A slowly evolving mean of the price-to-dividend ratio, its economic influences and predictive power for stock returns

Helmut Herwartz\textsuperscript{1}, Malte Rengel\textsuperscript{1,\textsuperscript{♭}} and Fang Xu\textsuperscript{2}

\textsuperscript{1}Department of Economics, Georg-August-University Göttingen
\textsuperscript{2}Department of Economics, University of Reading

October 1, 2014

Abstract

The potential forces behind persistent variations of the log price-to-dividend ratio (PtDR) and their implications for the prediction of stock returns have attracted a lively discussion in the literature. We estimate a gradually time-varying mean of the PtDR in the framework of a present value model, which is then used to adjust the PtDR in predictive regressions. In real time forecasting the proposed predictor outperforms the unadjusted PtDR and an adjustment of the PtDR by means of discrete shifts. We show that during the past 60 years this slowly evolving mean process is jointly shaped by the consumption risk, the demographic structure of the population and the proportion of firms with traditional dividend payout policy. 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.

\textsuperscript{*} Financial support from the German Research Foundation (Grant No. HE 2188/8-1) is gratefully acknowledged.
\textsuperscript{♭} Corresponding author. Email: mrengel@uni-goettingen.de
1 Introduction

In the late 1990s aggregate stock prices rose to unprecedented levels relative to any fundamental values. The logarithmic price-to-dividend ratio (PtDR), for example, has increased by 26% from 1980 to 2013. Even after a substantial decline since its peak in 2000, its level is still far away from its historical values. There is evidence for structural break(s) or instability in the mean of the PtDR (Lettau, Ludvigson and Wachter; 2008) and in the relation between the PtDR and future stock returns (Payea and Timmermann; 2006; Rapach and Wohar; 2006; Welch and Goyal; 2008). Empirical evidence indicates that the increasing mean of the PtDR could be due to a persistent fall in macroeconomic risk 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). Coping with the persistence of the PtDR, Lettau and Van Nieuwerburgh (2008) suggest a regime-switching model that allows discrete mean shifts. They show that deviations from shifting means of the PtDR carry predictive power for stock returns in-sample but fail to signal stock returns ex-ante compared with the historical average return as a benchmark predictor.

In this paper we consider a gradually time-varying mean of the PtDR that not only forecasts returns out-of-sample but also enables simultaneous testing of distinct determinants of the mean of the PtDR. We propose a latent variable reflecting the slowly evolving mean of the PtDR within a generalized version of the present value model introduced by Campbell and Shiller (1988). Particle filtering (e.g. Cappé, Godsill and Moulines; 2007) is employed to estimate this nonlinear state-space model. The gradually time-varying mean is entirely determined by the data and includes the (succession of) discrete mean shifts as a special case. The application of the state-space model in this paper differs from previous applications in forecasting returns that treat the expected return and expected dividend growth as two latent processes; see among others, Binsbergen and Koijen (2010) and Rytchkov (2012). We model the mean of the PtDR as a latent process, which can be interpreted as a combination of local means of expected returns and expected dividend growth. We do not assume an exogenous fixed mean of the PtDR as previous studies have. Instead, we use the present value model of the PtDR as an estimation equation rather than an identity restriction.1

We find that a gradually time-varying mean of the PtDR is strongly supported by log-likelihood diagnostics. The estimated long-term state has step-like patterns similar

---

1 In our framework, it is not straightforward to treat expected returns and expected dividend growth as latent state variables simultaneously.
to mean shifts with two structural breaks as suggested by Lettau and Van Nieuwerburgh (2008). In contrast to discrete mean shifts, however, the gradually time-varying mean evolves slowly over time. 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. We find that local deviations of the PtDR 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 PtDR adjustments by means of discrete mean shifts.

Extracting a gradually time-varying mean of the PtDR offers a unique opportunity to examine its potential influences and provides an economic interpretation of the out-of-sample predictive power of adjusting the PtDR by its slowly evolving mean. Following an error correction approach, we investigate the above mentioned three factors that have been documented to affect the PtDR in a long-run manner – consumption risk, the demographic structure, and the dividend payout policy of firms. We find that all three variables jointly shape the slowly evolving mean of the PtDR during the past 60 years, with consumption risk playing the most important role. A low consumption volatility risk drives down the 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 PtDR. 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 PtDR in the 1990s are consistent with increases in the middle-aged to young ratio during this same period. In addition to the macroeconomic and demographic influences, lowered dividends can affect the long-run relationship between stock price 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 PtDR. Nevertheless, among the three factors this has the smallest contribution in explaining the variations in the mean of the PtDR.

Section 2 illustrates the persistence of the PtDR, sketches its implications for the standard present value model, and introduces the state space model of the PtDR incorporating a gradually time-varying mean. The forecasting model, evaluation methods and forecasting performance are discussed in Section 3. In Section 4 we investigate the linkage between the gradually time-varying mean of the PtDR and its potential influences. Section 5 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).
2 A state space model of the PtDR

In this Section we first discuss the observed persistence of the PtDR and its implications for respective present value formulations. Then a latent gradually time-varying mean of the PtDR is formally derived and estimated, which is in line with the diagnosed stochastic trends governing the PtDR. Log-likelihood statistics support the view that the present value model of the PtDR incorporating a gradually time-varying mean outperforms the model with a constant mean.

2.1 Persistence of the PtDR

The persistent increase of stock prices relative to dividends from 1980 to 2000 can be seen from Figure 1. We find that the PtDR 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). Using annual CRSP data from 1926 to 2013 and S&P500 data from 1871 to 2013, Table 1 documents results from numerous unit root tests. The hypothesis of a non-stationary PtDR cannot be rejected with 5% significance by means of the ADF test and tests proposed by Phillips and Perron (1988) and Elliott, Rothenberg and Stock (1996).

Figure 1 about here

The PtDR 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 PtDR is rejected by means of the KPSS statistic (Kwiatkowski, Phillips, Schmidt and Shin; 1992). Moreover, we find that the PtDR can be better described by a non-stationary process than by a stationary process with a structural break at unknown timing. Testing the unit root hypothesis and allowing for a structural break in the mean of the PtDR as proposed by Perron and Vogelsang (1992) provides the relevant supporting evidence (see column 5 in Table 1, ‘PV’).

Table 1 about here

As noted by Campbell (2008), the persistence of the PtDR challenges the present value model in Campbell and Shiller (1988) that rests on the assumption of a stationary PtDR. Let \( P_t \) and \( D_t \) denote stock prices and the corresponding dividends in time \( t \), respectively.

---

2 It is worthwhile to mention that opposite to pure random walks diagnosed by common unit root tests, actual PtDR 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 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 PtDR, $\eta_t = \ln(P_t) - \ln(D_t)$,

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

where $d_t = \ln(D_t)$ and $\Delta$ 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 the linear approximation

$$r_{t+1} \approx \kappa - \eta_t + \rho \eta_{t+1} + \Delta d_{t+1}, \quad (2)$$

with

$$\rho \equiv 1/(1 + \exp(-\overline{\eta})) \text{ and } \kappa \equiv -\ln(\rho) - (1 - \rho) \ln(1/\rho - 1). \quad (3)$$

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 PtDR, $\overline{\eta}$ is unlikely to be constant and $\rho$ becomes also time-varying. Figure 2 illustrates the time variation of sample means of the PtDR from rolling time windows covering observations from the most recent 20 years.

Figure 2 about here

It is worthwhile to point out that a gradually time-varying mean of the PtDR is well in line with its diagnosed non-stationarity. Even when assuming that the PtDR is a stationary but near integrated process, a gradually time-varying mean could be regarded as a finite sample approximation of the local mean.

### 2.2 A state-space approximation

Taking a gradually time-varying mean of the PtDR into account, we modify the traditional present value model of the PtDR. Let $\tilde{\eta}_t$ denote the local mean employed to expand the Taylor approximation of the one-step-ahead stock returns in (2). Then, we obtain

$$r_{t+1} \approx \kappa_t - \eta_t + \rho_t \eta_{t+1} + \Delta d_{t+1}, \quad (4)$$

with both parameters ($\kappa_t$ and $\rho_t$) in (4) becoming time-specific, i.e.

$$\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t)) \text{ and } \kappa_t \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(1/\rho_t - 1). \quad (5)$$

To derive the present value formulation of the PtDR from (4), the following approximations similar to those in Lettau and Van Nieuwerburgh (2008) are adopted: $E_t[\rho_{t+i}] \approx \rho_t$, \ldots
$E_t[\kappa_{t+i}] \approx \kappa_t$ and $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_t]E_t[\eta_{t+i+1}]$. Simulation studies documented in Appendix C show that respective approximation errors are negligible for typical values of $\tilde{\eta}_t$. Taking the conditional expectation and iterating equation (4) forward provides the log-linear present value formulation of the PtDR

$$\eta_t \approx \frac{\kappa_t}{1 - \rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} E_t[\Delta d_{t+i}^e - r_{t+i}^e] + \lim_{i \to \infty} \rho_t^i E_t[\eta_{t+i}], \quad (6)$$

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 $r_{t+i}^f$. Changes in the long-term state of the PtDR affect the observed PtDR in a nonlinear fashion. A time-varying $\tilde{\eta}_t$ leads to a time-varying rather than a constant intercept term $\kappa_t/(1 - \rho_t)$, and the future return-adjusted dividend growth rates are discounted at time-varying rates $\rho_t$ rather than at a constant one. This casts a new light on the connection between stationary stock returns and dividend growth rates with the persistent PtDR. Under a constant $\eta_t$, the PtDR should be stationary if stock returns and dividend growth are stationary. However, under time-varying $\tilde{\eta}_t$, the PtDR could have a time trend through the intercept term. Allowing for a time-varying $\eta_t$ might explain PtDR dynamics which are not fully captured by time invariant valuation of return-adjusted cash flows.

An intuitive way to link equation (6) 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 PtDR conditional on past information. In this case, as shown in Lacerda and Santa-Clara (2010), the mean of the PtDR becomes time-varying and one can introduce directly a time index $t$ for the parameters $\rho$ and $\kappa$ in (3) in the traditional present value model to derive equation (6) (see also Figure 2). Against this background the proposed model offers a structural interpretation for this approach.

We employ a state space model to estimate the latent time-varying $\tilde{\eta}_t$. Assume a random disturbance term $\epsilon_t \sim N(0, \sigma^2_{\epsilon})$ 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 (6) by objective expectations conditional on the information set at the end of period $t$ ($\tilde{E}_t$), equation (6) is transformed into the measurement equation,

$$\eta_t = \frac{\kappa_t}{1 - \rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} \tilde{E}_t[\Delta d_{t+i}^e - r_{t+i}^e] + \epsilon_t, \quad (7)$$

where $\rho_t = 1/(1 + \exp(-\tilde{\eta}_t))$. The state equation formalizes a dynamic pattern for the latent process $\tilde{\eta}_t$, which is consistent with the diagnosed persistence of the PtDR,

$$\tilde{\eta}_t = \tilde{\eta}_{t-1} + u_t, \quad (8)$$
where \( u_t \sim N(0, \sigma_u^2) \), and the initialization \( \tilde{\eta}_0 \) is treated as a model parameter.

As a particular alternative state equation we consider a stationary first order autoregressive state process, i.e.

\[
\tilde{\eta}_t = \delta + \alpha \tilde{\eta}_{t-1} + u_t, \tag{9}
\]

where \(|\alpha| < 1\), \( u_t \sim N(0, \sigma_u^2) \) and \( \tilde{\eta}_0 = \delta/(1-\alpha) \).

A novel feature of the state space model in (7) coupled with (8) or (9) is that \( \tilde{\eta}_t \) can be estimated from filtered data. Compared with a framework of structural breaks, a continuously evolving steady state of the PtDR not only allows testing for its various determinants simultaneously, but might also enable a more successful out-of-sample forecasting of stock returns. In the framework of structural breaks subsample means are used to approximate the distinct steady states. Then, timing and magnitude of the breaks have to be estimated, which may lead to a weakened real time performance of break-adjusted forecasting schemes (Lettau and Van Nieuwerburgh; 2008). The state space model, however, provides a continuously evolving path of the mean of the PtDR. It can be easily extended into the future. In particular, it becomes unnecessary to locate break dates and magnitudes when using the PtDR to forecast stock returns out-of-sample. As it turns out, this distinctive characteristic is essential in improving the predictability of stock returns by means of the PtDR.

### 2.3 Model implementation

The state space model outlined in (7) and (8) is nonlinear. We use particle filtering (Cappé et al.; 2007) based on 3000 trajectories for an approximation of the models’ log-likelihood, subsequent parameter and state estimation. Following Lettau and Van Nieuwerburgh (2008) we focus mainly on the annual CRSP stock market index and corresponding dividends starting in 1926 to obtain the PtDR. S&P500 data dating back until 1871 are also analysed to address the robustness of the results. A detailed description of data sources and particle filtering can be found in Appendices A and B, respectively.

To formulate the objective expectations about future excess dividend growth rates and excess returns in (7), we follow Campbell and Vuolteenaho (2004) and employ low dimensional vector autoregressions (VARs) of order one comprised of the PtDR series \( (\eta_t) \), excess dividends growth rates \( (\Delta d_{t+1}) \), excess returns \( (r_{t+1}^e) \) and inflation \( \pi_t \).

The VAR based determination of \( (\bar{E}_t[\Delta d_{t+1}]) \) and \( (\bar{E}_t[r_{t+1}^e]) \) goes back to Campbell and Shiller (1988)
and Campbell (1991). Including the PtDR in the VAR provides unobservable market information about the future dividends and returns. The reduced form VAR is informative and at the same time general enough to be consistent with a present value relation with a gradually time-varying mean of the PtDR.

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 in VAR dynamics. Specifically, we evaluate in each forecast origin a set of VAR models \( \Omega_{t, \omega} = \{ \eta_\tau, \Delta d^e_\tau, r^e_\tau, \pi_\tau, \tau = t - \omega + 1, \ldots, t \} \) with alternative sample lengths \( \omega = 20, 21, \ldots, 30 \). From this set the particular VAR model with window size \( \omega \) is employed to determine ex-ante predictions \( \hat{\Delta d^e_{t+i}} \) and \( \hat{r^e_{t+i}} \) that minimize the root mean squared errors for the 10 most recent in-sample observations \( \{ \Delta d^e_m - r^e_m \}_{m=t-9} \).

The determination of the objective expectations in (7) requires an initialization period which is chosen to comprise 30 observations. Therefore, to evaluate the state space model in (7) 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.

2.4 Estimates and diagnostics

Turning to the evaluation of the state space model in (7) and (8) we find that the model with a gradually time-varying mean of PtDR outperforms its constant mean counterpart. The estimated model parameters are documented in Table 2, which also includes estimation results for the constant state benchmark model of Campbell and Shiller (1988). Considering the random walk (RW) as the state process (8), the log-likelihood value con-

---

[5] In related contexts, VAR based predictions have also been used to approximate price expectations, for instance by Sbordone (2002) and Rudd and Whelan (2006).

[6] 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 PtDR. However, a deterministic trend is at odds with stationary stock returns and dividend growth. Estimates for the latent process \( \tilde{\eta}_t \) display similar dynamics if a time trend is included. The estimated trend coefficient is quite small due to stationary returns and dividend growth. Although the path of forecasted return-adjusted dividend growth rate \( \Delta d^e_{t+i} - r^e_{t+i} \) diverges as \( i \to \infty \), the rate of this divergence is lower than the rate of convergence of \( \rho_i^{t-1} \) towards zero. Thus, with or without a trend in VAR doesn’t change the essence of the results. The 3-month Treasury Bill rate is employed to approximate the risk free rate and the CPI to measure inflation.

[7] To assess the robustness of outcomes we consider a set of robustness tests (i) using fitted errors regarding excess returns \( \{ r^e_m \}_{m=t-9} \) instead of \( \{ \Delta d^e_m - r^e_m \}_{m=t-9} \); (ii) using the five most recent observations in \( \{ \Delta d^e_m - r^e_m \}_{m=t-4} \) 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 identical.

At the implementation side, non-stationarity of the PtDR may cause explosive paths in forecasts of excess returns and excess dividends growth rates. For the VAR models selected by our benchmark adaptive scheme this turns out to be the case in 1917 and periods 1993-1996 and 1998-2001. For these years we restrict the autoregressive coefficient of the PtDR in the VAR and reestimate the system by means of EGLS until all eigenvalues of the characteristic polynomial are smaller than unity in modulus.
ditioned 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 (see the first row of Table 2). Similar log likelihood diagnostics favouring time variation in mean of the PtDR are obtained for S&P data, and if the static model is contrasted against a model with \( \tilde{\eta}_t \) specified as a stationary state process (9).

Table 2 about here

The lead of the more flexible model approach over the static benchmark present value model can be visualized by eyeballing the estimated patterns of \( \tilde{\eta}_t \) provided in Figure 3. The data speak against time invariance of the mean of the PtDR. In addition, estimating \( \tilde{\eta}_t \) 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). As it turns out both estimates of \( \tilde{\eta}_t \) lead to qualitatively identical results for the remaining empirical analysis. We concentrate on the estimates from a RW state equation henceforth, since it is marginally in the lead over a stationary AR (1) process (9) according to log likelihood diagnostics documented in Table 2.

Figure 3 about here

Obtaining the continuous movements of the mean of the PtDR enables a regression based economic interpretation of its determinants, which we provide in Section 4. Moreover, time variation in the mean of the PtDR is valuable for the ex-ante modeling of stock returns. In the next section, we analyse how \( \tilde{\eta}_t \) exploits the informational content of the PtDR in so-called predictive regressions.

3 Forecasting performance

We discuss predictive regression models for stock returns conditional on CRSP data, which is also used in Lettau and Van Nieuwerburgh (2008). Results for S&P data are qualitatively identical. Adjusting the PtDR 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 PtDR 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.
3.1 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),

\[ r_{t+1} = \beta_0 + \beta_1 (\eta_t - s_t) + v_{t+1}, \]

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 \( \Delta d_{t+1} \) for \( r_{t+1} \) in (10). To implement predictive regressions the PtDR \( (\eta_t) \) 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 \( \beta_1 = 0 \). The corresponding equation 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 by adjusting the PtDR by means of four alternative long-run states \( s_t \in \{ \bar{\eta}, \eta^{(1)}_t, \eta^{(2)}_t, \tilde{\eta}_t \} \). In the first specification the PtDR is centered by its (full sample) mean \( (\bar{\eta}) \). We refer to this setting as the ‘unadjusted’ PtDR since this model is equivalent to that of using the actual PtDR series in the predictive regressions. In the second and third specification, the PtDR is adjusted for one and two structural breaks \( (\eta^{(1)}_t \text{ and } \eta^{(2)}_t, \text{ respectively}) \). Following Lettau and Van Nieuwerburgh (2008) the supremum \( F \)-test (Bai and Perron; 1998, 2003) is employed to determine the timing of the breaks. 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 1993.8 Lastly, we adjust the PtDR by means of the gradually time-varying mean \( \tilde{\eta}_t \) which is extracted from the state space model in outlined in Section 2.

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 PtDR \( s_t \in \{ \bar{\eta}, \bar{\eta}_t, \tilde{\eta}_t \} \) – all of which are recursively estimated from the respective estimation samples. In the benchmark setting, the PtDR is centered with its mean \( \bar{\eta} \) from the estimation period. The second adjustment \( s_t = \bar{\eta}_t \) 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 \( \bar{\eta}_t \). Lastly, the PtDR is adjusted by \( \tilde{\eta}_t \) conditioning only on the information

8 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 \( (supF(1) = 18.12 \text{ and } supF(2) = 23.90) \). The null hypothesis of one break is rejected against the alternative of two breaks \( (supF(2|1) = 9.56) \) with 10% significance. We consider both one and two breaks to compare prediction outcomes with those in Lettau and Van Nieuwerburgh (2008). The applied test procedure is robust to serial correlation and heteroscedasticity, the trimming is 5% of the sample.
from the estimation periods.

The four alternative long run states of the PtDR entering the IS analysis, \( s_t \in \{\bar{\eta}_t, \eta_t^{(1)}, \eta_t^{(2)}, \tilde{\eta}_t\} \), are displayed in Figure 4 over the maximum sample period (1926 to 2013). The smoothly evolving mean \( \tilde{\eta}_t \) seems to be mostly close to the mean with two structural breaks \( \eta_t^{(2)} \). 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 random walk state equation (8) as an example, each particle is equal to \( \tilde{\eta}_{t-1}^{(i)} \) plus a draw from the error term with variance \( \sigma_u^2 \). Thus, being a (weighted) average of particles, \( \tilde{\eta}_t \) 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 PtDR in the OOS analysis. The core obstacle in using discrete break adjustments in OOS forecasting is determining the timing and magnitude of the breaks. The gradually time-varying mean \( \tilde{\eta}_t \) 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 \( \tilde{\eta}_t \) by means of weighted averaging, particles \( \tilde{\eta}_{t-1}^{(j)} \) are ranked according to the fit of the corresponding measurement equation for period \( t - 1 \). Particle \( \tilde{\eta}_{t-1}^{(j)} \) enters \( \tilde{\eta}_t \) with higher weight than particle \( \tilde{\eta}_{t-1}^{(k)} \) 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.

### 3.2 Forecast evaluation

Predicting stock returns in-sample the unadjusted PtDR provides a small \( R^2 \) of about 0.0392 (column 2 of the first panel in Table 3). Adjusting the PtDR 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 (\( \beta_1 \)) 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 PtDR by a slowly evolving mean does not outperform adjustments for discrete shifts in the mean. As an adaptive filtering process, \( \tilde{\eta}_t \) 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 PtDR.

Table 3 about here

Forecasting dividend growth, the unadjusted PtDR does not have much predictive power. 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

Adjusting the PtDR for discrete mean shifts in real time (OOS forecasting) fails to improve upon using historical average returns as benchmark predictors. In contrast, centering the PtDR 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 PtDR, centering with discrete mean shifts and centering around $\tilde{\eta}_t$ result 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). Results from mean absolute errors (MAEs) are qualitatively identical (not shown).

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

$$R^2_{\text{OOS}} = 1 - \frac{MSE_s}{MSE_{\bar{r}}},$$

(11)

where $MSE_{\bar{r}}$ denotes the mean squared forecast error from naive forecasts and $MSE_s$ is the corresponding statistic from alternative models (10) with $s_t \in \{\eta, \bar{\eta}, \tilde{\eta}\}$. Under 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 $R^2_{\text{OOS}} < 0$ ($R^2_{\text{OOS}} > 0$). Following Rapach, Strauss and Zhou (2010) the significance of $R^2_{\text{OOS}}$ is evaluated by means of the MSE-adjusted statistic in Clark and West (2006).\(^9\)

We find that in real time adjusting the PtDR by $\tilde{\eta}_t$ outperforms the historical average return in forecasting stock returns significantly. As can be seen from the second and

\[^9\] The MSE-adjusted statistic is based on

$$f_{\tau+1} = (r_{\tau+1} - \bar{r}_\tau)^2 - ((r_{\tau+1} - \tilde{r}_{\tau+1})^2 - (\bar{r}_\tau - \tilde{r}_{\tau+1})^2), \tau = \tau_0, \tau_0 + 1, \ldots, T - 1,$$

(12)
fourth row of Table 4, only adjusting the PtDR for the gradually time-varying mean ($\tilde{\eta}_t$) provides positive and significant $R^2_{\text{OOS}}$ statistics. We find the same evidence for the case of S&P500 data, forecasting excess returns instead of returns, or using different specifications to determine the window size of the VAR forecasting scheme for the state space model as already pointed out in footnote 7. The corresponding results are available upon request.

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.\textsuperscript{10} We find that using the gradually time-varying mean adjustment improves the forecasting strength of the PtDR throughout the entire sample period compared with the unadjusted PtDR or using break adjustments. The performance curve of adjusting the PtDR 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 ($s_t \in \{\eta, \tilde{\eta}_t, \tilde{\eta}_t\}$) start to underperform in comparison with the naive forecasts around 1957 and embark a negative trend. However, adjusting the PtDR by $\tilde{\eta}_t$ the respective performance curve falls least (black solid line). The performance curve of the unadjusted PtDR (grey solid line) falls a bit more than the one from the break adjusted PtDR (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 predictors outperform the naive forecast in periods with oil price shocks in 1973/1974. The good performance of the unadjusted PtDR during this period has also been noted by Welch and Goyal (2008). Centering the PtDR around its slowly evolving mean, however, obtains the only predictor that sustains this positive trend until 1994. The performance curves of both the unadjusted PtDR and the break adjusted PtDR 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 PtDR for its gradually time-varying mean dropping the least. The strongest performance deterioration (and weakest recovery since 1999) is observed when centering the PtDR around discrete shifts in mean.

where $\hat{\tau}_r$ denotes forecasts from the benchmark model with historical average returns, and $\hat{r}_{\tau+1}$ those from alternative models. Predicting dividend growth rates we apply the equivalent formula replacing $r_{\tau+1}$, $\hat{r}_\tau$ and $\hat{r}_{\tau+1}$ by the corresponding counterparts. For a regression of $f_{\tau+1}$ on a constant, Clark and West (2006) show that the corresponding $t$-statistic is asymptotically normally distributed even in the case of nested models. Under the alternative hypothesis the mean of $f_{\tau+1}$ (the constant coefficient) is greater than zero. It is worth noting that the alternative state processes are re-estimated at each period given available sample information. Thus, the varying mean processes of the PtDR change with each forecast origin.

\textsuperscript{10}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.
In summary, due to its adaptive potential in both turmoil and tranquil periods adjusting the PtDR by the slowly evolving mean offers superior ex-ante signaling in particular since the beginning of the new millennium.

Moreover, adjusting the PtDR by its gradually time-varying mean has a better performance than the naive forecasts based on historical mean returns throughout the sample period. This is visualized in Figure 6, which depicts OOS forecasting performance of various models through time using the PtDR adjusted for discrete shifts in mean as the benchmark. Although the break adjustment is not the best performing one (as discussed above), by using it as a benchmark one can easily compare the performance of historical mean forecasts with the one of the gradually time-varying mean adjusted PtDR. As can be seen from Figure 6, adjusting the PtDR by its gradually time-varying mean (black solid line) never had the dramatic falls as those from the historical mean (black dashed line). In the turbulent periods of the 1950s and the 1970s, the performance curve of the historical mean falls dramatically, since it failed to reflect the underlying changes in the economy. In early 1990s, the gradually time-varying mean adjusted PtDR adapts to structural changes quickly and outperforms the break adjustment while the historical mean lags behind.

In summary, we conclude that centering the PtDR around its slowly evolving mean provides a superior OOS forecasting performance compared with adjusting the PtDR for discrete shifts, and is also preferable to using historical means as a predictor.

4 Long-run determinants of the PtDR

Extracting a gradually time-varying mean of the PtDR provides a unique opportunity to examine its (joint) potential influences, and offers an economic interpretation for the out-of-sample predictive power of adjusting the PtDR by the gradually time-varying mean in predictive regressions. We investigate three important factors that have been documented to affect the PtDR 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 PtDR, and diagnose consumption volatility to be the most important influence. 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.
4.1 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. More recent studies such as those in Bansal and Yaron (2004) and Lettau et al. (2008) use recursive Epstein and Zin (1989) preferences, and demonstrate that economic agents dislike economic uncertainty, and 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 consumption volatility measures have good in-sample predictive power for the one-step ahead quarterly PtDR if historical volatilities 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. In this paper, 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 PtDR.

The consumption volatility is measured as $cr_t^W = \ln \left( \sum_{i=0}^{W-1} |c_{t-i}| \right)$, where $c_t$ denotes the centered annual growth rate of per capita consumption and $W$ is the size of rolling time windows. 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 7 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 (1984 to 1970), 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 boundary of 20 years is consistent with the so-called Kuznets swings in economic growth (e.g. Solomou; 2008).

---


12 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 $c_t$. The $p$-values of respective Ljung-Box statistics including 5 and 10 lags are 0.349 or 0.476, respectively.
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 expectations. A higher boundary of 30 years coincides with the estimated average duration of a regime of consumption volatility in Lettau et al. (2008).

Figure 7 about here

The upper right panel of Figure 8 depicts $c_{rt}^W$ 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 from the 1940s until the 1960s continuously. It remained relatively stable during the 1970s and 1980s and then decreased further from the 1990s till present. Comparing $c_{rt}^W$ with $\tilde{\eta}_t$ depicted in the upper left panel of Figure 8, it seems that consumption risk is negatively related to movements in the gradually time-varying mean of the PtDR throughout the entire sample for all considered window lengths $W$.

Figure 8 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 PtDR 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 PtDR 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 PtDR 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 8. 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 PtDR. The increases in $my_t$ in the 1950s and the 1980s correspond to the increases in $\tilde{\eta}_t$ in the 1960s and the 1990s, respectively.
Dividend payout policy  Dynamics of the PtDR can also be influenced by changes in the dividend payout policy by firms (see Fama and French; 2001; Robertson and Wright; 2006; Boudoukh, Michaely, Richardson and Roberts; 2007). The proportion of firms paying cash dividends fell from 66.5% in 1978 to 20.8% in 1999 (Allen and Michaely; 2003). Lowered dividends may result in persistent increases of the PtDR. Kim and Park (2013) show that the changing payout policy affects the long-run relationship between stock 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.

Decreasing cointegration coefficients linking log prices and log dividends are consistent with a non-stationary PtDR, which would be stationary if the cointegration coefficient were unity. 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 PtDR.

The firms with traditional payout policy are abbreviated as Type I firms following Kim and Park (2013). The lower right panel of Figure 8 depicts $t_{p_t}$, the proportion of Type I firms among all firms in the (CRSP) value-weighted market portfolio. Data on $t_{p_t}$ for the sample period from 1946 to 2008 has been kindly provided by C.J. Kim and C. Park. The proportion of firms with traditional payout policy decreases from 87% in 1946 to 35% in 2008. The falls in the proportion of Type I firms from 1980 are consistent with the increasing trend of the $\tilde{\eta}_t$ particularly since 1990s.

4.2 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 $\tilde{\eta}_t$, $c_{t}W$, $m_{yt}$ and $t_{p_t}$ are documented in Table 5. Almost all tests indicate first order integration of $\tilde{\eta}_t$, $c_{t}W$ and $t_{p_t}$ at conventional significance levels. The test regressions include a constant and a deterministic trend and refer to the periods from 1926 to 2013 in case of $\tilde{\eta}_t$, $m_{yt}$ and $c_{t}W$ and from 1946 to 2008 for $t_{p_t}$. The unit root hypothesis is rejected for all $c_{t}W$ measures with $W = 20, ..., 30$. Results for $c_{t}W$ with $W = 25$ are shown in Table 5 as an example. Although unit root tests hint at stationarity of $m_{yt}$, these results are to be taken with caution. Eyeballing $m_{yt}$ hardly supports regarding the process as mean stationary. 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 $m_{yt}$ can be found. We follow Favero et al. (2011) and treat $m_{yt}$ as a first order integrated process.

13 Test results are qualitatively identical if only a constant is included or the sample is reduced to the period from 1946 to 2008 for all variables.
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).\textsuperscript{14} With given presample values the SECM reads as

\begin{equation}
\Delta \tilde{\eta}_t = \delta_0 + \alpha (\tilde{\eta}_{t-1} + \beta_1 c_{t-1}^W + \beta_2 m_{t-1} + \beta_3 t_{t-1}) + \delta_1 \Delta c_{t}^W \\
+ \delta_2 \Delta t_{t} + \delta_3 \Delta m_{t} + \sum_{i=1}^{2} \phi_i \Delta \tilde{\eta}_{t-i} + \epsilon_t, \quad t = 1, 2, \ldots, T. \tag{13}
\end{equation}

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 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 $\delta_i$, $i = 1, 2, 3$, and $\psi_i$, $i = 1, 2$, with the lowest $t$-statistic and lacking 30% significance.\textsuperscript{15} 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 of variables below.

We find that the consumption risk, the demographic factor and the proportion of Type I firms are weakly exogenous and unaffected by their cointegration relation with $\tilde{\eta}_t$ for $W \in \{21, 23, \ldots, 30\}$. This supports the use of SECM in (13).\textsuperscript{16}

\textsuperscript{14} As a preliminary analysis of cointegration relations, we look at the possibility of bivariate cointegration relations between the gradually time-varying mean of the PtDR $\tilde{\eta}_t$ 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 PtDR jointly.

\textsuperscript{15} Using a liberal significance level for the removal of single variables from the model we believe that joint insignificance of the removed variables is likely for common (more conservative) significance levels, 5% say.

\textsuperscript{16} We test for weak exogeneity by adopting autoregressive models of order two augmented by the long-run relation between $\tilde{\eta}_{t-1}$, $c_{t-1}^W$, $m_{t-1}$ and $t_{t-1}$ as specified in (13), and looking at the significance of the corresponding adjustment coefficients. All respective adjustment coefficients are insignificant at the 10% level (not shown). Using $W = 25$ as an example, the respective $p$-values regarding the first differences of $c_{t}^W$, $t_{t}$ and $m_{t}$ are 0.315, 0.233 and 0.204.
We find evidence for a cointegration relation between $\tilde{\eta}_t$, $cr_t^W$, $my_t$ and $tp_t$, where $W \in 23, \ldots, 29$. 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 (13) with $W = 25$ as an example the $t$-statistic of the adjustment coefficient is -3.7217 while the 10% critical value is -3.4509. As can be seen from Table 6, $\hat{\alpha}$ estimates for $W \in 23, \ldots, 29$ are significant at 10% level. 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.

Significant effects from all three factors – consumption risk, the demographic structure and the dividend payout ratio – on the mean of the PtDR can be confirmed. Focusing on $W \in 23, \ldots, 29$, all estimated cointegration parameters are significant (an exception is $\hat{\beta}_3$ for $W = 28, 29$) and have the expected sign (see Table 6). Both consumption risk ($cr_t^W$) and the proportion of Type I firms ($tp_t$) have a negative influence on $\tilde{\eta}_t$ and, thus, the signs of $\hat{\beta}_1$ and $\hat{\beta}_3$ shall be positive. For the demographic factor ($my_t$), it is the opposite case and the sign for $\hat{\beta}_2$ shall be negative. 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 so-called cross-validation (CV) criteria (e.g. Picard and Cook; 1984). While augmenting (reducing) the set of explanatory variables in a regression 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 $\Delta \tilde{\eta}_t$, the left-hand side variable in the SECM (13). Specifically,

$$CV = \frac{1}{T} \sum_{t=1}^{T} \left| \Delta \tilde{\eta}_t - \hat{\Delta} \tilde{\eta}_t \right|, \quad (14)$$

where the forecast $\hat{\Delta} \tilde{\eta}_t$ for period $t$ is based on a model of $\Delta \tilde{\eta}_t$ that is estimated excluding the sample information (both dependent and explanatory variables) in period $t$. It is also referred as the leave-one-out estimator. In this sense, $\hat{\Delta} \tilde{\eta}_t$ is an out-of-sample forecast for

17 Johansen trace tests also confirm the results.
18 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 $W = 23, \ldots, 29$.
19 It is noteworthy that we obtain estimates with correct signs for all three cointegration parameters also for $W \in (3, 20)$ (not shown).
Thus, the better the fit of the model, the smaller the CV statistic.

To unravel the relative importance of each determinant \((cr_t^{W-1}, my_{t-1}, tp_{t-1})\), we consider three different sets of models to obtain \(\hat{\eta}_t\). The first is the SECM (13), which we refer to as the full model. The second set includes bivariate models of \(\hat{\eta}_t\) and one of the three determinants. And the third type includes trivariate models of \(\hat{\eta}_t\) including two of the three determinants. A particular determinant is regarded as more informative for the mean of the PtDR 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 7 about here

We find that among the three factors consumption risk is most informative for changes in the mean of the PtDR while changes in the payout policy of firms appear to be least informative. Table 7 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 \(cr_t^{W-1}\) in a bivariate model leads to markedly smaller loss than using \(my_{t-1}\) or \(tp_{t-1}\) for all different window sizes \(W\). Using \(W = 25\) as an example, the bivariate models with \(cr_t^{W-1}\), \(my_{t-1}\) 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, \(cr_t^{W-1}\) and \(my_{t-1}\) appear to be comparably informative for the changes of the mean of the PtDR. By removing \(cr_t^{W-1}\) 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.\(^{20}\)

5 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 gradually time-varying mean of the PtDR, and formalize a state space model to estimate its latent path. Log-likelihood statistics support the model. Adjusting the PtDR by its slowly evolving mean is fruitful in out-of-sample forecasting of stock returns. It outperforms

\(^{20}\)In an in-sample framework likelihood ratio (LR) statistics can assess the significance of distinct model fits based on residual variances. We adopt this framework to mean squared cross validation errors to underpin the strength of models’ CV statistics. We apply the 95% quantiles of a \(\chi^2\)-distribution with one degree of freedom as approximate critical values. 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 without \(tp_t\) are insignificant. Hence, this is further evidence that \(tp_t\) is less important than \(cr_t^{W}\) and \(my_t\) to shape \(\hat{\eta}_t\).
both adjusting the PtDR for structural mean shifts, and the historical average return as a common benchmark predictor.

We also provide evidence for economic influences on the mean of the PtDR. We find that consumption risk, the demographic structure of the population and firm’s dividend payout policy all play significant roles in shaping the slowly evolving mean of the PtDR. Among these determinants, 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 PtDR 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 PtDR 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. They contain the S&P500 index based on end-of-year closing prices and corresponding dividends for the period from 1871 to 2012. Annual dividends correspond to the sum of the four quarterly paid dividends within the corresponding year. For more details see Welch and Goyal (2008). The S&P500 index in 2013 is drawn from datastream (‘S&PCOMP’) and the corresponding dividend is provided by Robert J. Shiller and available from the internet.

**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 (CRSP). 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 2012 which is available from the internet. More details can be found in Welch and Goyal (2008). In 2013 we update the risk free rate by means of the 3-month US Treasury Bill rate in terms of the ‘secondary market’ quote, published by the Board of Governors of the Federal Reserve System and available from the internet.

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. For the period before 1913 Shiller calculated the series by means of data in Warren and Pearson (1935). 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.

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 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’). 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. To join the sum of non-durables and services specific consumption measures with the data of Barro and Ursua 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 \( c^{BEA}_t = 2.932 + 1.012 c^{BU}_t + \hat{nu}_t \) 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 (7) and (8) is highly nonlinear in the parameters, and the maximization of the corresponding log-likelihood function is not tractable analytically. Using Monte Carlo approximation techniques 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, Reeds, Wright and Wright (1998) for parameter estimation that does not depend on the gradient of the log-likelihood function. The particle filtering algorithm, specific for the state space model provided in Section 2, involves the following steps:

**Step (1): Initialization (t=1).** Sample \( N \) particles \( \tilde{\eta}_1^{(i)} \sim N(\tilde{\eta}_0, \sigma_u^2), i=1,...,N \) and determine importance weights and normalized weights, respectively, as

\[
    w_1^{(i)} = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2} \left( \frac{\xi_1^{(i)}}{\sigma} \right)^2 \right) \quad \text{and} \quad w_1^{(i)} = \frac{w_1^{(i)}}{\sum_{i=1}^N w_i^{(i)}}.
\]

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

1. Select \( N \) particles according to weights \( w_{t-1}^{(i)} \). Set accordingly \( \tilde{\eta}_{t-1}^{(i)} = \tilde{\eta}_1^{(i)} \) (resampling)

2. For all particles draw

\[
    \tilde{\eta}_t^{(i)} \sim N(\tilde{\eta}_{t-1}^{(i)}, \sigma_u^2), i = 1, ..., N,
\]

and determine raw and normalized weights, respectively, as

\[
    w_t^{(i)} = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2} \left( \frac{\xi_t^{(i)}}{\sigma} \right)^2 \right) \quad \text{and} \quad w_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_i^{(i)}}.
\]

3. go back to step ‘1’.

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

(i.) Approximation $E_t[\rho_{t+1}] \approx \rho_t$. Based on the empirical observations in Section 2 it becomes reasonable to assume $\eta_t$, and also $\bar{\eta}_t$, to follow a random walk. In principle, this martingale characteristic implies constant expectations of the gradually time-varying mean of the PtDR $\bar{\eta}_t$. To derive in (4) a function of returns $r_{t+1}$ which is linear in $\eta_t$ we apply a first order Taylor approximation based on $\rho_t \equiv 1/(1 + \exp(-\bar{\eta}_t))$. In consequence, $\rho_t$ is concave in $\bar{\eta}_t$ and therefore $E_t(\rho_{t+1}) \leq \rho_t$ by Jensen’s inequality. However, as displayed in the upper panel of Figure 9 the function $\rho_t \equiv 1/(1 + \exp(-\bar{\eta}_t))$ is approximately linear in the domain of $\bar{\eta}_t \in [2.8, 4.1]$.  

![Figure 9](image-url)

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

(ii.) Approximation $E_t[\kappa_{t+1}] \approx \kappa_t$. In the first order Taylor expansion $\kappa_t$ is determined as $\kappa_t \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(\rho_t - 1)$. Thus, $\kappa_t$ is also concave in $\rho_t$ and $E_t[\kappa_{t+1}] \leq \kappa_t$ by Jensen’s inequality. In the relevant domain of $\rho_t \in [0.943, 0.984]$ which is directly implied by $\bar{\eta}_t \in [2.8, 4.1]$ $\kappa_t = -\ln(\rho_t) - (1 - \rho_t) \ln(\rho_t - 1)$ is approximately linear as displayed in the middle panel of Figure 9. The maximal difference between $b\kappa_t^1(\rho_t) + (1-b)\kappa_t^2(\rho_t)$ and $\kappa_t^1(\rho_t) + (1-b)\kappa_t^2(\rho_t)$ 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 $E_t[\kappa_{t+1}] \approx \kappa_t$ is negligible.

(iii.) Approximation $E_t[\rho_{t+1}\eta_{t+1}] \approx E_t[\rho_{t+1}]E_t[\eta_{t+1}]$. To evaluate the magnitude of the error implied by this approximation we perform a simulation study. The parameter estimations $\bar{\eta}_0 = 2.892$ and $\sigma_u = 0.059$ from Table 2 are applied to simulate the process $\bar{\eta}_t = \bar{\eta}_{t-1} + u_t$ as a random walk, for $t = 1, \ldots, T$. To reflect the range of the empirical PtDR we bound the random walk by the minimum of the empirical PtDR (2.288) and the maximum of the empirical PtDR (4.495). The
process of $\rho_t = 1/(1 + \exp(-\eta_t))$ 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 $\eta_t$ we add a first order moving average process, obtaining

$$\eta_t = \eta_{\tilde{t}} + \alpha \omega_{t-1} + \omega_t. \quad (15)$$

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

To approximate $E_t[\rho_t, \eta_{t+i+1}]$ 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 $E_t[\rho_t+i\eta_{t+i+1}]$ 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 $t$ as $\hat{\psi}_{r,t+i}$, $\hat{\rho}_{r,t+i}$ and $\hat{\eta}_{r,t+i+1}$, for $r = 1, \ldots, R$ and $i = 1, \ldots, H$. To determine the approximation error of interest we use the following statistic

$$\tilde{\Omega}_i = \frac{1}{R(T - T_1 - T_2)} \sum_{r=1}^{R} \sum_{t=T_2+1}^{T} \Omega_{r,t+i}, \quad \text{for } i = 1, 2, \ldots, H, \quad (16)$$

where

$$\Omega_{r,t+i} = \left| \frac{df_{r,t+i}}{\psi_{r,t+i}} \right|, \quad (17)$$

with

$$|df_{r,t+i}| = |\hat{\psi}_{r,t+i} - \hat{\rho}_{r,t+i}\hat{\eta}_{r,t+i+1}|. \quad (18)$$

We set $R = 1000$, $T = 2000$, $T_1 = 500$, $T_2 = 500$ and $H = 100$. The lower panel of Figure 9 displays $\tilde{\Omega}_i$, for $i = 1, \ldots, 100$. With increasing forecasting horizons the average approximation error converges. It reaches not more than 1% for the 100-step ahead forecasting. As a result, the magnitude of this error is rather small.


Tables

<table>
<thead>
<tr>
<th>ADF</th>
<th>PP</th>
<th>DFGLS</th>
<th>PV</th>
<th>KPSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Statistics</td>
<td>CRSP 1926-2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.595</td>
<td>-1.604</td>
<td>-1.171</td>
<td>-3.553</td>
<td>0.801</td>
</tr>
<tr>
<td>Crit. val. at 5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.895</td>
<td>-2.895</td>
<td>-1.945</td>
<td>-4.580</td>
<td>0.463</td>
</tr>
<tr>
<td>Crit. val. at 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.585</td>
<td>-2.585</td>
<td>-1.614</td>
<td>-4.240</td>
<td>0.347</td>
</tr>
<tr>
<td>S&amp;P500 1871-2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.783</td>
<td>-1.841</td>
<td>-1.139</td>
<td>-3.737</td>
<td>0.906</td>
</tr>
<tr>
<td>Crit. val. at 5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.882</td>
<td>-2.882</td>
<td>-1.943</td>
<td>-4.580</td>
<td>0.463</td>
</tr>
<tr>
<td>Crit. val. at 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.578</td>
<td>-2.578</td>
<td>-1.615</td>
<td>-4.270</td>
<td>0.347</td>
</tr>
</tbody>
</table>

Table 1: Unit-root tests for the PtDR. 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 AR estimator is used to calculate the long-run variance. DFGLS refers to the test proposed by Elliott et al. (1996) where the BIC is applied to determine the lag length. 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>
<thead>
<tr>
<th>State equation</th>
<th>Time-varying $\bar{\eta}$</th>
<th>constant $\bar{\eta}$ ($\sigma_u = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRSP 1926-2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td>-</td>
<td>2.892</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.042</td>
<td>0.986</td>
</tr>
<tr>
<td>S&amp;P500 1901-2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.046</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimates and model evaluations. In each panel the upper row documents statistics from the model with a random walk (RW) as state equation (8) and the lower row statistics from the model applying a stationary AR(1) process as the state equation (9). Out-of-sample VAR forecasts are applied to evaluate $\hat{E}_t[\Delta d_{t+1}]$ and $\hat{E}_t[r_{t+1}]$ in (7).
Table 3: In-sample predictive regressions

This table documents statistics from regressions (10) with four alternative mean processes of the PtDR using CRSP data from 1926 to 2013. In the second column forecasts for returns (dividend growth) are conditioned on the unadjusted PtDR (using the overall sample mean $\overline{\eta}$). The third and fourth column contain the estimates based on the adjusted PtDR using one ($\overline{\eta}^{(1)}$) or two ($\overline{\eta}^{(2)}$) mean shifts, respectively. In the last column forecasts are conditioned on the PtDR adjusted by means of the smooth state process $\tilde{\eta}_t$. Newey and West (1987) robust $t$-statistics for coefficient estimates are presented in parentheses. The bandwidth is selected by means of the procedure proposed by Newey and West (1994).

<table>
<thead>
<tr>
<th></th>
<th>$\overline{\eta}$</th>
<th>$\overline{\eta}^{(1)}$</th>
<th>$\overline{\eta}^{(2)}$</th>
<th>$\tilde{\eta}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock Returns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.0928</td>
<td>0.0937</td>
<td>0.0939</td>
<td>0.1142</td>
</tr>
<tr>
<td></td>
<td>(4.5682)</td>
<td>(4.5205)</td>
<td>(6.0840)</td>
<td>(5.3922)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.0923</td>
<td>-0.2516</td>
<td>-0.4120</td>
<td>-0.2089</td>
</tr>
<tr>
<td></td>
<td>(-2.5610)</td>
<td>(-4.8718)</td>
<td>(-6.0994)</td>
<td>(-2.7885)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0392</td>
<td>0.1027</td>
<td>0.1751</td>
<td>0.0641</td>
</tr>
<tr>
<td>adj $R^2$</td>
<td>0.0279</td>
<td>0.0922</td>
<td>0.1654</td>
<td>0.0531</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.0456</td>
<td>0.0457</td>
<td>0.0458</td>
<td>0.0539</td>
</tr>
<tr>
<td></td>
<td>(3.3098)</td>
<td>(3.5017)</td>
<td>(3.3140)</td>
<td>(4.0314)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.0013</td>
<td>-0.0431</td>
<td>-0.1157</td>
<td>-0.0832</td>
</tr>
<tr>
<td></td>
<td>(-0.0426)</td>
<td>(-0.9694)</td>
<td>(-1.7106)</td>
<td>(-1.8706)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0000</td>
<td>0.0057</td>
<td>0.0260</td>
<td>0.0192</td>
</tr>
<tr>
<td>adj $R^2$</td>
<td>-0.0117</td>
<td>-0.0060</td>
<td>0.0146</td>
<td>0.0076</td>
</tr>
</tbody>
</table>

Table 4: Out-of-sample predictive regressions for stock returns

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 PtDR ($s_t = \overline{\eta}$), the PtDR adjusted by mean shifts ($s_t = \overline{\eta}^{(1)}$), and the PtDR adjusted by the smooth state ($s_t = \tilde{\eta}_t$). 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 (2006). CRSP data are considered.

<table>
<thead>
<tr>
<th></th>
<th>$\overline{r}$</th>
<th>$\overline{\eta}$</th>
<th>$\overline{\eta}^{(1)}$</th>
<th>$\tilde{\eta}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1946-2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1694</td>
<td>0.1751</td>
<td>0.1793</td>
<td>0.1683</td>
</tr>
<tr>
<td>$R^2_{oos}$</td>
<td>-0.0691</td>
<td>-0.1202</td>
<td>0.0121***</td>
<td></td>
</tr>
<tr>
<td>1946-2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1607</td>
<td>0.1681</td>
<td>0.1714</td>
<td>0.1589</td>
</tr>
<tr>
<td>$R^2_{oos}$</td>
<td>-0.0934</td>
<td>-0.1367</td>
<td>0.0232***</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Unit-root tests

This table displays results of unit root tests for gradually time-varying mean of the PtDR and its potential triggers. Test regressions include a constant and deterministic trend. 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. For further notes see Table 1.

<table>
<thead>
<tr>
<th>( \tilde{\eta}_t )</th>
<th>( \Delta \tilde{\eta}_t )</th>
<th>( cr^W_t )</th>
<th>( \Delta cr^W_t )</th>
<th>( my_t )</th>
<th>( \Delta my_t )</th>
<th>( \Delta^2 my_t )</th>
<th>( tp_t )</th>
<th>( \Delta tp_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.602</td>
<td>-1.542</td>
<td>-2.691***</td>
<td>-5.011***</td>
<td>-1.425</td>
<td>-12.928***</td>
<td>-1.438</td>
<td>-12.236***</td>
</tr>
<tr>
<td></td>
<td>0.168**</td>
<td>0.178</td>
<td>0.153</td>
<td>0.122*</td>
<td>0.08</td>
<td>0.097</td>
<td>0.262***</td>
<td>0.500**</td>
</tr>
</tbody>
</table>

Table 6: Cointegration Analysis

This table documents the estimates for the error correction model in (13) 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 (\( \alpha \)) 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>
<thead>
<tr>
<th>( W ) from ( cr^W_t )</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-0.07</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(-2.57)</td>
<td>(-3.28)</td>
<td>(-3.10)</td>
<td>(-3.91)</td>
<td>(-3.72)</td>
<td>(-3.64)</td>
<td>(-3.70)</td>
<td>(-3.62)</td>
<td>(-3.49)</td>
<td>(-3.33)</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.50</td>
<td>0.42</td>
<td>0.58</td>
<td>0.53</td>
<td>0.62</td>
<td>0.64</td>
<td>0.64</td>
<td>0.65</td>
<td>0.66</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>(2.35)</td>
<td>(2.50)</td>
<td>(3.29)</td>
<td>(4.05)</td>
<td>(4.20)</td>
<td>(4.56)</td>
<td>(4.51)</td>
<td>(4.64)</td>
<td>(4.49)</td>
<td>(4.30)</td>
<td>(3.96)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.92</td>
<td>-1.09</td>
<td>-0.90</td>
<td>-1.01</td>
<td>-0.92</td>
<td>-0.93</td>
<td>-0.98</td>
<td>-1.02</td>
<td>-1.09</td>
<td>-1.02</td>
<td>-1.07</td>
</tr>
<tr>
<td></td>
<td>(-1.93)</td>
<td>(-2.70)</td>
<td>(-2.47)</td>
<td>(-3.49)</td>
<td>(-3.12)</td>
<td>(-3.42)</td>
<td>(-3.51)</td>
<td>(-3.69)</td>
<td>(-3.74)</td>
<td>(-3.33)</td>
<td>(-3.26)</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.86</td>
<td>0.85</td>
<td>0.75</td>
<td>0.74</td>
<td>0.67</td>
<td>0.61</td>
<td>0.56</td>
<td>0.49</td>
<td>0.42</td>
<td>0.46</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(2.37)</td>
<td>(2.99)</td>
<td>(2.51)</td>
<td>(3.26)</td>
<td>(2.61)</td>
<td>(2.49)</td>
<td>(2.20)</td>
<td>(1.90)</td>
<td>(1.53)</td>
<td>(1.52)</td>
<td>(1.30)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.45</td>
<td>0.51</td>
<td>0.48</td>
<td>0.53</td>
<td>0.51</td>
<td>0.52</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>( \text{adj } R^2 )</td>
<td>0.39</td>
<td>0.45</td>
<td>0.43</td>
<td>0.47</td>
<td>0.45</td>
<td>0.46</td>
<td>0.46</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Table 7: Cross Validation 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 $\tilde{\eta}_t$ with all three determinants ($cr^W_t$, $my_t$, $tp_t$). Panel B shows the ratio of CVs from bivariate models of $\tilde{\eta}_t$ 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 $\tilde{\eta}_t$ with two determinants using the CV from the corresponding full model in Panel A as the benchmark.

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

Figure 1: The price-to-dividend ratio The PtDR from S&P500 data from 1871 to 2013 is depicted in the left hand side panel. The right hand side panel illustrates the PtDR from CRSP data from 1926 to 2013.

Figure 2: Rolling mean of the PtDR The Figure illustrates the rolling window sample mean of the price-to-dividend ratio ($\eta_t$) from CRSP data. The estimation windows includes the most recent 20 years of observations.
Figure 3: Gradually time-varying mean of the PtDR The figure depicts the estimated gradually time-varying mean of the PtDR $\tilde{\eta}_t$ (black solid line for the random walk and black dashed line for the stationary AR(1) state process) along with its time invariant counterpart (grey line).

Figure 4: The PtDR and distinct mean evaluations The figure presents the price-to-dividend ratio $\eta_t$ (grey solid line), the total sample average PtDR $\overline{\eta}$ (grey dotted line), the mean of the PtDR with one break $\overline{\eta}_t^{(1)}$ (grey dashed line), the mean of the PtDR with two breaks $\overline{\eta}_t^{(2)}$ (black dashed line), and the gradually time-varying mean of the PtDR $\tilde{\eta}_t$ (black solid line).
Figure 5: Out-of-sample forecasting performance This figure depicts the difference of the cumulative squared forecast errors of the naive prediction by means of historical average returns ($\bar{r}_t$) minus those of the three alternative predictive regressions. The considered predictors are the unadjusted PtDR (grey solid line with $s_t = \eta$), the PtDR adjusted by allowing for potential mean shifts (black dashed line with $s_t = \overline{\eta}$), and the PtDR adjusted by means of the smooth state (black solid line with $s_t = \tilde{\eta}$). The first forecasting period is 1946.

Figure 6: Out-of-sample forecasting performance This figure depicts the difference of the cumulative squared forecast errors of the prediction by means of the break adjusted PtDR ($s_t = \overline{\eta}$) minus those of the three alternative predictive regressions. The considered predictors are the historical mean (black dashed line with $\overline{\eta}$), the unadjusted PtDR (grey solid line with $s_t = \eta$) and the PtDR adjusted by means of the smooth state (black solid line with $s_t = \tilde{\eta}$). The first forecasting period is 1946.
Figure 7: Consumption growth This figure plots the absolute growth rate of per capita consumption with its HP trend (smoothing parameter 100).

Figure 8: Economic influences and the gradually time-varying mean of the PtDR The upper panels display the time-varying mean process of the PtDR and three distinct consumption risk measures (dashed from $W = 20$, solid from $W = 25$, dots from $W = 30$). The lower panels display the middle-aged to young ratio and the proportion of Type I firms.
Figure 9: Evaluation of approximations The upper panel displays $\rho_t = 1/(1 + \exp(-\eta_t))$ for $\eta_t \in [2.8, 4.1]$ and the middle panel $\kappa(\rho_t) \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(1/\rho_t - 1)$ for $\rho_t \in [0.943, 0.984]$. The lower panel indicates the approximation error ($\bar{\Omega}_i$) from $E_t[\rho_{t+i} \eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ for alternative forecast horizons $i = 1, \ldots, 100$. 