DECLARATORIA SULLA TESI DI DOTTORATO
Da inserire come prima pagina della tesi

Il sottoscritto

COGNOME Tamoni

NOME Andrea Giovanni

Matr. 1194539

Titolo della tesi:

Essays in Asset Pricing

Dottorato di ricerca in Finanza

Ciclo XXII

Tutor del dottorando Carlo Ambrogio Favero

Anno di discussione 2011

DICHIARA

sotto la sua responsabilità di essere a conoscenza:

1) che, ai sensi del D.P.R. 28.12.2000, N. 445, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici previsti dalla presente declaratoria e da quella sull'embargo;

2) che l'Università ha l'obbligo, ai sensi dell'art. 6, comma 11, del Decreto Ministeriale 30 aprile 1999 prot. n. 224/1999, di curare il deposito di copia della tesi finale presso le Biblioteche Nazionali Centrali di Roma e Firenze, dove sarà consentita la consultabilità, fatto salvo l'eventuale embargo legato alla necessità di tutelare i diritti di enti esterni terzi e di sfruttamento industriale/commerciale dei contenuti della tesi;

3) che il Servizio Biblioteca Bocconi archivierà la tesi nel proprio Archivio istituzionale ad Accesso Aperto e che consentirà unicamente la consultabilità on-line del testo completo (fatto salvo l'eventuale embargo);

4) che per l'archiviazione presso la Biblioteca Bocconi, l'Università richiede che la tesi sia consegnata dal dottorando alla Società NORMADEC (operante in nome e per conto dell'Università) tramite procedura on-line con contenuto non modificabile e che la Società Normadec indicherà in ogni piè di pagina le seguenti informazioni:

- tesi di dottorato (titolo *tesi)* ....Essays in Asset Pricing.................................. ;
- di *(cognome e nome del dottorando*) Andrea Giovanni Tamoni .... ;
- discussa presso l'Università commerciale Luigi Bocconi – Milano nell'anno ...2011.. (*anno di discussione);*
- La tesi è tutelata dalla normativa sul diritto d'autore (legge 22 aprile 1941, n.633 e successive integrazioni e modifiche). Sono comunque fatti salvi i diritti dell'Università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte;
- **solo nel caso sia stata sottoscritta apposita altra dichiarazione con richiesta di embargo**: La tesi è soggetta ad embargo della durata di .... mesi (indicare durata embargo);

5) che la copia della tesi depositata presso la NORMADEC tramite procedura on-line è del tutto identica a quelle consegnate/inviate ai Commissari e a qualsiasi altra copia depositata negli Uffici dell'Ateneo in forma cartacea o digitale e che di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;

6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi (legge 22 aprile 1941, n.633 e successive integrazioni e modifiche), ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto l'Università è in ogni caso esente da responsabilità di qualsivoglia natura, civile, amministrativa o penale e sarà dal sottoscritto tenuta indenne da qualsiasi richiesta o rivendicazione da parte di terzi;

7) **scegliere l'ipotesi 7a o 7b indicate di seguito**: 7a

7a)✓ che la tesi di dottorato non è il risultato di attività rientranti nella normativa sulla proprietà industriale, non è stata prodotta nell'ambito di progetti finanziati da soggetti pubblici o privati con vincoli alla divulgazione dei risultati; non è oggetto di eventuali registrazioni di tipo brevettale o di tutela, e quindi non è soggetta a embargo;

Oppure

7b) che la tesi di Dottorato rientra in una delle ipotesi di embargo previste nell'apposita dichiarazione "**RICHIESTA DI EMBARGO DELLA TESI DI DOTTORATO**" sottoscritta a parte.

Data 26/01/2011

F.to (indicare nome e cognome) Andrea Giovanni Tamoni

Essays in Asset Pricing

Dissertation in partial fulfillment of the requirements for the academic degree of Doctor of Philosophy in Finance (XXII cycle).

Andrea Giovanni Tamoni (1194539)
Università Commerciale Luigi Bocconi, Milan

Thesis Committee :

Carlo A. Favero, IGIER and Department of Finance, Bocconi University
Fulvio Ortu, IGIER and Department of Finance, Bocconi University
Sydney C. Ludvigson, Department of Economics, New York University

Contents

Acknowledgments

The writing of this dissertation has been an incredible journey and a monumental milestone in my academic life. I could not have embarked on this expedition and traveled this far without the passionate and continued support of advisors, colleagues, friends and family.

First of all, I am deeply indebted to my advisors Carlo A. Favero, Fulvio Ortu and Claudio Tebaldi for their helpful comments, suggestions and especially for having encouraged me during all these years. Part of this thesis has been written while I was visiting the Department of Economics at New York University. Sydney Ludvigson has been very nice in accepting to be part of my thesis committee: I wish to thank her for the helpful comments. I also thank David Backus, Thomas J. Sargent and Stanley E. Zin for their help and their suggestions; a special thought goes to Alberto Bisin.

I would also like to thank all of my friends for all the good times during these years. Many other people have been great friends in these years: among them, special thanks go to Alessandro, Andrea T., Arie, Federico, Felix. In particular I would like to thank Arie. I could not do the things I have done without his friendship, collaborations and many stimulating discussions.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. And most of all I would like to thank Serena. Her love, encouragement, support, and patience constantly inspired me to better myself and aim higher during all these years. I am at a loss for words. Thank you.

Preface

This dissertation consists of four essays in the field of asset pricing. Each of the papers stress the importance of classifying aggregate shocks that impinge an economy according to the duration of their effects. The separation between transient and permanent components is an extreme example of this classification: effects of transient components are expected to vanish esponentially faster than the persistent ones. While traditional linear time series analysis is based on Fourier spectral analysis in the first and second chapters we introduce new decomposition of time series which separates effects over different time scales. Such decomposition relies on the notion of multiresolution analysis, a spectral approach alternative to traditional Fourier spectral techniques whose application is limited to the analysis of stationary time series. Using such a decomposition we then re-examines, within a long-run risk paradigm, the price response to the aggregate risk, represented by shocks to the conditional mean and volatility of consumption growth, disentangling it over different horizons. The third and fourth chapters instead documents the existence in the price-dividend ratio of fluctuations with different periodicity and their impacts on stock market risk and return.

The first chapter : *Persistence Heterogeneity in Volatility and Long-Run Risk-Return Trade-Off*, reexamines the intertemporal risk-return relation using a more sensible empirical specification that is motivated by two concerns: first, long-run market returns is predicted by past long-run market variance; second, such empirical finding cannot be reproduced by simple aggregation of short-term risk-return models under a classical autoregressive process for variance (Bandi and Perron, 2008). We therefore supply a new asset pricing model with heterogeneous persistence in the consumption growth variance. Our framework considers the stochastic variance as the sum of many components realized over different time horizons. The heterogeneity of the model derives from the fact that at each time scale, the variance is described by a different autoregressive structure. Our model has three major implications. First, it suggests to study the relation between the conditional mean and conditional variance of stock market returns across different levels of persistence without necessarily implying a positive short-run risk-return relation. By disaggregating the mean and variance of returns across different levels of persistence we find evidence in favor of a positive long-run risk-return trade-off. Second, the estimation of the risk-return trade-off considering the information contained in all the frequency components delivers a risk-aversion coefficient of about 4 which lines up well with the theory. Third we find strong support in favor of the hypotheses that much of the variation in asset prices can be attributed to fluctuations in persistent components of consumption volatility. Incidentally these very same persistent components are a good indicator of economic uncertainty and are linked to the persistence stock market risk components. Thus we are able to identify well-known macroeconomic factors as key drivers of market returns in the long run. Last but not least movements in the short rate at business cycle frequencies are mostly explained by movements in risk.

The second chapter is titled *Long Run Risk and the Persistence of Consumption Shocks* and is joint work with Prof. Fulvio Ortu and Prof. Claudio Tebaldi. The essay proposes an extension of the long-run valuation model of R.Bansal and A.Yaron (2004) explicitly designed to optimally account for the persistence heterogeneity of shocks to consumption growth mean. 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. 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 and produce sizeable estimates for the equity premia.

The third chapter is titled *Demographics and the Term Structure of Stock Market Risk* and is joint work with Prof. Carlo A. Favero. The term structure of stock market risk depends on the predictability of stock market returns at different horizons. Intuitive reasoning, formal modeling and empirical evidence show that demographic trends are a low frequency information variable, that determines the slow evolving mean of the dividend-price ratio and has a forecasting power for stock market returns that increases with the horizon. A potential role for demographic variables has never been considered in the ongoing debate on the slope of the term structure of stock market risk. Therefore the chapter analyzes the consequences of demographics for the term structure of stock market risk. We show that the forward solution of the dynamic dividend growth model augmented with demographics delivers a negative sloping term structure of stock market risk. Direct regressions of returns at different horizons on the relevant predictors are much better suited to capture this feature of the model while VAR based multi-period iterated forecasts do not, as they are derived from a backward solution of a reduced form empirical model. These results are very little affected by parameters' uncertainty, as very few parameters are very precisely estimated in the relevant empirical model. Moreover, results are robust to the existence of "imperfect predictors" as forecast of returns in the direct regression approach involve only currently observable variables and no projections of future variables. Overall the combination of direct regression with the use of demographic trends leads us to find a steeply downward sloping term structure of stock market risk.

The last chapter is titled *Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns* and is joint work with Prof. Carlo A. Favero and Arie E. Gozluklu. The dynamic dividend growth model linking the log dividend yield to future expected dividend growth and stock market returns has been extensively used in the literature for forecasting stock returns. The empirical evidence on the performance of the model is mixed as its strength varies with the sample choice. This model is derived on the assumption of stationary log dividend-price ratio. The empirical validity of such hypothesis has been challenged in the recent literature (Lettau and Van Nieuwerburgh, 2008) with strong evidence on a time varying mean, due to breaks, in this financial ratio. In this paper, we pursue two distinct aims. First, we show using annual US data that the slowly evolving mean toward which the dividend price ratio is reverting is driven by a demographic factor consistent with the predictions of the theoretical model by Geanakoplos et al. (2004). We then exploit stability analysis for long-run economic relationships to construct an equilibrium dividend-price ratio. Second, relying upon the exogeneity and predictability of the demographic factor, we use our measure of disequilibrium obtained as the difference between the actual dividend yield and the equilibrium dividend yield to forecast market excess returns at different horizons (up to 10 years) and evaluate the forecasting performance of the model based on the corrected dividend-price ratio against different alternative specifications. We show that an empirical model using information in long-run factors overperforms virtually all alternative models proposed in the literature within the framework of the dynamic dividend growth model.

Part I

Persistence Heterogeneity in Volatility and Long-Run Risk-Return Trade-Off

1 Introduction

Merton (1973) in his seminal intertemporal capital asset pricing model (ICAPM), shows that the expected market excess return is a positive function of the markets conditional variance, which can be stated, ignoring a hedge component, as follows:

$$E_t\left[r_{t+1}\right] = \alpha + \gamma\ \sigma_{m,t}^2 \tag{1}$$

where r_{t+1} denotes excess market returns, $\sigma_{m,t}^2 \equiv Var_t(r_{t+1})$, γ is a parameter reflecting the relative risk aversion of the representative agent and, according to the model, α should be equal to zero. The expectation and the variance of the market excess return are conditional on the information available at the beginning of the return period, time t. We call this relation the short-run risk return trade-off since returns and variance are measured at the highest frequency possible.[1] Unfortunately, the trade-off has been hard to find in the data. Previous estimates of the relation between risk and return often have been insignificant and sometimes even negative, see Lettau and Ludvigson (2010) for a thorough discussion.[2]

Whereas much work has been devoted to assessing the validity of the classical short-run risk-return trade-off, the recent empirical findings in Bandi and Perron (2008) shows that although the dependence between excess market returns and past market variance is statistically mild at short horizons (thereby leading to a hard-to-detect risk-return trade-off, as in the existing literature) it increases with the horizon and is strong in the long-run (i.e., between 6 and 10 years). More formally consider regressions of the type:

$$r_{t,t+h} = \alpha_h + \beta_h \sigma_{m,t-h,t}^2 + \epsilon_{t,t+h}$$

where $r_{t,t+h}$ denotes excess market returns between months t and $t+h$, $\sigma_{m,t-h,t}^2$ denotes past market variance, and $\epsilon_{t,t+h}$ is a forecast error. For aggregation levels $h = 84, 96, 108$, and 120 (i.e. 7-10 years) the former relation reveals a pronounced dependence as highlighted by a monotonously increasing R^2 that peaks to 72% at the 10-year horizon. We call this relation the long-run risk-return trade-off.

Importantly, Bandi and Perron (2008) explore the long-run implications of traditional short-term risk-return models such as (1) augmented with a classical (autoregressive) process

[1] Tests of the short-run risk return trade-off have been mainly conducted with returns computed for a holding period of one month.

[2] French, Schwert and Stambaugh (1987), Baillie and DeGennaro (1990) and Campbell and Hentschel (1992) do find a positive albeit mostly insignificant relation between the conditional variance and the conditional expected return. In contrast, Campbell (1987) and Nelson (1991) find a significantly negative relation. Glosten, Jagannathan and Runkle (1993), Harvey (2001), and Turner et al. (1989) find both a positive and a negative relation depending on the method used.

for variance, and prove that simple temporal aggregation of such short-term specification cannot imply the long-term results.

In this paper, we therefore take a new look at the risk-return trade-off by proposing a consumption-based asset pricing models which, by allowing for richer consumption variance dynamics, reconcile these findings. Importantly our pricing model shed lights into the nature of long-run market risk, and the economic channel through which past market variance proxies for it.

Regarding the modelling of stochastic consumption variance, our approach builds on recent applications of multiresolution analysis to asset pricing and econometrics, see Ortu, Tamoni and Tebaldi (2010b). In particular the authors show that news 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. Intuitively, at short horizons, a variety of corporate and macroeconomic announcements affect investor forecasts of future cash flows and discount rates, and in turn the dynamics of daily returns. However equity markets also price diverse lower-frequency fundamentals such as demographics (Abel, 2003), technological innovation (Pastor and Veronesi, 2005), and variations in consumption, dividends, and macroeconomic uncertainty (Bansal and Yaron, 2004; Lettau, Ludvigson, and Wachter, 2004). To accommodate this diversity, the present paper embeds the persistence based decomposition developed in Ortu et al. (2010b) within an asset pricing equilibrium where the consumption variance is the sum of many autoregressive processes each one with its own level of persistence. Overall the variance is hit by exogenous shocks of heterogeneous durations and is described by a different autoregressive structure at each time scale.

The model links consumption and market risk to asset prices. However these relations are horizon dependent. In particular the model predicts a disaggregated risk-return relation for each level of persistence. With the goal of evaluating the empirical performance of the proposed asset pricing model, we apply the persistence based decomposition to both returns, conditional market and consumption variance, and we extract the components at different levels of persistence of these series. Using quarterly market return data from 1949 to 2009 we find a positive and statistically significant relation between risk and return at medium and long horizon. Once we aggregated back our series we obtain the very same results as in Bandi and Perron (2008). By accounting for the different frequency components in returns and consumption variance we also manage to estimate a positive risk aversion coefficient equal to 4.06, which lines up well with economic intuition about a reasonable level of risk aversion. Whereas the main difficulty in testing the ICAPM relation has been attributed to the fact that the conditional variance of the market is not observable and must be filtered from past returns and therefore the conflicting findings in past studies are mostly the result of differences in the approach to modeling the conditional variance, our findings show that what matters for the tests of the risk-return trade-off is not so much the way expected returns and conditional variance are measured but the fact that the relation must consider the persistence heterogeneity in the time series of interest all levels of persistence. Finally our model uncovers an important relation between macroeconomic uncertainty, consumption risk and asset prices. In particular our empirical results show that the "macroeconomic uncertainty", as measured by the persistent components of consumption growth variance, is an important source of aggregate risk and properly valued in asset markets.

Our study complements earlier research that investigates the risk-return relation using data at different frequencies. A notable contribution is Ghysels, Santa-Clara and Valkanov (2005) who study the predictability of stock returns over relative low frequencies (monthly or quarterly) from high frequency volatility estimates. More generally our paper is related to the the recent body of work that deal with the issue of estimating mixed frequency models. In this context, the MIDAS (Mixed Data Sampling) model of Ghysels, Sinko and Valkanov (2007) has recently gained considerable attention. The MIDAS method allows regressions of some low frequency variable onto high frequency variables. Our methodology differs from MIDAS in that we run regression of specific frequency components of the regressand onto the same frequency components of regressors. An alternative mixed frequency modeling approach is represented by the Markov-switching multi-fractal model suggested in Calvet and Fisher (2001) which allows to investigate the role of persistence heterogeneity in (volatility) time series. With this regard our work is close to Calvet and Fisher (2007) who investigate the role of persistence heterogeneity in volatility in a partial equilibrium set-up by means of non linear regime switching multifractal models. Similar to their work our modeling approach generate consumption volatility dynamics that accounts for the persistence heterogeneity. However the use of the persistence based decomposition allows us to retain analytical tractability and a linear specification.

The rest of the paper is organized as follows. Section 2 briefly motivates the new dynamic processes used for the state variables in our model. We then introduce the long-run risk asset pricing model with persistence heterogeneity in volatility components, and solves various functions of interest analytically. Section 3 examines empirically the implications of the models and sheds new light on the classical risk-return relation. We also investigate the model ability to explain the major problems facing the Bansal and Yaron (2004) and Bansal, Yaron and Kiku (2007) ones. Section 4 concludes. All derivations are in the appendix.

2 The Long-run Risk Model with Persistence Heterogeneity in Volatility

Before introducing our asset pricing model which will deliver a persistence-based risk return trade-off, we first show in subsection 2.1 that aggregation of short-term returns delivered by the standard long-run risk models, see Bansal and Yaron (2004), does not provide a first-order role for market payoff risk in long-run returns. We will then look at the time series and spectral properties of the market and consumption variance to motivate the new dynamic processes used for the state variables in our model. In subsection 2.2 we eventually introduce and solve the new long-run risks model with persistence heterogeneity in volatility in terms of the state variables and provide approximate solutions to functions of interest: the consumption-wealth ratio, market prices of risks, price-dividend ratio, and the market return volatility.

2.1 Persistence Heterogeneity in Consumption and Market Variance

In this section we first present a simplified version of the Bansal and Yaron (2004) model. This paired with empirical analysis allow us to define the key ingredients necessary for such long-run models to lead to the pricing implications for long-run expected returns.

In order to make our point we closely follow the approach of Bollerslev, Tauchen and Zhou (2009) and Zhou (2010) and suppose that the growth rate of consumption in the economy is not predictable. The data generating process for consumption and dividends is then as follows:

$$\begin{aligned} g_{t+1} &= \mu + \sigma_t \eta_{t+1} \\ gd_{t+1} &= \mu_d + \phi_d \sigma_t u_{t+1} \\ \sigma^2_{t+1} &= \sigma^2 + \rho(\sigma^2_t - \sigma^2) + \sigma_w w_{t+1} \end{aligned}$$

Relying on log-linear approximations for the log return on the market portfolio BY04 show that, in equilibrium, the log price-dividend ratio z^m_t is an affine function of the state variable, and that the variance the stock market volatility is proportional to the consumption uncertainty:

$$var_t(r_{m,t+1}) = \phi_d^2 \sigma_t^2 + \beta_{m,w} \sigma_w^2 \tag{2}$$

It is a noteworthy point that, given the assumed autoregressive dynamics for consumption variance, the market variance too follows a classical autoregressive process. Therefore we can apply the results in Proposition 3 in Bandi and Perron (2008) and we can conclude that dependence between excess market returns and autoregressive conditional variance cannot deliver the long-run risk-return results upon aggregation. Moreover note that in this simple set-up the stock market variance is proportional to the consumption uncertainty, see again (2), and thus it shares the same persistence properties. This is not the case in the data however. Although the market variance is highly positive dependent at short horizons as it is well-known (see for instance Bandi and Perron (2008)), we find an autocorrelation value equal to about 0.65 and relatively low compared with the consumption variance persistence. The persistence of consumption volatility is in fact much higher and close to 0.85 at quarterly horizons.[3] This is consistent with the literature investigating the volatility of U.S. economy, see for instance Justiniano and Primiceri (2008), that assume that the macroeconomic log-volatility behaves as a geometric random walk process. This evidence suggests that the speed of mean reversion towards a constant mean of the consumption variance is very different from that of stock returns variance. However it is not a simple matter of describing the consumption and returns volatilities with two autoregressive processes with different roots. In fact let's assume that realized stock market variance mean reverts at the speed of $\rho = 0.65$ at the time scale $\tau = 1$ quarter. Upon aggregation up to $T = 8$ quarters it is possible to show[4] that the market variance over T periods would then have an autocorrelation equal to 0.35. Once again this is not the case in the data since as soon as we aggregate stock market volatility at longer horizons than its persistence becomes quickly close to 1. The Figure 1 displays this fact by showing the realized stock market volatility at two different levels of aggregation, namely one quarter and two years. We thus argue that important long-run fluctuations are present in the stock market volatility but one needs to disentangle them from the high frequency components that at very short horizon bring down its persistence.

This statement is further supported by (spectral) variance decomposition of the different time series across frequencies. We report such decomposition in Tables 4 and 5 for the consumption volatility and the return volatility, respectively. The variance decomposition shows that the great part of the variability in consumption and returns volatility is explained by low frequency (i.e. high j) and high frequency (i.e. low j) components, respectively. This is also reflected in the autoregressive coefficient obtained by fitting an AR(1) process on the aggregate time series. By doing so one would capture the highest in variance component of the aggregate time series. Thus it is not surprising that the estimate autoregressive coefficient for returns variance is lower than the one for consumption variance. To uncover the long-run

[3]This persistence refers to the Bayesian filtered volatility estimated according to the procedure described in Appendix C. To ensure the robustness of this result we also measure consumption volatility following both Campbell (2003) and Bansal, Khatchatrian and Yaron (2005) methods. The general pattern of the aggregate persistence level using these methods is very similar to those using the Bayesian filtering method

[4]For any time series described by an AR(1) process, if ρ is the autocorrelation on a scale τ then upon aggregation we have the following expressions for the covariance and variance of the aggregate process $\sum_{s=1-T}^{T-1}(T-|s|)\sum_{s'=1-T}^{T-1}(T-|s'|)\rho^{|T+s-s'|}$ and $\sum_{s=1-T}^{T-1}(T-|s|)\sum_{s'=1-T}^{T-1}(T-|s'|)\rho^{|s-s'|}$.

relations between asset prices, macroeconomic uncertainty[5] and market risk one then needs a filtering procedure to disentangle the small in volatility but highly persistent components of market variance from the the highly volatile low persistent components.

The above evidence suggest that an asset pricing model to be successful in explaining the long-run risk return trade-offs must allow consumption and stock market variances to have different persistence levels and let however stock market volatility to have important low-frequency movements. Overall a more realistic description of the persistence properties of the relevant time series is in order. We then rely on the technique suggested in Ortu et al. (2010b) which decomposes the time series, in a sequence of 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. This allows each time series to be driven by different components with different degrees of persistence. Whereas in Ortu, Tamoni and Tebaldi (2010a) and Ortu et al. (2010b) we deal in detail with the tool to decompose the time series of interest into components with different levels of persistence here we briefly introduce this tool in Appendix B and we just use it in the next section to introduce persistence heterogeneity for the variance of consumption and dividend processes in a standard set-up close to the one used in Bansal and Yaron (2004) and Bansal et al. (2007).

2.2 The Long-run Risk Model with Persistence Heterogeneity in Volatility

Following the approach discussed in Appendix B, we incorporate in the standard long-run risk model the decomposition of time series into components with different levels of persistence so that the log consumption growth, g_t, and the log dividend growth, gd_t take the following form:

$$g_t = \sum_{j=1}^{J} g_{j,t}$$
$$gd_t = \sum_{j=1}^{J} gd_{j,t}$$

where $g_{j,t}$ and $gd_{j,t}$ denote the components with level of persistence j. Now, we closely follow the approach of Bollerslev et al. (2009) and Zhou (2010) and suppose that the growth rate of consumption in the economy is not predictable. The novelty now is to assume that each component of consumption growth, $g_{j,t}$ and of dividend growth, $gd_{j,t}$ is driven by its own

[5] We define as macroeconomic uncertainty the variance factors extracted from consumption growth rate.

stochastic volatility, $\sigma_{j,t}$, i.e.:

$$g_{j,t+2^j} = \sigma_{j,t}\ e^g_{j,t+2^j} \tag{3}$$
$$e^g_{j,t+2^j} \sim N(0,1)$$
$$gd_{j,t+2^j} = \varphi_{d,j}\sigma_{j,t}\ e^d_{j,t+2^j} \tag{4}$$
$$e^d_{j,t+2^j} \sim N(0,1)$$

with the shocks $e^g_{j,t+2^j}$ and $e^d_{j',t+2^{j'}}$ being correlated for $j = j'$ and uncorrelated otherwise. To close the dynamics of the model we assume that each of the stochastic variance components $\left\{\sigma^2_{j,t}\right\}_{j=1}^J$ follows its own multiscale autoregressive process, i.e.

$$\sigma^2_{j,t+2^j} = \rho_j\sigma^2_{j,t} + \varepsilon_{j,t+2^j} \tag{5}$$
$$\varepsilon_{j,t+2^j} \sim N\left(0, \left(\sigma^{(j)}\right)^2\right)$$

This is the key innovation of our model. We consider variances realized over different time horizons. The heterogeneity of the model derives from the fact that different autoregressive structures are present at each time scale. Importantly, equations (3) to (5) represent a natural way to incorporate persistence heterogeneity in the macroeconomic uncertainty framework while retaining its pedagogical simplicity. Note that in this simple specification of the model variance can go negative. To ensure the positivity of the variance process, one can pursue the approach of Barndorff-Nielsen and Shephard (2001) and assume that the innovations in the components of volatility processes follow a Gamma distribution. This would just slightly complicate the algebra without adding much to the intuition to the model. In the following we therefore assume that consumption volatility shocks are Gaussian to explain the major implications of the model. However when we estimate the consumption volatility components we impose positivity by estimating the log-volatility.

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

$$E_t\left[e^{m_{t+1}+r^i_{t+1}}\right] = 1 \tag{6}$$

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

$$m_{t+1} = \theta\log\beta - \frac{\theta}{\psi}g_{t+1} + (\theta-1)r^a_{t+1}\ , \tag{7}$$

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

In what follows we provide the basic steps to determine the pricing kernel and risk premia on the market portfolio in our long-run risk model with persistence heterogeneity.[6] Recall first that by the standard Campbell and Shiller (1988) log-linear approximation for returns one obtains:

$$\begin{aligned} r_{a,t+1} &= \kappa_0 + \kappa_1 z^a_{t+1} - z^a_t + g_{t+1} \\ r_{m,t+1} &= \kappa_{0,m} + \kappa_{1,m} z^m_{t+1} - z^m_t + gd_{t+1} \end{aligned} \tag{8}$$

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

$$\begin{aligned} z^a_{j,t} &= A_{0,j} + A_j \sigma^2_{j,t} \\ z^m_{j,t} &= A^m_{0,j} + A^m_j \sigma^2_{j,t} \end{aligned} \tag{9}$$

As long as A_j and A^m_j are not vanishing, these relations tells us that the variation in valuation ratios can be attributed to fluctuations in economic uncertainty. These relations together with equation 5 imply that measures of economic uncertainty (conditional variance of consumption) are predicted by valuation ratios.

The values of $A_{0,j}, A_j, A^m_{0,j}, A^m_j$ in terms of the parameters of the model are obtained from the Euler condition (6) after the log stochastic discount factors and the returns are all expressed in terms of the factors $\{\sigma^2_{j,t}\}^J_{j=1}$ and of the innovations $\{e^g_{j,t+2^j}\}_j$ and $\{\varepsilon_{j,t+2^j}\}_j$. In Appendix D we show that plugging these expressions for the stochastic discount factor and for the returns into the Euler equation and using the method of undetermined coefficients one obtains a set of equations for the coefficients $A_{0,j}, A_j, A^m_{0,j}, A^m_j$, the solution of which are

[6] All details behind our calculations are given in the Appendix D.

given by the following vectors of sensitivities:

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

where

$$M = diag\left(\rho_1,\dots,\rho_J\right)$$

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

To study the consequences of our model with persistence heterogeneity for the equity premium recall that the risk premium on any asset i satisfy, in this set-up, $E_t[r_{i,t+1}-r_{f,t}]+0.5\sigma^2_{r_{i,t}} = -cov_t(m_{t+1},r_{i,t+1})$. In Appendix we show that the innovations of the stochastic discount factor are given by

$$m_{t,t+1}-E_t[m_{t,t+1}] = -\left(\frac{\theta}{\psi}-\theta+1\right)\sum_{j=1}^{J}\sigma_{j,t}e^g_{j,t+2^j}+(\theta-1)\kappa_1\left(\sum_{j=1}^{J}A_j\varepsilon_{j,t+2^j}\right) \tag{10}$$

while analogous steps yield the following expressions for the return innovations

$$\begin{aligned}r_{a,t+1}-E_t[r_{a,t+1}] &= \sigma_{j,t}\odot e^g_{j,t+1}+\kappa_1\underline{A}\boldsymbol{\varepsilon}_{t+1}\\ r_{m,t+1}-E_t[r_{m,t+1}] &= \varphi_{d,j}\sigma_{j,t}\odot e^d_{j,t+1}+\underbrace{\kappa_{1,m}\underline{A_m}}_{\beta_{m,\varepsilon}}\boldsymbol{\varepsilon}_{t+1}\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\varepsilon}^{\intercal}_{t+1} &\equiv \left[\varepsilon_{1,t+2^1},\dots,\varepsilon_{J,t+2^J}\right]\\ \sigma_{j,t}\odot e^g_{j,t+1} &= \sum_{j=1}^{J}\sigma_{j,t}e^g_{j,t+2^j}\end{aligned}$$

and from which we can compute the conditional variance of $r_{a,t+1}$ and $r_{m,t+1}$:

$$\sigma_{a,t}^2 \equiv Var_t(r_{a,t}) = \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1^2 \underline{A}\mathbf{Q}\underline{A}' \tag{11}$$

$$\sigma_{m,t}^2 \equiv Var_t(r_{m,t+1}) = \sum_{j=1}^{J} \varphi_{d,j}^2 \sigma_{j,t}^2 + \kappa_1^2 \underline{A_m} Q \underline{A_m}' \tag{12}$$

With the innovations to the equilibrium returns at hand and using (10), one finally can compute the risk premia for the consumption claim asset, $r_{a,t+1}$ and for the market portfolio, $r_{m,t+1}$

$$E_t[r_{a,t+1} - r_{f,t}] + 0.5\sigma_{r_{a,t}}^2 = \lambda_g \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1 \underline{\lambda}_\varepsilon \mathbf{Q}\underline{A}' \tag{13}$$

$$E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma_{r_{m,t}}^2 = \lambda_g \varphi_{d,j} \sigma_{j,t}^2 \odot \rho_j + \kappa_{1,m} \underline{\lambda}_\varepsilon \mathbf{Q}\underline{A}'_m \tag{14}$$

where

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

Note that the λ's represent the market price of risk for each source of risk, namely $e_{j,t+1}^g$ and $\boldsymbol{\varepsilon}_{t+1}$. Moreover, the market prices of risks for variances are negative when $\gamma \geq 1/\psi$. This is consistent with the empirical evidence that the variance risk premium is negative. Intuitively, the negative sign of the variance risk premium indicates that investors regard increases in market volatility as unfavorable shocks to the investment opportunity. However, in contrast to the Bansal and Yaron (2004) and Bansal et al. (2007) models, the market variance risk premium is determined by all of the volatilities components. Finally, notice that when $\gamma = 1/\psi$ we obtain the results for power utility. In this case, all the risk premia other than that of the instantaneous consumption growth become zero.

Before testing the restrictions imposed by our pricing model, we comment on another key difference between the standard long-run risk model of Bansal and Yaron (2004) and our with persistence heterogeneity in variance. In BY04 stock prices are driven by persistent changes in the volatility of consumption, which in turn moves the stock market. However the BY04 model asserts a much stronger predictive power of dividend-price ratio on future stock return volatility than is found in the data, as demonstrated by Beeler and Campbell (2009). By interpreting price-dividend ratio and the volatilities of consumption as the sum of different components each one characterized by its own persistence level, our extension of the BY04 model overcomes this problem. The components of financial ratios can forecast components of market and consumption variance without necessarily requiring the aggregate financial ratio to forecast the aggregate market variance. We come back to this point in Section 33.3

3 Empirical evidence

In this section we provide new evidence that relates asset prices to consumption and stock market variance. Along this section we extract consumption volatility and its components using the technique described in Appendix C. We instead measure stock market risk by the realized variance obtained using high-frequency (i.e. daily in our case) return data. Appendix A contains a detailed description of the data.

3.1 Long-run Risk and Return Trade-off

In this Section we discuss the risk-return trade-off implied by our model. Whereas the body of empirical evidence on the risk-return relation is mixed and inconclusive, as discussed in the introduction, here we argue that the disagreement in the empirical literature on the risk-return relation is likely to be attributable to not properly considering the persistence heterogeneity present in the conditional mean and conditional variance of excess stock market returns.

In Appendix D we show that the following relation at a fixed level of persistence j between the component of expected market returns and the component of macroeconomic uncertainty, $\sigma^2_{j,t}$, holds:[7]

$$E_t[r_{m,t+2^j} - r_{f,t+2^j}] = \lambda_g \rho_j \varphi_{d,j} \sigma^2_{j,t} + const \tag{15}$$

Applying the persistence based decomposition described in Appendix B to the realized returns[8] and using the technique described in Appendix C to filter out the unobserved components of consumption variance, we can finally estimate the following set of regressions:

$$r_{m,t+2^j} - r_{f,t+2^j} = \alpha_j + \beta_j \sigma^2_{j,t} + \epsilon_{j,t+2^j} \quad j = 1, \ldots, J$$

[7]If one sums over j the relation (15) and applies the forward decomposition (B.4) to the right-hand side then (13) is retrieved.

[8]Although the theoretical risk-return relation is based on the expected (components of) excess return, following many empirical studies we employ (components of) realized excess returns to proxy for the latent variables.

The results are reported in Table 1. Although all regressions are run with an intercept, we do not report their point estimate since we always find insignificance of the intercept coefficients.[9] Note that the relation is not statistically significant both at short and medium horizons. However the results suggest that the (macroeconomic) risk-return relation holds true for the components at level of persistence $j = 5, 6, 7$ corresponding to horizons of $4-8$, $8-16$ and $16-32$ years, respectively. In Section 3.4 we show that these consumption variance components indeed proxy for real economic uncertainty such as long-run unemployment and/or productivity risk.

Since in our model consumption variance drives market risk, we can obtain a more familiar risk-return relation by first applying the decomposition (B.3) to the equation (12) and looking at the relation at a fixed level of persistence j to obtain

$$\sigma^2_{m,j,t} = \varphi^2_{d,j}\sigma^2_{j,t} + const \tag{16}$$

Substituting (16) into (15) one finally obtains the risk-returns trade-off disaggregated across levels of persistence:[10]

$$E_t[r_{m,t+2^j} - r_{f,t+2^j}] = \underbrace{\beta_j}_{\frac{\lambda_g \rho_j}{\varphi_{d,j}}} \sigma^2_{m,j,t} + const \tag{17}$$

To study of the risk-return relation we disaggregate across different levels of persistence the risk, proxied by realized variance, and returns and then we run the following regressions:

$$r_{j,t+2^j} = \alpha_j + \beta_j \sigma^2_{m,j,t} + \epsilon_{j,t+2^j} \quad j = 1, \dots, J$$

The results are reported in Table 2.[11] Note that the relation is not statistically significant both at very short horizon, i.e. $j = 1$, and at medium horizons $j = 4, 5$. However the results suggest that the risk-return relation holds true for the components at level of persistence $j = 6$ and $j = 7$ corresponding to horizons of $8-16$ and $16-32$ years, respectively. At these horizons the R^2 jumps to 17% and 32%. These components were also found to be significant drivers of expected returns in the case where consumption variance is used in the estimation. This points to an important link between (long-run) market uncertainty and macroeconomic

[9] We did not consider regressions for which the intercept α_j is constrained to be 0 although from a statistical standpoint, provided the restriction is true, the slope estimator is still estimated consistently but with increased precision.

[10] Once again if one sums over j the relation (17) and applies the forward decomposition (B.4) to the right-hand side then one obtains back (14).

[11] Once again we do not report α_j since their point estimate is not statistically significant

one.

From an economic point of view the above results provide evidence that in the long-run market returns are driven by market variance and we can state that at the market level there is compensation for aggregate risk as represented by market variance.

To further support the above statement, we now sum up the components of returns and variance at level of persistence J up to level $j = 6$, i.e. $\sum_{j=6}^{J} r_{j,t+2^j}$ and $\sum_{j=6}^{J} \sigma^2_{m,j,t}$, we obtain the cumulative sum of future returns $r_{t,t+h}$ and the cumulative sum of past market variance $\sigma^2_{t-h,t}$ where $h = 2^{j-1} = 8$ years. We plot in Figure 5 the two series obtained by summing the sixth and seventh components for returns and stock market variance. The comovement is visibly striking. These components of market variance explains 25% of the variation in long-run market returns and up to 42% when we replace the observation in October 1987 with the second highest observation in the sample.

Some remarks are in order. First of all the above evidence supports a strong relation of future long-run excess market returns with past long-run market variance. This is not in contrast with the existing literature which documents a hard-to-detect risk-return trade-off. In fact consistent with this findings our risk-return dependence is statistically mild at short horizons. Moreover these results are in line with those in Bandi and Perron (2008) who find a strongly significant correlation between $r_{t,t+h}$ and $\sigma^2_{t-h,t}$ for values of h between 6 and 10 years. Importantly we are providing a long-run model which support these results. In fact as noted in Bandi and Perron (2008) simple aggregation of short-term risk-return models under a classical (autoregressive) process for variance cannot imply our results. Therefore whereas traditional short-term risk-return models yield counter factual long-run implications our long-run risk model with persistence heterogeneity is free of this unsupported empirical implication. This is due to the fact that we do not impose a single autoregressive process for the aggregate variance observed at the highest frequency of observation but instead we use a multiscale autoregressive process to describe the variance components at each different level of persistence.

Second our methodology used to study the (long-run) risk-return trade-off can be compared to the MIDAS framework of Ghysels et al. (2005), Ghysels et al. (2007) and to the Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009). In fact all these methodology share the common idea to construct regressions combining data with different sampling frequencies. However our approach differs in many respect. In particular MIDAS[12]

[12] MIDAS come in different form, e.g. MIDAS regressions with polynomials (Ghysels, Santa-Clara and Valkanov, 2003a) and MIDAS with stepfunctions (Forsberg and Ghysels, 2004) whose HAR Model is a special case. However the basic ideas stay the same.

exploits high frequency (financial) data to predict low frequency (macro) data whereas our approach uses the components of the series at the same frequency level.

Finally the risk-return relation is involving conditional expected return and conditional expected risk. Previous research takes diverse approaches in measuring the expected return (e.g. Campello, Chen and Zhang (2008) and Pastor, Sinha and Swaminathan (2008)) and conditional variance (e.g., French et al. (1987) and Ghysels et al. (2005)). Importantly Ludvigson and Ng (2007) note that the estimated risk-return relation is likely to be highly dependent on the particular conditioning variables analyzed in any given empirical study.[13] Here we adopt a simple approach and use average realized excess returns as a proxy for expected equity returns[14] and then extract their components using the forward decomposition. We do so because we want to stress the importance of using persistence-consistent measures of returns and variance in investigating the intertemporal risk-return relation. Our results has highlighted that in order to detect a positive risk-return it is important to make sure that the measures of both the expected return and conditional variance of returns have the same level of persistence.[15]

3.2 The risk aversion

In Appendix B we show that the following relation at a fixed level of persistence j between the component of expected market returns on the aggregate consumption claim and conditional consumption variance holds true:

$$E_t[r_{a,t+2^j} - r_{f,t+2^j}] = \lambda_g \sigma^2_{j,t} + \kappa_1 [\underline{\lambda}_\varepsilon \mathbf{Q}]_j A_j \;\; j = 1, \ldots, J$$

This is a constrained set of relations where the parameter γ, that links the information content of the excess return components to the consumption growth variance ones, is common across persistence levels.

In this special case where the parameter to be estimated is equal at all levels of persis-

[13] Ludvigson and Ng (2007) discuss one potential remedy to this problem based on the methodology of dynamic factor analysis for large data sets, whereby a large amount of economic information can be summarized by a few estimated factors.

[14] This practice is justified on grounds that for sufficiently long horizons, the average return will "catch up and match" expected return on equity securities and relies on a belief that information surprises tend to cancel out over the period of the study. Thus ex post average excess equity returns provide for an easy-to-implement, seemingly unbiased estimate of expected equity risk premium.

[15] Alternatively in Merton's theoretical specification it is implicitly embedded that measures of both the expected excess return and conditional variance of returns are based on the same information set. Basically our procedure define "common information set" as "the information set where all the variables have the same level of persistence"

tence, one can solve the overlapping problem by adopting the technique suggested in Fadili and Bullmore (2002). In particular Fadili and Bullmore (2002) suggest to (sub)sample the components at level of persistence j with frequency 2^j[16] in order to get rid of the autocorrelation problem and then to apply to the so obtained sampled time series the generalized least squares estimator (GLS). [17] We report the results obtained using this approach in Table 4. The estimated ICAPM coefficient γ is 4.08 in the full sample, with a significant t-statistic. Most important, the magnitude of γ lines up well with the theory. According to the ICAPM, γ is the coefficient of relative risk aversion of the representative investor and a risk aversion coefficient of 2.606 matches a variety of empirical studies (see Hall (1988), and references therein). The significance of γ is robust in the subsamples, with estimated values of 1.547 and 3.748, and t-statistics always higher than 2. Importantly when the returns and (consumption) risk are disaggregated across levels of persistence, the intercept α is always insignificant. The estimated magnitude and significance of the risk aversion coefficient are remarkable in light of the ambiguity of previous results.

3.3 Volatility and Asset Prices

In this and the next section we show that macroeconomic uncertainty, defined as the common variance factors extracted from consumption growth rates, play an important role in determining asset prices, especially if the perceived macroeconomic uncertainty unravels slowly.

We first plot in Figure 3 the components of the consumption variance along with the ones of (minus) the price-dividend ratio. The filtered aggregated time-varying consumption growth volatility is very persistent with an autocorrelation coefficient of about 0.864. Nevertheless as we have argued in the previous sections, the volatility itself can be interpreted as the sum of different components each one characterized by its own level of persistence j. Figure 3 exhibits a striking correlation between the components of the macroeconomic risk and the stock market at levels of persistence higher than 5, i.e. for cyclical components with periodicity greater than $2^{5-1} = 16$ quarters. This preliminary evidence suggests that by not properly taking into account heterogeneity of persistence in consumption volatility

[16] Note that if we apply our decomposition to a time series with $T = 2^J$ elements we then obtain J components with T elements. If we subsample the components we obtain a new time series with $T/2 + T/4 + ... + T/2^J = T$ elements, that is the new sampled series has the same length of the original one.

[17] More precisely this estimator makes use of the decimated (not-redundant) Haar transform which yields a diagonalized covariance matrix of the regression errors, i.e. the off-diagonal elements can be set to zero. Diagonalization simplifies numerical identification of parameter estimates and implies that the WLS estimator is theoretically approximate to the best linear unbiased (BLU) estimator and can provide maximum likelihood estimates of both signal and noise parameters, namely γ and σ_η.

can obscure the true economic relation between asset prices and macroeconomic uncertainty.

We now move to analyze the underlying sources of risks that are driving asset prices. In particular we empirically test the relations in (9) which tell us that the component of economic uncertainty at level of persistence j should explain the corresponding component of the asset valuation ratios. We thus run the following projections:[18]

$$pd_{j,t} = \beta_0 + \beta_j \sigma^2_{j,t} + \epsilon_{j,t} \quad j = 1, \dots, J$$

where β_j should provide an estimate of A^m_j according to relations (9). Table 7 provides the estimates, t-statistics, and R^2 from the above componentwise regressions where the columns stand for different levels of persistence $j = 1, \dots, J = 7$. We observe that the estimates for $j = 5, 6, 7$ have significant robust t-statistics at the 5% level and the R^2 rises to 50% and 59% at level of persistence of $j = 6$ and $j = 7$ respectively. Hence the three components of consumption variance responsible for explaining the fluctuations in the price-dividend are very persistent with an half-life of $4, 8, 16$ years, respectively. Our results endorse the argument of Lettau, Ludvigson and Wachter (2008) where the authors argue that the increase in asset valuation ratios is not well described as a sudden jump upward, but instead occurs gradually over several years due to (close to) permanent fluctuations in macroeconomic volatility of the order of up to 30 years. Whereas our 7-th component of consumption variance has a similar half-life,[19] our estimation technique uncovers other fluctuations in volatility, namely the ones at level $j = 5$ and $j = 6$ which will turn out to be important both for the risk-return trade-off and for the real risk-free rate variation. Importantly the signs of these relations are, as predicted by our economic model, negative. Our evidence thus shows that a rise in economic uncertainty leads to a fall in asset prices. This is important because it highlights an often discussed but not verified view that aggregate economic uncertainty (i.e., real aggregate consumption volatility) has sizable effects on asset valuations and that financial markets dislike economic uncertainty.

In Fig 4 we plot the sum of the components at level of persistence $j = 5, 6, 7$ of consumption volatility and of (minus) price-dividend ratio. We recall that as shown in Appendix B by summing the component of any series from the highest level of persistence $J = 7$ back to the level j one obtains the original series aggregated over the horizon 2^{j-1}. Thus we can interpret the figure as plotting the standard deviation of consumption growth and the

[18]Using the components of consumption volatility instead of variance does not qualitatively affect any of our empirical results.

[19]In Lettau et al. (2008) the estimated low volatility state reached in the 1990s is expected to last about 125 quarters, over 30 years. Our estimated 7-th component has an half-life of 16 years and therefore the full cycle can last up to 32 years.

dividend-price ratio in 4-year windows. The correlation (of about 70%) between these two series is visibly striking.

Next, we document that in addition to the fact that economic uncertainty helps explaining the price-dividend ratio, current valuation ratios are useful in predicting future realized stock market volatility. This fact is implied by our model. In fact equation (12) tells us that the components of the consumption volatility are important for explaining the conditional market variance. Moreover we also know from equation (9) that the each component of the financial ratios is an affine function of the corresponding consumption volatility component. If we invert relation (9) to express the components of consumption volatility as a function of the components of financial ratio and if we substitute in (12) we obtain:

$$\sigma^2_{m,t} = \sum_{j=1}^{J} \varphi^2_{d,j} \underbrace{\frac{pd_{j,t} - A^m_{0,j}}{A^m_j}}_{\sigma^2_{j,t}} + const \tag{18}$$

where we recall that we define $\sigma^2_{m,t} \equiv Var_t(r_{m,t+1})$. Importantly the conditional market variance is no more predicted by the aggregate price-dividend ratio but instead it is possibly predicted by a combination of the components of the financial ratio. Of course the above relation holds only when $A^m_j \neq 0$. This is equivalent to say that the same components of price-dividends ratio that comove with consumption volatility, namely the one at level of persistence $j = 5, 6, 7$ are the candidates (assuming $\varphi_{d,j} \neq 0$) in predicting future market volatility.

Before bringing these relation to the data we apply the forward decomposition (B.4) to the conditional variance of returns, i.e. $\sigma^2_{m,t} = \sum_{j=1}^{J} \sigma^2_{m,j,t+2^j}$. We then consider the following additional set of projections:[20]

$$\sigma_{m,t+2^j,j} = \beta_0 + \beta_j pd_{j,t} + \epsilon_{j,t+2^j} \quad j = 1, \ldots, J$$

We remark that although these regressions resemble those run by Beeler and Campbell (2009), there the authors focus on the aggregate time series whereas here we study predictability at a specific level of persistence.

Table (8) reports the results. Importantly the only projection coefficients that are sig-

[20]Note that if we sum over levels j the right-hand side and the left-hand side and assuming β_j to be an estimate of $\frac{\varphi^2_{d,j}}{A^m_j}$ we obtain back relation (18). We also consider an alternative specification where each component $\sigma_{r,t+2^j,j}$ at level j is projected onto all the set of components $pd_{j,t}$. However the effect is mainly significant between the components with the same persistence level.

nificantly different from zero are those at levels of persistence $j = 5, 6, 7$, i.e. those horizons for which consumption volatility components comove with scaled prices, in agreement with our model.

Of course the question is now why the results in Table 8 obtained by disaggregating the time series deliver a different predictability pattern from the ones in Table 6 obtained for the aggregate series. We argue that the explanation can be found in the variance decomposition of the different time series across levels of persistence. As reported in Tables 4 and 5 the consumption variance decomposition shows that the great part of the variability in consumption and returns volatility is explained by low frequency (i.e. high j) and high frequency (i.e. low j) components, respectively. By disentangling the small in volatility but highly persistent component of market risk from the the highly volatile low persistent component our filtering procedure is able to uncover the relation between scaled prices and return volatility.

The signs of the coefficients are negative as predicted by our model only for the components $j = 6, 7$. The negative slope coefficients imply that a rise in the persistent components of stock market uncertainty leads to a fall in the corresponding components of asset valuations.

Taken together the results in Tables 7 and 8 lead to the conclusion that the long-run components at levels of persistence $j = 6, 7$ of the current price-dividend ratio embody useful information reflecting the macroeconomic uncertainty which in turn is useful for predicting the future stock market volatility. Overall this evidence suggests that fundamental measure of economic uncertainty, as captured by the persistent components of consumption variance, is priced in the market.

3.4 Identification of macroeconomic uncertainty

In this section we argue that those components of consumption variance that in the previous section we found to influence asset prices are indeed proxying for macroeconomic uncertainty. In order to do so we use a more direct measures of economic uncertainty available using the Survey of Professional Forecasters that do not rely on consumption data or a specific ARIMA model. In particular we follow the lead of Ang and Bekaert (2008) and measure the economic uncertainty by the uncertainty among professional forecasters regarding real GDP growth.[21] Figure 6 displays the sum of the components of our filtered consumption volatility relevant to explain the financial ratio together with the real uncertainty measure vr_t suggested in Ang and Bekaert (2008). Although our measure uses US consumption data to filter out, using a latent bayesian approach, the economic uncertainty process whereas

[21] The Appendix A describes the construction of such a measure.

vr_t is just combining information from a survey about next years real GDP growth, the two measures seem to comove in a consistent way. This evidence support the idea that using our latent bayesian approach together with componentwise regressions of financial ratio on consumption volatility we are indeed able to capture the macroeconomic risk.

A natural question is now what the consumption volatility really stands for. Empirically, this is a latent variables to be estimated, and it is difficult to link them to known macroeconomic risks. We therefore try to give single name to the long-run economic uncertainty by linking it to macroeconomic variables. In particular we follow the lead of Malkhozov and Shamloo (2010) and Benigno, Ricci and Surico (2010). Malkhozov and Shamloo (2010) show that shocks to volatility of productivity create movements in consumption volatility and Benigno et al. (2010) in turn show that, in a model of the labor market with asymmetric real-wage rigidities, movements in the variance of productivity growth influence the trend of unemployment. We therefore investigate whether consumption volatility is reflecting movements in long-run unemployment and/or the variance of productivity. Figures 7 and 8 report the series of our consumption volatility along with the trend in unemployment and the variance of productivity growth for a postwar sample of U.S. data, respectively. The time-series for the U.S. long-run mean of unemployment and the variance of productivity are obtained computing averages and variances over five-year rolling windows, where the window length has been chosen in such a way to match the persistence of our consumption volatility series. The correlation between long-run unemployment and consumption volatility is $\rho = 0.62$ and the correlation between long-run productivity variance and consumption volatility $\rho = 0.24$ (which increases to 0.32 when we exclude 2008Q1-2009Q4). Our long-run measure of consumption volatility reflects important low frequencies movements in the series: the fall in long-run unemployment towards the end of the 1980s/early 1990s and its rise during the first half of the 2000s. It is also apparent the Great Moderation in the variance of productivity growth which coincides with a sharp fall in the unemployment trend and a decrease in consumption volatility. Thus these Figures confirm the very interesting feature of the data that there is a strong positive association between long-run unemployment, the variance of productivity growth and the consumption volatility.

In this section we have argued that the components of consumption volatility that are reflected in the financial ratios do indeed proxy for macroeconomic uncertainty. Moreover we suggest the variance of productivity growth and the long-run mean of U.S. unemployment as significant determinants of this uncertainty. These results suggest that, since as we note in Section 3, financial market dislikes economic uncertainty and such uncertainty comoves positively with the long-run trend in unemployment, a shift to higher mean unemployment coincided with a fall in the level of share prices.

In this Section we study what are the implications of our long-run risk model with persistence heterogeneity in volatility for the relation between macroeconomic uncertainty and the risk-free rate. In particular in Appendix D.3 we derive the following expression for the risk-free rate

$$r_{f,t+1} = \beta_0 + \lambda_\eta^2 + \frac{(1-\theta)}{\theta}(\lambda_\eta)\sum_{j=1}^{J}\sigma_{j,t}^2 \tag{19}$$

The above relation entails that real interest rates reflect volatility. This is in line with the empirical finance literature (see Glosten et al. (1993), among others) that reports significant links between equity market volatility and short-term interest rate, the hypothesis behind this relation being that the short rate reflects inflation uncertainty, which in turn is likely to be correlated with aggregate economic uncertainty.

In order to bring the above relation (19) to the data, we first apply our decomposition (B.3) to the right-hand side and then we run a set of regressions of the components of the real short-term rate onto the components of consumption volatility. Importantly, as suggested by our model, we restrict the coefficients loading on the volatility components to be the same across levels of persistence. The results are reported in Table 9 and suggest that the risk-free rate compensates for fluctuations in risk. In the following we are going to argue that our results are mainly driven by the comovement between the very same components of consumption volatility that are reflected in asset prices, i.e. those at levels of persistence the $j = 5, \ldots, 7$ and the corresponding ones in the risk-free rate.

To start with we plot in Figure 9 the (aggregate) real risk-free rate and (minus) the approximation of consumption volatility at level of persistence $j = 5$, i.e. the $h = 2^{5-1} = 16$ quarters fluctuations in consumption volatility. The correlation between these two series is 0.23. This figure already highlights important comovements in postwar U.S. data between the (aggregate) interest rates and those components of consumption volatility that, in Section 3.4, we argued to proxy for macroeconomic risk.

However we recall that our set of (constrained) regressions consider the component-by-component relation between risk free rate and consumption volatility and thus suggest to look not at the aggregate series but instead at the risk-free rate components that correspond to the same levels of persistence as the ones that proxy for the macroeconomic risk, i.e. $j = 5, 6, 7$. With this regard our approach is similar to, and actually complements, the one of Atkeson and Kehoe (2008) where the authors decompose the observed postwar U.S. history of nominal interest rates into a secular and a business cycle component. These components of the short rate are intended to capture, respectively, the random walk movements in the Feds inflation target and (response to) exogenous changes in real risk. Their model in particular suggest that the business cycle component of the interest rates should move one for one with risk. Instead our approach says that the components with persistence of 4 years of the interest rate should reflect, not the aggregate, but instead the corresponding component of macroeconomic risk. We therefore undertake the following exercise. We plot in Figure 10 the business cycle component of the risk-free rate suggested by Atkeson and Kehoe (2008) together with approximation of (minus) consumption volatility at level of persistence $j = 5$. We use their approximation of the risk-free rate business cycle component (extracted using principal components analysis) because by doing so we are able to test both the fact that the comovements between the two series should become more apparent once they are filtered at the same time scale and also the robustness of our technique to alternative filtering choice. Indeed we observe that the correlation now rises to 0.40 which bring further support the thesis that over our persistency horizon 4 years, great part of the movements in the short rate come from movements in conditional variances.

The above results are particularly important for two reasons. First they potentially explain the findings of Canzoneri, Cumby and Diba (2007) who document that in those models imposing that the conditional variances of the variables that enter the Euler equation are constant, the Euler equation itself does a poor job of capturing the link between the short rate and the economy at business cycle frequencies. Since we just document that all of the movements in the short rate come from movements in conditional variances and not from conditional means, the failure of the Euler equation is not surprising. Second they suggest to view the central bank's policy changes, namely the short rate, as primarily intended to compensate for exogenous business cycle fluctuations in risk. Clearly as noted also in Atkeson and Kehoe (2008), this view differs substantially from the standard view, often summarized by a Taylor rule, where risk plays no role and, instead, the Feds policy is a function of its forecasts of economic variables such as expected real growth and expected inflation.

4 Conclusion

In this paper, we demonstrated that it is important to account for persistence heterogeneity in volatility to explain the different empirical pattern across investment horizons of the risk-return trade off, i.e. market risk appears in the long-run to be a major contributing factor with a much smaller role when adopting a short-run perspective.

To reconcile these seemingly contradictory findings we design and solve a consumption-based asset pricing model where at each time scale the components of consumption variance are described by a different autoregressive structure. Our disaggregated economic models delivers the interesting empirical finding of a positive and significant long-run dependence between expected excess market returns and past market variance without necessarily implying a positive short-horizon risk-return. Moreover our estimates for the risk aversion is around four.

We also obtain new results about the link between asset prices and macroeconomic uncertainty. We show that long horizons components of economic uncertainty, as measured by the persistent components of consumption variance, sharply explain valuation ratios. In particular asset valuations drop as economic uncertainty rises that is, financial markets dislike economic uncertainty. Finally we show that long-run macroeconomic volatility exerts a significant effect on the short-term interest rate. We therefore conclude that once the channels associated with fluctuating economic uncertainty and economic growth are disaggregated across levels of persistence, they become important for a reasonable interpretation of asset markets.

The paper offers several possible directions for further research. For example we believe that it would be interesting and extremely informative to apply the theoretical environment of the model to the cross section of returns to explain the findings of Parker and Julliard (2005) and Bandi, Garcia, Lioui and Perron (2010) that the contemporaneous consumption or market risk explains little of the variation in average returns across the 25 Fama-French portfolios, but that a measure of consumption or market risk at a horizon of three to five years explains a large fraction of this variation.

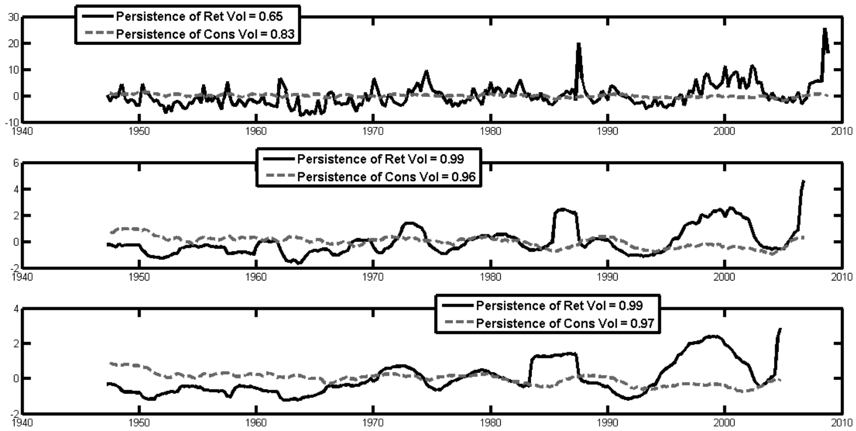


Figure 1: The Figure shows the stock market return realized volatility over a 1 and 8 quarters horizon. The sample spans the period 1947Q2-2009Q4.

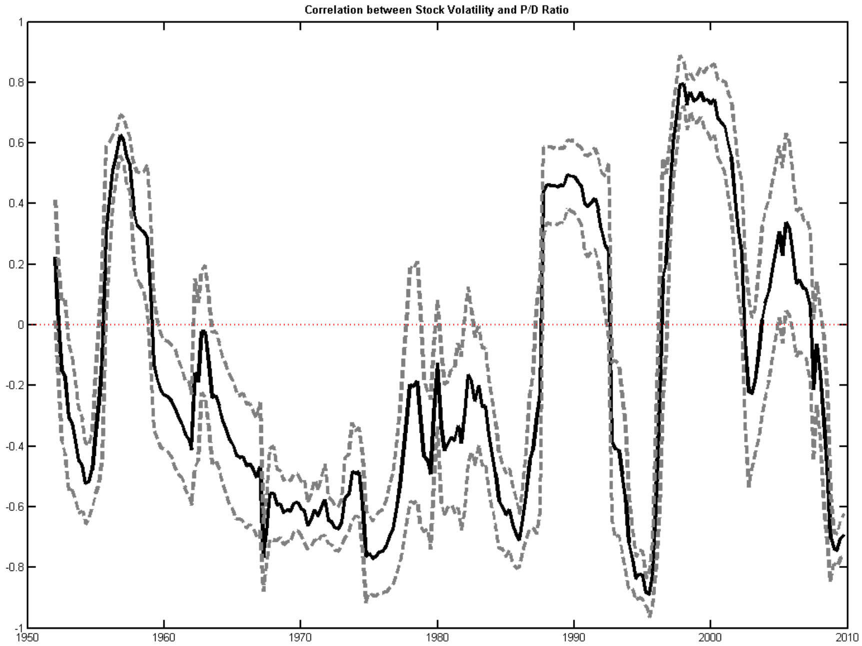


Figure 2: The Figure shows the quarterly time series of the correlation between the S&P 500 P/D ratio measured at the end of the quarter, and realized stock market volatility in that quarter constructed from daily returns. The correlation is based on a rolling sample of 5 years. Dashed lines are the 95% percent confidence bands. The sample spans the period 1947Q2-2009Q4.

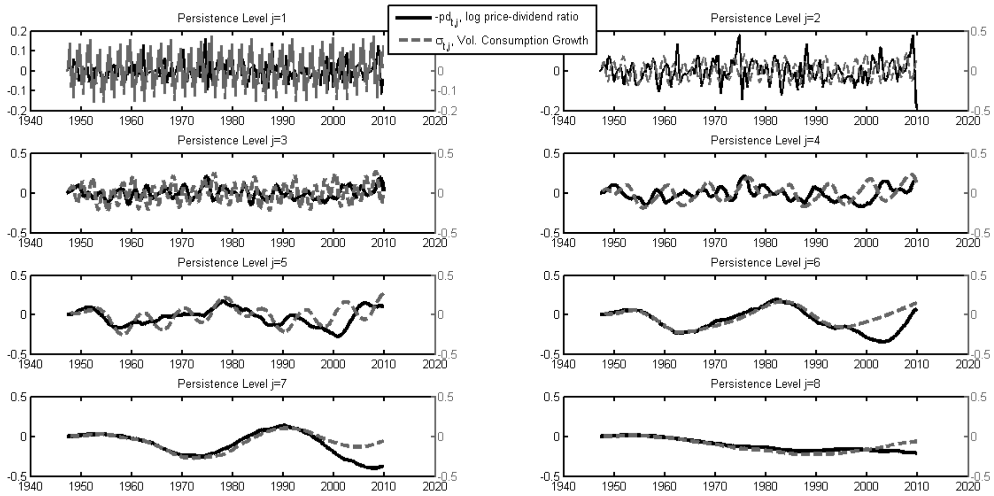


Figure 3: Time-scale decomposition for (minus) the log price-dividend pd_t and log consumption growth volatility σ_t based upon quarterly data. The sample spans the period 1947Q2-2009Q4.

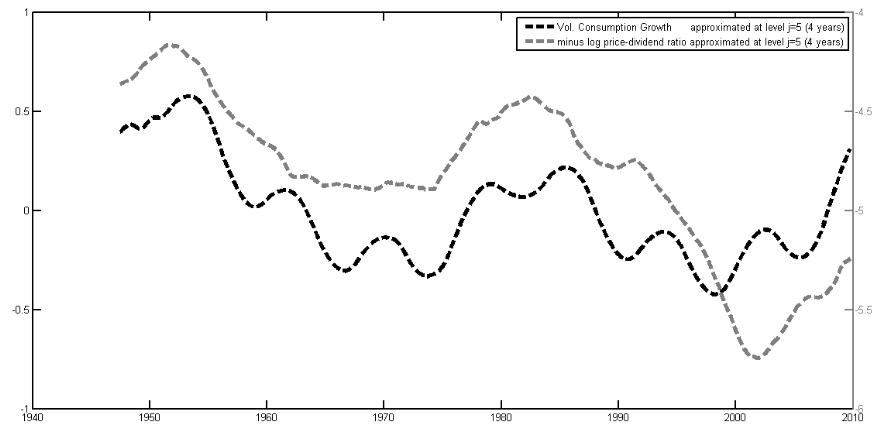


Figure 4: Approximation at level of persistence $j = 5$ of (minus) the log price-dividend pd_t and of log consumption growth volatility σ_t based upon quarterly data. Both series are demeaned and divided by their standard deviation. The sample spans the period 1947Q2-2009Q4. The $\bar{R}^2$ from a regression of (minus) the price-dividend onto the log consumption growth volatility is $\bar{R}^2 = 0.483$. The correlation is $\text{corr}(-\sum_{j=5}^{J} pd_{j,t}, \sum_{j=5}^{J} \sigma_{j,t}) = 0.696$.

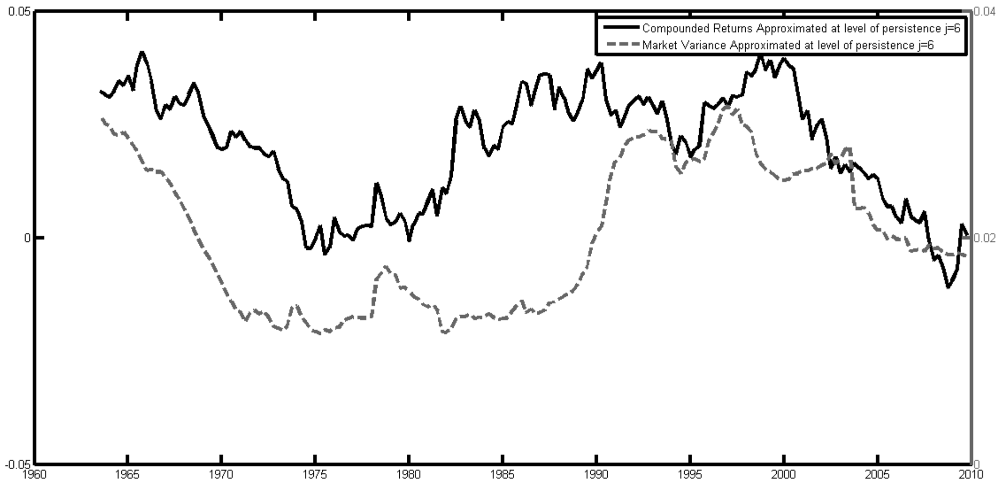


Figure 5: The Figure shows the future compounded log returns and the realized stock market volatility (constructed from daily returns). Both time series are approximated at level of persistence $j = 6$. Therefore we have 8-year ($2^{(}6-1$) quarters) return. The timing is as follows: if you invested one dollar on a given date, it tells you how much total return you would have made over the following eight years. The sample spans the period 1947Q2-2009Q4. The coefficient of determination from a regression of the future returns on the realized stock market variance is $R^2(\%) = 0.25$ and the corrected t-statistics of the coefficient loading on the risk measure is 3.41.

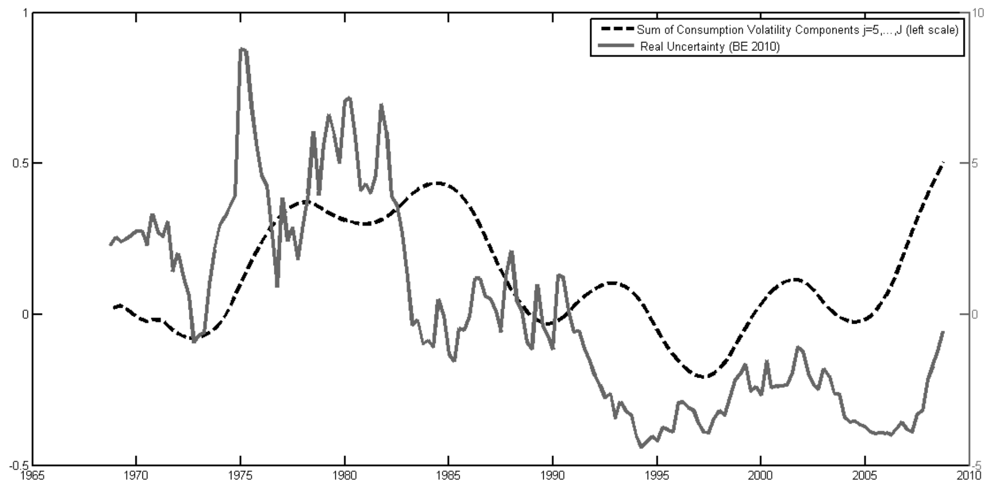


Figure 6: Approximation at level of persistence $j = 5$ of log consumption growth volatility σ_t and a measure of real macroeconomic uncertainty, vr_t, based upon the Survey of Professional Forecasters (SPF). The availability of the SPF data determines the starting point for our sample; the sample spans the fourth quarter of 1968 through the end of 2009. The correlation is $\operatorname{corr}(vr_{j,t}, \sum_{j=5}^{J} \sigma_{j,t}) = 0.404$.

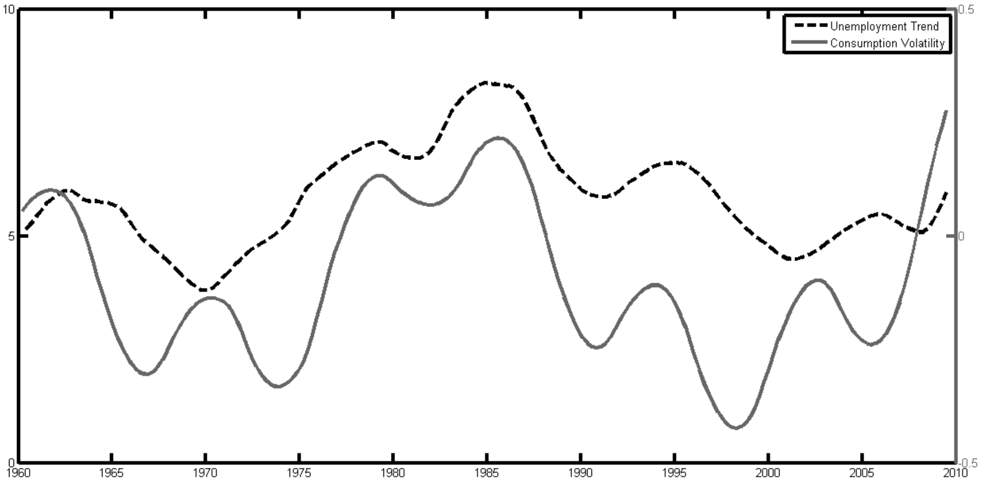


Figure 7: Approximation at level of persistence $j = 5$ of log consumption growth volatility σ_t and long-run unemployment computed using five-year rolling windows. Both series are demeaned and divided by their standard deviation.

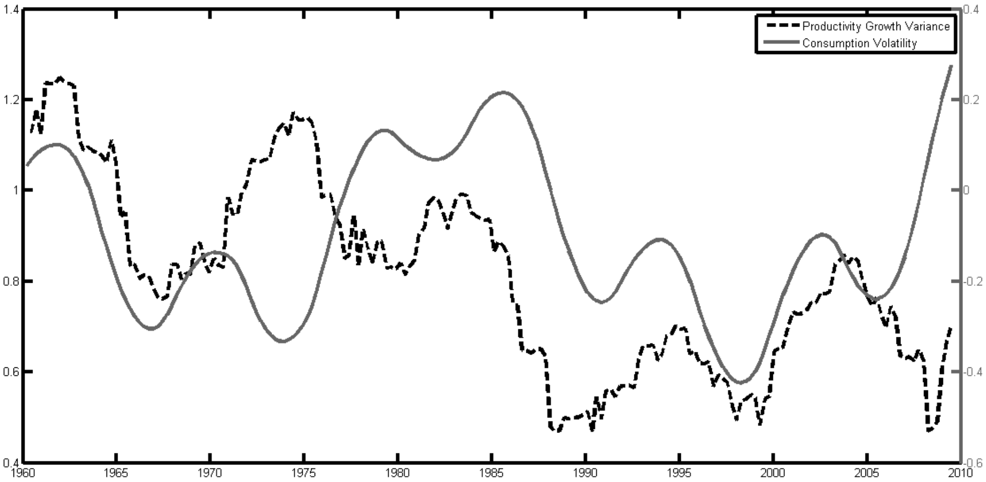


Figure 8: Approximation at level of persistence $j = 5$ of log consumption growth volatility σ_t and long-run variance of productivity growth computed using five-year rolling windows. Both series are demeaned and divided by their standard deviation.

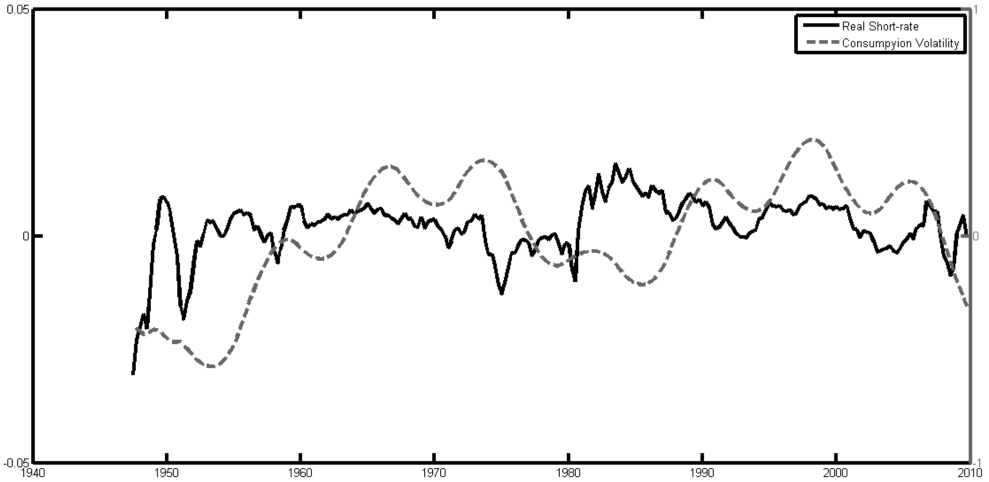


Figure 9: Short-rate (left scale) and consumption volatility (right scale) approximated to scale 5. The short-rate is the 3-month T-bill rate. The sample spans the period 1947Q2-2009Q4.

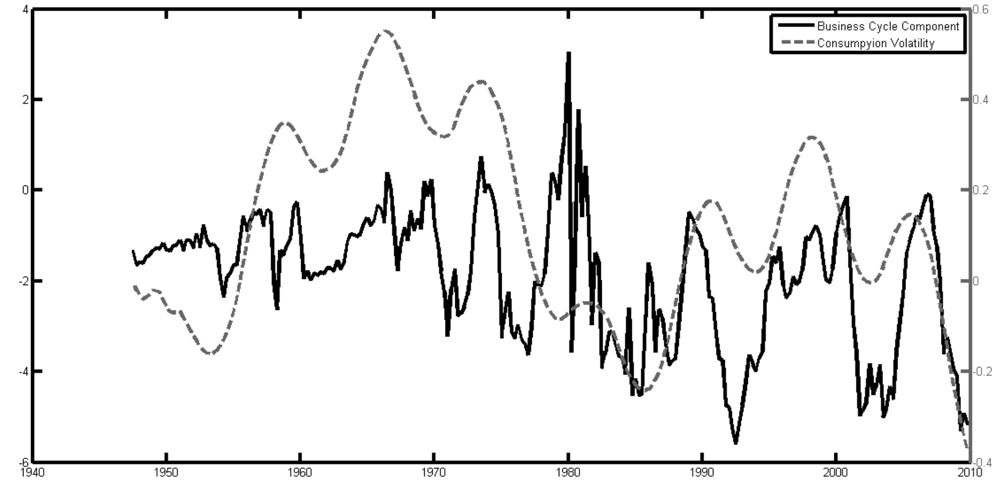


Figure 10: Business cycle component of the short-rate (left scale) and consumption volatility (right scale) approximated to scale 5. Both series are demeaned and divided by their standard deviation. The sample spans the period 1947Q2-2009Q4.

$z_t =$	Level of persistence j 1	2	3	4	5	6	7
	-7.60	-4.58	1.38	-0.82	3.32	4.92	2.45
${\sigma^2}_{j,t}$	(-2.44)	(-1.98)	(0.55)	(-0.27)	(2.95)	(4.70)	(2.20)
	[0.02]	[0.02]	[0.00]	[0.00]	[0.10]	[0.29]	[0.18]

Table 1: This table reports the results of componentwise predictive regressions of the components of stock market returns on the components of consumption variance $\sigma^2_{j,t}$. 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 sample is quarterly and spans the period 1947Q2-2009Q4.

$z_t =$	Level of persistence j 1	2	3	4	5	6	7
	7.05	11.74	12.49	1.76	-1.42	18.47	21.14
${\sigma^2}_{m,j,t}$	(1.57)	(5.57)	(4.41)	(0.35)	(-0.38)	(2.77)	(2.20)
	[0.02]	[0.13]	[0.11]	[0.00]	[0.00]	[0.17]	[0.32]

Table 2: This table reports the results of componentwise predictive regressions of the components of stock market returns on the components of stock market variance $\sigma^2_{m,j,t}$. 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 sample is quarterly and spans the period 1947Q2-2009Q4.

Sample	$r_{m,j,t+1} = \alpha + \gamma\sigma^2_{j,t}$ α	γ
1947Q2-2009Q4	0.03	4.08
	(1.44)	(2.10)
1977Q1-2009Q4	0.03	4.94
	(1.02)	(2.02)
1947Q2-1979Q4	-0.05	8.57
	(-1.55)	(2.27)

Table 3: This table displays the risk aversion estimates based on the persistence heterogeneity tests of the risk-return trade-off. The estimate are based on not-redundant Haar decomposition as suggested in Fadili and Bullmore (2002) based on 256 data points for the first column and 128 data points for the second and third columns.

	Component at level of persistence j						
j=	1	2	3	4	5	6	7
$\frac{Var(\sigma^2_{j,t})}{Var(\sum \sigma^2_{j,t})}$	0.108	0.121	0.141	0.086	0.123	0.233	0.188

Table 4: Contribution to total unconditional variance of the different details components $\sigma^2_{t,j}$ of the variance of log consumption growth. Note that $Var(\sum \sigma^2_{j,t}) = Var(\sigma^2_t)$

.

	Component at level of persistence j						
j=	1	2	3	4	5	6	7
$\frac{Var(\sigma^2_{m,j,t})}{Var(\sum \sigma^2_{m,j,t})}$	0.247	0.251	0.168	0.140	0.087	0.058	0.049

Table 5: Contribution to total unconditional variance of the different details components $\sigma^2_{m,j,t}$ of the variance of log market returns. Note that $Var(\sum \sigma^2_{m,j,t}) = Var(\sigma^2_t)$

.

Panel A: Consumption Volatility (1947.2-2009.4)

	Horizon h (in quarters)		
$z_t =$	4Q	12Q	20Q
pd_t	-0.620	-0.620	-0.551
	(-4.108)	(-4.766)	(-5.205)
	[0.215]	[0.408]	[0.457]

Panel B: Excess Return Volatility (1947.2-2009.4)

	Horizon h (in quarters)		
$z_t =$	4Q	12Q	20Q
pd_t	0.083	0.164	0.160
	(0.501)	(1.100)	(1.647)
	[0.000]	[0.035]	[0.055]

Panel C: Excess Return Volatility Estimated from Monthly Data (1947.2-2009.4)

	Horizon h (in quarters)		
$z_t =$	4Q	12Q	20Q
pd_t	0.047	0.122	0.180
	(1.011)	(1.384)	(1.704)
	[0.010]	[0.046]	[0.072]

Table 6: Panel A reports the results of h-period predictive regressions of the log of the sum of absolute residuals from an AR(1) model of consumption growth on the log price-dividend ratio. Panel B reports the results of h-period predictive regressions of the log of the sum of absolute residuals from an AR(1) model of excess returns on the log price-dividend ratio. Panel C reports the results of h-period predictive regressions of the excess return volatility (estimated from monthly data) on the log price-dividend ratio. For each regression, the table reports OLS estimates of the regressors, corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is quarterly and spans the period 1947Q2-2009Q4. All standard errors are Newey-West with lags equal to 2*(horizon-1). The monthly excess log return for this panel is calculated using CRSP data for stock returns and one month T-bills.

$z_t =$	Scale j						
	1	2	3	4	5	6	7
	0.06	0.06	0.04	-0.12	-0.41	-0.89	-1.03
pd_t	(2.14)	(1.49)	(0.80)	(-0.92)	(-2.85)	(-5.55)	(-5.52)
	[0.02]	[0.02]	[0.01]	[0.02]	[0.19]	[0.50]	[0.59]

Table 7: This table reports the results of regressions of the components of (log) price-dividend ratio $pd_{j,t}$ on the components of consumption growth volatility $\sigma_{c,j,t}$. 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 sample is quarterly and spans the period 1947Q2-2009Q4.

$z_t =$	Level of persistence j						
	1	2	3	4	5	6	7
	-0.01	-0.01	-0.02	0.01	0.01	-0.02	-0.02
pd_t	(-0.68)	(-0.76)	(-1.62)	(1.14)	(-2.43)	(-4.15)	(-3.88)
	[-0.00]	[-0.00]	[0.02]	[0.02]	[0.05]	[0.27]	[0.51]

Table 8: This table reports the results of predictive regressions of the components of stock market returns volatility $\sigma_{m,j,t}$ 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 sample is quarterly and spans the period 1947Q2-2009Q4.

	$r_{f,j,t+1} = \beta_0 + \beta_1 \sigma^2_{j,t} \quad j = 1, \ldots, J$
	-3.67
β_1	(-2.00)
	[0.08]

Table 9: This table reports the result of componentwise regression of the real short-rate (in percentage) on the components of consumption volatility $\sigma^2_{j,t}$, where the loading coefficient β_1 is restricted to be the same across all levels of persistence. The table reports OLS estimates of the regressors, t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample is quarterly and spans the period 1947Q2-2009Q4.

A Data

The data are constructed as follows. The variable to be predicted, i.e. the quarterly excess return, is computed using the continuously compounded daily return (including dividends) during the t-th month for the NYSE/AMEX index. These data are provided by CRSP/WRDS and cover the period from January 1947 to December 2009. The risk-free rate comes from the CRSP Fama Risk Free Rates data file. This data set covers the period from January 1947 to December 2008, and is based on the prices of one-month T-bills.

In Panel A and B of Table 6 we use a measure of realized volatility suggested by Bansal, Khatchatrian and Yaron (2005). We begin by fitting an AR(1) process for each variable y_t that we are interested in:

$$y_{t+1} = b_0 + b_1 y_t + u_{t+1}$$

Then we calculate K-period realized volatility as the sum of the absolute values of the residuals over K periods:

$$Vol_{t,t+K-1} = \sum_{k=0}^{K-1} |u_{t+k}|$$

Finally, we regress the log of K-period realized volatility onto the log price-dividend ratio:

$$\log|Vol_{t+1,t+K}| = \alpha + \beta pd_t + w_{t+K}$$

Since stock returns are measured more frequently than consumption and dividends, a better measure of volatility can be obtained by starting with daily data. In particular to obtain a measure of volatility for the excess return on the CRSP-VW index, we use the time-series variation of daily returns:

$$\sigma_{m,t} = \sqrt{\sum_{k \in t} r_{m,k}^2}$$

where $\sigma_{m,t}$ is the sample volatility of the market return in period t, $r_{m,k}$ is the daily CRSP-VW return minus the implied daily yield on the 3 month Treasury bill rate, k represents a day, and t is a quarter. In Panel C of Table 6 we follow this procedure.

The robustness measure of real growth uncertainty used in Section 3.4 is constructed as follows. As suggested in Ang and Bekaert (2008) we use the two imperfect SPF measures of uncertainty about future real growth to generate a real uncertainty index. First, respondents are asked to report their subjective assessment of the probability of negative real GDP growth over the next quarter. Assuming a binomial distribution for real GDP growth (+1.0 percent growth in expansion, -0.5 percent growth in contractions), we calculate the implied standard deviation of real growth for each respondent and then take the cross-sectional average in each quarter. This measure is denoted sd_t. The second measure, denoted $disp_t$, is the dispersion in respondents expectation for real GDP growth over the next four quarters; specifically, the difference between the 90th percentile and the 10th percentile of all responses. We aggregate these two measures by assuming they are noisy indicators of the true uncertainty vr_t. In the paper we thus use standard Kalman filter methods to extract the conditional (not smoothed) filtered estimates for vr_t.

B The Persistence Decomposition of a Time Series

Among the many ways to isolate the components with different levels of persistence in a time series, two common alternative approaches are the Fourier representation and the Beveridge-Nelson decomposition (see Beveridge and Nelson (1981), Canova (1998) and Christiano and Fitzgerald (2003)). The approach we propose here is, on one side, close to the Fourier representation in the sense that it permits to capture not only the transitory and permanent parts of our time series but also those fluctuations with less extreme persistence. Since it works in the time domain, on the other side, our suggested procedure maintains the intuitive appeal of the Beveridge-Nelson decomposition.

Given a time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ consider its sample mean over a window of past observations with size 2^j:

$$\pi_t^{(j)} = \frac{1}{2^j}\sum_{p=0}^{2^j-1} x_{t-p} \tag{B.1}$$

where $j \geq 1$ and $\pi_t^{(0)} \equiv x_t$. The component $x_{j,t}$ that identifies the fluctuations of the time series $\mathbf{x}$ with periodicity between 2^{j-1} and 2^j periods is then filtered out as follows:

$$x_{j,t} = \pi_t^{(j)} - \pi_{t-2^{j-1}}^{(j-1)} \tag{B.2}$$

In the sequel of the paper we refer to j as the level of persistence of the component $x_{j,t}$. Since in the empirical work one always deals with time series of finite length T, we also define the maximum (observable) level of persistence J as the the greatest integer such that $2^J \leq T$. Clearly given the data sample it is impossible to draw any inference about the persistence of shocks that last longer than 2^J periods.

Simple algebra shows that, for any given J, the generic element x_t, the components

$\{x_{j,t}\}_{j=1}^{J}$ and the permanent part $\pi_t^{(J)}$ are related via the following identity:

$$x_t = \sum_{j=1}^{J} x_{j,t} + \pi_t^{(J)} \tag{B.3}$$

This relation fits well with the intuition that current financial and economic quantities are the result of the overlay of past fluctuations with different periodicity, where these fluctuations can go from the extreme of an incoming flow of information at high frequency (low j) to the one of slowly moving structural changes (high j) like those induced by technological innovation or demographic trends.

The components $\{x_{j,t}\}_{j=1}^{J}$ can also be used to obtain an alternative decomposition of the original time series, i.e.

$$x_{t+1} = \sum_{j=1}^{J} x_{j,t+2^j} + \pi_{t+2^J}^{(J)} \tag{B.4}$$

Intuitively this decomposition, which in the sequel will be referred as the forward decomposition, can be interpreted as a way to reconstruct the realization of x_{t+1} from the effect that this realization will have at different horizons and it will turn out to be a fundamental tool in working out our model.

C Estimation of Consumption Volatility Components

We use the Bayesian estimator suggested by Jensen (2004) based on the Markov chain Monte Carlo sampler and the wavelet representation of the log-squared consumption to draw values of the latent volatilities from their joint posterior distribution.

We assume that the sum of consumption growth stochastic volatility components leads to an aggregate volatility with long-memory. Then let the long-memory stochastic volatility model (LMSV henceforth) be defined as:

$$\begin{aligned} y_t &= \sigma \exp\{h_t/2\}\xi_t \\ (1-L)^d h_t &= \mu + \sigma_\eta + \eta_t \end{aligned} \tag{C.1}$$

for $t = 1, \ldots, T$, and where $y_t = \Delta c_t$, h_t is the log volatility at time t that follows a stationary fractionally integrated process with the long-memory parameter, $|d| \leq \frac{1}{2}$, and L is the lag operator. The innovation terms ξ_t and η_t are standard normal white noise processes that are uncorrelated. The parameter $\sigma_e ta$ is the standard deviation of the log-volatility and σ is the modal instantaneous volatility which we assume without any loss of generality to be equal to one. Taking the logarithm of the squared returns, the LMSV model in Eq. (C.1) is transformed into the linear model

$$\log y_t^2 = h_t + \log_{\xi_t^2} \tag{C.2}$$

Following Kim et al. (1998) we use the offset mixture time series model

$$y_t^* = h_t + z_t \tag{C.3}$$

where $y_t^* = \log(y_t^2 + c)$ and the distribution of z_t is a mixture of normal densities with individual probabilities, means and variances chosen to approximate the complicated distribution of Eq. (C.2) innovations, $\log_{\xi_t^2}$.

The Bayesian approach to estimating the unknown parameters of the LMSV model's offset mixture representation is to augment the unknown parameters with the latent volatilities, h, and the mixture conditioning states, s. Thus under the Bayesian methodology we desire to make draws from the intractable posterior distribution $\pi(s, h, d, \sigma_\eta, \mu | y^*)$.

We project the LMSV model's offset mixture representation in Eq. (C.3) into the wavelet's time-scale space and obtain the linear relationship:

$$W_{j,k}^{y^*} = W_{j,k}^h + W_{j,k}^z \tag{C.4}$$

for $j = 1, \ldots, J$ and $k = 0, 1, \ldots, \frac{T}{2^j} - 1$. Like z_t in Eq. eq:mixturerepresentation, we approximate the wavelet transform of the $\log \chi^2_{(1)}$ random variable with wavelet coefficients, $W_{j,k}^z$ that are distributed as a mixture of normals but with a different number of orders, individual probabilities, means, and variances from z_t. It follows that the pdf of $W_{j,k}^z$ equals

$$f(W_{j,k}^z) = \sum_{i=1}^{k_w} q_i f_n(W_{j,k}^z | \mu_i, \rho_i^2)$$

where k_w is the number of mixture components, and for $i = 1, \ldots, k_w$, $q_i = P(s_{j,k=i})$, is the probability of state i, μ_i is the mean of state i, and ρ_i^2 is the variance of state i. Jensen (2004) discusses the selection of the mixture distributions order and moments.

Jensen (2000) shows that the wavelet basis is a 'near' diagonalizing operator in which the wavelet coefficients of a long-memory process will be an approximately independent multivariate Normal stochastic process with:

$$W_{j,k}^h \equiv \mathcal{N}(0, \sigma_d^2 2^{2d_j})$$

where σ^d is a function of d and $\sigma_e ta$. By projecting y^* into the time-scale space of the wavelet basis, our interest in drawing realization from $\pi(s, h, d, \sigma_\eta, \mu | y^*)$ has now turned into designing a MCMC simulator that makes draws from the posterior distribution $\pi(s, W^{(h)}, \beta | W^{(y^*)})$, where $\beta = (sigma_d^2, d)$, and $s = \{s_{1,1}, s_{1,2}, \ldots, s_{1,T/2}, s_{2,1}, \ldots, s_{J-1,1}, s_{J-1,2}\}$. Then cycling over the steps:

1. sample β from $\beta | W^{(y^*)}, W^{(h)}, s$
2. sample $W^{(h)}$ from $W^{(h)} | W^{(y^*)}, \beta, s$
3. sample s from $s | W^{(y^*)}, W^{(h)}, \beta$

In Step 2 the wavelet transform of h affords us an easy and efficient way of sampling $W^{(h)} | W^{(y^*)}, \beta, s$ from

$$\pi(W_{j,k}^{(h)} | W^{(y^*)}, \beta, s) \propto \pi(W_{j,k}^{(h)}) \pi(W^{(y^*)} | W^{(h)}, \beta, s)$$

where $\pi(W^{(y^*)} | W^{(h)}, \beta, s)$ is the likelihood function. The details of this distribution and its first and second moments can be found in Jensen (2004). For now it should be clear that by modeling W(h) as an independent series we sidestep the difficult task of having to sample from $h | y^*, d, \sigma_e ta, \mu, s$.

D The Valuation Approach: the details of the derivation

In this Section we show the steps to obtain the values of the financial ratios coefficients $A_{0,j}, A_j, A_{0,j}^m, A_j^m$ in terms of the parameters of the model. We then compute the equity premia on both the consumption claim asset and the market return. Finally we derive the risk-free rate.

D.1 The Financial Ratios

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

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

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

which is derived from (6) for the special case where the asset being priced is the aggregate consumption claim, i.e. $r_{i,t+1} = r_{a,t+1}$. We then express the log consumption growth g_{t+1} and the return $r_{a,t+1}$ in terms of the factors $\{x_{j,t}\}_j$ and of the innovations $\{e^g_{j,t+2^j}\}_j$ and $\{\varepsilon_{j,t+2^j}\}_j$. To do so we plug first the Campbell and Shiller (1988) approximation for log returns, see equation (8), into the above expression to obtain:

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

By the backward decomposition (B.3) applied to the (demeaned) price-consumption ratio at time t and by the forward decomposition (B.4) applied to the (demeaned) consumption growth and price-consumption processes at time $t+1$ we have:

$$z^a_t = \sum_{j=1}^{J} z^a_{j,t} \tag{D.1}$$

$$z^a_{t+1} = \sum_{j=1}^{J} z^a_{j,t+2^j} \tag{D.2}$$

$$g_{t+1} = \sum_{j=1}^{J} g_{j,t+2^j} \tag{D.3}$$

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

$$\sigma^2_t = Var_t(g_{t+1}) = Var_t\left(\sum_{j=1}^{J} g_{j,t+2^j}\right) = Var_t\left(\sum_{j=1}^{J}\sigma_{j,t}\, e^g_{j,t+2^j}\right) = \sum_{j=1}^{J}\sigma^2_{j,t}$$

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

Plugging the above expressions into the Euler condition yields:

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

Finally using the dynamics for the components of log consumption growth given in equation (3) together with our guess for the components of price-consumption ratio solution given in equation (9), rearranging terms and using the log normal properties of the shocks we obtain:

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

$$= E_t\left[\exp\quad\left(\theta\log\beta+\theta\left(1-\frac{1}{\psi}\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)+\right.\right.$$

$$\left.\left.\theta\left(\kappa_0+\kappa_1\left(\sum_{j=1}^{J}A_{0,j}+\sum_{j=1}^{J}A_j\sigma^2_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}A_{0,j}+\sum_{j=1}^{J}A_j\sigma^2_{j,t}\right)\quad\right)\quad\right]$$

$$= E_t\left[\exp\quad\left(\quad\theta(\log\beta+\kappa_0+(\kappa_1-1)\sum_{j=1}^{J}A_{0,j})+\ldots\right.\right.$$

$$\left.\left.\theta\left(1-\frac{1}{\psi}\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)+\theta\left(\kappa_1\sum_{j=1}^{J}A_j\sigma^2_{j,t+2^j}-\sum_{j=1}^{J}A_j\sigma^2_{j,t}\right)\quad\right)\quad\right]$$

$$= E_t\left[\exp\quad\left(\quad\theta(\log\beta+\kappa_0+(\kappa_1-1)\sum_{j=1}^{J}A_{0,j})+\ldots\right.\right.$$

$$\left.\left.\theta\left(1-\frac{1}{\psi}\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)+\theta\left(\kappa_1\sum_{j=1}^{J}A_j\underbrace{(e_jM\tilde{\Sigma}_t+e_j\varepsilon_{t+2^J})}_{\sigma^2_{j,t+2^j}}-\sum_{j=1}^{J}A_j\sigma^2_{j,t}\right)\quad\right)\quad\right]=1$$

where we defined $\tilde{\Sigma}_t\equiv[\sigma^2_{1,t},\ldots,\sigma^2_{J,t}]^\intercal$. Collecting terms in $\tilde{\Sigma}_t$ yields eventually a system of equations

$$e_j\left(0.5\left(\theta-\frac{\theta}{\psi}\right)^2+\theta A_j(\kappa_1M-\mathbb{I}_J)\right)=0$$

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

$$\underline{A}\quad\equiv\quad[A_1,\ldots,A_J]^\intercal$$

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

$$\underline{A}\quad=\quad0.5\frac{\left(\theta-\frac{\theta}{\psi}\right)^2}{\theta}\left(\mathbb{I}_J-\kappa_1M\right)^{-1}\underline{1}$$

To derive the expression for A^m_j we exploit once again the Euler condition

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

where now the asset being priced is the market return $r_{m,t+1}$. Following the same steps as above, and additionally using the Campbell and Shiller (1988) log-linear approximation for $r_{m,t+1}$, see equation (8) we can then rewrite the Euler equation as:

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

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

Let's focus on the term

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

This can be written

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

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

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

Plugging into the Euler equation the above simplifying expression, using the dynamics for the components of the log consumption growth and the log dividend growth given in formula (3) and (4), respectively and rearranging terms we have:

$$
\begin{aligned}
&E_t\Bigg[\exp\quad\Bigg(\theta\log\beta+\left(\theta-\frac{\theta}{\psi}-1\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)-0.5(1-\gamma)\left(\frac{1}{\psi}-\gamma\right)\left(\tilde{\Sigma}_t\right)+\dots\\
&\quad+\kappa_{0,m}+\kappa_{1,m}\left(\sum_{j=1}^{J}z^{m}_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}z^{m}_{j,t}\right)+\left(\sum_{j=1}^{J}gd_{j,t+2^j}\right)\quad\Bigg)\quad\Bigg]\\
&=E_t\Bigg[\exp\quad\Bigg(\theta\log\beta+\left(\theta-\frac{\theta}{\psi}-1\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)-0.5(1-\gamma)\left(\frac{1}{\psi}-\gamma\right)\left(\tilde{\Sigma}_t\right)+\dots\\
&\quad+\kappa_{0,m}+\kappa_{1,m}\left(\sum_{j=1}^{J}z^{m}_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}z^{m}_{j,t}\right)+\left(\sum_{j=1}^{J}\varphi_{d,j}\sigma_{j,t}e^{d}_{j,t+2^j}\right)\quad\Bigg)\quad\Bigg]\\
&=E_t\Bigg[\exp\quad\Bigg(\theta\log\beta+\left(\theta-\frac{\theta}{\psi}-1\right)\left(\sum_{j=1}^{J}\sigma_{j,t}e^{g}_{j,t+2^j}\right)-0.5(1-\gamma)\left(\frac{1}{\psi}-\gamma\right)\left(\tilde{\Sigma}_t\right)+\dots\\
&\quad+\kappa_{0,m}+\left(\kappa_{1,m}\Big(\sum_{j=1}^{J}A^{m}_{j}\underbrace{\left(e_jM\tilde{\Sigma}_t+e_j\varepsilon_{t+2^J}\right)}_{\sigma^2_{j,t+2^j}}\Big)-\sum_{j=1}^{J}A^{m}_{j}\Sigma_{j,t}\right)+\left(\sum_{j=1}^{J}\varphi_{d,j}\sigma_{j,t}e^{d}_{j,t+2^j}\right)\quad\Bigg)\quad\Bigg]=
\end{aligned}
$$

where in the last line we substitute for our guess for the components of log price-dividend ratio given in equation (9). Finally using the log normal property of the shocks, collecting all the $\sigma_{j,t}$ terms and defining the vectors

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

and

$$
\begin{aligned}
\underline{A}_m&\equiv[A^m_1,\dots,A^m_J]\\
\underline{\phi}&\equiv[\phi_1,\dots,\phi_J]
\end{aligned}
\tag{D.4}
$$

we obtain the following restriction in vector notation,

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

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

D.2 The Risk Premium and Return Volatility

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

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

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

Given the solution above for $z^a_{j,t}$ it is possible to derive the innovation to the return $r_{a,t+1}$ as a function of the evolution of the state variables and the parameters of the model. In particular the equilibrium return innovations can be found by plugging the expressions (D.1), (D.2) and (D.3) into the Campbell and Shiller (1988) approximation for log returns, see equation (8) to obtain

$$\begin{aligned} r_{a,t+1} - E_t[r_{a,t+1}] &= \left(\sum_{j=1}^{J} g_{j,t+2^j}\right) + \kappa_0 + \kappa_1 \left(\sum_{j=1}^{J} z_{j,t+2^j}\right) - \left(\sum_{j=1}^{J} z_{j,t}\right) - E_t[r_{a,t+1}] \\ &= \sum_{j=1}^{J} \sigma_{j,t} e^g_{j,t+2^j} + \kappa_1 \left(\sum_{j=1}^{J} A_j(e_j \varepsilon_{t+2^J})\right) \\ &= \sigma_{j,t} \odot e^g_{j,t+1} + \kappa_1 \underline{A} \boldsymbol{\varepsilon}_{t+1} \end{aligned} \tag{D.5}$$

where we define

$$\begin{aligned} \boldsymbol{\varepsilon}^{\mathsf{T}}_{t+1} &\equiv [\varepsilon_{1,t+2^1}, \ldots, \varepsilon_{J,t+2^J}] \\ \sigma_{j,t} \odot e^g_{j,t+1} &\equiv \sum_{j=1}^{J} \sigma_{j,t} e^g_{j,t+2^j} \end{aligned}$$

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

$$r_{a,t+2^j} - E_t[r_{a,t+2^j}] = \sigma_{j,t} e^g_{j,t+2^j} + \kappa_1 \left(A_j \varepsilon_{j,t+2^j}\right) \tag{D.6}$$

It is trivial to show that

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

which allows us to decompose the innovation in aggregate return into the sum of the innovation in the market return components. Further, it follows that the conditional variance of $r_{a,t+1}$ is:

$$Var_t(r_{a,t+1}) = \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1^2 \underline{A}\mathbf{Q}\underline{A}' \tag{D.7}$$

where we define

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

Analogous steps yield the following expression for the market return innovations

$$r_{m,t+1} - E_t[r_{m,t+1}] = \varphi_{d,j}\sigma_{j,t} \odot e_{j,t+1}^d + \underbrace{\kappa_{1,m}\underline{A}_m}_{\beta_{m,\varepsilon}} \boldsymbol{\varepsilon}_{t+1} \tag{D.8}$$

where we define

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

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

$$r_{m,t+2^j} - E_t[r_{m,t+2^j}] = \varphi_{d,j}\sigma_{j,t} e_{j,t+2^j}^d + \underbrace{\kappa_{1,m}\underline{A}_m}_{\beta_{m,\varepsilon}} \varepsilon_{j,t+2^j} \tag{D.9}$$

Using the expression (D.8) we can compute the conditional variance of $r_{m,t+1}$ as follows

$$Var_t(r_{m,t+1}) = \sum_{j=1}^{J} \varphi_{d,j}^2 \sigma_{j,t}^2 + \kappa_1^2 \underline{A_m} Q \underline{A_m}' \tag{D.10}$$

To find the innovations in the stochastic discount factor, we plug the expressions (D.1), (D.2) and (D.3), together with the dynamics for the components of log consumption growth given in equation (3) and our guess for the components of price-consumption ratio solution given in equation (9) into equation (7) to obtain:

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

Finally using the dynamics for our latent factors (5) we obtain

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

which implies

$$
\begin{aligned}
m_{t+1} - E_t[m_{t+1}] &= -\left(1 - \theta + \frac{\theta}{\psi}\right) \sum_{j=1}^{J} \sigma_{j,t} e^g_{j,t+2^j} + (\theta - 1)\kappa_1 \left(\sum_{j=1}^{J} A_j \varepsilon_{j,t+2^j} \right) \\
&= -\lambda_g \sum_{j=1}^{J} e^g_{j,t+2^j} - \sum_{j=1}^{J} \lambda_j \varepsilon_{j,t+2^j} \\
&= -\lambda_g \sigma_{j,t} \odot e^g_{j,t+1} - \underline{\lambda}_\varepsilon \boldsymbol{\varepsilon}_{t+1} \qquad \text{(D.11)}
\end{aligned}
$$

where

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

Using the formula (D.6) and the innovation in the SDF (D.11) we obtain the risk premium for the components of consumption claim asset

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

and using the formula for the return on aggregate wealth (D.5) and the innovation in the SDF (D.11) we obtain the risk premium for the consumption claim asset,

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

where $\sigma^2_{r_{a,t}}$ is defined in equation (11).

Similarly to what we have just done, using the formula (D.6) and the innovation in the SDF (D.11) the premia to the market return components become:

$$E_t[r_{m,t+2^j} - r_{f,t+2^j}] + 0.5\sigma^2_{r_{m,t,j}} = \lambda_g \varphi_{d,j} \sigma^2_{j,t} \rho_j + \kappa_{1,m}[\underline{\lambda}_n \mathbf{Q}]_j A^m_j$$

and using the formula for the innovations in the market return (D.8) and in the SDF (D.11) the premium to the market return becomes:

$$E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{m,t}} = \varphi_{d,j} \sigma^2_{j,t} \odot \rho_j + \kappa_{1,m} \underline{\lambda}_n \mathbf{Q} \underline{A}'_m$$

where $\sigma^2_{r_{m,t}}$ is defined in equation (12) and ρ_j is the correlation at level of persistence j between the stochastic discount factor shocks $e^g_{j,t+2^j}$ and the dividend growth shocks $e^d_{j,t+2^j}$.

D.3 Risk-Free Rate Dynamics

To obtain our expression for the risk-free rate we start by plugging the log short-term real interest rate $r_{f,t+1}$ for r^i_{t+1} into the Euler equation (6). Then by applying the forward decomposition (B.4) to the (demeaned) consumption growth and to the log returns processes at time $t+1$ we observe that the risk-free rate between t and $t+1$, $r_{f,t+1}$ satisfies the following condition:

$$E_t\left[\exp\left(\theta \log\beta - \left(\frac{\theta}{\psi}\right)\sum_{j=1}^{J} g_{j,t+2^j} + (\theta-1)\sum_{j=1}^{J} r_{a,j,t+2^j}\right)\right] = \exp(-r_{f,t+1})$$

where once again $r_{a,t+1}$ is the return on the asset that pays consumption as dividend. Taking logs on both sides and using the log normal properties of the shocks we can rewrite it as follows

$$
\begin{aligned}
r_{f,t+1} &= -\theta \log \beta + \frac{\theta}{\psi} E_t \left[\sum_{j=1}^{J} g_{j,t+2^j} \right] + (1-\theta) E_t \left[\sum_{j=1}^{J} r_{a,j,t+2^j} \right] \\
&\quad - \frac{1}{2} var_t \left[\frac{\theta}{\psi} \sum_{j=1}^{J} g_{j,t+2^j} + (1-\theta) \sum_{j=1}^{J} r_{a,j,t+2^j} \right] \\
&= -\log \beta + \frac{1}{\psi} E_t \left[\sum_{j=1}^{J} g_{j,t+2^j} \right] + \frac{(1-\theta)}{\theta} E_t \left[\sum_{j=1}^{J} r_{a,j,t+2^j} - r_f \right] \\
&\quad - \frac{1}{2\theta} var_t \left[\frac{\theta}{\psi} \sum_{j=1}^{J} g_{j,t+2^j} + (1-\theta) \sum_{j=1}^{J} r_{a,j,t+2^j} \right]
\end{aligned} \tag{D.12}
$$

where in the last line we substract $(1-\theta) r_{f,t}$ from both sides and divide by θ, where it is assumed that $\theta \neq 0$. Further to solve the above expression, note that

$$
var_t \left[\frac{\theta}{\psi} \sum_{j=1}^{h} g_{j,t+h} + (1-\theta) \sum_{j=1}^{h} r_{a,j,t+h} \right] = var_t(m_{t+1})
$$

Recall from (D.11) that

$$
m_{t+1} - E_t[m_{t+1}] = -\lambda_g \sigma_{j,t} \odot e^g_{j,t+1} - \underline{\lambda}_\varepsilon \boldsymbol{\varepsilon}_{t+1}
$$

and therefore

$$
var_t(m_{t+1}) = \lambda_g^2 \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1^2 (1-\theta)^2 \underline{A} \mathbf{Q} \underline{A}'
$$

Note that given the dynamics for log consumption growth (see equation (3)) we have that $E_t \left[\sum_{j=1}^{J} g_{j,t+2^j} \right] = 0$. Eventually, using the expression for the equity premium of return on aggregate wealth we obtain:

$$
r_{f,t+1} = -\log \beta + \lambda_g^2 \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1^2 (1-\theta)^2 \underline{A} \mathbf{Q} \underline{A}' + \frac{(1-\theta)}{\theta} \left(\lambda_g \sum_{j=1}^{J} \sigma_{j,t}^2 + \kappa_1 \underline{\lambda}_\varepsilon \mathbf{Q} \underline{A}' \right)
$$

References

Ang, Andrew and Geert Bekaert, "The Term Structure of Real Rates and Expected Inflation," *The Journal of Finance*, April 2008, *63*, 797849.

Atkeson, Andrew and Patrick J. Kehoe, "On the Need for a New Approach to Analyzing Monetary Policy," NBER Working Papers 14260, National Bureau of Economic Research, Inc August 2008.

Baillie, Richard T. and Ramon P. DeGennaro, "Stock Returns and Volatility," *Journal of Financial and Quantitative Analysis*, June 1990, *25* (02), 203–214.

Bandi, Federico M. and Benot Perron, "Long-run risk-return trade-offs," *Journal of Econometrics*, April 2008, *143* (2), 349–374.

___ **, Rene Garcia, Abraham Lioui, and Benoit Perron**, "A Long-Horizon Perspective on the Cross-Section of Expected Returns," *SSRN eLibrary*, 2010.

Bansal, Ravi, Amir Yaron, and Dana Kiku, "Risk for the Long Run: Estimation and Inference," Working Paper September 2007.

___ **and** ___ , "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 08 2004, *59* (4), 1481–1509.

___ **, Varoujan Khatchatrian, and Amir Yaron**, "Interpretable asset markets?," *European Economic Review*, 2005, *49* (3), 531 – 560.

Barndorff-Nielsen, Ole E. and Neil Shephard, "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics," *Journal Of The Royal Statistical Society Series B*, 2001, *63* (2), 167–241.

Beeler, Jason and John Y. Campbell, "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," NBER Working Papers 14788, National Bureau of Economic Research, Inc March 2009.

Benigno, Pierpaolo, Luca Antonio Ricci, and Paolo Surico, "Unemployment and Productivity in the Long Run: The Role of Macroeconomic Volatility," NBER Working Papers 16374, National Bureau of Economic Research, Inc September 2010.

Beveridge, Stephen and Charles R. Nelson, "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*, 1981, *7* (2), 151–174.

Bollerslev, Tim, George Tauchen, and Hao Zhou, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, November 2009, *22* (11), 4463–4492.

Calvet, Laurent and Adlai Fisher, "Forecasting multifractal volatility," *Journal of Econometrics*, 2001, *105* (1), 27 – 58.

Calvet, Laurent E. and Adlai J. Fisher, "Multifrequency news and stock returns," *Journal of Financial Economics*, October 2007, *86* (1), 178–212.

Campbell, John Y. and Ludger Hentschel, "No news is good news: An asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, June 1992, *31* (3), 281–318.

Campello, Murillo, Long Chen, and Lu Zhang, "Expected returns, yield spreads, and asset pricing tests," *Review of Financial Studies*, May 2008, *21* (3), 1297–1338.

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

Canzoneri, Matthew B., Robert E. Cumby, and Behzad T. Diba, "Euler equations and money market interest rates: A challenge for monetary policy models," *Journal of Monetary Economics*, October 2007, *54* (7), 1863–1881.

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

Corsi, Fulvio, "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 2009, *7* (2), 174–196.

Fadili, M.J and E.T Bullmore, "Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors," *NeuroImage, Volume 15, Number 1*, 2002.

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, "Expected stock returns and volatility," *Journal of Financial Economics*, September 1987, *19* (1), 3–29.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov, "MIDAS Regressions: Further Results and New Directions," *Econometric Reviews*, 2007, *26* (1), 53–90.

___ **, Pedro Santa-Clara, and Rossen Valkanov**, "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 2005, *76* (3), 509 – 548.

Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, December 1993, *48* (5), 1779–1801.

Hall, Robert E, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, April 1988, *96* (2), 339–57.

Harvey, Campbell R., "The specification of conditional expectations," *Journal of Empirical Finance*, December 2001, *8* (5), 573–637.

Jensen, Mark J., "Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility Models," *Journal of Time Series Analysis*, November 2004, *25* (6), 895–922.

Justiniano, Alejandro and Giorgio E. Primiceri, "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, June 2008, *98* (3), 604–41.

Lettau, Martin and Sydney C. Ludvigson, "Measuring and Modelling Variation in the Risk-Return Trade-off," *Handbook of Financial Econometrics*, 2010, *1*, 617–690.

___ , ___ , and Jessica A. Wachter, "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?," *Review of Financial Studies*, July 2008, *21* (4), 1653–1687.

Ludvigson, Sydney C. and Serena Ng, "The empirical risk-return relation: A factor analysis approach," *Journal of Financial Economics*, January 2007, *83* (1), 171–222.

Malkhozov, Aytek and Maral Shamloo, "Asset Prices in a News Driven Real Business Cycle Model," Technical Report March 2010.

Merton, Robert C, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, September 1973, *41* (5), 867–87.

Ortu, Fulvio, Andrea Tamoni, and Claudio Tebaldi, "Long-Run Risk and the Persistence of Consumption Shocks," Working Papers 14260 August 2010.

___ , ___ , and ___ , "Technical Appendix to "Long Run and the Temporal Aggregation of Risks" , URL: `http://www.igier.unibocconi.it/finance`," Technical Report March 2010.

Pastor, Lubos, Meenakshi Sinha, and Bhaskaran Swaminathan, "Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital," *Journal of Finance*, December 2008, *63* (6), 2859–2897.

Zhou, Hao, "Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty," *SSRN eLibrary*, 2010.

Part II
Long Run Risk and the Persistence of Consumption Shocks

1 Introduction

This paper contributes to the ongoing research on long-run risk in asset pricing. The contribution consists in a consumption-based asset pricing model in which long-run risk in consumption growth is priced and contributes to the equity premium as in Bansal and Yaron (2004) and yet the model is free of the empirically unsupported implication that consumption growth is predicted by the price-dividend ratio (see Constantinides and Ghosh (2008) and Beeler and Campbell (2009)).

Our extension is motivated by the empirical observation that consumption growth is generated both by *predictable* low frequency variations and by *non-predictable* highly volatile, high frequency idiosyncratic variations. This empirical evidence that motivates our extension is obtained by disaggregating consumption growth into cyclical components, classified by their level of persistence (or characteristic half-life). While consumption growth remains unpredictable at the aggregate level, it does contain cyclical components that are predictable. These predictable components, moreover, are highly correlated with well known structural drivers of consumption variability. On the longest side, for instance, demographic shocks are highly correlated with the component describing consumption growth variations that occurs on time scales which range between 16 and 32 years. On the intermediate side, long-run productivity growth explains cyclical variations with time scale between 8 and 16 years. On the shortest side, finally, an high frequency predictable component with a yearly half-life is found and can be identified with the well documented fourth quarter effect (see Moller and Rangvid (2010))).

As the above mentioned empirical evidence suggests, it is therefore important to develop an asset pricing model where consumption responds to shocks of heterogeneous durations. Inspired by this observation and to study the effect of such diversity, we introduce a parsimonious equilibrium model where a representative agent with Epstein-Zin preferences faces an exogenous consumption and dividend streams driven by many factors, each one operating over different time horizons. Given our preference choice, these same factors enter the stochastic discount factor of the agent thus affecting asset returns. Our model differs from the standard long-run risk along three main dimensions. First, whereas in the standard long-run risk model stock prices respond strongly to variation in future aggregate consumption and dividend growth, in our model this relation between price variations and the fluctuations in both consumption and cash flows growth is disaggregated across different levels of persistence in order to make the model consistent with the (lack of) empirical evidence of aggregate consumption predictability. Second, while predictability is induced exogenously by latent factors, our model identifies "observable" well-known drivers of specific components of consumption growth which are classified by their level of persistence. Third, since in our model the stochastic discount factor dynamics are driven by shocks with highly heterogeneous durations, we are able to characterize the dependence of the price of risk on the investment horizon and to reconstruct the entire term-structure of the risk-return trade-off.

The model builds on a decomposition of time series into the sum of components with different persistence levels.[1] This decomposition allows us to disaggregate consumption and dividend growth into different components each one driven by its own state variable. Consistent with the idea that each state variable operates at a specific frequency we model their dynamics with a multiscale autoregressive process, i.e. each autoregressive operates over a time interval of increasing length and its autoregressive coefficient uniquely identifies the persistence of the shocks. With this specification for the dynamics of the state variables our model is extremely parsimonious and tractable and yields closed-form expressions for equilibrium prices and return dynamics.

This approach generates very important implications for consumption predictability. In fact the presence in the same time series of highly persistent components with small volatility together with highly volatile components with low persistence can hide the predictable relation generated by the most persistent ones. From this point of view we interpret the findings of Beeler and Campbell (2009) as indicating the absence of predictability only for the low persistence components of consumption growth which are likely to provide the largest contribution to aggregate consumption volatility. On the other hand once we properly disaggregate the time series of interests across different levels of persistence, we do find that price components reflect future prospects of consumption components, as predicted by our model.

Our model classifies the shocks impinging the economy along two competing dimensions: their size as measured by their instantaneous volatility and their persistence as measured by their half life. Controlling simultaneously for these two dimensions we are able to obtain an interesting decomposition for the equity premium across different time horizons. In particular the term structure of equity premium implied by our model allows us to conclude that high frequency components which are responsible for most of the consumption growth variance produce a negligible contribution to equity premia; on the contrary low frequency, thin (in variance) components account for most of the long-run risk contribution to equity risk premia. Intuitively in the short-run the effect on prices of slow moving structural changes like those induced by technological innovation or by demographic trends, is completely hidden by the myopic and volatile reaction of markets to the incoming flow of information. However as the

[1] Whereas in this paper we concentrate on the valuation implications of such an interpretation, a formal discussion of the econometric methodological issues required to extract these components can be found in the companion paper OTT2010b.

valuation horizon increases, the effect of transitory shocks is averaged out while persistent structural trends emerge as the driving forces of long-run expectations and play a pivotal role in the rational valuation of assets. This suggests that our classification of shocks is potentially critical for asset valuation.

Finally, this paper contributes to the debate on the estimates of the intertemporal elasticity of substitution (IES) of the representative agent. By separating the consumption growth into cyclical components we obtain an estimate of the IES which is strictly greater than one and the Bansal and Yaron (2004)) hypothesis that the substitution effect dominates the income one, which means that the elasticity of intertemporal substitution must be greater than one, is empirically supported. If instead we do not account for persistence heterogeneity our estimated equation collapses to the one in Beeler and Campbell (2009) where the data provide strong statistical evidence against an IES greater than one.

The above results provide strong empirical support to the following conclusion: the decomposition of the time series into cyclical components classified by their level of persistence identifies predictable patterns present in aggregate consumption and dividends growth data which cannot be detected using traditional models. Importantly these predictable variations in consumption and cash flows are priced by the market and contribute significantly to the explanation of the equity premium.

The remainder of the paper is organized as follows. The next subsection concludes the introduction with a review of the literature. Section 22.1 revisits Bansal and Yaron (2004) to show briefly that long-run risk being priced generates the counterfactual implication that the consumption growth and the price-dividend ratio have the same level of persistence. This motivates the needs to find a viable method to separate a time series in components characterized by their levels of persistence. Section 22.2 introduces quickly such a method and then Section 22.3 uses it to filter the time series of interests in our long-run risk asset pricing model which accounts for persistence heterogeneity. Section 3 explores the main empirical findings obtained applying the new persistence based decomposition. Section 4 concludes.

1.1 Related literature

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.

In their seminal contribution to long run risk valuation Bansal and Yaron (2004) 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). The long-run risk theoretical framework has motivated many empirical tests on the presence of long run consumption risk in the data.[2] For example Parker and Julliard (2005) find 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. Along the same line 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.

Nevertheless some recent papers, namely Constantinides and Ghosh (2008) and Beeler and Campbell (2009), evaluate the long-run risks model and find reversal of earlier conclusions. The main contribution of our paper is to reconcile these evidence within a theoretical framework which still allows for long-run risk in consumption growth.

Other recent and interesting implementations of long-run risk model try to understand the source of persistent predictable component in consumption growth. With this respect 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 (greater than ten years), and are shown to carry over into a small, highly persistent component of aggregate consumption. Analogously Kaltenbrunner and Lochstoer (2010) and Croce (2010) show that consumption and savings decisions of agents in a production economy lead to low-frequency movements in consumption growth that are linked to the conditional mean of productivity growth. Similar to these authors we find that shifts in the long-run rate of productivity growth are one of the key factors driving the slow-moving consumption components.

With regard to the estimation of the intertemporal elasticity of substitution, the empirical literature has produced contradictory evidence on this point. On one hand Hall (1988) and Campbell and Mankiw (1989) estimated an extremely small value of IES on US data and Campbell (2003) summarizes these results and finds similar patterns in international data. On the other hand Attanasio and Weber (1993) and Beaudry and van Wincoop (1996) have found values of the IES higher than one using disaggregated cohort-level and state-level consumption data. Vissing-Jorgensen (2002), moreover, pointing out that many consumers

[2]Recent literature focuses on the asset pricing implications in equity, bond and currency markets refer, for instance, to Bansal and Shaliastovich (2010) and Koijen, Lustig, Nieuwerburgh and Verdelhan (2010b). We do not pursue this line of research further and we just focus on the equity markets.

do not participate actively in asset markets, finds that, in household data, the IES is greater than one for asset market participants. Similar to the latter studies we present empirical evidence on aggregation problems with the relationship between consumption growth and the real interest rate and we show that using disaggregated consumption data is key to find a value for the IES greater than one. However we add to this literature since the main driver of our results is the intrinsic persistence heterogeneity in consumption and not the differences in preferences and/or opportunity sets for different cohort or the different levels of stock market participation.

Finally our work, which builds on a decomposition of time series in a sequence of shocks classified by their level of persistence, 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. Our technique can also be related to the multiplicative permanent-transitory decomposition proposed in Hansen and Sheinkman (2009) and used in Hansen, Heaton and Li (2008) and Alvarez and Jermann (2005). For a formal analysis of the link between these two spectral approaches we refer the interesting reader to OTT2010b.

2 A Long Run Risk Model with Heterogeneous Persistence

In this section we first revisit Bansal and Yaron (2004) (BY04 henceforth) to show briefly why long-run risk being priced generates the counterfactual implication that the consumption growth and the price-dividend ratio have the same level of persistence. Motivated by this fact we then introduce our long-run valuation model which accounts for persistence heterogeneity. We focus our attention on the case in which second moments are constant although the further extension to stochastic volatility would entail no formal impediments. By assuming constant volatilities of log consumption growth and log dividend growth we are able to better concentrate our attention on the primary research question under debate, that is whether fluctuations in the conditional mean of consumption and dividend growth are indeed priced.

2.1 Long-run risk versus consumption and price-dividend persistence

We consider a simplified version of the BY04 in which aggregate consumption is equal to the aggregate dividend.[3]

[3] BY04 introduce a leverage effect between dividends and consumption in order to allow dividend growth to be more volatile than consumption growth and to allow for an imperfect correlation between consumption growth and dividend growth, as it is in the data. Our point holds regardless of this effect being present.

The data generating process for consumption is then as follows:

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

where g_{t+1} is the log rate of consumption growth and x_t is the state variable of the model. Relying on log-linear approximations for the log return on the market portfolio BY04 show that, in equilibrium, the log price-dividend ratio z_t^m is an affine function of the state variable, i.e.

$$z_t^m = A_{0,m} + A_{1,m}x_t \qquad (2)$$

and the equity premium is given by

$$E[r_{m,t+1} - r_{f,t}] + 0.5var(r_{m,t}) = \gamma\sigma^2 + (1-\theta)\kappa_{1,m}^2 A_{1,m}^2 \phi_e^2 \sigma^2 \qquad (3)$$

Observe now that, in equation (3), the long-run risk contribution to the expected excess returns is proportional to $A_{1,m}$. Thus, as long as the representative agent has non-trivial Epstein-Zin preferences, i.e. $\theta \neq 1$, long-run risk is priced in the BY04 model if and only if $A_{1,m}$ is different from zero. Considering equations (1) and (2) together, therefore, long-run risk is priced if and only if the consumption growth and the price-dividend ratio have the same decay rate in the autocorrelation function[4], i.e. the same persistence.

This implication of the model however is empirically rejected. On the one hand, in fact, the price-dividend is (close to) a unit root process, as it is well documented in the literature, see e.g. Torous, Valkanov and Yan (2004), Campbell and Yogo (2006) and Lettau and Nieuwerburgh (2008). On the other hand consumption growth resembles closely a white noise process and therefore does not share the high persistence of the price-dividend ratio. This point is well synthesized in Figure 1 that displays the very different statistical behavior of the demeaned price-dividend and consumption growth series. The figure shows that over the sample 1947Q2-2009Q4 the price-dividend ratio has crossed its mean value much less often

[4] This follows immediately from the fact that, given the equations in (1), the expression for autocovariance of the consumption growth at lag k is given by

$$cov(g_t, g_{t+k}) = \frac{\rho^k \phi_e^2 \sigma^2}{1-\rho^2}$$

and as long as $A_{1,m} \neq 0$ then the covariance for the price-dividend ratio will decay as ρ^k as well.

than the consumption growth. In fact Campbell and Shiller (2001) report that the price-dividend ratio has crossed its mean value only 29 times since 1872. The intervals between crossings for the price-dividend ratio range from one year to twenty years, the twenty-year interval being the one between 1950 and 1970. The persistence of the consumption growth is only moderate however, the half life of consumption growth shocks being 1 year.[5]

Not surprisingly the empirical literature has tested and rejected other implications of the BY04 model. Constantinides and Ghosh (2008), for instance, examine the ability of the model to explain the returns of the market portfolio and find that the average pricing error is substantial. This fact can be easily understood if one recalls that in order to price the average equity returns the stochastic discount factor must have a large permanent component (see Alvarez and Jermann, 2005 and Koijen, Lustig and Van Nieuwerburgh, 2010a). In BY04 however the persistence of the stochastic discount factor, which depends on the price-dividend ratio[6], is tied to the low persistence of the consumption growth. It is then not surprising that using this constrained pricing kernel they find that the model is rejected at the annual frequency over the 1930-2006 sample. Beeler and Campbell (2009), moreover, test the model by evaluating the ability of the log price-dividend ratio z_t^m, proxing for the latent state variable x_t, to predict consumption growth. As noted in Figure 1 however, consumption growth is, at quarterly horizon, close to a white noise and therefore it would be difficult to find evidence of a predictable persistent component at this frequency of observations. Not surprisingly the simple OLS regressions of consumption growth on the log price-dividend ratio run by Beeler and Campbell (2009) display relatively little, if any, predictability of consumption growth in the data.

A possible way to reconcile the findings of Beeler and Campbell (2009) with the long-run risk framework is to think of consumption growth as the sum of components with different levels of persistence, where the highly persistent ones contribute for a very small fraction to the total volatility of aggregate consumption growth and yet are predictable by the highly persistent components of the financial ratios. If this was the case the contemporaneous presence in the same time series of highly persistent components with small volatility together with

[5]Similarly Paseka and Theocharides (2010) find that, by relaxing the equilibrium restriction (2) that requires the price-dividend to be affine in the long-run risk variable, the persistence of the latent mean consumption growth corresponds to an half life of about 1.3 years for the $1934 - 2005$ period.

[6]In our simplified version of BY04 the log pricing kernel can be expressed in terms of observables, namely the aggregate log price-dividend ratio, and its lags, and consumption growth. In the general version of the BY04 model, the aggregate log price-dividend ratio and log interest rate are affine functions of the long-run risk variable and the conditional variance of its innovation. In this more general case Constantinides and Ghosh (2008) show that it is possible to express the log pricing kernel as an affine function of the aggregate log price-dividend ratio, log interest rate, and their lags, in addition to consumption growth. As we said in this paper we do not consider stochastic volatility since our interest is just in the the long-run risk channel and not in macroeconomic uncertainty.

highly volatile components with low persistence would generate a severe errors-in-variable problem in the Beeler and Campbell (2009) regression, a problem hiding the predictability relation which could eventually hold for specific components.[7] In order to see if this alternative way of interpreting the time series could help reconciling the above results, we first need to develop a tool to decompose the time series of interest into components with different levels of persistence. We briefly introduce this tool in the next section and then we use it in our asset pricing model to account for persistence heterogeneity.

2.2 A Persistence-based Decomposition of Time Series

The above discussion highlights the necessity to classify the components of a time series on the basis of their level of persistence. To give the basic intuition behind our decomposition we can think of applying a h-period moving average to our time series. The averaging action smooths the original time series by removing the high frequency components. The output of this moving-average filter is a new time series which conveys information only about those cyclical components with periodicity greater than h periods. Of course, by increasing the window length h of the moving-average filter we would be able to extract components that decay more and more slowly. This procedure, however, imposes only a lower bound on the persistence of the filtered components whereas our aim is to obtain a component with a well-defined level of persistence. Ideally, in fact, we would like to isolate those fluctuations of the original time series that lie within a specific band of frequencies. To do so one can consider the outputs of two moving-average filters with windows $h > h'$. Since the moving average filters yield two time series, one characterized by fluctuations longer than h' periods and one by fluctuations longer than h periods, the difference between these two time series should in principle identify the component reflecting the fluctuations of the original time series with periodicity between h' and h periods.

To lay down some basic notation useful in what follows, given a time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ consider its sample mean over a window of past observations with size 2^j:[8]

$$\pi_t^{(j)} = \frac{1}{2^j}\sum_{p=0}^{2^j-1} x_{t-p} \tag{4}$$

[7]We come back to this point in Section 33.2 and 33.3

[8]Note that the window of values over which the average is made increases exponentially with base two. The moving average filters with such a property are optimal in the sense that they satisfy the principles of multiresolution analysis (see e.g. Mallat (1989a), Mallat (1989b), Daubechies (1990) and Daubechies (1992)). See also OTT2010b.

where $j \geq 1$ and $\pi_t^{(0)} \equiv x_t$. Following the intuition given above, the component $x_{j,t}$ that identifies the fluctuations of the time series $\mathbf{x}$ with periodicity between 2^{j-1} and 2^j periods[9] is then filtered out as follows:

$$x_{j,t} = \pi_t^{(j)} - \pi_{t-2^{j-1}}^{(j-1)} \tag{5}$$

In the sequel of the paper we refer to j as the level of persistence of the component $x_{j,t}$. Since in the empirical work one always deals with time series of finite length T, we also define the maximum (observable) level of persistence J as the the greatest integer such that $2^J \leq T$. Clearly given the data sample it is impossible to draw any inference about the persistence of shocks that last longer than 2^J periods.[10]

Simple algebra shows that, for any given J, the generic element x_t, the components $\{x_{j,t}\}_{j=1}^J$ and the permanent part $\pi_t^{(J)}$ are related via the following identity:[11]

$$x_t = \sum_{j=1}^{J} x_{j,t} + \pi_t^{(J)} \tag{6}$$

This relation fits well with the intuition that current financial and economic quantities are the result of the overlay of past fluctuations with different periodicity, where these fluctuations can go from the extreme of an incoming flow of information at high frequency (low j) to the one of slowly moving structural changes (high j) like those induced by technological innovation or demographic trends. In Section 33.4 we foster this idea by showing that the predictable consumption components are indeed highly correlated with well known structural drivers of consumption variability.

The components $\{x_{j,t}\}_{j=1}^J$ can also be used to obtain an alternative decomposition of the original time series, i.e.

$$x_{t+1} = \sum_{j=1}^{J} x_{j,t+2^j} + \pi_{t+2^J}^{(J)} \tag{7}$$

Intuitively this decomposition, which in the sequel will be referred as the forward decomposition, can be interpreted as a way to reconstruct the realization of x_{t+1} from the effect that this realization will have at different horizons and it will turn out to be a fundamental tool in working out our model.

---

[9]For an interpretation of the cycle durations corresponding to the persistence level j in the case of quarterly time series see Table 4.

[10]In Appendix AA.2 we relate the persistence properties of the series $x_{j,t}$ and $\pi_t^{(J)}$ to their Fourier spectra.

[11]The algebra behind relations (6) and (7) is carried out in Appendix AA.1

A last comment is in order. In general the component $\pi_t^{(J)}$ is meant to capture the finite sample permanent component of the time series, in the sense that it conveys information about all those cyclical components with periodicity greater than 2^J periods including, potentially, the truly permanent component.[12] However in the rest of the paper we will focus on stationary time series for which it turns out that the component $\pi_t^{(J)}$ simply captures the rolling mean and contributes very little to the total variance of x_t and has therefore little, if any, explanatory power for the original time series. Figure 2 provides graphical evidence supporting this statement for the case of consumption growth. The top panel plots the demeaned consumption growth together with the sum of its components, excluding the permanent one. The two series are close to each other with a correlation of 0.97. The bottom panel shows instead the difference between the unconditional mean of consumption growth and $\pi_t^{(J)}$. This difference vanishes as the sample length increases.[13] A similar conclusion can be drawn for the dividend growth and the financial ratios series. This is why in the rest of the paper we focus on demeaned time series which allows us to neglect the component $\pi_t^{(J)}$ in both (6) and (7).

We now turn to our asset pricing model in which all the time series of economic significance are described as sum of components with different persistence levels, as in equations (6) or, alternatively, (7).

2.3 The Long-run Risk Model with Heterogeneous Persistence

Following the approach discussed in the previous section, we incorporate in the standard long-run risk model the decomposition of time series into components with different levels of persistence so that the log consumption growth, g_t, and the log dividend growth, gd_t take the following form:

$$g_t = \sum_{j=1}^{J} g_{j,t}$$
$$gd_t = \sum_{j=1}^{J} gd_{j,t}$$

[12]In OTT2010b we show that, for the case where the series of observations is infinite and if the time series had a component which persists beyond any time scale, for instance if the time series were I(1), one would have that π_t^{∞} would converge to the Beveridge-Nelson permanent component. Intuitively one could in principle keep on iterating over (4) and (5). However an arbitrary number of iterations of the moving average filter on the integrated time series will not be sufficient to remove such a permanent component.

[13]Intuitively equation (4) tells us that $\pi_t^{(J)}$ simply captures the rolling mean, which, for any stationary time series, asymptotically converges to the sample mean, i.e. $\pi_t^{(J)} \approx E[x_t]$ for t large enough.

where $g_{j,t}$ and $gd_{j,t}$ denote the components with level of persistence j as defined in the previous section. The novelty now is to assume that each component of consumption growth, $g_{j,t}$ and of dividend growth, $gd_{j,t}$ is driven by its own state variable, $x_{j,t}$, i.e.

$$\begin{aligned} g_{j,t+2^j} &= x_{j,t} + e^g_{j,t+2^j} \qquad (8) \\ e^g_{j,t+2^j} &\sim N\left(0, \sigma^2_{g,j}\right) \\ gd_{j,t+2^j} &= \phi_j x_{j,t} + e^d_{j,t+2^j} \qquad (9) \\ e^d_{j,t+2^j} &\sim N\left(0, \sigma^2_{d,j}\right) \end{aligned}$$

where we allow the shocks to be correlated across levels of persistence (for fixed time t) but not across time (for fixed persistence level j) and we assume the consumption shocks $e^g_{j,t+2^j}$ to be mutually independent from the dividend ones $e^d_{j,t+2^j}$. To close the dynamics of the model we assume that the components $\{x_{j,t}\}_{j=1}^J$ follow a multiscale autoregressive process, i.e.

$$\begin{aligned} x_{j,t+2^j} &= \rho_j x_{j,t} + \varepsilon_{j,t+2^j} \qquad (10) \\ \varepsilon_{j,t+2^j} &\sim N\left(0, \left(\sigma^{(j)}\right)^2\right) \end{aligned}$$

In words, we are modeling separately the conditional mean $x_{j,t}$ of each one of the components of consumption and dividends growth. Importantly, equations (8) to (10) represent a natural way to incorporate persistence heterogeneity in the long-run risk framework while retaining its pedagogical simplicity. On one side, in fact, these equations allow consumption growth to accommodate different degrees of persistence so to break the link between the autocorrelation of consumption and price-dividend. On the other side they maintain the simplicity of having, at each level of persistence j, only one variable driving the respective consumption component.

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

$$E_t\left[e^{m_{t+1}+r^i_{t+1}}\right] = 1 \qquad (11)$$

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

$$m_{t+1} = \theta \log \beta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r^a_{t+1} \,, \qquad (12)$$

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

In what follows we provide the basic steps to determine the pricing kernel and risk premia on the market portfolio in our long-run risk model with persistence heterogeneity.[14] Recall first that by the standard Campbell and Shiller (1988) log-linear approximation for returns one obtains:

$$
\begin{aligned}
r_{a,t+1} &= \kappa_0 + \kappa_1 z^a_{t+1} - z^a_t + g_{t+1} \\
r_{m,t+1} &= \kappa_{0,m} + \kappa_{1,m} z^m_{t+1} - z^m_t + gd_{t+1}
\end{aligned}
\tag{13}
$$

where z^a_t, z^m_t, denote the log price-consumption and the log price-dividend ratio respectively. Recalling our decomposition of consumption and dividends into components with different levels of persistence, and denoting with $z^a_{j,t}$, $z^m_{j,t}$, the components with persistence j of the (log) price-consumption ratio and (log) price-dividend ratio respectively, it is natural to conjecture that there exists *component by component* a linear relation between the financial ratios and our state variables $x_{j,t}$, i.e.

$$
\begin{aligned}
z^a_{j,t} &= A_{0,j} + A_j x_{j,t} \\
z^m_{j,t} &= A^m_{0,j} + A^m_j x_{j,t}
\end{aligned}
\tag{14}
$$

As long as A_j and A^m_j are not vanishing, these relations and equation (8) together imply that the components of price-consumption $z^a_{j,t}$ and price-dividends $z^m_{j,t}$ lead the component of consumption and dividends with the same level of persistence j.

The values of $A_{0,j}, A_j, A^m_{0,j}, A^m_j$ in terms of the parameters of the model are obtained from the Euler condition (11) after the log stochastic discount factors and the returns are all expressed in terms of the factors $\{x_{j,t}\}_{j=1}^J$ and of the innovations $\{e^g_{j,t+2^j}\}_j$ and $\{\varepsilon_{j,t+2^j}\}_j$. In Appendix B we show that plugging these expressions for the stochastic discount factor and for the returns into the Euler equation and using the method of undetermined coefficients one obtains a system of equations for the coefficients $A_{0,j}, A_j, A^m_{0,j}, A^m_j$, the solution of which

[14] All details behind our calculations are given in Appendix B.

is given by the following vectors of sensitivities:

$$\underline{A} = \left(1 - \frac{1}{\psi}\right) (\mathbb{I}_J - \kappa_1 M)^{-1} \underline{1}$$
$$\underline{A_m} = (\mathbb{I}_J - \kappa_{1,m} M)^{-1} \left(\underline{\phi} - \frac{1}{\psi}\underline{1}\right)$$

where

$$M = diag\left(\rho_1, \ldots, \rho_J\right)$$

i.e. the matrix collecting on the diagonal the persistence parameters of the components, $\mathbb{I}_J$ is the identity matrix, $\underline{\phi}$ is a column vector with entries $\phi_1, \ldots, \phi_J$ which reflect the exposures of the market dividends components to the consumption growth ones and $\underline{A}$ and $\underline{A_m}$ denote the column vectors with entries, $A_1, \ldots, A_J$, $A_1^m, \ldots, A_J^m$, respectively.

To study the implications of persistence heterogeneity for the equity premium recall that the risk premium on any asset i satisfies, in this set-up, $E_t[r_{i,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{i,t}} = -cov_t(m_{t+1}, r_{i,t+1})$. In the Appendix we show that the innovations of the stochastic discount factor are given by

$$m_{t+1} - E_t[m_{t+1}] = -\left(\frac{\theta}{\psi} - \theta + 1\right) e^g_{t+1} - \kappa_1(1-\theta)\underline{A} \cdot (\boldsymbol{\varepsilon}_{t+1}) \tag{15}$$

while analogous steps yield the following expressions for the return innovations

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

where

$$\boldsymbol{\varepsilon}^{\intercal}_{t+1} \equiv [\varepsilon_{1,t+2^1}, \ldots, \varepsilon_{J,t+2^J}]$$
$$e^g_{t+1} \equiv \sum_j e^g_{j,t+2^j}$$
$$e^d_{t+1} \equiv \sum_j e^d_{j,t+2^j}$$

With the innovations to the equilibrium returns at hand and using (15), one can finally compute the risk premia for the consumption claim asset, $r_{a,t+1}$ and for the market portfolio,

$r_{m,t+1}$, hence obtaining

$$E_t[r_{a,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{a,t}} = \lambda_g \sigma^2_g + \kappa_1 \underline{\lambda}'_\varepsilon \mathbf{Q} \underline{A}$$
$$E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{m,t}} = \kappa_{1,m} \underline{\lambda}'_\varepsilon \mathbf{Q} \underline{A_m}$$

where

$$\lambda_g \equiv \left(\frac{\theta}{\psi} - \theta + 1 \right)$$
$$\underline{\lambda}_\varepsilon \equiv \kappa_1 (1-\theta) \underline{A}$$
$$\mathbf{Q} = \mathbf{E}_t \left[\boldsymbol{\varepsilon}_{t+1} \boldsymbol{\varepsilon}'_{t+1} \right]$$
$$\sigma^2_g = Var\left(e^g_{t+1}\right)$$

The innovations $\varepsilon_{j,t+2^j}$ driving the components of consumption growth at scale j reflect the impact of current uncertainty many years into the future. In analogy with the interest rate literature, they are similar to forward rates and we therefore refer to $\boldsymbol{\varepsilon}_{t+1}$ as the term structure of risk. The parameter $\underline{\lambda}_\varepsilon$ determines the risk compensation for these innovations.[15] Whereas BY04 price only the shock with persistence 0.979 affecting the *aggregate* consumption growth, in our asset pricing model we price all the shocks, each characterized by its own level of persistence, driving the *components* of consumption growth. The expressions for $\underline{\lambda}_\varepsilon$ reveal that the degrees of persistence of consumption growth components affect the risk premium on the asset. In particular a rise in persistence increases $\underline{\lambda}_\varepsilon$. The exposure of the market return to these shocks is $\mathbf{Q}\underline{A_m}$. Importantly the exposure of the market return is determined simultaneously by the size of the shocks as measured by their instantaneous volatility, captured by $\mathbf{Q}$ and by their persistence as measured by their half life. It is therefore key in order to obtain the entire term structure of risk-return trade-offs to decompose the aggregate shocks that impinge an economy along these two competing dimensions. The process defined in equation (10) is very well suited to capture this fact since it classifies the shocks $\varepsilon_{j,t+2^j}$ both by its volatility $\sigma^{(j)}$ and its half-life as measured by ρ_j. This classification is key for our model because it allows to determine precisely the short-run and long run-dynamics. In fact although highly volatile shocks with low level of persistence j can dominate in the short-run, as the valuation horizon increases, the effect of these shocks is averaged out while persistent (high j) trends emerge and play a pivotal role. Thus we have obtained a picture where the asset valuation involves the full term structure of shocks driving the predictable components of consumption growth. In the next section we empirically investigate some of

[15] The parameter λ_g determine the risk compensation for the independent consumption shock e^g_{t+1} is present also in BY04 model and it is standard as λ_g equals the risk aversion parameter γ.

the implications of our asset pricing model with heterogeneous persistence.

3 Empirical implications

In this section we first analyze the statistical behavior of the components with different levels of persistence that can be filtered out of the relevant time series. We then use the time series disaggregated across levels of persistence to revisit the consumption predictability test proposed by Beeler and Campbell (2009) and to estimate the intertemporal elasticity of substitution (IES). Our main empirical finding can be summarized as follows: once consumption growth is disaggregated across different levels of persistence, consistently with our model, some of its components are predictable by financial ratios and the estimates of IES is found to be greater than one. Last but not least, setting risk aversion to the conservative value of $\gamma = 5$, we compute the term structure of risk premia implied by our model and we show that thin predictable components contribute significantly to the equity premium.

3.1 Persistence Based Decomposition of the Relevant Time Series

In order to evaluate the implications of the long-run risks model with persistence heterogeneity we look at four variables: the changes in log consumption and dividends, the log price-dividend ratio and the log price-consumption ratio. Following BY04 and Beeler and Campbell (2009)[16] we use data on US nondurables and services consumption from the Bureau of Economic Analysis. We make the standard "end-of-period" timing assumption that consumption during period t takes place at the end of the period. The price-dividend ratio and dividend growth rates are obtained from the CRSP files. All nominal quantities are converted to real, using the personal consumption deflator. We consider a postwar quarterly US series over the period 1947:Q2-2009:Q4 and for robustness a long-run annual series over the period 1930-2009. In what follows we report results only for the consumption growth and the price-dividend ratio series. The conclusions for the dividend process are quite in line with that of the consumption series, while the the price-consumption ratio behaves in quite the same way as the price-dividend series.

Since in Section 22.1 (recall in particular in Figure 1) we have already commented upon the very different statistical behavior of the consumption growth and price-dividend ratio, we now turn to the analysis of the components of these two time series. To do so we apply the decomposition described in Section 22.2 to the aggregate time series. Figure 3 and Figure 4 display the components of the consumption growth and the price-dividend ratio for the

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

quarterly and annual sample respectively.

As we move toward higher persistence levels, i.e. greater j, a common long-run behavior between the two series becomes apparent. This fact will be further explored in the next section when we discuss in details the predictability of consumption and dividend growth components by the financial ratios.

To further dig into the statistical properties of the components of the time series used in our empirical analysis, we focus on three key dimensions: the unconditional correlation, the half-life or persistence and, finally, the contribution of each component to the total unconditional variance of the aggregate series. First we inspect the pairwise correlation between the components of consumption growth and report Pearson's p-values[17] in Table 2. The Pearson's correlation test indicates that almost all of the consumption growth components are pairwise uncorrelated. The correlation is significant at standard levels only between the second and third and between the third and fourth components. For these components, however, the Pearson's correlation coefficients are no greater than 0.15. Overall these findings suggest that, although the components with different periodicity extracted with our approach could be in principle correlated across different levels of persistence, in practice we find very little evidence in favor of interaction across levels.[18]

Second we show that each component is stationary and has a well defined (in terms of interval) level of persistence. In order to do so we fit to each component an autoregressive process of order one. Estimates of the autoregressive coefficients and the R^2 are shown in Table 4 and 5 for the consumption process and price-dividend ratio, respectively. We verify that each component is strongly stationary and has a degree of persistence identified by the root of the autoregressive process.

Now recall that the filtering procedure described in Section 22.2 should identify, at level of persistence j, the fluctuations of the generic time series with periodicity belonging to the interval $[2^{j-1}, 2^j)$. To check the goodness of our method we denote with ρ_j the root of the autoregressive process of order one fitted onto the j-th component of the time series[19] and approximate the half-life $HL(j)$ of this component by the standard relation $HL(j) \simeq -ln(2)/ln(\rho_j)$. Then consistent with our discussion in Section 22.2 we should observe that the so computed half-life $HL(j)$ for the generic component at level of persistence

[17] Similar results are obtained using Spearman rank-order correlation coefficients test.

[18] We carry exactly the same analysis for the dividend growth and the financial ratios. Analogous results are obtained; in fact the correlations are significant only at similar low levels of persistence and are never greater than 0.15 for all the time series. Although we do not report them here they are available upon request.

[19] The autoregressive coefficients ρ_j for the consumption growth and the price-dividend ratio series are reported in Tables 4 and 5 respectively.

j is included in the interval $[2^{j-1}, 2^j)$. This is indeed the case and in particular we find that the half-life extrapolated from the autoregressive root is generally very close to the lower bound of the interval.[20]

Finally we report in Table 3 the contribution of the components $g_{j,t}$ and $z^m_{j,t}$ to the total variance of consumption growth and price-dividend respectively. We note that the highly persistent components of consumption growth yield a minor contribution to the total variance of the aggregate series. In particular each of the sixth and seventh components yield about 5% of total variance. The opposite happens for the price-dividend series: the components at levels 6 and 7 account for more than half of the total variance. This evidence contributes to explain why the aggregate time series of consumption and price-dividend have a very different persistence behavior. In fact, since the great part of the variability in consumption and price-dividend is explained by high frequency (i.e. low j) and low frequency (i.e. high j) components respectively, then the aggregate time series of consumption growth and price-dividend will resembles a white noise and a (close to) unit root process, respectively. The fact that the highly persistent components contribute for a very small fraction to the total volatility of aggregate consumption growth explains also why the predictability that exists at a specific level of persistence j disappears at the aggregate level. Indeed since the long-run components of consumption at level of persistence $j = 6$ and $j = 7$ are clearly overwhelmed by the high frequency noise then the comovements highlighted in Figure 3 between consumption growth and price-dividend ratio at these levels of persistence do not emerge unless we suitably separate the informative low frequency components from the noisy high frequency ones. We will make this argument more formal in subsection 3.3.

The above discussion and the evidence presented so far suggest therefore that the time series of interest in this paper can empirically be well represented as the sum of autoregressive components, each of which has an half-life belonging to a well defined interval. Intuitively this makes sense from an economic point of view because, for instance, consumption data results from an aggregation through time and across heterogeneous households. Even assuming that each component in this aggregation follows a simple autoregressive process, still the aggregation procedure would generate persistence heterogeneity.[21] Finally this section, and in particular Figures 3 and 4, confirm the importance of decomposing across levels of

[20]For instance the half-life of the consumption growth components at levels $j = 5, 6, 7$ is about $5, 11$ and 21 years, respectively. This numbers are consistent with the fact that these components should identify fluctuations with periodicity, measured in years, belonging to the interval [4, 8), [8, 16) and [16, 32) respectively.

[21]Aggregating these heterogeneous components results in the long memory property of aggregate consumption which is well documented in the literature (see e.g. Lippi and Zaffaroni (1998) and Thornton (2008)).

persistence the time series of interest in order to uncover important economic relations that would be otherwise hindered by the noisy components. In the next subsections we employ the disaggregated time series in order to test whether the empirical implications of our model are indeed supported by the data.

3.2 Predictability of Consumption and Dividend Growth

In the standard long-run risk model stock prices respond strongly to variation in expected future aggregate consumption growth. The innovation we propose in this paper consists in the fact that this relation holds componentwise even when it is not necessarily required to hold for the aggregate series.

In particular our long-run risk model with persistence heterogeneity implies that the components of price-consumption $z^a_{j,t}$ and price-dividends $z^m_{j,t}$ lead the components of consumption with the same level of persistence j (see relations (8) and (14)). In light of our model, therefore, we first disaggregate our variables across different levels of persistence and then we quantify the predictability at each level of persistence by running the following regressions:

$$g_{j,t+2^j} = \beta_{0,j} + \beta^g_{1,j} z^m_{j,t} + \varepsilon_{t+2^j,j} \tag{16}$$
$$g_{j,t+2^j} = \beta_{0,j} + \beta^g_{1,j} z^a_{j,t} + \varepsilon_{t+2^j,j}$$

Although these regressions resemble those run by Beeler and Campbell (2009) we remark that while they focus on the aggregate time series we study instead predictability at a specific level of persistence. Clearly from an empirical point of view it can be the case that the predictability relation holds for all, some of or none of the persistence levels j. It is notationally useful to let $S \subset \{1, .., J\}$ denote the set of persistence levels for which the components of consumption growth and dividend growth are led by the financial ratios. Hence the persistence levels belonging to S select the characteristic components of consumption growth which are predictable. Results for the quarterly sample are reported in Table 6 and Table 7. Table 6 shows that at levels of persistence $j = 3, 6, 7$ the coefficients on the price-dividend ratio are statistically significant at the 5% level and that the sixth and seventh components account for a great part of the variation in the future consumption growth at the corresponding scale, the R^2 being between 24% and 38% respectively. Empirically we have $S = \{3, 6, 7\}$, i.e. the components of the price-dividend ratio that actually lead the corresponding components of consumption growth have cycles of length, measured in years, belonging to the intervals $\{[1\ ,2], [8\ ,16], [16\ ,32]\}$. Table 7 shows that the same components of consumption growth that are predictable by the price-dividend ratio are also predictable by the price-consumption ratio. Finally, as a robustness check we perform the same consumption predictability test but now using annual data. Table 8 reports the results and shows that they are consistent with the ones obtained for the quarterly series: the component with level of persistence $j = 3, 6, 7$ turn out to be the only statistically significant ones.

Our long-run risk model with heterogeneous levels of persistence implies also that the same latent state variables $x_{j,t}$ that generate persistent variations in consumption growth at levels of persistence $j \in S$ should generate variations at the same levels of persistence in the dividend growth components. A natural test of this implication is to see if the components of price-consumption $z^a_{j,t}$ and price-dividends $z^m_{j,t}$ lead the corresponding dividend growth components. Table 9 reports the results[22] and shows that the very same components at levels $3, 6, 7$ that are significant for consumption growth are also statistically significant at the 5% level for the dividend growth. The sixth and seventh components, in particular, account for a great part of the variation in the expected future dividend growth at the corresponding scale, the R^2 being between 25% and 38% respectively.

Once again we remark that we are not providing evidence in favor of strong predictability of aggregate consumption growth but, in fact, we just provide evidence for the presence of persistent components in consumption which cause at the corresponding levels of persistence, stock price variability (relative to dividends). Therefore the above results are not in contrast with the ones of Beeler and Campbell (2009). The persistent variations in consumption growth, unless filtered, are indeed overwhelmed by measurement error and aggregate predictability therefore does not occur.

It is now interesting to compare the filtering and estimation approach used in this paper with two other techniques, namely cointegration and long-horizon regressions, which have been adopted to analyze long-run economic and financial relations.

The cointegrated approach has been used in recent empirical work, e.g. Bansal et al. (2007) and Ferson, Nallareddy and Xie (2010), to model the persistent component in consumption and dividend processes. In particular Bansal et al. (2007) argue that the cointegrating relation between dividends and consumption is a good measure of long-run consumption risks. Our long-run risk model with heterogeneity in persistence does not rule out this possibility. In fact the presence in consumption and dividends series of seasonal patterns with a very long half-life suggests to interpret the components at levels 6 and 7 as the common

[22]We report results only for the case where the regressor is the price-dividend ratio and the data are quarterly. Conclusions do not change when we use the price-consumption series as the regressor and/or when we use longer annual series.

trend driving the cointegrating relation between consumption and dividends. However our interpretation points to a kind of cointegration different from the traditional one used in Bansal et al. (2007) and Ferson et al. (2010), since our components are persistent but not permanent. Therefore our suggested cointegration should rely upon the notion of seasonal cointegration relationships (see Osborn (1993)), which is a generalization of the classic concept of cointegration proposed in the seminal work of Engle and Granger (1987) where we may consider cointegrating relations not only at zero frequency but also at other (two) frequencies connected with long-run cycles (Engle et al., 1993).[23] In this framework the parallel common movements in the components at persistence levels 6 and 7 in the dividends and consumption variables may be therefore interpreted as a cointegration at seasonal frequencies and long-run risk may well be captured by (seasonal) cointegration relations.

With regard to long-horizon regressions, we note that our approach, which uses filtered regressand and regressors in ordinary least squares, relates to the one of Torous et al. (2004)[24] and Bandi and Perron (2008) where the authors study predictability relations aggregating both the regressor and regressand over non-overlapping periods having the same length. This is very similar to what we do in equations (16) where we regress the consumption components obtained using aggregate consumption data in the time window $t+1$ to $t+2^j$ onto the price-dividend components in the time window $t-2^j$ to t.

From the methodological point of view, the same statistical caveats which are present in long-horizon predictive regressions also apply here. In particular Section 22.2 explains that our filtering procedure is based on moving-average filters. This smoothing operation in turns generates 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. To address this problem in this paper the standard errors are computed using the method proposed by Hansen and Hodrick (1980) to correct for serially correlated errors.[25]

Finally since our entire procedure relies on the linear filtering technique described in Section 22.2, to decompose our time series, it is important to check whether our results are

[23] According to the definition proposed primary by Hylleberg et al. (1990) seasonal cointegration means the cointegration only at seasonal frequencies. However in many theoretical and practical works (Johansen and Schaumburg, 1999), the term seasonal cointegration analysis corresponds to the cointegration analysis conducted at not only seasonal but nonseasonal frequencies as well. We will use this term in this more general meaning.

[24] In fact Torous et al. (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.

[25] Results are robust when standard errors are computed using Newey-West with optimal lag length estimators, similar to Beeler and Campbell (2009).

driven by the particular choice of filter made here. As a robustness check for the filtering procedure we use the bandpass filter described in Christiano and Fitzgerald (2003).[26] The choice of frequencies interval characterizing the band-pass filter is determined according to Section 22.2: 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 10. We find again evidence of a very long-lasting component in consumption growth (note in fact that the R^2 spikes at the scales 6 and 7).

In summary, we find that at suitably defined levels of persistence stock prices predict the long-run prospects for consumption and dividend growth. In the next section we reconcile our evidence of componentwise predictability with the one presented in the very recent long-run risk literature which rules out predictability at the aggregate level.

3.3 Componentwise versus Aggregate Consumption Predictability and the Errors-in-Variable Problem

Why does the componentwise predictability, i.e. the predictability we find when we disaggregate the consumption and dividend growth across different levels of persistence, wash away at the aggregate level? The main reason is that if one does not filter appropriately the time series of interest then the empirical results are plagued by an errors-in-variables problem which hides the predictability of components with a specific levels of persistence.

To understand where this errors-in-variables problem arises, consider the following simple case where consumption growth is the sum of a persistent component g_t^*, and an idiosyncratic one, and the price-dividend ratio is a noisy proxy for the state variable x_t driving the persistent component of consumption growth, i.e.

$$\begin{aligned} g_{t+1}^* &= \mu + x_t + \sigma\eta_{t+1} \\ g_t &= g_t^* + \nu_{2,t} \\ z_t^m &= A_{0,m} + A_{1,m}x_t + \nu_{1,t} \end{aligned} \tag{17}$$

where $\nu_{1,t}$ and $\nu_{2,t}$ are the consumption growth and price-dividend ratio idiosyncratic components, respectively. Note that this model differs from the standard long-run risk one in two dimensions. First, the relation (1) does not hold for the aggregate consumption but only for its persistent component g_t^*. Second, the key aspect of our simple model is that the components $\nu_{1,t}, \nu_{2,t}$ and the common component x_t across the two processes, have different levels of persistence. Similarly to the standard long-run risk model we assume that g_t^* is highly

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

persistent and explains a small fraction of the total variance of the consumption growth.

If one now looks for predictability at the aggregate level and runs the following regression:

$$g_{t+1} = \beta_0 + \beta_1 z_t^m + \epsilon_{t+1}$$

the predictive effect could be largely underestimated. In fact notice that the price-dividend ratio z_t^m covaries only with the small component g_t^* of consumption growth. This covariation manifests itself on a characteristic time scale much longer than the interval of observation and thus it could be hidden at short horizons by the large in volatility idiosyncratic component.

More formally the results are plagued by a typical errors-in-variable problem,[27] the OLS estimator being downward biased and inconsistent. In order to understand why this happens, let's substitute $g_t^* = g_t - \nu_{2,t}$ and $x_t = \frac{z_t^m - \nu_{1,t} - A_{0,m}}{A_{1,m}}$ into the first equation in (17). This yields

$$\begin{aligned} g_{t+1} &= \mu + \frac{z_t^m - \nu_{1,t} - A_{0,m}}{A_{1,m}} + \sigma\eta_{t+1} + \nu_{2,t+1} \\ &= \beta_0 + \beta_1 z_t^m + \underbrace{\sigma\eta_{t+1} + \nu_{2,t+1} - \beta_1\nu_{1,t}}_{\varepsilon_{t+1}} \end{aligned} \tag{18}$$

It is now easy to see that the presence of measurement error in g_t and z_t^m leads to an increase in the variance of the error term. In addition the regressor z_t^m and the error term ε_{t+1} in (18) are correlated and this covariance does not depend on the sample size, it does not vanish asymptotically, and hence the OLS estimator is downward biased and inconsistent. These facts contribute to explain why a simple regression on scaled stock prices does not reveal the long-run prospects for consumption.

Hence when the underlying economic relation holds only for the unobserved processes g_t^* and x_t and when the size of the component g_t^* measured by instantaneous (single period) volatility is very small compared to the instantaneous total volatility of consumption growth as in this example, the filtering of these latent components becomes a critical issue for the empirical analysis.[28] Our approach manages to uncover the true economic relation by pre-filtering out the idiosyncratic components $\nu_{1,t}$ and $\nu_{2,t}$ from the observables series g_t and z_t^m and then by using the filtered regressand g_t^* and regressor x_t in a simple OLS framework. The pre-filtering procedure exploits the key assumption that the idiosyncratic components and the predictable one have different levels of persistence.

[27]In the standard EIV set-up $\nu_{1,t}$ and $\nu_{2,t}$ are i.i.d homoskedastic shocks.

[28]The importance of filtering procedures to avoid EIV problem has been recently highlighted, see also Gencay and Gradojevic (2009).

In the previous section we highlighted the presence of consumption components predictability at specific levels of persistence, namely $j \in S = \{3, 6, 7\}$. It is interesting to investigate the existence of reasonable economic proxies for these consumption components. To search for these proxies we rely on time series that are economically significant, that are characterized by an half-life close to the one of the components they are to proxy for, and that are significantly correlated with such component.

First we focus on the third component filtered out of consumption growth, $g_{3,t}$, whose mean reversion is between one and two years. In order to identify this component with observable economic factors we follow the lead of Jagannathan and Wang (2007) and Moller and Rangvid (2010) who analyze the ability of the fourth-quarter consumption growth rate to predict expected excess returns on stocks. Importantly, this variable aims at capturing economic and financial choices happening with yearly frequency. Figure 5 reports the series $g_{3,t}$ and the one used by Moller and Rangvid (2010). Quite remarkably the correlation between these two series is 0.60 in our sample period.

The economic idea behind Jagannathan and Wang (2007) and Moller and Rangvid (2010) is motivated by an alignment between consumption and investment decisions in the fourth quarter. Indeed the infrequent points in time where investors decide to review their investments are most likely influenced by culture (such as Christmas) and institutional features (such as end-of-year bonuses and the tax consequences of capital gains and losses, which both occur mainly in the fourth quarter of the year). If the points in time when consumption and investment decisions are taken coincide, such as in the fourth quarter, a clear relation between consumption decisions and stock prices should emerge. Our third component seems to be a good candidate to capture this effect.

We now turn to the sixth and seventh components, which are slow moving series with an half life of about 8 and 16 years, respectively. To search for a valid proxy for the sixth component we follow the lead of the recent macro-finance literature suggesting that technology prospects should be positively related with aggregate consumption. For instance Garleanu et al. (2009) argue that consumption growth over long-horizons should reveal the position of the economy with respect to the technological cycle. In particular the authors show that predictable components of consumption that occur at cycles between 10 and 15 years are due to the presence of large infrequent embodied technology shocks. Similarly Kaltenbrunner and Lochstoer (2010) and Croce (2010) investigate the implications of long-run risk in a general equilibrium production economy and show how shocks to productivity growth generate predictable movements in consumption growth. Moreover Hsu and Huang (2010) show that changes in technology prospects are risk factors which explain the growth of aggregate consumption. Based on this motivating evidence we investigate whether our consumption component at level of persistence $j = 6$ plays the role of shocks to productivity growth. We plot in Figure 6 the sixth component of consumption growth together with the long-run multifactor productivity index.[29] The correlation between the two series is a comforting 0.64 giving further support to the idea that highly persistent time-variation in (components of) consumption growth, i.e. long-run risk, can reflect permanent technology shocks.

As for the seventh component, we look at the literature linking demographic fluctuations to long-run stock prices. For instance, Geanakoplos, Magill and Quinzii (2004) and more recently Favero, Gozluklu and Tamoni (2010) have shown that changes in the distribution of the population age account for long-run cycles in the U.S. stock market. Geanakoplos et al. (2004) analyze an overlapping generation model in which the demographic structure mimics the pattern of live births in the US. Importantly live births in the US have featured alternating twenty-year periods of boom and busts, and therefore are consistent with the half-life of our seventh component. One implication of their model is that real stock prices are positively related to the ratio of "middle aged population to population of young adults" (the so called MY ratio) even when investors are forward-looking and rational. Favero, Gozluklu and Tamoni (2010) provide empirical evidence for the presence of a slowly evolving trend in the log dividend-price ratio, determined by the MY ratio. Figure 6 reports the seventh component of consumption growth along with the series for MY. The correlation in the full sample is equal to 0.44. Note that the demographic variable still leaves place for unexplained variability in the slow-moving component of consumption growth. A possible way to improve the identification of the seventh component would entail considering persistent improvements in the degree of risk sharing among households or regions (see Lustig and Van-Nieuwerburgh (2006)) or the persistent changes in the tax code (see McGrattan and Prescott (2005)).

Summing up we have shown that the predictable components of consumption growth are highly correlated with well known structural drivers of consumption variability. On the longest side we found demographic shocks to be highly correlated with the component describing consumption growth variations that occurs on time scales which range between 16 and 32 years. On the intermediate side, long-run productivity growth explains cyclical variations with time scale between 8 and 16 years. On the shortest side, finally, we find an high frequency predictable component with a yearly half-life that can be identified with the well documented fourth quarter effect.

[29]Data are from the Bureau of Labor Statistics; the sample spans 1948-2008. In particular, the index measures the value-added output per combined unit of labor and capital input in private business and private nonfarm business. Available at ftp://ftp.bls.gov/pub/special.requests/opt/mp/prod3.mfptablehis.zip.

In this Section we aim at estimating the intertemporal elasticity of substitution (IES) by separating the consumption growth into cyclical components. To obtain the empirical relation that allows us to estimate the IES we derive first the expression for the risk-free rate in our long-run risk model. This allows us to obtain, under the maintained assumption of non stochastic second moments, a simple linear relation that links the components of the real interest rate to the ones of consumption growth via the intertemporal elasticity of substitution.

The starting point of this analysis is the following expression which is derived in Appendix B.3:[30]

$$r_{f,t+1} = \frac{1}{\psi}\sum_j x_{j,t} \tag{19}$$

In order to bring this relation to the data we need two further steps. First we rewrite relation (19) by applying the forward decomposition (7) to its left-end side. By doing so we obtain

$$\sum_{j=1}^{J} r_{f,j,t+2^j} = \frac{1}{\psi}\sum_j x_{j,t}$$

Then using the equation (8) to link the latent component $x_{j,t}$ with the consumption growth component and introducing measurement errors, the following set of J testable implications obtains:[31]

$$r_{f,j,t+2^j} = \frac{1}{\psi} g_{j,t+2^j} + \sigma_{f,j}\ \eta_{j,t+2^j} \qquad j = 1, \ldots, J \tag{20}$$

This is a set of relations which is constrained by the condition that the coefficient linking in the linear regression the information content of the risk-free rate components to the consumption growth ones, must be the same at all levels of persistence. Table 11 displays the results of the above constrained system of equations. The first row shows that when we decompose the risk-free rate and consumption growth across the different levels of persistence we estimate the IES to be significant at standard levels and equal to 4.762. The reported t-statistics are based on GMM corrected standard errors in order to cope with the overlapping

[30] The relation in Appendix B.3 includes also a constant α_f. As already said in Section 22.2 we focus on demeaned time series which allows us to neglect α_f and also the component $\pi_t^{(J)}$ in both (6) and (7).

[31] Indeed using the equation (8) one obtains $\sum_{j=1}^{J} r_{f,j,t+2^j} = \frac{1}{\psi}\sum_j g_{j,t+2^j} - \frac{1}{\psi}\sum_j e^g_{j,t+2^j}$ and theoretically one could draw inference on ψ both from the loadings on the consumption growth components and from the variance of the innovations. However observe that once we add measurement errors at each time scale the volatilities of the innovations $e^g_{j,t+2^j}$ and the volatilities of the measurement errors cannot be separated on the basis of the information set that we have. We therefore draw inference based only on the loading coefficients on the consumption growth components.

observations problem introduced by our moving-averages. In this special case where the parameter to be estimated is equal at all levels of persistence, one can solve the overlapping problem by adopting the technique suggested in Fadili and Bullmore (2002). In particular Fadili and Bullmore (2002) suggest to (sub)sample the components at level of persistence j with frequency 2^j[32] in order to get rid of the autocorrelation problem and then to apply to the so obtained sampled time series the generalized least squares estimator (GLS). [33] We report the results obtained using this approach in the second and third rows of Table 11. The second row uses the sample period 1978Q1-2009Q4, that is exactly 128 data points. The third row uses the sample period 1948Q1-2009Q4, which is composed of 248 data points. and 8 data points are missing to reach the critical dimension of 256 points. We fill these 8 points with either a sequence of zeros or by using reflecting boundaries. The results are unaltered.[34] Importantly the estimates that we obtain are strongly significant, all above one and close to the value obtained in the case where overlapping data have been used.

It is useful to compare our results with the standard regression approach[35] originally suggested by Hansen and Singleton (1983) to estimate the elasticity of intertemporal substitution and followed by Hall (1988) and Campbell and Mankiw (1990), among many others. Observe that under the assumption that the conditional expectation of the consumption growth components, i.e. the latent variables $\{x_{j,t}\}_{j=1}^{J}$, sum up to the expectation of the aggregate consumption, i.e. the single latent variable x_t used in BY04, formally if $\sum_j x_{j,t} = x_t$, then equation (19) can be rewritten as

$$r_{f,t+1} = \frac{1}{\psi} x_t$$

and using the BY04 dynamics for consumption growth (see equation (1)) to proxy for the

---

[32] Note that if we apply our decomposition to a time series with $T = 2^J$ elements we then obtain J components with T elements. If we subsample the components we obtain a new time series with $T/2 + T/4 + ... + T/2^J = T$ elements, that is the new sampled series has the same length of the original one.

[33] More precisely this estimator makes use of the decimated (not-redundant) Haar transform which yields a diagonalized covariance matrix of the regression errors, i.e. the off-diagonal elements can be set to zero. Diagonalization simplifies numerical identification of parameter estimates and implies that the WLS estimator is theoretically approximate to the best linear unbiased (BLU) estimator and can provide maximum likelihood estimates of both signal and noise parameters, namely $\frac{1}{\psi}$ and σ_η.

[34] Fadili and Bullmore (2002) study in detail the effect of artifactual inter-coefficient correlations introduced by boundary correction at the limits of the data and show that WLS is unbiased over a wide range of data conditions and its efficiency closely approximates theoretically derived limits

[35] More in details instrumental variables (IV) are used.

latent variable x_t, the set of relations (20) collapse to the standard regression[36]

$$r_{f,t+1} = \frac{1}{\psi} g_{t+1} + \sigma_f \eta_{t+1} \tag{21}$$

Empirical tests carried out by Hall (1988) and Campbell and Mankiw (1990) find a low estimate of ψ which contradicts the assumption of the long-run risk model that assets are priced by a representative agent with an elasticity of intertemporal substitution greater than one. A potential explanation of these results is that (particularly in postwar quarterly data) the real interest rate is very volatile relative to predictable variation in consumption growth (see Beeler and Campbell (2009) for a discussion). Therefore unless we disentangle the highly volatile noise component from the low volatile informative ones it would be difficult to estimate properly the IES.

This is exactly what we do when we apply the persistence based decomposition before running the regressions. In fact compared to regression (21), equation (20) mandates to consider all of the latent variables driving the consumption components. Thus the testable implications of our model require to properly take into account the heterogeneity in consumption growth generated by the mixture of highly volatile and the slowly evolving components (see again equation (19)). By doing so a robust estimate larger than one is obtained producing empirical support to a key hypothesis of the long-run risk valuation approach. Our empirical findings are in agreement with previous studies, e.g. Attanasio and Weber (1993), Beaudry and van Wincoop (1996) and Vissing-Jorgensen (2002) who find values for ψ higher than one. In fact similar to these studies we present empirical evidence on aggregation problems with the relationship between consumption growth and the real interest rate and we show that using disaggregated consumption data is key to find a value for the IES greater than one. However whereas these studies focus on consumption data disaggregated at cohort-level, state-level and household level, respectively, we suggest persistence heterogeneity as an additional key dimension along which consumption can be disaggregated.

It is important to observe that these studies use consumption data disaggregated at cohort-level, state-level and household level respectively.

3.6 The term structure of equity market risk premium

In subsection 3.2 we find that stock prices reveal the long-run prospects for consumption and dividend growth once the persistence level is properly taken into account. The ultimate

[36] Alternatively, one can reverse the regression and estimate $\Delta c_{t+1} = \beta_0 + \psi r_{f,t+1} + \eta_{t+1}$. However, if ψ is large as it will turn out to be the case in our empirical exercise, then it is better to estimate the equation reported in the text.

relevance of the predictability effects in the components of consumption and dividends is related to the ability of these "thin persistent effects" to generate sizeable risk premia. It is therefore crucial to quantify the contribution of the predictable components to risk premia within our long-run risk model with persistence heterogeneity.

Recall that in our model the equity premium for the market portfolio $r_{m,t+1}$ satisfies:

$$
\begin{aligned}
E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{m,t}} &= \kappa_{1,m}\underline{\lambda}_\varepsilon \mathbf{Q}\underline{A_m} \\
&= \kappa_1 \underbrace{\kappa_1(1-\theta)\underline{A}}_{\underline{\lambda}_\varepsilon} \mathbf{Q}\ \underline{A_m}
\end{aligned} \tag{22}
$$

where

$$
\begin{aligned}
\underline{A} &= \left(1 - \frac{1}{\psi}\right)(\mathbb{I}_J - \kappa_1 M)^{-1}\underline{1} \\
\underline{A_m} &= (\mathbb{I}_J - \kappa_{1,m} M)^{-1}\left(\underline{\phi} - \frac{1}{\psi}\underline{1}\right)
\end{aligned}
$$

The aggregate risk premia reflect the risk exposures $\mathbf{Q}\ \underline{A_m}$ and the risk prices $\underline{\lambda}_\varepsilon$ at all levels of persistence. We are then able to characterize the contribution to the equity premium of the risk components at each different horizon. This is similar in spirit to Hansen and Sheinkman (2009) and Borovicka, Hansen, Hendricks and Scheinkman (2009) where the authors are interested in the entire term structure of risk prices. Moreover note that whereas the risk prices tend to increase with the level of persistence, the risk exposures do not since as we have already seen the highly persistent predictable components contribute for a very small fraction to the total volatility of aggregate consumption growth. This yields non trivial dynamics and in particular the equity premium does not need to increase with the level of persistence.

Aside from ψ which has been already taken care of in the previous section, to compute the equity premium we need an estimate for M, $\mathbf{Q}$ and $\underline{\phi}$. With these parameters at hand we can obtain the equity premium for calibrated values of γ.

The risk aversion parameter is set to $\gamma = 5$ whereas according to the previous section 3.5 we set $\widehat{\psi} = 4.76$.[37] The persistence levels matrix M and the innovations' variance-covariance matrix $\mathbf{Q}$ are obtained by fitting a vector autoregressive system to the price-dividend components where the matrix M is restricted to be diagonal. Finally we need an estimates of $\underline{\phi}$. Recalling the equations for the consumption growth dynamics (8) and

[37]Our conclusions do not change significantly even when we use the conservative value $\psi = 1.5$ suggested in BY04.

the linear relations (14) for the financial ratios it is immediate to verify that the coefficient $\hat{\beta}^{g}_{1,j}$ obtained by regressing componentwise the consumption growth on the price-dividend produces an estimate of the coefficient $\frac{1}{A_{j,m}}$ (see Table 6). By the same token, using the equations for the dividend growth (9) and together with equations (14), we see that the coefficient $\hat{\beta}^{gd}_{1,j}$ estimated from the regression of the components of log dividend growth on the components of log price-dividend ratio yields an estimate of $\frac{\phi_j}{A_{j,m}}$ (see Table 9). Therefore an indirect estimate of ϕ_j is given by

$$\hat{\phi}_j = \frac{\hat{\beta}^{gd}_{1,j}}{\widehat{\beta}^{g}_{1,j}}$$

With these parameter values at hand we compute the equity premium. Results are reported in Table 12. Notice that the equity premium can 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 6 and Table 9 we observe that the risk exposure can be different from zero only at scale 3, 6 and 7. 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 our long-run risk model with persistence heterogeneity, long-run risk offers a plausible solution to the equity premium puzzle.

4 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 errors-in-variable problem generated by the heterogeneity of persistence.

This paper shows that a long-run risk model, 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 trade-offs in asset valuation. On the empirical side it is clear that the next step to verify the credibility of the long-run risk with persistence heterogeneity is the extension of the analysis to bond prices and to state dependent volatility. 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.

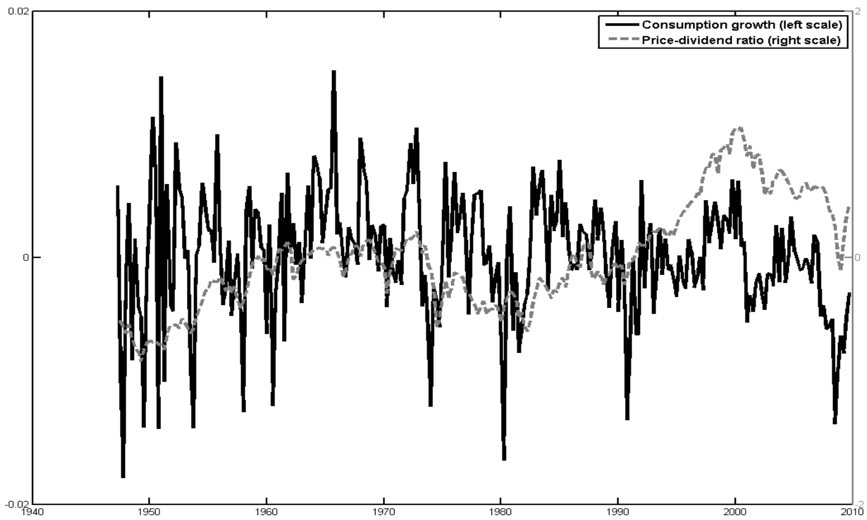


Figure 1: The figure displays the series of US consumption growth (nondurables and services) from the Bureau of Economic Analysis and the log price-dividend ratio (dashed line).

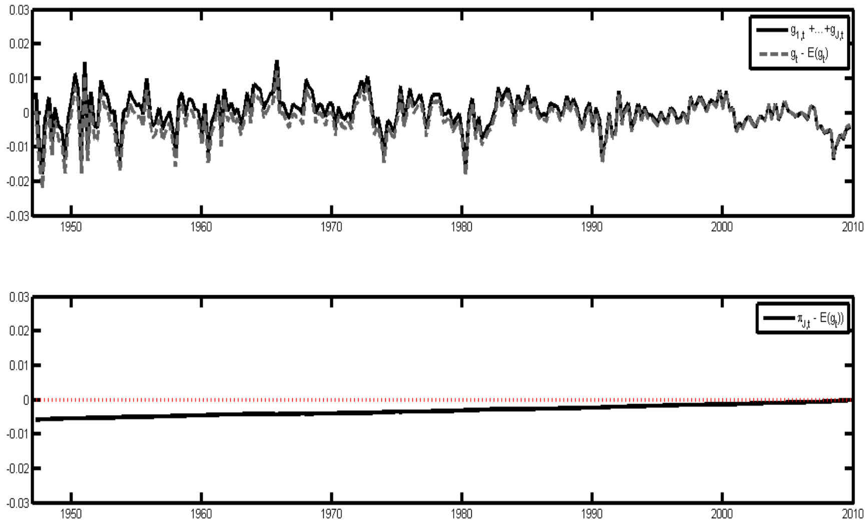


Figure 2: The figure displays the effect of removing the permanent component $\pi_{J,t}$. The top panel reports the *demeaned* series of US consumption growth (nondurables and services) from the Bureau of Economic Analysis consumption growth (dashed line) together with the sum of its components, excluding the permanent one (solid line). The correlation between the two series is 0.97. The bottom panel shows the difference between the permanent component $\pi_{J,t}$ and the sample mean of the consumption growth. This difference vanishes as the sample length increases. Both panels are obtained for $J = 8$.

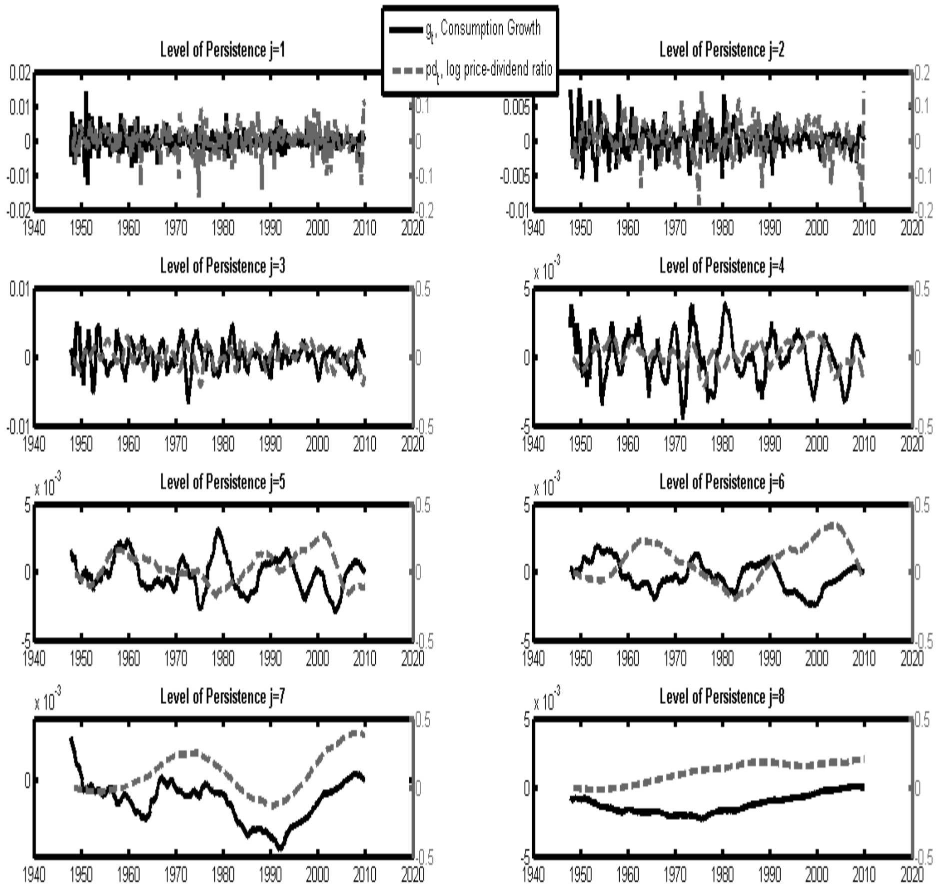


Figure 3: Time-scale decomposition for the log price-dividend pd_t and log consumption growth based upon quarterly data. The figure displays the components of consumption growth $g_{j,t+2^j}$ obtained using the forward decomposition along with the components of the price-dividend ratio $z^m_{j,t}$ obtained using the backward decompostion. The sample spans the period 1947Q2-2009Q4.

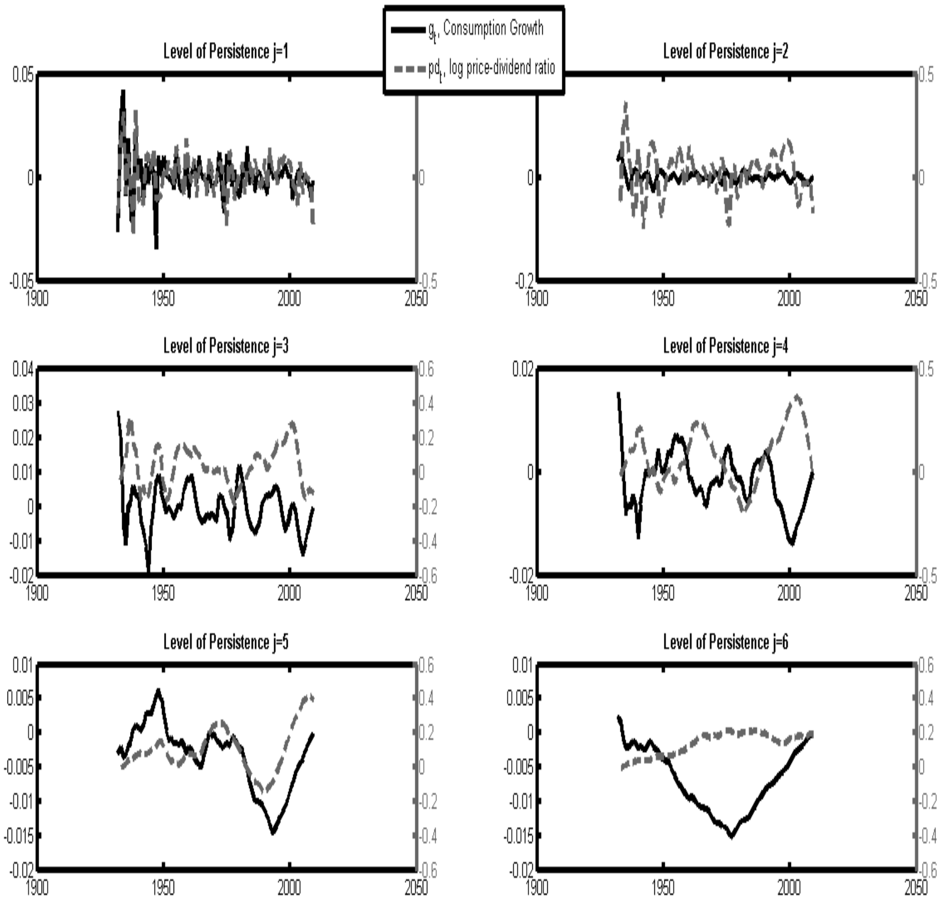


Figure 4: Time-scale decomposition for the log price-dividend pd_t and log consumption growth g_t based upon annual data. The figure displays the components of consumption growth $g_{j,t+2^j}$ obtained using the forward decomposition along with the components of the price-dividend ratio $z^m_{j,t}$ obtained using the backward decompostion. The sample spans the period 1930-2009.

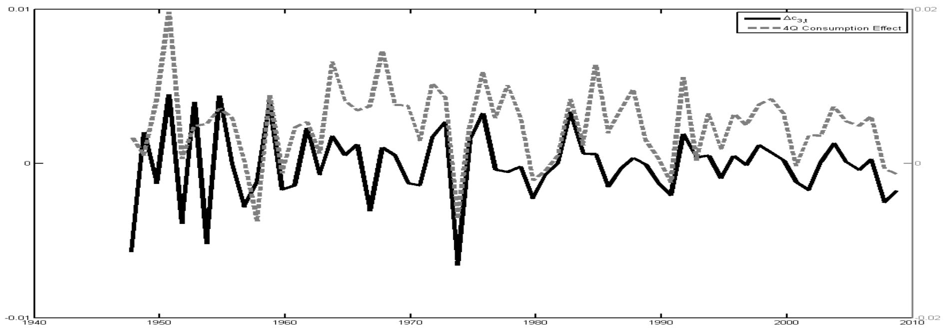


Figure 5: The figure displays the third component of consumption growth, $g_{3,t}$ along with the real consumption from the third quarter of a calendar year to the fourth quarter as suggested in Rangvid (2010).

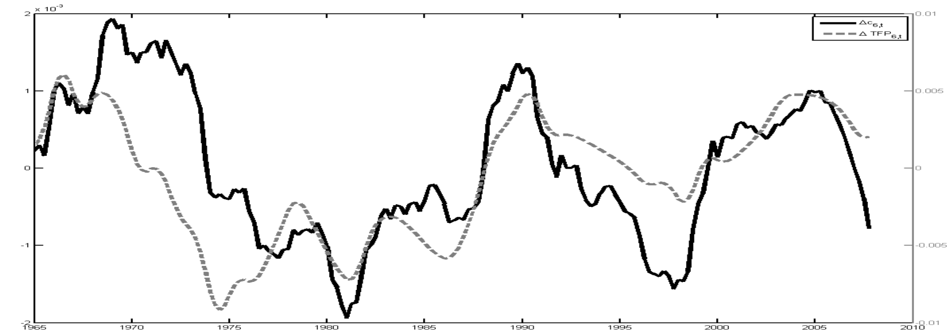


Figure 6: The figure displays the sixth component of consumption growth, $g_{6,t}$ and total factor productivity, $\Delta TFP_{6,t}$.

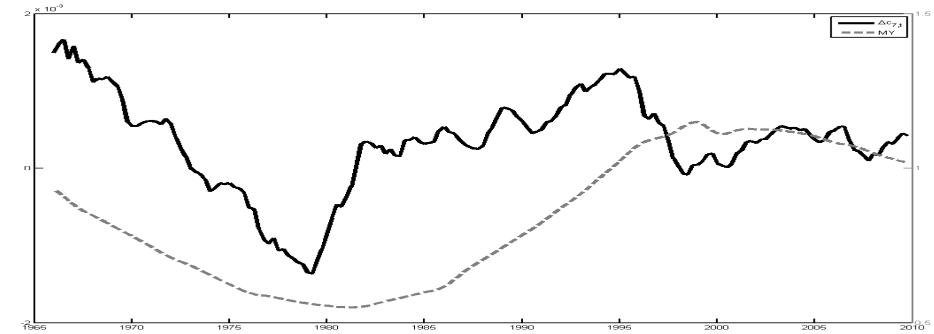


Figure 7: The figure displays the seventh component of consumption growth, $g_{7,t}$ along with a demographic variable, MY_t, the middle-aged to young ratio proposed in Geanakoplos et al. (2004).

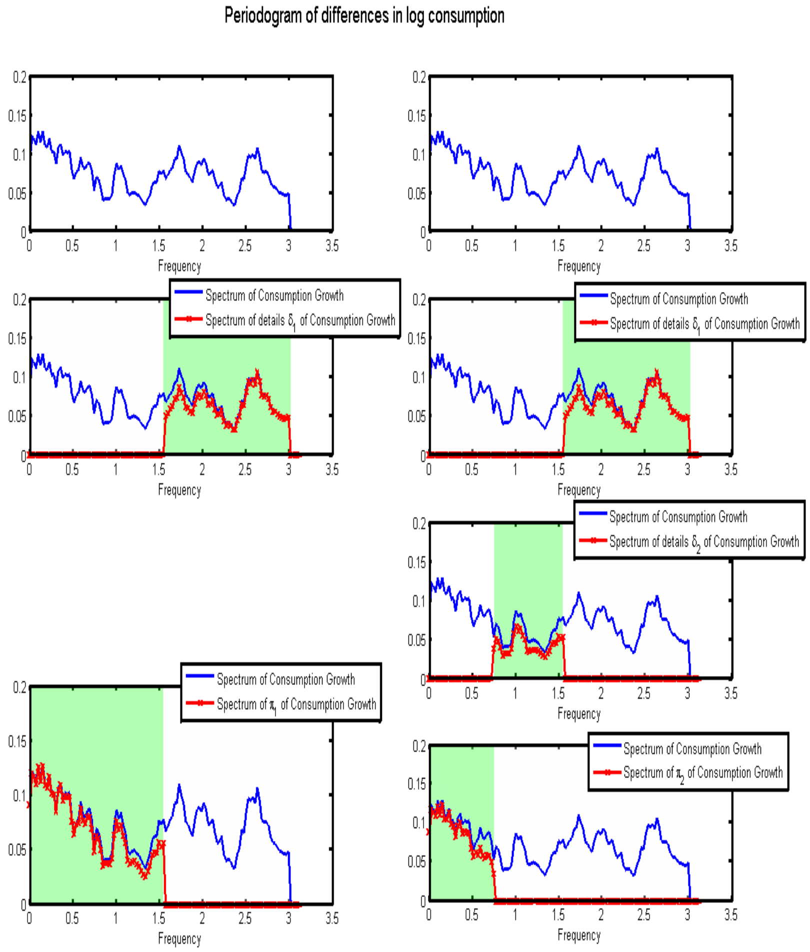


Figure 8: The figure displays the effects of the persistence based decomposition of the consumption time series applied up to level $J = 1$ (left panels) and $J = 2$ (right panels). In particular the top panels displays the smoothed periodogram the consumption process for the data. An equally weighted "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 $\pi_t^{(1)}$.

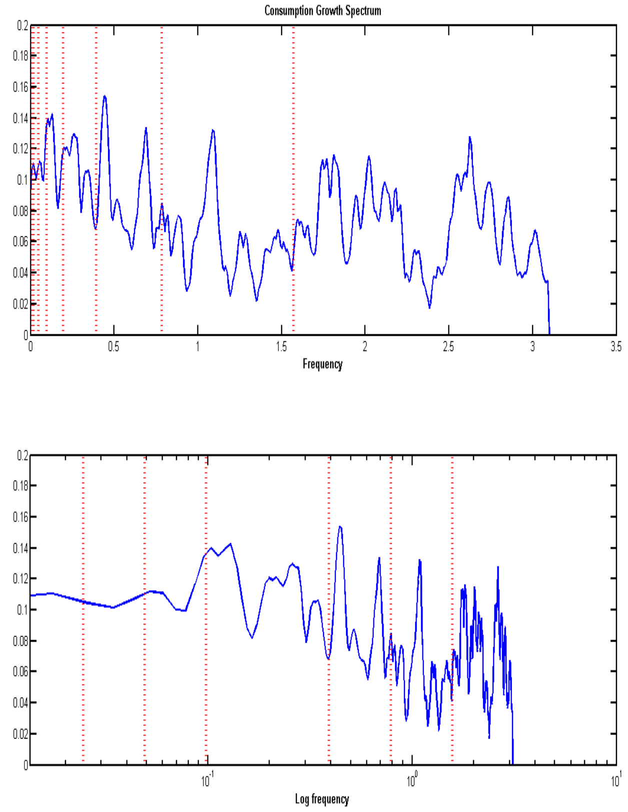


Figure 9: The figure displays the smoothed periodogram the consumption process for the data together with the intervals $\left[\frac{f_{max}}{2^j}\ \frac{f_{max}}{2^{j-1}}\right)$ $j = 1, \ldots, 8$. An equally weighted "nearest neighbor" kernel was used to perform the smoothing, equally weighting the 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.

Component	Quarterly-frequency resolution
$x_{1,t}$	$1-2$ quarters
$x_{2,t}$	$2-4$ quarters
$x_{3,t}$	$1-2$ years
$x_{4,t}$	$2-4$ years
$x_{5,t}$	$4-8$ years
$x_{6,t}$	$8-16$ years
$x_{7,t}$	$16-32$ years
$x_{8,t}$	$32-64$ years
$\pi_t^{(8)}$	>64 years

Table 1: Frequency interpretation of the component $x_{j,t}$ at level of persistence j. We assume the original time series x_t to be observed at quarterly intervals.

	Persistence level j							
$j =$	1	2	3	4	5	6	7	8
$g_{1,t+1}$	-	0.0495	-0.0040	-0.0100	-0.0257	-0.0194	-0.0221	-0.0129
	-	(0.435)	(0.949)	(0.875)	(0.685)	(0.760)	(0.728)	(0.839)
$g_{2,t+1}$	-	-	**0.1310**	-0.0372	-0.0632	-0.0430	-0.0413	-0.0373
	-	-	(0.0380)	(0.558)	(0.319)	(0.498)	(0.515)	(0.557)
$g_{3,t+1}$	-	-	-	**0.1320**	-0.0646	0.0074	-0.0841	-0.0643
	-	-	-	(0.0366)	(0.308)	(0.907)	(0.184)	(0.310)
$g_{4,t+1}$	-	-	-	-	0.0801	0.0523	-0.0906	-0.1223
	-	-	-	-	(0.206)	(0.409)	(0.152)	(0.067)
$g_{5,t+1}$	-	-	-	-	-	0.1016	-0.0699	-0.2025
	-	-	-	-	-	(0.108)	(0.269)	(0.110)
$g_{6,t+1}$	-	-	-	-	-	-	0.0483	-0.2757
	-	-	-	-	-	-	(0.446)	(0.121)
$g_{7,t+1}$	-	-	-	-	-	-	-	-0.3018
	-	-	-	-	-	-	-	(0.089)

Table 2: This table reports the Pearson's correlation coefficients among consumption growth components.

	Component at persistence level j							
$j =$	1	2	3	4	5	6	7	8
$\frac{Var(g_{j,t})}{Var(\sum g_{j,t})}$	0.314	0.183	0.147	0.112	0.066	0.050	0.052	0.075
$\frac{Var(z_{j,t}^m)}{Var(\sum z_{j,t}^m)}$	0.017	0.028	0.045	0.065	0.120	0.249	0.305	0.171

Table 3: Contribution to total unconditional variance of the different details components $g_{j,t}$ of the log consumption growth. Note that $Var(\sum g_{j,t}) = Var(g_t)$ and $Var(\sum z_{j,t}^m) = Var(z_t^m)$

.

$z_t =$	ρ_j^g	$\bar{R}^2$
$g_{1,t+1}$	-0.454	[0.20]
	(-5.40)	
$g_{2,t+1}$	0.350	[0.12]
	(8.12)	
$g_{3,t+1}$	0.746	[0.55]
	(14.93)	
$g_{4,t+1}$	0.885	[0.81]
	(28.12)	
$g_{5,t+1}$	0.966	[0.93]
	(44.91)	
$g_{6,t+1}$	0.985	[0.97]
	(65.20)	
$g_{7,t+1}$	0.992	[0.99]
	(94.06)	
$g_{8,t+1}$	0.998	[0.99]
	(130.05)	

Table 4: This table reports the results of regressions of each of the components of 1-period ahead consumption growth $g_{j,t+1}$ on its own lagged components $g_{j,t}$. 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.

$z_t =$	ρ_j^{pd}	$\bar{R}^2$
$pd_{1,t+1}$	0.286	[0.10]
	(1.92)	
$pd_{2,t+1}$	0.763	[0.43]
	(12.45)	
$pd_{3,t+1}$	0.892	[0.79]
	(20.01)	
$pd_{4,t+1}$	0.921	[0.93]
	(31.51)	
$pd_{5,t+1}$	0.954	[0.99]
	(68.10)	
$pd_{6,t+1}$	0.981	[0.99]
	(64.40)	
$pd_{7,t+1}$	0.992	[1.00]
	(164.54)	
$pd_{8,t+1}$	0.998	[1.00]
	(327.63)	

Table 5: 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 1947Q2-2009Q4.

	Persistence level j							
$j=$	1	2	3	4	5	6	7	8
	0.31	-0.49	-0.73	0.16	-0.17	-0.35	0.28	0.40
pd_t	(0.74)	(-1.75)	(-2.88)	(0.50)	(-0.85)	(-1.93)	(2.56)	(1.51)
	[0.00]	[0.01]	[0.06]	[0.01]	[0.02]	[0.24]	[0.38]	[0.01]

Table 6: This table reports the results of predictive regressions of the components of consumption growth $g_{j,t+2^j}$ 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 1947Q2-2009Q4.

	Persistence level j							
$j=$	1	2	3	4	5	6	7	8
	0.40	-0.25	-0.67	-0.03	-0.11	-0.17	0.27	-0.11
pc_t	(1.22)	(-1.11)	(-2.95)	(-0.09)	(-0.77)	(-1.88)	(3.50)	(0.73)
	[0.00]	[0.00]	[0.06]	[0.00]	[0.01]	[0.07]	[0.50]	[0.02]

Table 7: This table reports the results of predictive regressions of the components of 1-period ahead consumption growth $g_{j,t+2^j}$ 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 1947Q2-2009Q4.

	Persistence level j					
$j=$	1	2	3	4	5	6
	5.95	-2.63	1.10	-2.50	1.58	-4.81
pd_t	(4.19)	(-3.35)	(1.10)	(-2.69)	(2.90)	(-1.51)
	[0.38]	[0.08]	[0.04]	[0.39]	[0.19]	[0.49]

Table 8: 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 annual and spans the period 1930-2009.

$j =$	Persistence level j 1	2	3	4	5	6	7	8
	0.39	-0.58	-0.78	0.17	-0.17	-0.37	0.29	0.31
pd_t	(0.86)	(-1.91)	(-2.96)	(0.51)	(-0.87)	(-1.98)	(2.55)	(1.17)
	[0.00]	[0.02]	[0.06]	[0.01]	[0.02]	[0.25]	[0.38]	[0.05]

Table 9: This table reports the results of predictive regressions of the components of dividend growth $gd_{j,t+2^j}$ 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 1947Q2-2009Q4.

$j =$	Persistence level j 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 10: 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 1947Q2-2008Q4. The components are extracted using the default filter recommended in Christiano and Fitzgerald (1999).

Asset	$r_{f,j,t+1} = \alpha_f + \frac{1}{\psi} g_{t+1,j}$ Sample	$\hat{\psi}$
$r_{f,t+1}$	1947Q2-2009Q4	4.762
		(6.505)
$r_{f,t+1}$	1978Q1-2009Q4	4.594
		(3.037)
$r_{f,t+1}$	1947Q2-2009Q4	5.076
		(2.707)

Table 11: This table displays the EIS estimates using the risk free rate. The first row reports two-stage least squares estimates. The instruments are the components of consumption growth and of the returns for the asset. The second and third rows report the estimate based on not-redundant persistence-based decomposition as suggested in Fadili and Bullmore (2002) based on 128 and 256 data points respectively.

Scale $j =$	Half-life (Years)	Q_j (1.0e-005)	Risk Exposure (1.0e-003)	Risk Price	Risk Premium (%)
1	0.08	0.31	1.072	4.67	0.50
2	0.44	0.18	0.712	12.12	0.86
3	1.52	0.15	0.592	32.33	1.91
4	3.63	0.12	0.652	96.03	6.29
5	4.57	0.07	0.288	168.69	4.86
6	12.5	0.05	0.140	181.71	2.51
7	18.77	0.05	0.068	183.28	1.25
8	33.27	0.07	0.016	183.84	0.26

Table 12: This table reports equity premium (in %) $E_t[r_{m,t+1} - r_{f,t}]$ decomposed by level of persistence. We set $\psi = 1.5$ and $\gamma = 5$.

A The Persistence Based Decomposition

A.1 The Backward and Forward Decomposition

In this section we show that the identities in equations (6) and (7) hold true. It is useful to carry out the calculations in the simple case where $J = 2$ which is enough to get the point and does not mess up the algebra. Generalization are straightforward and can be found in Daubechies (1992), Daubechies (1990), Mallat (1989a) and Mallat (1989b).

When $J = 2$ the identity (6) becomes

$$x_t = x_{2,t} + x_{1,t} + \pi_t^{(2)}$$

To construct the components $x_{j,t}$ and $\pi_t^{(2)}$ we use equations (4) and (5). Using these equations one obtains

$$\begin{aligned}
\pi_t^{(2)} &= \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4} \\
x_{2,t} &= \frac{x_t + x_{t-1} - x_{t-2} - x_{t-3}}{4} \\
x_{1,t} &= \frac{x_t - x_{t-1}}{2}
\end{aligned}$$

It is trivial to verify that (6) holds. In order to show that relation (7) holds it is useful to introduce a linear transformation which maps a block of 2^J observations, i.e. $\mathbf{x} = \{x_{t-i}\}_{i=0}^{2^J-1}$ into its components $\{x_{j,t}\}_{j=1}^{J}$ with persistence j and the permanent component $\pi_t^{(J)}$.[38] This linear mapping can be explicitly represented using a matrix of size 2^J, $\mathcal{T}^{(J)}$. While we refer once again to Daubechies (1992), Daubechies (1990), Mallat (1989a) and Mallat (1989b) 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 components with different levels of persistence j is given by:

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

In order to verify that the above matrix is the transformation we are after, we first let $\underline{X}_t^{(2)}$ be the vector that collects the elements of our time series from time $t - 2^2 + 1$ to time t

[38]Indeed the linear transformation maps the original windowed time series with 2^J elements into exactly 2^J components. This is because at level of persistence j we will have exactly $2^J/2^j$ independent fluctuations.

(hence backward from time t), i.e.

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

and then we postmultiply it by the matrix $\mathcal{T}^{(2)}$

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

to obtain

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

We have therefore verified that the vector $\underline{\widetilde{X}}_t^{(2)}$ that stacks the sequence of components $x_{2,t}, x_{1,t-2}, x_{1,t}$ below the persistent component $\left\{\pi_t^{(2)}\right\}$ can be obtained as a linear transformation of our original time series. Importantly we remark that our transformation matrix $\mathcal{T}^{(2)}$ does not depend on the time t from which we start collecting (backward) the time series observations.

Next we show that the forward persistence decomposition in equation (7) which is reported here for reader convenience holds:

$$x_{t+1} = \sum_{j=1}^{J} x_{j,t+2^j} + \pi_{t+2^J}^{(J)}$$

Once again we give a simple example for the case $J = 2$. In order to obtain the components $x_{1,t+2}, x_{2,t+4}, \pi_{t+4}^{(2)}$ we can still use the transformation matrix given by equation (A.1) but now we apply it to the vector that collects the elements of our time series from time $t+1$ to time $t+2^J$ (hence forward from time $t+1$), i.e.

$$\underline{X}_{t+1}^{(2)} = \begin{bmatrix} x_{t+4} \\ x_{t+3} \\ x_{t+2} \\ x_{t+1} \end{bmatrix}$$

By doing so we obtain

$$\pi^{(2)}_{t+4} = \frac{x_{t+4} + x_{t+3} + x_{t+2} + x_{t+1}}{4}$$
$$x_{2,t+4} = \frac{-x_{t+4} - x_{t+3} + x_{t+2} + x_{t+1}}{4}$$
$$x_{1,t+2} = \frac{-x_{t+2} + x_{t+1}}{2}$$
$$x_{1,t+4} = \frac{-x_{t+4} + x_{t+3}}{2}$$

Finally we can check that for $J = 2$

$$\begin{aligned} x_{t+1} &= \sum_{j=1}^{J} x_{j,t+2^j} + \pi^{(J)}_{t+2^J} \\ &= x_{2,t+4} + x_{1,t+2} + \pi^{(2)}_{t+4} \end{aligned}$$

holds true.

A.2 A Frequency Interpretation of the Persistence-based Decomposition

The filtering procedure described in Section 22.2 and the persistence properties of the series $\pi^{(j)}_t$ and $x_{j,t}$ can be usefully visualized in the frequency domain in terms of their Fourier spectra.[39] As an example we apply equations (4) and (5) for the case $J = 1$ and $J = 2$ to the consumption growth time series and we report the results in the left and right columns of Figure 8 respectively.

The top subplot of Figure 8 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 $\boldsymbol{\pi}^{(1)}_t = \{\pi^{(1)}_t\}_{t\in\mathbb{Z}}$. We clearly see that the effect of the simple 2-period moving average is to halve the spectrum and to keep the lowest part. Nevertheless the high frequency part of the spectrum is recovered by the component $\boldsymbol{x}_{1,t} = \{x_{1,t}\}_{t\in\mathbb{Z}}$. This can be seen in the mid left panel where the shadowed region represent the spectrum of $\mathbf{x}_{1,t}$. In some sense we are reassured that, for $J = 1$ we recover a simple permanent-transitory decomposition.

The right column of Figure 8 shows the case $J = 2$ where we filter out the first two components $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$ of the the aggregate consumption growth. From (5) we know that $\boldsymbol{x}_{2,t}$ is obtained as the difference of $\boldsymbol{\pi}^{(2)}_t$ and $\boldsymbol{\pi}^{(1)}_t$ reported in the bottom right and

[39]Indeed in the frequency representation, a 2^j-period moving average operator works as a low band pass filter which removes all those components whose frequency is higher than 2^j.

left panels, respectively. The third left panel displays exactly this operation and confirms our previous intuition that by taking the difference between $\boldsymbol{\pi}_t^{(2)}$ and $\boldsymbol{\pi}_t^{(2)}$, we are able to identify the fluctuations of the original time series $\boldsymbol{x}$ that lies in the well defined frequency range $[1/4,\ 1/2)$.[40] The bottom panels of Figure 8 show that the Fourier spectrum of the time series $\boldsymbol{\pi}_t^{(2)}$ differs from that of $\boldsymbol{\pi}_t^{(1)}$ because the dyadic averaging operation gets rid of the components at frequencies larger than $1/2^2$ while the low frequency components are essentially left unaffected. Therefore we conclude that increasing the window of values over which the average is made is equivalent to focus one's attention on lower and lower frequencies.

B The Long-run Risk Model with Persistence Heterogeneity

In this Section we show the steps to obtain the values of the financial ratios coefficients $A_{0,j}, A_j, A_{0,j}^m, A_j^m$ in terms of the parameters of the model. We then compute the equity premia on both the consumption claim asset and the market return. Finally we derive the risk-free rate.

B.1 The Financial Ratios

We solve first for the price-consumption coefficients $A_{0,j}, A_j$ and then for the price-dividend ones $A_{0,j}^m, A_j^m$. To obtain the values of the coefficients $A_{0,j}, A_j$ we exploit the Euler condition

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

which is derived from (11) for the special case where the asset being priced is the aggregate consumption claim, i.e. $r_{i,t+1} = r_{a,t+1}$. We then express the log consumption growth g_{t+1} and the return $r_{a,t+1}$ in terms of the factors $\{x_{j,t}\}_j$ and of the innovations $\{e_{j,t+2^j}^g\}_j$ and $\{\varepsilon_{j,t+2^j}\}_j$. To do so we plug first the Campbell and Shiller (1988) approximation for log returns, see equation (13), into the above expression to obtain:

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

[40] Each component has in fact a corresponding Fourier spectrum localized in the finite interval of frequencies $\left[\frac{f_{max}}{2^j}\ \frac{f_{max}}{2^{j-1}}\right)$ where f_{max} is the maximum frequency of observations, and in our case quarterly.

By the backward decomposition (6) applied to the (demeaned) price-consumption ratio at time t and by the forward decomposition (7) applied to the (demeaned) consumption growth and price-consumption processes at time $t+1$ we have:

$$z_t^a = \sum_{j=1}^{J} z_{j,t}^a \tag{B.1}$$

$$z_{t+1}^a = \sum_{j=1}^{J} z_{j,t+2^j}^a \tag{B.2}$$

$$g_{t+1} = \sum_{j=1}^{J} g_{j,t+2^j} \tag{B.3}$$

Plugging the above expressions into the Euler condition yields:

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

Finally using the dynamics for the components of log consumption growth given in equation (8) together with our guess for the components of price-consumption ratio solution given in equation (14), rearranging terms and using the log normal properties of the shocks we obtain:

$$\begin{aligned}
&E_t\Bigg[\exp\quad\Bigg(\theta\log\beta + \theta\left(1-\frac{1}{\psi}\right)\left(\sum_{j=1}^{J} g_{j,t+2^j}\right) + \\
&\quad\theta\left(\kappa_0 + \kappa_1\left(\sum_{j=1}^{J} z_{j,t+2^j}\right) - \left(\sum_{j=1}^{J} z_{j,t}\right)\right)\quad\Bigg)\quad\Bigg] \\
&= E_t\Bigg[\exp\quad\Bigg(\theta\log\beta + \theta\left(1-\frac{1}{\psi}\right)\left(\sum_{j=1}^{J} x_{j,t} + \sum_{j=1}^{J} e^g_{j,t+2^j}\right) + \\
&\quad\theta\left(\kappa_0 + \kappa_1\left(\sum_{j=1}^{J} A_{0,j} + \sum_{j=1}^{J} A_j x_{j,t+2^j}\right) - \left(\sum_{j=1}^{J} A_{0,j} + \sum_{j=1}^{J} A_j x_{j,t}\right)\right)\quad\Bigg)\quad\Bigg]
\end{aligned}$$

$$= E_t\left[\exp\left(\theta(\log\beta + \kappa_0 + (\kappa_1 - 1)\sum_{j=1}^{J} A_{0,j}) + \dots\right.\right.$$

$$\left.\left.\theta\left(1 - \frac{1}{\psi}\right)\left(\sum_{j=1}^{J} x_{j,t} + \sum_{j=1}^{J} e^{g}_{j,t+2^j}\right) + \theta\left(\kappa_1\sum_{j=1}^{J} A_j x_{j,t+2^j} - \sum_{j=1}^{J} A_j x_{j,t}\right)\right)\right]$$

$$= E_t\left[\exp\left(\theta(\log\beta + \kappa_0 + (\kappa_1 - 1)\sum_{j=1}^{J} A_{0,j}) + \dots\right.\right.$$

$$\left.\left.\theta\left(1 - \frac{1}{\psi}\right)\left(\sum_{j=1}^{J} x_{j,t} + \sum_{j=1}^{J} e^{g}_{j,t+2^j}\right) + \theta\left(\kappa_1\sum_{j=1}^{J} A_j \underbrace{(e_j M\tilde{X}_t + e_j\varepsilon_{t+2^J})}_{x_{j,t+2^j}} - \sum_{j=1}^{J} A_j x_{j,t}\right)\right)\right] = 1$$

where we defined $\tilde{X}_t \equiv [x_{1,t}, \dots, x_{J,t}]^\intercal$. Collecting terms in $\tilde{X}_t$ yields eventually a system of equations

$$e_j\left(\left(1 - \frac{1}{\psi}\right) + A_j(\kappa_1 M - \mathbb{I}_J)\right) = 0$$

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

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

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

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

To derive the expression for A^m_j we exploit once again the Euler condition

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

where now the asset being priced is the market return $r_{m,t+1}$. Following the same steps as above, and additionally using the Campbell and Shiller (1988) log-linear approximation for

$r_{m,t+1}$, see equation (13) we can then rewrite the Euler equation as:

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

Let's focus on the term

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

This can be written (neglecting error terms that are going to be captured by the constant using the law of log normal distribution and neglecting constant terms) as follows

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

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

$$
\begin{aligned}
&= \theta \left(1 - \frac{1}{\psi}\right) \left(\sum_{j=1}^{J} x_{j,t} \right) - (\theta - 1) \left(1 - \frac{1}{\psi}\right) \underline{1} \tilde{X}_t \\
&= \left(1 - \frac{1}{\psi}\right) \underline{1} \tilde{X}_t
\end{aligned}
$$

Plugging into the Euler equation the above simplifying expression, using the dynamics for the components of the log consumption growth given in formula (8) and rearranging terms we have:

$$
\begin{aligned}
&E_t\Bigg[\exp\Bigg(\theta\log\beta+\left(1-\frac{1}{\psi}\right)\underline{1}\tilde{X}_t-\left(\sum_{j=1}^{J}g_{j,t+2^j}\right)\\
&\quad+\kappa_{0,m}+\kappa_{1,m}\left(\sum_{j=1}^{J}z^m_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}z^m_{j,t}\right)+\left(\sum_{j=1}^{J}gd_{j,t+2^j}\right)\Bigg)\Bigg]\\
&=E_t\Bigg[\exp\Bigg(\theta\log\beta+\left(1-\frac{1}{\psi}\right)\underline{1}\tilde{X}_t-\left(\sum_{j=1}^{J}x_{j,t}+\sum_{j=1}^{J}e^g_{j,t+2^j}\right)\\
&\quad+\kappa_{0,m}+\kappa_{1,m}\left(\sum_{j=1}^{J}z^m_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}z^m_{j,t}\right)+\left(\sum_{j=1}^{J}gd_{j,t+2^j}\right)\Bigg)\Bigg]\\
&=E_t\Bigg[\exp\Bigg(\theta\log\beta-\left(\frac{1}{\psi}\right)\underline{1}\tilde{X}_t-\sum_{j=1}^{J}e^g_{j,t+2^j})\\
&\quad+\kappa_{0,m}+\kappa_{1,m}\left(\sum_{j=1}^{J}z^m_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}z^m_{j,t}\right)+\left(\sum_{j=1}^{J}gd_{j,t+2^j}\right)\Bigg)\Bigg]
\end{aligned}
$$

Finally, analogously to what we have done for the return on the consumption claim, using the dynamics for the components of the log dividend growth given in equation (9) together with our guess for the components of log price-dividend ratio given in equation (14), rearranging terms and using the log normal properties of the shocks we obtain:

$$
\begin{aligned}
&E_t\Bigg[\exp\Bigg(\theta\log\beta+\kappa_{0,m}+(\kappa_{1,m}-1)\sum_{j=1}^{J}A^m_{0,j}-\left(\frac{1}{\psi}\right)\underline{1}\tilde{X}_t-\sum_{j=1}^{J}e^g_{j,t+2^j})+\\
&\left(\kappa_{1,m}\left(\sum_{j=1}^{J}A_j x_{j,t+2^j}\right)-\left(\sum_{j=1}^{J}A^m_j x_{j,t}\right)\right)+\left(\sum_{j=1}^{J}\phi_j x_{j,t}+\sum_{j=1}^{J}e^d_{j,t+2^j}\right)\Bigg)\Bigg]\\
&=E_t\Bigg[\exp\Bigg(\theta\log\beta+\kappa_{0,m}+(\kappa_{1,m}-1)A^m_{0,-1}+(\kappa_{1,m}-1)\sum_{j=1}^{J}A^m_{0,j}-\left(\frac{1}{\psi}\right)\underline{1}\tilde{X}_t-\sum_{j=1}^{J}e^g_{j,t+2^j})+\\
&\left(\kappa_{1,m}\left(\sum_{j=1}^{J}A^m_j\underbrace{(e_jM\tilde{X}_t+e_j\boldsymbol{\varepsilon}_{t+2^J})}_{x_{j,t+2^j}}\right)\right)-\left(\sum_{j=1}^{J}A^m_j x_{j,t}\right)+\left(\sum_{j=1}^{J}\phi_j x_{j,t}+\sum_{j=1}^{J}e^d_{j,t+2^j}\right)\Bigg)\Bigg]=1
\end{aligned}
$$

Define

$$
\begin{aligned}
\underline{A_m}&\equiv[A^m_1,\ldots,A^m_J]^\intercal\\
\underline{\phi}&\equiv[\phi_1,\ldots,\phi_J]^\intercal
\end{aligned}
\tag{B.4}
$$

In vector notation we have

$$\underline{A}^m(\kappa_{1,m}M - \mathbb{I}_J) = \frac{1}{\psi}\underline{1} - \underline{\phi}$$
$$\underline{A}^m = \left(\underline{\phi} - \frac{1}{\psi}\underline{1}\right)(\mathbb{I}_J - \kappa_{1,m}M)^{-1}$$

B.2 The Risk Premia

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

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

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

The equilibrium return innovations can be found by plugging the expressions (B.1), (B.2) and (B.3) into the Campbell and Shiller (1988) approximation for log returns, see equation (13) to obtain

$$\begin{aligned} r_{a,t+1} - E_t[r_{a,t+1}] &= \left(\sum_{j=1}^{J} g_{j,t+2^j}\right) + \kappa_0 + \kappa_1\left(\sum_{j=1}^{J} z_{j,t+2^j}\right) - \left(\sum_{j=1}^{J} z_{j,t}\right) - E_t[r_{a,t+1}] \\ &= \sum_{j=1}^{J} e^g_{j,t+2^j} + \kappa_1\left(\sum_{j=1}^{J} A_j\varepsilon_{j,t+2^j}\right) \\ &= e^g_{t+1} + \kappa_1\underline{A}\boldsymbol{\varepsilon}_{t+1} \end{aligned} \tag{B.5}$$

where in the second line we use our solution for the price-consumption ratio and in the third line we define

$$\boldsymbol{\varepsilon}^{\intercal}_{t+1} \equiv [\varepsilon_{1,t+2^1}, \ldots, \varepsilon_{J,t+2^J}]$$
$$e^g_{t+1} \equiv \sum_j e^g_{j,t+2^j}$$

Analogous steps yield the following expression for the market return innovations

$$r_{m,t+1} - E_t[r_{m,t+1}] = e^d_{t+1} + \kappa_{1,m}\underline{A_m} \cdot \boldsymbol{\varepsilon}_{t+1} \tag{B.6}$$

To find the innovations in the stochastic discount factor, we plug the expressions (B.1), (B.2) and (B.3), together with the dynamics for the components of log consumption growth given in equation (8) and our guess for the components of price-consumption ratio solution given in equation (14) into equation (12) to obtain:

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

Finally using the dynamics for our latent factors (10) we obtain

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

which implies

$$
\begin{aligned}
m_{t+1} - E_t[m_{t+1}] &= -\left(1 - \theta + \frac{\theta}{\psi}\right) \sum_{j=1}^{J} e^g_{j,t+2^j} + (\theta-1)\kappa_1 \left(\sum_{j=1}^{J} A_j \varepsilon_{j,t+2^j}\right) \\
&= -\lambda_\eta \sum_{j=1}^{J} e^g_{j,t+2^j} - \sum_{j=1}^{J} \lambda_j \varepsilon_{j,t+2^j} \\
&= -\lambda_\eta e^g_{t+1} - \underline{\lambda}_n \boldsymbol{\varepsilon}_{t+1} \qquad \text{(B.7)}
\end{aligned}
$$

where

$$
\begin{aligned}
\lambda_\eta &\equiv \left(\frac{\theta}{\psi} - \theta + 1\right) \\
\underline{\lambda}_n &\equiv \kappa_1 (1-\theta) \underline{A}
\end{aligned}
$$

Using the formula for the return on aggregate wealth (B.5) and the innovation in the SDF (B.7) we obtain the risk premium for the consumption claim asset,

$$E_t[r_{a,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{a,t}} = \lambda_\eta \sigma^2_{\eta,t} + \kappa_1 \underline{\lambda}_n \mathbf{Q} \underline{A}'$$

where

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

Similarly to what we have just done, using the formula for the innovations in the market return (B.6) and in the SDF (B.7) the premium to the market return becomes:

$$E_t[r_{m,t+1} - r_{f,t}] + 0.5\sigma^2_{r_{m,t}} = \kappa_{1,m} \underline{\lambda}_\varepsilon \mathbf{Q} \underline{A_m}$$

B.3 The Risk-Free Rate and The Intertemporal Elasticity of Substitution

To obtain our expression for the risk-free rate we start by plugging the log short-term real interest rate $r_{f,t+1}$ for r^i_{t+1} into the Euler equation (11). Then by applying the forward decomposition (7) to the (demeaned) consumption growth and to the log returns processes at time $t+1$ we observe that the risk-free rate between t and $t+1$, $r_{f,t+1}$ satisfies the following condition:

$$E_t\left[\exp\left(\theta\log\delta - \left(\frac{\theta}{\psi}\right)\sum_{j=1}^{J} g_{j,t+2^j} + (\theta-1)\sum_{j=1}^{J} r_{a,j,t+2^j}\right)\right] = \exp(-r_{f,t+1})$$

where once again $r_{a,t+1}$ is the return on the asset that pays consumption as dividend. Taking logs on both sides and using the log normal properties of the shocks we can rewrite it as follows

$$\begin{aligned} r_{f,t+1} = &-\theta\log\beta + \frac{\theta}{\psi}E_t\left[\sum_{j=1}^{h} g_{j,t+2^j}\right] + (1-\theta)E_t\left[\sum_{j=1}^{J} r_{a,j,t+2^j}\right] \\ &-\frac{1}{2}var_t\left[\frac{\theta}{\psi}\sum_{j=1}^{J} g_{j,t+2^j} + (1-\theta)\sum_{j=1}^{J} r_{a,j,t+2^j}\right] \end{aligned}$$

$$
\begin{aligned}
&= -\log\beta + \frac{1}{\psi}E_t\left[\sum_{j=1}^{J} g_{j,t+2^j}\right] + \frac{(1-\theta)}{\theta}E_t\left[\sum_{j=1}^{h} r_{a,j,t+2^j} - r_f\right] \\
&- \frac{1}{2\theta}var_t\left[\frac{\theta}{\psi}\sum_{j=1}^{h} g_{j,t+2^j} + (1-\theta)\sum_{j=1}^{h} r_{a,j,t+2^j}\right]
\end{aligned}
\tag{B.8}
$$

where in the last line we subtract $(1-\theta)r_{f,t}$ from both sides and divide by θ, where it is assumed that $\theta \neq 0$. Further to solve the above expression, note that

$$
var_t\left[\frac{\theta}{\psi}\sum_{j=1}^{J} g_{j,t+2^j} + (1-\theta)\sum_{j=1}^{J} r_{a,j,t+2^j}\right] = var_t(m_{t+1})
$$

Now we show that in our homoskedastic version of the long-run risk with persistence heterogeneity the variance of the stochastic discount factor $var_t(m_{t+1})$ is constant (not function of time). Indeed recall from (B.7) that we have

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

from which we can compute the variance as follows

$$
var_t(m_{t+1}) = \lambda_\eta^2\sigma^2 + \kappa_1^2(1-\theta)^2\underline{A}\mathbf{Q}\underline{A}'
$$

Therefore using the dynamics for log consumption growth, namely (8) and reported here for reader's convenience

$$
\begin{aligned}
g_{t+1} &= \sum_{j=1}^{J} g_{j,t+2^j} \\
g_{j,t+2^j} &= x_{j,t} + e^g_{j,t+2^j} \\
e^g_{j,t+2^j} &\sim N\left(0, \sigma^2_{g,j}\right)
\end{aligned}
$$

and taking conditional expectation we rewrite (B.8) as follows:

$$
r_{f,t+1} = \alpha_f + \frac{1}{\psi}\sum_j x_{j,t}
$$

where α_f is meant to capture the unconditional mean of the risk-free rate and ψ is the IES.

References

Alvarez, Fernando and Urban J. Jermann, "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, November 2005, *73* (6), 1977–2016.

Bandi, Federico M. and Benot Perron, "Long-run risk-return trade-offs," *Journal of Econometrics*, April 2008, *143* (2), 349–374.

Bansal, Ravi and Amir Yaron, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 08 2004, *59* (4), 1481–1509.

___ **and Ivan Shaliastovich**, "Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets," *SSRN eLibrary*, 2010.

___ **, Robert Dittmar, and Dana Kiku**, "Cointegration and Consumption Risks in Asset Returns," *Review of Financial Studies*, 2007, *22* (3), 1343–1375.

___ **, Robert F. Dittmar, and Christian T. Lundblad**, "Consumption, Dividends, and the Cross Section of Equity Returns," *Journal of Finance*, 08 2005, *60* (4), 1639–1672.

Beeler, Jason and John Y. Campbell, "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," NBER Working Papers 14788, National Bureau of Economic Research, Inc March 2009.

Borovicka, Jaroslav, Lars Peter Hansen, Mark Hendricks, and Jose A. Scheinkman, "Risk Price Dynamics," NBER Working Papers 15506, National Bureau of Economic Research, Inc November 2009.

Calvet, Laurent E. and Adlai J. Fisher, "Multifrequency news and stock returns," *Journal of Financial Economics*, October 2007, *86* (1), 178–212.

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

___ **and N. Gregory Mankiw**, "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," NBER Working Papers, National Bureau of Economic Research, Inc May 1990.

___ **and Robert J. Shiller**, "Valuation Ratios and the Long-run Stock Market Outlook: An Update," Cowles Foundation Discussion Papers 1295, Cowles Foundation, Yale University 2001.

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

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

Constantinides, George M. and Anisha Ghosh, "Asset Pricing Tests with Long Run Risks in Consumption Growth," NBER Working Papers 14543, National Bureau of Economic Research, Inc December 2008.

Croce, Mariano M., "Long-Run Productivity Risk: A New Hope for Production-Based Asset Pricing?," *SSRN eLibrary*, 2010.

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

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

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

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

Fadili, M.J and E.T Bullmore, "Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors," *NeuroImage, Volume 15, Number 1*, 2002.

Favero, Carlo A., Arie E. Gozluklu, and Andrea Tamoni, "Long-Run Factors and Fluctuations in Dividend/Price," *JFQA Forthcoming*, 2010.

Ferson, Wayne E., Suresh Nallareddy, and Biqin Xie, "The Out-of-sample Performance of Long-Run Risk Models," NBER Working Papers, National Bureau of Economic Research, Inc 2010.

Garleanu, Nicolae B., Stavros Panageas, and Jianfeng Yu, "Technological Growth and Asset Pricing," NBER Working Papers 15340, National Bureau of Economic Research, Inc September 2009.

Geanakoplos, John, Michael Magill, and Martine Quinzii, "Demography and the Long-Run Predictability of the Stock Market," *Brookings Papers on Economic Activity*, 2004, *35* (2004-1), 241–326.

Gencay, Ramazan and Nikola Gradojevic, "Errors in Variables Estimation with No Instruments," Working Paper Series wp30 09, Rimini Centre for Economic Analysis January 2009.

Hall, Robert E, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, April 1988, *96* (2), 339–57.

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

Hansen, Lars Peter and Jose A Sheinkman, "Long-term Risk: An Operator Approach," *Econometrica*, November 2009, *77* (1), 177–234.

——, **John C. Heaton, and Nan Li**, "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 04 2008, *116* (2), 260–302.

Hsu, Po-Hsuan and Dayong Huang, "Technology prospects and the cross-section of stock returns," *Journal of Empirical Finance*, January 2010, *17* (1), 39–53.

Jagannathan, Ravi and Yong Wang, "Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns," *Journal of Finance*, 08 2007, *62* (4), 1623–1661.

Kaltenbrunner, Georg and Lars A. Lochstoer, "Long-Run Risk Through Consumption Smoothing," *Review of Financial Studies*, 2010.

Koijen, Ralph S.J., Hanno Lustig, and Stijn Van Nieuwerburgh, "The Cross-Section and Time-Series of Stock and Bond Returns," Working Paper 15688, National Bureau of Economic Research January 2010.

——, ——, ——, **and Adrien Verdelhan**, "Long-Run Risk, the Wealth-Consumption Ratio, and the Temporal Pricing of Risk," *SSRN eLibrary*, 2010.

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

Lettau, Martin and Stijn Van Nieuwerburgh, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, July 2008, *21* (4), 1607–1652.

Lippi, Marco and Paolo Zaffaroni, "Aggregation of Simple Linear Dynamics: Exact Asymptotic Results," STICERD - Econometrics Paper Series /1998/350, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE April 1998.

Lustig, Hanno and Stijn Van-Nieuwerburgh, "Can Housing Collateral Explain Long-Run Swings in Asset Returns?," NBER Working Papers 12766, National Bureau of Economic Research, Inc December 2006.

Mallat, Stephane G., "Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R)," September 1989, *315* (1).

——, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, *11*, 674–693.

Moller, Stig V. and Jesper Rangvid, "The fourth-quarter consumption growth rate," Technical Report 2010.

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

Paseka, Alex and George Theocharides, "Predictability in Consumption Growth and Equity Returns: A Bayesian Investigation," *Financial Review*, 2010, *45* (1), 167–203.

Thornton, Michael, "Incomplete Information, Long Memory and Aggregation in Consumption under the Permanent Income Hypothesis," Working Paper 2008.

Torous, Walter, Rossen Valkanov, and Shu Yan, "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business*, October 2004, *77* (4), 937–966.

Vissing-Jorgensen, Annette, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, August 2002, *110* (4), 825–853.

Part III

Demographics and the Term Structure of Stock Market Risk

1 Introduction

This paper examines the consequences for the term structure of stock market risk, defined as the per period conditional variance of cumulative returns, given the investor's information set, of the significance of demographics in capturing the time varying mean of the dividend-price ratio and in predicting stock market returns. To our knowledge demographic variables has never been considered in the ongoing debate on the slope of the term structure of stock market risk. This is a surprising omission as demographic data, by their nature, have a predictive power for stock market returns that it increases with the horizon.

The term structure of stock market risk is of special relevance to strategic asset allocation when risk is measured on the basis of a predictive model for returns at different horizons. There are two basic elements determining the shape of risk as the horizon at which returns are defined gets larger. On the one hand, the longer the horizon the more distant is the future to predict and therefore the more uncertainty should matter. On the other hand, if the "information" in fundamentals for stock market returns emerges only in the long-run, while the short-run is dominated by "noise", then the longer is the horizon the less is the "noise" in the variable to be predicted.[1] Demographic trends are by the nature very smooth, they do not contribute to the short-run noise but they are a natural candidate to capture the information that emerges in the long-run.

The empirical evidence on stock market return predictability associated with demographic information is already fairly extensive in the empirical literature (see Geanakoplos, Magill and Quinzii (2004), Ang and Maddaloni (2005), Favero, Gozluklu and Tamoni(2010)), but the term structure of stock market risk is not simply the other side of the coin of predictability of returns.

The strategic asset allocation literature (e.g. Campbell and Viceira (2002), Campbell and Viceira (2005) and Schotman, Tschernig and Budek (2008) to name just a few), is based on a stationary Vector Autoregressive (VAR) specification for predictors and returns that captures time-variation in the investment opportunity set, and constitutes the input into the optimal asset allocation decision of a long-horizon investor. In practice, long-horizon returns are predicted via multi-step ahead projections of a VAR model in which the dividend-price ratio is used as a predictor of returns and no demographic variables are included. Importantly Campbell and Viceira (2005) (henceforth CV) and more recently Schotman et al. (2008) have shown that whereas absence of predictability entails a flat term structure of risk, predictability per se does not lead to a downward sloping term structure of risk. In

[1] The use of the terms "noise" and "information" has been inspired by our reading of Chapter 3 of Taleb (2001).

fact risk at the different horizons is determined by three components: i.i.d uncertainty, mean reversion, uncertainty about future predictors. Without predictability the entire term structure is flat at the level determined by i.i.d uncertainty. This is the classical situation where portfolio choice is independent of the investment horizon. For a negative slope of the term structure of risk to emerge one needs the mean reversion component to compensate the increase with the horizon in the uncertainty about future predictors. It is an empirical issue, however, which term is more important at economically relevant horizons. Overall, the slope of the term structure of risk depends on the significance of predictors in explaining returns, on the contemporaneous correlation between the innovations in the equations for predictors and returns, on the variance of returns and predictors, and on the persistence of the predictors. High negative contemporaneous correlation between innovations in the returns and in the predictors, paired with significant predictability and low persistence in the predictors generate a steeply declining term structure of stock market risk. CV find that empirically a very high negative correlations between innovations compensate the high persistence in the predictors (the dividend-price ratio) to generate a negatively sloping term structure of risk. Pastor and Stambaugh (2008) and Pastor and Stambaugh (2009) (henceforth PS) extend the framework of CV to consider two additional sources of risk: one reflecting uncertainty around the mean of the process generating returns and one reflecting parameters' uncertainty. Within this extended framework, PS do find a positive slope for the term structure of stock market risk, despite the evidence of predictability.

The contribution of our paper to this debate is twofold.

First, following Geanakoplos et al. (2004) and Favero, Gozluklu and Tamoni (2010), we augment the set of predictors of stock market returns by considering the dividend-price ratio together with a demographic variable, MY, the ratio of middle-aged to young population, that captures the slow moving but time-varying mean of the US dividend-price ratio. The quick mean reversion of the dividend-price ratio towards a time varying mean determined by demographics has important implications for the predictability of returns and the slope of the term structure of stock market risk.

Second, we argue that the importance of such variable can be best understood in the framework of the dynamic dividend growth model (Campbell and Shiller 1988) with a time varying linearization point driven by demographics, and that in this context direct regression rather than multi-step ahead forecasting is the natural way to predict long-horizon returns.

The best way to introduce our work is to refer the reader to Figures 1(a)-fig:Dem-c. Figure 1(a) illustrates, over about one century of US data, the relationship between 1-year real stock market returns and MY. Figure 1(b) relates again demographics and stock market fluctuations, but 20-year real annualized returns are now considered instead of 1-year return.

The comovement between demographics and stock market returns is negligible for annual returns and remarkable for 20-year returns. Figure 1(c) rationalizes the positive relation between long-horizons stock market returns and the demographic variable in terms of the negative relation between the dividend-price ratio and MY. If MY captures the time varying-mean of the dividend price ratio, being negatively related to it[2], and stock market returns react to correct deviations of the dividend-price from its equilibrium level, then future stock market returns are positively related to MY.

When demographic trends are used to model the slow moving fluctuations in the dividend-price ratio a natural decomposition of this variable into an high volatility "noise" component, and a low-volatility "information" component naturally emerges.

This paper embeds the decomposition in a small "structural" model that allows for an explicit role for demographics in the dynamic dividend growth model. In particular, the dividend-price ratio is made a function of a temporary "noise" component and of a persistent "information" component related to demographics. The term structure of stock market risk is then derived by the simultaneous estimation of a system for stock market returns at different horizon obtained by the forward solution of the model. We show that the forward solution of the dynamic dividend growth model (Campbell and Shiller 1988) augmented with demographics does naturally progressively eliminate the noise component as the horizon increases. The explicit comparison of our results with the traditional VAR-based methods to derive the term structure of stock market risk shows that the combination of direct regressions methods with the inclusion of the demographic variable in the information set relevant for long-horizon regressions makes the term structure of stock market risk steeply downward sloping.

Importantly, the slope of the term structure of stock market estimated within our proposed framework remains downward sloping even when the two additional sources of risk proposed by PS are considered.

The paper is organized as follows. The first section places our contribution in the literature. In the second section a simple structural model linking demographics and the dynamic dividend growth model is introduced to derive the term structure of stock market risk via direct estimation of a system for returns at different horizon. In the third section, the term structure of stock market risk derived from the forward-looking solution of the model and estimated via direct regressions is compared with the one derived from the backward solution of the VAR adopted by the CV and estimated via the multi-step iterated forecast. We then

[2]We shall discuss the sign of this relation and provide an intepretation for it in the next section of the paper.

consider the impact on our results of considering the two additional sources of uncertainty introduced by PS. The last section concludes.

2 Related Literature

This paper adds to a considerable literature on the relation between the predictability of stock market returns and the term structure of stock market risk.

In describing the "verdict of history" on asset returns on a long-sample (1802-1996) of US historical data Siegel (2007), pp.32, pointed out that "*...stocks are riskier than fixed-income investment over short-term holding periods. But once the holding period increases to between 15 and 20 years, the standard deviation of average annual returns,..., become lower than the standard deviation of average bond and bill returns...*".

This statement on unconditional second moments has been strengthened by Campbell and Viceira (2002, 2005) who exploited the predictability of returns by estimating VAR models for returns and predictors and by using VAR-based multi-period iterated forecasts to find that the conditional variance of stock return does not grow in proportion with the investment horizon but it grows more slowly. As a consequence the term-structure of stock market risk is downward sloping and the findings by Siegel on the property of the unconditional distribution of stocks returns are extended and strengthened when the conditional distribution of returns is used to measure stock market risk.

However, the downward sloping term-structure of stock market risk, has been recently questioned by Pastor and Stambaugh (2008, 2009) who show that, allowing for coefficient uncertainty and imperfect predictors in a Campbell-Viceira type of VAR, the conditional variance of stock returns does increase with the horizon and it can even exceed both the unconditional variance and the one period ahead conditional variance.

The VAR-based approach to measure the term structure of stock market risk uses the (log) dividend-price ratio as the predictor for stock market returns at different horizons. This specification has its foundation in the dynamic dividend growth (DDG) model proposed by Campbell and Shiller (1988). In fact, the DDG model predicts that (log) dividends and prices share a common stochastic trend and that deviations of (log) prices from the common trend in (log) dividends summarize expectations of either stock market returns, or dividend growth or some combination of the two. The empirical investigation of the DDG has established a number of relevant results.

First, the log dividend-price ratio, dp_t, does not have important long-horizon forecasting power for future discounted dividend-growth (Campbell (1991), Cochrane (1991), Campbell, Lo and Mackinlay (1997), Cochrane (2001) and Cochrane (2008)). Second, dp_t is a very persistent time-series and forecasts stock market returns and excess returns over horizons of many years (Fama and French (1988), Campbell and Shiller (1988), Cochrane (2001, Ch. 20), and Cochrane (2008)). Third, the very high persistence of dp_t has led some researchers to question the evidence of its forecasting power for returns, especially at short-horizon. Indeed careful statistical analysis that takes full account of the persistence in dp_t provides little evidence in favor of stock market returns forecastability (Nelson and Kim (1993); Stambaugh (1999); Ang and Bekaert (2007); Valkanov (2003); Goyal and Welch (2003) and Welch and Goyal (2008)). Structural breaks have also been found in the relation between dp_t and future returns (Neely and Weller (2000) and Weller(2000) and Paye and Timmermann (2006), Rapach and Wohar (2006)). With this regard a recent strand of the empirical literature has related the contradictory evidence on the dynamic dividend growth model to the potential weakness of its fundamental hypothesis that log dividend-price ratio is a stationary process (Lettau and Nieuwerburgh (2008), henceforth LVN). LVN use a century of US data to show evidence on the breaks in the constant mean $\overline{dp}$. The potential time-variation of the linearization point creates a link between demographics and the DDG model.

Favero, Gozluklu and Tamoni (2010) have explained the structural breaks found by LVN in the dividend-price process with demographic trends. They point out that theoretical model by Geanakoplos et al. (2004) (henceforth GMQ) predicts that a specific demographic variable, MY, the ratio of middle-aged to young population, determines fluctuations in the dividend yield.

GMQ consider an overlapping generation model in which the demographic structure mimics the pattern of live births in the US, that have featured alternating twenty-year periods of boom and busts. They conjecture that the life-cycle portfolio behavior (Bakshi and Chen 1994), which suggests that agents should borrow when young, invest for retirement when middle-aged, and live off their investment once they are retired, plays an important role in determining equilibrium asset prices. Consumption smoothing by the agents, given the assumed demographic structure requires that when the MY ratio is small (large), there will be excess demand for consumption (saving) by a large cohort of retirees (middle-aged) and for the market to clear, equilibrium prices of financial assets should adjust, i.e. decrease (increase), so that saving (consumption) is encouraged for the middle-aged. As the dividend-price ratio is negatively related to fluctuations in prices, the model predicts a negative relation between this variable and MY. When the GMQ model is taken to the data via the conjecture that fluctuations in MY could capture a slowly evolving mean in dp_t within the dynamic dividend growth model (Favero, Gozluklu and Tamoni (2010)), strong evidence is found in favour of using this variable together with the dividend-price ratio in long-run forecasting regressions for stock market returns. Interestingly, the fluctuations in

MY match very well the break-points in the mean of dp_t identified by LVN in the fifties and the nineties. This paper differs from Favero, Gozluklu and Tamoni (2010) in that it concentrates on the implications of the relation between demographics and the dividend-price ratio for the term structure of stock market risk. To our knowledge, this step has never been taken in the available literature. We propose an empirical strategy potentially capable of identifying separately the importance of demographic variables for high-frequency and low-frequency fluctuations in asset prices. Investigations on the interaction between asset prices and demographic variables have traditionally concentrated either on high-frequency or low frequency fluctuations but have never considered an empirical framework based on the dynamic dividend growth model, capable of accommodating both of them, with a different role (see Erb, Harvey and Viskanta (1997), Poterba (2001), Goyal (2004), Ang and Maddaloni (2005) and DellaVigna and Pollet (2007)).

3 The Dynamic Dividend Growth Model, Demographics and the Term Structure of Stock Market Risk.

The objective of this section is to propose a new model to measure the term structure of stock market risk.

Consider the continuously compounded stock market return from time t to time $t+1$, $\mathbf{r}_{t+1}$. Define μ_t, the conditional expected log return given information up to time t, as follows:

$$\mathbf{r}_{t+1} = \mu_t + \mathbf{u}_{t+1}$$

where $\mathbf{u}_{t+1}$ is the unexpected log return. Define the k-period cumulative return from period $t+1$ through period $t+k$, as follows:

$$\mathbf{r}_{t,t+k} = \sum_{i=1}^{k} \mathbf{r}_{t+i}$$

The term structure of risk is defined as the conditional variance of cumulative returns, given the investor's information set, scaled by the investment horizon

$$\Sigma_r(k) \equiv \frac{1}{k} Var(\mathbf{r}_{t,t+k} \mid D_t) \tag{1}$$

where $D_t \equiv \sigma\{z_k : k \leq t\}$ consists of the full histories of returns as well as predictors that investors use in forecasting returns.

In the light of the results of the empirical investigations on the DDG model and on the evidence of the relation between demographics and the dividend price ratio dp_t, we consider the following small "structural" model:[3]

$$\begin{aligned}
\Delta d_{t+1} &= \varepsilon_{1,t+1} &(2)\\
dp_{t+1} &= \varphi_{22} dp_t + \varphi_{23} MY_{t+1} + \varepsilon_{2,t+1} &(3)\\
r^s_{t+1} &= \Delta d_{t+1} - \rho\left[dp_{t+1} - \overline{dp}_{t+1}\right] + \left[dp_t - \overline{dp}_t\right] + \varepsilon_{3,t+1} &(4)\\
\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} &\sim \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{matrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{matrix}\right]
\end{aligned}$$

Equation (2) specifies the process for the dividend growth as a white noise, where we label $\varepsilon_{1,t+1}$ the innovation to real dividend growth. This simple parameterization is fully consistent with the evidence of very little predictability of dividend growth.[4]

Equation (3) specifies the process for the dividend-price ratio as fluctuating around a time-varying mean determined by the age structure of the population, MY, a slowly evolving highly predictable variable (the Bureau of Census makes available through its web page projections of this variable up to 2050). Such a modification is justified by the theoretical model of Geanakoplos et al. (2004) and by the empirical evidence provided in Favero, Gozluklu and Tamoni (2010). MY constitutes the information component of the dividend-price ratio and there is no uncertainty attached to it: we take it as an exogenous variable whose path for the relevant future is known. However, the dividend-price is also affected by some short-term idiosyncratic noise $\varepsilon_{2,t+1}$. If $|\varphi_{22}| < 1$, then the dividend-price is mean reverting toward a long-run trend determined by the information variable and the effect of the noise shock on the process is only temporary. In fact, our empirical results will show that the speed of mean reversion of the dividend-price ratio toward its long-run mean determined by demographic trends is much higher than that of the dividend-price process itself. Importantly, stability analysis conducted via the Quandt-Andrews test (see Andrews, 1993) for unknown breakpoints confirms the evidence of instability discussed in LVN for the parameters of a simple autoregressive process for dp_{t+1}, while the null of no-break cannot be rejected when the autoregressive model is augmented with MY_{t+1}. [5] This reduced persistence in the financial

[3] The model representation omits constants that have been included in the estimated version.

[4] Following the approach of Lacerda and Santa-Clara (2010) we also tried a specification where agents forecast the dividend growth rate from the average of past dividend growth rates, i.e. $E_t[\Delta d_{t+k}] = \bar{g}_t$. Results, available upon requests, are unaffected.

[5] The Quandt-Andrews test for unknown breakpoints (with a trimming of 10 per cent of the observations) takes a Maximum Wald statistic of 20.06 in 1954 with a tail probability of 0.001 for the parameters in the autoregressive process for dp_t. When the same test is applied to equation (3) the Maximum Wald statistic

ratio is important for two reasons. First, it makes inference less problematic, as there is little doubt on the stationarity of dividend-price ratio around a demographic trend. Second, the reduced half-life of the shocks to the short-term idiosyncratic noise in dividend-price ratios has a direct impact on the term structure of stock market risk.

Equation (4) is an extended version of the Campbell-Shiller log linear approximation of real returns proposed by to allow for time-varying steady-state growth rates and returns. In Appendix B of their paper, (5) derive the following log-linear approximation of returns:[6]

$$\left(r^s_{t+1} - \overline{r}_{t+1}\right) = \left(\Delta d_{t+1} - \overline{\Delta d}_{t+1}\right) - \rho_t \left[dp_{t+1} - \overline{dp}_{t+1}\right] + \left[dp_t - \overline{dp}_t\right] - \Delta \overline{dp}_{t+1} \qquad (5)$$

We obtain our equation (4) from (5) by assuming that of the three processes for returns, dividend growth and the dividend-price ratio only the last one is persistent $\left(\overline{r}_t = \overline{\Delta d}_t = const\right)$, that the time-varying mean of the dividend-price ratio is very slowly evolving, i.e. $\Delta \overline{dp}_{t+1} \approx 0$ and that the linearization parameter is constant, $\rho_t = \rho$.[7] We introduce an error term $\varepsilon_{3,t+1}$ to capture the effect of our approximation.

We report in Figure 2 the three endogenous time series in our small structural model. The graphical evidence suggests that the speed of mean reversion towards a constant mean of the dividend-price ratio is very different from that of annual real returns and annual real dividend growth. We model this feature of the data by introducing a time-varying mean for the dividend-price ratio, driven by demographics. The figure makes clear that without this step it would be very hard to reconcile the time-series properties of dp_t with those of r^s_t and Δd_t. This feature of the data, rather overlooked in the literature, is at the heart of the controversy on the predictive power of the dividend-price ratio for stock market returns.

By solving eq. (4) forward we obtain:

$$\begin{aligned}\sum_{j=1}^{m} \rho^{j-1}\left(r^s_{t+j}\right) &= \left[dp_t - \overline{dp}_t\right] + \sum_{j=1}^{m} \rho^{j-1}\left(\Delta d_{t+j}\right) - \rho^m \left[dp_{t+m} - \overline{dp}_{t+m}\right] \\ &\quad + \sum_{j=1}^{m} \rho^{j-1}\left(\varepsilon_{1,t+j} + \varepsilon_{3,t+j}\right)\end{aligned} \qquad (6)$$

Eq. (6) clearly shows that deviations of the dividend-price ratio from its equilibrium value at time t have a predictive power for m-period ahead stock market returns (and/or dividend

takes a value 11.68 with a tail probability of 0.076 and the null of parameters stability cannot be rejected.

[6] We define $r^s_{t+1} = \ln\left(\frac{P_{t+1}+D_{t+1}}{P_t}\right)$, $dp_t = \ln\left(\frac{D_t}{P_t}\right)$.

[7] Rytchkov (2008) estimates a system of equation similar to ours and study how sensitive ML parameters are to variation in this parameter. He concludes that there is almost no sensitivity to the choice of ρ (see Table 1 in his paper).

growth) that increases with the horizon, as the larger is m the smaller is the effect of future noise in the dividend-price ratio $\left[dp_{t+m} - \overline{dp}_{t+m}\right]$. However, this term cannot be ignored in the computation of the term structure of stock market risk that considers typically horizons from 1-year onwards. To bring (6) to the data, an observable counterpart of the time varying linearization value for the dividend-price must be considered. Consistently with (3), we assume that the relevant linearization value for computing returns from time t to time $t+m$ is the conditional expectation of the dividend-yield for time $t+m$, given the information available at time t. We then have

$$\sum_{j=1}^{m}\rho^{j-1}\left(r^{s}_{t+j}\right) = dp_t - \left[\varphi_{22}^{m} dp_t + \sum_{j=1}^{m}\varphi_{22}^{j-1}\varphi_{23}MY_{t+m+1-j}\right] + u_{t+m} \tag{7}$$

$$= \left(1-\varphi_{22}^{m}\right) dp_t - \sum_{j=1}^{m}\varphi_{22}^{j-1}\varphi_{23}MY_{t+m+1-j} + u_{t+m}$$

$$u_{t+m} = \sum_{j=1}^{m}\rho^{j-1}\left(\varepsilon_{1,t+j}+\varepsilon_{3,t+j}\right) - \rho^{m}\sum_{j=1}^{m}\varphi_{22}^{j-1}\varepsilon_{2,t+m+1-j}$$

Note that the relevance of the noisy component $\varepsilon_{2,t+m+1-j}$ of the dividend-price ratio in the distribution of m-period returns decreases with the horizon: as the horizon gets longer the mean-reversion of the dividend-yield process around the information variable makes the informative content of this variable dominant. The speed at which the effect of the noise is dampened depends on the speed of mean reversion of the dividend process and on the discount parameter ρ. However, even for values of ρ close to unity, the mean reversion in dividend-prices is sufficient to cause a dampening over the horizon of the effect of the noise $\varepsilon_{2,t}$. The remaining components of the noise in m-period returns are the uncertainty in the dividend process and in the real returns. These components die out much more slowly than the noise $\varepsilon_{2,t+j}$ and become persistent when ρ approaches the unit value. Eq. (7) implies that the fit of direct predictive regressions projecting returns at different horizon on the information available at time t should improve with the horizon. It also predicts that the residuals of such predictive regressions have a moving-average component that should be taken care of in estimation. This is a well-known result (see for example, Valkanov (2003)). Interestingly, the model also predicts that the coefficient on the dividend-yield in the projections of long-horizon returns on this variable should be increasing with the horizon.

We measure the term structure of stock market risk by estimating the following "structural" system of eleven equations:[8]

$$
\begin{aligned}
\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(r_{t+j}^{s}\right) &= \delta_{0,m}+\frac{1}{\sqrt{m}}\left(1-\varphi_{22}^{m}\right)dp_t-\frac{\varphi_{23}}{\sqrt{m}}\left(\sum_{j=1}^{m}\varphi_{22}^{j-1}MY_{t+m+1-j}\right)+u_{t+m} \qquad (8)\\
m &= 1,...,10\\
dp_{t+1} &= \varphi_{20}+\varphi_{22}dp_t+\varphi_{23}MY_{t+1}+\varepsilon_{2,t+1}
\end{aligned}
$$

The specification in (8) slightly differs from the model in that we use as a dependent variable the unweighted annualized period-returns ($\rho = 1$). This is because the objective of our exercise is to compare the term structure of stock market risk obtained by direct regression and by iterative multi-step iterated VAR based forecasts. To assess the potential cost of the several approximations we have introduced to derive our structural system an unrestricted version of (8) is also estimated to perform a test of the validity of the relevant restrictions:

$$
\begin{aligned}
\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(r_{t+j}^{s}\right) &= \delta_{0,m}+\frac{\delta_{1m}}{\sqrt{m}}dp_t+\frac{\delta_{2m}}{\sqrt{m}}\left(\sum_{j=1}^{m}\varphi_{22}^{j-1}MY_{t+m+1-j}\right)+u_{t+m} \qquad (9)\\
m &= 1,...,10\\
dp_{t+1} &= \varphi_{20}+\varphi_{22}dp_t+\varphi_{23}MY_{t+1}+\varepsilon_{2,t+1}
\end{aligned}
$$

Note that (8) and (9) are both specified with $\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(r_{t+j}^{s}\right)$ as the dependent variable to obtain directly the conditional annualized standard error of returns from the standard error of the regression.

We estimate the model on a dataset of annual observations for the period 1910-2009. The data are from Welch and Goyal (2008)[9], who provide detailed descriptions of the data and their sources. Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends. To compute real returns we calculate inflation rate from the CPI (all urban consumers). The predictor for the equity premium is the dividend-price ratio, computed as the difference between the log of dividends paid on the S&P 500 index and log of stock prices (S&P 500 index), where dividends are measured using a one-year

[8]Our "structural" estimation is similar to that by Van Binsbergen and Koijen (2009) with two main differences: equations at all relevant horizons are simultaneously estimated and all variables included in the model are observable.

[9]The data are available at www.bus.emory.edu/AGoyal/Research.html.

moving sum.

The results of the estimation are reported in Table 1. The GMM method allows to estimate the restricted model parameters φ_{22} and φ_{23} by taking into account the MA(m-1) structure of the error terms in computing standard errors. We also report estimates of the standard error that uses the SURE method proposed by Pesaran, Pick and Timmermann (2010) to account for serial correlation in the residuals of the multi-period direct forecasting model.[10] The Table shows an highly significant effect of MY both in the equation for dp_t and in all ten predictive regressions. The performance of the restricted model, that estimates only two parameters in addition to eleven constants, is very similar in term of adjusted R^2 and standard error of the equations to that of the unrestricted model that estimates twenty more parameters and the restrictions are not rejected by the relevant chi-square test.

The estimates of the parameters φ_{22} and φ_{23} show that demographics are clearly significant in explaining the dividend-price ratio and that the dividend-price ratio is clearly mean reverting around a mean determined by MY. Figure 3 brings more evidence on this issue by reporting dp_t along with the time-varying linearization point used in the model and the breaks identified by Lettau and Nieuwerburgh (2008).

Finally, the term structure of stock market risk described by the estimation of the structural system of direct regression is steeply downward sloping as it can be read directly off the standard errors of regressions reported in Table 1.

4 Direct regression versus VAR multistep-ahead forecasts

Our empirical results on the term structure of stock market risk differ rather importantly from those derived in the literature based on VAR models. To illustrate the point, we consider first a simple representation of the VAR adopted by CV by estimating a model for continuously compounded total stock market returns, r_t^s, and the log dividend-price, dp_t:

$$(z_t - E_z) = \Phi_1 (z_{t-1} - E_z) + \nu_t$$
$$\nu_t \sim \mathcal{N}(0, \Sigma_\nu)$$

[10]See the Appendix on Robustness for a full description of the results.

where

$$
\begin{aligned}
z_t &= \begin{bmatrix} r_t^s \\ dp_t \end{bmatrix}, E_z = \begin{bmatrix} E_{r^s} \\ E_{d-p} \end{bmatrix} \\
\Phi_1 &= \begin{bmatrix} 0 & \varphi_{1,2} \\ 0 & \varphi_{2,2} \end{bmatrix} \\
\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} &\sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix} \right]
\end{aligned}
$$

The bivariate model for returns and the predictor features a restricted dynamics such that only the lagged predictor is significant to determine current returns ($\varphi_{1,1} = 0$) and the predictor is itself a strongly exogenous variable ($\varphi_{2,1} = 0$).

Given the VAR representation and the assumption of constant Σ_ν

$$
\begin{aligned}
Var_t\left[(z_{t+1} + ... + z_{t+k}) \mid D_t\right] &= \Sigma_\nu + (I + \Phi_1)\Sigma_\nu(I + \Phi_1)' + \\
&\quad (I + \Phi_1 + \Phi_1^2)\Sigma_\nu(I + \Phi_1 + \Phi_1^2)' + ... \\
&\quad +(I + \Phi_1 + ... + \Phi_1^{k-1})\Sigma_\nu(I + \Phi_1 + ... + \Phi_1^{k-1})'
\end{aligned} \tag{10}
$$

from which we can derive:

$$
\begin{aligned}
\Sigma_r(k) &= \frac{1}{k}\sum_{i=0}^{k-1} D_i \Sigma D_i' \\
D_i &= I + \Phi_1 \Xi_{i-1} \quad i > 0 \\
\Xi_i &= \Xi_{i-1} + \Phi_1^i \quad i > 0 \\
D_0 &\equiv I, \quad \Xi_0 \equiv I
\end{aligned}
$$

Note that, under the chosen specification of the matrix Φ_1 we can write the generic term $D_i \Sigma D_i'$, as follows:

$$
D_i \Sigma D_i' = \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}' & M_{22} \end{array}\right) \tag{11}
$$

$$
\begin{aligned}
M_{11} &= \Sigma_{1,1} + \Phi_{1,2}\Xi_{i-1}^{(22)}\Sigma_{1,2}' + \Sigma_{1,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' + \Phi_{1,2}\Xi_{i-1}^{(22)}\Sigma_{2,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' \\
M_{12}' &= \Xi_i^{(22)}\Sigma_{1,2}' + \Xi_i^{(22)}\Sigma_{2,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' \\
M_{22} &= \Xi_i^{(22)}\Sigma_{2,2}\Xi_i^{(22)\prime}
\end{aligned}
$$

where we have used the fact that

$$\begin{aligned}
\Xi_i &= \sum_{j=0}^{i} \Phi_1^i \\
&= \left(\begin{array}{c|c} 0 & \phi_{1,2} \sum_{j=0}^{i-1} \phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^{i} \phi_{2,2}^j \end{array}\right)
\end{aligned}$$

and

$$\begin{aligned}
D_i &= I + \Phi_1 \Xi_{i-1} \\
&= \left(\begin{array}{c|c} I & \phi_{1,2} \sum_{j=0}^{i-1} \phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^{i} \phi_{2,2}^j \end{array}\right)
\end{aligned}$$

Eq. (11) implies that, in our simple bivariate example, the term structure of stock market risk takes the form

$$\sigma_r^2(k) = \sigma_1^2 + 2\varphi_{1,2}\sigma_{1,2}\psi_1(k) + \varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k) \tag{12}$$

where

$$\begin{aligned}
\psi_1(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\sum_{i=0}^{l} \varphi_{2,2}^i \quad\quad k > 1 \\
\psi_2(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\left(\sum_{i=0}^{l} \varphi_{2,2}^i\right)^2 \quad k > 1 \\
\psi_1(1) &= \psi_2(1) = 0
\end{aligned}$$

The total stock market risk can be decomposed in three components: i.i.d uncertainty, mean reversion, uncertainty about future predictors. Without predictability ($\varphi_{1,2} = 0$) the entire term structure is flat at the level σ_1^2. This is the classical situation where portfolio choice is independent of the investment horizon. Since the first and third term are always positive, and in particular the third term increases with the horizon when the autoregressive coefficient in the dividend-price process is positive, then the possible downward slope of the term structure of risk crucially depends on the second term. A necessary condition for having a downward slope requires the contemporaneous presence of both predictability, i.e. $\varphi_{1,2} \neq 0$, and a negative correlation between the innovations in dividend-price ratio and in stock market returns, i.e. $(\sigma_{1,2}) < 0$. Overall, the slope of the term structure of risk depends on the significance of the dividend-price in explaining returns, on the contemporaneous correlation between the innovations in the equations for the dividend-price and returns, on the variance of returns and the dividend-price, and on the persistence of the dividend-price.

Table 2 summarizes the results of the estimation of the system. The estimation results confirm the noisy nature of 1-year stock market returns and the high persistence of the dividend-price ratio. The covariance structure of the innovations is such that the unexpected log excess stock returns are highly negatively correlated with the innovations in the log dividend price ratio. Figure 4 plots the term structure of risk resulting from the estimation of the restricted VAR and its decomposition. The evidence of a downward sloping curve with risk halving from the one-year to the thirty year horizon replicates the results in Campbell and Viceira (2002), based on the estimation of a larger model including bond and stock excess returns, the nominal and real risk-free rate together with the dividend-yield and the yield spread as predictors.

The slope of the term structure of the stock market risk estimated within a traditional VAR features a much smaller decline than that obtained by direct regression of our structural model with demographics. In fact, risk at one year horizon is estimated at about the same level in both models, as the one-period ahead standard deviation is of about 20 per cent. However, from the one-year horizon onward the term structure based on the structural model with demographics declines much more steeply to reach an annualized standard deviation of about 11 per cent at the 10-year horizon against an estimate fifty per cent larger of about 16 per cent from the VAR model.

4.1 Understanding the difference

Why are our results so different from those based on traditional VAR analysis? There are two reasons: the inclusion of MY_t in the investor's information set relevant for long-horizon and the derivation of the term structure of risk via direct regression.

To assess the first effect we estimate the standard Campbell-Viceira VAR augmented with the exogenous MY variable. Results are reported in Table 4.

The comparison of Tables 2 and 4 illustrates clearly that the inclusion of MY in the Campbell-Viceira VAR model would cause a downward shift and a steepening of the term structure of stock market risk as it reduces the persistence of dp_t with respect to the traditional specification omitting demographics (the parameter $\varphi_{2,2}$ of the own lag of dp_t is reduced from .89 to .73 when MY_t is included in the VAR specification) and it also reduces its conditional and unconditional variance (the one-step ahead volatility σ_2 goes from .22 to .21). Interestingly, the correlation between the residuals of the equation for one-year returns and the equation for the dividend- price ratio is very little affected by the inclusion of MY_t in the specification for the dynamics of the predictor (as it goes from -.86 to -.85).

As for the second effect, the use of the direct regression rather than iterative multi-step ahead forecast from the VAR has a less immediate implication that is however empirically and theoretically relevant. To illustrate this point we re-write, consistently with equation (7), the term structure of stock market risk implied by the direct regression of returns at different horizons on the relevant predictors as follows:

$$\begin{aligned}
\sigma_r^2(k) &= \psi_1(k)\left(\sigma_1^2+\sigma_3^2\right)+\psi_2(k)\sigma_2^2 \\
\psi_1(k) &= \frac{1}{k}\sum_{j=1}^{k}\rho^{2(j-1)} \\
\psi_2(k) &= \frac{\rho^{2k}}{k}\sum_{j=1}^{k}\varphi_{22}^{2(j-1)}
\end{aligned} \tag{13}$$

This term structure is downward sloping as the effect of the noisy component of the dividend-price dies out as the horizon m increases.

Consider now the case in which a VAR is fitted to the data generated by eqs. (2)-(4):

$$\begin{aligned}
r_{t+1}^s &= \varphi_{10}+\varphi_{12}dp_t+\varphi_{13}MY_{t+1}+\varepsilon_{1,t+1}+\varepsilon_{3,t+1}-\rho\varepsilon_{2t+1} \\
dp_{t+1} &= \varphi_{20}+\varphi_{22}dp_t+\varphi_{23}MY_{t+1}+\varepsilon_{2,t+1}
\end{aligned}$$

As noted by Cochrane (2008b), deliciously, the regression and "structural "model match almost perfectly.[11] However, the term structure of stock market risk derived by backward projection of the VAR on the information available at time t delivers a different shape from that obtained by the direct regression:

$$\begin{aligned}
\sigma_r^2(k) &= \left(\sigma_1^2+\sigma_3^2+\rho^2\sigma_2^2\right)-2\varphi_{1,2}\rho\sigma_2^2\psi_1(k)+\varphi_{1,2}^2\sigma_2^2\psi_2(k) \\
\psi_1(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\sum_{i=0}^{l}\varphi_{22}^i \quad k>1 \\
\psi_2(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\left(\sum_{i=0}^{l}\varphi_{22}^i\right)^2 \quad k>1 \\
\psi_1(1) &= \psi_2(1)=0
\end{aligned} \tag{14}$$

Note that, as long-run returns are obtained by aggregating high-frequency returns and the relevant model is solved backward, the effect of the noisy component does not decrease with the forecasting horizon. Direct comparison of (13) with (14) shows that the VAR based

[11] Recall that MY is an exogenous variable and therefore we can use MY_{t+1} in the VAR.

backward-looking term structure of risk produces biased estimates of DDG based forward looking term structure of stock market risk.

In order to assess the empirical relevance of the two effects, i.e. the inclusion of the demographic variable in the information set and the use of direct estimation of a forward-looking model rather than iterative multi-step forecast, we compare in Figure 5 the term structure of risk derived by direct estimation of a model including MY_t, with the term structure of risk based on the recursive iteration of two VARs, one with and the other without MY_t. The results show clearly that MY_t plays an important role in determining the conditional mean of the system but also that the use of direct estimation of a forward-looking model rather than iterative recursive multi-step forecast is a source of a major shift in the measured term structure of stock market risk. Consistently with the prediction of the dynamic dividend growth model, such shift is far from being a parallel one.

5 Adding two more sources of uncertainty

Pastor and Stambaugh (2008, 2009) illustrate how the possibility of "imperfect predictors" in the predictive system adds two more sources of uncertainty to the VAR based estimation of the term structure of stock market risk and it empirically changes the slope estimated by Campbell-Viceira. The traditional VAR setup is modified by PS by introducing a predictive relationship linking stock market returns to an unobserved variable μ_t, that in turns is only imperfectly related to the observed dividend-price. In this case, the relevant empirical model can be written as follows:

$$\begin{aligned}
(r_t^s - E_{r^s}) &= (\mu_{t-1} - E_{r^s}) + u_{1,t} \qquad (15)\\
(dp_t - E_{dp}) &= \varphi_{22}\,(dp_{t-1} - E_{dp}) + u_{2,t}\\
(\mu_t - E_{r^s}) &= \varphi_{32}\,(dp_t - E_{dp}) + u_{3,t}\\
\begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} &\sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \begin{matrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{matrix} \end{bmatrix}
\end{aligned}$$

from which the following VAR representation is derived:

$$\begin{aligned}
(z_t - E_z) &= \Phi_1\,(z_{t-1} - E_z) + \nu_t \qquad (16)\\
\nu_t &\sim \mathcal{N}(0, \Sigma_\nu)
\end{aligned}$$

where

$$
\begin{aligned}
z_t &= \begin{bmatrix} r_t^s \\ dp_t \\ \mu_t \end{bmatrix}, E_z = \begin{bmatrix} E_{r^s} \\ E_{dp} \\ E_{r^s} \end{bmatrix} \\
\Phi_1 &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & \varphi_{22} & 0 \\ 0 & \varphi_{32}\varphi_{22} & 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{bmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{bmatrix} \sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \begin{matrix} \sigma_1^2 & \sigma_{12} & \varphi_{32}\sigma_{12} + \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \varphi_{32}\sigma_2^2 + \sigma_{23} \\ \varphi_{32}\sigma_{12} + \sigma_{13} & \varphi_{32}\sigma_2^2 + \sigma_{23} & \varphi_{32}^2\sigma_2^2 + \sigma_3^2 + 2\varphi_{32}\sigma_{23} \end{matrix} \end{bmatrix}
$$

the term structure of stock market risk takes now the form

$$
\sigma_r^2(k) = \sigma_1^2 + 2\sigma_{13} + \sigma_3^2 + 2\varphi_{3,2}(\sigma_{12} + \sigma_{23})\psi_1(k) + \varphi_{2,2}^2\sigma_{2,2}^2\psi_2(k)
$$

where

$$
\begin{aligned}
\psi_1(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\sum_{i=0}^{l}\beta_{22}^i \quad k > 1 \\
\psi_2(k) &= \frac{1}{k}\sum_{l=0}^{k-2}\left(\sum_{i=0}^{l}\beta_{22}^i\right)^2 \quad k > 1 \\
\psi_1(1) &= \psi_2(1) = 0
\end{aligned}
$$

Note that the specification of the relevant VAR to project the term structure of risk involves an unobservable variable. VAR estimation needs that to deal with this problem that it is best solved within a Bayesian framework. Such a framework in turn generates naturally another source of volatility, namely parameters uncertainty. In the PS framework the term-structure of risk can be decomposed into five components: the original three in eq. (12) plus other two, one reflecting uncertainty around the mean of the process generating returns and one reflecting parameters' uncertainty.

Table 3 shows the results from the estimation of the three-variate predictive system. Within this Bayesian framework the prior beliefs on the correlation between innovations in the equation for returns and innovations in the equation for expected returns (i.e. ρ_{ν_1,ν_3}, the Stambaugh Correlation) substantially affects estimates of expected returns as well as various inferences about predictability. In our estimation we impose the belief $\rho_{\nu_1,\nu_3} < 0$ following the evidence of Campbell (1991) and of Van Binsbergen and Koijen (2009). Interestingly, Robertson and Wright (2009) show that for a plausible range of ARMA parameters the Stambaugh Correlation is bounded away from zero and very close to (minus) unity. This leads us to specify an informative prior on ρ_{ν_1,ν_3} that the implied prior on $\rho^2_{\nu_1,\nu_3}$ has 99.9% of its mass above 0.5, with a mean of about 0.77.[12] The Table 3 reports the R^2 in the regression of r_{t+1} on $E[r_{t+1}|D_t]$ for the predictive system and shows that this R^2 is higher than the R^2 in Table 2 because $dp_t \in D_t$ and therefore the estimates of the expected returns from the predictive system are at least as precise as the estimates from the predictive regression and VAR. Moreover the R^2 (not reported) from a regression of μ_t on dp_t larger than 0.5 receive very little posterior probability suggesting that the predictor is not perfectly correlated with the latent expected return and therefore the predictive system has superior ability in extracting information and forming the proxy for the true unobservable μ_t.

Figure 6 plots the conditional variance $Var[r_{t,t+k}|D_t]$ and its components for the three-variate predictive system with the dividend price as an observable predictor. It is interesting to note that the three components of $Var[r_{t,t+k}|\mu_t, \Theta, D_t]$[13], namely the i.i.d. (top right panel), the mean reverting (mid left panel) and the uncertainty about future values of μ_t (mid right panel) are fairly similar to the one we compute under the VAR approach (see Figure fig:TSriskComp). Therefore the sum of these three contribution, namely $Var[r_{t,t+k}|\mu_t, \Theta, D_t]$ is almost identical under the CV and the PS approaches. Nevertheless Pastor and Stambaugh (2009) show that other two important blocks affect the conditional variance $Var[r_{t,t+k}|D_t]$: one is the predictor imperfection that reflects the uncertainty about the current conditional expected returns and the other is the estimation risk that reflects uncertainty about parameters in Φ_1. In particular, when we add the predictor imperfection component to the $Var[r_{t,t+k}|\mu_t, \phi, D_t]$ we obtain the dashed line in top-left panel of Figure 6 which shows no evidence for a downward sloping term structure of stock-market risk.

In the light of these results it important to evaluate the effect of the introduction of the new sources of risk on the term structure measured by direct regression within our small structural model. The imperfect predictors problem would affect the interpretation of the error term in the ten estimated equations for returns at horizon from 1-year to 10-year but it would not change their estimated standard deviations. In fact in the direct regression approach the term structure of stock market risk is obtained by projecting returns at different horizon $t+m$ on observable variables at time t. If these observable variables are imperfect predictors, then the variance of the direct regression residual will reflect this feature of the data.

[12]As noted in PS 2008 this prior reflects the belief that at least half of the variance of market returns is due to discount rate news.

[13]We indicate with Θ the full set of parameters.

Parameter uncertainty could be an issue, as in the previous section we derived the term structure of stock market risk by keeping the estimated parameters fixed without considering uncertainty surrounding them. We have therefore recomputed the term structure by allowing for parameter uncertainty both in the restricted and unrestricted version of our model. The results, reported in Figure 7, show that parameters uncertainty add very little uncertainty to the conditional distribution of stock market returns in our framework. This evidence is not surprising, as the estimation results reported in Table (1) document that the term structure of risk derived from our structural system is based on the estimation of very few parameters, all them very well determined. To provide further evidence on the relative performance of our small structural model with demographics and the three-variable VAR with one imperfect predictor proposed by PS we report in Figure 8 actual and predicted returns at the 1-year and 10-year horizon. At the shor-end of the term structure the information from fundamentals is totally blurred by the "noise" and the structural model with demographics does not overperfrom the three-variate VAR. However, at the long-end the "information" generated by using the predictor determined by demographics in the structural model prevails and generates a sizeable overperformance. More precisely at the horizon of 10 years, the mean square error (MSE) and mean absolute error (MAE) for the predictive system are equal to 0.0023 and 0.0409, respectively, whereas the MSE and MAE for the direct regression are equal to 0.0012 and 0.0280, respectively.

6 Conclusions

We started this paper by arguing that the emergence of relative importance of "information" versus "noise" in the determination of stock market returns at different horizon could generate a downward sloping term structure of stock market risk. We have shown that this is indeed the case when i) a demographic variable is used to capture the slow-moving information component in the dividend-price ratio and in stock market returns and ii) direct regressions based on the structural estimation of a forward-looking specification consistent with the dynamic dividend growth model is adopted. We have also shown that the use of backward-looking iterated multi-step forecasts to derive the term structure of risk leads to an underestimation of the importance of the emergence of "information" as the horizon increases.

A Robustness

The effect of overlapping observations in the system described in equation (7) is that the error term is serially correlated, as it features a moving average structure, and therefore conventional OLS standard errors are incorrect. In Table 1 the reported standard errors are corrected for the presence of serial correlation, using a Generalised Method of Moments (GMM) estimator. As an additional test, we examined non-overlapping regression for the unrestricted system. This approach of course does not capture all of the information in the data but it has the advantage of being transparent and providing a baseline estimate. The results, available upon request from the corresponding author, show that the parameters δ_{1m} and δ_{2m} are all significant and very close to the values reported in Table 1. The problem now is that, e.g. at horizon $m = 10$ we have only ten observations. To deal with this problem we have adopted the approach proposed by Pesaran, Pick, and Timmermann (2010) that allows to capture all of the information in the data by choosing the appropriate weighting matrix to accommodate overlapping observations. In particular, following the notation of Pesaran, Pick, and Timmermann (2010), for the fixed horizon m we can estimate $\beta = \{\delta_{1m}, \delta_{2m}\}$ from pooled regressions of m non-overlapping regressions:

$$y_{j+(i-1)m} = \beta x_{j+(i-1)m-m} + v_{i,j} \quad i = 1, 2, \ldots, \frac{T}{m} \textit{ and } j = 1, 2, \ldots, m$$

where $y_{j+(i-1)m}$ is the series of m-period cumulated returns sampled every m period with offset j and analogously $x_{j+(i-1)h-h}$ are the sampled regressors MY and dp_t. Now we can view the m regressions as a set of seemingly unrelated regression equations, allowing for cross dependence of the errors. For each horizon m we estimate $\hat{\beta} = (\delta_{0m}, \delta_{1m}, \delta_{2m})$ as $[W'\Sigma_u^{-1}W]^{-1}W'\Sigma_u^{-1}\tilde{y}$ and $Var(\hat{\beta}) = [W'\Sigma_u^{-1}W]^{-1}$ where $\tilde{y}$ and W are the reordered matrices of regressand and regressors and Σ_u is the appropriate GLS covariance matrix whose analytic form is reported in Pesaran, Pick, and Timmermann (2010). Importantly as Pesaran, Pick, and Timmermann (2010) highlight, since the direct regression is derived from the forward solution of the Campbell-Shiller return log-linearization, the GLS covariance matrix and the MA coefficients of the errors term can be linked to a set of deeper parameters, namely $\{\varphi_{22}, \rho\}$ by just recalling the following

$$u_{t+m} = \sum_{j=1}^{m} \rho^{j-1} \left(\varepsilon_{1,t+j}\right) - \rho^m \sum_{j=1}^{m} \varphi_{22}^{j-1} \varepsilon_{2,t+m+1-j}$$

The corrected t-stat obtained by imposing the above error terms structure when we estimate the GLS covariance matrix are reported in Table 1 in the line GLS-PPT. Alternatively we estimate the GLS covariance matrix Σ_u without restrictions. Results are almost identical.

Table 1: System Estimation (1910-2009)

$$dp_{t+1} = \varphi_{20} + \varphi_{22} dp_t + \varphi_{23} MY_{t-j} + \varepsilon_{2t+1}$$

$$\text{UM:} \frac{1}{\sqrt{m}} \sum_{j=1}^{m} r^s_{t+j} = \delta_{0m} + \frac{\delta_{1m}}{\sqrt{m}} dp_t + \frac{\delta_{2m}}{\sqrt{m}} \left(\sum_{j=1}^{m} \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m} \quad m = 1, .., 10$$

$$\text{RM:} \frac{1}{\sqrt{m}} \sum_{j=1}^{m} r^s_{t+j} = \delta_{0m} + \frac{1}{\sqrt{m}} \left(1 - \varphi_{22}^m\right) dp_t - \frac{\varphi_{23}}{\sqrt{m}} \left(\sum_{j=1}^{m} \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m}$$

	horizon m in years										
UM		1	2	3	4	5	6	7	8	9	10
δ_{1m}		0.18	0.38	0.48	0.61	0.70	0.73	0.78	0.86	0.89	0.89
$(t-stat)$		(3.87)	(5.57)	(5.61)	(6.96)	(7.99)	(7.17)	(7.25)	(8.73)	(7.63)	(6.15)
$(GLS-PPT)$		(−)	(4.54)	(5.21)	(5.77)	(7.43)	(5.19)	(7.03)	(8.61)	(7.57)	(2.67)
δ_{2m}		0.41	0.52	0.56	0.64	0.69	0.70	0.74	0.78	0.81	0.83
$(t-stat)$		(3.23)	(3.69)	(3.85)	(4.28)	(4.55)	(4.69)	(4.78)	(4.93)	(4.94)	(5.01)
$(GLS-PPT)$		(−)	(3.85)	(3.50)	(3.87)	(4.21)	(4.10)	(4.06)	(4.10)	(4.19)	(4.84)
φ_{22}	0.61										
$(t-stat)$	(9.21)										
φ_{23}	−0.83										
$(t-stat)$	(−3.79)										
RM											
φ_{22}	0.76										
$(t-stat)$	(19.31)										
φ_{23}	−0.53										
$(t-stat)$	(4.41)										
χ^2_{12}	13.45	χ^2_{20}	17.19								
	(0.34)		(0.64)								
σ_{DepVar}		0.195	0.198	0.187	0.185	0.181	0.174	0.172	0.173	0.171	0.168
$\sigma_{u_{t+m}}$	UM	0.188	0.179	0.164	0.152	0.140	0.131	0.125	0.118	0.112	0.109
$\sigma_{u_{t+m}}$	RM	0.189	0.179	0.164	0.152	0.141	0.133	0.127	0.120	0.115	0.112
$adjR^2$	UM	−	0.18	0.24	0.33	0.41	0.44	0.48	0.54	0.57	0.58
$adjR^2$	RM	0.06	0.18	0.23	0.32	0.40	0.42	0.46	0.52	0.54	0.55

Table 1: This table compares the univariate OLS long-horizon regression coefficients, to the GMM estimates that impose the restrictions suggested by the present-value model with demographics. The estimation is by GMM, where the moments are the OLS normal conditions. Standard errors are by Newey-West with optimal bandwidth selection. The first-stage weighting matrix is the identity matrix. $GLS-PPT$ is the t-stat that explicitly accounts for the MA(m-1) errors structure by using a GLS covariance matrix as suggested in Pesaran, Pick and Timmermann (2010). σ_{DepVar} is the annualized unconditional standard deviation. $\sigma_{u_{t+m}}$ is the annualized conditional standard deviation of the compounded (over m periods) returns, i.e. our measure of stock market risk. The effective sample period is 1910-2009.

Table 2: A simple bivariate VAR (1910-2009)

$$\left(r^s_{t+1} - E_{r^s}\right) = \varphi_{12}\left(dp_t - E_{dp}\right) + \nu_{1,t+1}$$
$$\left(dp_{t+1} - E_{dp}\right) = \varphi_{22}\left(dp_t - E_{dp}\right) + \nu_{2,t+1}$$

φ_{12} $(t-stat)$	φ_{22} $(t-stat)$	χ^2_2 $\varphi_{11}=0,\varphi_{21}=0$	σ_1	σ_2	$\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$	$adjR^2_{r^s_{t+1}}$	$adjR^2_{dp_{t+1}}$
0.067	0.892	5.673	0.194	0.219	-0.856	0.02	0.78
(1.70)	(18.80)	(0.06)					

Table 2: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.

Table 3: A three-variate VAR with imperfect predictors (1910-2009)

$$\left(r^s_{t+1} - E_{r^s}\right) = \left(\mu_t - E_{r^s}\right) + \nu_{1,t+1}$$
$$\left(dp_{t+1} - E_{dp}\right) = \varphi_{22}\left(dp_t - E_{dp}\right) + \nu_{2,t+1}$$
$$\left(\mu_{t+1} - E_{r^s}\right) = \varphi_{33}\left(\mu_t - E_{r^s}\right) + \nu_{3,t+1}$$

φ_{22} $(low-upp)$	φ_{33} $(low-upp)$	σ_1	σ_2	σ_3
0.847	0.939	0.187	0.285	0.171
[0.757 0.934]	[0.811 0.993]	[0.166 0.214]	[0.251 0.327]	[0.088 0.335]
	Pred R^2	$\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$	$\frac{\sigma_{13}}{\sigma_{11}\sigma_{33}}$	$\frac{\sigma_{23}}{\sigma_{22}\sigma_{33}}$
	0.04	-0.631	-0.655	0.451
		[-0.732 -0.502]	[-0.841 -0.325]	[0.205 0.623]

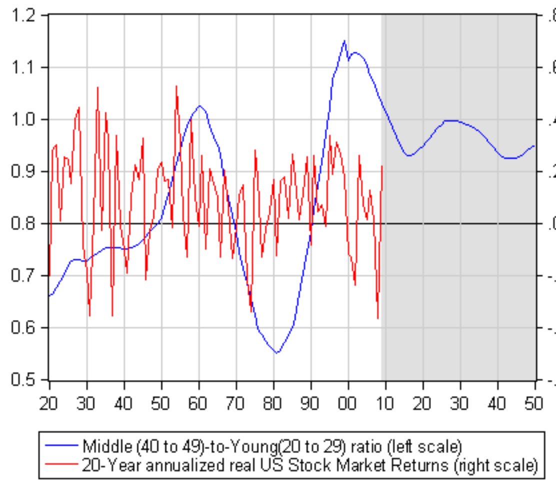
Table 3: This table shows the posterior median and [0.025 0.975] quantile obtained with the predictive system described in Pastor and Stambaugh (2008). The sample period is 1910-2009.

Table 4: A bi-variate VAR with MY (1910-2009)

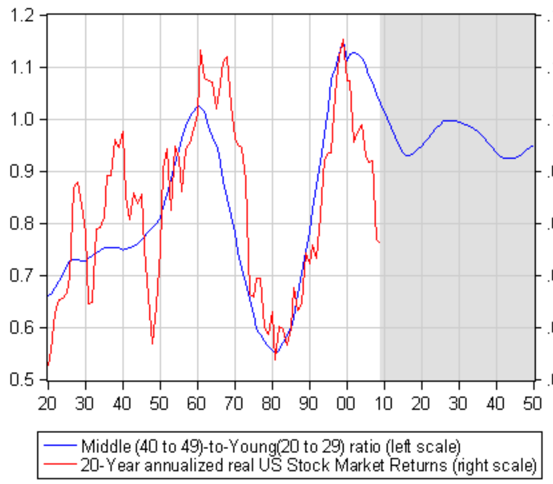
$$r^s_{t+1} = \varphi_{10} + \varphi_{12} dp_t + \varphi_{13} MY_{t+1} + \nu_{1,t+1}$$
$$dp_{t+1} = \varphi_{20} + \varphi_{22} dp_t + \varphi_{23} MY_{t+1} + \nu_{2,t+1}$$

$\underset{(t-stat)}{\varphi_{12}}$	$\underset{(t-stat)}{\varphi_{13}}$	$\underset{(t-stat)}{\varphi_{22}}$	$\underset{(t-stat)}{\varphi_{23}}$	$\underset{\varphi_{11}=0,\varphi_{21}=0}{\chi^2_2}$	σ_1	σ_2	$\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$	$adjR^2_{r^s_{t+1}}$	$adjR^2_{dp_{t+1}}$
0.179	0.410	0.728	-0.603	4.74	0.188	0.207	-0.846	0.07	0.80
(3.07)	(2.67)	(11.33)	(-3.56)	(0.09)					

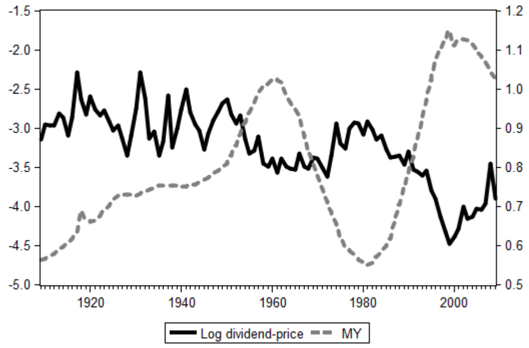
Table 4: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.



(a) 1-year real US stock market returns and demographic trends.



(b) 20-year ex-post real US stock market returns and demographic trends. Returns over the holding period from t to $t+h$ are plotted along with MY_{t+h}.



(c) The demographic trend MY and the log dividend-price ratio.

Figure 1: Stock market returns, the dividend-price and demography.

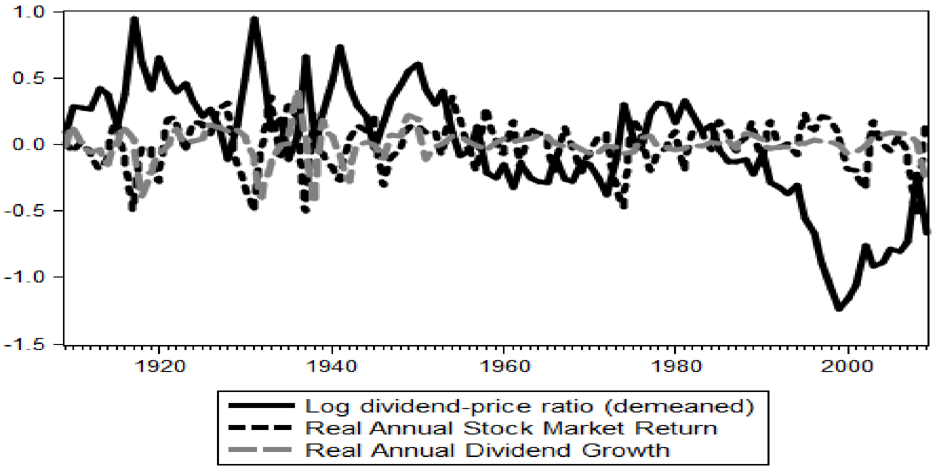


Figure 2: The three endogenous variables in our model.

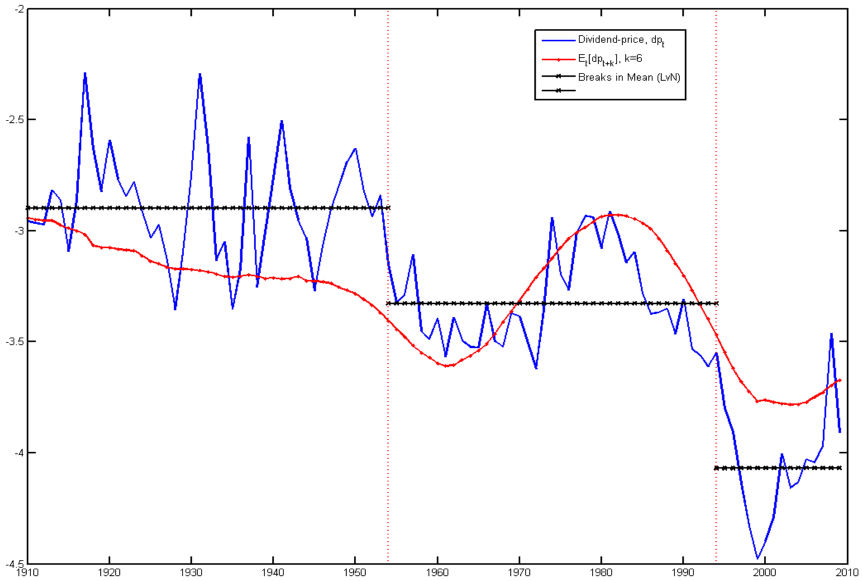


Figure 3: dp_t along with the time-varying linearization point used in our model and the breaks identified by LVN.

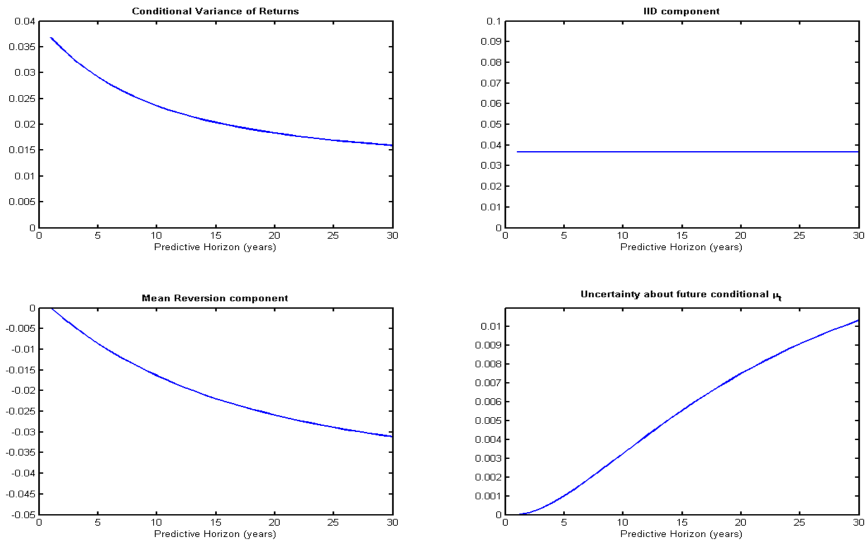


Figure 4: The term structure of stock market risk from a bi-variate VAR.

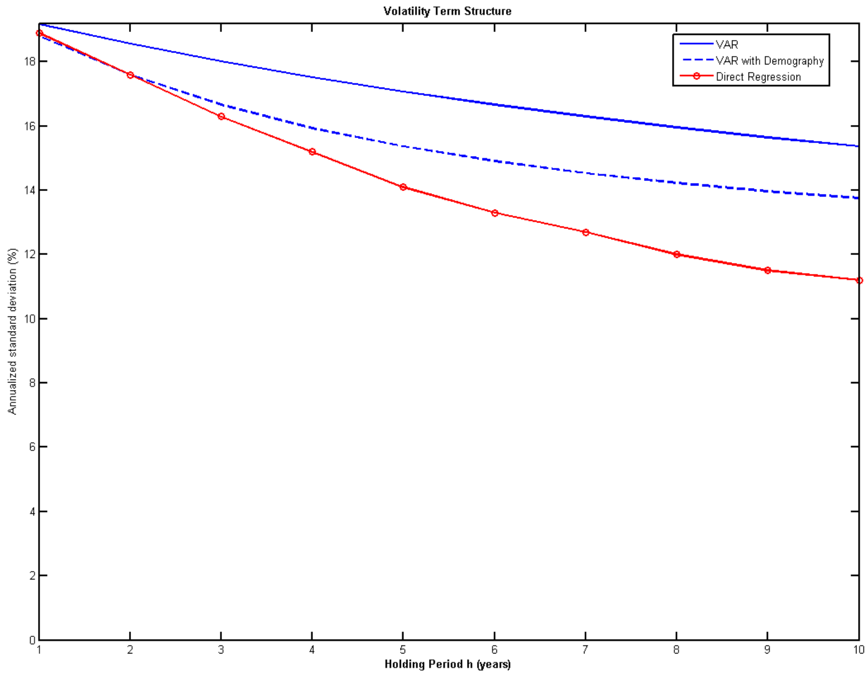


Figure 5: Three alternative measures of the TS of stock market risk.

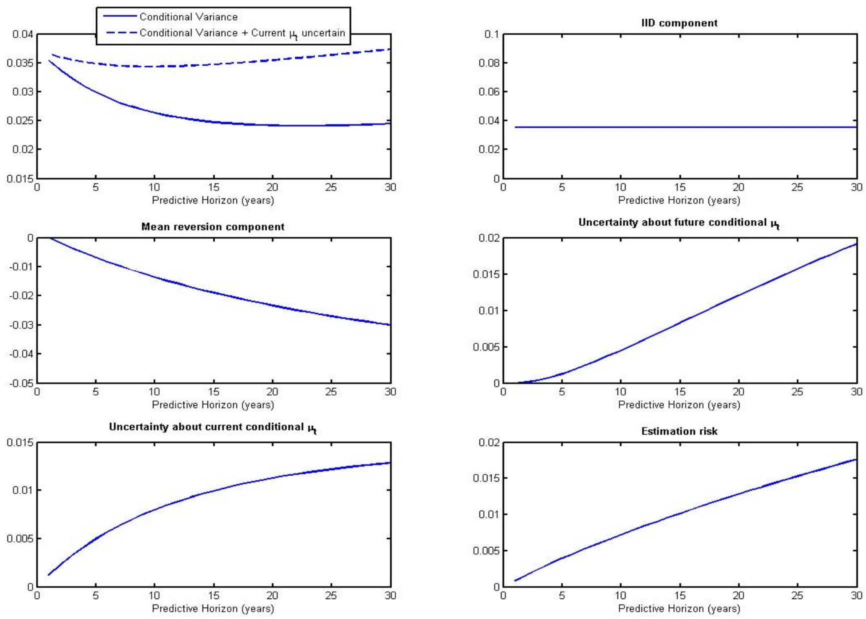


Figure 6: The term structure of US stock market risk from a three-variate VAR with an unobservable component.

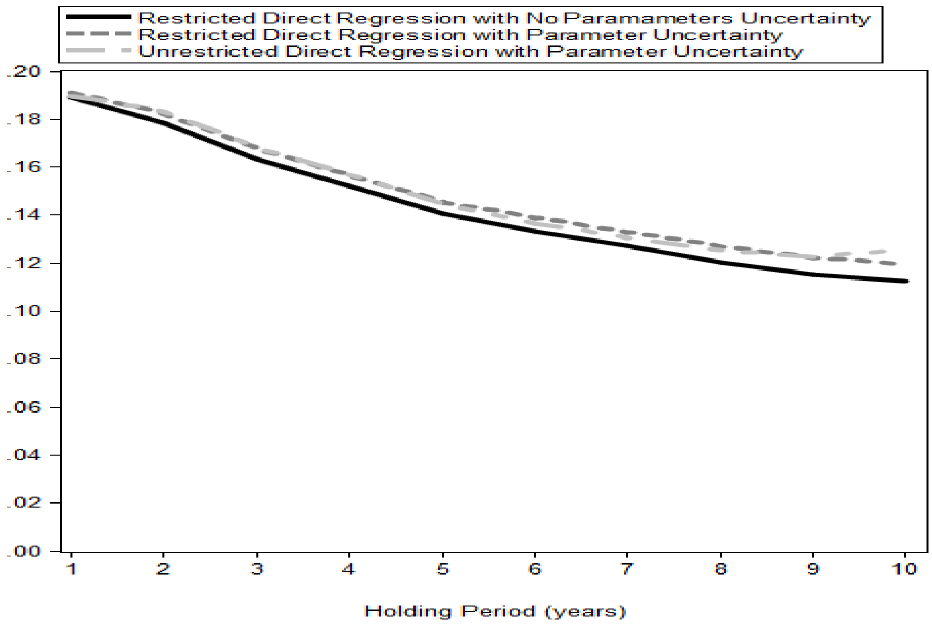


Figure 7: The term structure of US stock market risk in a structural model with and without parameters uncertainty.

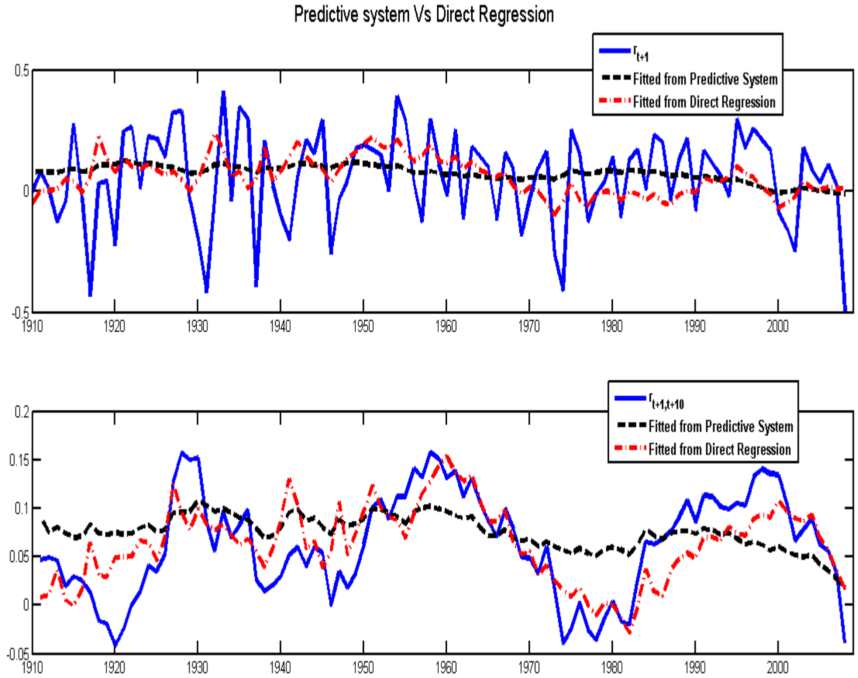


Figure 8: Noise and Information reconsidered. The top panel plots one-year ahead realized returns along with the forecast obtained from both the predictive system and the direct regression approach. The bottom panel plots ten-year ahead realized returns along with the forecast obtained from both the predictive system and the direct regression approach.

References

Ang, Andrew and Angela Maddaloni, "Do Demographic Changes Affect Risk Premiums? Evidence from International Data," *Journal of Business*, 2005, *78* (1), 341–380.

___ **and Geert Bekaert**, "Stock Return Predictability: Is It There?," *Review of Financial Studies*, 2007, *20* (3), 651–707.

Bakshi, Gurdip S and Zhiwu Chen, "Baby Boom, Population Aging, and Capital Markets," *Journal of Business*, April 1994, *67* (2), 165–202.

Binsbergen, Jules H. Van and Ralph S. Koijen, "Predictive Regressions: A Present-Value Approach," *SSRN eLibrary*, 2009.

Campbell, John Y., "A Variance Decomposition for Stock Returns," *Economic Journal*, 1991, *101* (405), 157–79.

___ **and Luis M. Viceira**, "Strategic Asset Allocation," 2002.

Campbell, John Y and Luis M Viceira, "The Term Structure of the Risk-Return Tradeoff," CEPR Discussion Papers 4914, C.E.P.R. Discussion Papers February 2005.

Campbell, John Y. and Robert J. Shiller, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1988, *1* (3), 195–228.

___ **, Andrew W. Lo, and A. Craig Mackinlay**, *The Econometrics of Financial Markets*, Princeton University Press, June 1997.

Cochrane, John H., "Volatility tests and efficient markets : A review essay," *Journal of Monetary Economics*, June 1991, *27* (3), 463–485.

___ , *Asset Pricing.*, Princeton University Press, Princeton, NJ, 2001.

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

DellaVigna, Stefano and Joshua M. Pollet, "Demographics and Industry Returns," *American Economic Review*, 2007, *97* (5), 1667–1702.

Erb, Claude B., Campbell R. Harvey, and Tadas E. Viskanta, "Demographics and International Investment," *Financial Analysts Journal*, 1997.

Fama, Eugene F. and Kenneth R. French, "Dividend yields and expected stock returns," *Journal of Financial Economics*, October 1988, *22* (1), 3–25.

Favero, Carlo A., Arie E. Gozluklu, and Andrea Tamoni, "Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns," *Journal of Financial and Quantitative Analysis, forthcoming*, 2010, pp. –.

Geanakoplos, John, Michael Magill, and Martine Quinzii, "Demography and the Long-Run Predictability of the Stock Market," *Brookings Papers on Economic Activity*, 2004, *35* (2004-1), 241–326.

Goyal, Amit, "Demographics, Stock Market Flows, and Stock Returns," *Journal of Financial and Quantitative Analysis*, March 2004, *39* (01), 115–142.

___ **and Ivo Welch**, "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 2003, *49* (5), 639–654.

Lacerda, Filipe and Pedro Santa-Clara, "Forecasting Dividend Growth to Better Predict Returns," Technical Report February 2010.

Lettau, Martin and Stijn Van Nieuwerburgh, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, July 2008, *21* (4), 1607–1652.

Neely, Christopher J. and Paul Weller, "Predictability in International Asset Returns: A Reexamination," *The Journal of Financial and Quantitative Analysis*, 2000, *35* (4), 601–620.

Nelson, Charles R. and Myung J. Kim, "Predictable Stock Returns: The Role of Small Sample Bias," *The Journal of Finance*, 1993, *48* (2), 641–661.

Pastor, Lubos and Robert F Stambaugh, "Predictive Systems: Living with Imperfect Predictors," *Journal of Finance, forthcoming*, February 2008, pp. –.

___ **and** ___ , "Are Stocks Really Less Volatile in the Long Run?," NBER Working Papers 14757, National Bureau of Economic Research, Inc March 2009.

Paye, Bradley S. and Allan Timmermann, "Instability of return prediction models," *Journal of Empirical Finance*, 2006, *13* (3), 274 – 315.

Pesaran, M. Hashem, Andreas Pick, and Allan G Timmermann, "Variable Selection, Estimation and Inference for Multi-period Forecasting Problems," CEPR Discussion Papers 7139, C.E.P.R. Discussion Papers June 2010.

Poterba, James M., "Demographic Structure And Asset Returns," *The Review of Economics and Statistics*, November 2001, *83* (4), 565–584.

Rapach, David E. and Mark E. Wohar, "In-sample vs. out-of-sample tests of stock return predictability in the context of data mining," *Journal of Empirical Finance*, March 2006, *13* (2), 231–247.

Robertson, Donald and Stephen Wright, "The Limits to Stock Return Predictability," Technical Report 2009.

Rytchkov, Oleg, "Filtering Out Expected Dividends and Expected Returns, working paper," 2008.

Schotman, Peter C., Rolf Tschernig, and Jan Budek, "Long Memory and the Term Structure of Risk," *Journal of Financial Econometrics*, Fall 2008, *6* (4), 459–495.

Siegel, Jeremy J., *Stocks for the Long Run*, McGraw-Hill Companies, November 2007.

Stambaugh, Robert F., "Predictive regressions," *Journal of Financial Economics*, December 1999, *54* (3), 375–421.

Valkanov, Rossen, "Long Horizon Regressioms: Theoretical Results and Applications," *Journal of Financial Economics*, 2003, *68*, 201–232.

Welch, Ivo and Amit Goyal, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, July 2008, *21* (4), 1455–1508.

Part IV

Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns

1 Introduction

This paper characterizes the relationship between the dividend-price ratio and stock market returns in a model where a demographic variable, MY_t, the middle-aged to young ratio, captures the slowly evolving component in the dividend-price ratio. Interest in this model is partly motivated by the very high persistence of the dividend-price ratio that makes long-horizon regressions hard to interpret MY_t allows to extract from the log dividend-price a stationary variable capturing time-variation in the investment opportunity set and to specify a more reliable forecasting model for long-horizon stock market returns. Demographics are a very natural input into a forecasting model of long-horizon returns, and, consequently, into the optimal asset allocation decision of a long-horizon investor. We interpret MY_t as the information component that drives long-horizon stock market fluctuations after the noise in short-horizon stock market fluctuations subsides.

The empirical relevance of the dividend-price ratio for predicting long-run stock market returns is one of the most debated issues in financial econometrics. In fact, this variable regularly plays an important role in recent empirical literature that has replaced the long tradition of the efficient market hypothesis (Fama, 1970) with a view of predictability of returns (see, for example, Cochrane, 2007). However, there is an ongoing debate on the robustness of return predictability and its potential use from a portfolio allocation perspective (Boudoukh et al. (2008), Goyal and Welch (2008)).

Most of the available evidence on predictability can be framed within the dynamic dividend growth model proposed by Campbell and Shiller (1988). This model relies on a loglinearized version of one-period returns on the stock portfolio. Under the assumption of its stationarity and of the validity of a standard transversality condition, the log of the price-dividend ratio, dp_t, is expressed as a linear function of the future discounted dividend growth, $\Delta \mathrm{d}_{t+j}$ and of future returns, h^s_{t+j} :

$$\mathrm{dp}_t = \overline{\mathrm{dp}} + \sum_{j=1}^{\infty} \rho^{j-1} \mathrm{E}_t[(\mathrm{h}^s_{t+j} - \bar{\mathrm{h}}) - (\Delta \mathrm{d}_{t+j} - \bar{\mathrm{d}})] \tag{1}$$

where $\overline{\mathrm{dp}}$, the mean of the dividend-price ratio, $\bar{\mathrm{d}}$, the mean of dividend growth rate, $\bar{\mathrm{h}}$, the mean of log return and ρ are constants.

Under the maintained hypothesis that stock market returns, and dividend-growth are covariance-stationary, Eq. (1) says that the log dividend-price ratio is stationary, i.e. the (log) price and the (log) dividend are cointegrated with a (-1,1) cointegrating vector, and that deviations of (log) prices from the common trend in (log) dividends summarize expectations of either stock market returns, or dividend growth or some combination of the two.

The empirical investigation of the dynamic dividend growth model has established a few empirical results:

(i) dp_t is a very persistent time-series and forecasts stock market returns and excess returns over horizons of many years (Fama and French (1988), Campbell and Shiller (1988), Cochrane (2005, 2007).

(ii) dp_t does not have important long-horizon forecasting power for future discounted dividend-growth (Campbell (1991), Campbell, Lo and McKinlay (1997) and Cochrane (2001)).

(iii) the very high persistence of dp_t has led some researchers to question the evidence of its forecasting power for returns, especially at short-horizons. Careful statistical analysis that takes full account of the persistence in dp_t provides little evidence in favour of the stock-market return predictability based on this financial ratio (Nelson and Kim (1993); Stambaugh (1999); Ang and Bekaert (2007); Valkanov (2003); Goyal and Welch (2003) and Goyal and Welch (2008)). Structural breaks have also been found in the relation between dp_t and future returns (Neely and Weller (2000), Paye and Timmermann (2006) and Rapach and Wohar (2006)).

(iv) More recently, Lettau and Ludvigson (2001, 2005) have found that dividend growth and stock returns are predictable by long-run equilibrium relationships derived from a linearized version of the consumer's intertemporal budget constraint. The excess consumption with respect to its long run equilibrium value is defined by the authors alternatively as a linear combination of labour income and financial wealth, cay_t, or as a linear combination of aggregate dividend payments on human and non-human wealth, cdy. cay_t and cdy_t are much less persistent than dp_t, they are predictors of stock market returns and dividend-growth, and, when included in a predictive regression relating stock market returns to dp_t, they swamp the significance of this variable. Lettau and Ludvigson (2005) interpret this evidence in the light of the presence of a common component in dividend growth and stock market returns. This component cancels out from (1), cay_t and cdy_t are instead able to capture it as the linearized intertemporal consumer budget constraint delivers a relationship between excess consumption and expected dividend growth or future stock market returns that is independent from their difference.

A recent strand of the empirical literature has related the contradictory evidence on the dynamic dividend growth model to the potential weakness of its fundamental hypothesis that the log dividend-price ratio is a stationary process (Lettau and Van Nieuwerburgh (2008), LVN henceforth). LVN use a century of US data to show evidence on the breaks in the constant mean $\overline{dp}$. We report the time series of US data on dp_t over the last century in Figure 1. As a matter of fact, the evidence from univariate test for non-stationarity and bivariate cointegration tests does not lead to the rejection of the null of the presence of a unit-root in dp_t[1].

As shown in Figure 1, LVN identify two statistically significant breaks in the mean of dp_t in 1954 and 1991. They then provide evidence that deviations of dp_t from its time-varying mean have a much stronger forecasting power for stock market returns than deviations of dp_t from a constant mean[2]. This evidence for time-variation in the mean of the dividend-price ratio has been also confirmed by Johannes et al. (2008), who estimate the process for log dividend-price ratio within a particle filtering framework.

So far the evidence towards a slowly evolving mean in dp_t has been reported as a pure statistical fact. LVN give some hints on possible causes for the breaks arising from economic fundamentals due to technological innovation, changes in expected return, etc. but do not explore the possible effects of fundamentals any further. The idea of correcting dp_t to reduce its persistence has been also pursued by an alternative strand of research that relates the apparent non-stationarity of this variable to a shift in corporate payout policies. Boudoukh et al. (2007) provide a new measure of the cash flow based net payout yield (dividends plus repurchases minus issuances are used instead of dividends to construct the relevant ratio) that is more quickly mean reverting than the dividend-price ratio. Yet, this suggested measure is unlikely to explain the full decrease in this financial ratio as argued by LVN. Moreover other financial ratios such as earning-price ratio witness similar declines.

The aim of our paper is to investigate the possibility that the slowly evolving mean in the log dividend-price is related to demographic trends. We first illustrate how the theoretical model by Geanakoplos, Magill and Quinzii (2004, henceforth, GMQ) implies that a specific demographic variable, MY_t, the proportion of middle-aged to young population, explains fluctuations in the dividend yield.

GMQ consider an overlapping generation model where the demographic structure mimics the pattern of live births in the US, which have featured alternating twenty-year periods of

[1]The Dickey-Fuller test for the null of non-stationarity delivers an observed statistics of -2.34 when computed over the full sample 1911-2008 and a value of -1.72 when computed over the sample 1955-2008. This evidence is confirmed by the implementation of the Johansen (1991) test on a bivariate VAR for p_t and d_t, that does not reject the null hypothesis of at most zero cointegrating vectors over the full-sample and the post-war subsample.

[2]These results are confirmed by the search for possible structural breaks in the cointegrating relationship based on the application of the recursive test based on the non zero-eigenvalues suggested in Hansen and Johansen (1999). The eigenvalue shows a remarkable amount of variability over the examination period with indication of three break points around 1950, 1980, 2000.

boom and busts. They conjecture that the life-cycle portfolio behavior (Bakshi and Chen (1994)), which suggests that agents should borrow when young, invest for retirement when middle-aged, and live off from their investment once they are retired, plays an important role in determining equilibrium asset prices. In their model, the demographic structure requires that when the MY ratio is small (large), there will be excess demand for consumption (saving) by a large cohort of retirees (middle-aged). For the market to clear, equilibrium prices of financial assets and therefore the dividend-price ratio should adjust, i.e. decrease (increase), so that saving (consumption) is encouraged for the middle-aged. We take the GMQ model to the data via the conjecture that fluctuations in MY_t could capture a slowly evolving mean in dp_t within the dynamic dividend growth model. Demographic trends should capture the slowly evolving mean in dp_t and then, deviations of dp_t from $\overline{\mathrm{dp}_t}$ could be used as a potential predictor for long-term stock market returns and dividend growth. Our empirical strategy has the potential for identifying separately the importance of demographic variables for high-frequency and low-frequency fluctuations in asset prices. Investigations conducted in the literature on the interaction between asset prices and demographic variables have traditionally concentrated either on high-frequency or low frequency fluctuations but have never considered an empirical framework based on the dynamic dividend growth model, capable of accommodating both of them, with a different role (see Erb et al. (1996), Poterba (2001), Goyal (2004), Ang and Maddaloni (2005) and DellaVigna and Pollet (2006)).

We first use long-run predictive regressions and cointegration analysis to assess the statistical significance of MY_t in a dynamic dividend growth model. The robustness of our results is evaluated by comparing the predictive power of the dividend-price ratio corrected for demographics with that of the dividend-price ratio, the dividend-price ratio corrected for breaks in mean (LVN) and the cash flow based net payout yield (Boudoukh et al. (2007)). The role of MY_t is then further investigated against different alternative specifications, in particular those based on cay_t and cdy_t. Finally, the availability of long-run projections for MY_t is exploited to derive predictions of long-run equity returns up to 2050.

2 Demography and the Dividend-Price Ratio: The GMQ Model

GMQ analyze an overlapping generation model in which the demographic structure mimics the pattern of live births in the US. Live births in the US have featured alternating twenty-year periods of boom and busts. They consider an OLG exchange economy with a single good (income) and three periods; young, middle-aged, retired. Each agent (except retirees) has an endowment and labor income w= (w^y,w^m,0). There are two types of financial instruments, a riskless bond and a risky asset, which allow agents to redistribute income over time. In the simplest version of the model, dividends and wages are deterministic, hence the bond and the risky asset are perfect substitutes. GMQ assume that in *odd (even)* periods a large (small) cohort N(n) enters the economy, therefore in every odd (even) period there will be {N,n,N} ({n,N,n}) cohorts living. They conjecture that the life-cycle portfolio behavior (Bakshi and Chen (1994)) which suggests that agents should borrow when young, invest for retirement when middle-aged, and live off from their investment once they are retired, plays important role in determining equilibrium asset prices.

Let q_o (q_e) be the bond price and $\{c^o_y, c^o_m, c^o_r\}$ ($\{c^e_y, c^e_m, c^e_r\}$) the consumption stream in the odd (even) period. The agent born in an odd period then faces the following budget constraint

$$c^o_y + q_o c^o_m + q_o q_e c^o_r = w^y + q_o w^m \tag{2}$$

and in an even period

$$c^e_y + q_e c^e_m + q_o q_e c^e_r = w^y + q_e w^m \tag{3}$$

Moreover, in equilibrium the following resource constraint must be satisfied

$$\begin{aligned} N c^o_y + n c^o_m + N c^o_r &= N w^y + n w^m + D \quad (4) \\ n c^e_y + N c^e_m + n c^e_r &= n w^y + N w^m + D \quad (5) \end{aligned}$$

where D is the aggregate dividend for the investment in financial markets. If q_o were equal to q_e, the agents would choose to smooth their consumption, i.e. $c^i_y = c^i_m = c^i_r$ for i = o, e, but then for values of wages and aggregate dividend calibrated from US data the equilibrium condition above would be violated leading to excess demand either for consumption or saving. To illustrate this point we refer to the calibration provided by GMQ; take N = 79, n = 69 as the size (in millions) of Baby Boom (1945-64) and Baby Bust (1965-84) generations (thus, we obtain in an even period a high MY ratio of $MY_t = \frac{N}{n} = 1.15$, and in odd period $MY_t = \frac{n}{N} = 0.87$ (see Figure 2)) and $w^y = 2$, $w^m = 3$ to match the ratio (middle to young cohort) of the average annual real income in US. We can calculate the total wage in even and odd periods using $Nw^y + nw^m$ for odd periods and $nw^y + Nw^m$ for even periods, and then given the average ratio (0.19) of dividend to wages we compute the aggregate dividends. Assuming an annual discount factor of 0.97, which translates to a discount of 0.5 in the model of 20-year periods, if $q_o = q_e = 0.5$ were to hold and agents smooth their consumption, from the budget constraint (eq. 6-7) we obtain $c^i_y = c^i_m = c^i_r = \bar{c} = 2$, but then the resource constraint (eq. 8-9) above would have been violated. For instance, an agent from the Baby Bust generation would enter in an even period in the model, i.e. (n, N, n) and high MY ratio, and faces the following aggregate resource constraint: $n(c^e_y - w^y) + N(c^e_m - w^m) + nc^e_r - D = 69 \times (2-2) + 79(2-3) + 69 \times 2 - 70 = -11$, where $D = 0.19(\frac{375+365}{2}) = 70$. This leads to excess saving in the economy. For equilibrium conditions to hold, the model implies that asset prices should increase and hence discourage saving in the economy (the experience we observed during the 90's in US). When the MY ratio is small (large), i.e. an odd (even) period, there will be excess demand for consumption (saving) by a large cohort of retirees (middle-aged) and for the market to clear, equilibrium prices of financial assets should adjust, i.e. decrease (increase), so that saving (consumption) is encouraged for the middle-aged. Thus, letting q_t^b be the price of the bond at time t, in a stationary equilibrium, the following holds

$$\begin{aligned} \mathrm{q}_t^b &= \mathrm{q}_o \text{ when period odd} \\ \mathrm{q}_t^b &= \mathrm{q}_e \text{ when period even} \end{aligned}$$

together with the condition $q_o < q_e$. Moreover the model predicts a positive correlation between MY and market prices, consequently a negative correlation with the dividend yield.

So, since the bond prices alternate between q_o and q_e, then the price of equity must also alternate between q_t^e and q_t^e as follows

$$\begin{aligned} \mathrm{q}_o^{eq} &= \mathrm{Dq}_o + \mathrm{Dq}_o\mathrm{q}_e + \mathrm{Dq}_o\mathrm{q}_e\mathrm{q}_o + ... \\ \mathrm{q}_e^{eq} &= \mathrm{Dq}_e + \mathrm{Dq}_e\mathrm{q}_o + \mathrm{Dq}_e\mathrm{q}_o\mathrm{q}_e + ... \end{aligned}$$

which implies

$$\begin{aligned} \mathrm{DP}_o &= \frac{\mathrm{D}}{\mathrm{q}_o^{eq}} = \frac{1 - \mathrm{q}_o\mathrm{q}_e}{\mathrm{q}_o\mathrm{q}_e + \mathrm{q}_o} \\ \mathrm{DP}_e &= \frac{\mathrm{D}}{\mathrm{q}_e^{eq}} = \frac{1 - \mathrm{q}_o\mathrm{q}_e}{\mathrm{q}_o\mathrm{q}_e + \mathrm{q}_e} \end{aligned}$$

where DP_o (DP_e) is the dividend-price ratio implied by low (high) MY in the model for odd (even) periods.

3 The Empirical Evidence

The GMQ model provides a foundation for a long-run negative relationship between the dividend-price ratio and demography. GMQ define the empirical counterpart of the MY_t ratio as the proportion of the number of agents aged 40-49 to the number of agents aged 20-29, which serves as a sufficient statistic for the whole population pyramid. We report the MY_t ratio in Figure 2. Interestingly, this variable shows highly persistent dynamics and a twin peaked behavior, with peaks and troughs around 1950, 1980, 2000, close to the break points in $\overline{\mathrm{dp}_t}$ identified by LVN.

To combine the GMQ model with the dynamic dividend growth we consider the derivation of LVN, who allow for a time varying mean in the linearization and consider MY_t as the potential determinant of this slowly evolving process.

$$\begin{aligned} \mathrm{dp}_t &= \overline{\mathrm{dp}_t} + \sum_{j=1}^{\infty} \rho^{j-1} \mathrm{E}_t[(\mathrm{h}^s_{t+j} - \bar{\mathrm{h}}) - (\Delta \mathrm{d}_{t+j} - \bar{\mathrm{d}})] \\ \overline{\mathrm{dp}_t} &= \beta_0 + \beta_1 \mathrm{MY}_t + \mathrm{u}_t \end{aligned} \tag{6}$$

Inserting GMQ into the dynamic dividend growth model leads to the prediction that the (log) dividend-price adjusted for demographics should be significant in the long-horizon forecasting regression for real stock market returns, the real dividend growth, and their difference. MY_t should also be significant in explaining the persistence of the dividend-price, and the variable predicted to be stationary in this extended model is not the dividend-price but a combination between price, dividends and MY_t. We investigate the hypothetical cointegrating relation between dividend, prices and MY_t, by running the Johansen (1988) procedure on a cointegrating system based on the vector of variables $\mathbf{y}_t' = \begin{bmatrix} \mathrm{d}_t & \mathrm{p}_t & \mathrm{MY}_t \end{bmatrix}$.

3.1 Long-Horizon Forecasting Regressions

We report in Table 1, 2 and 3 the evidence from the long-horizon forecasting regression. To make our evidence directly comparable with that reported in Lettau and Ludvigson (2005) we consider predictive regressions for annual data with horizons ranging from one to six years. We consider the annual data for the S&P 500 index from 1909 to 2008 taken from Robert Shiller's website, dividends are twelve-month moving sums of dividends paid on the S&P 500 index. These series coincide with those used in Goyal and Welch (2008), and made available at Amit Goyal's website. A full description of all data used in our empirical analysis is provided in the Data Appendix.

Table 1, 2 and 3 report the evidence for forecasting returns, dividend growth, returns adjusted for dividend growth, based on the following three models:

$$\begin{aligned}
\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}) &= \beta_0+\beta_1\mathrm{p}_t+\beta_2\mathrm{d}_t+\beta_3\mathrm{MY}_t+\varepsilon_{t,t+j}\\
\sum_{j=1}^{k}(\Delta\mathrm{d}_{t+j}) &= \beta_0+\beta_1\mathrm{p}_t+\beta_2\mathrm{d}_t+\beta_3\mathrm{MY}_t+\varepsilon_{t,t+j}\\
\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}-\Delta\mathrm{d}_{t+j}) &= \beta_0+\beta_1\mathrm{p}_t+\beta_2\mathrm{d}_t+\beta_3\mathrm{MY}_t+\varepsilon_{t,t+j}\\
\mathrm{k} = & \quad 1,...,6
\end{aligned}$$

In each forecasting regression MY_t is measured at the end of the forecasting period. We report heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) to account for overlapping observations where the bandwidth has been selected following the procedure described in Newey and West (1994). Alternatively, we also conduct a (wild) bootstrap exercise (Davidson and Flachaire (2008)) to compute p-values. To take care of the potential effect on statistical inference in finite sample of the use of overlapping data we also report the rescaled t-statistic recommended by Valkanov (2003) for the hypothesis that the regression coefficient on the dividend-price adjusted for the effect of demographics is zero. We report test of predictability at each horizon but we also compute joint tests across horizons based on SUR estimation and report in the last row the relevant χ^2 statistics with associated p-values.

The evidence can be summarized as follows:

i) MY_t is always significant along withp_t and d_t in all the forecasting regressions for real stock market returns (Panel A). The adjusted R^2 of the predictive regression increases with the horizon from 0.09 at the 1-year horizon to 0.54 at the 6-year horizon. Consistently with the prediction of the GMQ model, the effect of MY is negative on the slowly evolving mean of the dividend-price and hence positive for expected returns at all horizons.

ii) MY_t is never significant in the forecasting regressions for real dividend growth (Panel B). The adjusted R^2 of the predictive regression declines with the horizon from 0.15 at the 1-year horizon to 0.06 at the 6-year horizon.

iii) MY_t is always significant along with p_t and d_t in all the forecasting regressions for real stock market returns adjusted for real dividend growth (Panel C). The adjusted R^2 of the predictive regression increases with the horizon from 0.26 at the 1-year horizon to 0.67 at the 6-year horizon. The evidence of the strongest predictability of $\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}-\Delta\mathrm{d}_{t+j})$ is fully consistent with the dynamic dividend growth model. Such evidence, paired with that on the different forecastability of the two components of stock market returns adjusted for dividend growth, rules out the dominance of a common stochastic component for the determination of the dynamics of dividend growth and stock market returns.

iv) MY_t dominates alternative approaches proposed in the literature to capture an evolving mean in the dividend-price ratio. In the last rows of each panel of the Tables 1, 2 and 3 we report the results of augmenting the long-run forecasting regressions based on the GMQ model with alternative filtered dividend-price series. In particular we consider, dp_t^{LVN}, the (log) dividend-price corrected for breaks in LVN and dp_t^{BMRR}, the cash flow based net payout yield (dividends plus repurchases minus issuances) proposed by Boudoukh et al. (2007).

Overall the long-run forecasting regressions lend strong support to the inclusion of MY_t in the traditional dynamic dividend growth model. The dividend-price corrected for a slowly long-run mean, determined by MY_t, predicts long-run stock market returns and long-run stock market returns adjusted for dividend growth, but it does not predict long-run dividend growth. The R^2 associated to the relevant predictive regressions increases with the horizon. This evidence of a positive relation between predictability and the forecasting horizon is interesting, in that both the dynamic dividend growth model and the GMQ model establish a predictive relation for long-run returns. In fact, the most natural horizon for the GMQ model is one generation, i.e. about twenty years. Of course, it is difficult to establish some evidence via predictive regressions for twenty years returns, as we have only one century of data. To give the reader a visual impression on the relationship between real stock market returns and MY_t at a frequency as close as possible to that implied by the relevant models, we report in Figure 3 MY_t and 20-year real stock market returns. We find the graphical evidence interesting and fully consistent with the statistical evidence from the long-run regressions at higher frequencies.

3.2 Cointegration

The evidence of forecasting power of a linear combination of dividend, prices and MY_t for forecasting long-run returns and long-run returns adjusted for dividend growth, provides indirect evidence of stationarity of such a combination. The validity of this hypothesis can be further investigated by running the Johansen (1988, 1991) procedure on a cointegrating system based on the vector of variables $\mathbf{y}_t' = \left[\begin{array}{ccc} \mathrm{d}_t & \mathrm{p}_t & \mathrm{MY}_t \end{array} \right]$. We then test for cointegration within a three-variate VAR[3].

[3]See Appendix A for the details of the specification of our statistical model. In a previous version of this paper we allow for a presence of a technology-driven trend (GMQ, p.6), proxied by Total Factor Productivity, in the long-run equilibium relationship. We have decided to exclude TFP from the cointegrating relationship

We report in Table 4 and 5 the evidence for the full-sample and for the sub-sample 1955-2008. The results lead to the rejection of the null of at most zero cointegrating vectors, while the null of at most one cointegration vector cannot be rejected. The evidence in favour of one cointegrating vector in which all variables are always significant confirms that the high persistence of the dividend-price is matched by the high persistence of MY_t. Using the augmented Dickey-Fuller test, the null of a unit root in MY cannot be rejected. The coefficients determining the adjustment in presence of disequilibrium in the Vector Error Correction model confirm the evidence from the forecasting regressions reported in the previous section: stock market returns adjust in presence of disequilibrium. The significance of MY_t increases in the second sub-sample, where LVN found the two breaks in dp_t[4].

Figures 4.A - 4.B provide a graphical assessment of the capability of MY_t of capturing the slowly evolving mean of dp_t. Figure 4.A reports the residuals from our cointegrating vector, along with dp_t, and the deviations of dp_t from $\overline{\mathrm{dp}}^{LVN}$, the shifting mean identified by LVN. Figure 4.B reports residuals from our cointegrating vector with the cycle of dp_t, obtained by applying an Hodrick-Prescott filter to the original series. The graphical evidence illustrates how the cointegration based correction matches the break-based correction in LVN (2008) and the cycle obtained by applying the HP filter. It is important to note that while the cointegration based analysis can be promptly used for forecasting, the same does not apply to both the HP filter and the correction for breaks.

Overall we take the evidence of long-run forecasting regressions and cointegration analysis as consistently supportive of the GMQ model. Two more remarks are in order before we move forward.

First, in the GMQ model, bond and stock are perfect substitutes, therefore the evaluation of the performance of MY_t in forecasting yields to maturity of long-term bonds seems a natural extension of our empirical investigation. In fact, the debate on the so-called FED model (Lander et al. (1997)) of the stock market, based on a long-run relation between

on the basis of two arguments i) the presence of a technology driven trend in the dividend pice ratio is very hard to justify theoretically ii) the TFP trend does not attract any significance when included in the long-run forecasting regressions discussed in the previous section. We are grateful to an anonymos referee for attracting our attention on this point.

[4]We have also investigated the stability of the cointegrating relationship by using the recursively calculated eigenvalues and the tests for constancy of the parameters in the cointegrating space proposed by Nyblom (1989), Hansen and Johansen (1999) and Warne et al. (2003). The results, available upon request, show no evidence of instability.

the price-earning ratio and the long-term bond yield, brings some interesting evidence on this issue. The FED model is based on the equalization, up to a constant, between long-run stock and bond market returns. This feature is shared by the GMQ framework, and it requires a constant relation between the risk premium on long-term bonds and stocks. It has been shown that, although the FED model performs well in periods where the stock and bond market risk premia are strongly correlated, some measure of the fluctuations in their relative premium is necessary to model periods in which volatilities in the two markets have been different (see, for example, Asness (2003)). As a consequence, to put MY_t at work to explain bond yields, some modelling of the relative bond/stock risk premia is also in order. We consider this as an interesting extension beyond the scope of this paper which is on our agenda for future work.

Second, although MY_t is the GMQ model consistent measure of demographics, there are a number of different potential measures for demographic trends. We have therefore conducted robustness analysis of our cointegration results to the introduction of different measures of demographic structure of the population and productivity trends. The results, discussed in Appendix B, are supportive of our preferred specification.

4 MY, CAY and CDY

In the light of the evidence reported in the previous section it is interesting to reconsider point iv) in the introduction and evaluate the significance of the introduction of MY_t in the dynamic dividend growth model against cay_t and cdy_t. As stated in the introduction, Lettau and Ludvigson (2001, 2005) have found that dividend growth and stock returns are predictable by long-run equilibrium relationships derived from a linearized version of the consumer's intertemporal budget constraint. The excess consumption with respect to its long run equilibrium value is defined by Lettau and Ludvigson (2001) alternatively as a linear combination of labour income and financial wealth, cay_t, or as a linear combination of aggregate dividend payments on human and non-human wealth, cdy_t. cay_t and cdy_t are much less persistent time-series than dp_t, they are predictors of both stock returns and dividend-growth, and when included in a predictive regression relating stock market returns to dp_t, they swamp the significance of this variable.

Evaluating the effect of the inclusion of cay_t and cdy_t in the long-run forecasting regressions that also include MY_t is important for a number of reasons. First, it is a parsimonious way of evaluating the model with MY_t against all financial ratios traditionally adopted to predict returns. In fact, Lettau and Ludvigson (2001,2005) show the superior performance in predicting long-run returns of cay_t and cdy_t with respect to all the traditionally adopted financial ratios, such as the detrended short term interest rate (Campbell (1991), Hodrick

(1992)), the log dividend earnings ratio and the log price earning ratio (Lamont (1998)), the spread of long term bond yield (10Y) over 3M Treasury bill, and the spread between the BAA and the AAA corporate bond rates. Second, it would allow further investigation on the presence of a common component in dividend and stock market returns suggested by Lettau and Ludvigson (2005) but not consistent with our findings in Table 3, that witness the significance of MY_t for predicting long-run returns and long-run returns adjusted for dividend growth. Third, it could shed further light on the relative importance of cay_t and cdy_t and MY_t for predicting returns and dividend growth in the dynamic dividend growth model. Note that a joint significance of cay_t or cdy_t and MY_t in long-run forecasting regressions for real stock market returns is fully consistent with the GMQ model if the significance of MY_t is interpreted in the light of its role as a predictor for $\overline{\mathrm{dp}_t}$ while cay_t or cdy_t are taken as predictors of long term expectations of real returns and dividend growth.

We report the relevant evidence in Tables 6, 7 and 8. cay_t and cdy_t are estimated by Lettau and Ludvigson (2001,2005) as cointegrating residuals for the systems (c_t, a_t, y_t) and (c_t, d_t, y_t), where c_t is log consumption, y_t is log labor income, a_t is log asset wealth (net worth), d_t is log stock market dividends. We have taken the cointegrating relationship directly from Lettau and Ludvigson (2005): $\mathrm{cay}_t = \mathrm{c}_t - 0.33\mathrm{a}_t - 0.57\mathrm{y}_t$, $\mathrm{cdy}_t = \mathrm{c}_t - 0.13\mathrm{d}_t - 0.68\mathrm{y}_t$. The evidence clearly indicates that the significance of dp_t corrected for MY_t in the long-horizon regressions is not reduced by the augmentation of the model with cay_t and cdy_t. These two variables, and in particular cdy_t, have strong predictive power for dividend growth. Therefore, the evidence that the best predictive model for long-horizon stock returns is the one combining dividend-price with the demographic variable and cdy_t is indeed fully consistent with an interpretation based on the Dynamic Dividend Growth model where MY_t explains the slowly evolving component of the mean of the dividend-price and cdy_t acts as a predictor of dividend-growth. Such an interpretation is supported by the long-horizon regressions for stock returns adjusted for dividend growth, in which both MY_t and cdy_t enter with highly significant coefficients of the opposite sign, positive for MY_t and negative for cdy_t.

In Figure 5.A and 5.B we plot $\mathrm{dp}_t - \overline{\mathrm{dp}_t}$ against cay_t and cdy_t, respectively. We derive $\overline{\mathrm{dp}_t}$ by using the coefficients from Table 1 (k=3)[5], while cay_t and cdy_t series are taken from Lettau and Ludvigson (2005). The graph shows positive but not too strong correlation between $\mathrm{dp}_t - \overline{\mathrm{dp}_t}$ and cay_t (cdy_t), of 0.57 (0.18). This evidence is consistent with our

[5] We take 3-year horizon as a representative example, results for other horizons remain qualitatively similar (available upon request). We use the coefficients from table Table 1, the results are very similar when Table 6 is used instead.

inclusion of MY_t in the dynamic dividend growth model and the derivation of cay_t and cdy_t from the consumer's intertemporal budget constraint. Consider for example $\mathrm{dp}_t - \overline{\mathrm{dp}_t}$ and cay_t, they have a common component, which is the weighted sum of future returns, but they are also determined by idiosyncratic components: future dividend growth and future consumption growth, respectively.

4.1 Out-of-Sample Evidence

In this section we follow Goyal and Welch (2008), and analyze the performance of cdy_t and MY_t adjusted dividend-price ratio from the perspective of a real-time investor. We therefore consider out-of-sample evidence, for the 1-year, 2-year, and 3-year horizons, and we compare the performance of the bivariate model based on the combination of the two predictors with that of the two univariate models based on each predictor and the univariate models based on dp_t and dp_t^{LVN}.

We run rolling forecasting regressions for the one, three and five years ahead horizon by using 1955-1981 as an initialization sample. The forecasting period beginning in 1982 includes the anomalous period of late 90's where the sharp increase in the stock market index weakens the forecasting power of financial ratios. In particular, we consider both the univariate models and the bivariate encompassing model; we compare the forecasting performance with the historical mean benchmark. In the first two columns of Table 9 we report the adjusted $\bar{\mathrm{R}}^2$ and t-statistics using the full sample 1955-2008. Then we report mean absolute error (MAE) and root mean square error (RMSE) calculated based on the residuals in the forecasting period, namely 1982-2008. The first column of the out-of- sample panel reports the out-of-sample R^2 statistics (Campbell and Thomson (2008)) which is computed as

$$\mathrm{R}^2_{OS} = 1 - \frac{\sum_{t=t_0}^{T}(\mathrm{r}_t - \hat{\mathrm{r}}_t)^2}{\sum_{t=t_0}^{T}(\mathrm{r}_t - \bar{\mathrm{r}}_t)^2}$$

where $\hat{r}_t$ is the forecast at t -1 and $\bar{r}_t$ is the historical average estimated until t -1. In our exercise, $\mathrm{t}_0 = 1982$ and T= 2008. If R^2_{OS} is positive, it means that the predictive regression has a lower mean square error than the prevailing historical mean. In the last column, we report the Diebold-Mariano (DM) *t*-test for checking equal-forecast accuracy from two nested models for forecasting h-step ahead excess returns.

$$\mathrm{DM} = \sqrt{\frac{(\mathrm{T} + 1 - 2 * \mathrm{h} + \mathrm{h} * (\mathrm{h} - 1))}{\mathrm{T}}} * \left[\frac{\bar{\mathrm{d}}}{\widehat{\mathrm{se}}(\bar{\mathrm{d}})}\right]$$

where we define e^2_{1t} as the squared forecasting error of prevailing mean, and e^2_{2t} as the squared forecasting error of the predictive variables, $\mathrm{d}_t = \mathrm{e}^2_{1t} - \mathrm{e}^2_{2t}$, i.e. the difference between the two forecast errors, $\bar{\mathrm{d}} = \frac{1}{\mathrm{T}} \sum_{t=t_0}^{T} \mathrm{d}_t$ and $\widehat{\mathrm{se}}(\bar{\mathrm{d}}) = \frac{1}{\mathrm{T}} \sum_{\tau=-(h-1)}^{h-1} \sum_{t=|\tau|+1}^{T} (\mathrm{d}_t - \bar{\mathrm{d}}) * (\mathrm{d}_{t-|\tau|} - \bar{\mathrm{d}})$. A positive DM t-test statistic indicates that the predictive regression model performs better than the historical mean.

We report in Figure 6 the cumulative squared prediction errors of the historical mean minus the cumulative squared prediction error of our best forecasting model

We use all available data from 1910 until 1954 for initial estimation and then recursively calculate the cumulative squared prediction errors until the sample end, namely 2008.

Overall, the results reported in Table 9 and Figure 6 confirm the evidence from the forecasting regressions, with a clear indication that the model combining cdy_t and MY_t adjusted dividend-price ratio dominates all alternative specifications, both within-sample and out-of-sample.

5 Long-Run Equity Premium Projections

An interesting feature of MY_t is that long-run forecasts for this variable are readily available. In fact, the Bureau of Census (BoC) provide projections up to 2050 for MY_t. In this section we combine a long-run horizon regression with the cointegrating system estimated in section 2 to construct a model that can be simulated to generate long-run equity premium projections.

We concentrate on 5-year excess returns and estimate the following model:

$$
\begin{aligned}
\sum_{j=1}^{5} (\mathrm{h}_{t+j}^{s} - \mathrm{r}_{f,t+j}) &= c_1 + c_2 (\mathrm{p}_t - c_3 \mathrm{d}_t - c_4 \mathrm{MY}_t) + \mathrm{u}_{1t} \qquad (7) \\
\begin{bmatrix} \Delta \mathrm{p}_{t+1} \\ \Delta \mathrm{d}_{t+1} \end{bmatrix} &= \begin{bmatrix} c_5 \\ c_{10} \end{bmatrix} + \begin{bmatrix} c_6 \\ c_{11} \end{bmatrix} \begin{bmatrix} 1 & -c_3 & -c_4 \end{bmatrix} \begin{bmatrix} \mathrm{p}_t \\ \mathrm{d}_t \\ \mathrm{MY}_t \end{bmatrix} + \\
& \begin{bmatrix} c_7 & c_8 & c_9 \\ c_{12} & c_{13} & c_{14} \end{bmatrix} \begin{bmatrix} \Delta \mathrm{p}_t \\ \Delta \mathrm{d}_t \\ \Delta \mathrm{MY}_t \end{bmatrix} + \begin{bmatrix} \mathrm{u}_{2t} \\ \mathrm{u}_{2t} \end{bmatrix},
\end{aligned}
$$

(7) Combines a long-run forecasting regression for five year excess returns, defined as the difference between returns on the S&P500 and the risk-free rate[6], with the equations for $\Delta \mathrm{p}_{t+1}$, $\Delta \mathrm{d}_{t+1}$ in the cointegrated VAR estimated in Section 2. Equity premium projections are obtained by forward simulation of the first equation. This requires projections for the

[6]See the Data Appendix for a detailed description of the construction of our risk-free rate.

three right-hand side variables. We obtain them directly from the BoC for MY_t and by forward simulation of the CVAR estimated in Section 2 for p_t and d_t.Three comments are in order on the specification of (7). First, omitting an equation for MY_t from the model used for projections requires (strong) exogeneity of this variable: we believe in the validity of such an assumption. Second, we impose cross-equation restrictions in order to have the same estimates of the coefficients determining the long-run equilibrium of the system in the equation for excess-returns and in the equation for 1-year returns and dividend growth. Third, we did not report the results based on the inclusion of cdy_t in our forecasting model. In fact, the long-horizon forecast for this variable do rapidly converge to its historical mean to leave the variability of projections of the risk-premium to be dominated by projections for MY_t. Moreover, as pointed out by Goyal and Welch (2008), this variable might suffer from look-ahead bias, as the cointegrating coefficients are computed using full-sample estimates.

The estimates are fully consistent with those reported in Table 1 and 4[7]. Figure 7 illustrates the results from the projection of the model.

Over the sample up to 2008 we report (pseudo) out-of-sample 5-year annualized equity premium forecasts and its realizations. The model consistently performs very well with only two exceptions: the 1929 crisis and the boom market at the end of the millennium. We then conduct the out-of-sample exercise by estimating the model with data up to 2008, and then by solving it forward stochastically to obtain out-of-sample forecasts until 2050. Our simulation predicts a rapid stock market recovery for the next two years followed by fluctuations of the risk premium around a mean of 5.02 per cent, just below the historical average. The width of the 95 per cent confidence intervals points to the existence of a sizeable amount of uncertainty around point estimates. Interestingly, the model does not foresee a dramatic market meltdown, a "doomsday" scenario, due to a collective exit from the stock market by the retired baby boomers. This evidence is a natural outcome of the GMQ model which relies on the cyclicality of U.S. age structure.

[7] All the evidence reported for the long-run forecasting regressions are based on real equity returns, the dependent variable consistent with the dynamic dividend growth model. Results are robust when excess returns are used as a dependent variable instead of real returns.

6 Conclusions

This paper has documented the existence of a slowly evolving trend in the mean dividend-price ratio determined by a demographic variable, MY_t, the proportion of middle-age to young population. We have shown that MY_t captures well a slowly evolving component in the mean dividend-price ratio and it is strongly significant in long-horizon regressions for real stock market returns.

A model including MY_t overperforms all alternative models for forecasting returns. The best forecasting model for real stock market returns found in our work is the one combining MY_t with cdy_t, a variable constructed by Lettau and Ludvigson (2005) to capture excess consumption with respect to its long run equilibrium value. We take this evidence as strongly supportive of the Dynamic Dividend Growth model with an evolving mean, determined by MY_t. In fact, the model predicts that long-horizon returns should depend on the deviations of the dividend-price ratio from its mean and on long-run dividend growth. We show that MY_t models the mean of the dividend-price ratio while cdy_t is a predictor of long-horizon dividend-growth, confirming the evidence in Lettau and Ludvigson (2005). We provide evidence that an important component of time-varying expected returns is captured by allowing the mean of the dividend-price ratio to fluctuate MY_t. The importance of such a component increases with the forecasting horizon.

The empirical results we have reported should be of special relevance to the strategic asset allocation literature (e.g. Campbell and Viceira (2002)), in which the log dividend-price ratio is often used in VAR models as a stationary variable capturing time-variation in the investment opportunity set, and as an input into the optimal asset allocation decision of a long-horizon investor. In a companion paper (Favero and Tamoni (2010)) we show that allowing for the presence of MY_t in the VAR models that are used to estimate the time profile of stock market return and its volatility does cast new light on the hot debate on the safety of stock market investment for the long-run (Pastor and Stambaugh (2009)).

Finally, by exploiting the exogeneity and the predictability of MY_t, we have also provided projections for equity risk premia up to 2050. Our simulations point to an average equity risk premium of about five per cent for the period 2010-2050.

FIGURE 1 **Time Series of Log Dividend-Price Ratio**

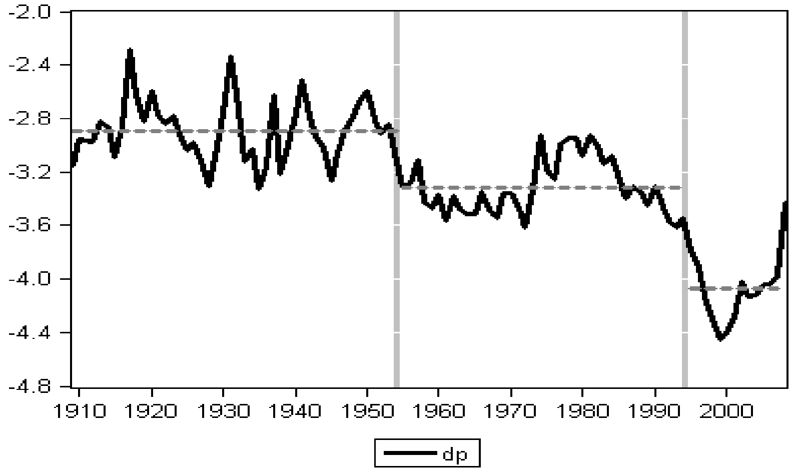


Figure 1 plots the time series of log dividend-price ratio (dp_t). Two vertical lines indicate the breaks in 1954 and 1994 identified by LVN. Sample 1909 - 2008. Annual data.

FIGURE 2 **Time Series of Middle-Young (MY) Ratio**

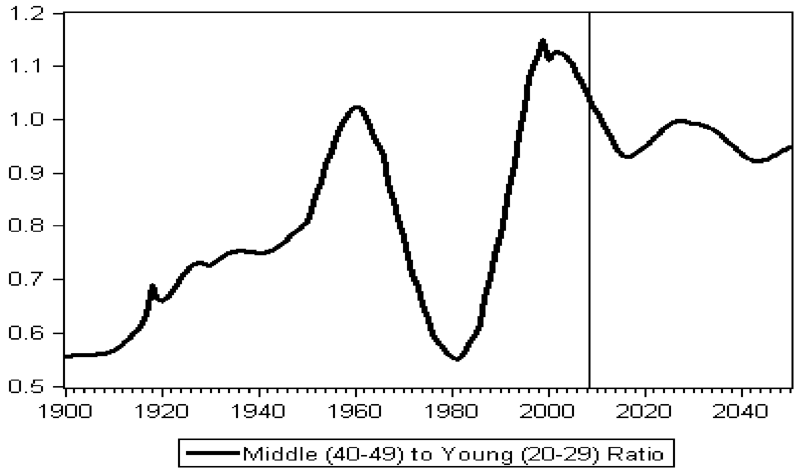


Figure 2 plots the time series of middle-young (MY) ratio. The vertical line in 2008 indicates the end of in-sample data and the start of Bureau of Census projections. Sample 1909-2050. Annual data.

FIGURE 3 **MY and 20-year Annualized Real US Stock Market Returns**

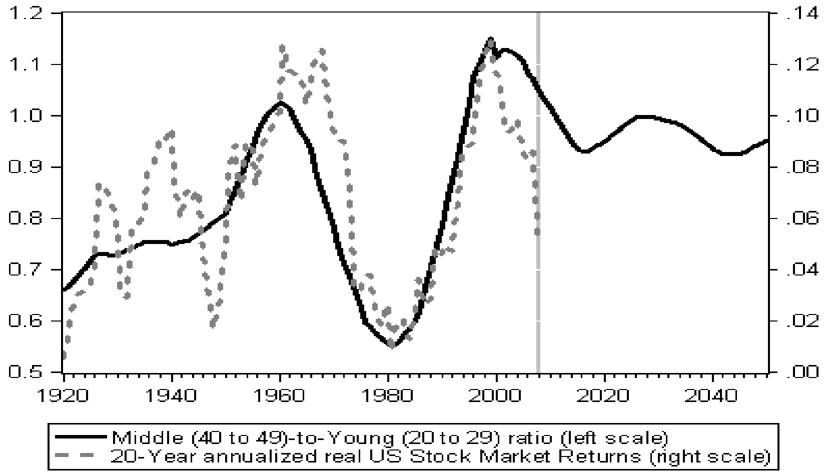
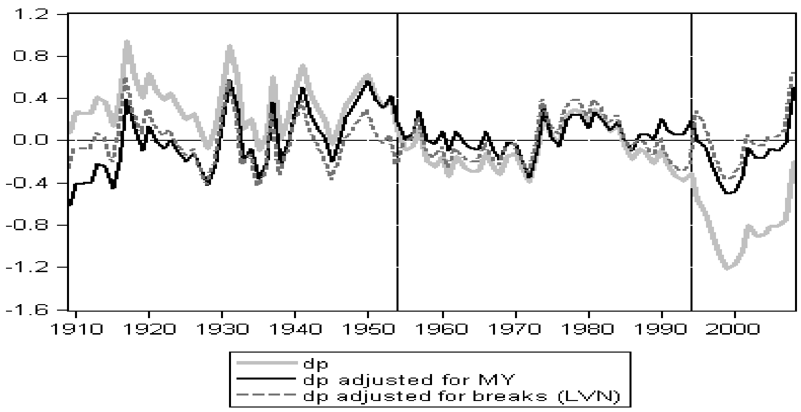


Figure 3 plots the middle-young ratio (MY) and the annualized real US stock market returns. The vertical line in 2008 indicates the end of in-sample data and the start of Bureau of Census projections. Sample 1920-2050. Annual data.

FIGURE 4

Graph A. Alternative Measures of the Cycle in dp$_t$



Graph B. Hodrick and Prescott (HP) Filtered Cycle in dp$_t$

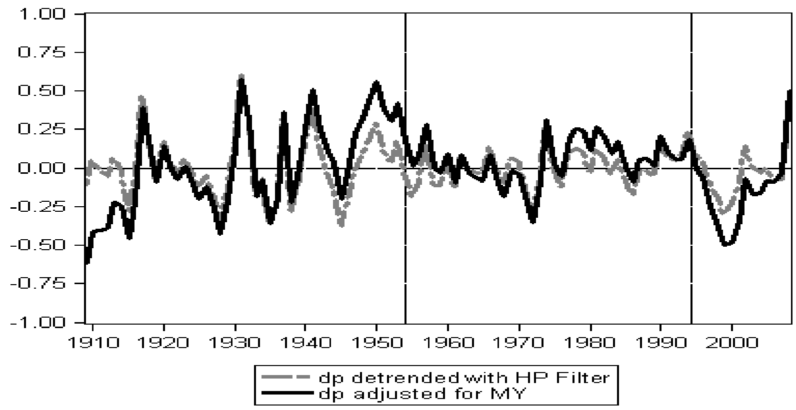
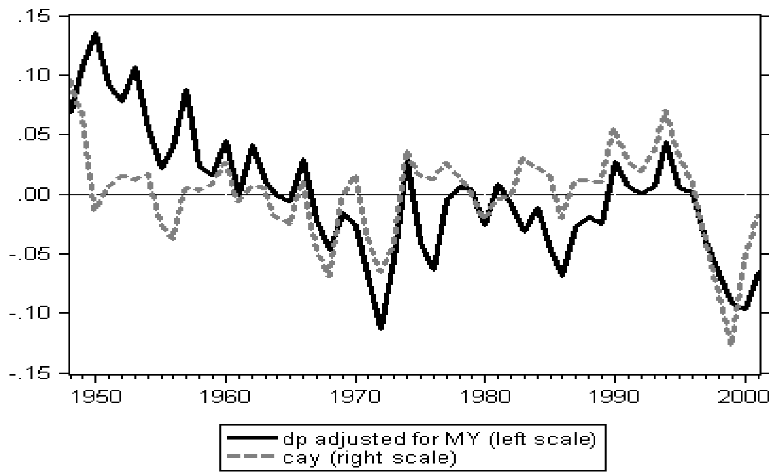


Figure 4.A plots dividend-price ratio, dp$_t$, dp$_t$ adjusted for breaks (LVN) and fluctuations of dp$_t$ around a time-varying mean determined by MY$_t$. We estimate a vector error correction model following Johansen procedure to determine the cointegrating vector between dp$_t$ and MY$_t$ (see Table 4 Panel A). Figure 4.B illustrates an alternative measure of the cycle in dp$_t$ using Hodrick and Prescott (HP) filter with a smoothing parameter equal to 100 (Jaimovich and Siu (2008)). Two vertical lines indicate the breaks in 1954 and 1994 identified by LVN. Sample 1910-2008. Annual data.

FIGURE 5

Graph A. cay$_t$ vs. dp$_t$ adjusted for MY$_t$



Graph B. cdy$_t$ vs. dp$_t$ adjusted for MY$_t$

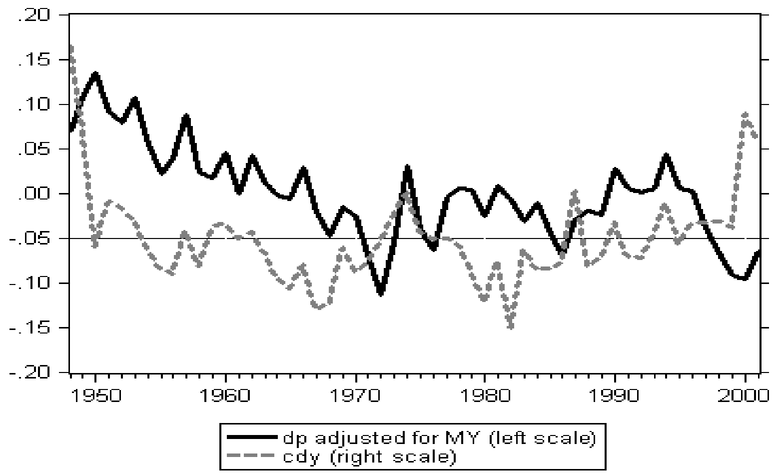


Figure 5.A plots cay$_t$ and dp$_t$ adjusted for MY$_t$. Figure 5.B plots cdy$_t$ and dp$_t$ adjusted for MY$_t$. cay$_t$ and cdy$_t$ are the annual series taken from Martin Lettau's website, dp$_t$ is adjusted for MY$_t$ using the coefficients estimated from predictive regressions reported in Table 1 (k=3). Sample 1910-2008. Annual data.

FIGURE 6 **Out-of-Sample Predictive Performance**

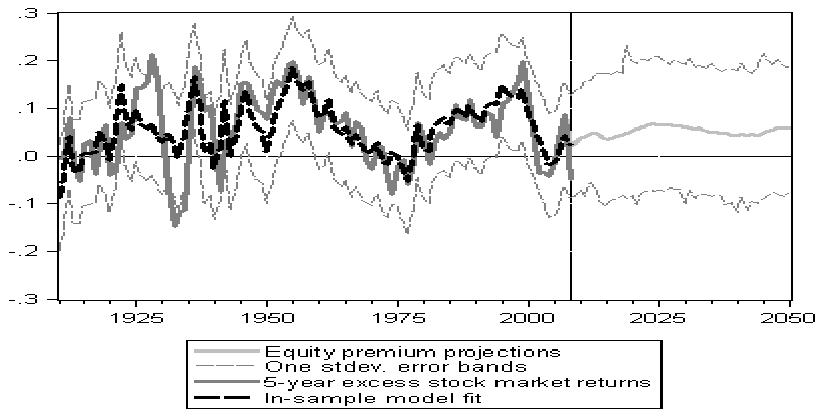


Figure 6 plots the difference between the cumulative RMSE of forecasts based on the historical prevaling mean and forecasting models based on either dp_t (dashed line) or dp_t adjusted for MY (solid line). The estimation sample is 1910-1954 and the forecasts cover the period 1955-2008. Annual data.

FIGURE 7 **Long-Run Equilibrium Projections**

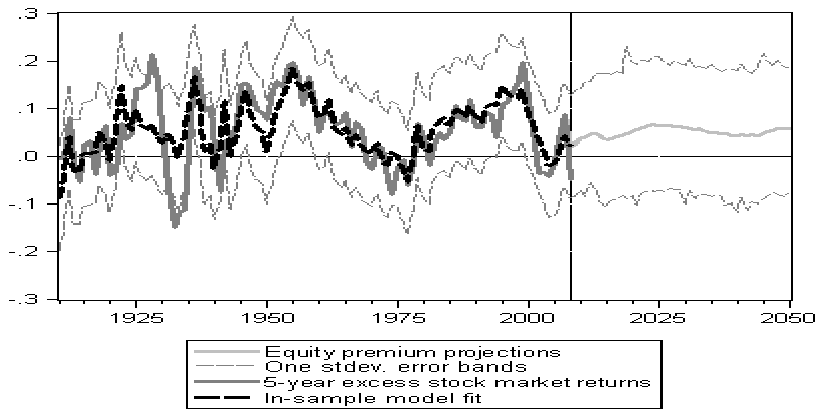


Figure 7 plots 5-year stock market return (solid dark gray line), in-sample prediction (dashed black line) and out-of-sample projections for excess returns (solid gray line) along with 95% confidence intervals (dashed gray lines). The vertical line in 2008 indicates the end of in-sample data and the start of the projections. Sample 1910-2050. Annual data.

TABLE 1

Long-horizon regression. Sample 1910 -2008. Annual data.

Panel A. k-period regressions for real stock returns

$$\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}) = \beta_0 + \beta_1 \mathrm{p}_t + \beta_2 \mathrm{d}_t + \beta_3 \mathrm{MY}_t + \varepsilon_{t,t+j}$$

χ^2 (t-stat) 15.63 (0.000) horizon k in years

	1	2	3	4	5	6
β_1 (t-stat)	−0.27 (−4.000)	−0.28 (−6.287)	−0.23 (−6.563)	−0.21 (−9.077)	−0.19 (−9.352)	−0.16 (−7.824)
β_2 (t-stat)	0.30 (3.679)	0.32 (5.748)	0.26 (5.801)	0.24 (7.974)	0.21 (8.291)	0.17 (7.065)
β_3 (t-stat)	0.44 (4.065)	0.45 (5.875)	0.37 (6.509)	0.34 (7.627)	0.29 (7.865)	0.26 (7.710)
$\beta_1 = -\beta_2$ (t-stat / w.b.p-values)	−0.19 (−3.645 / 0.002)	−0.20 (−5.439 / 0.000)	−0.16 (−6.016 / 0.000)	−0.16 (−7.652 / 0.000)	−0.14 (−8.584 / 0.002)	−0.13 (−7.740 / 0.008)
$t/\sqrt{T}$ - test	{0.30*}	{0.49**}	{0.59**}	{0.75**}	{0.89**}	{0.95**}
$\beta_3 \mid (\beta_1 = -\beta_2)$ (t-stat)	0.42 (3.447)	0.44 (4.636)	0.38 (5.059)	0.36 (6.006)	0.33 (6.578)	0.29 (6.915)
Adj. R^2	0.09	0.25	0.33	0.44	0.52	0.54
Adj. $R^2(\beta_1 = -\beta_2)$	0.07	0.20	0.27	0.37	0.46	0.50
F-statistic	4.10	12.18	16.74	26.62	36.74	39.48

Panel B. Testing MY against alternative models

$$\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}) = \beta_0 + \beta_1 \mathrm{p}_t + \beta_2 \mathrm{d}_t + \beta_3 \mathrm{MY}_t + \beta_4 \mathrm{Z}_t + \varepsilon_{t,t+j}$$

$\mathrm{Z}_t =$		1	2	3	4	5	6
dp_t^{LVN}	β_3 (t−stat)	0.44 (4.149)	0.45 (6.138)	0.39 (7.242)	0.36 (9.009)	0.32 (10.383)	0.28 (10.449)
	β_4 (t−stat)	0.07 (0.665)	0.09 (1.128)	0.05 (0.665)	0.03 (0.626)	0.01 (0.140)	−0.03 (−0.636)
dp_t^{BMRR}	β_3 (t−stat)	0.44 (2.850)	0.46 (4.849)	0.39 (6.229)	0.35 (8.561)	0.31 (10.286)	0.27 (10.925)
	β_4 (t−stat)	0.51 (3.602)	0.37 (2.945)	0.21 (2.443)	0.09 (1.496)	0.02 (0.407)	−0.03 (−1.103)

Table 1 reports in parentheses t-statistic based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994). For the univariate model with the restriction ($\beta_1 = -\beta_2$), $\sum_{j=1}^{k}(\mathrm{h}^s_{t+j}) = \beta_0 + \beta_1(\mathrm{p}_t - \mathrm{d}_t + \frac{\beta_3}{\beta_1}\mathrm{MY}_t) + \varepsilon_{t,t+j}$, we also conduct a (wild) bootstrap exercise (Davidson and Flachaire (2008)) to compute p-values and report for β_1 t/$\sqrt{\mathrm{T}}$-test suggested in Valkanov (2003) in curly brackets. Significance at the 5% and 1% level of the t /$\sqrt{\mathrm{T}}$ test using Valkanov's (2003) critical values is indicated by * and **, respectively. The χ^2 statistics with associated p-values is for the joint tests of significance across all different horizon within a SUR estimation framework.

TABLE 2

Long-horizon regressions. Sample 1910 - 2008. Annual data.

Panel A. k-period regressions for real dividend-growth

$$\sum_{j=1}^{k}(\Delta d_{t+j}) = \beta_0 + \beta_1 p_t + \beta_2 d_t + \beta_3 MY_t + \varepsilon_{t,t+j}$$

χ^2 (t-stat): 2.66 (0.015) — horizon k in years

	1	2	3	4	5	6
β_1	0.19	0.10	0.05	0.03	0.02	0.01
(t-stat)	(2.097)	(1.827)	(1.400)	(1.222)	(1.006)	(0.803)
β_2	−0.23	−0.13	−0.07	−0.04	−0.03	−0.02
(t-stat)	(−2.196)	(−1.996)	(−1.585)	(−1.441)	(−1.308)	(−1.153)
β_3	−0.15	−0.03	0.04	0.07	0.08	0.08
(t-stat)	(−1.250)	(−0.371)	(0.780)	(1.863)	(2.428)	(2.755)
$\beta_1 = -\beta_2$	0.10	0.05	0.02	0.01	−0.00	−0.00
(t-stat / w.b.p-values)	(1.580 / 0.240)	(1.118 / 0.433)	(0.638 / 0.353)	(0.298 / 0.171)	(−0.018 / 0.096)	(−0.234 / 0.106)
$\beta_3 \mid (\beta_1 = -\beta_2)$	−0.12	−0.03	0.03	0.06	0.06	0.06
(t-stat)	(−1.118)	(−0.366)	(0.714)	(1.700)	(2.067)	(2.134)
Adj. R^2	0.15	0.09	0.05	0.04	0.05	0.06
Adj. $R^2(\beta_1 = -\beta_2)$	0.06	0.03	0.02	0.02	0.03	0.03
F-statistic	6.87	4.05	2.73	2.52	2.66	3.04

Panel B. Testing MY against alternative models

$$\sum_{j=1}^{k}(\Delta d_{t+j}) = \beta_0 + \beta_1 p_t + \beta_2 d_t + \beta_3 MY_t + \beta_4 Z_t + \varepsilon_{t,t+j}$$

Z_t =		1	2	3	4	5	6
dp_t^{LVN}	β_3	−0.09	−0.00	0.05	0.07	0.08	0.09
	(t-stat)	(−0.877)	(−0.033)	(0.993)	(1.886)	(2.511)	(2.886)
	β_4	−0.24	−0.14	−0.08	−0.05	−0.02	0.00
	(t-stat)	(−2.412)	(−2.094)	(−1.520)	(−0.942)	(−0.321)	(0.024)
dp_t^{BMRR}	β_3	−0.18	−0.05	−0.00	0.03	0.04	0.06
	(t-stat)	(−1.304)	(−0.697)	(−0.001)	(0.772)	(1.752)	(2.509)
	β_4	0.09	0.14	0.16	0.16	0.15	0.08
	(t-stat)	(0.879)	(1.981)	(2.122)	(2.206)	(2.977)	(2.304)

Table 2 reports in parentheses t-statistic based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994). For the univariate model with the restriction $(\beta_1 = -\beta_2)$, $\sum_{j=1}^{k}(\Delta d_{t+j}) = \beta_0 + \beta_1(\mathrm{p}_t - \mathrm{d}_t + \frac{\beta_3}{\beta_1}\mathrm{MY}_t) + \varepsilon_{t,t+j}$, we also conduct a (wild) bootstrap exercise (Davidson and Flachaire (2008)) to compute p-values. The χ^2 statistics with associated p-values is for the joint tests of significance across all different horizon within a SUR estimation framework.

TABLE 3

Long-horizon regressions. Sample 1910 - 2008. Annual data.

Panel A. k-period regressions for real stock returns adjusted for dividend growth

$$\sum_{j=1}^{k}(h^s_{t+j} - \Delta d_{t+j}) = \beta_0 + \beta_1 p_t + \beta_2 d_t + \beta_3 MY_t + \varepsilon_{t,t+j}$$

χ^2 (t-stat)	27.94 (0.000)	horizon k in years				
	1	2	3	4	5	6
β_1 (t-stat)	−0.46 (−6.159)	−0.39 (−8.285)	−0.28 (−9.393)	−0.24 (−10.344)	−0.20 (−14.944)	−0.17 (−12.523)
β_2 (t-stat)	0.53 (6.127)	0.45 (8.234)	0.33 (9.157)	0.28 (10.022)	0.24 (13.150)	0.20 (10.940)
β_3 (t-stat)	0.59 (4.631)	0.47 (5.313)	0.33 (5.297)	0.27 (5.697)	0.22 (6.202)	0.18 (6.361)
$\beta_1 = -\beta_2$ (t-stat / w.b.p-values)	−0.28 (−4.961 / 0.001)	−0.24 −6.603 / 0.000)	−0.18 (−7.345 / 0.000)	−0.16 (−8.516 / 0.000)	−0.14 (−13.217 / 0.000)	−0.12 (−12.976 / 0.002)
$t/\sqrt{T}$ - test	{0.47**}	{0.67**}	{0.77**}	{0.91**}	{1.04**}	{1.12**}
$\beta_3 \mid (\beta_1 = -\beta_2)$ (t-stat)	0.55 (3.610)	0.47 (4.206)	0.35 (4.204)	0.31 (4.646)	0.27 (5.160)	0.23 (5.406)
Adj. R^2	0.24	0.45	0.51	0.60	0.66	0.67
Adj. $R^2(\beta_1 = -\beta_2)$	0.16	0.29	0.35	0.43	0.50	0.53
F-statistic	11.49	27.74	35.07	50.25	65.08	67.56

Panel B. Testing MY against alternative models

$$\sum_{j=1}^{k}(h^s_{t+j} - \Delta d_{t+j}) = \beta_0 + \beta_1 p_t + \beta_2 d_t + \beta_3 MY_t + \beta_4 Z_t + \varepsilon_{t,t+j}$$

Z_t =		1	2	3	4	5	6
dp_t^{LVN}	β_3 (t-stat)	0.53 (4.111)	0.45 (5.041)	0.34 (5.453)	0.29 (6.172)	0.24 (7.314)	0.19 (8.584)
	β_4 (t-stat)	0.31 (2.618)	0.23 (2.739)	0.13 (1.915)	0.08 (1.512)	0.02 (0.513)	−0.03 (−0.719)
dp_t^{BMRR}	β_3 (t-stat)	0.62 (4.100)	0.51 (6.105)	0.39 (6.490)	0.32 (7.313)	0.27 (8.837)	0.21 (9.256)
	β_4 (t-stat)	0.42 (2.294)	0.22 (1.881)	0.05 (0.704)	−0.07 (−1.734)	−0.13 (−4.349)	−0.11 (−4.035)

Table 3 reports in parentheses t-statistic based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994). For the univariate model with the restriction ($\beta_1 = -\beta_2$), $\sum_{j=1}^{k}(h^s_{t+j} - \Delta d_{t+j}) = \beta_0 + \beta_1(p_t - d_t + \frac{\beta_3}{\beta_1}MY_t) + \varepsilon_{t,t+j}$, we also conduct a (wild) bootstrap exercise (Davidson and Flachaire (2008)) to compute p-values and report for β_1 $t/\sqrt{T}$-test suggested in Valkanov (2003) in curly brackets. Significance at the 5% and 1% level of the $t/\sqrt{T}$-test using Valkanov's (2003) critical values is indicated by * and **, respectively. The χ^2 statistics with associated p-values is for the joint tests of significance across all different horizon within a SUR estimation framework.

TABLE 4

Cointegrated VAR Estimates. Sample 1910-2008. Annual Data.

Panel A. Cointegrating vector	p_t	d_t	MY_t	C
β	-1.00	1.21	1.107	2.16
$(s.e.)$		(0.035)	(0.25)	
Panel B. Error Correction Model	Δp_t	Δd_t	ΔMY_t	
α	0.29	-0.12	0.007	
$(s.e.)$	(0.096)	(0.046)	(0.007)	
Adj. R^2	0.126	0.43	0.63	
Panel C. Cointegration Test	Trace	p-value	Max eigen	p-value
Hypothesized No of CE(s)				
None	29.68	0.05	22.86	0.028
At Most 1	6.82	0.59	6.75	0.51

TABLE 5

Cointegrated VAR Estimates. Sample 1955-2008. Annual Data.

Panel A. Cointegrating vector	p_t	d_t	MY_t	C
β	-1.00	1.248	1.14	2.07
$(s.e.)$		(0.035)	(0.14)	
Panel B. Error Correction Model	Δp_t	Δd_t	ΔMY_t	
α	0.63	-0.02	0.03	
$(s.e.)$	(0.16)	(0.035)	(0.015)	
Adj.R^2	0.22	0.40	0.75	
Panel C. Cointegration Test	Trace	p-value.	Max eigen	p-value
Hypothesized No of CE(s)				
None	28.71	0.06	19.48	0.08
At Most 1	9.22	0.34	8.81	0.30

Table 4 and 5 report the trace and max eigenvalue statistics obtained from Johansen cointegration test with linear trend in the data. We report the coefficients of the cointegrating vector β and the weights α (see Appendix A) for the whole sample (Table 4) and for the post-war sample (Table 5). The reported p-values for the relevant null to test for cointegration are McKinnon-Haugh-Michelis (1999) p-values. The lag length in the VAR model is chosen according to optimal information criteria, i.e. sequential LR test, Akaike (AIC), Schwarz (SIC), Hannan-Quinn (HQ) information criterion.

TABLE 6

cay and cdy. Sample 1948-2008. Annual Data.

k-period regressions for real stock returns

$$\sum_{j=1}^{k}(\mathrm{h}^{s}_{t+j}) = \beta_0 + \beta_1 \mathrm{p}_t + \beta_2 \mathrm{d}_t + \beta_3 \mathrm{MY}_t + \beta_4 \mathrm{z}_t + \varepsilon_{t,t+j}$$

		horizon k in years					
	z_t	1	2	3	4	5	6
β_1 $(t-stat)$	cay_t	-0.41 (-4.050)	-0.29 (-4.694)	-0.21 (-5.570)	-0.20 (-7.222)	-0.19 (-8.257)	-0.15 (-8.790)
	cdy_t	-0.51 (-6.487)	-0.40 (-6.442)	-0.29 (-6.840)	-0.24 (-8.592)	-0.21 (-11.549)	0.19 (-12.403)
β_2 $(t-stat)$	cay_t	0.50 (3.994)	0.36 (4.492)	0.25 (5.230)	0.24 (6.373)	0.22 (7.067)	0.17 (7.384)
	cdy_t	0.63 (6.210)	0.48 (6.134)	0.34 (6.287)	0.28 (7.143)	0.25 (9.064)	0.22 (9.832)
β_3 $(t-stat)$	cay_t	0.66 (4.194)	0.48 (5.077)	0.36 (6.327)	0.32 (8.823)	0.29 (10.967)	0.24 (13.000)
	cdy_t	0.781 (5.467)	0.57 (5.774)	0.42 (6.146)	0.35 (7.756)	0.31 (9.394)	0.26 (10.067)
β_4 $(t-stat)$	cay_t	2.06 (1.336)	2.41 (3.329)	1.94 (3.111)	0.87 (1.192)	0.71 (1.471)	1.15 (3.739)
	cdy_t	-0.51 (-0.718)	0.25 (0.490)	0.27 (1.102)	0.04 (0.225)	0.06 (0.498)	0.44 (3.246)
Adj. R^2	cay_t	0.35	0.61	0.70	0.75	0.82	0.87
	cdy_t	0.34	0.56	0.65	0.74	0.81	0.86

Table 6 reports the OLS estimates from k-period regressions for real stock returns. Each column reports a different horizon, odd (even) rows refer to $\mathrm{z}_t = \mathrm{cay}_t$ (cdy_t). The reported t-statistics are based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994).

TABLE 7

cay and cdy. Sample 1948-2008. Annual data.

k-period regressions for real dividend-growth

$$\sum_{j=1}^{k}(\Delta d_{t+j}) = \beta_0 + \beta_1 \mathrm{p}_t + \beta_2 \mathrm{d}_t + \beta_3 \mathrm{MY}_t + \beta_4 \mathrm{z}_t + \varepsilon_{t,t+j}$$

		horizon k in years					
		1	2	3	4	5	6
β_1 $(t{-}stat)$	cay_t	0.10 (1.649)	0.06 (1.481)	0.03 (0.977)	0.01 (0.451)	0.01 (0.268)	0.01 (0.251)
	cdy_t	−0.03 (−0.528)	−0.04 (−0.981)	−0.03 (−1.392)	−0.02 (−1.353)	−0.01 (−1.353)	−0.01 (−0.608)
β_2 $(t{-}stat)$	cay_t	−0.14 (−1.733)	−0.09 (−1.574)	−0.04 (−1.109)	−0.02 (−0.626)	−0.01 (−0.466)	−0.01 (−0.463)
	cdy_t	0.01 (0.233)	0.03 (0.698)	0.03 (1.071)	0.02 (0.940)	0.01 (0.586)	0.01 (0.284)
β_3 $(t{-}stat)$	cay_t	−0.07 (−0.838)	−0.02 (−0.363)	0.02 (0.542)	0.04 (1.286)	0.05 (1.806)	0.05 (2.406)
	cdy_t	0.02 (0.188)	0.04 (0.669)	0.05 (1.479)	0.05 (2.033)	0.04 (2.238)	0.04 (2.416)
β_4 $(t{-}stat)$	cay_t	2.92 (2.503)	2.49 (3.176)	1.62 (4.377)	0.99 (3.359)	0.61 (2.084)	0.39 (1.244)
	cdy_t	1.16 (1.561)	1.00 (2.115)	0.69 (4.500)	0.59 (4.853)	0.56 (4.272)	0.52 (3.670)
Adj. R^2	cay_t	0.33	0.42	0.40	0.28	0.20	0.19
	cdy_t	0.20	0.27	0.30	0.31	0.33	0.35

Table 7 reports the OLS estimates from k-period regressions for real dividend-growth. Each column reports a different horizon, odd (even) rows refer to $\mathrm{z}_t = \mathrm{cay}_t$ (cdy_t). The Table reports in parentheses t-statistic based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994).

TABLE 8

cay and cdy. Sample 1948-2008. Annual data.

k-period regressions for real stock returns adjusted for dividend growth

$$\sum_{j=1}^{k}(\mathrm{h}^s_{t+j} - \Delta \mathrm{d}_{t+j}) = \beta_0 + \beta_1 \mathrm{p}_t + \beta_2 \mathrm{d}_t + \beta_3 \mathrm{MY}_t + \beta_4 \mathrm{z}_t + \varepsilon_{t,t+j}$$

		horizon k in years					
		1	2	3	4	5	6
β_1 $(t-stat)$	cay_t	-0.51 (-4.796)	-0.36 (-5.417)	-0.24 (-7.799)	-0.22 (-9.769)	-0.19 (-9.506)	-0.15 (-10.203)
	cdy_t	-0.49 (-6.547)	-0.36 (-8.500)	-0.25 (-9.022)	-0.21 (-9.930)	-0.20 (-13.824)	-0.18 (-15.141)
β_2 $(t-stat)$	cay_t	0.65 (4.713)	0.45 (5.209)	0.29 (7.635)	0.26 (9.596)	0.23 (8.871)	0.18 (9.070)
	cdy_t	0.62 (6.220)	0.45 (7.653)	0.31 (7.887)	0.26 (8.109)	0.24 (10.452)	0.21 (11.669)
β_3 $(t-stat)$	cay_t	0.72 (4.487)	0.50 (5.255)	0.34 (6.670)	0.29 (9.758)	0.24 (10.237)	0.19 (10.029)
	cdy_t	0.77 (6.020)	0.53 (6.801)	0.37 (6.419)	0.31 (7.152)	0.27 (8.580)	0.21 (8.883)
β_4 $(t-stat)$	cay_t	-0.86 (-0.508)	-0.08 (-0.079)	0.329 (0.501)	-0.12 (-0.158)	0.10 (0.169)	0.757 (2.278)
	cdy_t	-1.67 (-2.428)	-0.74 (-1.295)	-0.422 (-2.141)	-0.55 (-3.636)	-0.50 (3.818)	-0.08 $-0.722)$
Adj. R^2	cay_t	0.28	0.50	0.63	0.72	0.79	0.87
	cdy_t	0.32	0.52	0.64	0.74	0.81	0.85

Table 8 reports the OLS estimates from k-period regressions for real stock returns adjusted for dividend-growth. Each column reports a different horizon, odd (even) rows refer to $\mathrm{z}_t = \mathrm{cay}_t$ (cdy_t). The Table reports in parentheses t-statistic based on heteroskedastic and autocorrelated consistent (HAC) covariance matrix estimators using Bartlett kernel weights as described in Newey and West (1987) where the bandwidth has been selected following the procedure described in Newey and West (1994).

	In-Sample				Out-of-Sample			
Panel A (k=1)	R^2	t-stat	MAE	RMSE	R^2_{OS}	MAE	RMSE	DM
dp_t	3.03	1.64	12.92	16.17	−11.22	14.54	18.60	−17.43
dp_t^{LVN}	6.36	2.64	11.93	16.20	−5.25	13.58	18.09	−8.09
cdy_t	−1.47	0.48	12.03	14.00	−16.11	13.24	15.12	−19.97
dp_t^{DT}	19.48	4.37	10.08	15.32	11.20	10.91	16.27	4.68
$dp_t^{DT}+cdy_t$	38.64	5.97/ − 0.32	8.71	10.97	28.61	9.83	11.86	14.36
Hist.Mean	−	−	12.92	16.70	−	13.40	17.63	−
Panel B (k= 2)	R^2	t-stat	MAE	RMSE	R^2_{OS}	MAE	RMSE	DM
dp_t	5.46	1.70	15.72	20.71	−55.77	24.21	30.94	−4.01
dp_t^{LVN}	15.74	2.45	14.20	19.91	−11.72	19.99	26.20	−3.30
cdy_t	2.92	1.13	16.00	21.93	−55.19	20.04	27.45	−1.33
dp_t^{DT}	48.32	7.88	12.20	17.52	35.52	14.39	19.90	3.11
$dp_t^{DT}+cdy_t$	62.63	6.32/1.40	10.58	14.17	40.85	13.16	16.94	6.08
Hist. Mean	−	−	16.19	21.91	−	18.65	24.79	−
Panel C (k= 3)	R^2	t-stat	MAE	RMSE	R^2_{OS}	MAE	RMSE	DM
dp_t	6.34	1.99	18.26	24.87	−88.56	33.98	43.59	1.32
dp_t^{LVN}	10.76	1.70	17.91	24.81	−25.89	28.40	35.61	−4.41
cdy_t	5.91	1.41	19.45	26.94	−41.43	26.64	33.70	−1.97
dp_t^{DT}	49.73	6.70	13.22	19.00	43.95	18.15	23.32	2.90
$dp_t^{DT}+cdy_t$	64.89	7.24/2.99	13.67	16.99	48.54	17.06	20.33	2.50
Hist. Mean	−	−	19.38	26.65	−	25.26	31.74	−

TABLE 9

Predictive Performance. Sample 1948-2008. Annual data.

Table 9 presents statistics on k-year ahead forecast errors (in-sample and out-of-sample) for stock returns. The first column lists the regressors in both univariate and bivariate predictive regressionsdp_t, log dividend-price ratio, dp_t^{LVN}, dp_t corrected for breaks in mean (LVN), dp_t^{DT}, dp_t adjusted for MY_t and cdy_t, cointegrated vector suggested by Lettau and Ludvigson (2005). The sample starts in 1948 and we construct first forecast in 1982. All numbers are in percent. RMSE is the root mean square error, MAE is the mean absolute error. DM is the Diebold and Mariano (1995) t-statistic for difference in MSE of the unconditional forecast and the conditional forecast. The out-of-sample R^2_{OS} compares the forecast error of the historical mean with the forecast from predictive regressions.

A The Statistical Model for Cointegration Analysis

We consider the following statistical model:

$$\begin{aligned} \mathbf{y}_t &= \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t \\ & \mathbf{y}_t \text{ is a } m \times 1 \text{ vector of variables} \end{aligned}$$

This model can be re-written as follows

$$\begin{aligned} \Delta \mathbf{y}_t &= \mathbf{\Pi}_1 \Delta \mathbf{y}_{t-1} + \mathbf{\Pi}_2 \Delta \mathbf{y}_{t-2} + ... \mathbf{\Pi}_{n-1} \Delta \mathbf{y}_{t-n+1} + \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{u}_t \\ &= \sum_{i=1}^{n-1} \mathbf{\Pi}_i \Delta \mathbf{y}_{t-i} + \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{u}_t, \end{aligned}$$

where:

$$\begin{aligned} \mathbf{\Pi}_i &= -\left(I - \sum_{j=1}^{i} \mathbf{A}_j \right), \\ \mathbf{\Pi} &= -\left(I - \sum_{i=1}^{n} \mathbf{A}_i \right). \end{aligned}$$

Clearly the long-run properties of the system are described by the properties of the matrix Π. There are three cases of interest:

1. rank $(\Pi) = 0$. The system is non-stationary, with no cointegration between the variables considered. This is the only case in which non-stationarity is correctly removed simply by taking the first differences of the variables;

2. rank $(\Pi) =$ m, full. The system is stationary;

3. rank $(\Pi) =$ k $<$ m. The system is non-stationary but there are k cointegrating relationships among the considered variables. In this case $\mathbf{\Pi} = \alpha\beta'$, where α is an (m $\times$ k) matrix of weights and β is an (k $\times$ m) matrix of parameters determining the cointegrating relationships.

Therefore, the rank of Π is crucial in determining the number of cointegrating vectors. The Johansen procedure is based on the fact that the rank of a matrix equals the number of its characteristic roots that differ from zero. The Johansen test for cointegration is based on the estimates of the two characteristic roots of $\mathbf{\Pi}$ matrix. Having obtained estimates for the parameters in the $\mathbf{\Pi}$ matrix, we associate with them estimates for the m characteristic roots and we order them as follows $\lambda_1 > \lambda_2 > ... \lambda_m$. If the variables are not cointegrated, then the rank of Π is zero and all the characteristic roots equal zero. In this case each of the expression $\ln(1-\lambda_i)$ equals zero, too. If, instead, the rank of Π is one, and $0 < \lambda_1 < 1$, then $\ln(1-\lambda_1)$ is negative and $\ln(1-\lambda_2) = \ln(1-\lambda_3) = ... = \ln(1-\lambda_m) = 0$. The Johansen test for cointegration in our bivariate VAR is based on the two following statistics that Johansen derives based on the number of characteristic roots that are different from zero:

$$\begin{aligned}\lambda_{\text{trace}}(\mathrm{k}) &= -\mathrm{T}\sum_{i=k+1}^{m}\ln\left(1-\widehat{\lambda}_i\right),\\ \lambda_{\max}(\mathrm{k},\mathrm{k}+1) &= -\mathrm{T}\ln\left(1-\widehat{\lambda}_{k+1}\right),\end{aligned}$$

where T is the number of observations used to estimate the VAR. The first statistic tests the null of at most k cointegrating vectors against a generic alternative. The test should be run in sequence starting from the null of at most zero cointegrating vectors up to the case of at most m cointegrating vectors. The second statistic tests the null of at most k cointegrating vectors against the alternative of at most k +1 cointegrating vectors. Both statistics are small under the null hypothesis. Critical values are tabulated by Johansen (1991) and they depend on the number of non-stationary components under the null and on the specification of the deterministic component of the VAR.

B Robustness Analysis for the Cointegrating Evidence

To assess the robustness of our cointegrating relationship in identifying the low frequency relation between stock market and demographics, we evaluate the effect of augmenting our baseline relation with an alternative demographic factor. Research in demography has recently concentrated on the economic impact of the *demographic dividend* (Bloom et al., 2003; Mason&Lee, 2005). The demographic dividend depends on a peculiar period in the demographic transition phase of modern population in which the lack of synchronicity between the decline in fertility and the decline in mortality typical of advanced economies has an impact on the age structure of population. In particular a high support ratio is generated, i.e. a high ratio between the share of the population in working age and the share of population economically dependent. Empirical evidence has shown that the explicit consideration of the fluctuations in the support ratio delivers significant results in explaining economic performance (see Bloom et al., 2003). The concept of *Support Ratio (SR)* has been precisely defined by Mason and Lee (2005) as the ratio between the number of effective number of producers, L_t, over the effective number of consumers, N_t (Mason&Lee, 2005). In practice we adopt the following empirical proxy:

$$\mathrm{SR} = \mathrm{a2064}/(\mathrm{a019}+\mathrm{a65ov})$$

where a2064 : Share of population between age 20-64, a019 : Share of population between age 0-19, a65ov: Share of population age 65+[8]. SR did not attract a significant coefficient when we augmented our cointegrating specification with this variable.

[8] We have checked robustness of our results by shifting the upper limit of the producers to the age of 75. This is consistent with the evidence on the cross-sectional age-wealth profile from Survey of Consumer Finances, provided in Table 1 of Poterba (2001), which shows that the population share between 64-74 still holds considerable amount of common stocks. Results are available upon request.

C Description of all Time-series used in our Empirical Investigation.

Stock Market Prices: S&P 500 index yearly prices from 1909 to 2008 are from Robert Shiller's website, we take december observations.
Stock Market Dividends: Dividends are twelve-month moving sums of dividends paid on the S&P 500 index. They are from the Robert Shiller website for the period 1900-2008. These series coincide with those used in Goyal and Welch (2008), and made available at http://www.bus.emory.edu/AGoyal/Research.html.
Stock Market Returns: For S&P 500 index, to construct the continuously compounded return r_t, we take the ex-dividend-price P_t add dividend D_t over P_{t-1} and take the natural logarithm of the ratio.
Risk-free Rate: We download secondary market 3-Month Treasury Bill rate from St.Louis (FRED) from 1934-2008. The risk-free rate for the period 1920 to 1933 is from New York City from NBER's Macrohistory data base. Since there was no risk-free short-term debt prior to the 1920's, we estimate it following Goyal and Welch (2008). We obtain commercial paper rates for New York City from NBER's Macrohistory data base. These are available for the period 1871 to 1970. We estimate a regression for the period 1920 to 1971, which yielded

$$\mathrm{T-billRate} = -0.004 + 0.886 \times \text{ Commercial Paper Rate.}$$

Therefore, we instrument the risk-free rate for the period 1909 to 1919 with the predicted regression equation. Hence we build our dependent variable which is the equity premium ($r_{m.t} - r_{f,t}$), i.e., the rate of return on the stock market minus the prevailing short-term interest rate in the year t -1 to t. Second, we construct the independent variables commonly used in the long horizon stock market prediction literature; namely
Log Dividend-Price Ratio (dp_t): the difference between the log of dividends and the log of prices.
Consumption, wealth, income ratio (cay): The series is taken from Lettau and Ludvigson (2001). Data are available from Martin Lettau's website at annual frequency from 1948 to 2001.
Consumption, dividend, income ratio (cdy): The series is taken from Lettau and Ludvigson (2005). Data are available from Martin Lettau's website at annual frequency from 1948 to 2001.
Demographic Variables The U.S annual population estimates series are collected from U.S Census Bureau and the sample covers estimates from 1900-2050.

DATA SOURCES

Amit Goyal's Website: http://www.bus.emory.edu/AGoyal/Research.html
Martin Lettau's Website: http://faculty.haas.berkeley.edu/lettau/
Andrew Mason's Website: http://www2.hawaii.edu/˜amason/
Michael R. Roberts' Website: http://finance.wharton.upenn.edu/˜mrrobert/
Robert Shiller's Website: http://www.econ.yale.edu/˜shiller/
Bureau of Labor Statistics Webpage: http://www.bls.gov/data/
FRED: http://research.stlouisfed.org/fred2/
NBER Macrohistory Data Base: http://www.nber.org/databases/macrohistory/contents/
US Census Bureau: http://www.census.gov/popest/archives/

References

[1] Abel, A. B."The Effects of a Baby Boom on Stock Prices and Capital Accumulation in the Presence of Social Security." *Econometrica*, 71 (2003), 2, 551-578.

[2] Abel, A. B. "Will Bequests Attenuate the Predicted Meltdown in Stock Prices When Baby Boomers Retire?" *Review of Economics and Statistics*, 83 (2001), 4, 589-595.

[3] Ang, A. and G. Bekaert. "Stock Return Predictability: Is It There?" *The Review of Financial Studies*, 20 (2007), 651–707.

[4] Ang, A. and A. Maddaloni."Do Demographic Changes Affect Risk Premiums? Evidence from International Data." *Journal of Business*, 78 (2005), 341-380.

[5] Asness, C. "Fight the Fed Model: the Relationship between Future Returns and Stock and Bond Market Yields." *Journal of Portfolio Management*, 30 (2003), 11-24.

[6] Bakshi, G. S., and Z. Chen."Baby Boom, Population Aging, and Capital Markets." *Journal of Business*, 67 (1994), 2, 165-202.

[7] Beaudry, P., and F. Portier. "Stock Prices, News and Economic Fluctuations." *American Economic Review*, 96 (2006), 1293-1307.

[8] Bloom, D. E.; D. Canning; and J. Sevilla."The Demographic Dividend. A new Perspective on the Economic Consequences of Population Change." *Rand Corporation*, Santa Monica (2003).

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

[10] Boudoukh, J.; R. Michaely; M. Richardson; and M. Roberts. "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing." *Journal of Finance*, (2007), 877-915.

[11] Brooks, R. J. "Asset-Market Effects of the Baby Boom and Social-Security Reform." *American Economic Review*, 92 (2002), 2, 402-406.

[12] Brooks, R J. "What Will Happen to Financial Markets When The Baby Boomers Retire?" *Computing in Economics and Finance*, 92 (2000), Society for Computational Economics.

[13] Campbell, J. Y. "A Variance Decomposition for Stock Returns." *Economic Journal*, 101 (1991), 157–179.

[14] Campbell, J. Y.; A. W. Lo; and C. MacKinlay. "The Econometrics of Financial Markets." *Princeton University Press*, Princeton, NJ (1997).

[15] Campbell, J. Y., and S. B. Thomson. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *The Review of Financial Studies*, 21(2008), 1509-1531.

[16] Campbell, J. Y. "A Comment on James M. Poterba's Demographic Structure and Asset Returns", *The Review of Economics and Statistics*, 83 (2001), 4, 585-588.

[17] Campbell, J. Y., and R. Shiller. "Stock Prices,Earnings, and Expected Dividends." *Journal of Finance*, 43 (1988), 661-676.

[18] Cochrane, J. H. "Asset Pricing." *Princeton University Press*, Princeton, NJ (2005).

[19] Campbell, J. Y., and L. M. Viceira. "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors." *Oxford University Press* (2002).

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

[21] Dalla Vigna S., and J.Pollet. "Demographics and Industry Returns." *American Economic Review*, 97(2007), 1167-1702.

[22] Davidson, R., and E. Flachaire. "The Wild bootstrap, Tamed at Last.", *Journal of Econometrics*,146(2008), 1, 162-169.

[23] Erb, C. B.; C. R. Harvey, and T. E. Viskanta. "Demographics and International Investment." *Financial Analysts Journal*, 53 (1996), 4,14-28.

[24] Fama, E., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3-26.

[25] Favero C.A. and A.Tamoni. "Demographics and the Term Structure of Stock Market Risk", available at http: \www.igier.unibocconi.it\favero, (2010).

[26] Geanakoplos, J.; M. Magill; and M. Quinzii. "Demography and the Long Run Behavior of the Stock Market." *Brookings Papers on Economic Activities*, 1(2004), 241-325.

[27] Goyal, A., and I. Welch. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *The Review of Financial Studies*, 21(2008),4, 1455-1508.

[28] Goyal, A., and I. Welch. "Predicting the Equity Premium with Dividend Ratios." *Management Science*, 49(2003), 5, 639–654.

[29] Goyal, A. "Demographics, Stock Market Flows, and Stock Returns." *Journal of Financial and Quantitative Analysis*, 39 (2004), 1, 115-142.

[30] Hansen, H., and S. Johansen. "Some Tests for Parameter Constancy in Cointegrated VAR-Models.", *The Econometrics Journal*, 2 (1999), 306-333.

[31] Hodrick, R., and E. C. Prescott. "Postwar U.S. Business Cycles: An Empirical Investigation.", *Journal of Money, Credit, and Banking*, 29 (1997), 1, 1-16.

[32] Hodrick, R. "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement." *Review of Financial Studies*, 5 (1992), 357–386.

[33] Jaimovich, N. and H. E. Siu. "The Young, the Old, and the Restless: Demographics and Business Cycle Volatility." *American Economic Review*, 99 (2009), 3, 804–826.

[34] Johannes, M.; Korteweg, A.; and N. Polson, "Sequential Learning, Predictive Regressions, and Optimal Portfolio Returns." *SSRN working paper* (2008).

[35] Johansen, S. "Statistical Analysis of Cointegrating Vectors.", *Journal of Economic Dynamics and Control*,12 (1988), 231–254.

[36] Johansen, S. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, 56 (1991), 1551–1580.

[37] Juselius, K. "The Cointegrated VAR Model: Methodology and Applications." *Oxford University Press* (2006).

[38] Lander, J.; A. Orphanides; and M. Douvogiannis. "Earnings Forecasts and the Predictability of Stock Returns: Evidence from trading the S&P." *Journal of Portfolio Management*, 23(1997), 4, 24-35.

[39] Lettau, M., and S. Ludvigson. "Expected Returns and Expected Dividend Growth.", *Journal of Financial Economics*, 76 (2005), 583-626.

[40] Lettau, M., and S. Ludvigson. "Consumption, Aggregate Wealth and Expected Stock Returns." *Journal of Finance*, 56 (2001), 3, 815-849.

[41] Lettau, M., and S. Van Nieuwerburgh. "Reconciling the Return Predictability Evidence." *The Review of Financial Studies*, 21 (2008), 4, 1607-1652.

[42] Lewellen, J. "Predicting Returns with Financial Ratios." *Journal of Financial Economics*, 74 (2004), 209–35.

[43] Macunovich, D. J. "Birth Quake." *Chicago University Press* (2002).

[44] MacKinnon, J. G.; A. A. Haug; and L. Michelis. "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration." *Journal of Applied Econometrics,* 14 (1999), 563–577.

[45] Mason, A., and R. Lee. "Reform and Support Systems for the Elderly in Developing Countries: Capturing the Second Demographic Dividend." *Genus*, 2 (2005), 11-35.

[46] Nelson, C. C. and M. J. Kim. "Predictable Stock Returns: The Role of Small Sample Bias." *Journal of Finance*, 43 (1993), 641–661.

[47] Neely C.J., and P.A. Weller. "Predictability in International Asset Returns: A Reexamination." *Journal of Financial and Quantitative Analysis*, 35 (2000), 601-620.

[48] Newey, W. K. and K. D. West. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.", *Econometrica*, 55 (1987), 3, 703-08.

[49] Newey, W. K., and D. K. West. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, 61 (1994), 631-653.

[50] Nyblom, J. "Testing for the Constancy of Parameters over Time." *Journal of the American Statistical Association*, 84 (1989), 223-230.

[51] Paye, B.S., and A. Timmermann. "Instability of Return Prediction Models." *Journal of Empirical Finance*, 13 (2006), 274-315

[52] Pastor, L. and R. F. Stambaugh. "Are Stocks Really Less Volatile in the Long Run?" *Working Paper*, 14757 (2009).

[53] Poterba, J. M. "Demographic Structure and Asset Returns." *The Review of Economics and Statistics*, 83 (2001), 4, 565-584.

[54] Rapach D.E., and M.E.Wohar. "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data-Mining." *Journal of Empirical Finance*, 13 (2006), 231-247.

[55] Shiller, R. J. "Irrational Exuberance" *Princeton University Press* (2005).

[56] Stambaugh, R. F. "Predictive Regressions." *Journal of Financial Economics*, 54 (1999), 375–421.

[57] Valkanov, R. "Long-Horizon Regressions: Theoretical Results and Applications." *Journal of Financial Economics*, 68 (2003), 201–232. 33

[58] Yoo, P. S. "Population Growth and Asset Prices.", *Federal Reserve Bank of St. Louis Working Paper No.* (1997)-016A,.