Local trends in price-to-dividend ratios - assessment, predictive value and determinants*

HELMUT HERWARTZ, MALTE RENGEL and FANG XU

April 6, 2016

Abstract

Persistent variations of the log price-to-dividend ratio (PD) and their economic determinants have attracted a lively discussion in the literature. We suggest a gradually time-varying state process to govern the persistence of the PD. The adopted state space approach offers favourable model diagnostics and finds particular support in out-of-sample stock return prediction. We show that this slowly evolving mean process is jointly shaped by the consumption risk, the demographic structure and the proportion of firms with traditional dividend payout policy during the past 60 years. In particular, the volatility of consumption growth plays the dominant role.

Keywords: Price-to-dividend ratio, stock return prediction, consumption risk, dividend payments, demographics, nonlinear state space model, particle filtering.

JEL Classification: C53, C58, E44, G12, G17.

* We are grateful for helpful suggestions from two anonymous referees and the editor (Pok-sang Lam). We thank participants at the 1st IAAE conference at Queen Mary University of London, the 9th International Symposium on Econometric Theory and Applications in Seoul, the 5th International IFABS Conference in Nottingham, the annual meeting of the Verein für Socialpolitik, Düsseldorf and several research seminars for useful comments and discussions of an earlier draft of the paper titled *A slowly evolving mean of the price-to-dividend ratio, its economic influences and predictive power for stock returns.* Financial support from the German Research Foundation (Grant No. HE 2188/8-1) is gratefully acknowledged.
Helmut Herwartz is a Professor at Georg-August-University Göttingen (Email: hherwartz@uni-goettingen.de). Malte Rengel obtained his PhD from Christian-Albrechts-University of Kiel in 2014 (Email: mrengel@uni-goettingen.de). Fang Xu is a Lecturer at University of Reading (Email: fang.xu@reading.ac.uk).

The log price to dividend ratio (PD) in the US embarked an upward trend since the 1980s that deviates markedly from its historical average in the preceding century. Market valuation reached unprecedented levels relative to any fundamental values in the new millennium. There is cumulating evidence for structural break(s) or instability in the mean of the PD (Lettau, Ludvigson, and Wachter 2008) and in the relation between the PD and future stock returns (Paye and Timmermann 2006, Rapach and Wohar 2006, Welch and Goyal 2008). Coping with the persistence of the PD, Lettau and Van Nieuwerburgh (2008) suggest a regime-switching model that allows discrete mean shifts. In particular, they show that deviations from shifting means of the PD carry predictive power for stock returns in-sample, but fail to signal stock returns ex-ante compared with the historical average return. Empirical evidence indicates that the increasing mean of the PD could be due to a persistent deceleration of macroeconomic risks which can be measured by the volatility of consumption growth rates (Bansal and Yaron 2004, Bansal, Khatchatrian, and Yaron 2005, Lettau et al. 2008, Bansal, Kiku, and Yaron 2010), changes in demographic structures of the population (Geanakoplos, Magill, and Quinzii 2004, Favero, Gozluklu, and Tamoni 2011), and the dividend pay-out policy by firms (Fama and French 2001, Robertson and Wright 2006, Boudoukh, Richardson, and Whitelaw 2008, Kim and Park 2013).

In this paper we consider a gradually time-varying mean of the PD that enables simultaneous testing of distinct determinants of persistent patterns characterizing the PD. In the framework of a nonlinear state space model we estimate a latent process reflecting the slowly evolving mean of the PD within a generalized version of the present value model introduced by Campbell and Shiller (1988). In explaining persistence, the consideration of local PD means firstly allows for a comparison with the (succession of) discrete mean shifts, particularly in terms of out-of-sample predictive power for stock returns. Secondly, it offers an opportunity to look for common trends linking the local PD mean, consumption risk, the demographic structure, and the dividend payout policy of firms. Using a nonlinear sate space model can be linked to recent applications of linear state-space models (with Kalman filtering) for modelling stock returns (e.g. Binsbergen and Koijen 2010, Rytchkov 2012). Binsbergen and Koijen (2010), for example, treat the expected return and expected dividend growth as two latent processes. However, these studies assume an exogenous fixed mean of the PD and, thus, are not necessarily consistent with the observed persistence in the PD. Addressing the persistence of the PD explicitly, the latent process considered in this work can be interpreted as a combination of local means of expected returns and expected dividend growth. Owing to intrinsic non-linearity, the Kalman filter doesn't offer optimal solutions. We adopt particle filtering (e.g. Cappé, Godsill, and Moulines 2007), a flexible Monte Carlo technique, for consistent log-likelihood assessment, inferential and model selection issues.

We find that a gradually time-varying mean of the PD is strongly supported by log-likelihood diagnostics. The estimated long-term state has some step-like patterns similar to mean shifts with two structural breaks as suggested by Lettau and Van Nieuwerburgh (2008). Importantly, the slowly evolving process allows a simple projection towards the future, and straightforward implementation of standard predictive regressions for stock returns conditional on this information. Local deviations of the PD from its gradually time-varying mean carry out-of-sample predictive power. Using the out-of-sample degree of explanation based on the root mean squared error (RMSE) (Goyal and Welch 2003), we confirm the significance of the out-of-sample forecasting performance in comparison with both historical average returns and PD adjustments using discrete mean shifts. As economic underpinnings of both PD persistence and out-of-sample predictive content, we find that consumption risk, the demographic structure and the dividend payout policy of firms jointly shape the slowly evolving mean of the PD during the past 60 years. The adopted error correction approach allows the data to determine the transmission channel among the observed trends in financial markets and the underlying economy. All long run determinants of the PD are diagnosed weakly exogenous. A low consumption volatility risk drives down equity premia and pushes up the stock price (Bansal and Yaron 2004). The decreasing volatility in the consumption growth rate has the highest contribution in explaining the increasing mean of the PD. A high middle-aged to young ratio, leading to excess demand for saving, drives up the equilibrium asset prices (Geanakoplos et al. 2004).

The significant increases in the mean of the PD in the 1990s are consistent with increases in the middle-aged to young ratio during this period. In addition to the macroeconomic and demographic influences, lowered dividends can affect the long-run relationship between stock prices and dividends (Kim and Park 2013). The fall in the proportion of firms that payout a significant fraction of their earnings in the form of dividends since the 1980s is consistent with the increasing mean of the PD. Nevertheless, among the three factors this has the smallest contribution in explaining the variations in the mean of the PD.

The rest of the paper is organized as follows. Section 1 illustrates the persistence of the PD, sketches its implications for the standard present value model, and introduces the state space model of the PD incorporating a gradually time-varying mean. The forecasting model, evaluation methods and forecasting performance are discussed in Section 2. In Section 3 we investigate the linkage between the gradually time-varying mean of the PD and its potential influences. Section 4 concludes. Appendices provide detailed descriptions of the data (Appendix A), the particle filtering approach (Appendix B), and approximation errors involved in the derivation of the present value model (Appendix C).

1 A STATE SPACE MODEL OF THE PD

In this section we first discuss the observed persistence of the PD and its implications for respective present value formulations. Then a latent gradually time-varying mean of the PD is formally derived and estimated, which is in line with the diagnosed trends governing the PD. Log-likelihood statistics support the view that the present value model of the PD incorporating a gradually time-varying mean outperforms the model with a constant mean and models with discrete mean shifts. We consider annual data for the period 1926 to 2013 from the Center for Research in Security Prices (CRSP) and S&P500 data from 1871 to 2013, see Appendix A for detailed information.

1.1 Persistence of the PD

The persistent increase of stock prices relative to dividends from 1980 to 2000 can be seen from Figure 1. We find that the PD can be well described by a non-stationary process, which confirms findings in previous studies; see for example Campbell (1999), Herwartz and Morales-Arias (2009) and Park (2010). Table 1 documents results from numerous unit root tests. The hypothesis of a non-stationary PD cannot be rejected with 5% significance by means of the ADF test, tests proposed by Phillips and Perron (1988), Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001).[1]

Figure 1 about here

The PD is unlikely to be a stationary process even taking into account the power weakness of unit root tests under near integration. As can be seen from the last column of Table 1, the null hypothesis of stationarity of the PD is rejected by means of the KPSS statistic (Kwiatkowski *et al.* 1992). Moreover, testing the unit root hypothesis as proposed by Perron and Vogelsang (1992) we find that the PD can be better described by a non-stationary process than by a stationary process with a structural break at unknown timing (see column 5 in Table 1, 'PV').

Table 1 about here

As noted by Campbell (2008), the persistence of the PD challenges the present value model in Campbell and Shiller (1988) that rests on the assumption of a stationary PD. Let P_t and D_t denote stock prices and the corresponding dividends in time t, respectively. The total log-return, realized at the end of period $t+1$, $r_{t+1} = \ln(P_{t+1} + D_{t+1}) - \ln(P_t)$, can be formulated as a nonlinear function of the PD, $\eta_t = \ln(P_t) - \ln(D_t)$,

$$r_{t+1} \quad = \quad -\eta_t + \ln\left(\exp(\eta_{t+1}) + 1\right) + \Delta d_{t+1}, \tag{1}$$

where $d_t = \ln(D_t)$ and Δ is shorthand for the first difference operator such that, e.g., $\Delta d_t = d_t - d_{t-1}$. A first order Taylor expansion around a fixed steady state $\overline{\eta}$ provides

[1] It is worthwhile to mention that opposite to pure random walks diagnosed by common unit root tests, actual PD processes cannot grow to any level. Recently, bounded non-stationary processes have attracted interest in the econometric literature (Cavaliere 2005). Cavaliere and Xu (2014) have proposed a novel ADF based approach to test for unit roots in the presence of bounds. The critical values of such tests are smaller (i.e., larger in absolute value) than those of unit root tests neglecting the bounded nature of a variable of interest. Thus, if common unit root tests hint at non-stationarity, bounded non-stationarity will be diagnosed once the bounds are taken into account.

the linear approximation

$$r_{t+1} \simeq \kappa - \eta_t + \rho\eta_{t+1} + \Delta d_{t+1}, \tag{2}$$

with $\rho \equiv 1/(1+\exp(-\overline{\eta}))$ and $\kappa \equiv -\ln(\rho) - (1-\rho)\ln(1/\rho - 1)$. In the empirical analysis, the constant parameter $\overline{\eta}$ is assumed to be known and commonly approximated by the sample mean (e.g. Campbell 1999). Under persistent behaviour of the PD, $\overline{\eta}$ is unlikely to be constant and ρ becomes also time-varying. Figure 2 illustrates the time variation of sample means of the PD from rolling time windows covering observations from the most recent 20 years.

Figure 2 about here

1.2 A State-Space Approximation

Taking a gradually time-varying mean of the PD into account, we modify the traditional present value model of the PD. Let $\widetilde{\eta}_t$ denote the local mean employed to expand the Taylor approximation of the one-step-ahead stock returns in (1). We obtain

$$r_{t+1} \simeq \kappa_t - \eta_t + \rho_t\eta_{t+1} + \Delta d_{t+1}, \tag{3}$$

with both parameters (κ_t and ρ_t) in (3) becoming time-specific, i.e.,

$$\rho_t \equiv 1/(1+\exp(-\widetilde{\eta}_t)) \text{ and } \kappa_t \equiv -\ln(\rho_t) - (1-\rho_t)\ln(1/\rho_t - 1). \tag{4}$$

To derive the present value formulation of the PD from (3), the following approximations similar to those in Lettau and Van Nieuwerburgh (2008) are adopted: $E_t[\rho_{t+i}] \approx \rho_t$, $E_t[\kappa_{t+i}] \approx \kappa_t$ and $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$. Simulation studies documented in Appendix C show that combined approximation errors are negligible for typical values of $\widetilde{\eta}_t$. Taking the conditional expectation and iterating equation (3) forward, provides the log-linear present value formulation of the PD

$$\eta_t \simeq \frac{\kappa_t}{1-\rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} E_t[\Delta d_{t+i}^e - r_{t+i}^e] + \lim_{i\to\infty} \rho_t^i E_t[\eta_{t+i}], \tag{5}$$

where superscripts e symbolize the excess of dividend growth rates ($\Delta d_{t+i}^e = \Delta d_{t+i} - r_{t+i}^f$) or of returns ($r_{t+i}^e = r_{t+i} - r_{t+i}^f$) over the risk-free interest rate r_t^f. Changes in the long-term state of the PD affect the observed PD in a nonlinear fashion. A time-varying $\widetilde{\eta}_t$ leads to a time-varying rather than a constant intercept term $\kappa_t/(1-\rho_t)$. Future return-adjusted dividend growth rates are discounted at time-varying rates ρ_t rather than at a constant one.

An intuitive way to link equation (5) to the traditional present value model in Campbell and Shiller (1988) is to reconsider it from the perspective of an investor who can only quantify the mean of the PD conditional on past information. In this case, as shown in Lacerda and Santa-Clara (2010), the mean of the PD becomes time-varying and one can introduce directly a time index t for the parameters ρ and κ in the traditional present value model to derive equation (5) (see also Figure 2). The proposed model offers a structural interpretation for this approach.

We employ a state space model to estimate the latent time-varying $\widetilde{\eta}_t$. Assume a random disturbance term $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ to capture eventual rational bubbles, approximation errors, and other influences in $\lim_{i\to\infty} \rho_t^i E_t[\eta_{t+i}]$. Further substituting E_t in (5) by objective expectations conditional on information available at the end of period t ($\widetilde{E}_t$), equation (5) is transformed into the measurement equation,

$$\eta_t = \frac{\kappa_t}{1-\rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} \widetilde{E}_t[\Delta d_{t+i}^e - r_{t+i}^e] + \epsilon_t, \tag{6}$$

with κ_t and ρ_t being nonlinear functions of $\widetilde{\eta}_t$ (see (4)). The state equation formalizes a dynamic pattern for the latent process $\widetilde{\eta}_t$, which is consistent with the diagnosed persistence of the PD,

$$\widetilde{\eta}_t = \widetilde{\eta}_{t-1} + u_t, \tag{7}$$

where $u_t \sim N(0, \sigma_u^2)$, and the initialization $\widetilde{\eta}_0$ is treated as a model parameter. Henceforth, we refer to this model as the randam walk (RW) state specification. As a particular alternative state equation we consider a stationary first order autoregressive (AR(1)) state process, i.e.,

$$\widetilde{\eta}_t \;=\; \delta + \alpha \widetilde{\eta}_{t-1} + u_t, \tag{8}$$

where $|\alpha| < 1$, $u_t \sim N(0, \sigma_u^2)$ and $\widetilde{\eta}_0 = \delta/(1-\alpha)$.[2] Coping with the nonlinear relation between the PD and its latent mean process, the state space model in (6) coupled with (7) or (8) is estimated by means of particle filtering (Cappé et al. 2007). We sketch this method in the context of the considered state space model below. A more detailed discussion can be found in Appendix B.

Compared with a framework of structural breaks, a continuously evolving steady state of the PD not only allows to test for its various determinants, but may, particularly, improve out-of-sample forecasting of stock returns. In-sample forecasting of stock returns with structural breaks takes advantage of the maximum of available information, since subsample means can be easily determined conditional on full samples. For out-of-sample forecasting, however, the performance of break adjusted schemes is weakened since timing and magnitude of the breaks are unknown and have to be estimated (Lettau and Van Nieuwerburgh 2008). In contrast, the state space model with particle filtering uses (mainly) past information to estimate the latent mean of the PD. Although this feature might be informationally inferior for in-sample forecasting, it has an edge over the break adjusted schemes in out-of-sample prediction. Estimates of the latent state ($\widetilde{\eta}_t$) continuously adapt to new information, and can be easily extended into the future. It becomes unnecessary to locate break dates and magnitudes.

Compared with recent applications of linear state-space models for forecasting stock returns (e.g. Binsbergen and Koijen 2010, Rytchkov 2012) the above nonlinear state-space

[2] We allow for many alternative specifications of the state process: a random walk with variance breaks, a non-stationary AR(2) model, a random walk with leptokurtic innovations or a moving average structure of the error term and an autoregressive process with parameter shifts. Irrespective of these alternatives the estimated state processes come very close to the one extracted from the pure random walk model in (7).

model has two important characteristics.

First, the measurement equation in (6) formalizes the PD (η_t) as a nonlinear function of the latent state ($\widetilde{\eta}_t$). Hence, the Kalman filter - the natural tool for estimation and inference in linear state-space models - might, at best, provide an approximate representation of the PD in the present context. In light of marked efficiency losses involved with such linear approximations, Monte Carlo (MC) approaches such as particle filtering are preferable to assess accuracy of nonlinear models (see e.g. Fernández-Villaverde and Rubio-Ramírez (2005) and Fernández-Villaverde and Rubio-Ramírez (2007) for the case of dynamic stochastic general equilibrium models).[3] Particle filtering weights a set of samples (particles) in accordance with their time local log-likelihood contributions (see e.g. Ristic, Arulampalam, and Gordon (2004) and Doucet, De Freitas, and Gordon (2001)). Thus, it does not require the restrictive assumptions of the Kalman filter. With an increasing number of samples, likelihood assessment by means of the particle filter is consistent and, hence, as precise as warranted by the analyst. In this framework likelihood based tests (Vuong 1989) may conveniently substitute common (quasi) ML test statistics based on the Kalman filter for model selection and inferential issues.

Second, the present-value relationship among the PD, expected returns and dividend growth serves as an estimation equation in this paper. Under the assumption of a constant and known mean of the PD, the traditional present value relationship (Campbell and Shiller 1988) serves as an identity restriction in state space models similar to the one in Binsbergen and Koijen (2010). This enables identification of two latent state variables - e.g., expected returns and expected dividend growth in Binsbergen and Koijen (2010). However, the imposition of a constant mean of the PD could lead to biased estimates for the latent expected return and dividend growth if the actual PD is persistent. Consistent with the diagnosed trends governing the PD, we formalize a latent varying mean of the PD to maximize the informational content of the present value model in (6). The variations in the steady-state PD reflect a combination of variations in the steady-state expected

[3] Similar caveats for Kalman filter based approximations are known from the literature on stochastic volatility models (Kim, Shephard, and Chib 1998, Pitt and Shephard 1999). For our generalized present value model we are not aware of a linear approximation that might be evaluated by means of Kalman recursions.

returns and expected dividend growth.[4]

1.3 Model Implementation

We use particle filtering based on 3000 trajectories for an approximation of the models' log-likelihood, subsequent parameter and state estimation. Before applying the particle filter, we need to determine objective expectations about future excess dividend growth rates and excess returns in (6). We follow Campbell and Vuolteenaho (2004) and employ low dimensional vector autoregressions (VARs) of order one comprised of the PD series, excess dividends growth rates, excess returns and inflation ($\{\eta_i, \Delta d_i^e, r_i^e, \pi_i\}_{i=1}^t$).[5] The VAR based determination of objective expectations ($\widetilde{E}_t[\Delta d_{t+i}^e]$ and $\widetilde{E}_t[r_{t+i}^e]$) goes back to Campbell and Shiller (1988) and Campbell (1991).[6] Including the PD in the VAR provides unobservable market information about the future dividends and returns. The reduced form VAR is also flexible by timely updating most relevant market information. To evaluate $\widetilde{\eta}_t$, objective expectations - $\widetilde{E}_t[\Delta d_{t+i}^e]$ and $\widetilde{E}_t[r_{t+i}^e]$ with $i > 0$ - are determined conditional on most recent information available in time t.[7]

We take an adaptive approach to the choice of the VAR sample size such that it not only provides efficient parameter estimates under structural invariance of VAR dynamics, but also responds to structural changes. Specifically, we evaluate in each forecast origin VAR models conditioning on samples $\Omega_{t,\omega} = \{\eta_\tau, \Delta d_\tau^e, r_\tau^e, \pi_\tau, \tau = t-\omega+1, \ldots, t\}$ with alternative lengths ($\omega = 20, 21, \ldots, 30$). To determine ex-ante predictions $\widehat{\Delta d}_{t+i}^e$ and $\hat{r}_{t+i}^e$, we employ the VAR with the particular window size ω that minimizes the root mean squared errors for the ten most recent in-sample observations $\{\Delta d_m^e - r_m^e\}_{m=t-9}^{m=t}$.[8] The

[4] It is noticeable that it is not straightforward in this framework to treat expected returns or/and expected dividend growth as additional latent state variables simultaneously.

[5] The inclusion of inflation accounts for the eventual effects of money illusion on equity prices.

[6] In related contexts, VAR based predictions have also been used to approximate price expectations, for instance by Sbordone (2002) and Rudd and Whelan (2006). By means of a theoretical model on the generation of inflation expectations Branch (2004) shows that economic agents use more often VAR forecasts for expectation formation in comparison with adaptive or naive prediction rules.

[7] Note that we include a constant in the VAR model. One may argue that a VAR model with a deterministic trend might be more suitable to model a persistent PD. Estimates for the latent process ($\widetilde{\eta}_t$) display similar dynamics if a time trend is included. Using a trend in the VAR doesn't change the essence of the results.

[8] To assess the robustness of outcomes we consider a set of robustness tests (i) using fitted errors regarding excess returns $\{r_m^e\}_{m=t-9}^{m=t}$ instead of $\{\Delta d_m^e - r_m^e\}_{m=t-9}^{m=t}$; (ii) using the five most recent observations in $\{\Delta d_m^e - r_m^e\}_{m=t-4}^{m=t}$ to compute the RMSE; or (iii) using the mean absolute error criterion instead of RMSE. The corresponding results with regard to the evaluation of the state space model are quantitatively almost

initialization period for VAR forecasting comprises 30 observations. Following Lettau and Van Nieuwerburgh (2008) we focus mainly on annual CRSP stock market data.[9] To evaluate the state space model from 1926 (a common starting period in the literature), we joined the CRSP data starting in 1926 with the S&P500 data before this period. In addition, we consider the S&P500 and estimate the state space model for the sample period 1901 to 2013.

1.4 Estimates and Diagnostics

The estimated parameters and diagnostic statistics for numerous model specifications including the time-varying mean model are documented in Table 2. For purposes of comparison we also estimate the constant mean model (Campbell and Shiller 1988) and the model of discrete mean shifts proposed by Lettau and Van Nieuwerburgh (2008). The latter is a regime-switching model and employs the supremum F-test (Bai and Perron 1998, 2003) to determine the timing of the breaks as proposed by Lettau and Van Nieuwerburgh (2008). In the case of one shift it is diagnosed to occur in 1992, and in the case of two breaks the respective locations are 1955 and 1993.[10]

Table 2 about here

We adopt the BIC and the Vuong statistic for model comparison (the last two columns of Table 2). A smaller BIC indicates the superiority of a model. The Vuong statistic is particularly helpful in isolating idiosyncratic contributions to diagnostics of non-nested models. It is calculated taking the time-varying state model with the RW state equation as baseline specification. Negative statistics indicate a lead of the baseline model. We refrain from using formal likelihood ratio (LR) tests for model comparison, since common

identical.

9 The 3-month Treasury Bill rate is employed to approximate the risk free rate and the CPI to measure inflation.

10 These break points are close to those diagnosed in Lettau and Van Nieuwerburgh (2008) who analyse a slightly distinct sample period (1926 to 2004). The null hypothesis of no break is rejected with 1% significance against one or two breaks ($supF(1|0) = 18.12$ and $supF(2|0) = 23.90$). The null hypothesis of one break is rejected against the alternative of two breaks ($supF(2|1) = 9.56$) with 10% significance. The applied test procedure is robust to serial correlation and heteroscedasticity, the trimming is 5% of the sample. ML estimates of the break points differ only slightly from those detected by means of supremum tests, and all subsequent results are qualitatively identical. Detailed results are not shown due to space considerations and are provided by the authors upon request.

χ^2 critical values may lack applicability. For instance, the constant mean model is at the bound of the variance parameter in the more flexible RW specification.

We find that estimating $\widetilde{\eta}_t$ conditional on either a RW or a stationary AR(1) state equation provides very similar results. Both implied state processes can only be differentiated marginally in the early and later sample periods (see Figure 3). It turns out that both estimates of $\widetilde{\eta}_t$ lead to qualitatively identical results for the remaining empirical analysis. We concentrate on the estimates from a RW state equation henceforth, since it is in the lead over a stationary AR (1) process according to log likelihood based diagnostics.

Figure 3 about here

Moreover, the model with a gradually time-varying mean of the PD outperforms its constant mean counterpart and models with discrete mean shifts. Considering the RW state process (7), the log-likelihood value conditioned on CRSP data for a time-varying state model is about 168.4 while the respective statistic for the constant state model is 23.9. The lead of the more flexible model approach over the static benchmark present value model can be visualized by eyeballing the estimated patterns of $\widetilde{\eta}_t$ provided in Figure 3. Both BIC and Vuong statistics are supportive for the time-varing mean model over both the constant mean model and models with discrete shifts. Results for S&P data yield identical conclusions as those for CRSP data.

2 FORECASTING PERFORMANCE

Time variation in the mean of the PD is valuable for the ex-ante modeling of stock returns. In this section, we analyse how $\widetilde{\eta}_t$ exploits the informational content of the PD in so-called predictive regressions. We discuss predictive regression models for stock returns conditional on CRSP data. Results for S&P data are qualitatively identical. Adjusting the PD by means of its slowly evolving mean provides better out-of-sample forecasts in terms of the RMSE and the out-of-sample R^2 compared with centering the PD with discrete mean shifts or using the historical average of returns as the predictor. In the following we describe in-sample (IS) and out-of-sample (OOS) forecasting designs, and discuss in detail the forecasting performance of competing approaches.

2.1 Predictive Regressions

The predictability of stock returns is evaluated by means of common predictive regressions of the following type (see e.g. Lettau and Van Nieuwerburgh 2008),

$$r_{t+1} = \beta_0 + \beta_1(\eta_t - s_t) + v_{t+1}, \tag{9}$$

where r_{t+1} denotes the total log-returns and v_{t+1} is an error term. We also use the predictive regressions to assess the predictability of dividend growth rates, substituting Δd_{t+1} for r_{t+1} in (9). To implement predictive regressions, the PD (η_t) is adjusted by alternative state processes (s_t) such that 'centered' observations ($\eta_t - s_t$) are considered to predict stock returns. Under the null hypothesis of no predictability, $\beta_1 = 0$. Predictability of return adjustments towards an equilibrium among prices and dividends imply $\beta_1 < 0$. Imposing $\beta_1 = 0$ serves as the benchmark model (see e.g. Welch and Goyal 2008). For IS analysis, the corresponding naive predictor is the full sample mean return. For OOS analysis the naive predictor is the historical average return obtained up to the forecast origin.

In the *IS analysis* we compare forecasting specifications obtained from four alternative long-run states $s_t \in \{\bar{\eta}, \overline{\overline{\eta}}_t^{(1)}, \overline{\overline{\eta}}_t^{(2)}, \widetilde{\eta}_t\}$. In the first specification the PD is centered by its (full sample) mean ($\overline{\eta}$). We refer to this setting as the 'unadjusted' PD since this model is equivalent to that of using the actual PD series in the predictive regressions. In the second and third specification, the PD is adjusted for one and two structural breaks ($\overline{\overline{\eta}}_t^{(1)}$ and $\overline{\overline{\eta}}_t^{(2)}$, respectively). Lastly, we adjust the PD by means of the gradually time-varying mean $\widetilde{\eta}_t$, which is filtered from the state space model outlined in Section 1.

Initializing the *OOS analysis*, the first forecasting regressions use 20 years of data. Then the estimation windows are expanded recursively as in Lettau and Van Nieuwerburgh (2008). We consider three corresponding adjustments for the PD - $s_t \in \{\overline{\eta}, \overline{\overline{\eta}}_t, \widetilde{\eta}_t\}$ – all of which are recursively estimated from the respective estimation samples. In the benchmark setting, the PD is centered with its mean $\overline{\eta}$ from the estimation period. The second adjustment $s_t = \bar{\bar{\eta}}_t$ corresponds to the case of discrete mean shifts. We apply supremum F-tests and rely on the 10% significance level to determine the mean shift processes $\bar{\bar{\eta}}_t$. Lastly, the PD is adjusted by $\widetilde{\eta}_t$ conditioning only on the information from the estimation periods.

Figure 4 about here

The four alternative long run states of the PD entering the IS analysis, $s_t \in \{\bar{\eta}, \bar{\bar{\eta}}_t^{(1)}, \bar{\bar{\eta}}_t^{(2)}, \widetilde{\eta}_t\}$, are displayed in Figure 4. The smoothly evolving mean $\widetilde{\eta}_t$ seems to be mostly close to the mean with two structural breaks $\bar{\bar{\eta}}_t^{(2)}$. However, the former lags behind the latter after the diagnosed break dates (1955 and 1993). This reveals the nature of the particle filtering applied to the non-linear state space model. Although the parameters of the state space model are estimated conditioning on the full sample information for the IS analysis, the estimated latent process is mainly based on past information. Using the RW state equation (7) as an example, each particle is equal to $\widetilde{\eta}_{t-1}^{(i)}$ plus a draw from the error term with variance σ_u^2. Thus, being a (weighted) average of particles, $\widetilde{\eta}_t$ is mainly determined from the past information. This contributes to the slowly evolving nature of the estimated gradually time-varying mean, which does not show much advantage for the IS analysis, but could be crucial for the predictive power of the PD in the OOS analysis. The core obstacle in using discrete break adjustments in OOS forecasting is to determine the timing and magnitude of the breaks. The gradually time-varying mean $\widetilde{\eta}_t$ overcomes these difficulties. When there are no marked structural changes, it evolves around a relatively stable level. In response to persistent movements, it adapts and incorporates the new information gradually. Specifically, to obtain an update for $\widetilde{\eta}_t$ by means of weighted averaging, particles $\widetilde{\eta}_{t-1}^{(i)}$ are ranked according to the fit of the corresponding measurement equation for period $t-1$. Particle $\widetilde{\eta}_{t-1}^{(j)}$ enters $\widetilde{\eta}_t$ with higher weight than particle $\widetilde{\eta}_{t-1}^{(k)}$ when the error term in the measurement equation for the former is smaller than the one for the latter. Along the updating steps the fittest particles survive. Readers may consult Appendix B for more details.

2.2 *Forecast Evaluation*

Results for in-sample analysis of predictive performance are documented in Table 3. Predicting stock returns in-sample, the unadjusted PD provides a small R^2 of about 0.0392. Adjusting the PD for shifts improves the explanatory content of predictive regressions markedly. The R^2 statistics increase to 0.1027 and 0.1751 for means with one and two shifts, respectively (column 3 and 4). The magnitude and the statistical significance of the estimated predictive coefficient (β_1) increase as well. This evidence confirms findings in Lettau and Van Nieuwerburgh (2008). As expected, with an in-sample degree of explanation of about 0.0641 (column 5), adjusting the PD by a slowly evolving mean does not outperform adjustments for discrete shifts in the mean. As an adaptive filtering process, $\widetilde{\eta}_t$ mainly depends on past information even in the in-sample setting. In contrast, the break adjustments take into account the full sample information and ex-post minimize squared approximation errors for the actual PD.

Table 3 about here

It is worth mentioning that the adjustment of the predictor variable by means of a (discretely/continuously) varying mean in the PD mitigates the persistence, thereby supporting the convenience of (9) for testing stock return predictability. If the predictor variable is persistent and its innovations are positively correlated with returns, coefficient estimates $\hat{\beta}_1$ from (9) are downward biased under the null hypothesis of no predictability (e.g. Stambaugh 1999). The AR(1) coefficient for the observed PD is about 0.94 for CRSP data. This persistence measure reduces to $0.77, 0.67$, and 0.82 for the PD adjusted with means implied by one and two structural breaks and the smoothly evolving mean, respectively. Also the correlations between the innovations in the predictive regression and those in an AR(1) regression of the predictor variable are moderate for all considered predictor variables (between 0.53 to 0.67). We consider the Q-statistics of Campbell and Yogo (2006) for a bias robust test of the predictive power for returns. The estimates of β_1 in (9) remain significantly different from zero for the PD adjusted for a time-varying mean or discrete shifts, but becomes insignificant for the unadjusted PD.[11]

Forecasting dividend growth, the unadjusted PD lacks predictive content. The predictive coefficient is not statistically significant at the 5% level and the R^2 is negligible (column 2 of lower panel). Neither break adjustments nor gradually time-varying mean adjustments improve the performance in a sizeable manner (column 3-5 of the lower panel). This evidence is in line with results from Lettau and Van Nieuwerburgh (2008) for a similar sample period (1927 to 2004).

Table 4 about here

Results for real time (OOS) forecasting performance are documented in Table 4. Adjusting the PD for discrete mean shifts fails to improve upon using historical average returns as benchmark predictors. In contrast, centering the PD around the gradually time-varying mean obtains the smallest RMSE statistic among all predictors (last column in Table 4). Considering the full sample period from 1946 to 2013 (first panel in Table 4), the naive benchmark, the unadjusted PD, centering with discrete mean shifts and centering around $\widetilde{\eta}_t$ results in RMSE statistics of 0.1694, 0.1751, 0.1793 and 0.1683, respectively. The same ranking of the RMSEs holds if the sample period ends at 2004, as considered by Lettau and Van Nieuwerburgh (2008) (lower panel in Table 4).[12]

To evaluate the statistical significance of the forecasting performance of alternative predictors compared with using historical average returns as naive forecasts, we consider an OOS degree of explanation (Welch and Goyal 2008),

$$R^2_{oos} = 1 - \frac{MSE_s}{MSE_{\bar{r}}}, \tag{10}$$

where $MSE_{\bar{r}}$ denotes the mean squared forecast error from naive forecasts and MSE_s is the corresponding statistic from alternative models (9) with $s_t \in \{\overline{\eta}, \overline{\overline{\eta}}_t, \widetilde{\eta}_t\}$. Under

[11] The $Q-$statistic can be interpreted as robust confidence intervals for slope estimates $\hat{\beta}_1$. Following the convention to consider the upper bound of confidence intervals with 90% coverage probability to provide evidence with 5% significance against $H_0 : \beta_1 \geq 0$, we find that upper bounds are 0.222, -0.088, -0.266 and -0.056 when using $\overline{\eta}, \overline{\overline{\eta}}_t^{(1)}$, $\overline{\overline{\eta}}_t^{(2)}$ and $\widetilde{\eta}_t$ to adjust η_t in (9), respectively. Further results on bias corrected inference in predictive regressions are available from the authors upon request.

[12] Results from mean absolute errors (MAEs) are qualitatively identical (not shown).

the hypothesis of less (more) accurate forecasts from alternative model specifications compared with naive predictions, the MSE of the benchmark model is smaller (larger) than that of the alternative model, which corresponds to $R^2_{oos} < 0$ ($R^2_{oos} > 0$). Following Rapach, Strauss, and Zhou (2010) the significance of R^2_{oos} is evaluated by means of the MSE-adjusted statistic in Clark and West (2007).[13]

Adjusting the PD in real time by $\widetilde{\eta}_t$ outperforms the historical average return in forecasting stock returns significantly. As can be seen from the second and fourth row of Table 4, only adjusting the PD for the gradually time-varying mean ($\widetilde{\eta}_t$) provides positive and significant R^2_{oos} statistics. We find the same evidence for S&P500 data (corresponding results are available upon request).

Figure 5 about here

Following Welch and Goyal (2008) we provide further insights into OOS forecasting performance over time and depict the difference of the cumulative squared forecasting errors of naive forecasts minus those of the alternative models in Figure 5.[14] We find that using the gradually time-varying mean adjustment improves the forecasting strength of the PD throughout the entire sample period compared with the unadjusted PD or using break adjustments. The performance curve of adjusting the PD by its gradually time-varying mean falls least during periods with structural changes, and has the longest positive trend during the relatively tranquil periods. The two ex-post identified structural changes occur in 1955 and 1993. In a real time forecasting situation, all three predictors ($s_t \in \{\overline{\eta}, \overline{\overline{\eta}}_t, \widetilde{\eta}_t\}$) start to underperform in comparison with the naive forecasts around 1957 and embark a negative trend. However, adjusting the PD by $\widetilde{\eta}_t$, the respective performance curve falls least (black solid line). The performance curve of the unadjusted PD (grey solid line) falls a bit more than the one from the break adjusted PD (black dashed line). The former reaches its bottom around 1968 and starts to follow the positive trend that is led by the performance lines of the two adjusted predictors. All three

[13] It is worth to point out that the alternative state processes are re-estimated at each period given available sample information. Thus, the varying mean processes of the PD change with each forecast origin.

[14] One can look at the performance for any OOS periods by redrawing a horizontal line at the start of OOS periods. If the curve terminates at a higher (lower) point at the end of OOS periods, the alternative model has a lower (higher) RMSE over the OOS periods of interest.

predictors outperform the naive forecast in periods with oil price shocks in 1973/1974. The good performance of the unadjusted PD during this period has also been noted by Welch and Goyal (2008). Centering the PD around its slowly evolving mean, however, obtains the only predictor that sustains this positive trend until 1994. The performance curves of both the unadjusted PD and the break adjusted PD reach their peaks in the early 1980s and start to fall since then. From 1994, the performance of all three predictors drops dramatically, with adjusting the PD for its gradually time-varying mean dropping the least. The strongest performance deterioration (and weakest recovery since 1999) is observed when centering the PD around discrete shifts in mean. All in all, due to its adaptive potential in both turmoil and tranquil periods, adjusting the PD by the slowly evolving mean offers superior ex-ante signalling.

Table 5 about here

It is worth noting that it is rare to observe a predictor outperforming the historical average return for forecasting returns out of sample for the entire period from 1946 to 2013. We have considered 13 other popular predictors as documented in Goyal and Welch (2003).[15] Only one of these predictors, the earnings to price ratio, has outperformed the historical average return showing a significantly positive R^2_{oos} statistic (see the column under 'e/p' in Table 5). This result confirms the importance of a time-varying mean of the PD in predicting returns in real time. Also, in-sample R^2 statistics are quite small for the considered period. Only six predictors obtain a R^2 statistic in excess of 0.01 with the highest R^2 being 0.0621 for a predictive regression based on net equity expansion (see the column under 'ntis' in Table 5).

3 LONG-RUN DETERMINANTS OF THE PD

The generalization of the present value model in Campbell and Shiller (1988) and its state space representation in Section 1 allows to extract a gradually time-varying mean of the PD from stock market data. Given the evidence in favour of trends to govern the PD

[15] Data are obtained from Amit Goyal's website. We consider all predictors for which data are available from 1926 to 2013.

locally in terms of likelihood diagnostics and OOS forecasting performance, it is tempting to address if this trend is shared by economic fundamentals. Diagnosing a cointegrating relation among the local mean of the PD and economic fundamentals would establish a powerful link among financial markets and the underlying economy. Moreover, complementing a cointegration analysis with indications of weak exogeneity is informative to address the transmission channel between the considered variables towards equilibrium.[16] We investigate three important factors that have been individually documented to affect the PD in a long-run manner – consumption risk, the demographic structure of the population, and the dividend payout policy of firms. We find that all three factors jointly shape the slowly evolving mean of the PD, and diagnose consumption volatility to be the most important influence. Violations of the long run equilibrium among the four series are mostly channelled through adjustments of the PD. In the following we discuss the considered factors and provide evidence from a cointegration analysis to assess their explanatory content. A detailed description of the variables is given in Appendix A.

3.1 The Three Long-Term Determinants

Consumption risk The influence of macroeconomic uncertainty on asset prices and equity premia has been long recognized in the asset pricing literature.[17] More recent studies such as Bansal and Yaron (2004) and Lettau et al. (2008) use recursive Epstein and Zin (1989) preferences, and demonstrate that a rise in consumption volatility can raise the expected return and lower asset prices. Empirically, Lettau et al. (2008) show low frequency evidence while Bansal et al. (2005) provide higher frequency evidence on the contribution of lower consumption volatilities to higher asset prices particularly since the 1990s. Bansal et al. (2005) show that measures of consumption volatility have good in-sample predictive power for the one-step ahead quarterly PD if historical volatilities

[16] As a conceptual alternative, one might consider to rewrite the right hand side of the state equation (7) as a linear index of considered trending economic indicators, and estimate all model parameters in one step. Following such lines, however, the considered economic processes would govern the PD in mean, by assumption, leaving little room for the data to object against the model specification and implicit transmission patterns. Moreover, it is not straightforward to contrast the varying mean model against the static benchmark in Campbell and Shiller (1988) in such a one-step framework if economic indicators are trending.

[17] See, among others, Gennotte and Marsh (1992), Giovannini (1989) and Kandel and Stambaugh (1990).

are extracted from short time windows of one or two years of consumption data. Lettau et al. (2008) argue in favour of a regime change in consumption risk to explain a regime change in asset valuations. The estimated regime is very persistent. The lower volatility regime reached in the early 1990s is expected to last for 30 years.

We adopt the consumption risk measure used by Bansal et al. (2005) in a low-frequency manner, in order to explain the gradually time-varying mean (low-frequency movements) of the PD. The consumption volatility is measured as $cr_t^W = \ln\left(\sum_{i=0}^{W-1} |co_{t-i}|\right)$, where co_t denotes the centered annual growth rate of per capita consumption and W is the size of rolling time windows.[18] We employ data on the per capita personal consumption expenditures on non-durable goods and services of the Bureau of Economic Analysis starting in 1929. To initialize time series of consumption risk we combine this series with the historical data on real per capita consumption recently collected by Barro and Ursua (2008).

To measure macroeconomic risk at low frequency one has to select W such that respective time windows carry informational content beyond short-run cycles. Figure 6 displays the absolute consumption growth with its Hodrick Prescott trend (the smoothing parameter is $\lambda = 100$). This trend visualizes the cyclical pattern of the consumption volatility. Counting from trough to trough, the length of the cycles are 30 (1840 to 1870), 18 (1870 to 1888), 22 (1888 to 1910), 44 (1910 to 1954), and 27 years (1954 to 1981). The 44 year cycle seems to be the odd one out, and could be regarded as containing two adjacent cycles – a 17 year cycle (1910 to 1927) around the WWI era and a 27 year cycle (1927 to 1954) around the WWII era. The durations of remaining cycles range from about 20 to 30 years. This forms our focus on alternative window lengths $W = 20, 21, \ldots, 30$ to calculate time local long-run consumption volatility cr_t^W. A lower bound of 20 years is consistent with the so-called Kuznets swings in economic growth (e.g. Solomou 2008). In addition, as argued by Geanakoplos et al. (2004) it is reasonable to assume that agents consider a 20 year horizon to incorporate demographic trends in long term asset price

[18] To calculate the volatility we use consumption growth directly instead of respective AR(1) regression residuals as considered in Bansal et al. (2005), since we do not detect any significant pattern of serial correlation in annual quotes of co_t.

expectations. An upper bound of 30 years coincides with the estimated average duration of a regime of consumption volatility in Lettau et al. (2008).

Figure 6 about here

The upper right panel of Figure 7 depicts cr_t^W with $W = 20, 25, 30$ as examples. The shapes of all three consumption measures are similar, and become a bit smoother as the window size W increases. It appears that macroeconomic uncertainty decreased continuously from the 1940s until the 1960s. It remained relatively stable during the 1970s and 1980s and then decreased further from the 1990s till present. Comparing cr_t^W with $\widetilde{\eta}_t$ depicted in the upper left panel of Figure 7, it seems that consumption risk is negatively related to movements in the gradually time-varying mean of the PD throughout the entire sample for all considered window lengths W.

Figure 7 about here

Demographics By means of an overlapping generation model Geanakoplos et al. (2004) provide the foundation for a long-run positive relationship between the PD and demographic trends. They argue that agents' incentives for holding equity vary over the life cycle. While the younger population intends to consume and willingly borrows for this purpose, the middle aged population concentrates more on saving and consumes these savings after retirement. The overall shape of the population pyramid is measured by means of the so-called middle-aged to young ratio (my_t). Geanakoplos et al. (2004) show that when my_t is large, there is excess demand for saving and equilibrium asset prices should increase to encourage consumption and to clear the market. This is consistent with price increases in the US stock market during the 1990s. Favero et al. (2011) demonstrate empirically the joint significance of my_t and the PD by means of long-horizon predictive regressions for stock returns, and diagnose a cointegration relationship between log dividends, log prices and my_t. These findings support the view that a slowly evolving mean of the PD could be driven by my_t.

Empirically my_t is defined as the ratio of the population aged 40-49 to the 20-29 year old, which is depicted in the lower left panel of Figure 7. Data is obtained from the US

Census Bureau. The middle-aged to young ratio shows a marked U-turn since the 1960s. This is mainly influenced by the baby boom after WWII. Beginning with the 1960s the baby boom generation affected the statistics for the young population, thereby reducing my_t. For the same reason, the ratio has been increasing since the 1980s when the baby boom generation became middle-aged. The twin peaks around 1960 and 2000 in my_t are related to the two major increases in the PD. The increases in my_t in the 1950s and the 1980s correspond to the increases in $\widetilde{\eta}_t$ in the 1960s and the 1990s, respectively.

Dividend payout policy Dynamics of the PD can also be influenced by changes in the dividend payout policy by firms (see Fama and French 2001, Robertson and Wright 2006, Boudoukh *et al.* 2007). The proportion of firms paying cash dividends fell from 66.5% in 1978 to 20.8% in 1999 (Allen and Michaely 2003).

Kim and Park (2013) show that the changing dividend payout policy affects the long-run relationship between prices and dividends. Both the proportion of firms that pay out a significant fraction of their earnings in the form of dividends and the cointegration coefficient between stock prices and dividends have followed a decreasing trend since the 1950s. A time-varying cointegration coefficient is an alternative interpretation of persistence governing the PD, and, hence, it is consistent with a time-varying mean of the PD in this paper. If the proportion of firms with traditional payout policy results in the time-varying cointegration coefficient between prices and dividends, it should also influence the gradually time-varying mean of the PD.

The proportion of firms with traditional payout policy among all firms in the (CRSP) value-weighted market portfolio (tp_t) decreases from 87% in 1946 to 35% in 2008, as can be seen in the lower right panel of Figure 7.[19] The downward trend in the proportion of firms with traditional payout policy since 1980 is consistent with the acceleration of $\widetilde{\eta}_t$ particularly since the 1990s. Lowered dividends may result in persistent increases of the PD.

[19] Data on tp_t for the sample period from 1946 to 2008 has been kindly provided by C.J. Kim and C. Park.

3.2 Cointegration Analysis

Unit Root Diagnosis First, we consider the individual characteristics of each variable by means of unit root tests. Unit root diagnostics for levels and first differences of $\widetilde{\eta}_t$, cr_t^W, my_t and tp_t are documented in Table 6. Almost all tests indicate first order integration of $\widetilde{\eta}_t$, cr_t^W and tp_t at conventional significance levels. The unit root hypothesis is not rejected for all cr_t^W measures with $W = 20, .., 30$. Results for cr_t^W with $W = 25$ are shown in Table 6 as an example. Although unit root tests hint at stationarity of my_t, these results are to be taken with caution. Eyeballing my_t hardly supports mean stationarity of this process. The null hypothesis of stationarity is rejected with 10% significance by means of the KPSS test. When longer ranges of data are considered, evidence on unit roots governing my_t can be found. We follow Favero et al. (2011) and treat my_t as a first order integrated process.

Table 6 about here

Error Correction Model To test for a cointegration relation among all four variables and to estimate the cointegration coefficients, we employ the conditional single equation error correction model (SECM).[20] With given presample values the SECM reads as

$$\begin{aligned} \Delta\widetilde{\eta}_t &= \delta_0 + \alpha\left(\widetilde{\eta}_{t-1} + \beta_1 cr_{t-1}^W + \beta_2 my_{t-1} + \beta_3 tp_{t-1}\right) + \delta_1 \Delta cr_t^W \\ &\quad + \delta_2 \Delta tp_t + \delta_3 \Delta my_t + \sum_{i=1}^{2} \phi_i \Delta\widetilde{\eta}_{t-i} + e_t,\ t = 1, 2, \ldots, T. \end{aligned} \tag{11}$$

The SECM specifies error correction dynamics conditional on current adjustments of weakly exogenous variables. It allows efficient inference by means of simple (non-linear) least squares estimation (see also Kremers, Ericsson, and Dolado 1992). As a particular merit it offers a parsimonious representation that does not suffer from weakened estimator precision in comparison with full dimensional maximum likelihood estimation of a vector

[20] As a preliminary analysis of cointegration relations, we look at the possibility of bivariate cointegration relations between the gradually time-varying mean of the PD $\widetilde{\eta}_t$ and each of the three long run determinants – consumption risk, demographics and the proportion of firms with traditional payout policy. We do not find evidence in support of any of the three bivariate long run relations (not shown). This hints at the importance of taking into account all three different influences on the PD jointly.

error correction model (Boswijk 1995, Johansen 1992). Model parsimony is beneficial in the present case of limited sample information. The estimation period starts in 1946 and ends in 2008 due to the nonavailability of the dividend payout ratio for earlier and later periods. To improve upon estimation uncertainty further, we apply a sequential estimation procedure eliminating in each step the short term coefficients δ_i, $i = 1, 2, 3$, and ψ_i, $i = 1, 2$, with the lowest t-statistic and lacking 30% significance.[21] Adopting a general-to-specific model composition, we start the model reduction from the SECM including two lags of the dependent variable which are necessary to capture patterns of serial correlations when testing for cointegration or weak exogeneity.

We find evidence for a cointegration relation between $\widetilde{\eta}_t$, cr_t^W, my_t and tp_t, where $W \in 23, ..., 29$.[22] For a significant cointegration relation, the absolute value of the t-statistic of the adjustment coefficient $\hat{\alpha}$ has to be larger than a respective non-standard critical value. For the specification in (11) with $W = 25$ as an example, the t-statistic of the adjustment coefficient is -3.72 while the 10% critical value is -3.45. As can be seen from Table 7, $\hat{\alpha}$ estimates for $W \in 23, ..., 29$ are smaller than zero with 10% significance. Estimating consumption risk from time windows of lengths $W = 20, 21, 22, 30$, we obtain similar degrees of explanation (see R^2 and adjusted R^2) and estimates for the cointegration coefficients, although a significant cointegration relation cannot be diagnosed within the SECM. Using the SECM in (11) is supported by diagnosing weak exogeneity of consumption risk (for $W \in 21, 23, \ldots, 30$), of the demographic factor and of the proportion of firms with traditional payout policy.[23] Hence, violations of the long run equilibrium among the PD and its persistent determinants are channelled through adjustments of the PD.

Table 7 about here

Significant effects from all three factors – consumption risk, the demographic structure

[21] Using a liberal significance level for the removal of single variables from the model, joint insignificance of the removed variables is likely for common (more conservative) significance levels, 5% say.

[22] Johansen trace tests confirm the diagnosis of a cointegration rank of unity with 5% significance.

[23] We test for weak exogeneity by means of autoregressive models of order two augmented by the long-run relation between $\widetilde{\eta}_{t-1}$, cr_{t-1}^W, my_{t-1} and tp_{t-1} as specified in (11). All respective adjustment coefficients are insignificant at the 10% level (not shown). Using, e.g., $W = 25$ to quantify consumption risk, the respective p-values for testing responses of Δcr_t^W, Δtp_t and Δmy_t are 0.315, 0.233 and 0.204.

and the dividend payout ratio – on the mean of the PD can be confirmed. Focusing on $W \in 23, ..., 29$, all estimated cointegration parameters are significant (an exception is $\hat{\beta}_3$ for $W = 28, 29$) and have the expected sign (see Table 7).[24] Both consumption risk (cr_t^W) and the proportion of firms with traditional payout policy (tp_t) have a negative influence on $\widetilde{\eta}_t$ and, thus, the signs of $\hat{\beta}_1$ and $\hat{\beta}_3$ in (11) shall be positive. For the demographic factor (my_t), it is the opposite case and the sign for $\hat{\beta}_2$ shall be negative.[25] In addition, the variations in the estimates for the coefficients attached to cr_t^W and my_t are small - ranging from 0.53 to 0.68 and from -1.09 to -0.92 respectively. Estimates for the coefficient of tp_t exhibit some larger variation and range between 0.42 and 0.74.

Cross Validations To gauge the relative importance of each long run determinant in a systematic way within the SECM approach, we employ cross-validation (CV) criteria (e.g. Picard and Cook 1984). While augmenting (reducing) the set of explanatory variables in regressions trivially goes along with gains (losses) in terms of in-sample model fit, CV criteria exhibit a nontrivial relation between a model's dimensionality and predictive content. The CV statistic is calculated as the mean absolute forecast error for $\Delta\widetilde{\eta}_t$, the left-hand side variable in the SECM (11). Specifically,

$$CV = \frac{1}{T}\sum_{t=1}^{T}\left|\Delta\widetilde{\eta}_t - \widehat{\Delta\widetilde{\eta}_t}\right|, \qquad (12)$$

where the forecast $\widehat{\Delta\widetilde{\eta}_t}$ for period t is based on a model of $\Delta\widetilde{\eta}_t$ that is estimated leaving out sample information (both dependent and explanatory variables) in period t. In this sense, $\widehat{\Delta\widetilde{\eta}_t}$ is an out-of-sample forecast for $\Delta\widetilde{\eta}_t$.

To unravel the relative importance of each determinant ($cr_{t-1}^W, my_{t-1}, tp_{t-1}$), we consider three different sets of models to obtain $\widehat{\Delta\widetilde{\eta}_t}$. The first is the SECM (11), to which we refer as the full model. The second set includes bivariate models of $\widetilde{\eta}_t$ and one of the

[24] To explore the sensitivity of these results, we also apply the dynamic least squares (DOLS) approach proposed by Stock and Watson (1993) to evaluate the sign and significance of the cointegration parameters. Test regressions include one lead and one lag of differentiated variables. DOLS estimates support significant influences of cr_t^W, my_t and tp_t of the right sign for $W = 23, \ldots, 29$.

[25] It is noteworthy that we obtain estimates with correct signs for all three cointegration parameters also for $W \in (3, 20)$ (not shown).

three determinants. And the third type includes trivariate models of $\widetilde{\eta}_t$ including two of the three determinants. A particular determinant is regarded as more informative for the mean of the PD if either its CV statistic from the bivariate model is close to that of the full model, or the CV statistic from the trivariate model without this determinant indicates a deterioration of the CV statistic.

Table 8 about here

Among the three factors consumption risk is most informative for changes in the mean of the PD while changes in the payout policy of firms appear to be least informative. Table 8 documents CV statistics from the full model (Panel A) along with the ratio of the CV statistics from the bivariate (Panel B) and trivariate model (Panel C) to those from the corresponding full model. Focusing on Panel B, we can see that using cr_{t-1}^W in a bivariate model leads to markedly smaller loss than using my_{t-1} or tp_{t-1} for all window sizes W. Using $W = 25$ as an example, the bivariate models with cr_{t-1}^W, my_{t-1} or tp_{t-1} have higher CV statistics than those of the full model by 5.2%, 9.7% and 13.7%, respectively. Conditional on the statistics documented in Panel C, cr_{t-1}^W and my_{t-1} appear to be comparably informative for the changes of the mean of the PD. By removing cr_{t-1}^W or my_{t-1} from the full model, the corresponding CV statistics increase by similar proportions (around 10% for most window sizes). In contrast, the removal of tp_{t-1} shows little effect on the CV outcome.[26]

4 CONCLUSIONS

In this paper, we consider a slowly evolving mean of the price-to-dividend ratio in the US, which is inspired by persistent dynamics of this series. We relax the assumption of a constant mean in the present value model (Campbell and Shiller 1988) towards a

[26] In an in-sample framework likelihood ratio (LR) statistics can assess the significance of distinct model fits based on residual variances. We apply the 95% quantile of a χ^2-distribution with one degree of freedom as approximate critical values to adopt this framework to mean squared cross validation errors. All CV statistics from bivariate models are significantly different from the respective statistics of full models. In contrast, LR-type statistics for the trivariate models without cr_t^W or my_t are significant while those of models excluding tp_t are insignificant. Hence, this is further evidence that tp_t is less important than cr_t^W and my_t to shape $\widetilde{\eta}_t$.

gradually time-varying mean of the PD, and formalize a state space model to estimate its latent path. Log-likelihood statistics support the model. Adjusting the PD by its slowly evolving mean is fruitful in out-of-sample forecasting of stock returns. It outperforms both adjusting the PD for structural mean shifts, and the historical average return as a common benchmark predictor. A cointegration analysis underpins that trends in the PD are shared with persistent patterns governing consumption risk, the demographic structure of the population and firm's dividend payout policy. While these determinants play significant roles in jointly shaping the slowly evolving mean of the PD, consumption risk turns out to be the dominant force.

As future research it would be interesting to compare the gradually time-varying mean of the PD from different markets and to uncover potential common components in their variations. International risk sharing could be one potential (global) determinant. As Artis and Hoffmann (2008) have pointed out, international risk sharing has increased since financial markets became more integrated in the 1980s. This might have played an important role in determining variations in the long-run PD of different markets in this period.

APPENDICES

In the following we provide further details on the analyzed data (Appendix A), the particle filtering approach to the estimation of the state space model (Appendix B), and discuss the approximation errors involved when deriving the observation equation of the state space model by means of a Taylor expansion (Appendix C).

APPENDIX A: DATA DESCRIPTION

S&P500 Stock Market Indices and Dividends. Annual series are provided by Amit Goyal and available on the internet.[27] They contain the S&P500 index based on end-of-year closing *prices* and corresponding *dividends* for the period from 1871 to 2013.

[27] http://www.hec.unil.ch/agoyal/

Annual dividends correspond to the sum of the four quarterly paid dividends within the corresponding year. For more details see Welch and Goyal (2008).

CRSP Stock Market Indices and Dividends. From 1926 to 2013 we apply annual end-of-year returns based on the weighted market portfolio (NASDAQ, NYSE, AMEX) of the Center for Research in Security Prices. We follow Lettau and Van Nieuwerburgh (2008) and calculate the prices from the return excluding dividend payments and the dividends from the dividend yield D_t/P_{t-1}. From 1871 to 1925 we apply the end-of-year S&P500 index and corresponding dividends employed in Welch and Goyal (2008) and described above. Annual dividends correspond to the sum of the four quarterly paid dividends within the corresponding year.

Interest Rates and Inflation. Similar to Campbell and Vuolteenaho (2004), we use a *short term rate* based on 3-month US Treasury Bills of the Federal Reserve System to approximate the risk-free rate. We employ the series provided by Amit Goyal for the period from 1871 to 2013 which is available from the internet.[28] More details can be found in Welch and Goyal (2008).

The annual *inflation* series from 1871 to 2013 is extracted from the consumer price index for all urban consumers as provided by Robert J. Shiller.[29] For more details see Shiller (1992, 2005).

Other Macroeconomic Variables. The ratio of the 40-49 over the 20-29 year aged *population* is determined by Census annual population data collected from Datastream (period since 1950, 'USPOP24Y' for the 20-24 year old agents, 'USPOP29Y' for the 25-29, 'USPOP44Y' for the 40-44 and 'USPOP49Y' for the 45-49). Data for the period before 1950 are directly from the US Census Bureau.[30]

Annual quotes of real per capita *consumption* (1929 to 2013) are derived from the sum of the personal consumption expenditures on nondurables and services of the Bureau of

[28] http://www.hec.unil.ch/agoyal/
[29] http://www.econ.yale.edu/~shiller/data.htm
[30] http://www.census.gov/popest/archives/pre-1980/PE-11.html

Economic Analysis which are two subgroups of the US total personal consumption expenditures (Tables 2.3.5. 'personal consumption expenditures by major type of product' and '2.3.4. price indexes for personal consumption expenditures by major type of product').[31] The total US population is drawn from the sources described above (the corresponding datastream code is 'USPOPTO.'). For periods before 1929 we use the series of real per capita total consumption collected by Barro and Ursua (2008). This series ranges from 1834 to 2009 and is available from the net.[32] To join the sum of non-durables and services specific consumption measures with the data of Barro and Ursua (2008), we regress the sum of both BAE series (co_t^{BEA}) on a constant and the series of Barro and Ursua (co_t^{BU}) in the overlapping sample (1929-2009) and estimate the pre-1929 data from the latter source. The estimated regression is $co_t^{BEA} = 2.932 + 1.012 co_t^{BU} + \hat{\nu}_t$ with a R^2 of 0.998.

For information regarding the measurement of the share of firms paying traditional dividends the reader may consider Kim and Park (2013).

APPENDIX B: PARTICLE FILTERING

The state space model of the price-to-dividend ratio in (6) and (7) is highly nonlinear in the latent state, and the maximization of the corresponding log-likelihood function is not tractable analytically. Using particle filtering (a Monte Carlo technique) it becomes possible to derive an approximative log-likelihood value by means of particle filtering. We apply the standard particle filter described in Cappé et al. (2007) (Algorithm 3, bootstrap filter) and an optimization technique based on the simplex search method of Lagarias *et al.* (1998) for parameter estimation that does not depend on the gradient of the log-likelihood function. The applied particle filtering algorithm involves the following steps:

Step (1): Initialization (t=1). Sample N particles $\underline{\widetilde{\eta}}_1^{(i)} \sim N(\widetilde{\eta}_0, \sigma_u^2)$, *i=1,...,N, and*

[31] http://www.bea.gov/itable/index.cfm

[32] http://scholar.harvard.edu/barro/publications/macroeconomic-crises-1870-bpea

determine importance weights and normalized weights, respectively, as

$$\underline{w}_1^{(i)} = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{1}{2}\left(\underline{\epsilon}_1^{(i)}/\sigma_\epsilon\right)^2\right) \quad \text{and} \quad w_1^{(i)} = \frac{\underline{w}_1^{(i)}}{\sum_{i=1}^{N} \underline{w}_1^{(i)}}.$$

Step (2): Iteration (t=2,...,T).

1. *Select N particles according to weights $w_{t-1}^{(i)}$. Set accordingly $\widetilde{\eta}_{t-1}^{(i)} = \underline{\widetilde{\eta}}_{t-1}^{(i)}$ (resampling)*

2. *For all particles draw*

$$\underline{\widetilde{\eta}}_t^{(i)} \sim N(\widetilde{\eta}_{t-1}^{(i)}, \sigma_u^2),\ i = 1, ..., N,$$

   and determine raw and normalized weights, respectively, as

$$\underline{w}_t^{(i)} = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{1}{2}\left(\underline{\epsilon}_t^{(i)}/\sigma_\epsilon\right)^2\right) \quad \text{and} \quad w_t^{(i)} = \frac{\underline{w}_t^{(i)}}{\sum_{i=1}^{N} \underline{w}_t^{(i)}}.$$

3. *go back to step '1'.*

Averaging over non-normalized weights $\underline{w}_t^{(i)}$ yields estimates of the contribution of ϵ_t to the Gaussian likelihood function, while averaging over draws $\widetilde{\eta}_t^{(i)}$, i.e., $\hat{\widetilde{\eta}}_t = \frac{1}{N}\sum_{i=1}^{N} \widetilde{\eta}_t^{(i)}$, results in estimates of $\widetilde{\eta}_t$, for $t = 1, ..., T$.

The so-called systematic resampling is used to compute uniformly distributed random numbers to implement the resampling step. This technique is described in Robert and Casella (2005). Doucet and Johansen (2009) argue that such a technique reduces the noise introduced by resampling, and it is commonly employed in the related literature.

APPENDIX C: APPROXIMATION ERRORS

As outlined in the main text, the following approximations have been made to derive the present value representation of the PD (see also Lettau and Van Nieuwerburgh 2008): (i) $E_t[\rho_{t+i}] \approx \rho_t$, (ii) $E_t[\kappa_{t+i}] \approx \kappa_t$ and (iii) $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$. While these approximations facilitate present value determination, eventual approximation errors are rather small for the variation ranges of involved variables. Given the observed range of $\widetilde{\eta}_t \in [2.8, 4.1]$, the respective parameter supports for both ρ and κ are implicit. In the following, we show that the local linear approximations of the underlying nonlinear functions leading to (i) and (ii) result in small approximation errors due to the very small local degree of concavity of the functions. A simulation exercise with data driven parameter settings further confirms that the average approximation error for (iii) is negligible. Finally, we evaluate the total approximation errors from all three approximations by comparing the right-hand side of the present value equation of the PD in (6) with its exact counterpart (without the approximations). Simulation results confirm that the total approximation error is small.

C.1 Approximation $E_t[\rho_{t+i}] \approx \rho_t$

Based on the empirical observations in Section 1, it becomes reasonable to assume η_t, and also $\widetilde{\eta}_t$, to follow a random walk. In principle, this martingale characteristic implies constant expectations of the gradually time-varying mean of the PD $\widetilde{\eta}_t$. To derive in (3) a function of returns r_{t+1} which is linear in η_t we apply a first order Taylor approximation based on $\rho_t \equiv 1/(1 + \exp(-\widetilde{\eta}_t))$. In consequence, ρ_t is concave in $\widetilde{\eta}_t$ and therefore $E_t(\rho_{t+i}) \leq \rho_t$ by Jensen's inequality. However, as displayed in the upper panel of Figure C1 the function $\rho_t \equiv 1/(1 + \exp(-\widetilde{\eta}_t))$ is approximately linear in the domain of $\widetilde{\eta}_t \in [2.8, 4.1]$.

Figure C1 about here

To evaluate the degree of concaveness of $\rho_t \equiv 1/(1 + \exp(-\widetilde{\eta}_t))$ and the impact of the approximation error we compute the difference between $b\rho(\widetilde{\eta}_t^1) + (1 - b)\rho(\widetilde{\eta}_t^2)$ and $\rho(b\widetilde{\eta}_t^1 + (1 - b)\widetilde{\eta}_t^2)$ for any $b \in [0, 1]$ and $\widetilde{\eta}_t^1, \widetilde{\eta}_t^2 \in [2.8, 4.1]$. The maximal error is 0.0061 in absolute terms and, thus, relatively small.

C.2 Approximation $E_t[\kappa_{t+i}] \approx \kappa_t$

In the first order Taylor expansion κ_t is determined as $\kappa_t \equiv -\ln(\rho_t) - (1-\rho_t)\ln(1/\rho_t - 1)$. Thus, κ_t is also concave in ρ_t and $E_t[\kappa_{t+i}] \leq \kappa_t$ by Jensen's inequality. In the relevant domain of $\rho_t \in [0.943, 0.984]$ implied directly by $\widetilde{\eta}_t \in [2.8, 4.1]$, $\kappa_t = -\ln(\rho) - (1-\rho)\ln(1/\rho - 1)$ is approximately linear as displayed in the middle panel of Figure C1. The maximal difference between $b\kappa(\rho_t^1) + (1-b)\kappa(\rho_t^2)$ and $\kappa(b\rho_t^1 + (1-b)\rho_t^2)$ for any $b \in [0,1]$ and $\rho_t^1, \rho_t^2 \in [0.943, 0.984]$ is 0.0064 in absolute terms and relatively small. Thus, the approximation error in $E_t[\kappa_{t+i}] \approx \kappa_t$ is negligible.

C.3 Approximation $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$

To evaluate the magnitude of the error implied by this approximation we perform a simulation study. The parameter estimations $\widetilde{\eta}_0 = 2.892$ and $\sigma_u = 0.059$ from Table 2 are applied to simulate the process $\widetilde{\eta}_t = \widetilde{\eta}_{t-1} + u_t$ as a random walk, for $t = 1, \ldots, T$. To reflect the range of the empirical PD we bound the random walk by the minimum of the empirical PD (2.288) and the maximum of the empirical PD (4.495). The process of $\rho_t = 1/(1+\exp(-\widetilde{\eta}_t))$ is simulated subsequently. The innovations u_t are generated as $N(0, \sigma_u)$. Further, we separate the dataset into $t = 1, \ldots, T_1, T_1+1, \ldots, T$ and neglect the first T_1 observations as initialization period. To simulate η_t we add a first order moving average process, obtaining

$$\eta_t = \widetilde{\eta}_t + \alpha\omega_{t-1} + \omega_t. \tag{C1}$$

The moving average specification for ω_t accounts for correlation of leads and lags. The parameter α and the standard deviation of ω_t are estimated based on our empirical data ($\alpha = 0.726$ and $\sigma_\omega = 0.188$). We draw innovations ω_t from $N(0, \sigma_\omega)$.

To approximate $E_t[\rho_{t+i}\eta_{t+i+1}]$ we define $\psi_t = \rho_t\eta_{t+1}$ and forecast at each point in time t $\hat{\psi}_{t+i}$ by means of estimated AR(10) processes, for $i = 1, \ldots, H$. With regard to $E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ we estimate AR(10) processes for each series separately and determine at each point in time t the corresponding forecasts $\hat{\rho}_{t+i}$ and $\hat{\eta}_{t+i+1}$, for $i = 1, \ldots H$. To estimate all AR(10) models we separate the data set as $t = T_1+1, \ldots, T_2, T_2+1, \ldots, T$ and apply a recursive window starting with the first $T_2 - T_1$ observations. Thus, in total we are left with $T - T_1 - T_2$ periods for which we compute H forecasts. If the approximation error implied by setting $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ is small the product of the two separate forecasts $\hat{\rho}_{t+i}$ and $\hat{\eta}_{t+i+1}$ should come close to the forecast $\hat{\psi}_{t+i}$.

This procedure is repeated R times and calculations are stored at each point in time as $\hat{\psi}_{r,t+i}$, $\hat{\rho}_{r,t+i}$ and $\hat{\eta}_{r,t+i+1}$, for $r = 1, \dots, R$ and $i = 1, \dots, H$. To determine the approximation error of interest we use the following statistic

$$\bar{\Omega}_i = \frac{1}{R(T - T_1 - T_2)} \sum_{r=1}^{R} \sum_{t=T_2+1}^{T} \Omega_{r,t+i}, \quad \text{for } i = 1, 2, \dots, H, \tag{C2}$$

where

$$\Omega_{r,t+i} = \frac{|df_{r,t+i}|}{\hat{\psi}_{r,t+i}} \text{ and } |df_{r,t+i}| = |\hat{\psi}_{r,t+i} - \hat{\rho}_{r,t+i}\hat{\eta}_{r,t+i+1}|. \tag{C3}$$

We set $R = 1000$, $T = 2000$, $T_1 = 500$, $T_2 = 500$ and $H = 100$. The lower panel of Figure C1 displays $\bar{\Omega}_i$, for $i = 1, \dots, 100$. With increasing forecast horizons the average approximation error converges. It reaches not more than 1% for 100-step ahead forecasting. As a result, the magnitude of this error is rather small.

C.4 Overall Assessment

Although the approximation errors from C.1-C.3 are small individually, one can argue that they might accumulate when iterating equation (3) with these approximations forward. To check for this possibility, we compare the present value of the PD in (5) based on approximations C.1-C.3,

$$\begin{aligned}
\eta_t^{(1)} &= \kappa_t(1 + \rho_t + \rho_t^2 + \cdots + \rho_t^{T-1}) + E_t[\Delta d_{t+1} - r_{t+1}] + \rho_t E_t[\Delta d_{t+2} - r_{t+2}] \\
&+ \rho_t^2 E_t[\Delta d_{t+3} - r_{t+3}] + \cdots + \rho_t^{T-1} E_t[\Delta d_{t+T} - r_{t+T}] + \rho_t^T E_t[\eta_{t+T}],
\end{aligned}$$

with its exact counterpart,

$$
\begin{aligned}
\eta_t^{(2)} &= \kappa_t + \rho_t E_t[\kappa_{t+1}] + \rho_t E_t[\rho_{t+1}\kappa_{t+2}] + \cdots + \rho_t E_t[\rho_{t+1}\rho_{t+2}\cdots\rho_{t+T-2}\kappa_{t+T-1}] \\
&+ E_t[\Delta d_{t+1} - r_{t+1}] + \rho_t E_t[\Delta d_{t+2} - r_{t+2}] + \rho_t E_t[\rho_{t+1}(\Delta d_{t+3} - r_{t+3})] + \cdots \\
&+ \rho_t E_t[\rho_{t+1}\rho_{t+2}\cdots\rho_{t+T-2}(\Delta d_{t+T} - r_{t+T})] + \rho_t E_t[\rho_{t+1}\rho_{t+2}\cdots\rho_{t+T-1}\eta_{t+T}].
\end{aligned}
$$

For this MC experiment, we simulate $\widetilde{\eta}_t$ as a random walk (see C.3), and η_t according to (C1). The processes ρ_t and κ_t are determined subsequently according to (4). We build objective expectations of the varying mean of the PD by forecasting $\{\widetilde{\eta}_{t+i}\}_{i=1}^{T-1}$ with the estimated AR(1) processes. Forecasts of $\widetilde{\eta}_t$ are bounded by the empirical minimum and maximum. Objective expectations of $\{\rho_{t+i}, \kappa_{t+i}\}_{i=1}^{T-1}$ are then calculated accordingly. The expected return-adjusted dividend growth rate is simulated as an AR(2) process based on estimates from observed data, i.e.,

$$
(\Delta d_t - r_t) = -0.051 - 0.3(\Delta d_{t-2} - r_{t-2}) + e_t, \text{ with } e_t \sim N(0, 0.164).
$$

The expected PD at period $t+T$ is taken from the simulated process. We generate 1000 replications with $t = 1000$ to take advantage of the consistency of AR parameter estimates and $T = 1000$ to capture high horizons of future cash flows. As documented in Table C1, the total approximation errors (absolute difference between $\eta_t^{(1)}$ and $\eta_t^{(2)}$) are very small, with a mean of 0.0448. In relative terms this is about 1% of the mean of the exact formulation ($\eta_t^{(2)}$).

Table C1 about here

LITERATURE CITED

Allen, Frankin, and Roni Michaely. (2003) "Payout Pollicy." In *North-Hoolad Handbook of Economics*, edited by George Constantinides, Milton Harris, and Rene Stulz, pp 337–429. Elsevier.

Artis, Michael J., and Mathias Hoffmann. (2008) "Financial Globalization, International Business Cycles and Consumption Risk Sharing." *Scandinavian Journal of Economics*, 110(3), 447–471.

Bai, Jushan, and Pierre Perron. (1998) "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1), 47–78.

Bai, Jushan, and Pierre Perron. (2003) "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics*, 18(1), 1–22.

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

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

Bansal, Ravi, Dana Kiku and Amir Yaron. (2010) "Long Run Risks, the Macroeconomy, and Asset Prices." *American Economic Review*, 100(2), 542–46.

Barro, Robert J., and Jose F Ursua. (2008) "Macroeconomic Crises since 1870." *Brookings Papers on Economic Activity*, 39(1), 255–350.

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

Boswijk, H. Peter. (1995) "Efficient Inference on Cointegration Parameters in Structural Error correction models." *Journal of Econometrics*, 69(1), 133–158.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts. (2007) "On the importance of measuring payout yield: Implicatins for empirical asset pricing." *Journal of Finance*, 62, 877–915.

Boudoukh, Jacob, Matthew Richardson, and Robert F Whitelaw. (2008) "The Myth of Long-Horizon Predictability." *Review of Financial Studies*, 41, 1577–1605.

Branch, William A. (2004) "The Thoery of Rationally Heterogeneous Expecations: Evidence from Survey Data on Inflation Expectations." *Economic Journal*, 114, 592–621.

Campbell, John Y. (1991) "A Variance Decomposition for Stock Returns." *Economic Journal*, 101(405), 157–179.

Campbell, John Y. (1999) "Asset Prices, Consumption, and the Business Cycle." In *Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford, vol. 1, pp. 1231 – 1303. Elsevier.

Campbell, John Y. (2008) "Viewpoint: Estimating the Equity Premium." *Canadian Journal of Economics*, 41(1), 1–21.

Campbell, John Y, and Robert J. Shiller. (1988) "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1(3), 195–228.

Campbell, John Y., and Tuomo Vuolteenaho. (2004) "Inflation Illusion and Stock Prices." *The American Economic Review*, 94(2), pp. 19–23.

Campbell, John Y., and Motohiro Yogo. (2006) "Efficient Tests of Stock Return Predictability." *Journal of Financial Economics*, 81(1), 27–60.

Cappé, Olivier, Simon Godsill, and Eric Moulines. (2007) "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo." *Proceedings of the IEEE*, 95(5).

Cavaliere, Giuseppe. (2005) "Limited Time Series with A Unit Root." *Econometric Theory*, 21(05), 907–945.

Cavaliere, Giuseppe, and Fang Xu. (2014) "Testing for Unit Roots in Bounded Time Series." *Journal of Econometrics*, 178(Part 2), 259–272.

Clark, Todd E., and Kenneth D. West. (2007) "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics*, 138, 291–311.

Doucet, Arnaud, and Adam M. Johansen (2009) "A Tutorial on Particle Filtering and Smoothing: Fifteen years Later." In *Handbook of Nonlinear Filtering*, edited by Dan Crisan and Boris Rozovsky. Oxford University Press.

Doucet, Arnaud, Nando De Freitas, and Neil Gordon. (2001) *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag.

Elliott, Graham, Thomas J Rothenberg, and James H. Stock. (1996) "Efficient Tests for an Autoregressive Unit Root." *Econometrica*, 64(4), 813–36.

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

Ericsson, Neil R., and James G. MacKinnon. (2002) "Distributions of Error Correction Tests for Cointegration." *Econometrics Journal*, 5(2), 285–318.

Fama, Eugene F., and Kenneth R. French. (2001) "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay." *Journal of Financial Economics Letters*, 60, 3–43.

Favero, Carlo A., Arie E. Gozluklu, and Andrea Tamoni. (2011) "Demographic Trends, the Dividend-Price Ratio, and the Predictability of Long-Run Stock Market Returns." *Journal of Financial and Quantitative Analysis*, 46(05), 1493–1520.

Fernández-Villaverde, Jesús, and Juan F. Rubio-Ramírez. (2005) "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood." *Journal of Applied Econometrics*, 20(7), 891–910.

Fernández-Villaverde, Jesús, and Juan F. Rubio-Ramírez. (2007) "Estimating Macroeconomic Models: A Likelihood Approach." *The Review of Economic Studies*, 74(4), 1059–1087.

Geanakoplos, John, Michael Magill, and Martine Quinzii. (2004) "Demography and the Long-Run Predictability of the Stock Market." *Brookings Papers on Economic Activity*, 2004(1), 241–307.

Gennotte, Gerard, and Terry A. Marsh. (1992) "Variations in Economic Uncertainty and Risk Premium on Capital Assets." *European Economic Review*, 37, 1021–1041.

Giovannini, Alberto. (1989) "Uncertainty and Liquidity." *Journal of Monetary Economics*, 23, 239–258.

Goyal, Amit, and Ivo Welch. (2003) "Predicting the Equity Premium with Dividend Ratios." *Management Science*, 49(5), pp. 639–654.

Herwartz, Helmut, and Leonardo Morales-Arias. (2009) "In-Sample and Out-of-Sample Properties of International Stock Return Dynamics Conditional on Equilibrium Pricing Factors." *European Journal of Finance*, 15(1), 1–28.

Johansen, Soren. (1992) "Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis." *Journal of Econometrics*, 52(3), 389–402.

Kandel, Shmuel, and Robert F. Stambaugh. (1990) "Expectations and Volatility of Consumption and Asset Returns." *Review of Financial Studies*, 3, 207–232.

Kim, Chang-Jin, and Cheolbeom Park. (2013) "Disappearing Dividends: Implications for the Dividend-Price Ratio and Return Predictability." *Journal of Money, Credit and Banking*, 45(08), 933–952.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib. (1998) "Stochastic volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, 65(3), 361–393.

Kremers, Jeroen J. M., Neil R. Ericsson, and Juan J. Dolado. (1992) "The Power of Cointegration Tests." *Oxford Bulletin of Economics and Statistics*, 54(3), 325–48.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. (1992) "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root : How Sure are We that Economic Time Series Have a Unit Root?" *Journal of Econometrics*, 54(1-3), 159–178.

Lacerda, Filipe, and Pedro Santa-Clara. (2010) "Forecasting Dividend Growth to Better Predict Returns." Working Paper. Universidade Nova de Lisboa.

Lagarias, Jeffrey C., James A. Reeds, Margaret H. Wright, and Paul E. Wright. (1998) "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions." *SIAM Journal of Optimization*, 9, 112–147.

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

Lettau, Martin, Sydney C. Ludvigson, and Jessica A. Wachter. (2008) "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?" *Review of Financial Studies*, 21, 1653–1687.

Newey, Whitney K, and Kenneth D. West. (1987) "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3), 703–08.

Newey, Whitney K, and Kenneth D. West. (1994) "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, 61(4), 631–53.

Ng, Serena, and Pierre Perron. (2001) "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica*, 1519–1554.

Park, Cheolbeom. (2010) "When does the Dividend-Price Ratio Predict Stock Returns?" *Journal of Empirical Finance*, 17(1), 81–101.

Paye, Bradley S., and Allan Timmermann. (2006) "Instability of Return Prediction Models." *Journal of Empirical Finance*, 13, 274–315.

Perron, Pierre, and Timothy J. Vogelsang. (1992) "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity." *Journal of Business and Economic Statistics*, 10(3), 301–20.

Phillips, Peter C. B., and Pierre Perron. (1988) "Testing for a Unit Root in Time Series Regression." *Biometrika*, 75(2), 335–346.

Picard, Richard R., and R. Dennis Cook. (1984) "Cross-Validation of Regression Models." *Journal of the American Statistical Association*, 79(387), 575–583.

Pitt, Michael K., and Neil Shephard. (1999) "Filtering via simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association*, 94(446), 590–599.

Rapach, David E., and Mark E. Wohar. (2006) "Structural Breaks and Predictive Regressin Models of Aggregate US Stock Returns." *Journal of Financial Eocnometrics*, 4, 238–274.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. (2010) "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23(2), 821–862.

Ristic, Branko, Sanjeev Arulampalam, and Neil Gordon. (2004) *Beyond the Kalman Filter*. Artech House.

Robert, Christian P., and George Casella. (2005) *Monte Carlo Statistical Methods*. 2nd edn. Springer Texts in Statistics.

Robertson, Donald, and Stephen Wright. (2006) "Dividends, Total Cash Flow to Shareholders, and Predictive Return Regressions." *The Review of Economics and Statistics*, 88(1), 91–99.

Rudd, Jeremy, and Karl Whelan. (2006) "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?" *American Economic Review*, 96(1), 303–320.

Rytchkov, Oleg. (2012) "Filtering Out Expected Dividends and Expected Returns." *Quarterly Journal of Finance*, 2(3), 1250012:1–1250012:56.

Sbordone, Argia M. (2002) "Prices and unit labor costs: A new test of price stickiness." *Journal of Monetary Economics*, 49(2), 265–292.

Shiller, Robert J. (1992) *Market Volatility*. MIT Press.

Shiller, Robert J. (2005) *Irrational Exuberance*. 2nd edn. Princeton: Princeton University Press.

Solomou, S. N. (2008) In *New Palgrave Dictionary of Economics*, 2nd edn., vol. 4, edited by Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

Stambaugh, Robert F. (1999) "Predictive Regressions." *Journal of Finance*, 54, 375–421.

Stock, James H, and Mark W. Watson. (1993) "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica*, 61(4), 783–820.

Vuong, Quang H. (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57(2), 307–333.

Welch, Ivo, and Amit Goyal. (2008) "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction" *Review of Financial Studies*, 21(4), 1455–1508.

Tables

Table 1: UNIT-ROOT TESTS FOR THE PD

	ADF	PP	DF_{GLS}	MZ_t	PV	$KPSS$
			CRSP 1926-2013			
Test Statistics	-1.595	-1.604	-1.171	-1.204	-3.553	0.801
Crit. val. at 5%	-2.895	-2.895	-1.945	-1.980	-4.580	0.463
Crit. val. at 10%	-2.585	-2.585	-1.614	-1.620	-4.240	0.347
			S&P500 1871-2013			
Test Statistics	-1.783	-1.841	-1.139	-1.078	-3.737	0.906
Crit. val. at 5%	-2.882	-2.882	-1.943	-1.980	-4.580	0.463
Crit. val. at 10%	-2.578	-2.578	-1.615	-1.620	-4.270	0.347

Notes: The upper panel covers test results for CRSP data ranging from 1926 to 2013 and the lower panel those for S&P500 data ranging from 1871 to 2013. Test regressions include a constant. ADF refers to the Augmented Dickey-Fuller test where the lag selection criterion is the BIC. For PP, the test statistic considered in Phillips and Perron (1988), the spectral OLS AR estimator is used to calculate the long-run variance. DF_{GLS} refers to the test proposed by Elliott et al. (1996) where the BIC is applied to determine the lag length. MZ_t refers to the modified Pillips-Perron Z_t statistic introduced in Ng and Perron (2001) which is implemented by means of a spectral OLS AR estimator and BIC based order selection. For the PV-test proposed by Perron and Vogelsang (1992) the innovation outlier model is applied and the lag length is determined by means of a t-test procedure. A Bartlett Kernel is applied in the $KPSS$-test of Kwiatkowski *et al.* (1992).

Table 2: MODEL ESTIMATES AND DIAGNOSTICS

Model	δ	α	$\widetilde{\eta}_0$	$\widetilde{\eta}_1$	$\widetilde{\eta}_2$	σ_u	σ_ϵ	log-lik	BIC	V-stat
					CRSP 1926-2013					
$\overline{\eta}$	-	-	3.409	-	-	-	0.184	23.9	-38.94	-7.669
$\overline{\overline{\eta}}_t^{(1)}$	-	-	3.129	3.668	-	-	0.100	77.9	-142.46	-5.874
$\overline{\overline{\eta}}_t^{(2)}$	-	-	2.921	3.199	3.670	-	0.096	81.6	-145.24	-5.530
$\widetilde{\eta}_{RW,t}$	-	-	**2.892**	-	-	**0.059**	**0.026**	**168.4**	**-323.84**	-
$\widetilde{\eta}_{AR,t}$	0.042	0.986	-	-	-	0.063	0.025	166.5	-314.99	-6.889
					S&P500 1901-2013					
$\overline{\eta}$	-	-	3.298	-	-	-	0.215	13.3	-17.08	-8.396
$\overline{\overline{\eta}}_t^{(1)}$	-	-	3.028	3.650	-	-	0.117	81.8	-149.50	-7.753
$\overline{\overline{\eta}}_t^{(2)}$	-	-	2.878	3.234	3.673	-	0.098	101.9	-184.84	-5.912
$\widetilde{\eta}_{RW,t}$	-	-	**3.074**	-	-	**0.076**	**0.033**	**196.1**	**-378.01**	-
$\widetilde{\eta}_{AR,t}$	0.046	0.985	-	-	-	0.075	0.031	195.8	-372.72	-2.776

Notes: This table documents estimation and diagnostic results for the static present value model ($\overline{\eta}$), specifications presuming one ($\overline{\overline{\eta}}_t^{(1)}$) and two ($\overline{\overline{\eta}}_t^{(2)}$) mean shifts of the PD, the benchmark model with time varying mean and RW state equation ($\widetilde{\eta}_{RW,t}$), and a corresponding model with AR state process ($\widetilde{\eta}_{AR,t}$). Estimates $\widetilde{\eta}_0$, are either initial means or the unconditional mean of the PD. Similarly, $\widetilde{\eta}_1$ and $\widetilde{\eta}_2$ are mean levels in models with discrete shifts. Diagnostic statistics comprise the log-likelihood, the BIC and Vuong tests (V-stat). Results for the benchmark model are displayed in bold face.

Table 3: IN-SAMPLE PREDICTIVE REGRESSIONS

	$\overline{\eta}$	$\overline{\overline{\eta}}_t^{(1)}$	$\overline{\overline{\eta}}_t^{(2)}$	$\widetilde{\eta}_t$
		Stock Returns		
β_0	0.0928	0.0937	0.0939	0.1142
	(4.6562)	(4.5205)	(6.0840)	(5.3922)
β_1	−0.0923	−0.2516	−0.4120	−0.2089
	(−2.5610)	(−4.8718)	(−6.0994)	(−2.7885)
R^2	0.0392	0.1027	0.1751	0.0641
adj R^2	0.0279	0.0922	0.1654	0.0531
		Dividend Growth		
β_0	0.0456	0.0457	0.0458	0.0539
	(3.3098)	(3.5017)	(3.3140)	(4.0314)
β_1	−0.0013	−0.0431	−0.1157	−0.0832
	(−0.0426)	(−0.9694)	(−1.7106)	(−1.8706)
R^2	0.0000	0.0057	0.0260	0.0192
adj R^2	−0.0117	−0.0060	0.0146	0.0076

Notes: This table documents statistics from regressions (9) with four alternative mean processes of the PD using CRSP data from 1926 to 2013. In the second column forecasts for returns (dividend growth) are conditioned on the unadjusted PD (using the overall sample mean $\overline{\eta}$). The third and fourth column contain the estimates based on the adjusted PD using one ($\overline{\overline{\eta}}_t^{(1)}$) or two ($\overline{\overline{\eta}}_t^{(2)}$) mean shifts, respectively. In the last column forecasts are conditioned on the PD adjusted by means of the smooth state process $\widetilde{\eta}_t$. Newey and West (1987) robust t-statistics for coefficient estimates are presented in parantheses. The bandwidth is selected by means of the procedure proposed by Newey and West (1994).

Table 4: OUT-OF-SAMPLE PREDICTIVE REGRESSIONS FOR STOCK RETURNS

	$\overline{r}_t$	$\overline{\eta}$	$\overline{\overline{\eta}}_t$	$\widetilde{\eta}_t$
		1946-2013		
RMSE	0.1694	0.1751	0.1793	0.1683
R^2_{oos}		−0.0691	−0.1202	0.0121***
		1946-2004		
RMSE	0.1607	0.1681	0.1714	0.1589
R^2_{oos}		−0.0934	−0.1367	0.0232***

Notes: This table documents the OOS forecasting performance of the naive prediction by means of historical average returns ($\overline{r}_t$) and alternative predictive regressions with the unadjusted PD ($s_t = \overline{\eta}$), the PD adjusted by mean shifts ($s_t = \overline{\overline{\eta}}_t$), and the PD adjusted by the smooth state ($s_t = \widetilde{\eta}_t$). Root mean squared errors (RMSE) and OOS R^2 statistics (R^2_{oos}) are shown. R^2_{oos} is constructed against the naive forecasting scheme ($\overline{r}_t$). Statistical significance levels of R^2_{oos} at the 1%, 5%, 10% level denoted by ***, **, * are based on the MSE-adjusted statistic proposed by Clark and West (2007). CRSP data are considered.

Table 5: FORECASTING PERFORMANCE OF POPULAR PREDICTORS

	dfr	dfy	$infl$	d/e	lty	$svar$	e/p
R^2	0.0027	0.0001	0.0073	0.0015	0.0025	0.0000	0.0463
R^2_{oos}	−0.0446	−0.0112	−0.0422	−0.0367	−0.0665	−0.0065	0.0409**
	d/y	tms	tbl	ltr	b/m	$ntis$	$\widetilde{\eta}_t$
R^2	0.0356	0.0170	0.0000	0.0190	0.0538	0.0621	0.0641
R^2_{oos}	−0.1222	−0.0751	−0.0643	−0.0713	−0.0172	−0.0793	0.0121***

Notes: This table documents the forecasting performance of popular predictors for stock returns with S&P500 annual data from 1926 to 2013. While the first line documents the R^2 from the in-sample exercise, the second line provides the R^2_{oos} from the out-of-sample forecasting. Statistical significance levels of R^2_{oos} at the 1%, 5%, 10% level denoted by ***, **, * are based on the MSE-adjusted statistic proposed by Clark and West (2007). The predictors are the default return spread (dfr), the default yield spread (dfy), inflation (infl), the dividend payout ratio(d/e), the long term yield (lty), stock variance (svar), the earning price ratio (e/p), the dividend yield (d/y), the term spread (tms), the T-Bill rate (tbl), the long term return (ltr), the book to market ratio(b/m), and net equity expansion (ntis). For benchmarking purposes results for $\widetilde{\eta}_t$ are also documented in the Table.

Table 6: UNIT-ROOT TESTS

	ADF_t	PP_t	DF_{GLS}	MZ_t	$KPSS$
$\widetilde{\eta}_t$	−1.817	−1.837	−1.602	−1.543	0.168**
$\Delta\widetilde{\eta}_t$	−5.023***	−5.023***	−4.842***	−3.779***	0.178
cr^W_t	−2.458	−2.448	−1.542	−1.888	0.127*
Δcr^W_t	−5.720***	−5.725***	−2.691***	−3.862***	0.153
my_t	−5.007***	−40.829***	−5.011***	−40.353***	0.122*
Δmy_t	−1.547	−1.690	−1.425	−1.449	0.08
$\Delta^2 my_t$	−13.062***	−13.042***	−12.928***	−4.240***	0.097
tp_t	−2.184	−2.986	−1.438	−1.278	0.262***
Δtp_t	−13.966***	−13.966***	−12.236***	−3.179**	0.500**

Notes: This table displays results of unit root tests for gradually time-varying mean of the PD and its potential triggers. Test regressions for variables in levels include a constant and deterministic trend. Test regressions for variables in differences include a constant. The sample ranges from 1926 to 2013 and 1946 to 2008 in case of tp_t. Significance at 1%, 5%, 10% level is denoted by ***, **, *, respectively. Test results are qualitatively identical if only a constant is included for variables in levels or the sample is reduced to the period from 1946 to 2008 for all variables. For further notes see Table 1.

Table 7: COINTEGRATION ANALYSIS

W from cr^W_t	20	21	22	23	24	25	26	27	28	29	30
α	-0.07	-0.09	-0.09	**-0.12**	**-0.11**	**-0.11**	**-0.12**	**-0.12**	**-0.12**	**-0.11**	-0.11
	(-2.57)	(-3.28)	(-3.10)	(-3.91)	(-3.55)	(-3.72)	(-3.64)	(-3.70)	(-3.62)	(-3.49)	(-3.33)
β_1	**0.50**	**0.42**	**0.58**	**0.53**	**0.62**	**0.64**	**0.64**	**0.65**	**0.66**	**0.68**	**0.67**
	(2.35)	(2.50)	(3.29)	(4.03)	(4.20)	(4.56)	(4.51)	(4.64)	(4.49)	(4.30)	(3.96)
β_2	**-0.92**	**-1.09**	**-0.90**	**-1.01**	**-0.92**	**-0.93**	**-0.98**	**-1.02**	**-1.09**	**-1.02**	**-1.07**
	(-1.95)	(-2.70)	(-2.47)	(-3.49)	(-3.12)	(-3.42)	(-3.51)	(-3.69)	(-3.74)	(-3.33)	(-3.26)
β_3	**0.86**	**0.85**	**0.75**	**0.74**	**0.67**	**0.61**	**0.56**	**0.49**	0.42	0.46	0.43
	(2.37)	(2.99)	(2.51)	(3.26)	(2.61)	(2.49)	(2.20)	(1.90)	(1.53)	(1.52)	(1.30)
R^2	0.45	0.51	0.48	0.53	0.51	0.52	0.51	0.51	0.51	0.51	0.50
adj R^2	0.39	0.45	0.43	0.47	0.45	0.46	0.46	0.46	0.45	0.45	0.44

Notes: This table documents the estimates for the error correction model in (11) including consumption risk measures for distinct window sizes W. t-statistics appear in brackets below the corresponding estimates. Based on critical values from surface regressions provided by Ericsson and MacKinnon (2002), adjustment coefficients (α) that are significant at 10% level are highlighted. Also significant cointegration coefficients ($\beta_1, \beta_2, \beta_3$) at the 10% level are highlighted.

Table 8: CROSS VALIDATION

W from cr_t^W	23	24	25	26	27	28	29
			Panel A: CV from the full model				
	0.0220	0.0216	0.0216	0.0215	0.0217	0.0221	0.0226
			Panel B: Relative CV for the bivariate models				
With cr_t^W	1.0429	1.0458	1.0524	1.0675	1.0638	1.0557	0.9994
With my_t	1.0958	1.0946	1.0971	1.1022	1.0902	1.0717	1.0672
With tp_t	1.1346	1.1348	1.1374	1.1427	1.1303	1.1111	1.0434
			Panel C: Relative CV for the trivariate models				
Without cr_t^W	1.0928	1.1018	1.1043	1.1094	1.0974	1.0788	1.0749
Without my_t	1.0801	1.0932	1.1022	1.1155	1.1128	1.1028	1.0270
Without tp_t	0.9867	0.9967	0.9956	1.0011	0.9973	0.9866	0.9832

Notes: This table documents the cross validation (CV) statistics with distinct window sizes W for consumption risk measures. Panel A displays the CVs using the full model of $\widetilde{\eta}_t$ with all three determinants (cr_t^W, my_t, tp_t). Panel B shows the ratio of CVs from bivariate models of $\widetilde{\eta}_t$ with one determinant and the CVs from the corresponding full model in Panel A. These quotients are referred as relative CV. Similarly, Panel C shows the relative CV from trivariate models of $\widetilde{\eta}_t$ with two determinants using the CV from the corresponding full model in Panel A as the benchmark.

Table C1: EVALUATION OF APPROXIMATION ERRORS

Statistic	$\eta_t^{(2)}$	$\eta_t^{(1)}$	$\eta_t^{(2)} - \eta_t^{(1)}$
Mean	3.0271	2.9823	0.0448
Standard deviation	0.5933	0.6345	0.1084

Notes: This table documents MC simulation results based on 1000 replications. Column-$\eta_t^{(2)}$ is based on the exact formulation of the right-hand side of the present value model of the PD. Column-$\eta_t^{(1)}$ provides its approximation based on the three approximations C.1-C.3.

Figures

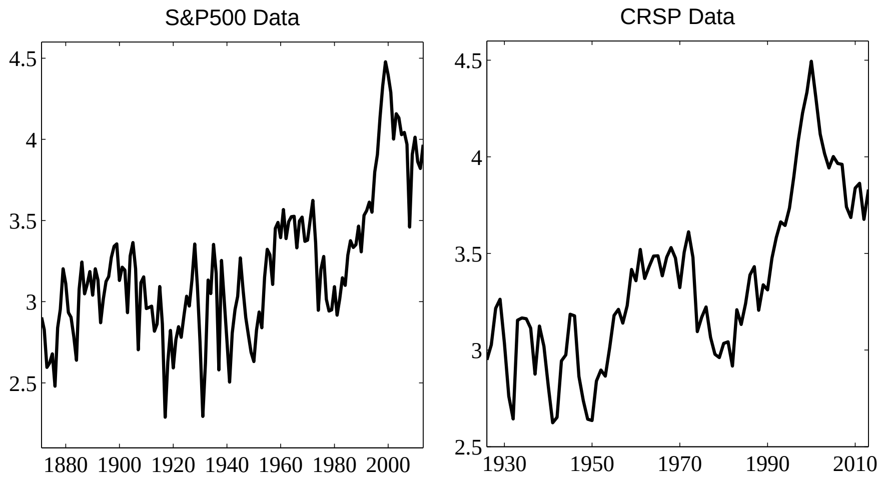

Figure 1: The Price-to-Dividend Ratio

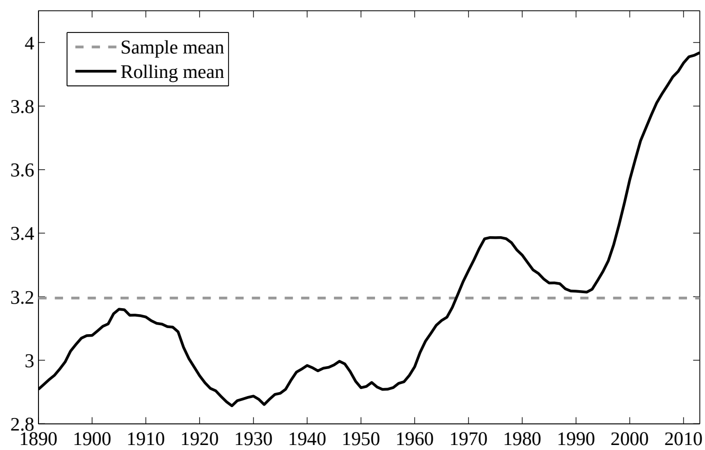

Figure 2: Rolling Mean of the PD

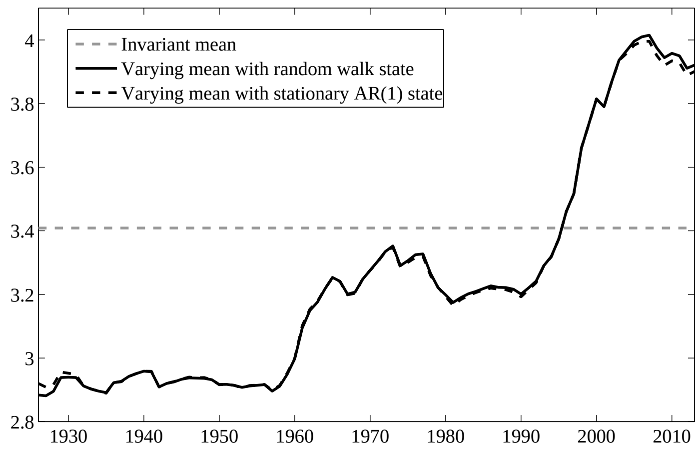

Figure 3: Estimated Means of the PD

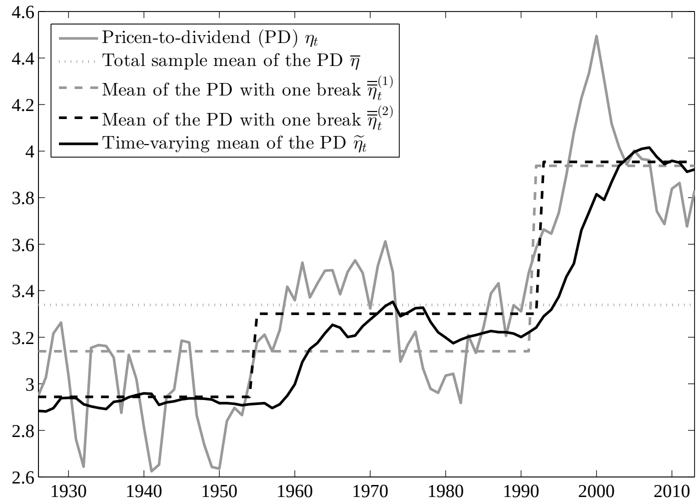

Figure 4: The PD and Distinct Mean Evaluations

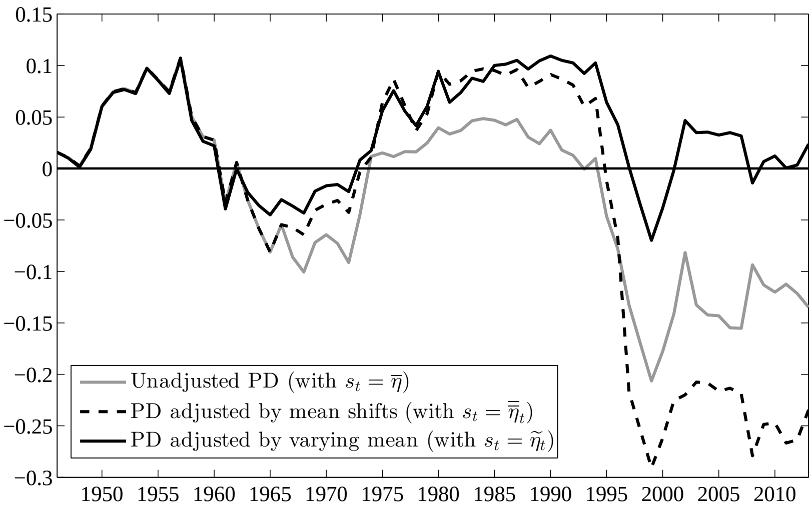

Figure 5: Out-of-Sample Forecasting Performance

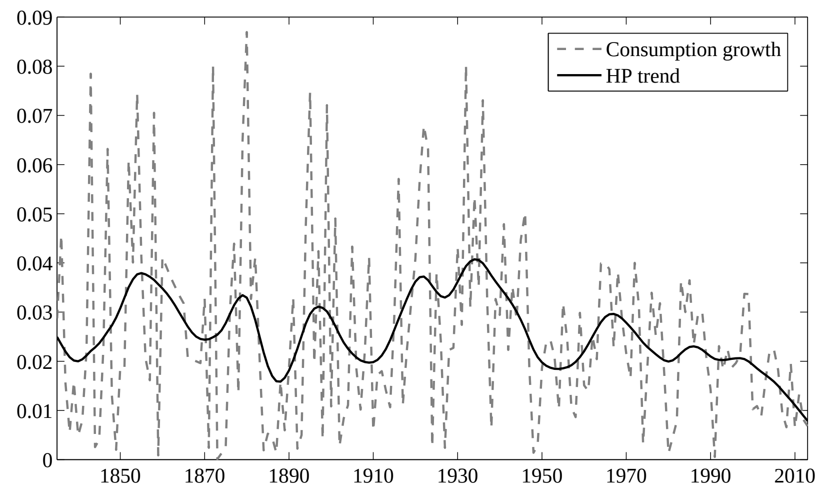

Figure 6: Absolute Growth Rate of per Capita Consumption

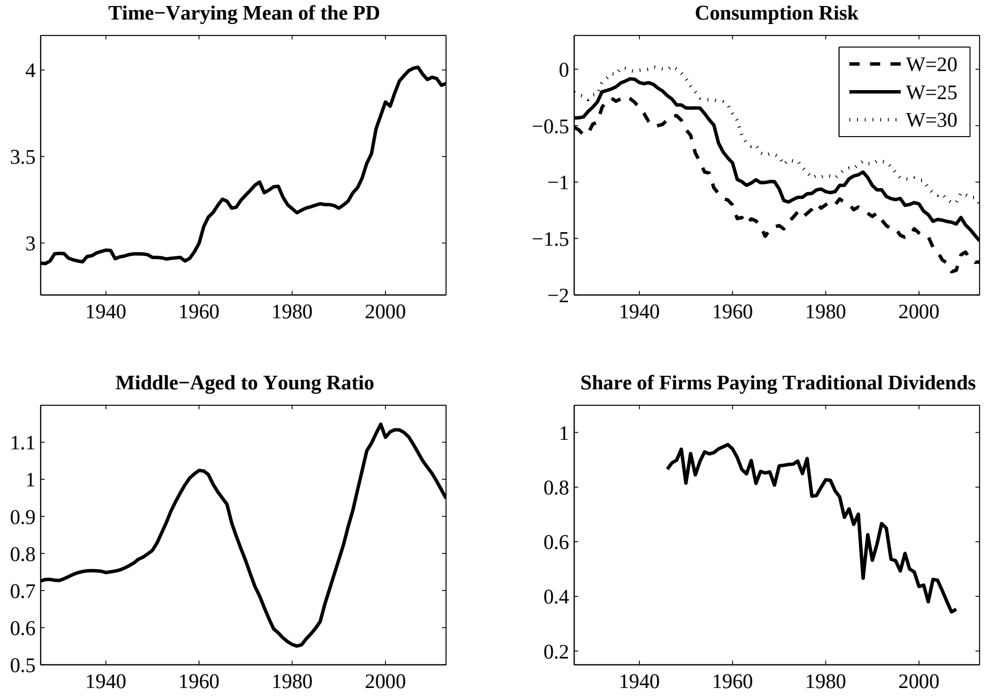

Figure 7: Economic Influences and the Gradually Time-Varying Mean of the PD

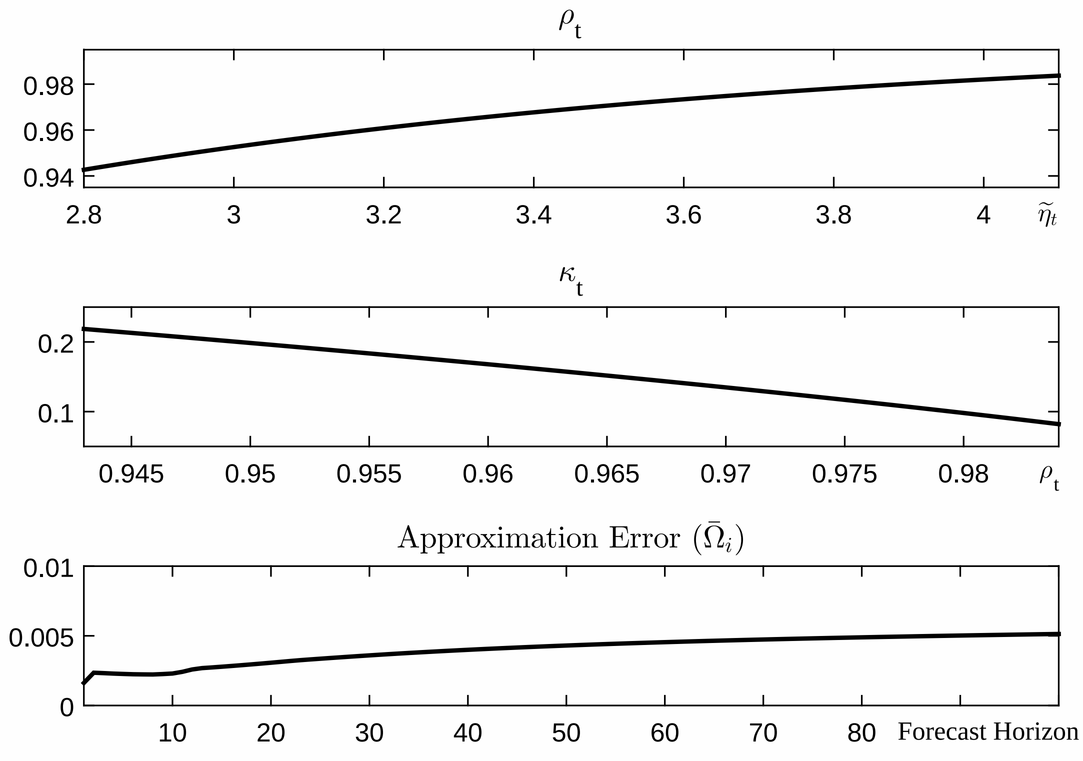

Figure C1: Evaluation of Approximations