the different scale j respectively. The risk compensation

[11]The parameters that are being introduced depend only from the level of persistence j while they are independent from the second index n which in the wavelet notation parametrizes the lag of the detail with respect to the reference time t. For this specific class of vectors the following notational convention is assumed.

When the dependence with respect to the lag index is irrelevant they can be considered as vectors with J components obtained considering a linear combination of the elements $e^{(j)}$ $j = -1, 1, .., J$ of the canonical basis.

Alternatively when the dependence with respect to the lag index n matters, it can be taken into account, using the following definition of the vectors $e^{(j)}$:

The vectors $e^{(j)}$, $j = 1, .., J$ are composed as follows:

$$\begin{aligned}
\left[e^{(j)}\right]_k &= 1 \quad 2^{J-j} < k \leq 2^{J-j+1} \\
\left[e^{(j)}\right]_k &= 0 \quad otherwise
\end{aligned}$$

The vector $e^{(-1)}$ (corresponding to the scale component) is a vector of 2^J elements, the first element equals 1 all the other elements equal 0.

[12]Detailed calculations are given in a technical appendix available on-line at the authors homepage, see Ortu, Tamoni, and Tebaldi (2010).

for the e^g_{t+1} shock is standard as λ_η equals the risk aversion parameter γ. The equilibrium return becomes

$$r_{a,t+1} - E_t[r_{a,t+1}] = e^g_{t+1} + \kappa_1 \underline{A} \cdot \boldsymbol{\varepsilon}_{t+1}$$

Correspondingly the conditional variance of $r_{a,t+1}$ is given by:

$$\begin{aligned} Var_t(r_{a,t+1}) &= \sigma^2_\eta + \kappa_1^2 \underline{A} \mathbf{Q} \underline{A}' \\ \sigma^2_\eta &= Var\left(e^g_{t+1}\right) \\ \mathbf{Q} &= \mathbf{E}_t\left[\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}'_{t+1}\right] \end{aligned}$$

and the risk premium for the consumption claim asset, $r_{a,t+1}$ is given by

$$\begin{aligned} E_t[r_{a,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{a,t}} &= -cov_t(m_{t,t+1}, r_{a,t+1}) \\ &= \lambda_\eta \sigma^2_{\eta,t} + \kappa_1 \underline{\lambda}_n \mathbf{Q} \underline{A}' \end{aligned}$$

Similarly the innovation to the market return is obtained as follows: the equilibrium return innovation is

$$r_{m,t+1} - E_t[r_{m,t+1}] = e^d_{t+1} + \kappa_{1,m} \underline{A_m} \cdot \boldsymbol{\varepsilon}_{t+1} \tag{22}$$

while the conditional variance of $r_{m,t+1}$ is

$$\begin{aligned} Var_t(r_{m,t+1}) &= \sigma^2_u + \kappa^2_{1,m} \underline{A_m} \mathbf{Q} \underline{A_m}' \\ \sigma^2_u &= Var\left(e^d_{t+1}\right) \end{aligned}$$

Thus we have obtained a picture where the asset valuation involves the full term structure of shocks driving the predictable components of consumption growth.

5 Empirical implications

In this section we revisit the empirical tests which have produced the most controversial results and raised concerns on the effective presence of long run risk in the data. In light of the above discussion the same tests are performed disaggregating economic variables across different levels of persistence in order to avoid the error-in-variable problem. This generates an extended long run risk picture described by a complete term structure of risk-return tradeoffs.

5.1 Predictability of Consumption and Dividend Growth

BY04 model implies that long-run consumption and dividend growth should be highly persistent and predictable from stock prices. In order to evaluate the performance of the long-run risks model, Beeler and Campbell (2009) run the following regression:

$$g_{t+h} = \beta_0 + \beta_1 dp_t + \varepsilon_{t+h}$$

for $h = 1, 3, 5$ years as done in Beeler and Campbell (2009). Results are reported in Table 7 for US quarterly series over the postwar period 1947:2-2007:4 and in Table 8 for US annual data over the period 1930-2006. The results show that stock prices do not predict consumption.

[Insert Table 7 and 8 about here.]

As discussed in the motivating example, heterogeneity of persistence can generate an error-in-variable problem and reduce the discriminatory power of the test on linear predictability. Using the PBD decomposition to the (log) consumption growth g_{t+h} and (log) price-dividend pd_t time series it is possible to separate the components of consumption growth and the price-dividend ratio across different levels of persistence. Results are shown in Figure 6 and in Figure 7 for the quarterly and annual sample respectively. It is interesting to note that as we move toward higher scale, i.e. higher persistence, we can identify a common long-run behavior between the two series. We quantify the predictability at each level of persistence by running a regression component by component, namely:

$$g_{j,t+1} = \beta_0 + \beta_1 dp_{j,t} + \varepsilon_{t+1,j}$$

Results are reported in Table 10 and Table 15.

[Insert Figure 6 and 7 about here.]

[Insert Table 10 and 15 about here.]

Table 10 shows that at level of persistence $j = 3, 6, 7$ the linear predictability tests is statistically significant at 10% level, the sixth and seventh components account for a great part of variation (the R^2 are between 23% and 30%) in the expected future consumption growth at the corresponding scale. Hence on the basis of the empirical results we set $S = \{3, 6, 7\}$ corresponding to cycle length of $\{[12], [816], [1632]\}$ years[13]. We exclude the permanent component -1 since the dividend price variable has been verified to be stationary and to have a zero permanent component. Table 15 reports the results for the same test performed on annual data. Results are consistent: the component at scale corresponding to $1-2$ years, $8-16$ years and $16-32$ years turn to be statistically significant[14].

Note that we are not providing evidence in favor of strong predictability of aggregate consumption growth but just evidence for a persistent component in consumption which cause at the corresponding level of persistence, stock price variability (relative to dividends). This persistent variation in consumption growth, unless filtered, is overwhelmed by measurement error. To better quantify this statement we report in Table 9 the contribution of the details δ_t^j to the total variance of consumption growth. We note that the highly persistent component yields a minor contribution to total variance.

It is important to stress that the whole procedure which has carried to the above result relies on a linear filtering (decomposing our time series using the PBD) and the use the filtered regressand and regressors in ordinary least squares (OLS).

As a robustness check for the filtering procedure we use the bandpass filter described in Christiano and Fitzgerald (2003) [15] The choice of frequencies interval characterizing the band-pass filter is determined according to Section 3: we band-pass the consumption growth and price-dividend ratio over the interval $\left[\frac{f_{max}}{2^j} \; \frac{f_{max}}{2^{j-1}}\right)$ $j = 1, \ldots, 8$. Results are reported in Table 11. We find again evidence for a very long-lasting component in consumption growth (note in fact that the R^2 spikes at scale 6 and 7).

[Insert Table 11 about here.]

The statistical filtering procedure has been discussed extensively, the use of the ordinary least squares on the sequence of details deserves further comments. As already stated in Section 3 the use of the redundant transform increases the size of the sample of observations, but generates a higher autocorrelation in the data. This is a critical issue for the OLS procedure since both the regressor and the regressand become highly persistent and imposes the use of properly modified statistical significance indicators.

While a rigorous asymptotic theory is currently under investigation, we remark that the statistical problem which arises in this situation is very similar to the one which is usually

[13]For an interpretation of the corresponding cycle duration see Table 7.

[14]We do not consider the results for scale 6 given that the components at this scale have very low variability and their trendy behavior makes statistical inference problematic.

[15]Results are practically identical under Baxter and King's (1999) band-pass filter.

considered in the statistical analysis of the so called "long-horizon predictive regressions".

Long-run regressions were made popular in the context of return predictability by influential articles such as Fama and French (1988) and Campbell and Shiller (1988) to name a few. For instance when attempting to predict stock returns over longer horizons, often covering several years, rather than on a month-to-month basis, the evidence in favor of predictability generally appears much stronger. In fact, the estimated regression coefficients tend to increase almost linearly with the forecasting horizon. However the long-horizon predictive regression has been found to be subject to many econometric problems at the same time. These problems arise from the persistence of the forecasting variable[16], overlapping observations in the dependent variable of the predictive regression, and the strong correlation between residuals in the predictive regression and innovations in the process of the dividend-price ratio (see, for example, Hodrick (1992), Nelson and Kim (1993), Stambaugh (1999), and Ferson, Sarkissian, and Simin (2003)). As a result, conventional t-statistics have rejection rates that are well above their nominal levels and new test measures or methods to correct such econometric problems have been proposed to test for the presence of predictability (for example, Ang and Bekaert (2007), Lanne (2002), Valkanov (1999), Lewellen (2004), Torous, Valkanov, and Yan (2004), and Campbell and Yogo (2006), Boudoukh, Richardson, and Whitelaw (2008) and many others). In our paper the standard error are computed using the method proposed by (Hansen and Hodrick 1980) to correct for serially correlated errors.

Note that the approach used in this paper is consistent with the findings of Torous, Valkanov, and Yan (2004) and Bandi and Perron (2008): both the regressor and regressand are aggregated over non-overlapping periods having the same length. In fact Torous, Valkanov, and Yan (2004) shows that the OLS estimator is consistent when both the regressor and the regressand are aggregated over non-overlapping periods (cases 2 and 4 in their paper), i.e., regressing a long-horizon variable against the other.

In the long-run risks model, the persistent variation in consumption growth should create similar variation in dividend growth. Hence we test ability of the log price-dividend ratio components to lead the dividend growth components. We undertake this exercise in Table 12 and 13[17]. The first Table replicates the results of Beeler and Campbell (2009) and shows that dividend growth predictability is absent in postwar quarterly data, in striking contrast with the BY04 calibration of the long-run risks model. In Table 13 we apply our PBD and find that the the components at levels $3, 6, 7$ are statistically significant at 10% with sixth and seventh are statistically significant at 5% level and, account for a great part of variation (the R^2 are between 29% and 35%) in the expected future consumption growth at the corresponding scale.

[16]For instance the list of potential predictors of stock market includes, inter alia, dividend yields, interest rates, book-market values, price-earnings ratios.

[17]We report results only for quarterly data over the period 1947Q2-2007Q4. Similar results are obtained for annual data over the period 1930-2006 and are available upon request.

[Insert Table 12 and 13 about here.]

In summary, we find that at fixed levels of persistence stock prices predict the long-run prospects for consumption and dividend growth. The ultimate relevance of the predictability effects in the components is related to the ability of the "thin persistent effects" to generate sizeable risk premia. To quantify effects on prices it is necessary to exploit the information in the data on the representative agent preference parameters.

5.2 Equity Premium and Intertemporal Elasticity of Substitution

The objective of this section is analyze the effects of the persistence heterogeneity on the tests which elicit the elasticity of intertemporal substitution of the representative agent from market data. In order to do that we recall that $\underline{A_m}$ are given by the following relation:

$$\underline{A_m} = \left(\underline{\phi} - \frac{1}{\psi}\right)\left((I_h - \kappa_1 M)\right)^{-1}$$

thus ψ can be expressed as a function of the estimates for $\underline{A_m}$, $\underline{\phi}$ and M.

We now focus on the estimation of $\underline{A_m}$, $\underline{\phi}$. Recall the following relation $\forall j \in S$

$$\begin{aligned} g_{j,t+2^j} &= x_{j,t} + e^g_{j,t+2^j} \\ gd_{j,t+2^j} &= \phi_j x_{j,t} + e^d_{j,t+2^j} \end{aligned} \tag{23}$$

and

$$\begin{aligned} z_{m,j,t} &= A_{0,j,m} + A_{j,m} x_{j,t} \\ z_{j,t} &= A_{0,j} + A_j x_{j,t} \end{aligned} \tag{24}$$

Consider the regression component by component of the 1-period ahead log consumption growth against the log price-dividend ratio:

$$g_{j,t+2^j} = \beta_0 + \beta_{j,c} z_{m,j,t} + \varepsilon_{t+2^j}. \tag{25}$$

Using equations (23) and (24) it is immediate to verify that the regression coefficient $\widehat{\beta}_{j,c}$ produces an estimate of the coefficient $\frac{1}{A_{j,m}}$.
By the same token we regress components by components the 1-period ahead log dividend growth on the log price-dividend ratio

$$gd_{j,t+2^j} = \beta_0 + \beta_{j,gd^m} z_{m,j,t} + \varepsilon_{t+2^j} \tag{26}$$

and this yields $\hat{\beta}_{j,gd^m}$ equal to $\frac{\phi_j}{A_{j,m}}$. The results for regressions (25) and (25) are given in Table 10 and in Table 13, respectively. In conclusion an estimate of ϕ_j is given by

$$\phi_j = \frac{\hat{\beta}_{j,gd^m}}{\widehat{\beta}_{j,c}}$$

The estimation of M has already been discussed in sub-section 3.1, see Table 6.

With all our parameters at hands we then obtain the estimation of $\widehat{\psi}^{(j)}$ at different levels of persistence which reported in Table 7. Inspection of these estimates reveals that heterogeneity of persistence causes heterogeneity in the value of this parameter which we recall describes the preferences of the representative agent. Of course the representative agent hypothesis imposes that a single coefficient is sufficient to describe the aggregate preferences. If we consider that all the estimates are equally relevant irrespectively of the statistical significance of the linear regression, then the minimum square estimate of the IES value falls below 1 and is given by $\hat{\psi} = 0.66$. On the contrary the minimum square estimate which relies only on the significant estimates at 10% level $\hat{\psi}^{10\%} = 3.3893$ from Table 10 and Table 13 scale $j \in S = \{3, 6, 7\}$. If the level of significance is raised to 5% and only the components $j = 6, 7$ are retained the minimum square estimate is given by $\hat{\psi}^{5\%} = 1.43$. Note that these results have a straight economic interpretation agents are likely to be less elastic in response to high frequencies variations and more elastic in response to low frequencies (e.g., life cycle) variations.

[Insert Table 7 about here.]

After controlling for persistence these findings suggests that the conflicting evidence on the magnitude of the elasticity of intertemporal substitution seems to be consistent with a framing effect: the representative agent shows different preferences when facing intertemporal substitution of short- or long- term consumption risk. In particular our results suggest that the elasticity of intertemporal rate of substitution based on low persistence components are consistent with a representative agent for which the wealth effect dominates the intertemporal substitution one. On the contrary the determination based on high persistence components detects an elasticity parameter larger than 1, consistent with the long run risk picture where it is necessary that the intertemporal substitution dominates the wealth effect.

6 The term structure of equity risk premia.

6.1 Market risk premium

The key verification that the Extended Long Run Risk picture reconciles empirical findings with the calibration of BY04 relies on the analysis of the risk premia.

Equation (21) together with equation (22) and the well known relation

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

imply the following relation for the market portfolio $r_{m,t+1}$:

$$\begin{aligned} & E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{m,t}} \\ &= \kappa_1 \underline{\lambda}_n \mathbf{Q} \underline{A_m}' \\ &= \kappa_1 \underbrace{\kappa_1(1-\theta)\underline{A}}_{\underline{\lambda}_n} \mathbf{Q}\ \underline{A_m}' \end{aligned} \tag{27}$$

The risk premia reflect both risk exposure, $\underline{A_m}$, and risk prices, $\underline{\lambda}$, associated with the different horizons. We just recall that we have

$$\begin{aligned} \underline{A_m} &= \left(\underline{\phi} - \frac{1}{\psi}\right)((I_h - \kappa_1 M))^{-1} \\ \underline{A} &= \left(\underline{1} - \frac{1}{\psi}\right)((I_h - \kappa_1 M))^{-1} \end{aligned}$$

Therefore in order to compute the equity premium we need to obtain an estimate for M, $\mathbf{Q}$ and $\underline{\phi}$ and θ (and therefore γ and ψ).

We set the risk aversion parameter $\gamma = 5$ whereas according to the previous section we set a conservative estimate $\psi = 1.5$ equal to the original BY04 and close to $\widehat{\psi}^{5\%} = 1.43$. Results with $\widehat{\psi}^{10\%} = 3.3893$ are reported and are very similar. The M and $\mathbf{Q}$**,** the innovations' variance-covariance matrix is obtained from the vector autoregressive system of dividend-price components. The estimation of $\underline{A_m}$ requires the computation of the parameters $\underline{\phi}$. The procedure is the same outlined in Section 5.2: we use the relation $\phi_j = \frac{\hat{\beta}_{j,gd^m}}{\hat{\beta}_{1,c}}$ where $\widehat{\beta}_{j,c}$ and the $\hat{\beta}_{j,gd^m}$ are reported in Table 10 and Table 13 respectively[18].

Results are reported in Table (7). We can note that the equity premium doesn't necessarily increases with the level of persistence. Moreover notice that the equity premium can

[18]In order to compute ϕ_j we keep only the $\widehat{\beta}_{j,c}$ and the $\hat{\beta}_{j,gd^m}$ that are significant at 10% level, otherwise we set $\phi_j = 0$.

be zero at some scale and this can happen either because the price of risk is zero or because the risk exposure is zero. For the market portfolio using the results in Table 10 and Table 13 we observe that the risk exposure can be different from zero only at scale 3, 6 and 7. We highlight that the unique parameter which is calibrated is the risk aversion coefficient, which is set to the reasonable value of $\gamma = 5$. Hence we can conclude that on the basis of our empirical findings, within this Extended Long Run Risk model, Long Run Risk offers a plausible solution to the equity premium puzzle.

[Insert Table (7) about here.]

6.2 The value premium

We now carry out the same analysis but for the expected returns of the growth and value portfolios. It is in fact interesting to analyze the relation between dividends on characteristic sorted portfolios and components in aggregate consumption growth. In fact in a consumption-based model, differences in risk compensation on assets should be explained by differences in the *exposure* of assets' cash flows to consumption. Following the same line of development of Hansen, Heaton, and Li (2008) we consider the portfolios constructed from stocks with different ratios of book value to market value of equity. Details of the construction of the portfolios and cash flows can be found in Hansen, Heaton, and Li (2005). In particular we focus on portfolios 1 (value) and 5 (growth) (see, e.g., Fama and French 1992) and analyze their covariation with consumption growth. It is well documented that the one period average returns to portfolios of high book-to-market stocks (value portfolios) are substantially larger than those of portfolios of low book-to-market stocks (growth portfolios). In particular we find in the data that the one-period realized returns to the growth and values portfolios is 4.9 percent and 10.1 percent, respectively (see table 18).

[Insert Table 18 about here.]

A preliminary analysis is performed computing the correlation[19] between the cash flow growth rate components for Portfolios 1 and 5 and consumption growth rate components $Corr\left(gd_j^{(i)}, g_j^{(i)}\right)$ $i = 1, 5$ $j = 1, .., 9$. Table 19 report the results. We find that, for scale $j \geq 5$, the correlation are significantly different from zero and, more important, the dividend growth components of the value portfolio, $gd_{j,t}^{(5)}$ are positively related with consumption growth, $g_{j,t}$, whereas the dividend growth components of the growth portfolio $gd_{j,t}^{(1)}$ are negatively correlated with consumption growth. Hence this preliminary analysis seems to lead to a similar finding as in Hansen, Heaton, and Li (2008), that is the cash flows of value

[19] We use Pearson's correlation coefficient but similar results were obtained when using the Spearman's correlation coefficient.

portfolios exhibit positive comovement in the long run with macroeconomic shocks (proxied by our smoothed consumption growth rate) whereas the growth portfolios show little (or negative) covariation with these shocks and these findings suggest that the book-to-market portfolios can be possibly sorted along macroeconomic exposures across firms.

In order to compute the expected premium on those portfolios we apply the same procedure outlined for the market risk premium. We use the relation $\phi_j = \frac{\hat{\beta}_{j,gd^i}}{\hat{\beta}_{j,c}}$ where $i = 1, 5$ using the $\widehat{\beta}_{j,c}$ and the $\hat{\beta}^{(5)}_{j,gd^i}$ reported in Table 10 and Table 20 Panel A (for the value portfolio) and $\hat{\beta}^{(1)}_{j,gd^i}$ reported in Table 20 Panel B(for the growth portfolio). In particular, if we keep only the $\widehat{\beta}_{j,c}$ and the $\hat{\beta}_{j,gd^i}$ that are significant at 10% level, the estimate of ϕ_j for the value and growth portfolios are different from zero only at scale 3 and 6.

Figure 8 displays decomposition of expected returns for the growth, value, and market portfolios described as a function of horizon and produces the term structure of risk premia over alternative horizons.

[Insert Figure 8 about here.]

The risk premium produced by the growth and value portfolios are substantially different (and much larger for the value portfolio) at scale 3 and 6 accounting for a 2% and 3% respectively and thus yielding a total of 5% similar to the one we find in data. It is also intresting to note that the (difference in) expected returns to the the value portfolio and the growth portfolio increase with horizon and, importantly, the exposure to long-run macroeconomic risk spikes at different horizon. Moreover we observe at long horizon that the growth portfolio carries a negative risk premium that therefore seems to act as an hedge instrument whereas the value portfolio carries a positive risk premium. This is consistent with our previous analysis (see Table 19) where at long horizon the correlation of the growth (value) portfolio with consumption growth were negative (positive).

[Insert Table 20 and 19 about here.]

Notice that in this Extended Long Run Risk Model the variability in the expected returns for the value/growth portfolio and the market portfolio is by construction explained by the different exposure to long-run macroeconomic risk. Indeed the market portfolios risk premium spikes at very long horizon (scale 6 and 7), differently from the value and growth portfolios that spikes both at for short horizons (scale 3) and medium horizon (scale 6). It has to be remarked that the model predicts that the value premium manifests itself over a time scale of 8 years which of course has important implications for asset valuation over long horizons.

7 Conclusions

The above considerations prove that a classification of shocks based on the persistence based decomposition improves the discriminatory power of empirical tests on long run risk valuation models getting rid of an error in variable problem generated by the heterogeneity of persistence.

This paper shows that an Extended Long Run Risk picture where the effects of persistence heterogeneity is properly taken into account offers a credible explanation to many empirical results which seemed to contradict the long run valuation picture. Our results clearly indicate that any systematic empirical test of a Long Run Risk model must classify shocks across two competing dimensions, their size as measured by volatility and their persistence as measured by their half life.

Our proposal, the use of a persistent based decomposition, seems to offer interesting developments. It has to be remarked that pros and cons of filtering procedures have been discussed in the macroeconomic literature (see for instance Christiano and Fitzgerald (2003) and Canova (1998)) where it has been observed that sometimes results are not robust with respect to different choices of the filtering criterion. In fact the decomposition procedure introduces an additional source of model risk, hence an uncertainty averse agent should take it into account while forming expectations. In our analysis we assumed that the representative agent is uncertainty indifferent leaving for for future research the analysis of the effects of ambiguity aversion on valuation.

There's a number of research questions which are left open by our research both on the methodological and on the empirical side. In our understanding a systematic analysis of the asymptotic large sample theory for the persistence based decomposition would offer a natural and general framework to analyze the term structure of risk-return tradeoffs in asset valuation. On the empirical side it is clear that the next step to verify the credibility of the Extended Long Run Risk picture is the extension of the analysis to bond prices. Preliminary research on this side has produced promising results but unavoidably requires the extension of the model with the inclusion of inflation among priced risks.

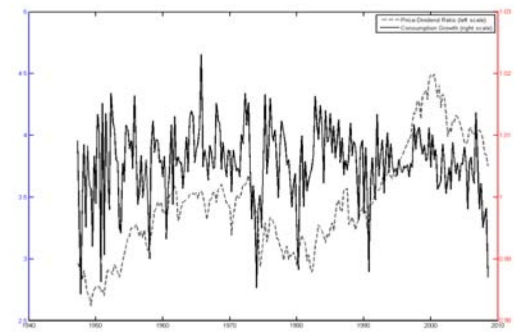

Figure 1: The figure displays the series of US consumption growth (nondurables and services) from the Bureau of Economic Analysis and the log price-dividend ratio (dashed line).

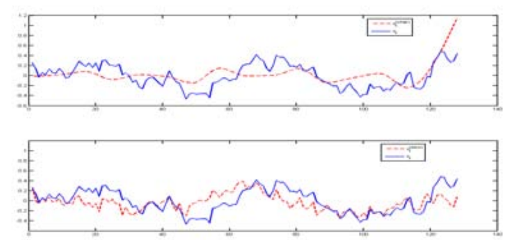

Figure 2: The figure displays (a realization of) pd_t together with the components $x_t^{(char)}$ (top panel) and $x_t^{(idyo)}$ (bottom panel).

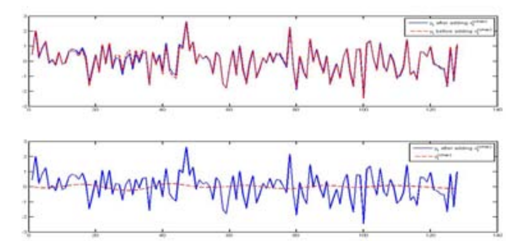

Figure 3: The figure displays (a realization of) y_t before adding $x_t^{(char)}$ and y_t after having added the persistent component $x_t^{(char)}$ of the signal x_t.

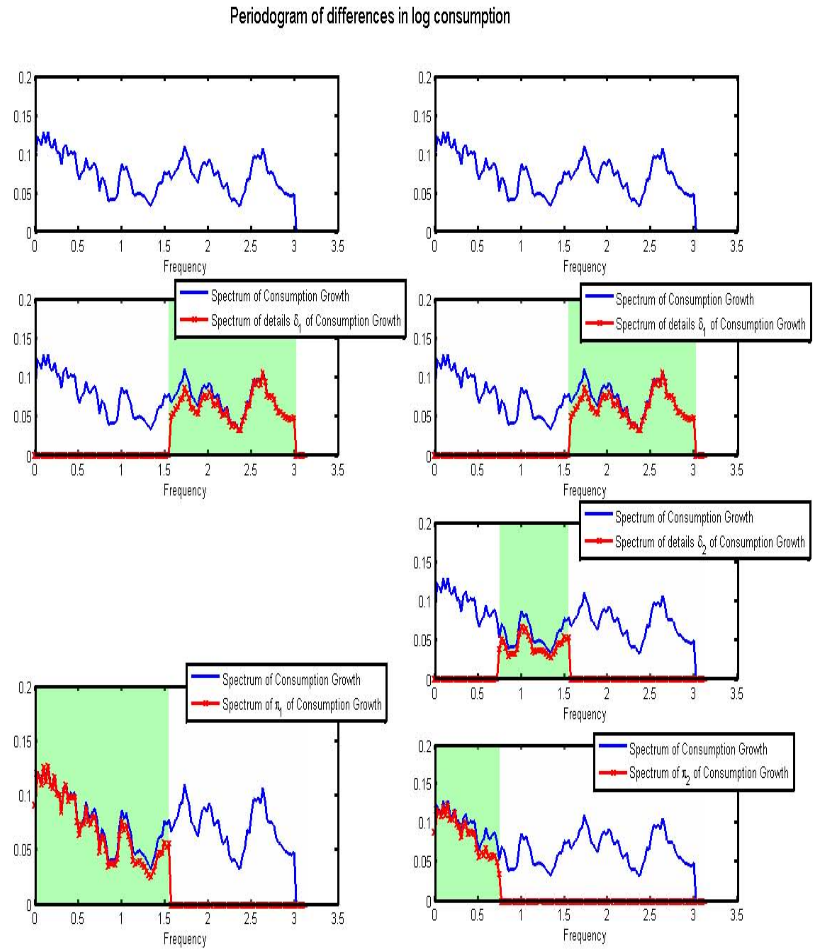

Figure 4: The figure displays the effects of the persistence based decomposition of the consumption time series applied up to level $J = 1$ (left panels) and $J = 2$ (right panels). In particular the top panels displays the smoothed periodogram the consumption process for the data. An equally weighted "nearest neighbor" kernel was used to perform the smoothing, equally weighting the 4 nearest frequencies. The bottom right panel displays the Fourier spectrum of the time series $\pi_t^{(2)}$ whereas the bottom left panel displays the Fourier spectrum of the time series $\pi_t^{(1)}$.

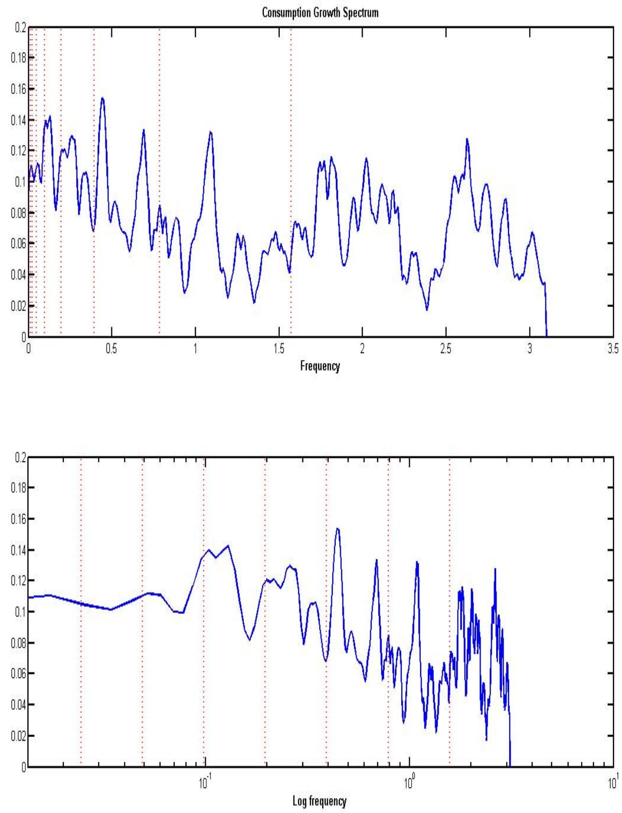

Figure 5: The figure displays the smoothed periodogram the consumption process for the data together with the intervals $\left[\frac{f_{max}}{2^j} \; \frac{f_{max}}{2^{j-1}}\right)$ $j = 1, \ldots, 8$. An equally weighted "nearest neighbor" kernel was used to perform the smoothing, equally weighting the 4 nearest frequencies. In the top panel linear scale is used for frequencies whereas in the bottom panel logarithmic scale is used for the X-axis.

	Total Number of observation		
$z_t =$	T=64	T=128	T=256
pd_t	0.541	0.361	0.215
	(0.921)	(0.937)	(0.868)
	[0.02]	[0.01]	[0.00]

Table 1: This table reports the results of $g_{t+1} = \beta_0 + \beta_1 pd_t + \epsilon_{t+1}$. The table reports OLS estimates of the regressors, t-statistics in parentheses and R^2 statistics in square brackets. The values are from 5000 simulations.

	Total Number of observation		
$z_t =$	T=64	T=128	T=256
$pd_t^{(char)}$	0.999	0.999	1.000
	(3.318)	(4.790)	(6.467)
	[0.17]	[0.16]	[0.15]

Table 2: This table reports the results of the (filtered) regression $g_{t+1}^{(char)} = \widetilde{\beta}_0 + \widetilde{\beta}_1 pd_t^{(char)} + \widetilde{\epsilon}_{t,t+1}$. The table reports OLS estimates of the regressors, t-statistics in parentheses and R^2 statistics in square brackets. The values are from 5000 simulations. $v = 10^{-1}$

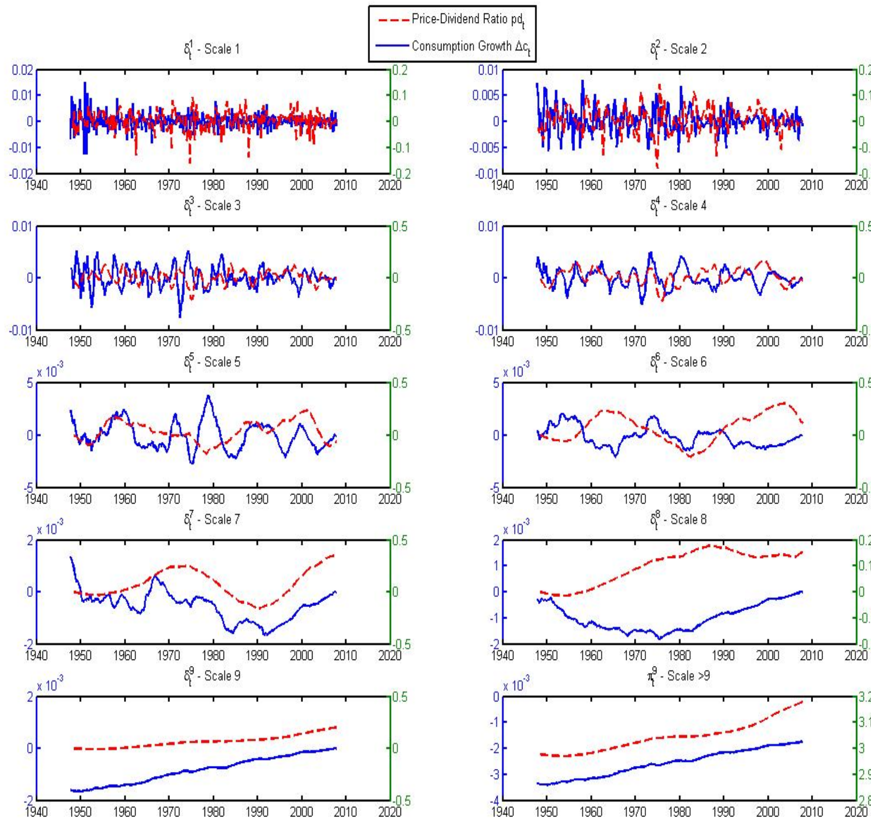

Figure 6: Time-scale decomposition for the log price-dividend pd_t and log consumption growth based upon quarterly data. The sample spans the period 1947Q2-2007Q4.

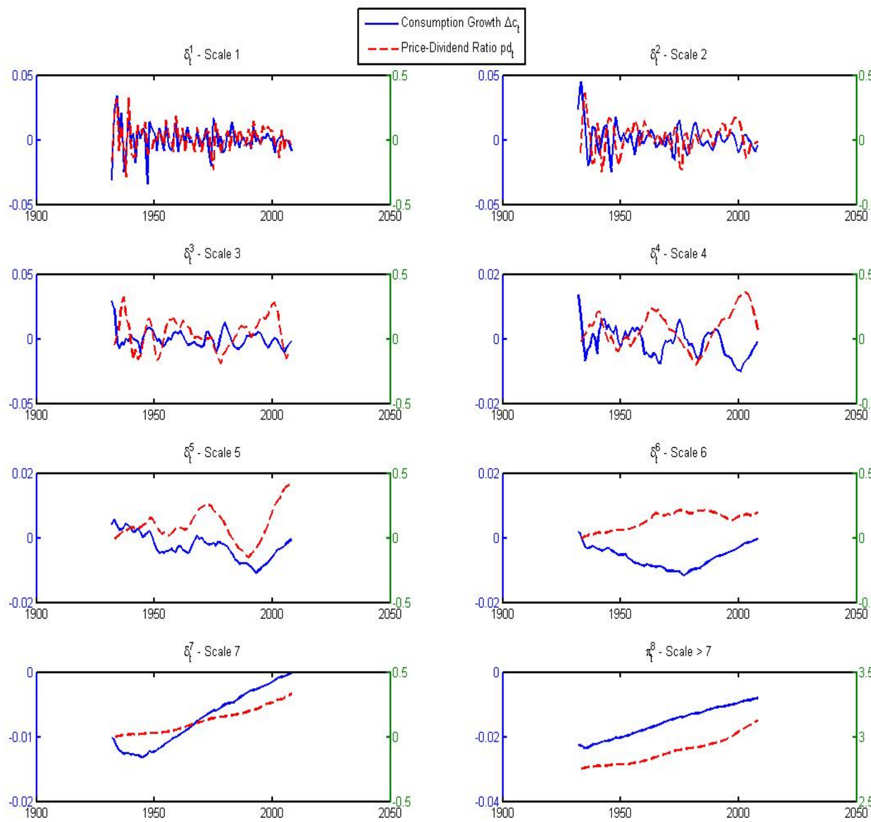

Figure 7: Time-scale decomposition for the log price-dividend pd_t and log consumption growth Δc_t based upon annual data. The sample spans the period 1930-2008.

Component	Quarterly-frequency resolution
δ_t^1	$1-2$ quarters
δ_t^2	$2-4$ quarters
δ_t^3	$1-2$ years
δ_t^4	$2-4$ years
δ_t^5	$4-8$ years
δ_t^6	$8-16$ years
δ_t^7	$16-32$ years
δ_t^8	$32-64$ years
δ_t^9	$64-128$ years
π_t^9	>128 years

Table 3: Frequency interpretation of time scales.

	Scale j								
$z_t =$	1	2	3	4	5	6	7	8	$\bar{R}^2$
$\Delta c_{1,t+1}$	$-$**0.459**	-0.406	-0.261	-0.325	-0.308	-0.328	-0.409	-0.770	[0.38]
	(-5.15)	(-5.46)	(-3.45)	(-3.28)	(-4.08)	(-2.26)	(-2.18)	(-2.04)	
$\Delta c_{2,t+1}$	0.534	**0.294**	-0.158	-0.154	-0.135	-0.154	-0.195	-0.360	[0.65]
	(8.09)	(8.09)	(-3.89)	(-3.36)	(-5.65)	(-2.19)	(-2.37)	(-1.86)	
$\Delta c_{3,t+1}$	0.016	0.314	**0.693**	-0.123	-0.086	-0.088	-0.068	-0.116	[0.68]
	(0.43)	(15.07)	(24.71)	(-3.82)	(-2.20)	(-1.66)	(-1.42)	(-1.74)	
$\Delta c_{4,t+1}$	-0.018	-0.024	0.185	**0.862**	-0.080	-0.042	-0.069	-0.116	[0.83]
	(-1.11)	(-1.21)	(7.53)	(27.82)	(-2.39)	(-0.93)	(-1.15)	(-1.21)	
$\Delta c_{5,t+1}$	0.007	-0.003	-0.001	0.083	**0.956**	-0.040	-0.023	-0.036	[0.93]
	(0.71)	(-0.31)	(-0.05)	(4.77)	(52.17)	(-1.60)	(-0.80)	(-0.70)	
$\Delta c_{6,t+1}$	-0.002	0.010	0.013	-0.004	-0.027	**0.977**	-0.025	-0.042	[0.96]
	(-0.44)	(1.66)	(1.88)	(-0.56)	(2.47)	(65.04)	(-1.34)	(-1.41)	
$\Delta c_{7,t+1}$	0.002	0.001	0.003	0.009	0.007	0.016	**0.993**	-0.013	[0.99]
	(0.92)	(0.64)	(1.00)	(2.47)	(1.70)	(2.50)	(96.47)	(-0.81)	
$\Delta c_{8,t+1}$	0.000	0.001	0.001	0.001	0.002	0.003	0.008	**1.00**	[0.99]
	(0.11)	(1.34)	(1.13)	(0.33)	(1.07)	(1.14)	(2.84)	(138.8)	

Table 4: This table reports the results of regressions of each of the components of 1-period ahead consumption growth $\Delta c_{j,t+1}$ on all lagged components of (log) consumption growth $\Delta c_{i,t}$ for $i = 1, \ldots, 8$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q2-2007Q4.

$z_t =$	$\Delta c_{j,t}$	$\bar{R}^2$
$\Delta c_{1,t+1}$	-0.460 (-5.42)	[0.21]
$\Delta c_{2,t+1}$	-0.309 (10.68)	[0.09]
$\Delta c_{3,t+1}$	0.733 (14.81)	[0.53]
$\Delta c_{4,t+1}$	0.885 (31.14)	[0.78]
$\Delta c_{5,t+1}$	0.961 (50.63)	[0.92]
$\Delta c_{6,t+1}$	0.982 (64.40)	[0.96]
$\Delta c_{7,t+1}$	0.994 (131.13)	[0.99]
$\Delta c_{8,t+1}$	0.995 (130.05)	[0.99]

Table 5: This table reports the results of regressions of each of the components of 1-period ahead consumption growth $\Delta c_{j,t+1}$ on its own lagged components $\Delta c_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q2-2007Q4.

$z_t =$	$pd_{j,t}$	$\bar{R}^2$
$pd_{1,t+1}$	0.057 (-5.42)	[0.10]
$pd_{2,t+1}$	0.651 (11.96)	[0.42]
$pd_{3,t+1}$	0.888 (20.74)	[0.79]
$pd_{4,t+1}$	0.965 (31.51)	[0.93]
$pd_{5,t+1}$	0.966 (68.10)	[0.99]
$pd_{6,t+1}$	0.998 (64.40)	[0.99]
$pd_{7,t+1}$	0.999 (164.54)	[1.00]
$pd_{8,t+1}$	0.997 (327.63)	[1.00]

Table 6: This table reports the results of regressions of each of the components of 1-period ahead $pd_{j,t+1}$ on its own lagged components $pd_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q2-2007Q4.

$z_t =$	Horizon h (in quarters) 4Q	12Q	20Q
	0.002	-0.001	-0.002
pd_t	(0.539)	(-0.107)	(-0.127)
	[0.005]	[0.000]	[0.001]

Table 7: This table reports the results of h-period predictive regressions of the consumption growth on price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is quarterly and spans the period 1947Q2-2007Q4.

$z_t =$	Horizon h (in years) 1	3	5
	0.010	0.010	-0.004
pd_t	(1.444)	(0.434)	(-0.181)
	[0.05]	[0.01]	[0.00]

Table 8: This table reports the results of h-period predictive regressions of the consumption growth on price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2008.

	Component at scale j							
j=	1	2	3	4	5	6	7	8
$\frac{Var(\Delta c_{t,j})}{Var(\sum \Delta c_{t,j})}$	0.395	0.203	0.148	0.104	0.057	0.030	0.022	0.041

Table 9: Contribution to total variance of the different details components $\Delta c_{t,j}$ of the log consumption ratio. Note that $Var(\sum \Delta c_{t,j}) = Var(\Delta c_t)$

	Scale j							
$z_t =$	1	2	3	4	5	6	7	8
	0.34	-0.70	-0.89	0.16	-0.02	-0.32	0.23	0.04
pd_t	(0.62)	(-1.85)	(-2.62)	(0.38)	(-0.08)	(-1.60)	(2.96)	(0.30)
	[0.00]	[0.02]	[0.07]	[0.00]	[0.00]	[0.23]	[0.30]	[0.00]

Table 10: This table reports the results of predictive regressions of the components of consumption growth $\Delta c_{j,t+2^j}$ on the components of (log) price-dividend ratio $pd_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q2-2007Q4.

	Scale j							
$z_t =$	1	2	3	4	5	6	7	8
	0.93	-1.63	-1.58	1.78	-1.05	-0.53	1.68	-0.81
pd_t	(1.47)	(-2.70)	(-1.78)	(2.55)	(-2.01)	(-2.34)	(19.24)	(-1.06)
	[0.01]	[0.03]	[0.02]	[0.03]	[0.02]	[0.06]	[0.61]	[0.01]

Table 11: This table reports the results of predictive regressions of the components of consumption growth on the components of (log) price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is quarterly and spans the period 1947Q2-2007Q4. The components are extracted using the default filter recommended in Christiano and Fitzgerald (1999).

$z_t =$	Horizon h (in quarters) 4Q	12Q	20Q
	0.009	0.006	0.009
pd_t	(0.390)	(0.096)	(0.108)
	[0.005]	[0.000]	[0.001]

Table 12: This table reports the results of h-period predictive regressions of the dividend growth on price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is quarterly and spans the period 1947Q2-2007Q4.

	Scale j							
$z_t =$	1	2	3	4	5	6	7	8
	-8.90	1.69	-2.83	-0.34	0.58	8.09	-9.23	5.97
pd_t	(-0.81)	(0.30)	(-1.66)	(-0.19)	(0.26)	(1.96)	(-2.15)	(1.10)
	[0.00]	[0.00]	[0.02]	[0.00]	[0.00]	[0.35]	[0.29]	[0.04]

Table 13: This table reports the results of predictive regressions of the components of dividend growth $\Delta d_{j,t+2^j}$ on the components of (log) price-dividend ratio $pd_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q4-2007Q4.

	Scale j							
$z_t =$	1	2	3	4	5	6	7	8
	0.54	-0.66	-0.91	0.12	0.06	-0.13	0.20	-0.36
pc_t	(1.04)	(-1.82)	(-3.03)	(0.33)	(0.22)	(-1.88)	(7.63)	(-5.50)
	[0.00]	[0.02]	[0.08]	[0.00]	[0.00]	[0.06]	[0.60]	[0.55]

Table 14: This table reports the results of predictive regressions of the components of 1-period ahead consumption growth $\Delta c_{j,t+2^j}$ on the components of (log) price-consumption ratio $pc_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q4-2007Q4.

$z_t =$	Scale j					
	1	2	3	4	5	6
	5.86	-2.28	-0.05	-1.70	1.09	-2.68
pd_t	(4.53)	(-3.17)	(-0.06)	(-2.09)	(1.88)	(7.32)
	[0.34]	[0.06]	[0.00]	[0.26]	[0.13]	[0.31]

Table 15: This table reports the results of predictive regressions of the components 1-period ahead of consumption growth on the components of (log) price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is annual and spans the period 1930-2008.

	Scale j for which $\phi_j \neq 0$			
$j =$	{1,. . . ,7}	{3,6,7}	{6,7}	{7}
	0.66	3.38	1.45	3.36

Table 16: This table reports the estimate of ψ obtained by inverting the relation $\underline{A_m} = \left(\underline{\phi} - \frac{1}{\psi}\right)((I_h - \kappa_1 M))^{-1}$. Given the relation (24) between the price-dividend and the process x_t we estimate the matrix of persistence coefficients of the price-dividend components to approximate M.

	Scale j							
$j =$	8	7	6	5	4	3	2	1
	0	1.5	3.3	0	0	-0.00	0	0
	0	1.6	3.6	0	0	-0.14	0	0

Table 17: This table reports equity premium (in %) $E_t[r_{m,t+1} - r_{f,t}]$ decomposed by level of persistence. Given the relation (24) between the price-dividend and the process x_t we estimate the matrix of persistence coefficients of the price-dividend components to approximate M in a similar way to what has been done in Table 5. We calibrate the terms $\mathbf{Q}$ to the innovation variance-covariance matrix of the dividend price components. The results in the first row are obtained when we set $\Psi = 1.5$ whereas the results in the second row are obtained when we set $\Psi = 3.3$

	Portfolio Returns					
	B1	B2	B3	B4	B5	Market
$Returns(\%)$	4.92	6.45	7.57	8.84	10.14	6.11
$gd(\%)$	1.89	2.11	3.26	3.99	5.08	2.44

Table 18: Summary statistics: portfolio returns and cash flow growth. The variable $B1$ the lowest book-to-market quintile. The data are sampled at the quarterly frequency and cover the first quarter of 1959 through fourth quarter of 2007. Rates of return and log dividend growth are given in annual units.

	Portfolio Risk Measures							
	1	2	3	4	5	6	7	8
	-0.171	-0.070	-0.112	0.165	0.328	0.256	0.519	0.923
$\rho(gd_{5,j}, g_{c,j,t})$	(0.01)	(0.28)	(0.08)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
	0.064	0.041	0.073	0.222	0.096	-0.535	-0.821	-0.969
$\rho(gd_{1,j}, g_{c,j,t})$	(0.32)	(0.53)	(0.26)	(0.00)	(0.15)	(0.00)	(0.00)	(0.00)

Table 19: The table presents correlation between the components of cash flow growth for portfolio 5, $gd_{5,j}$, and 1, $gd_{1,j}$, with the components of consumption growth. P-values for testing the hypothesis of no correlation are in parentheses.

	Scale j							
	1	2	3	4	5	6	7	8
$z_t =$	Panel A: Portfolio 5							
	-2.26	15.12	-7.55	-3.01	6.79	-1.54	0.21	-1.11
pd_t	(-0.43)	(4.54)	(-2.27)	(-0.88)	(2.62)	(-1.81)	(0.17)	(-1.00)
	[0.00]	[0.09]	[0.03]	[0.01]	[0.12]	[0.04]	[0.00]	[0.02]
$z_t =$	Panel B: Portfolio 1							
	-5.34	2.32	-6.45	4.18	-2.30	1.31	0.02	-1.55
pd_t	(-1.29)	(0.91)	(-1.69)	(1.68)	(-1.49)	(1.77)	(0.02)	(-3.14)
	[0.00]	[0.00]	[0.03]	[0.03]	[0.04]	[0.10]	[0.00]	[0.17]

Table 20: This table reports the results of predictive regressions of the components of Portfolio 1 and 5 cash flows growth $\Delta d_{j,t+2^j}$ on the components of (log) price-dividend ratio $pd_{j,t}$. For each regression, the table reports OLS estimates of the regressors, Hansen and Hodrick corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The effective sample is quarterly and spans the period 1947Q4-2007Q4.

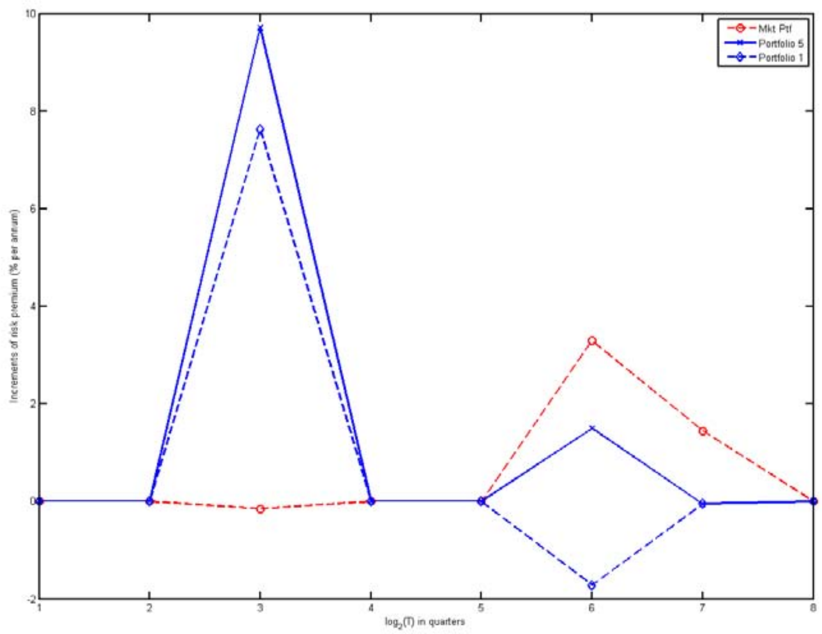

Figure 8: Rates of return are given in annual percentage rates.

A A motivating example

The behavior just observed, namely the biased point estimate and the low t-statistics, can be re conducted to the well-known errors-in-variable (EIV) problem in time series analysis. Indeed suppose the true model for consumption growth rate g_{t+1} be:

$$\begin{aligned} g_{t+1} &= x_t^{(char)} + \sigma\varepsilon_{t+1} \\ x_t &= x_t^{(char)} + x_t^{(spur)} \end{aligned} \tag{A.1}$$

but we wrongly assume the following model to hold:

$$g_{t+1} = \beta_0 + \beta_1 x_t + \eta_{t+1} \tag{A.2}$$

and we try to estimate the true parameter $\beta = 1$ from (A.2). The computation of the OLS slope[20]:

$$\begin{aligned} \widehat{\beta}_1 &= \frac{\sum x_t g_{t+1}}{\sum x_t^2} \\ &= \frac{\sum (\beta x_t^{(char)} + \sigma\varepsilon_{t+1}) x_t}{\sum x_t^2} \\ &= \frac{\sum \beta \; P^{(char)}(x_t) \; x_t}{\sum x_t^2} + \frac{\sum \sigma\varepsilon_{t,t+1} x_t}{\sum x_t^2} \end{aligned}$$

Now assuming that $x_t^{(char)}$, $x_t^{(spur)}$ and ε_{t+1} are mutually independent thus yielding

$$\begin{aligned} plim\left(\frac{1}{T}\sum x_t^2\right) &= \sigma^2_{x^{(char)}} + \sigma^2_{x^{(spur)}} \\ plim\left(\frac{1}{T}\sum P^{(char)}(x_t)\; x_t\right) &= plim\left(\frac{1}{T} P^{(char)}(x_t)\; (P^{(char)}(x_t) + (\mathbb{I} - P^{(char)})(x_t))\right) = \sigma^2_{x^{(char)}} \\ plim\left(\frac{1}{T}\sum x_t \varepsilon_{t,t+1}\right) &= 0 \end{aligned}$$

It then follows that:

$$plim\; \widehat{\beta}_1 = \beta\left(\frac{\sigma^2_{x^{(char)}}}{\sigma^2_{x^{(char)}} + \sigma^2_{x^{(spur)}}}\right)$$

Therefore the p-limit of the OLS slope $\hat{\beta}$ is closer to zero than the true slope β, due to the phenomenon called *attenuation bias.* This is also an example of specification error. Indeed the relevant model is (A.1) but for data reasons we used (4).

[20]For easy of notation we do not include a constant in the regression, i.e. we assume $\beta_0 = 0$

Assume that a projector operator $Proj$ which is able to select $x_t^{(char)}$ out of x_t can be identified, then if we apply it to both sides of (A.2) we obtain:

$$P^{(char)}(g_{t+1}) = \widetilde{\beta}_1 P^{(char)}(x_t) + P^{(char)}(\varepsilon_{t+1}) \tag{A.3}$$

The bias will be now:

$$\begin{aligned}
\widehat{\widetilde{\beta}}_1 - \beta &= \frac{\sum_t P^{(char)}(x_t) P^{(char)}(g_{t+1})}{\sum P^{(char)}(x_t) P^{(char)}(x_t)} - \beta \\
&= \frac{\sum_t P^{(char)}(x_t)[\beta x_t^{(char)} + \sigma P^{(char)}(\eta_{t,t+1})]}{\sum_t P^{(char)}(x_t) P^{(char)}(x_t)} \\
&= \frac{\sum_t P^{(char)}(x_t) \sigma P^{(char)}(\varepsilon_{t,t+1})}{\sum_t P^{(char)}(x_t) P^{(char)}(x_t)}
\end{aligned}$$

where in the second passage we have applied the Projection operator to the true model (A.1) for g_{t+1}. If now we assume that

$$plim \left(\frac{1}{T} \sum P^{(char)}(x_t)\ P^{(char)}(\varepsilon_{t,t+1}) \right) = 0$$

Then we manage to estimate consistently the β.

B Temporal aggregation and a persistence based isometric operator.

In this section Theorem 4 is proved. The proof is based on the application of the abstract Wold theorem. Consider the Hilbert spaces

$$\begin{aligned}\mathcal{H}_{\infty}(\mathbf{x}) &= \left\{ Z = \sum_{n\in\mathbb{Z}} \alpha_n x_n \mid \sum_{n\in\mathbb{Z}} |\alpha_n|^2 < +\infty, \quad \langle Z^1, Z^2 \rangle = \sum_{n\in\mathbb{Z}} \alpha_n^1 \alpha_n^2 \right\} \\ \mathcal{H}_t(\mathbf{x}) &= \left\{ Z = \sum_{k\in\mathbb{N}} \alpha_{t-k} x_{t-k}, \ \in \mathcal{H}_{\infty}^2(\mathbf{x}) \right\}\end{aligned}$$

where $\mathcal{H}_{\infty}(\mathbf{x})$ is the Hilbert space of square integrable linear combinations of elements of the time series $\mathbf{x}_t$ which corresponds at the sequence of elements of the time series $\mathbf{x}$ observed up to time t with the standard scalar product inherited from $l^2(\mathbb{Z})$. As stated in the theorem hypotheses, $x_{t-k} = \Delta y_{t-k} = y_{t-k} - y_{t-k-1}$ where $\mathbf{y}$ is an $I(1)$ process. Without loss of generality we can assume $\mu = 0$ and correspondingly $\mathbf{x}$ to be a weakly stationary process. The persistence based decomposition of x_t in terms of a linear combination of uncorrelated innovations is produced applying the abstract Wold theorem to a different isometric operator. The new isometry operator is given by the composition of the following operators:

Definition 6 *The mean operator M centered at time t is defined by:*

$$\begin{aligned} M &: \mathcal{H}_t(\mathbf{x}) \to \mathcal{H}_t(\mathbf{x}) \\ \mathbf{x} &= \{x_{t-k}\}_{k\in\mathbb{N}} \to M\mathbf{x} = \{\widetilde{x}_{t-k}\}_{k\in\mathbb{N}}, \\ \widetilde{x}_{t-k} &= \frac{x_{t-k} + x_{t-k-1}}{2}\end{aligned}$$

the dyadic dilation operator centered at time t is defined by:

$$\begin{aligned} D &: \mathcal{H}_t(\mathbf{x}) \to \mathcal{H}_t(\mathbf{x}) \\ \widetilde{\mathbf{x}} = \{\widetilde{x}_{t-k}\}_{k\in\mathbb{N}} &\to \mathbf{x}^{(1)} = \left\{ x_{t-k}^{(1)} \right\}_{k\in\mathbb{N}} = D\widetilde{\mathbf{x}} \\ x_{t-k}^{(1)} &= \sqrt{2}\widetilde{x}_{t-2k}\end{aligned}$$

The rescaling operator R centered in t is defined by:

$$R = D \circ M$$

i.e. as the composition of the dyadic dilation operator D with the averaging one M.

The rescaling operator R defines a Wold decomposition of the Hilbert space $\mathcal{H}_t(\mathbf{x})$ in fact the following lemma holds:

Lemma 7 (Isometry_Lemma) *The rescaling operator is isometric in* $\mathcal{H}_t(\mathbf{x})$.

Proof. The rescaling operator R can be represented as the composition of two operators: a rotation matrix *Rot* obtained as the direct sum of two by two rotation matrices:

$$Rot : \bigoplus_{k=0}^{+\infty} \begin{bmatrix} \frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and a projection on the odd components

$$P^{Odd} : \{x_{-n}\}_{n\in\mathbb{N}} \to \{x_{-2n-1}\}_{n\in\mathbb{N}}$$

Their joint action produces

$$P^{Odd} \circ Rot : (x_t, x_{t-1}, x_{t-2}....) \to \left(\frac{x_t + x_{t-1}}{\sqrt{2}}, \frac{x_{t-2} + x_{t-3}}{\sqrt{2}}..... \right)$$

$$Rot : (x_t, x_{t-1}, x_{t-2}....) \to \left(\frac{x_t + x_{t-1}}{\sqrt{2}}, \frac{x_t - x_{t-1}}{\sqrt{2}}, \frac{x_{t-2} + x_{t-3}}{\sqrt{2}}, ... \right)$$

$$P^{Odd} : \left(\frac{x_t + x_{t-1}}{\sqrt{2}}, \frac{x_t - x_{t-1}}{\sqrt{2}}, \frac{x_{t-2} + x_{t-3}}{\sqrt{2}}..... \right) \to \left(\frac{x_t + x_{t-1}}{\sqrt{2}}, \frac{x_{t-2} + x_{t-3}}{\sqrt{2}}..... \right)$$

and this implies that:

$$R = P^{Odd} \circ Rot$$

Direct verification proves that:

$$\langle R\mathbf{x}, R\mathbf{x} \rangle_{\mathcal{H}_t(\mathbf{x})} = \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}_t(\mathbf{x})},$$

in fact by definition:

$$\left\| \frac{x_{t-2k} + x_{t-2k-1}}{\sqrt{2}} \right\|_{\mathcal{H}_t(\mathbf{x})} = \|x_{t-k}\|_{\mathcal{H}_t(\mathbf{x})} \quad \forall k \in \mathbb{N}$$

■

As a consequence the abstract Wold theorem admits a realization in $\mathcal{H}_t(\mathbf{x})$ where the rescaling operator R acts isometrically.

Then application of the abstract Wold theorem grants the existence of the following decomposition:

$$\mathcal{H}_t(\mathbf{x}) = \bigoplus_{j=1}^{+\infty} R^j \mathcal{W}^R \oplus H^\infty$$

where

$$\mathcal{W}^R = \mathcal{H}_t(\mathbf{x}) \ominus R\mathcal{H}_t(\mathbf{x})$$

The generic element of the "innovation " subspace in the new decomposition is given by:

$$\delta_t^{(1)} \propto x_t - LP_{Rx_t}(x_t)$$

it corresponds to the detail which is removed by dyadic averaging. The detail $\delta_t^{(1)}$ has to verify the following conditions:

$$\begin{aligned} x_t &= LP_{\left\{Rx_t, \delta_t^{(1)}\right\}}(x_t) \\ 0 &= \left\langle \delta_t^{(1)}, Rx_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \end{aligned}$$

and for an arbitrary k all the constraints are satisfied by:

$$\delta_{t-k}^{(1)} = \frac{x_{t-k} - x_{t-k-1}}{2} \tag{B.1}$$

The detail subspace defines the information of the time series which is "removed" by aggregation or equivalently the "innovation" which is discovered when the resolution is increased. Note that the "normalized innovation" would be $\sqrt{2}\delta_{t-2k}^{(1)}$. The detail subspace for $\delta_{t-4k}^{(2)}$ is determined by the conditions

$$\begin{aligned} x_t &= LP_{\left\{\delta_t^{(1)}, \delta_t^{(2)}, R^2 x_t\right\}}(x_t) \\ 0 &= \left\langle \delta_t^{(1)}, Rx_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \\ 0 &= \left\langle \delta_t^{(2)}, R^2 x_t \right\rangle_{\mathcal{H}_t(\mathbf{x})} \end{aligned}$$

hence

$$\delta_t^{(2)} \propto \pi_t^{(1)} - LP_{(R^2x)_t}(x_t)$$

and a possible choice is given by:

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

The $j-th$ iteration of the above construction implies that at resolution j the (unnormalized) detail, $\delta_t^{(j)}$ is given by:

$$\begin{aligned} \delta_{t-2^j k}^{(j)} &= \left(\frac{(R^j x)_{t-2^j k}}{\sqrt{2^j}} - \frac{(R^{j+1} x)_{t-2^j k}}{\sqrt{2^{j+1}}} \right) \\ \delta_{t-2^j k}^{(j)} &= \pi_{t-2^j k}^{(j-1)} - \frac{\pi_{t-2^j k}^{(j-1)} + \pi_{t-2^j (k+1)}^{(j-1)}}{2} = \frac{\pi_{t-2^j k}^{(j-1)} - \pi_{t-2^j (k+1)}^{(j-1)}}{2} \end{aligned}$$

By the assertion of the abstract Wold theorem the space H^{∞} is determined by

$$H^{\infty} = \cap_{j=0,..,+\infty} R^{j} \mathcal{H}_t\left(\mathbf{x}\right)$$

and can be identified by that component which is not removed by an arbitrary number of averaging operations. Theorem 4 can be easily proved applying the CLT for stationary time series

Proof. Observe that the projection $\pi^{(\infty)}$ of $\mathbf{x}$ on $H^{(\infty)}$ is given by:

$$\pi^{(\infty)} = \lim_{J\to\infty} \sqrt{2^j}\pi^{(J)} = \lim_{J\to\infty} \sqrt{2^j}\left(\frac{\sum_{n=0}^{2^j} x_{t-n} - \mu}{2^j}\right)$$

the existence of this limit can be derived from the well known central limit theorem for stationary processes presented in Hall and Heyde (1980) Corollary 5.2 pg.135. That result applies to the $I\left(0\right)$ class of processes and proves that the limiting procedure converges in distribution and $\pi^{(\infty)}$ has the limit distribution corresponding to a normal law with variance $2\pi f\left(0\right)$ where $f\left(0\right)$ is the zero frequency component of the spectrum $f\left(\lambda\right)$ which characterizes the stationary process. Since x_t is an linear $I\left(0\right)$ process, it has an absolutely summable infinite moving average representation, this implies

$$f\left(0\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} \gamma\left(k\right) = \psi\left(1\right)^2$$

hence point 1 is proved. Point 2 is immediately verified since each detail time series is obtained by a recursive application of a finite difference filter to the original time series which is wide sense stationary. ■

C Computational Algorithms and Procedures

C.1 Redundant Persistence-Based Decomposition

A **redundant** transform based on an n-length input time series X_t has an n length resolution scale for each of the resolution levels that we consider.

The non-decimated a trous wavelet algorithm can be described simply as follows. Consider the creation of the first wavelet resolution level. The scale coefficients $\pi_t^{(j)}$ are generated from the scale coefficients $\pi_t^{(j-1)}$ by convolving the latter with a low-pass filter h:

$$\pi_t^{(j+1)} = \sum_{l=-\infty}^{+\infty} h(l)\pi_{t+2^j l}^{(j)}$$

where the finest scale is the original series: $\pi_t^{(0)} = X_t$. The choice of Haar wavelet implies $h = (\frac{1}{2}, \frac{1}{2})$:

$$\begin{aligned} \pi_t^{(1)} &= \frac{1}{2}[\pi_t^{(0)} + \pi_{t-1}^{(0)}] \\ \delta_t^{(1)} &= \pi_t^{(0)} - \pi_t^{(1)} \end{aligned}$$

At the level of aggregation j we get:

$$\begin{aligned} \pi_t^{(j+1)} &= \frac{1}{2}[\pi_t^{(j)} + \pi_{t-2^j}^{(j)}] \\ \delta_t^{(j+1)} &= \pi_t^{(j)} - \pi_t^{(j+1)} \end{aligned}$$

At any time point, t, we never use information (time-wise) after k in calculating the wavelet coefficient, thus the algorithm produces an adapted (non anticipative) decomposition.

C.2 Isometric Haar Matrix Transform

If we let $W = \{\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(J)}, \pi^{(J)}\}$ represents the details coefficients then we can write

$$\mathbf{W} = \mathcal{T}^{(2^J)}\mathbf{X}$$

where $\mathbf{W}$ is a column vector of length $t = 2^J$ whose n-th elements is the n-th details coefficient and $\mathcal{T}^{(J)}$ is an $T \times T$ real-valued matrix defining the transform and satisfying ${\mathcal{T}^{(J)}}^T \mathcal{T}^{(J)} = I_T$. It is important to recall that the n-th details coefficient is associated with a particular scale and with a particular set of times. We give now an example of the transformation matrix $\mathcal{T}^{(J)}$ for the case $J = 2$ and $J = 3$. For the case $J = 2$ the matrix takes the following form

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

For the case $J = 3$ the matrix takes the following form

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

References

ALVAREZ, F., AND U. J. JERMANN (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73(6), 1977–2016.

ANG, A., AND G. BEKAERT (2007): "Stock Return Predictability: Is It There?," *Review of Financial Studies*, 20(3), 651–707.

BACKUS, D., B. ROUTLEDGE, AND S. ZIN (2009): "The Cyclical Component of US Asset Returns," Working Paper 10597.

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

BANSAL, R., R. DITTMAR, AND D. KIKU (2007): "Cointegration and Consumption Risks in Asset Returns," *Review of Financial Studies*, 22(3), 1343–1375.

BANSAL, R., R. F. DITTMAR, AND C. T. LUNDBLAD (2005): "Consumption, Dividends, and the Cross Section of Equity Returns," *Journal of Finance*, 60(4), 1639–1672.

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

BANSAL, R., A. YARON, AND D. KIKU (2007): "Risk for the Long Run: Estimation and Inference," Working paper.

BEELER, J., AND J. Y. CAMPBELL (2009): "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," NBER Working Papers 14788, National Bureau of Economic Research, Inc.

BEVERIDGE, S., AND C. R. NELSON (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'," *Journal of Monetary Economics*, 7(2), 151–174.

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

BOUDOUKH, J., M. RICHARDSON, AND R. F. WHITELAW (2008): "The Myth of Long-Horizon Predictability," *Review of Financial Studies*, 21(4), 1577–1605.

BROCKWELL, P. J., AND R. A. DAVIS (1996): *Introduction to time series and forecasting.* Springer.

CALVET, L. E., AND A. J. FISHER (2007): "Multifrequency news and stock returns," *Journal of Financial Economics*, 86(1), 178–212.

Campbell, J. Y., and R. J. Shiller (2001): "Valuation Ratios and the Long-run Stock Market Outlook: An Update," Cowles Foundation Discussion Papers 1295, Cowles Foundation, Yale University.

Campbell, J. Y., and L. M. Viceira (2005): "The Term Structure of the Risk-Return Tradeoff," CEPR Discussion Papers 4914, C.E.P.R. Discussion Papers.

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

Canova, F. (1998): "Detrending and business cycle facts," *Journal of Monetary Economics*, 41(3), 475–512.

Christiano, L. J., and T. J. Fitzgerald (2003): "The Band pass filter," *International Economic Review*, 44(2), 435–465.

Constantinides, G. M., and A. Ghosh (2008): "Asset Pricing Tests with Long Run Risks in Consumption Growth," NBER Working Papers 14543, National Bureau of Economic Research, Inc.

Daubechies, I. (1990): "The wavelet transform, time–frequency localization and signal analysis," *IEEE Trans. Info. Theory*, 36(5), 961–1005.

Daubechies, I. (1992): *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*. Soc for Industrial & Applied Math.

Engle, R. F. (1974): "Band Spectrum Regression," *International Economic Review*, 15(1), 1–11.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–69.

——— (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99(2), 263–86.

Garleanu, N. B., S. Panageas, and J. Yu (2009): "Technological Growth and Asset Pricing," NBER Working Papers 15340, National Bureau of Economic Research, Inc.

Gencay, R., and Y. Fan (2008): "Unit Root Tests with Wavelets," *Econometric Theory, Forthcoming.*

Gencay, R., and N. Gradojevic (2009): "Errors in Variables Estimation with No Instruments," Working Paper Series wp30 09, Rimini Centre for Economic Analysis.

Gencay, R., F. Selcuk, and B. Whitcher (2001): *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116(2), 260–302.

Hansen, L. P., and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *The Journal of Political Economy*, 88(5), 829–853.

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

Kaltenbrunner, G., and L. A. Lochstoer (2007): "Long-Run Risk Through Consumption Smoothing," *SSRN eLibrary*.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Lettau, M., S. Ludvigson, and J. Wachter (2007): "The declining equity premium: what role does macroeconomic risk play?," *Review of Financial Studies*, 21(4), –.

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

Mallat, S. G. (1989a): "Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R)," 315(1).

Mallat, S. G. (1989b): "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Ortu, F., A. Tamoni, and C. Tebaldi (2010): "Technical Appendix to "Long Run and the Temporal Aggregation of Risks" , URL: `http://www.igier.unibocconi.it/finance`," Discussion paper.

Parker, J. A., and C. Julliard (2005): "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy*, 113(1), 185–222.

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

Torous, W., R. Valkanov, and S. Yan (2004): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business*, 77(4), 937–966.

Valkanov, R. (1999): "Long Horizon Regressioms: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 201–232.

Wold, H. (1938): *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist and Wiksell.

Yu, J. (2009): "The Long and the Short of Asset Prices: Using Long Run Consumption-Return Correlations to Test Asset Pricing Models," *SSRN eLibrary*.