Market Expectations in the Cross Section of Present Values *

Bryan Kelly†
University of Chicago
Booth School of Business

Seth Pruitt‡
Federal Reserve
Board of Governors

This Version: January 2011

Abstract

Returns and dividend growth for the aggregate US stock market are highly and robustly predictable. Using information extracted from the cross section of price-dividend ratios, we find an in-sample and out-of-sample forecasting R^2 as high as 30% for both series at the annual frequency. We present a general economic framework linking aggregate market expectations to disaggregated valuation ratios in a dynamic latent factor system. To derive our forecasts we use a new regression-based filter to extract factors driving aggregate expected returns and dividend growth from the cross section of price-dividend ratios. Our findings shed new light on the dynamic processes of market discount rates and growth expectations.

*We thank Amit Goyal for sharing data with us, Gray Calhoun, John Cochrane, George Constantinides, Gene Fama, Rob Engle, Xavier Gabaix, Jim Hamilton, Ralph Koijen, Juhani Linnainmaa, Hanno Lustig, Toby Moskowitz, Lubos Pastor, Stijn Van Nieuwerburgh and seminar participants at Barclays Capital, Chicago (Booth), Duke (Economics), the Federal Reserve Board, Michigan (Ross), Northwestern (Kellogg), UCSD (Rady), UNC (Kenan-Flagler) and Yale (SOM) for helpful comments.

†Mail: 5807 S. Woodlawn Ave., Chicago, IL 60637. E-mail: bryan.kelly@chicagobooth.edu. Phone: (773) 702-8359.

‡Mail: Board of Governors of the Federal Reserve System, Washington, DC 20551. E-mail: seth.j.pruitt@frb.gov. Phone: (202) 452-2344. This paper reflects the views and opinions of the authors and does not reflect on the views or opinions of the Federal Reserve Board, the Federal Reserve System, or their respective staffs.

Explaining market price behavior of the US capital stock is among the most fundamental challenges facing economists. The present value relationship between prices, discount rates and future cash flows has proved a valuable lens for understanding price changes. It reveals that price changes are wholly driven by fluctuations in investors' expectations of future returns and cash flow growth. Understanding asset prices amounts to understanding the dynamic behavior of these expectations.

The most common approach to measuring aggregate return and cash flow expectations is predictive regression. As suggested by the present value relation, research has found the aggregate price-dividend ratio to be among the most informative predictive variables for describing investor expectations. Typical estimates find that about 10% of annual return variation and 1% of annual dividend growth variation can be explained by forecasts based on the aggregate price-dividend ratio.[1] In this paper we find that reliance on aggregate quantities drastically understates the degree of predictability in both returns and cash flow growth, and hence understates the volatility of investor expectations. Our estimates suggest that as much as 30% of variation in both annual market returns and dividend growth can be explained by expanding the conditioning information set to include disaggregated market data.

To harness information contained in individual asset prices, we represent the cross section of asset-specific price-dividend ratios as a dynamic latent factor model. We relate these disaggregated present values to aggregate expected market returns and dividend growth. Our model highlights the idea that the same dynamic state variables driving aggregate expectations also govern the dynamics of the entire panel of asset-specific price-dividend ratios. This restriction allows us to exploit rich cross-sectional information to extract estimates of the market's expectations for aggregate returns and dividend growth period-by-period.

Our approach attacks a vexing problem in empirical asset pricing: How to exploit a wealth

[1] See Cochrane (2005) and Koijen and Van Nieuwerburgh (2010) for surveys of return and cash flow predictability evidence.

of predictors in relatively short time series?[2] We propose a solution using the methodology of Kelly and Pruitt (2011). The estimator, called the three-pass regression filter (3PRF), is a simple regression-based procedure designed to forecast a single time series using a large panel of predictors in a parsimonious way. The 3PRF constructs forecasts for market returns (or dividend growth) as a linear combination of disaggregated price-dividend ratios with weights based on their covariance with the variable being forecasted. It also identifies the forecast contribution of each asset.

Our procedure identifies and aggregates information about expected returns and growth rates distributed throughout cross section of price-dividend ratios. This information, hitherto unused in the present value context, gives rise to several surprising conclusions. First, we document a high and robust degree of predictability that is unprecedented in the literature. Applying our methodology to a cross section of portfolios formed on the basis of firm size and book-to-market ratio, we find a predictive R^2 for both annual market returns and dividend growth that consistently ranges between 20% and 30%. This is true over both the commonly studied post-war sample (1946-2009) as well as the sample including the pre-war era (1930-2009). In contrast to many alternative predictors, predictability of the same high magnitude is also found purely out-of-sample, especially for market returns. Predictability is strongest when we extract three factors from the cross section, and in some cases we find high levels of predictability when even a single factor is extracted. We show that limiting portfolios to only include stocks that previously paid dividends greatly enhances evidence of predictability. While perhaps this is a natural restriction to impose in a framework built upon price-dividend ratios (which require non-zero dividends to avoid degeneracy), the important feature of dividend paying firms is stability in their market and cash flow behavior. This stability is a

[2] In his AFA presidential address, Cochrane (2011) highlights the importance of this question:

> We have a bunch of univariate regressions - one return on the left, and one variable on the right... First, we have to move past treating extra variables one or two at a time, and understand which of these variables are really important. Alas, huge multiple regression is impossible. So the challenge is, how to answer the great multiple-regression question, without actually running huge multiple regressions?

key ingredient to the robust predictive relationship that we identify.

Second, we find that the best linear combination of price-dividend ratios for forecasting returns has two features. First, it uses a size spread by going long small stocks and short big stocks. Second, it goes long stocks with moderate book-to-market ratios (from the 20^{th} to 80^{th} percentile) and short extreme growth and extreme value stocks. We argue that this hump shape in portfolio weights is driven by the high (low) duration of value (growth) stocks, and is an equity analogue of the bond return forecasts of Cochrane and Piazzesi (2005). The best combinations for forecasting dividend growth are straight size and value spreads that go long small value stocks and short large growth stocks.

Third, our estimates shed new light on the dynamic processes for expected one-year-ahead returns and dividend growth rates. Conditional on the cross section of past price-dividend ratios, the volatility of expected one-year returns is 9.3%, two-thirds higher than the 5.6% volatility found by conditioning on the aggregate price-dividend ratio alone and more than twice as high as estimates from the aggregate present value state space. We also find less persistence in expected returns, with estimated autocorrelation of 61.8%, compared with over 90% based on estimates from aggregate data. The evidence for expected market dividend growth is similar. We find one-year-ahead expected growth rate volatility of 4.4% and autocorrelation of 56.9%. This is compared to annual volatility of only 0.77% and autocorrelation of 92.5% when conditioning on the aggregate price-dividend ratio. Our estimates for expected growth rates build upon van Binsbergen and Koijen's (2010) findings of high dividend growth predictability and volatile growth expectations. We find weak contemporaneous correlations between one-year expected market returns and dividend growth, but find stronger evidence in favor of serial cross-correlations between the two series. This dynamic structure among short horizon expectations poses an interesting challenge to fundamentals-based models of discount rates.

Fourth, our findings suggest that there is rich "horizon structure" in one-year expected returns. Recent research has modeled one-year expectations with first-order autoregressions.[3] In the present value context, this implies that 100% of variation in the aggregate price-dividend ratio is driven by next year's expected return and dividend growth rate. Our estimates reject this single-horizon structure. We find that short horizon expectations explain less than 20% of aggregate price-dividend ratio variation and are primarily responsible for its high frequency fluctuations. We then show that incorporating one-year expectations at more distant horizons helps describe low frequency, persistent movements in the price-dividend ratio. This point is related to another surprising conclusion from our analysis. Our method predicted negative one-year-ahead expected returns four times after World War II, each corresponding to NBER recessions and severe negative realized returns. Two of these dates, the 2001 tech crash and the 2008 financial crisis, occurred during our out-of-sample period, providing an interesting test of whether negative expected returns are a vestige of in-sample overfit. In both 2001 and 2008 (and *only* in those instances) cross-sectional information indeed forecasted negative returns out-of-sample. This procyclicality in short term expected returns is difficult to reconcile with standard structural asset pricing models.

Why do disaggregated prices produce such accurate forecasts? To illustrate the advantages of cross section information, consider a simple CAPM example.[4] In particular, one period expected market returns μ_t and dividend growth g_t are the two common factors in the economy, and the price-dividend ratio of any asset i is

$$pd_{i,t} = a_i - b_{i,\mu}\mu_t + b_{i,g}g_t + e_{i,t} \tag{1}$$

[3]See, for example, Cochrane (2008a), Pástor and Stambaugh (2009) and van Binsbergen and Koijen (2010).

[4]The present value system in Equations 1 and 2 obtains as a special case of the model in Section I. It arises in an economy where μ_t and g_t each follow an AR(1), individual expected returns obey an exact one factor model as in the CAPM, $\mu_{i,t} = \mu_{i,0} + c_{i,\mu}\mu_t$, and individual expected dividend growth obeys a one factor model, $g_{i,t} = g_{i,0} + c_{i,g}g_t + e_{i,t}$. This special case has been considered by van Binsbergen and Koijen (2010) and Cochrane (2008a).

while the market price-dividend ratio is

$$pd_t = a - b_\mu \mu_t + b_g g_t. \tag{2}$$

Equation 2 highlights the predictive relationship between pd_t, realized market returns ($r_{t+1} = \mu_t + \epsilon^r_{t+1}$) and dividend growth ($\Delta d_{t+1} = g_t + \epsilon^d_{t+1}$). However, it also evokes limitations of the aggregate system for understanding market expectations. Predictive regressions of r_{t+1} on pd_t take the form

$$\mathbb{E}_t[r_{t+1}|pd_t] = \hat{a} + \hat{b}pd_t = \hat{a} + \hat{b}(b_\mu \mu_t + b_g g_t)$$

and thus are contaminated by dividend growth information. The reciprocal problem arises in forecasting Δd_{t+1}. To overcome this difficulty, researchers have taken present value approaches that account for the joint relationship among pd_t, μ_t and g_t (Cochrane 2008b, Lettau and Van Nieuwerburgh 2008, van Binsbergen and Koijen 2010). While this begins to disentangle the link between prices and expectations, these joint systems continue to rely solely on aggregate variables. Because both μ_t and g_t are latent, each adds noise to the signal extraction problem of the other.[5] If there exist other signals for μ_t and g_t in the economy, incorporating them will improve estimates of the latent expectations. This is where disaggregated valuation ratios in Equation 1 become a valuable information source. Each $pd_{i,t}$ provides a new, non-redundant signal for both μ_t and g_t.

The 3PRF conveniently reduces the many available signals to an optimal forecast of returns or dividend growth. Kelly and Pruitt (2011) show how the 3PRF may be interpreted (and indeed even implemented) as a series of ordinary least squares regressions. In the first stage, individual price-dividend ratios for each asset are regressed on observable (but noisy) proxies for the true, latent factors. Next, in each time period, we run second stage

[5] This remains true despite the absence of measurement error in the aggregate price-dividend expression, as pointed out by Fama and French (1988).

cross-sectional regressions of assets' price-dividend ratios on firm-specific regression coefficients estimated in the first stage. In the final stage, aggregate return and dividend growth realizations are regressed on the fitted factors from the second stage, delivering our final filtered estimates for unobservable return and dividend growth expectations. The final-stage predictor is a discerningly-constructed linear combination of disaggregated price-dividend ratios that parsimoniously incorporates information from individual valuation ratios into predictions of future aggregate returns and dividend growth.

Building on this multi-stage OLS interpretation of the filter, the preceding CAPM example is also useful to develop intuition for how the 3PRF works. Each $pd_{i,t}$ is a function of only the *expected* portion of returns and dividend growth and is uncorrelated with their unanticipated future shocks. Therefore, first stage time series regression coefficients of $pd_{i,t}$ on r_{t+1} and Δd_{t+1} (which serve as observable proxies for the latent factors μ_t and g_t) describe how each firm's price-dividend ratio depends on the true factors μ_t and g_t. When the coefficients $b_{i,\mu}$ and $b_{i,g}$ differ across i, fluctuations in μ_t and g_t cause the cross section of price-dividend ratios to fan out and compress over time. The first-stage coefficient estimates provide a map from the cross-sectional distribution of pd_i's to the latent factors. Second-stage cross section regressions of $pd_{i,t}$ on first-stage coefficients use this map to estimate the factors at each point in time. Because the first-stage regression takes an errors-in-variables form, second-stage regressions estimate a rotation of the latent expectation vector $(\mu_t, g_t)'$. The third-stage projects realized returns and growth onto the factor space, which is unaffected by factor rotation, to deliver consistent estimates of μ_t and g_t.

The earliest predecessors of this study use predictive regressions based on the aggregate price-dividend ratio, including Rozeff (1984), Campbell and Shiller (1988), Fama and French (1988). A large subsequent literature has evolved testing the predictive relation, including Stambaugh (1986, 1999), Hodrick (1992), Goetzmann and Jorion (1993), Nelson and Kim (1993), Kothari and Shanken (1997), Lewellen (2004), Paye and Timmermann (2006), and Pástor and Stambaugh (2009a).

More directly, our paper builds on recent literature that exploits the present value relation to identify market expectations for returns and dividends, including van Binsbergen and Koijen (2010), Ghosh and Constantinides (2010), Ferreira and Santa-Clara (2010), Cochrane (2008a,b), Pástor, Sinha, and Swaminathan (2008), Rytchkov (2008), Campbell and Thompson (2008), Lettau and Van Nieuwerburgh (2008), Ang and Bekaert (2007), Lettau and Ludvigson (2005), Brennan and Xia (2005) and Menzly, Santos and Veronesi (2002). While these papers focus on aggregate present value models, the key to our analysis is incorporating cross-sectional information. Kelly (2010), Hansen, Heaton, and Li (2008), Pástor and Veronesi (2003, 2006), Kiku (2006), and Vuolteenaho (2002) also model valuation ratios for individual assets, though we are the first to exploit a factor structure in price-dividend ratios to form market return and cash flow forecasts.

Our three-pass regression filter is reminiscent of two-pass regression used in tests of cross-sectional beta-pricing models (see Fama and MacBeth 1973, Shanken 1992 and Jagannathan and Wang 1998).[6] Both techniques rely on cross-sectional dispersion of financial variables to infer market risk prices. There are two key differences between our three-pass latent variable filter and two-pass return tests. First, our cross section is constituted of price-dividend ratios rather than returns. Second, we string together period-by-period estimates from second-stage regressions to construct our key predictor variable, as opposed to averaging second-stage output over time to find a single risk price. Our technique is also related to Chowdhry, Roll, and Xia (2005), who construct estimates of the latent inflation time series from the cross section of returns using two-pass regression, and Polk, Thompson and Vuolteenaho (2006) who use a CAPM-motivated two-pass approach to forecasting returns. Methodologically, our procedure builds upon a rich literature on latent variable filters beginning with Kalman (1960). The filter we use is less computationally intensive and can be less sensitive to misspecification. The 3PRF is also closely linked to principal components estimators such as Stock and Watson (2002). The fundamental difference between our approach and principal

[6] Kelly and Pruitt (2011) formalize the sense in which the Fama-MacBeth procedure can be interpreted as a latent factor filter.

components is that the 3PRF forms predictive factors based on covariance between the forecast target and individual predictors, whereas principal components seeks to maximize the explained variation in the cross section of predictors. As we show, state space filters and principal components have less forecasting success in the present value setting.

In the next section we present an economic framework for the cross section of present values. Section II introduces our three-pass regression filter estimation. In Section III we present empirical findings, compare alternative methodologies and discuss our results. We present our conclusions in Section IV. Details about the three-pass filter, as well as additional proofs, technical assumptions and other details, are relegated to the appendix.

I The Cross-Sectional Present Value System

We assume that one-period expected log returns and log dividend growth rates across assets and over all horizons are linear in a set of common factors[7]

$$\begin{aligned} \mu_{i,t} &= \mathbb{E}_t[r_{i,t}] \quad = \gamma_{i,0} + \boldsymbol{\gamma}_i' \boldsymbol{F}_t \\ g_{i,t} &= \mathbb{E}_t[\Delta d_{i,t}] = \delta_{i,0} + \boldsymbol{\delta}_i' \boldsymbol{F}_t + \varepsilon_{i,t}. \end{aligned} \tag{3}$$

Equation 3 states that, conditional on time t information, expected one-period returns and growth rates at date t are driven solely by the $(K_F \times 1)$ vector of factors $\boldsymbol{F}_t$. Factor loading vectors $\boldsymbol{\gamma}_i$ and $\boldsymbol{\delta}_i$ summarize how market expectations respond to movements in the underlying economic factors. We assume that assets' expected returns are determined by systematic factors and possess no idiosyncratic behavior.[8] The same restriction is not imposed for indi-

[7]Factor models are analytically tractable and are sufficiently general to subsume a wide range of models considered in the asset pricing literature. Asset pricing models, both theoretical and empirical, link individual expected returns to aggregate expected returns either directly, as in the CAPM (Sharpe 1964, Lintner 1965, Treynor 1961) and Fama-French model (Fama and French 1993), or indirectly via common state variables, as in Merton's (1973) ICAPM. Similarly, theoretical models commonly assume a factor structure in dividend growth (Connor 1984, Bansal and Viswanathan 1993, and Bansal, Dittmar, and Lundblad 2005, among others).

[8]This assumption can be relaxed. Allowing for an orthogonal idiosyncratic component in firms' expected returns has no impact on the development or implementation of our approach.

vidual expected dividend growth, which may possess an idiosyncratic component, $\varepsilon_{i,t}$. The aggregate market obeys the same structure, with no i subscripts, and we impose that the factor model is exact for aggregate dividend growth:

$$\begin{aligned} \mu_t &= \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{F}_t \\ g_t &= \delta_0 + \boldsymbol{\delta}'\boldsymbol{F}_t. \end{aligned} \tag{4}$$

Realized returns and growth rates are equal to their conditional expectations plus an unforecastable shock:

$$\begin{aligned} r_{i,t+1} &= \mu_{i,t} + \eta^r_{i,t+1} \\ \Delta d_{i,t+1} &= g_{i,t} + \eta^d_{i,t+1}. \end{aligned}$$

Finally, we assume that the factor vector evolves as a first order vector autoregression[9]

$$\boldsymbol{F}_{t+1} = \boldsymbol{\Lambda}_1 \boldsymbol{F}_t + \boldsymbol{\xi}_{t+1}. \tag{5}$$

The log price-dividend ratio for any asset is tied to its future return and dividend growth rate according to the Campbell and Shiller (1988) approximate present value identity, which states

$$pd_{i,t} = \kappa_i + \rho_i pd_{i,t+1} - r_{i,t+1} + \Delta d_{i,t+1},$$

where $\rho_i = \frac{\exp(\mathbb{E}pd_{it})}{1+\exp(\mathbb{E}pd_{it})}$ and $\kappa_i = \log\left(1 + \exp\left(\mathbb{E}pd_{it}\right)\right)$. Iterating this equation forward and taking time t expectations delivers the following *ex ante* expression for the price-dividend ratio

$$pd_{i,t} = \frac{\kappa_i}{1-\rho_i} + \sum_{j=1}^{\infty} \rho_i^{j-1} \mathbb{E}_t[-r_{i,t+j} + \Delta d_{i,t+j}].$$

This weighted sum of expected one-period returns and growth rates over all future horizons,

[9] That $\boldsymbol{F}_t$ is a first order process is without loss of generality since any higher order vector autoregression can be written as a VAR(1).

combined with (3) and (5), reduces the *ex ante* price-dividend ratio to

$$pd_{i,t} = \phi_{i,0} + \boldsymbol{\phi}_i' \boldsymbol{F}_t + \varepsilon_{i,t} \tag{6}$$

where expressions for $\phi_{i,0}, \boldsymbol{\phi}_i$ and $\varepsilon_{i,t}$ are provided in Appendix A.A. Equations 4 and 6 unify disaggregated valuation ratios and aggregate expectations via a common factor model. They also provide a framework for utilizing cross section information to achieve our ultimate goal of precisely estimating conditional expected market returns and dividend growth. By a derivation similar to that for (6), the Campbell-Shiller equation for the aggregate market is

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \left(-\mathbb{E}_t[\mu_{t+j}] + \mathbb{E}_t[g_{t+j}]\right) \\ &= \phi_0 + \boldsymbol{\phi}' \boldsymbol{F}_t. \end{aligned} \tag{7}$$

The first equality in (7) captures a fundamental feature of the aggregate price-dividend ratio: It is an amalgam of expected returns and expected growth over all future horizons. The framework we use here preserves the generality of this statement. Recent literature has posited that expected one-period returns and dividend growth follow an AR(1). This places a restriction among expected one-period returns and growth rates at different horizons. In particular, it implies that $\mathbb{E}_t[\mu_{t+j}]$ is a linear function of μ_t for all $j > 1$ (and similarly for g_{t+j}), hence $\mathbb{E}_t[\mu_{t+j}]$ and μ_t are perfectly correlated for any j. This further implies that the aggregate price-dividend ratio can be represented by Equation 2, so that *any and all* variation in pd_t is due to fluctuations in expected returns and dividend growth one period ahead, μ_t and g_t.[10] By relaxing this restriction, our framework allows expectations at different horizons to influence valuation ratios differently.[11]

[10] See Cochrane (2008a) and van Binsbergen and Koijen (2010). Gabaix (2010) also arrives at an expression closely related to (2) using modified AR(1), or linearity-generating, processes.

[11] It is not a complete relaxation, however. The factors $\boldsymbol{F}_t$ are the only conduits linking future expectations and current valuations. In other words, conditional expectations at any horizon must solely be a function of the factor vector. Expectations for alternative horizons may be differentially tied to the factors and as a result may be imperfectly correlated (thus inconsistent with an AR(1) for μ_t).

II The Three-Pass Regression Filter

In this section we outline our empirical methodology, which is based on Kelly and Pruitt's (2011) three-pass regression filter. Interested readers can refer to that paper for detailed econometric development and proofs of results stated below. Assumptions underlying the stated results are given in Appendix A.B.

To ease the algebraic development we first establish notation. The 3PRF produces forecasts using three sets of inputs. The first input is the forecasting target, which in general takes the form $y_{t+1} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t + \eta_{t+1}$. We will focus primarily on two targets, aggregate market returns and dividend growth, implying

$$y_{t+1} = \begin{cases} \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{F}_t + \eta^r_{t+1} & \text{if} \quad y_{t+1} = r_{m,t+1} \\ \delta_0 + \boldsymbol{\delta}'\boldsymbol{F}_t + \eta^d_{t+1} & \text{if} \quad y_{t+1} = \Delta d_{m,t+1}. \end{cases}$$

Defining $\underset{(T\times K_F)}{\boldsymbol{F}} = [\boldsymbol{F}_1, \boldsymbol{F}_2, ..., \boldsymbol{F}_T]'$, the matrix representation of y_{t+1} is

$$\begin{aligned} \underset{(T\times 1)}{\boldsymbol{y}} &= [y_2, y_3, ..., y_{T+1}]' \\ &= \boldsymbol{\iota}\beta_0 + \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\eta}. \end{aligned}$$

The second input is the cross section of price-dividend ratios $pd_{i,t} = \phi_{i,0} + \boldsymbol{\phi}_i'\boldsymbol{F}_t + \varepsilon_{i,t}$ $(i = 1, ..., N)$. These are arranged into the vector $\boldsymbol{x}_t = (pd_{1,t}, ..., pd_{N,t})'$ and stacked as

$$\begin{aligned} \underset{(T\times N)}{\boldsymbol{X}} &= [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T]' \\ &= \boldsymbol{\iota}\boldsymbol{\phi}_0' + \boldsymbol{F}\boldsymbol{\Phi}' + \boldsymbol{\varepsilon} \end{aligned}$$

where $\underset{(N\times 1)}{\boldsymbol{\phi}_0} = [\phi_{1,0}, \phi_{2,0}, ..., \phi_{N,0}]'$ and $\underset{(N\times K_F)}{\boldsymbol{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]'$.

The third input is a set of K_Z proxies for the latent factors driving the present value system in 6. Proxies are *observable* linear functions of the true factors plus potentially some noise. We denote proxies as $\boldsymbol{z}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{\omega}_t$. It is immediately clear that despite the latency of $\boldsymbol{F}_t$ there exist many viable proxies for $\boldsymbol{F}_t$ within the present value system. For example, each price-dividend ratio qualifies as a proxy. Future realized returns and dividend growth for any asset are also potential proxies since they are linear functions of $\boldsymbol{X}_t$ plus noise. In our empirical analysis, we use the price-dividend ratio, return and dividend growth of the aggregate market as proxies. We return to the question of proxy selection in more detail below. The matrix representation for proxies is

$$\begin{aligned} \underset{(T\times K_Z)}{\boldsymbol{Z}} &= [\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_T]' \\ &= \boldsymbol{\iota}\boldsymbol{\lambda}_0' + \boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\omega}. \end{aligned}$$

The three-pass regression filter is the solution to a modified recursive least squares problem. The typical recursive least squares solution is the Kalman filter, which infers latent factor processes by aggregating long histories of signals into forecasts for future states. The 3PRF drastically simplifies the Kalman solution by foresaking past data and instead drawing inference for factors by aggregating over a large cross section of signals. The cost of sacrificing time aggregation is that in order to obtain consistent forecasts of the system (as the Kalman filter does) we require a sufficiently large cross section. In cross-sectionally data rich environments such as asset pricing, this is an easy trade off to make. It avoids maximum likelihood estimation of the Kalman filter, which is effectively infeasible with large cross sections, and instead uses much simpler OLS estimation to take advantage of the many asset prices available for analysis.

We first define the estimator and state some of its properties, then discuss the intuition behind the approach.

Definition 1 (Kelly and Pruitt (2011) Lemma 1)**.** *The three-pass regression filter forecast* *of* $\boldsymbol{y}$ *using cross section* $\boldsymbol{X}$ *and proxies* $\boldsymbol{Z}$ *is*[12]

$$\hat{\boldsymbol{y}} = \boldsymbol{\iota}\bar{y} + \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{X}' \boldsymbol{J}_T \boldsymbol{Z} \left(\boldsymbol{Z}' \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{X}' \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{X}' \boldsymbol{J}_T \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}' \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{X}' \boldsymbol{J}_T \boldsymbol{y}. \quad (8)$$

$\boldsymbol{J}_L \equiv \boldsymbol{I}_L - L^{-1}\boldsymbol{\iota}_L\boldsymbol{\iota}_L'$, *where* $\boldsymbol{I}_L$ *is the* L*-dimensional identity matrix and* $\boldsymbol{\iota}_L$ *is a* L*-vector of ones.*

The assumptions for our main convergence result can be found in the appendix and are quite weak. We require that the signal portion of the proxies (the vector $\boldsymbol{\Lambda F}_t$) spans the factor space, which amounts to $\boldsymbol{\Lambda}$ being full rank. Additional technical assumptions impose that second moments are finite and probability limits are well-behaved, that there is limited time series and cross-sectional autocorrelation among elements of the residual matrices $\boldsymbol{\eta}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\omega}$, and that innovations to returns and dividend growth are orthogonal to lagged factors and price-dividend ratios.

Proposition 1 (Kelly and Pruitt (2011) Proposition 1)**.** *Let Assumptions 1, 2, 3 and 4 hold. Then the three-pass regression filter forecaster of* $\boldsymbol{y}$ *using cross section* $\boldsymbol{X}$ *and proxies* $\boldsymbol{Z}$ *converges in probability to the infeasible best forecast,* $\boldsymbol{\iota}\beta_0 + \boldsymbol{F\beta}$*, as* $T, N \to \infty$*.*

Definition 1 states the solution to a restricted recursive least squares problem as derived in Kelly and Pruitt (2011). The estimator takes on an intuitive interpretation as a series of ordinary least squares regressions (from which its name derives) according to the following steps.[13] In the first stage, run a time series regression of the log price-dividend ratio for each stock i on the observable proxies

$$pd_{i,t} = \hat{\phi}_{i,0} + \hat{\boldsymbol{\phi}}_i' \boldsymbol{Z}_t + e_{i,t}.$$

[12]The $\boldsymbol{J}_L$ terms in the expression of $\hat{\boldsymbol{y}}$ appropriately handle constant terms in $\boldsymbol{X}$, $\boldsymbol{y}$ and $\boldsymbol{Z}$ when calculating the sample covariances (the cross product matrices composing $\hat{\boldsymbol{y}}$). When the input matrices are mean zero, $\boldsymbol{J}_L$ terms drop out reducing the expression to $\hat{\boldsymbol{y}} = \boldsymbol{\iota}\bar{y} + \boldsymbol{X}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{y}$.

[13]In fact, the series of regressions described in this section exactly replicates the estimator defined in Equation 8. This has the advantage that the 3PRF may be implemented within any software package that incorporates OLS.

The resulting estimate $\hat{\boldsymbol{\phi}}_i$ describes the sensitivity of each $pd_{i,t}$ to the latent factors. Because proxies like the market price-dividend ratio and realized return imperfectly represent the true factors, $\hat{\boldsymbol{\phi}}_i$ is a biased description of assets' factor sensitivities. Ultimately this does not affect forecasts as subsequent steps compensate for proxy imperfection to "de-bias" estimates.

In the second stage, for each period t, run a cross-sectional regression of assets' price-dividend ratios on their factor loadings estimated in the first stage

$$pd_{i,t} = \hat{c}_t + \hat{\boldsymbol{F}}_t{}' \hat{\boldsymbol{b}}_i + w_{i,t}.$$

Here, the first stage loadings become the independent variables, and the latent factors $\boldsymbol{F}_t$ are the coefficients to be estimated. The bias in $\hat{\boldsymbol{\phi}}_i$ from the first stage carries over to the second stage, whose estimates $\hat{\boldsymbol{F}}_t$ are asymptotically equal to a rotated version of the factor vector $\boldsymbol{F}_t$. The first two stages exploit the factor nature of the system to draw inferences about the underlying factors. As the factors fluctuate, the cross section of price-dividend ratios fans out and compresses over time. If the true factor loading $\boldsymbol{\phi}_t$ were known we could estimate the latent factors consistently by simply running cross section regressions period-by-period. Since $\boldsymbol{\phi}_t$ is unknown, the first-stage regression coefficients provide a preliminary description of how each pd_i depends on $\boldsymbol{F}_t$. This first stage regression sketches a map from the cross-sectional distribution of pd_i's to the latent factors. Second-stage cross section regressions of pd_i's on first-stage coefficients use this map to produce estimates of the factors at each point in time. Because the first-stage regression takes an errors-in-variables form, the second-stage regressions produce an estimate for a unique but unknown rotation of the latent expectation vector $(\mu_t, g_t)'$.

The third and final step in the filter runs predictive regressions of realized returns and dividend growth rates on the lagged factor rotations estimated in the second stage. The purpose of the third regression step is to "un-rotate" the second stage factor estimates and asymptotically recover the exact linear combinations of latent factors that constitute expectations of aggregate market returns and dividend growth. The second stage estimates converge to $\boldsymbol{\Gamma F}_t$, where $\boldsymbol{\Gamma}$ is square and nonsingular. The final stage forecasting regression fit mechanically converges to $\mathbb{E}_t[y_{t+1}] = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma F}_t$. This is a matrix analogue to the fact that least squares fitted values are invariant to scalar multiples of regressors.

This final predictive regression is the culmination of the multi-asset present value system. It parsimoniously combines information from individual assets' valuation ratios to arrive at a prediction of future aggregate returns and dividend growth. The ultimate predictor, $\hat{\boldsymbol{F}}_t$, is a discerningly-constructed linear combination of disaggregated price-dividend ratios that collapses the cross section system to its fundamental driving factors. The R^2 from the final step regression summarizes the predictive power of the multi-asset present value model; that is, the predictive accuracy of market expectations embodied in valuation ratios.

II.A Comparing Predictive Contribution of Individual Price-Dividend Ratios

We may rewrite Equation 8 as $\hat{\boldsymbol{y}} = \boldsymbol{\iota}\bar{y} + \boldsymbol{XB}$.[14] $\boldsymbol{B}$ is the $N \times 1$ vector of each price-dividend ratio's contribution to the total forecast. This vector, available in a convenient closed form, defines the best linear combination of price-dividend ratios for forecasting returns or dividend growth according to the three-pass regression filter.

To evaluate the forecasting fit, we calculate the predictive $R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}$. The 3PRF predictive R^2 is equal to the R^2 of the third stage regression. This fact highlights the sense in which the 3PRF is a dimension-reduced reduced forecast. The third stage regression forecasts returns and dividend growth using only the K_Z factors extracted from the cross section (we set $K_Z = 1$ or 3 in our empirical analysis), thus our ultimate 3PRF predictions and associated R^2 are interpreted as those of a K_Z regressor system.

[14]The loading matrix is $\boldsymbol{B} = \boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{XX}'\boldsymbol{XX}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{XX}'\boldsymbol{y}$. The stated expressions suppress $\boldsymbol{J}$ matrices, as in the zero constant case. In the general case we have $\hat{\boldsymbol{y}} = \boldsymbol{\iota}\bar{y} + \boldsymbol{J}_T\boldsymbol{XJ}_N\boldsymbol{B}$, where $\boldsymbol{B} = \boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{XJ}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{XJ}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{XJ}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{y}$.

II.B Proxy Variable Selection and Parsimony

Proxies can be decomposed into two components. The "signal" component is a linear combination of the underlying factors, and the "noise" component is an orthogonal shock. When selecting proxies for the latent factors, the theory outlined above requires only that proxies' signal components span the factor space of the cross section system. An important question, then, is how many proxies to use? The philosophy of parsimony underlying the factor framework seeks to summarize the cross section system with a small number of fundamental drivers. We consider specifications of up to three factors in our implementation.

When the factors driving expected returns or expected dividend growth are only a subset of $\boldsymbol{F}_t$, the requirement that proxies span the factor space may be weakened. For example, consider a case in which factors can be partitioned as $\boldsymbol{F}_t = (\boldsymbol{f}'_t, \boldsymbol{h}'_t)'$, where only $\boldsymbol{f}_t$ drives expected returns and $\boldsymbol{h}_t$ affects only expected growth.[15] In this case, it is unnecessary to proxy for $\boldsymbol{h}_t$ when forecasting returns. Therefore, the three-pass filter may asymptotically recover μ_t despite using a number of proxies inferior to the dimension of $\boldsymbol{F}_t$. The condition for this to work is that the proxies, like the forecast target, must be independent of the remaining $\boldsymbol{h}_t$.

Proposition 2 (Kelly and Pruitt (2011) Proposition 2). *Let Assumptions 1, 2 and 3 hold. Partition the factor vector as $\boldsymbol{F}_t = (\boldsymbol{f}'_t, \boldsymbol{h}'_t)'$ and assume that y_{t+1} and $\boldsymbol{Z}_t$ have zero loadings on $\boldsymbol{h}_t$. Finally, assume that the rank of matrix $\boldsymbol{\Lambda}$ is equal to the dimension of $\boldsymbol{f}_t$. Then the three-pass regression filter forecaster of $\boldsymbol{y}$ using cross section $\boldsymbol{X}$ and proxy $\boldsymbol{Z}$ converges in probability to the infeasible best forecast, $\boldsymbol{\iota}\beta_0 + \boldsymbol{F}\boldsymbol{\beta}$, as $T, N \to \infty$.*

This result is important for three reasons. First, it can improve the parsimony of an empirical forecasting model. Proposition 2 states that it is possible to proxy for only a subset of the factors without decreasing forecasting power. Second, this highlights differences in our approach versus the principal components-based forecasting approach of Stock and Watson (2002). The 3PRF can produce optimal forecasts even if the principal components important

[15] This corresponds to appropriately placed zeros in the $\boldsymbol{\gamma}$ vector.

for explaining cross section variation are irrelevant for forecasting returns or growth. Principal components, on the other hand, is designed to explain variation in the cross section. It then uses the dominant drivers of this variation for forecasting, even if those drivers are unrelated to the forecast target.

Third, this result offers guidance for selecting the most appropriate proxies. For instance, when forecasting returns, ideal proxies are those that depend *only* on the factors driving μ_t. They should also be independent of those variables driving other expectation that contribute to $pd_{i,t}$ but that do not affect μ_t. The realized market return is a proxy that satisfies these conditions since, by definition, it is equal to μ_t plus an orthogonal shock. This also cautions against using the aggregate price-dividend ratio as a lone proxy since it can produce poor forecasts of both returns and dividend growth. In fact, whenever using a single proxy, the optimal proxy is the forecast target itself. Intuitively, proxying for expected returns with realized returns gives the 3PRF the best chance to identify the part of cross sectional variation in price-dividend ratios that is due to expected returns (similarly for expected growth rates). Corollary 1 and Proposition 3 of Kelly and Pruitt (2011) formalize this logic.

To clarify the rationale in a simplified setting, we return to the CAPM example economy presented in the introduction. Recall that in this example economy the cross section system obeys $pd_{i,t} = a_i - b_{i,\mu}\mu_t + b_{i,g}g_t + e_{i,t}$ and the aggregate price-dividend ratio is given by $pd_{i,t} = a - b_\mu \mu_t + b_g g_t$. Consider forecasting the market return r_{t+1} using two potential implementations of the three-pass regression filter, first using pd_t as a proxy and then using r_{t+1} as a proxy. To simplify the example, we consider population regressions and assume that μ_t and g_t are independent. We also assume that loadings $b_{i,\mu}$ and $b_{i,g}$ are independently distributed in the cross section.

Case 1: Aggregate Price-Dividend Ratio as Proxy

The first-stage time series regression of each $pd_{i,t}$ series on pd_t produces coefficients

$$coeff_{i,1} = \frac{b_{i,\mu}\sigma_\mu^2 b_\mu + b_{i,g}\sigma_g^2 b_g}{\sigma^2(pd_t)}.$$

Note that both $b_{i,\mu}$ and $b_{i,g}$ enter the first-stage estimate even though the ultimate forecasting target is unrelated to g_t. This is because the proxy differs from the target, r_{t+1}, which opens the door to contamination from factors that have no relation to r_{t+1}. The second-stage cross section regression of $pd_{i,t}$ on $\mathit{coeff}_{i,1}$ produces a time series that takes the form[16]

$$\mathit{coeff}_{t,2} = z_\mu \mu_t + z_g g_t.$$

Because $b_{i,\mu}$ and $b_{i,g}$ appear in the first-stage estimate, cross-sectional regressions pick up variation in the dispersion of price-dividend ratios that is due to both μ_t and g_t. As a result, both μ_t and g_t appear in the forecasting variable $\mathit{coeff}_{t,2}$ rendering this forecast suboptimal. The return depends only on μ_t, thus the presence of g_t in the forecasting variable ultimately results in noisy estimates of the latent μ_t.

Case 2: Realized Market Return as Proxy

The first pass regression of $pd_{i,t}$ on r_{t+1} delivers

$$\mathit{coeff}_{i,1} = \frac{b_{i,\mu}\sigma_\mu^2}{\sigma^2(r_{t+1})}.$$

Because r_{t+1} correlates only with the part of $pd_{i,t}$ that is due to μ_t, the coefficient $\mathit{coeff}_{i,1}$ depends on $b_{i,\mu}$ but not $b_{i,g}$. Thus, the second-stage regression produces the forecasting variable

$$\mathit{coeff}_{t,2} = \frac{\sigma_\mu^2 + \sigma_\epsilon^2}{\sigma_\mu^2}\mu_t.$$

Because this predictor variable is a constant scalar time μ_t, it generates optimal forecasts of r_{t+1} in the final stage.

While our benchmark empirical results consider three proxies, we also examine forecasta-

[16]Where

$$z_\mu = \frac{\sigma^2(pd_t)\sigma^2(b_{i,g})b_g\sigma_g^2 g_t \sigma^2(b_{i,\mu})b_\mu\sigma_\mu^2}{\sigma^2(b_{i,\mu})b_\mu^2\sigma_\mu^4 + \sigma^2(b_{i,g})b_g^2\sigma_g^4} \quad \text{and} \quad z_g = \frac{\sigma^2(pd_t)\sigma^2(b_{i,g})b_g\sigma_g^2 g_t \sigma^2(b_{i,g})b_g\sigma_g^2}{\sigma^2(b_{i,\mu})b_\mu^2\sigma_\mu^4 + \sigma^2(b_{i,g})b_g^2\sigma_g^4}.$$

bility based on one proxy. Our proxy choice in this case, based on the preceding argument, uses future realized returns as the sole proxy in return forecasts, and the sole proxy for dividend growth is the realized future growth rate.

II.C Implementing the Three-Pass Regression Filter

Throughout our empirical analysis we consider three different approaches for implementing the three-pass regression filter that vary in the information sets used to construct forecasts. To outline the differences in information sets it is convenient to work within the three-stage regression construction of the filter rather than the direct formulation in (8).

We refer to our basic procedure as the *full information* filter. In this version, first-stage regressions use the full time series of data to estimate factor loadings. Second-stage regressions that produce the predictor variable at each time t use only price-dividend ratio data at time t and constant factor loadings estimated in the first stage. Finally, third-stage predictive regressions are run in-sample.

In the full information version it is possible that first-stage regressions introduce a small sample bias in our predictors since first-stage factor loadings are based on the full time series. This is analogous to small sample bias in standard predictive regression (Stambaugh 1986, Nelson and Kim 1993), or in the preliminary parameter estimation step of a Kalman-filtered state space. This can favor false detection of predictability, though the effect fades with sample size as each future observation has vanishing importance for first-stage coefficient estimates. Since we are constrained to work with a relatively small sample of 64 years, however, we take the possibility of small sample-induced look-ahead bias very seriously.

As a first alternative to the full information procedure, we calculate a *no-peek* version of our filter that reduces the effect of small sample bias while sacrificing minimal precision in estimating first-stage coefficients. In each period t beginning with $t = 1$, we run first-stage time series regressions omitting observations for dates $\{t + 1, t + 2, t + 3\}$.[17] We

[17] We omit the subsequent three years of data to account for persistence in the data. We choose three years

Table I: Implementation Summary

Version	First Stage	Third Stage
Full Information	Same for all t: estimated from full sample $\{1,...,T\}$.	Same for all t: estimated from full sample $\{1,...,T\}$.
No-Peek	For each t: estimated from sample $\{1,...,t,t+4,...,T\}$.	Same for all t: estimated from full sample $\{1,...,T\}$.
Recursive Out-of-sample	For each $t \geq T_{train}$: estimated from sample $\{1,...,t\}$.	For each $t \geq T_{train}$: estimated from sample $\{1,...,t\}$.

Notes: Summary of three-pass regression filter implementation schemes. T_{train} denotes the time series length of the training sample for the pure out-of-sample procedure. In all cases, the second stage regresses the time t cross section of valuation ratios on the first-stage coefficients for time t, and the time $t+1$ predictor is formed by multiplying the time t second-stage result by the time t third-stage coefficients.

construct the date t observation for our predictor variable from the second-stage cross section regression of time t price-dividend ratios on first-stage estimates (which ignore observations $\{t+1, t+2, t+3\}$). At $t+1$ we re-estimate first-stage coefficients dropping observations for dates $\{t+2, t+3, t+4\}$, and calculate the date $t+1$ value of our forecasting variable from the second-stage regression, and so on. Once we have exhausted the time series, we run a single third-stage forecasting regression based on the no-peek predictor, which has been constructed to explicitly preclude any unduly favorable effect from small samples. It is important to note that no-peek estimates sacrifice observations from an already short sample, which can non-negligibly increase sampling error. Ultimately, we cannot distinguish if differences in forecasting results between full information and no-peek estimates are due to changes in look-ahead bias or sampling variation.[18]

Our second and more stark alternative to the full information filter is a pure out-of-

to balance the benefits of avoiding look-ahead bias against the costs of losing observations from an already short sample. Results are insensitive to alternative choices such as dropping one year or five years of data.

[18] More precisely, the no-peek procedure has two effects on the filter's estimates. It eliminates the effect of small-sample bias by omitting observations that are eventually forecasted from preliminary parameter estimation. This decreases the absolute correlation between our predictor variable and forecasting target, giving a cleaner measurement of forecasting power. This effect on the R^2 is useful since it clarifies assessments of our predictor's power. A costly side-effect is that the sample is shortened by 5%, thereby weakening the precision of our estimates and obscuring correlation between the predictor and target. This tends to also decrease the R^2, which is harmful to our ability to detect the predictive power. Unfortunately, both of these effects ultimately decreases the no-peek R^2 relative to the full information case, making it difficult to ascertain whether the effect is due to look-head bias or increased estimation noise.

sample analysis. The procedure we use is a standard recursive out-of-sample estimation scheme which has been well-studied in the literature and affords us "encompassing" tests for the statistical significance of out-of-sample performance (see, for example, Clark and McCracken 2001 and Goyal and Welch 2008). It proceeds as follows. We split the 1946-2009 sample at 1980,[19] using the first 34 observations as a training sample and the last 30 observations as the out-of-sample horizon.[20] Beginning with $t = 34$, we estimate first-stage factor loadings using observations $\{1, ..., t\}$. Then, for each period $\tau \in \{1, ..., t\}$, we estimate the time τ value of our predictor variable using the cross section of valuation ratios at τ and first-stage coefficients (which are based on data $\{1, ..., t\}$). We then estimate the predictive coefficient in a third-stage forecasting regression of realized returns (or dividend growth) for periods $\{2, ..., t\}$ on our predictor from $\{1, ..., t-1\}$. Finally, our out-of-sample forecast of the $t+1$ return is the product of the third-stage predictive coefficient and the time t second-stage result. At time $t+1$, we construct our forecast of the return at $t+2$ by repeating the entire three stage procedure using data from $\{1, ..., t+1\}$. This process is iterated forward each year until the entire time series has been exhausted.

For the reader's reference, we summarize the key characteristics of these three procedures in Table I. The differences lie in the sample that is used to estimate the first- and third-stage coefficients. In the full information procedure, these first-stage coefficients are the same for every time t; in the no-peek and out-of-sample procedures, these first-stage coefficients are different for each t due to the varying estimation samples. For every procedure, the second stage is the same: A time t cross section regression of valuation ratios on the first-stage coefficients. In all cases, the eventual predictor is formed the same way: the second-stage result for time t is multiplied by the third-stage predictive coefficient. In the full information and no-peek procedures, these third-stage coefficients are the same for every time t; in the out-of-sample procedure these third-stage coefficients are different for each t.[21]

[19] For our empirical analysis using the 1930-2009 sample, we split the sample at 1970.

[20] In Figure 3 we show that our findings are robust to alternative split dates.

[21] In the exposition of our filter, proxy time subscripts are the same as those for valuation ratios. However, when we use the realized return or dividend growth as a proxy, the time $t+1$ realizations will be lined up

Because of the multi-step nature of our estimator, inference for our ultimate forecasts must take into account sampling variation in parameter estimates from earlier stages. We therefore calculate test statistics via bootstrap. In all cases, we build pseudo-samples by resampling price-dividend ratios, returns and dividend growth by the circular block bootstrap of Politis and Romano (1992). The reported p-values are based on the sampling distribution of estimates that arises by running the three-pass filter for each pseudo-sample. By using blocks, our inference will be robust to serial correlation in our data series.[22]

For our out-of-sample results we rely on the test of forecast encompassing derived by Clark and McCracken (2001). This tests the null hypothesis that two predictors provide the same out-of-sample forecasting performance. When this statistic, we are testing the denoted predictor versus the historical mean of the target series.[23] We report significance levels as found from Clark and McCracken's (2001) appendix tables, where critical values for the 10%, 5% and 1% levels are provided. The notation "$< x$" represents the smallest significance level x for which the encompassing test statistic exceeds the critical value.

III Empirical Results

III.A Data

Our empirical analysis examines market return and dividend growth predictability by applying the three-pass regression filter to various sets of test portfolios. We examine the post-war sample (1946-2009) to be comparable with earlier work on this topic (e.g. Cochrane 2008b), as well as the extended (1930-2009) sample. We focus on annual data to avoid seasonality

with time t price-dividend ratios in the first-stage. When our proxies are period-ahead realizations, we must appropriately lag our data an extra period. Therefore, when realized returns or dividend growth are proxies, the first-stage time t regression observation consists of the time $t-1$ cross section and the time t realized proxies.

[22] We use five year blocks. Reported inference is qualitatively unchanged across different block length choices, including the length chosen by Politis and White's (2004) automatic selection procedure.

[23] This test compares our time t forecast of the $t+1$ realization against the forecast based on the target variable's mean estimated through time t.

in dividends, as is common in the literature. We construct portfolio-level log price-dividend ratios, forming sets of six, nine, twelve, 25 and 36 portfolios on the basis of underlying firms' market equity and book-to-market ratio. Our portfolios closely mimic the formation method of Fama and French (1993), particularly their no-look-ahead construction. However, we make two important modifications.

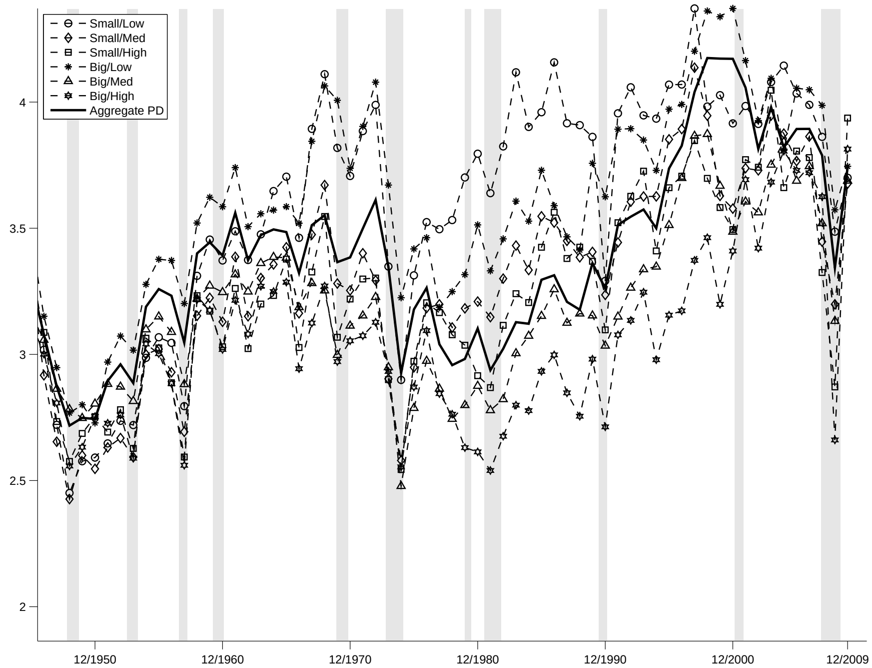
A major impediment to analysis of present value relations in the cross section is the absence of dividend payments (and hence undefined price-dividend ratios) for a substantial fraction of firms. The fraction of firms that paid dividends in 1946, 1980 and 2008 was 86%, 64% and 36%, though these fractions are substantially higher, 97%, 93% and 76%, when weighted by market capitalization. While concerns about declining numbers of dividend-paying firms are partly mitigated by portfolio aggregation, in many cases portfolios can be dominated by non-dividend payers, resulting in an erratic and highly inflated price-dividend ratio for that portfolio. To address this, we focus on dividend paying firms. When forming portfolios, we only assign a stock to a portfolio in month t if at paid positive dividends in the twelve months prior to t. This greatly increases[24] the fraction of firms in our portfolios with well-defined price-dividend ratios, while continuing to condition portfolio formation only on past publicly available information. We refer to this sample as "past dividend payers." Appendix A.E repeats our main analysis using portfolios formed from the full CRSP cross section.

Our second modification is that we rebalance portfolios every month. Fama and French rebalance once per year for a variety of reasons, such as limiting portfolio turnover. This can be important for mitigating concerns about portfolio formation transactions costs in many asset pricing settings. For our purposes, portfolios are solely a means of compiling information from the evolving cross section into a fixed set of assets that span our entire sample. By reforming portfolios each month, we limit the amount of drift in the characteristics of portfolios in to better preserve predictive links over time.

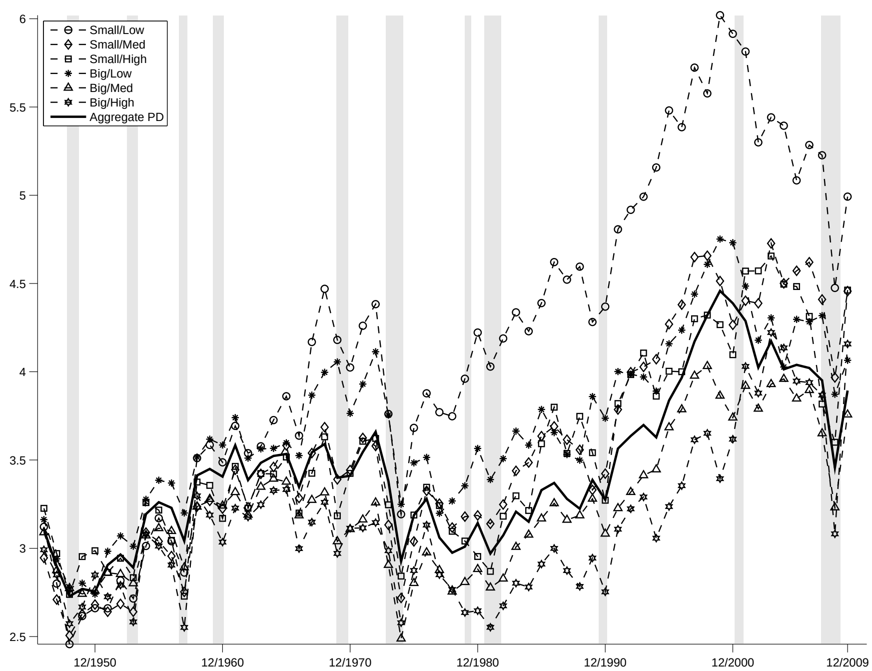
[24]Dividend paying behavior is highly persistent among US firms. Paying dividends in the past twelve months strongly predicts that a firm will pay dividends in the subsequent twelve months.

Figure 1: Portfolio Price-Dividend Ratios

(a) Past Dividend Payers



(b) All Stocks



Notes: Annual data from CRSP, 1940-2009. Log price-dividend ratios for six size and book-to-market portfolios constructed according to Section III.A, either using past dividend payers (a) or all stocks (b). Aggregate is log price-dividend ratio for the value-weighted market portfolio. NBER recession dates are represented by the shaded area.

Table II: PORTFOLIO STABILITY: ALL STOCKS VS. PAST DIVIDEND PAYERS

		Portfolio					
	Mean	1	2	3	4	5	6
B/M	1.10	1.12	1.06	1.13	1.11	1.05	1.16
ME (scaled)	1.15	1.27	1.27	1.26	0.97	1.01	1.12
ME (detrended)	1.55	1.82	1.93	1.67	0.95	1.26	1.69
ME (HP filtered)	1.50	1.52	1.65	1.58	1.16	1.22	1.86
B_r	1.97	1.31	1.29	2.80	1.07	2.05	3.26
$B_{\Delta d}$	1.79	1.59	1.19	2.31	1.01	1.95	2.67

Notes: Annual data from CRSP, 1946-2009. The tables report ratios of time series standard deviation for portfolio characteristics when portfolios are constructed using all stocks (numerator) or only using past dividend payers (denominator), thus entries represent the relative volatility of characteristics for all stock portfolios relative to dividend payer portfolios. Columns correspond to six size and book-to-market sorted portfolios. Also reported is an average of the relative volatility across portfolios for each characteristic. The first three characteristics describe variability in the total market equity of portfolios each month while controlling for level differences across portfolios. "Scaled" market equity normalizes each portfolio's monthly market equity by its time series average market equity. "Detrended" market equity removes a linear trend from each portfolio's market equity before calculating time series volatility. "HP filtered" market equity removes the non-cyclical trend component from each portfolio's market equity before calculating time series volatility. The next characteristic is the within-portfolio average book-to-market ratio. The last two characteristics are the predictive loadings of market returns (B_r) and aggregate dividend growth ($B_{\Delta d}$) from the 3PRF (using three proxies) on each portfolio price-dividend ratio. A time series for these loadings is formed using a rolling 30-year window from 1946-2009.

Figure 1 plots log price-dividend ratios for six size and book-to-market portfolios as well as the aggregate market series for the past dividend payers and all stocks samples. A prominent feature of these series is time-variation in their dispersion and ordering, which admit the possibility of interesting factor dynamics that we explore going forward.

Forming portfolios of dividend paying firms substantially improves stability in portfolio characteristics and stability in predictive relationships between individual portfolio price-dividend ratios and realizations for market returns and dividend growth. In Table II, we report a characteristic stability comparison when portfolios are composed of all stocks versus only past dividend payers. We report the ratio of time series standard deviation for each characteristics when portfolios are constructed using all stocks (numerator) or only using past dividend payers (denominator). The first three characteristics describe time series variability in the market value of portfolios. To account for the level differences in portfolio value over time and in the cross section, we consider three alternative calculations of size volatility. First, we look at "scaled" market equity, which normalize each portfolio's monthly value by its time series average market equity. We also compare market equity fluctuations after removing a linear time trend from each portfolio's market equity ("detrended"), as well as size volatility after removing the non-cyclical trend component of market equity based with the Hodrick-Prescott filter. On average, the size characteristic of all stock portfolios is between 15% and 55% more volatile then when portfolios are formed from past dividend payers. Portfolio book-to-market ratios for all stocks are on average 10% more variable. To measure stability in the predictive relationships with aggregate quantities, we calculate predictive loadings of market returns (B_r) and aggregate dividend growth ($B_{\Delta d}$) from the 3PRF (using three proxies) on each portfolio price-dividend ratio. A time series for these loadings is formed using a rolling 30-year window from 1946-2009. Volatility of predictive loadings for return and dividend growth forecasts are 97% and 79% higher for portfolios of all stocks versus past dividend payers.

There is an inherent tension between the consistency results of Section II, which use large N asymptotics, and realistic limitations of the data. Our methodology requires a fixed set of assets that maintain some constancy in their asset pricing character for the full length of sample. As in much empirical asset pricing research we achieve this by dividing the cross section into portfolios. However, increasing the number of portfolios is not fully congruous with the asymptotic theory, which assumes that signal quality does not depend on the number of signals. In reality, the underlying data is fixed, so forming more portfolios increases the noisiness of the signals at the same time it increases the number of signals because individual stock data becomes sliced ever more finely. We choose a moderate range of portfolios, between six and 36, to balance the tradeoff between signal count and signal noise.

Simultaneous two-dimension sorts á la Fama and French (1993) successfully generate portfolios with stable market value and book-to-market ratios, but can result in boundary portfolios with very few assets. Tables A4 and A3 in Appendix A.E detail characteristics of the portfolio configurations we consider. For even a five-by-five sort, some extreme size and value portfolios have very few stocks in certain months, which can result in erratic return and dividend growth behavior in those periods (and hence erratic price-dividend ratios). To limit the effect of these observations, portfolio data are treated as missing if there are four or fewer stocks in the portfolio in a given month. This data filter affects no observations for the six, nine and twelve portfolio sample, 0.36% of observations in the 25 portfolio sample, and 1.12% of observations in the 36 portfolio sample. Further details regarding our portfolio construction are provided in Appendix A.D

III.B Market Return Predictability

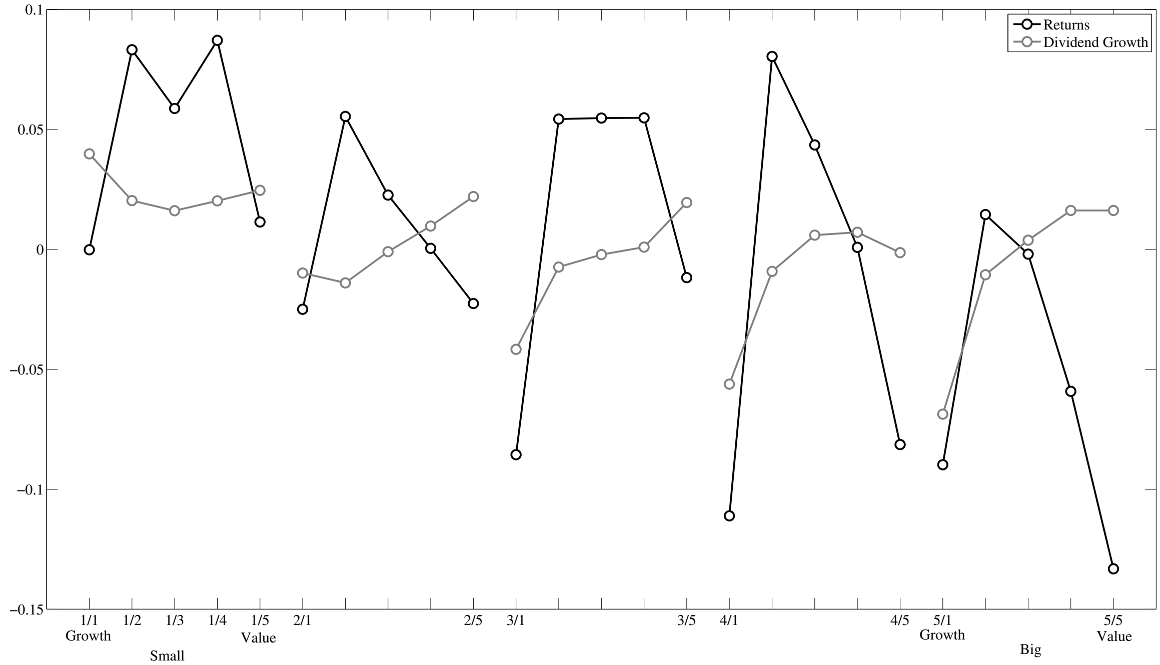
Table III presents our primary market return forecasting results based on the cross section of size and book-to-market portfolios at various levels of disaggregation. The left reports values from the three-pass filter using three latent factor proxies: One-period-ahead market returns and dividend growth and the contemporaneous aggregate price-dividend ratio. As discussed in Section II, these are natural proxies in the present value framework, and with three proxies these results are analogous to a forecasting regression with three regressors. The right panel reports results using one proxy variable - the market return itself. Section II.B explains that, when using a single proxy, the optimal choice is the forecast target itself.

The filter achieves a striking degree of predictability. The full information version (first column of each panel) generates a predictive R^2 for market returns ranging between 22.34% and 33.5% with three proxies and between 13.9% and 26.3% for a single proxy. The bootstrap p-value testing statistical significance for predictability is less than 0.001 in each case.

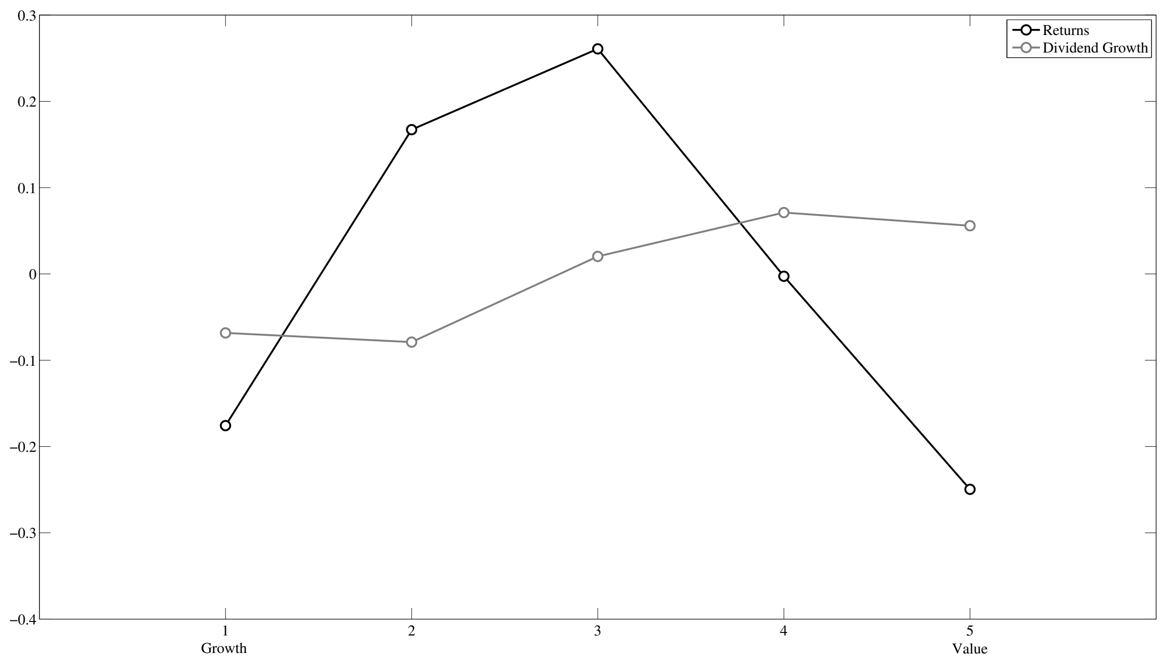
Figure 2.a plots the predictive loadings of future market returns and dividend growth

Figure 2: Predictive Loadings on Individual Price-Dividend Ratios

(a) 25 Size/Book-to-Market Portfolios



(b) 5 Book-to-Market Portfolios



Notes: Annual data from CRSP, 1946-2009. Plotted are the predictive loadings of market returns (black) and dividend growth (grey) on 25 size and book-to-market (a) and 5 book-to-market (b) portfolio price dividend ratios. Loadings are calculated as the 3PRF coefficient matrix $\boldsymbol{B}$ presented in Section II.A. The 3PRF is implemented using the realized market return, realized market dividend growth, and market price-dividend ratio as proxies.

Table III: Market Return Predictions (1946-2009)

	Three Factors			One Factor		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
			6 Portfolios			
R^2 (%)	29.94	24.60	41.08	26.25	23.09	18.16
Test	*11.30*	*10.17*	*11.28*	*10.53*	*10.75*	*3.29*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.002*	*0.002*	*< 0.01*
			9 Portfolios			
R^2 (%)	33.52	23.25	29.80	23.89	17.82	19.07
Test	*14.04*	*9.68*	*9.86*	*9.26*	*7.58*	*3.60*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.003*	*0.008*	*< 0.01*
			12 Portfolios			
R^2 (%)	28.33	20.86	22.88	24.63	17.38	8.15
Test	*12.45*	*9.59*	*5.02*	*9.33*	*7.26*	*1.18*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.003*	*0.009*	*< 0.10*
			25 Portfolios			
R^2 (%)	22.35	13.30	16.02	13.91	5.86	11.36
Test	*8.03*	*3.97*	*5.55*	*4.73*	*2.35*	*3.04*
p-val	*< 0.001*	*0.012*	*< 0.01*	*0.033*	*0.130*	*< 0.01*
			36 Portfolios			
R^2 (%)	29.35	18.37	23.02	25.54	13.17	20.22
Test	*9.34*	*6.39*	*7.54*	*9.04*	*4.81*	*4.27*
p-val	*< 0.001*	*0.001*	*< 0.01*	*0.004*	*0.032*	*< 0.01*

Notes: Annual data from CRSP, 1946-2009. The training window for out-of-sample forecasts is 34 years (out-of-sample forecasts begin in 1981). Portfolios are formed on the basis of market equity and book-to-market value using past dividend payers. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using the realized market return as sole proxy.

on each portfolio price-dividend ratio (the $\boldsymbol{B}$ coefficient in Section II.A). Dividend growth loadings are in effect straight size and value spreads, being long small value stocks and short large growth stocks. Market return forecasts are also long small stocks and short large stocks. The pattern of return loadings across book-to-market portfolios forms a hump shape, establishing a long position in stocks with moderate book-to-market ratios while being short extreme growth and value stocks. Figure 2.b shows that loadings of portfolios sorted only on book-to-market demonstrate the same patterns. A potential explanation for the hump shape across book-to-market portfolios lies in the high cash flow duration of growth stocks and the low duration of value stocks (Dechow, Sloan and Soliman 2004). The prices of assets with differing cash flow durations respond differently to persistent (but eventually mean-reverting) discount rate shocks. Consider stocks A, B and C that each pay a single dividend at year 1, 5 and 50, respectively. A positive shock to the discount rate equates with a drop in all asset prices. However, stock A's only dividend is paid in year 1, which makes it relatively insensitive to the shock. Because the discount rate is persistent, its effect aggregates over time, leading to a bigger drop in the price of stock B's year 5 dividend. Stock C, which pays a dividend only in year 50, is insulated from the shock because at very long horizons the effect of mean-reversion dominates. As the Campbell-Shiller decomposition and the pricing structure in Section I show, differential exposure to cash flow shocks is not the only driver of variation in price-dividend ratios across assets. There may be factors driving the cross section of pd_i's that are irrelevant for discount rates. The hump shape is therefore a natural combination of assets with varied durations. It ties neatly to the intuition behind the 3PRF, which forms optimal forecasts by weighting price-dividend ratios in a way that maximizes exposure to factors driving returns while netting out the effect of irrelevant factors.[25]

Results from the no-peek procedure are reported in the second column of Table III. The no-peek R^2 ranges between 13.3% and 24.6% with three proxies, and forecasting ability under this procedure remains highly statistically significant with bootstrap p-values ranging from 0.012 to below 0.001. While the no-peek procedure effectively removes the influence of small sample bias on estimated predictability, it also shortens our first-stage sample by three years, or 5% of our total sample horizon. Both of these effects ultimately decrease the no-peek R^2 relative to the full information case. Nonetheless, no-peek forecasts remain extraordinarily powerful. In the single proxy case the no-peek results are only slightly weaker.

The right column of each panel reports out-of-sample market return predictability for the three-pass filter. Our method obtains an out-of-sample R^2 ranging between 16.0% and 41.1% in the three factor case, and reaches as high as 20.2% with a single proxy. Out-of-sample

[25]Similar logic has been used in the models of Xiong and Yan (2010) and Lettau and Wachter (2010) for generating the "tent-shaped" Cochrane and Piazzesi (2005) bond premium loadings.

forecasting power is significant at the 0.010 level for all sets of portfolios according to the Clark and McCracken (2001) test. This finding is in stark contrast to out-of-sample results for much of the return predictability literature. We make comparisons with predictors from the prior literature below.

Results in Table III study the post-war period. Past literature has demonstrated much weaker predictability evidence in the sample extended back to the Great Depression (Chen 2009, Koijen and Van Nieuwerburgh 2010). In Table V we analyze this longer sample, extracting both one and three factors in-sample and out-of-sample. Predictability remains strong and highly significant in the pre-war period, particularly in the three proxy case.

How do our market return forecasts compare with predictors proposed in earlier literature? Table IV compares predictive accuracy of our approach with an extensive collection of alternative predictors that have been considered in the literature. In particular, we explore forecasts from 17 predictors studied in a recent return predictability survey by Goyal and Welch (2008) as well as the break-adjusted aggregate price-dividend ratio of Lettau and Van Nieuwerburgh (2008). The table considers in-sample (upper panel) and out-of-sample (lower panel) forecasts from each regressor individually. We also report multivariate forecast results using all predictors simultaneously, and using the best combination of three predictors (chosen from all permutations). The best univariate in-sample forecasts come from the unadjusted and break-adjusted price-dividend ratio ($R^2 = 10.8\%$ and 15.7%, respectively, with bootstrap p-values= 0.009 and 0.001). Only *cay* (Lettau and Ludvigson 2001) and the break-adjusted price-dividend ratio achieve a positive out-of-sample R^2.[26] Multivariate forecasts are stronger in-sample but demonstrate overfit symptoms. When all variables are used, the out-of-sample R^2 is -84.4%, far worse than the performance of the historical mean. The best combination of three predictors[27] gives an in-sample R^2 of 29.3%, but fails

[26] Note that neither *cay* nor the break-adjusted price-dividend ratio forecasts are purely out-of-sample. Each is subject to some degree of look-ahead bias since a two-sided moving average is used to construct *cay*, and future break dates must be selected to adjust the price-dividend ratio.

[27] Despite a negative R^2, the "Top 3" combination statistically outperforms forecasts based on the historical mean return at the 10% significance level.

Table IV: Market Return Predictions: Comparison with Common Alternative Predictors

	Alternative Predictors																				3PRF	
	dfy	infl	svar	de	lty	tms	tbl	dfr	pd	dy	ltr	ep	bm	ik	ntis	eqis	cay	pd*	*All*	*Top 3*	3 factor	1 factor
											In-Sample											
R^2 (%)	3.20	0.24	0.61	0.98	0.11	1.03	0.04	0.78	10.78	8.03	2.00	6.84	6.13	4.33	0.30	1.04	9.46	15.68	46.20	29.28	29.94	26.25
Test	*2.02*	*0.14*	*0.38*	*0.60*	*0.07*	*0.63*	*0.02*	*0.48*	*7.37*	*5.33*	*1.25*	*4.48*	*3.98*	*2.72*	*0.18*	*0.64*	*6.37*	*11.34*	*2.66*	*8.14*	*12.27*	*9.42*
p-val	*0.161*	*0.705*	*0.542*	*0.441*	*0.798*	*0.429*	*0.879*	*0.492*	*0.009*	*0.024*	*0.268*	*0.038*	*0.050*	*0.104*	*0.670*	*0.427*	*0.014*	*0.001*	*0.005*	*< 0.001*	*< 0.001*	*0.003*
											Out-of-Sample											
R^2 (%)	−4.41	−1.67	−10.84	−3.30	−4.55	−0.12	−4.17	−17.75	−2.30	−4.46	−5.76	−5.54	−6.18	−1.15	−10.59	−25.78	11.27	5.10	−84.36	−3.01	41.08	18.16
Test	*0.31*	*−0.96*	*−1.49*	*−1.13*	*−1.26*	*−0.82*	*−1.29*	*−2.19*	*1.94*	*1.09*	*−0.25*	*1.31*	*0.78*	*−0.51*	*−1.36*	*−2.30*	*1.53*	*3.92*	*−0.68*	*1.72*	*11.28*	*3.29*
p-val	–	–	–	–	–	–	–	–	*< 0.05*	*< 0.10*	–	*< 0.10*	–	–	–	–	*< 0.05*	*< 0.01*	–	*< 0.10*	*< 0.01*	*< 0.01*

Notes: Annual data from CRSP, 1946-2009. Out-of-sample forecasts begin in 1981. Alternative predictors, taken from Goyal and Welch (2008) with data updated through 2009, are the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the dividend payout ratio (de), the long term yield (lty), the term spread (tms), the T-bill rate (tbl), the default return spread (dfr), the price-dividend ratio (pd), the dividend yield (dy), the long term rate of returns (ltr), the earning price ratio (ep), the book to market ratio (bm), the investment to capital ratio (ik), the net equity expansion ratio (ntis), the percent equity issuing ratio (eqis), and the *ex post* consumption-wealth-income ratio (cay). We also include the break-adjusted price-dividend ratio (pd*) of Lettau and Van Nieuwerburgh (2008), which takes out regime-specific means for 1946-1954, 1955-1994, and 1995-2009. In-sample results report F-statistics in *italics* and associated p-values (for univariate specifications, the p-value from the F-statistic is identical to the p-value for the t-statistic). For three-pass regression filter (3PRF) results, F-statistics are bootstrapped using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing statistic in *italics*, with p-values from Clark and McCracken's (2001) critical value appendix tables (if below 10%). This tests the null hypothesis of no out-of-sample forecast improvement over the historical mean. *All* uses every alternative predictor (we drop pd in order to avoid including pd and pd* simultaneously). *Top 3* uses the best-performing in-sample trio of alternate predictors, ntis, eqis, and pd*. The 3PRF results replicate evidence from Table III based on six portfolios of past dividend payers. The column labeled "3 factor" extracts three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. The column labeled "1 factor' extracts a single factor using the realized market return as sole proxy.

Table V: Market Return Predictions (1930-2009)

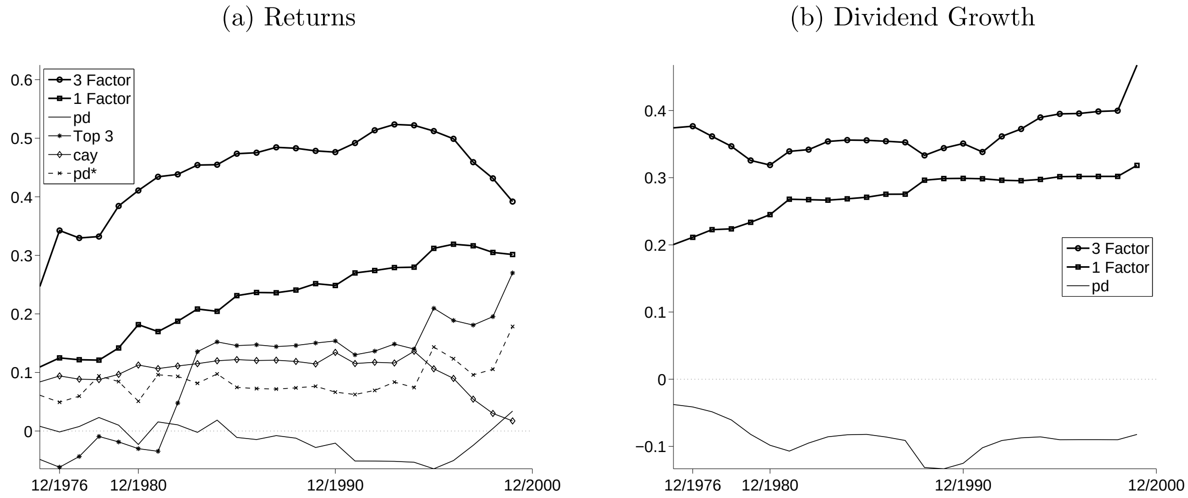
	Three Factors			One Factor		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
6 Portfolios						
R^2 (%)	15.86	11.98	19.38	13.24	1.93	11.64
Test	*5.09*	*2.87*	*5.24*	*5.39*	*0.82*	*2.01*
p-val	*0.003*	*0.042*	*< 0.01*	*0.023*	*0.367*	*< 0.05*
9 Portfolios						
R^2 (%)	16.12	5.95	15.40	11.71	0.59	9.93
Test	*5.53*	*1.76*	*3.67*	*3.81*	*0.21*	*1.31*
p-val	*0.002*	*0.161*	*< 0.05*	*0.055*	*0.648*	*< 0.10*
12 Portfolios						
R^2 (%)	19.47	8.82	12.97	17.83	1.32	7.44
Test	*9.58*	*3.65*	*4.34*	*7.38*	*0.78*	*0.83*
p-val	*< 0.001*	*0.016*	*< 0.01*	*0.008*	*0.379*	–
25 Portfolios						
R^2 (%)	19.01	4.02	9.27	18.20	1.18	4.48
Test	*8.09*	*1.39*	*2.78*	*5.12*	*0.33*	*0.80*
p-val	*< 0.001*	*0.252*	*< 0.05*	*0.026*	*0.567*	–
36 Portfolios						
R^2 (%)	24.77	9.87	13.81	25.23	6.92	2.62
Test	*7.37*	*2.94*	*5.85*	*7.36*	*3.26*	*0.69*
p-val	*< 0.001*	*0.039*	*0.01*	*0.008*	*0.075*	–

Notes: Annual data from CRSP, 1930-2009. The training window for out-of-sample forecasts is 40 years (out-of-sample forecasts begin in 1971). Portfolios are formed on the basis of market equity and book-to-market value using all stocks. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using the realized market return as sole proxy.

to produce a positive out-of-sample R^2. Our one factor forecast dominates each of the 18 alternative predictors in-sample and out-of-sample. Our three factor forecasts outperform the best combination of three alternative predictors, both in-sample out-of-sample.

The out-of-sample performance dichotomy between our predictor and standard alternatives is robust to different splits for training and test samples. To show this, Panel A of Figure 3 plots the out-of-sample R^2 sequence that results from varying the sample split year over 1976 to 2000, comparing our cross section-based predictor with various alternatives. The minimum training sample considered is 30 years, or less than half of the total sample.

Figure 3: Out-of-Sample R^2 by Sample Split Date



Notes: Annual CRSP data, 1946-2009. The out-of-sample R^2 is plotted above the year for which the in-sample training sample ends. Estimation is recursive; i.e., for every t, the forecast for $t+1$ is based on time t information and coefficients estimated using data only through time t. The lines labeled "3 factor" are based on forecasts from three extracted factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. The lines labeled "1 factor" are due to a single extracted factor using the realized market return as sole proxy in Panel A and realized dividend growth as sole proxy in Panel B. The forecast from the unadjusted and break-adjusted (Lettau and Van Nieuwerburgh 2008) aggregate price-dividend ratios are denoted pd and pd*, respectively, and the *ex post* consumption-wealth-income ratio (Lettau and Ludvigson 2001) is denoted cay. In Panel A, the "Top 3" is the best in-sample trio of the alternative predictors considered in Table IV.

The out-of-sample performance of our filter-based forecasts is uniformly superior to alternatives. From 1976 onwards, our out-of-sample R^2 is above 30% and 10% when extracting three factors and one factor, respectively.

The tension between the statistical benefits of a larger cross section at the cost of deterioration in portfolio properties, discussed in Section III.A, is evident from the performance across portfolio configurations. All results show that moving from a single, aggregate price-dividend ratio to disaggregated series dramatically boosts the ability to identify variability in conditional expectations of returns and dividend growth. Compiling stocks into a small number of portfolios averages away idiosyncrasies and produces continuous asset series with stable characteristics and predictive content throughout the sample. When the cross section is sliced more thinly, the benefits of aggregation disappear and the resulting noise makes it more difficult to extract return/growth-relevant factors from the panel.

Table VI: Market Dividend Growth Predictions (1946-2009)

	Three Factors			One Factor		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
			6 Portfolios			
R^2 (%)	40.70	33.39	31.88	9.63	9.23	8.20
Test	*26.55*	*17.47*	*10.32*	*3.52*	*4.03*	*1.08*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.065*	*0.049*	*< 0.10*
			9 Portfolios			
R^2 (%)	35.74	29.49	18.04	9.02	8.65	7.21
Test	*14.10*	*11.78*	*5.91*	*3.26*	*3.63*	*0.91*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.076*	*0.061*	–
			12 Portfolios			
R^2 (%)	30.52	22.51	10.19	12.18	13.44	8.38
Test	*12.72*	*10.86*	*3.36*	*5.02*	*7.33*	*0.76*
p-val	*< 0.001*	*< 0.001*	*< 0.05*	*0.029*	*0.009*	–
			25 Portfolios			
R^2 (%)	35.63	20.18	5.66	8.04	6.67	8.26
Test	*11.54*	*7.98*	*1.90*	*2.97*	*2.72*	*1.09*
p-val	*< 0.001*	*< 0.001*	*< 0.10*	*0.090*	*0.104*	*< 0.10*
			36 Portfolios			
R^2 (%)	40.86	20.48	10.72	11.13	11.87	9.27
Test	*12.53*	*8.81*	*1.97*	*4.45*	*4.80*	*1.02*
p-val	*< 0.001*	*< 0.001*	*< 0.10*	*0.039*	*0.032*	*< 0.10*

Notes: Annual data from CRSP, 1946-2009. The training window for out-of-sample forecasts is 34 years (out-of-sample forecasts begin in 1981). Portfolios are formed on the basis of market equity and book-to-market value using past dividend payers. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using realized dividend growth as sole proxy.

III.C Dividend Growth Predictability

We next evaluate aggregate dividend growth predictability using the three-pass regression filter. Table VI presents our dividend forecast results for the 1946-2009 sample based on three extracted factors. Using three proxies, annual dividend growth is forecastable with an in-sample R^2 ranging from 30.5% to 40.9%, and out-of-sample R^2 reaching 31.9%. Dividend forecastability using a single factor is weaker, but remains statistically significant out-of-sample. Recently, van Binsbergen and Koijen provided evidence of in-sample forecastability of dividend growth in the post-war sample, which our method strongly identifies. The

Table VII: Market Dividend Growth Predictions (1930-2009)

	Three Factors			One Factor		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
			6 Portfolios			
R^2 (%)	15.40	6.41	−23.31	5.20	4.28	7.72
Test	*4.91*	*2.30*	–	*3.32*	*3.34*	*5.23*
p-val	*0.004*	*0.084*	–	*0.072*	*0.071*	< *0.01*
			9 Portfolios			
R^2 (%)	21.80	9.60	−25.34	6.32	6.09	6.13
Test	*7.15*	*2.60*	–	*3.49*	*3.99*	*5.24*
p-val	< *0.001*	*0.058*	–	*0.066*	*0.049*	< *0.01*
			12 Portfolios			
R^2 (%)	17.42	7.87	−112.49	5.66	10.57	5.22
Test	*7.28*	*5.13*	–	*3.31*	*7.57*	*4.40*
p-val	< *0.001*	*0.003*	–	*0.073*	*0.007*	< *0.01*
			25 Portfolios			
R^2 (%)	19.64	8.19	−138.80	2.41	4.91	2.28
Test	*5.36*	*4.58*	–	*1.15*	*2.26*	*2.49*
p-val	*0.002*	*0.005*	–	*0.287*	*0.136*	< *0.05*
			36 Portfolios			
R^2 (%)	19.54	7.54	−79.98	0.04	0.02	−0.77
Test	*4.84*	*2.30*	–	*0.02*	*0.02*	–
p-val	*0.004*	*0.084*	–	*0.896*	*0.898*	–

Notes: Annual data from CRSP, 1930-2009. The training window for out-of-sample forecasts is 40 years (out-of-sample forecasts begin in 1971). Portfolios are formed on the basis of market equity and book-to-market value using past dividend payers. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using realized dividend growth as sole proxy.

surprising finding of Table VI is the out-of-sample dividend growth predictibility. Unlike market return forecasts, however, Table VII shows that evidence of out-of-sample dividend growth predictability is absent in the longer 1930-2009 sample. Previous literature has documented large differences in dividend growth predictability in the pre- and post-war samples (Chen 2009, Koijen and Van Nieuwerburgh 2010), and our findings suggest the same.

III.D State Space Representation

The central impetus for this study is the inclusion of cross section information in a present value framework to better discern market expectations. Furthermore, our development of the three-pass filter is largely motivated by the restrictiveness of fully specifying a state space model and the computational difficulty in estimating large cross section systems using standard filter techniques. Nonetheless it is instructive to understand the relationship between the results from our procedure and the state space approach. To this end, we examine Kalman filter estimates for latent return and dividend growth expectations using a relatively small and computationally tractable present value system. We present details regarding the state space representation of our system in Appendix A.C. Transition equations are described by a first order vector autoregression for latent factors $\boldsymbol{F}_t$. Realized returns, realized dividend growth and the cross section of price-dividend ratios constitute the observable measurement equations. To simplify computation, we use the set of six portfolios (past dividend payers) from our preceding analysis. To estimate our state space, we numerically maximize the likelihood with a simulated annealing algorithm.[28] A summary of implied expected market returns and dividend growth compared with the three-pass filter and the aggregate market state space system of van Binsbergen and Koijen is reported in Table VIII and Figure A1.[29] Our cross-sectional present value state space produces a predictive R^2 of 16.6% for market returns and 20.8% for aggregate dividend growth. The implied volatility of expected returns is 9.0%, and their implied persistence is 78.4% (first order autocorrelation). The fitted expected return and expected dividend growth series from the Kalman filter and three-pass regression filter show close similarity in the alternative approaches estimates for expectations.[30] The correlation between expected return fits from the two approaches is 60.0%, and the correlation between the two expected dividend growth series is 73.2%. This

[28] Detail on the 51 parameter estimates and their standard errors are available from the authors upon request.

[29] We replicate van Binsbergen and Koijen's estimation for the cash-reinvested dividends case using data for 1946-2009. Our estimates are very similar to their findings, which are based on data from 1946-2007.

[30] See in Figure A1.

Table VIII: Summary Statistics: Expected Market Return and Dividend Growth Estimates

	Realized	Expected 3PRF IS	Expected 3PRF OOS	Expected Cross Section PV State Space	Expected Aggregate PD	Expected Aggregate PV State Space
				Market Returns		
Mean (%)	10.24	10.23	15.17	10.56	10.23	10.08
Standard Deviation (%)	17.07	9.34	8.67	9.02	5.61	4.32
Autocorrelation	−0.081	0.618	0.476	0.784	0.925	0.927
Predictive R^2 (%)	–	29.94	41.08	16.55	10.78	9.63
				Market Dividend Growth		
Mean (%)	5.51	5.51	5.41	5.55	5.51	5.56
Standard Deviation (%)	6.90	4.40	4.03	3.25	0.77	3.09
Autocorrelation	0.340	0.569	0.642	0.748	0.925	0.181
Predictive R^2 (%)	–	40.70	31.88	20.79	1.23	18.34

Notes: Summary statistics of realized market dividend growth and one-year-ahead expected market dividend growth estimated by various methods. The columns labeled"3PRF" use six size and book-to-market portfolios constructed using past dividend payers, either in-sample (IS) or out-of-sample (OOS). The "Cross Section PV State Space" is presented in Appendix A.C. "Aggregate PD" is the predictive regression using the aggregate price-dividend ratio. "Aggregate PV State Space" is a reproduction of van Binsbergen and Koijen (2010).

analysis highlights the fact that cross section information is beneficial to the latent variable estimation problem independent of our benchmark method. That it is difficult to apply the Kalman filter for larger numbers of observation equations in turn highlights the function of the three-pass filter: It allows us to exploit a finer cross-sectional information set than may be feasible with traditional methods. The fact that the detected predictive power is slightly weaker than our full information case is consistent with the fact that our Kalman estimates impose a complete specification on the factors driving price-dividend ratios (one that is likely to be violated in the data) while the three-pass filter does not.

III.E Discussion

Our estimated one-year-ahead expected return process differs from past research in several qualitatively important ways. Table VIII presents the mean, volatility, persistence and predictive accuracy of five alternative estimates one-year return expectations. First, we find that expected returns are much more volatile than previously estimated. The volatility of one-year-ahead expected returns is two-thirds higher than estimates based on standard aggregate price-dividend ratio predictive regressions (annual volatility of 9.3% versus 5.6%). It is more than double the expected return volatility estimated from the aggregate present value state space (4.3% per year). Second, one-year-ahead expected returns mean revert more quickly than previously believed. The first order autocorrelation is 61.8% at the annual frequency, compared to standard aggregate regression and state space estimates of over 90%. Higher volatility and lower persistence of our one-year-ahead expected returns contribute to forecasts of realized returns that are nearly three times more accurate than aggregate price-dividend-based forecasts. These features of the expected return process also imply that the persistent, low frequency fluctuations that drive the majority of price-dividend ratio variation must be coming from persistent expectations of returns (or dividend growth) at horizons beyond one year. Thus, our findings suggest that there is rich structure in one-year expected returns over various horizons. Recent research has modeled one-year expectations with first-order autoregressions,[31] which imply in the present value context that 100% of variation in the aggregate price-dividend ratio is driven by next year's expected market return and dividend growth rate. Comparing our fitted one-year-ahead expected aggregate return and dividend growth estimates to the aggregate price-dividend ratio suggests that these short horizon expectations explain less than 20% of total price-dividend ratio variation, and are primarily responsible for its high frequency fluctuations. Incorporating one-year expectations at more distant horizons is necessary to describe the low frequency, persistent movements in the price-dividend ratio.

Figure 4 demonstrates this point with plots of three series. The first is the aggregate price-dividend ratio (black line) and the second (light grey dashed line) is the portion of the price-dividend ratio explained by our estimates of one-year-ahead expectations ($\hat{\mu}_t = \hat{\mathbb{E}}_t[r_{t+1}]$

[31] See, for example, Cochrane (2008a), Pástor and Stambaugh (2009) and van Binsbergen and Koijen (2010).

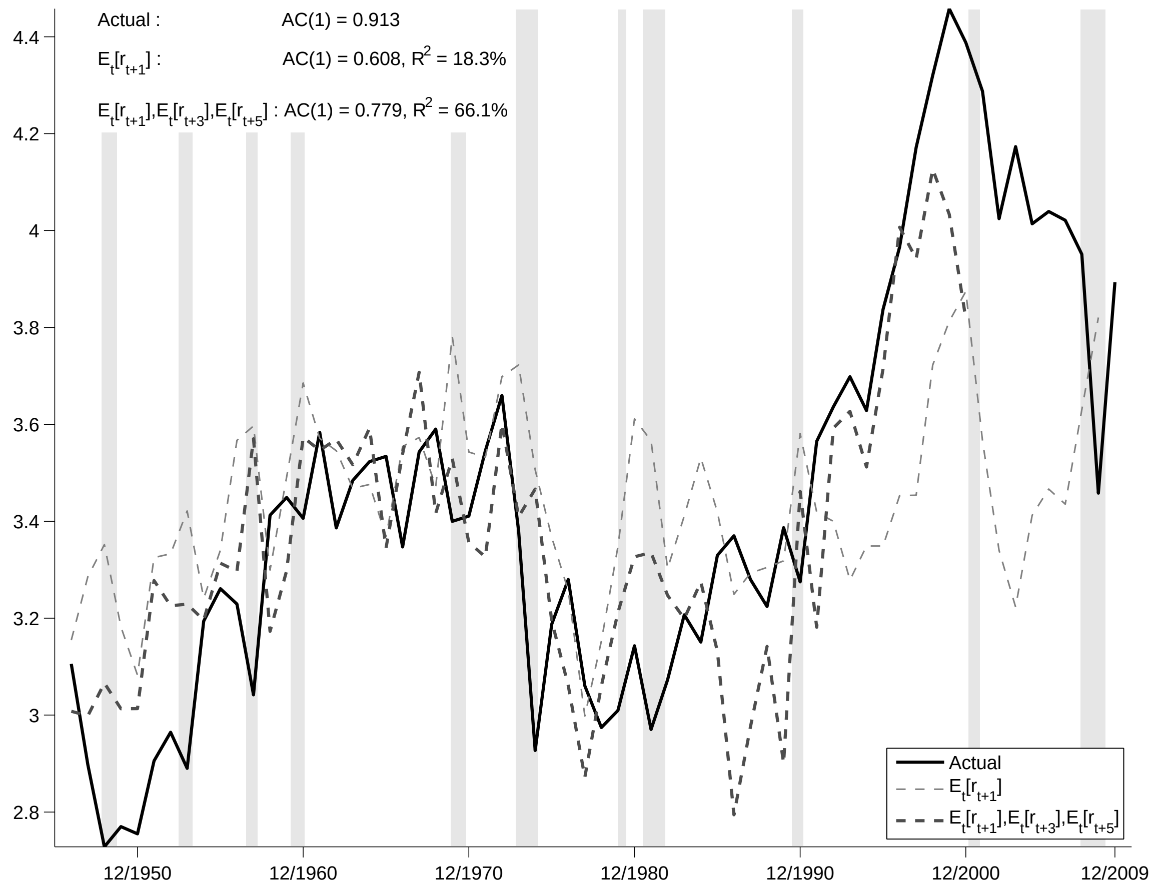
and $\hat{g}_t = \hat{\mathbb{E}}_t[\Delta d_{t+1}]$). These account for approximately 18% of the price-dividend ratio's total variance. As the figure shows, short term expectations are related to high frequency deviations from a slow moving curve. The third series (heavy grey dashed line) shows the fraction of the price-dividend ratio that can be explained when the three-pass filter is used to simultaneously estimate expectations at horizons one, three and five years ahead ($\hat{\mathbb{E}}_t[r_{t+1}], \hat{\mathbb{E}}_t[r_{t+3}]$ and $\hat{\mathbb{E}}_t[r_{t+5}]$). These longer term expectations contribute to persistent movements in the price-dividend ratio and increase its total explained variation to 66%.[32]

Allowing for a "horizon structure" in one-year expectations like that in Equation 3 breaks the tight link between short term expectations and the price-dividend ratio implied by an AR(1) model. Indeed, our analysis rejects the rigidity of models that imply valuation ratios such as those in Equation 2 ($pd_t = a - b_\mu \mu_t + b_g g_t$) in favor of a representation in the spirit of Equation 7 ($pd_t = \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \left\{-\mathbb{E}_t[\mu_{t+j}] + \mathbb{E}_t[g_{t+j}]\right\}$). In other words, our estimates of one-year-ahead expected returns and dividend growth represent an economic history that cannot be generated by a simple autoregression in expectations. Nor do they match predictions from the aggregate price-dividend ratio alone. Figure 5 plots realized market returns alongside our expected return estimates and estimates from the aggregate price-dividend ratio. The correlation between our estimates and estimates from the aggregate price-dividend ratio regression is only 3.0%. Our estimates of μ_t substantially diverge from OLS estimates during important economic episodes such as the high uncertainty of the 1970's following the oil crisis and the technology boom of the late 1990's. Amid the tech boom, OLS-based expectations fell gradually beginning in 1996, bottoming out in 2000 and climbing through 2001 and 2002. In contrast, three-pass filter forecasts show that discount rates remained high after 1996 and only began to fall in 2000.

Our method predicted negative one-year-ahead expected returns four times after World War II, each corresponding to NBER recession dates: 1970, 1973-1975 (oil crisis), 2001

[32] To simultaneously estimate multiple horizon expectations, we use the three-pass filter with 25 portfolios and a total of seven proxies. The proxies are the aggregate price-dividend ratio along with realized returns and dividend growth one, three and five years ahead. Using the second-stage filter output, we forecast realized returns and growth at the stated horizons.

Figure 4: Aggregate Price-Dividend Ratio: Actual and Explained

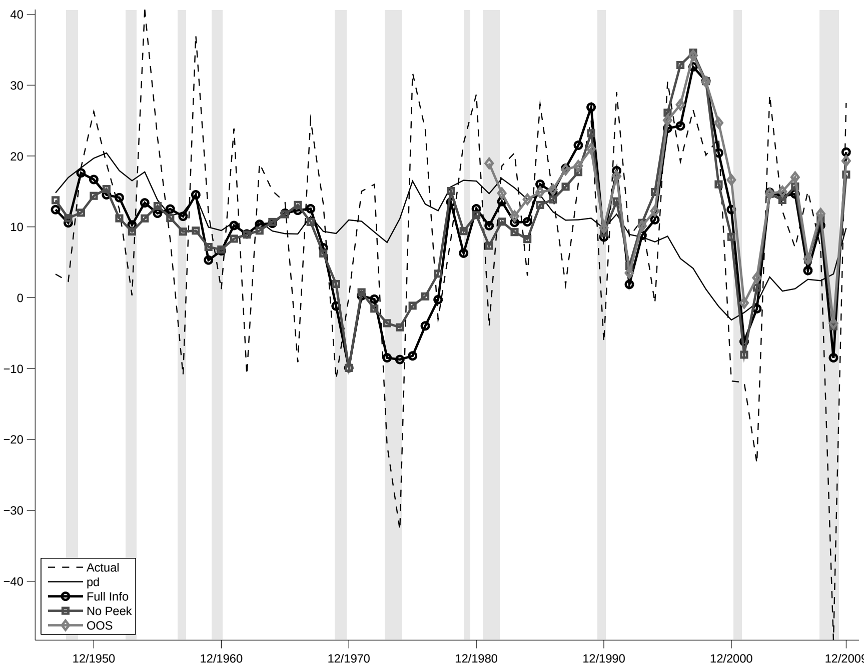


Notes: CRSP data, 1946-2009. "Actual" is the aggregate price-dividend ratio. $E_t[r_{t+1}]$ gives the price-dividend ratio explained only one-year ahead expectations (including $E_t[\Delta d_{t+1}]$, the legend lists only returns is used for notational convenience). $E_t[r_{t+1}], E_t[r_{t+3}], E_t[r_{t+5}]$ gives the price-dividend ratio explain by one-, three-, and five-year ahead expectations (including $E_t[\Delta d_{t+1}], E_t[\Delta d_{t+3}], E_t[\Delta d_{t+5}]$, the legend lists only returns is used for notational convenience).

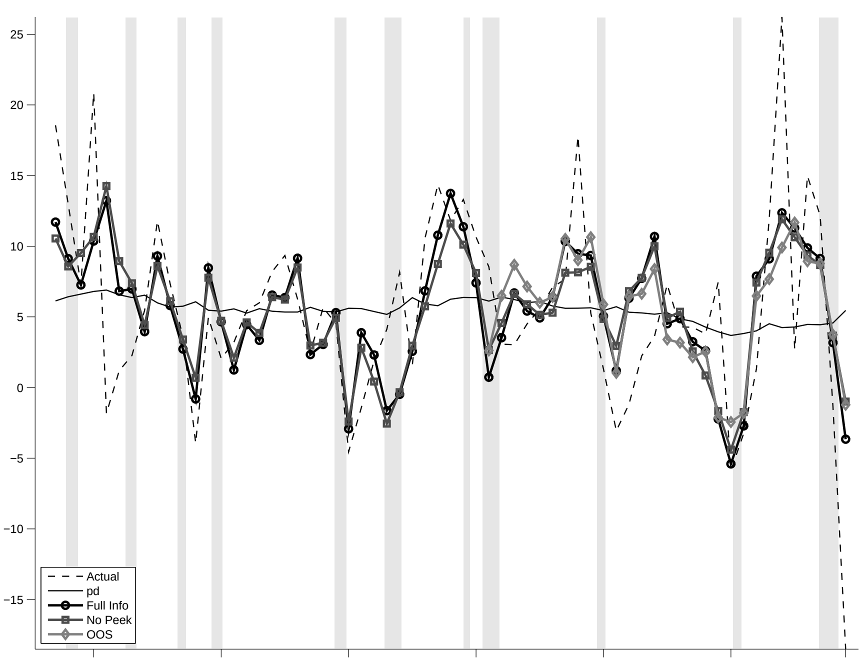
(technology-led market correction) and 2007-2009 (financial crisis). Realized returns during these four episodes also corresponded to the four most severe annual market declines in the post-war era. Two of these dates, the tech crash and the financial crisis, fall within our out-of-sample period, providing an interesting out-of-sample test of whether negative expected returns are due to in-sample overfit. On the contrary, the figure shows that in-sample and out-of-sample forecasts are closely synchronized. In both 2001 and 2008 (and *only* in those years), cross-sectional price information generated *out-of-sample* expected returns that were indeed negative.

Figure 5: Actual and Expected Market Returns and Dividend Growth

(a) Returns

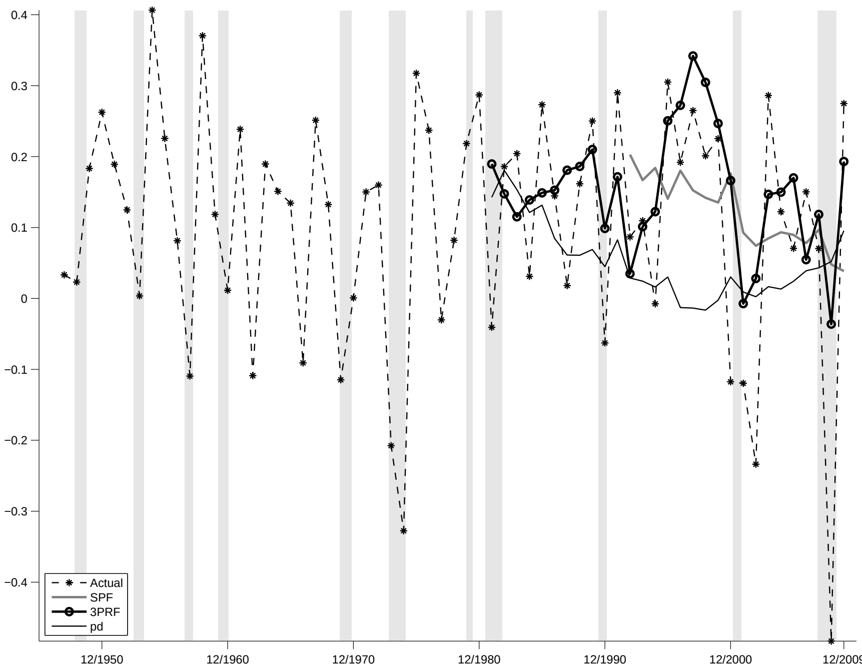


(b) Dividend Growth



Notes: Annual CRSP data, 1946-2009. Vertical axis is in percent. Expectations are aligned with the realizations they forecast; that is, a forecast $\hat{\mu}_t$ of r_{t+1} is plotted at year $t+1$ in the figure. Realizations are labeled "Actual." Expectations conditioned on the aggregate price-dividend ratio are labeled "pd." Expectations given by the 3PRF using six size and book-to-market portfolios of dividend payers are labeled "Full Info," "No Peek," and "OOS" for the three implementation versions, using three factors in all cases. NBER recession dates are represented by the shaded area.

Figure 6: Survey of Professional Forecasters Expected Market Returns

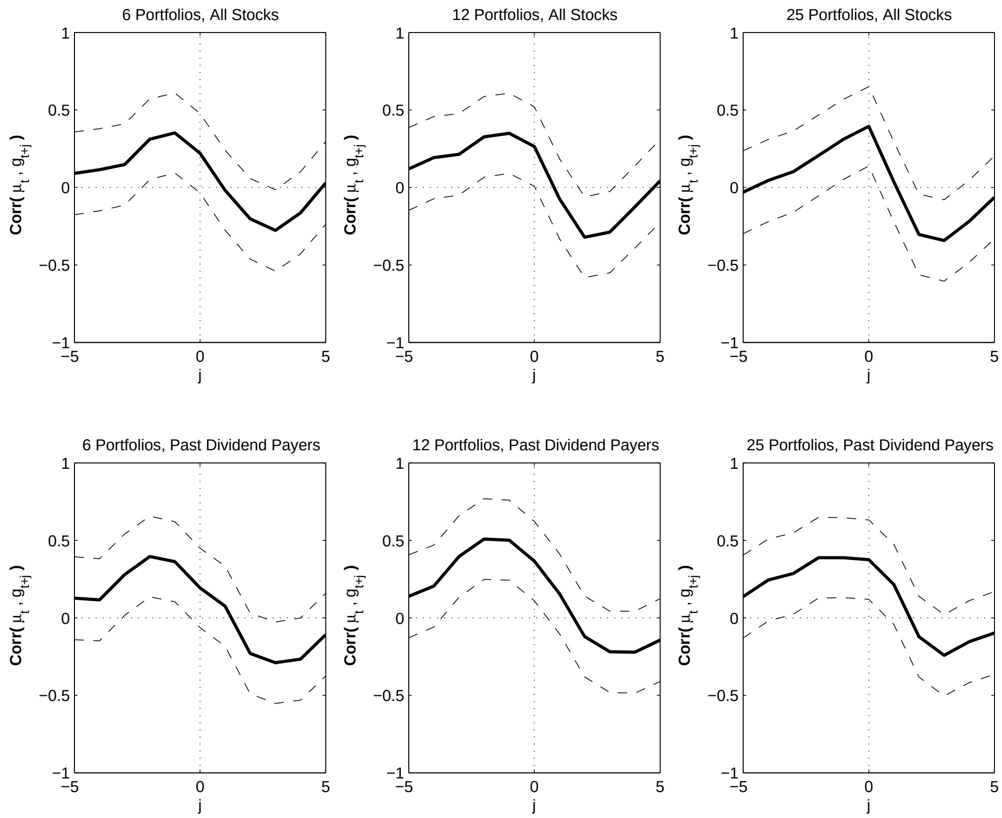


Notes: Annual CRSP data, 1946-2009. Vertical axis is in percent. Expectations are aligned with the realizations they forecast; that is, a forecast $\hat{\mu}_t$ of r_{t+1} is plotted at year $t+1$ in the figure. Realized market returns are labeled "Actual." Expectations coming from the Survey of Professional Forecasters's Stock Return Forecasts are labeled "SPF," available since 1991. "3PRF" labels expectations estimated by the 3PRF using six size and book-to-market portfolios of dividend payers, out-of-sample. Expectations given by an out-of-sample predictive regression on the aggregate price-dividend ratio are labeled "pd." NBER recession dates are represented by the shaded area.

Other than relying on forecasting power to evaluate the quality of our estimates, is there a way to compare one-year-ahead expected return estimates to expectations held by the market in real time? The *Survey of Professional Forecasters* maintains a survey database of forecasts for S&P 500 returns made by economists, finance professionals and other forecasters beginning in 1992.[33] Figure 6 plots the history of expected returns based on professionals' real time forecasts for the market return against our estimated series. The correlation of our in-sample and out-of-sample estimates with the mean survey forecast is 43.7% and 35.8%, respectively. The correlation between in-sample and out-of-sample aggregate price-dividend

[33]The survey is taken in January of each year. We line these up with our forecasts based on data one month earlier. Survey responses forecast the average rate of return on the S&P 500 index over the next ten years. For heuristic purposes, we translate this response to an implied one year forecast assuming that expected returns obey a first-order autoregression with unconditional mean equal to the full-sample average return. This transformation has no effect on the series' correlation with other time series.

Figure 7: Cross-Correlations of Expected Market Returns and Dividend Growth



Notes: Annual data from CRSP, 1946-2009. Cross correlations, holding μ constant and varying the lag/lead of g: left side gives correlation of μ with lagged values of g, right side gives correlation of μ with leads of g. Uses the realized market return, realized market dividend growth, and aggregate price-dividend ratio as proxies. Two standard error bands from normality assumption.

OLS predictions and survey expectations is 0.3% and -49.2%.

Another implication of our cross-sectional approach is that we can independently extract estimates of expected returns and dividend growth that are not mechanically correlated. This allows us to provide new estimates of dependence between short term return and growth expectations. Figure 7 plots estimates of unconditional cross-correlations between one-year-ahead expected returns and dividend growth at multiple leads and lags. The contemporaneous correlations between the series are between 20% and 40% and are only marginally significant. However, expected returns are positively correlated with lagged expected dividend growth and negatively correlated with future expected dividend growth. Serial correlations are larger in magnitude than the contemporaneous correlation and appear statistically significant. This structure poses an interesting challenge to fundamentals-based models of discount rates.

IV Conclusion

We use a general and tractable dynamic latent factor model to express assets' price-dividend ratios as a function of economic fundamentals. The same factors that drive price ratios also determine aggregate expectations of market returns and dividend growth, enabling us to use rich cross-sectional information in estimating latent market expectations. To estimate these latent processes, we adapt a new, easily implemented filtering procedure. By extracting information from the cross section of price-dividend ratios we are able to construct remarkably accurate forecasts of returns and dividend growth rates both in-sample and out-of-sample. The resulting estimates reveal several important facets of the time series of expected returns and dividend growth that may be used to guide future models of cash flows and discount rates.

References

Ang, Andrew, and Geert Bekaert. 2007. "Stock Return Predictability: Is it There?" *Review of Financial Studies*, 20(3): 651.

Bansal, Ravi, and S. Viswanathan. 1993. "No Arbitrage and Arbitrage Pricing: A New Approach." *Journal of Finance*, 48(4): 1231–1262.

Bansal, Ravi, Robert Dittmar, and Christian Lundblad. 2005. "Consumption, Dividends, and the Cross Section of Equity Returns." *Journal of Finance*, 60(4): 1639–1672.

Brennan, Michael, and Yihong Xia. 2005. "Persistence, Predictability, and Portfolio Planning." *Working Paper, Rodney L. White Center.*

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

Campbell, John, and Samuel Thompson. 2008. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21(4): 1509–1531.

Chen, Long. 2009. "On the Reversal of Return and Dividend Growth Predictability: A Tale of Two Periods." *Journal of Financial Economics*, 92(1): 128–151.

Chowdhry, Bhagwan, Richard Roll, and Yihong Xia. 2005. "Extracting Inflation from Stock Returns to Test Purchasing Power Parity." *American Economic Review*, 95(1): 255–276.

Clark, Todd, and Michael McCracken. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, 105(1): 85–110.

Cochrane, John. 2005. *Asset Pricing.* Princeton University Press.

Cochrane, John. 2008*a*. "State-Space vs. VAR Models for Stock Returns." *Working Paper, University of Chicago.*

Cochrane, John. 2008*b*. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies*, 21(4): 1533.

Cochrane, John. 2011. "Discount Rates." *Working Paper, University of Chicago.*

Cochrane, John, and Monika Piazzesi. 2005. "Bond Risk Premia." *American Economic Review*, 95(1): 138–160.

Connor, Gregory. 1984. "A Unified Beta Pricing Theory." *Journal of Economic Theory*, 34(1): 13–31.

Davis, James, Eugene Fama, and Kenneth French. 2000. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *Journal of Finance*, 55(1): 389–406.

Dechow, Patricia, Richard Sloan, and Mark Soliman. 2004. "Implied Equity Duration: A New Neasure of Equity Risk." *Review of Accounting Studies*, 9(2): 197–228.

Fama, Eugene, and James MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *The Journal of Political Economy*, 81(3): 607–636.

Fama, Eugene, and Kenneth French. 1988. "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, 96(2): 246–73.

Fama, Eugene, and Kenneth French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33(1): 3 – 56.

Ferreira, Miguel, and Pedro Santa Clara. 2010. "Forecasting Stock Market Returns: The Sum of the Parts Is More Than the Whole." *Journal of Financial Economics*, forthcoming.

Gabaix, Xavier. 2010. "Linearity-Generating Processes: A Modelling Tool Yielding Closed Forms for Asset Prices." *Working Paper, NYU.*

Ghosh, Anisha, and George Constantinides. 2010. "The Predictability of Returns with Regime Shifts in Consumption and Dividend Growth." *Working Paper, University of Chicago.*

Goetzmann, William, and Philippe Jorion. 1993. "Testing the Predictive Power of Dividend Yields." *Journal of Finance*, 48(2): 663–679.

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

Hansen, Lars, John Heaton, and Nancy Li. 2008. "Consumption Strikes Back? Measuring Long-Run Risk." *Journal of Political Economy*, 116(2).

Hodrick, Robert. 1992. "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement." *Review of Financial studies*, 5(3): 357.

Jagannathan, Ravi, and Zhenyu Wang. 1998. "A Note on the Asymptotic Covariance in Fama-MacBeth Regression." *Journal of Finance*, 53(2): 799–801.

Kalman, Rudolf. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering*, 82(1): 35–45.

Kelly, Bryan. 2010. "Risk Premia and the Conditional Tails of Stock Returns." *Working Paper, Chicago Booth.*

Kelly, Bryan, and Seth Pruitt. 2011. "The Three-Pass Regression Filter: A New Approach to Forecasting with Many Predictors." *Working Paper, University of Chicago.*

Kiku, Dana. 2006. "Is the Value Premium a Puzzle?" *Working Paper, The Wharton School.*

Koijen, Ralph, and Stijn Van Nieuwerburgh. 2010. "Predictability of Returns and Cash Flows." *Working Paper, University of Chicago.*

Kothari, S.P., and Jay Shanken. 1997. "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." *Journal of Financial Economics*, 44(2): 169–203.

Lettau, Martin, and Jessica Wachter. Forthcoming. "The Term Structures of Equity and Interest Rates." *Journal of Financial Economics.*

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

Lettau, Martin, and Sydney Ludvigson. 2005. "Expected Returns and Expected Dividend Growth." *Journal of Financial Economics*, 76(3): 583–626.

Lewellen, Jonathan. 2004. "Predicting Returns with Financial Ratios." *Journal of Financial Economics*, 74(2): 209–235.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *The Review of Economics and Statistics*, 47(1): 13–37.

Menzly, Lior, Tano Santos, and Pietro Veronesi. 2004. "Understanding Predictability." *Journal of Political Economy*, 112(1): 1–47.

Merton, Robert. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, 41(5): 867–887.

Nelson, Charles, and Myung Kim. 1993. "Predictable Stock Returns: The Role of Small Sample Bias." *Journal of Finance*, 48(2): 641–661.

Pástor, Ľuboš, and Pietro Veronesi. 2003. "Stock Valuation and Learning about Profitability." *Journal of Finance*, 58(5): 1749–1790.

Pástor, Ľuboš, and Pietro Veronesi. 2006. "Was There a Nasdaq Bubble in the Late 1990s?" *Journal of Financial Economics*, 81(1): 61–100.

Pástor, Ľuboš, and Robert Stambaugh. 2009. "Predictive Systems: Living with Imperfect Predictors." *Journal of Finance*, 64(4): 1583–1628.

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

Paye, Bradley, and Allan Timmermann. 2006. "Instability of Return Prediction Models." *Journal of Empirical Finance*, 13(3): 274–315.

Politis, Dimitris, and Halbert White. 2004. "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews*, 23(1): 53–70.

Politis, Dimitris, and Joseph Romano. 1992. "A Circular Block-Resampling Procedure for Stationary Data." *Exploring the Limits of Bootstrap*, 263–270.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho. 2006. "Cross-Sectional Forecasts of the Equity Premium." *Journal of Financial Economics*, 81(1): 101–141.

Rozeff, Michael. 1984. "Dividend Yields Are Equity Risk Premiums." *Journal of Portfolio Management*, 11(1): 68–75.

Rytchkov, Oleg. 2008. "Filtering Out Expected Dividends and Expected Returns." *Working Paper, Temple University.*

Shanken, Jay. 1992. "The Current State of the Arbitrage Pricing Theory." *Journal of Finance*, 47(4): 1569–74.

Sharpe, William. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance*, 19(3): 425–442.

Stambaugh, Robert. 1986. "Bias in Regressions with Lagged Stochastic Regressors." *Working Paper, University of Chicago.*

Stambaugh, Robert. 1999. "Predictive Regressions." *Journal of Financial Economics*, 54(3): 375–421.

Stock, James H., and Mark W. Watson. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Treynor, Jack. 1961. "Toward a Theory of Market Value of Risky Assets." *Unpublished manuscript.*

van Binsbergen, Jules, and Ralph Koijen. 2010. "Predictive Regressions: A Present-Value Approach." *Journal of Finance*, forthcoming.

Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance*, 57(1): 233–264.

Xiong, Wei, and Hongjun Yan. 2010. "Heterogeneous expectations and bond markets." *Review of Financial Studies*, 23(4): 1433.

A Appendix

A.A Derivation of Present Value System

$$
\begin{aligned}
pd_{i,t} &= \kappa_i + \rho_i pd_{i,t+1} - r_{i,t+1} + \Delta d_{i,t+1} \\
&= \frac{\kappa_i}{1-\rho_i} + \sum_{j=0}^{\infty} \rho_i^j (-r_{i,t+j+1} + \Delta d_{i,t+j+1}) \\
&= \frac{\kappa_i}{1-\rho_i} + \sum_{j=0}^{\infty} \rho_i^j \mathbb{E}_t(-\mu_{i,t+j} + g_{i,t+j}) \\
&= \frac{\kappa_i}{1-\rho_i} + \sum_{j=0}^{\infty} \rho_i^j \mathbb{E}_t \left[-(\gamma_{i,0} + \boldsymbol{\gamma}_i' \boldsymbol{F}_{t+j}) + (\delta_{i,0} + \boldsymbol{\delta}_i' \boldsymbol{F}_{t+j}) + \varepsilon_{i,t+j} \right] \\
&= \frac{\kappa_i - \gamma_{i,0} + \delta_{i,0}}{1-\rho_i} + \sum_{j=0}^{\infty} \rho_i^j \left(\mathbb{E}_t[\boldsymbol{\iota}' \boldsymbol{\Upsilon}_i' \boldsymbol{F}_{t+j} + \varepsilon_{i,t+j}] \right) \\
&= \frac{\kappa_i - \gamma_{i,0} + \delta_{i,0}}{1-\rho_i} + \sum_{j=0}^{\infty} \rho_i^j \boldsymbol{\iota} \boldsymbol{\Upsilon}_i' \boldsymbol{\Lambda}_1^j \boldsymbol{F}_t + \varepsilon_{i,t} \\
&= \frac{1}{1-\rho_i} (\kappa_i - \gamma_{i,0} + \delta_{i,0}) + \boldsymbol{\iota}' \boldsymbol{\Upsilon}_i' \left(I - \rho_i \boldsymbol{\Lambda}_1 \right)^{-1} \boldsymbol{F}_t + \varepsilon_{i,t} \\
&= \phi_{i,0} + \boldsymbol{\phi}_i' \boldsymbol{F}_t + \varepsilon_{i,t}
\end{aligned}
$$

where we have defined $\boldsymbol{\Upsilon}_i = (\boldsymbol{\gamma}_i, \boldsymbol{\delta}_i)$, $\boldsymbol{\iota} = (-1, 1)'$, $\boldsymbol{\phi}'_i = \boldsymbol{\iota}'\boldsymbol{\Upsilon}'_i (I - \rho_i \boldsymbol{\Lambda}_1)^{-1}$, and $\phi_{i,0} = \frac{1}{1-\rho_i}(\kappa_i - \gamma_{i,0} + \delta_{i,0})$.

A.B Three Pass Regression Filter Assumptions

Assumption 1 (Factor Structure)**.** *The data are generated as:*

$$\boldsymbol{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \qquad y_{t+1} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t + \eta_{t+1} \qquad \boldsymbol{z}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{\omega}_t$$

$$\boldsymbol{X} = \boldsymbol{\iota}\boldsymbol{\phi}'_0 + \boldsymbol{F}\boldsymbol{\Phi}' + \boldsymbol{\varepsilon} \qquad \boldsymbol{y} = \boldsymbol{\iota}\beta_0 + \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\eta} \qquad \boldsymbol{Z} = \boldsymbol{\iota}\boldsymbol{\lambda}'_0 + \boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\omega}.$$

Assumption 2 (Factors and Loadings)**.** *For $\tilde{\boldsymbol{\phi}}_i \equiv (\phi_{i,0}, \boldsymbol{\phi}'_i)'$ and $\bar{\phi} < \infty$*

1. *$T^{-1}\sum_{t=1}^T \boldsymbol{F}_t \xrightarrow[T\to\infty]{p} \boldsymbol{\mu}$ and $T^{-1}\sum_{t=1}^T \boldsymbol{F}_t \boldsymbol{J}_T \boldsymbol{F}'_t \xrightarrow[T\to\infty]{p} \boldsymbol{\Delta}_F$.*
2. *$|\tilde{\boldsymbol{\phi}}_i| \leq \bar{\phi}\ \forall i$.*
3. *$N^{-1}\sum_{i=1}^N \tilde{\boldsymbol{\phi}}_i \boldsymbol{J}_N \tilde{\boldsymbol{\phi}}'_i \xrightarrow[N\to\infty]{p} \begin{bmatrix} B_0 & \boldsymbol{B}'_1 \\ \boldsymbol{B}_1 & \boldsymbol{\mathcal{B}} \end{bmatrix}$ with $\boldsymbol{\mathcal{B}}$ nonsingular.*

Assumption 3 (Error Moments)**.** *There exists a constant $A < \infty$ such that*

1. *$T^{-1}\sum_{t=1}^T \varepsilon_{i,t} \xrightarrow[T\to\infty]{p} \boldsymbol{0}\ \forall i$ and $N^{-1}\sum_{i=1}^N \varepsilon_{i,t} \xrightarrow[N\to\infty]{p} \boldsymbol{0}\ \forall t$*
2. *$T^{-1}\sum_{t=1}^T \boldsymbol{\omega}_t \xrightarrow[T\to\infty]{p} \boldsymbol{0}$ and $T^{-1}\sum_{t=1}^T \boldsymbol{\eta}_t \xrightarrow[T\to\infty]{p} \boldsymbol{0}$.*
3. *$T^{-1}\sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\eta}'_t \xrightarrow[T\to\infty]{p} \boldsymbol{0}, T^{-1}\sum_{t=1}^T \boldsymbol{F}_t \boldsymbol{\eta}'_t \xrightarrow[T\to\infty]{p} \boldsymbol{0}$ and $T^{-1}\sum_{t=1}^T \boldsymbol{F}_t \boldsymbol{\omega}'_t \xrightarrow[T\to\infty]{p} \boldsymbol{0}$.*
4. *$T^{-1}\sum_{t=1}^T \varepsilon_{i,t}\omega_{k,t} \xrightarrow[T\to\infty]{p} \gamma(i,k)$, and $\lim_{N\to\infty}\sup_k \sum_{i=1}^N |\gamma(i,k)| \leq A$.*
5. *$T^{-1}\sum_{t=1}^T \varepsilon_{i,t}\varepsilon_{j,t} \xrightarrow[T\to\infty]{p} \delta(i,j) = \delta(j,i)$, and $\lim_{N\to\infty}\sup_j \sum_{i=1}^N |\delta(i,j)| \leq A$.*
6. *$N^{-1}\sum_{i=1}^N \varepsilon_{i,t}\varepsilon_{i,s} \xrightarrow[N\to\infty]{p} \kappa(t,s) = \kappa(s,t)$, and $\lim_{T\to\infty}\sup_s \sum_{t=1}^T |\kappa(s,t)| \leq A$.*

Assumption 4 (Rank Condition)**.** *The matrix $\boldsymbol{\Lambda}$ is nonsingular.*

A.C State Space Representation

Transition Equations: We assume that the dynamics of the underlying latent factors are governed by

$$\boldsymbol{F}_{t+1} = \boldsymbol{\Lambda}_1 \boldsymbol{F}_t + \boldsymbol{\xi}_{t+1}. \tag{A1}$$

The state vector is comprised of the factor vector $\boldsymbol{F}_t$ and its first lag,

$$\boldsymbol{S}_{t+1} = \boldsymbol{H}\boldsymbol{S}_t + \boldsymbol{\Gamma}\boldsymbol{\xi}_{t+1}$$

where $\boldsymbol{S}_t = \begin{pmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{X}_t \end{pmatrix}$, $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{0}_{n_x} & \boldsymbol{I}_{n_x} \\ \boldsymbol{0}_{n_x} & \boldsymbol{\Lambda}_1 \end{pmatrix}$ and $\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{0}_{n_x} \\ \boldsymbol{I}_{n_x} \end{pmatrix}$.

Measurement Equations: Realized returns, realized dividend growth and the cross section of price-dividend ratios constitute the measurement vector.

$$\boldsymbol{Y}_t = \boldsymbol{M}_0 + \boldsymbol{M}_1 \boldsymbol{S}_t + \boldsymbol{w}_t$$

where $\boldsymbol{Y}_t = \begin{pmatrix} \Delta d_t \\ r_t \\ \boldsymbol{pd}_t \end{pmatrix}$, $\boldsymbol{M}_0 = \begin{pmatrix} \delta_0 \\ \gamma_0 \\ \boldsymbol{\phi}_0 \end{pmatrix}$, $\boldsymbol{M}_1 = \begin{pmatrix} \boldsymbol{\delta}' & \boldsymbol{0}_{1\times n_x} \\ \boldsymbol{\gamma}' & \boldsymbol{0}_{1\times n_x} \\ \boldsymbol{0}_{n\times n_x} & \boldsymbol{\phi} \end{pmatrix}$, $\boldsymbol{w}_t = \begin{pmatrix} \eta_t^d \\ \eta_t^r \\ \boldsymbol{\varepsilon}_t \end{pmatrix}$.

Here, $\boldsymbol{pd}_t$, $\boldsymbol{\phi}_0$, $\boldsymbol{\phi}$, $\boldsymbol{\varepsilon}_t$ denote the stacked system of individual price-dividend ratios and their coefficients and idiosyncrasies.

Error Variances: The state error variance matrix is given by

$$\boldsymbol{\Sigma} = Var(\boldsymbol{\xi}_t) = \boldsymbol{D}_s \boldsymbol{C}_s \boldsymbol{D}_s$$

where $\boldsymbol{D}_s$ is a diagonal matrix of standard deviations for each $\boldsymbol{S}$ shock and $\boldsymbol{C}_s$ is an equicorrelation matrix parameterized by ρ_s. Measurement error variance is given by

$$\boldsymbol{R} = Var(\boldsymbol{w}_t) = \boldsymbol{D}_w \boldsymbol{C}_w \boldsymbol{D}_w$$

where $\boldsymbol{D}_w = diag(\sigma_d, \sigma_r, \boldsymbol{\sigma}_i)$ is a diagonal matrix of standard deviations for shocks to dividend growth, returns, and individual price-dividend ratios ($\boldsymbol{\sigma}_i$ is shorthand for the vector of stacked price-dividend ratio shock standard deviations). The correlation matrix of $\boldsymbol{w}_t$ is

$$\boldsymbol{C}_w = \begin{pmatrix} 1 & \rho_{d,r} & \boldsymbol{\rho}'_{d,i} \\ \rho_{d,r} & 1 & \boldsymbol{\rho}'_{r,i} \\ \boldsymbol{\rho}_{d,i} & \boldsymbol{\rho}_{r,i} & \boldsymbol{\rho}_i \end{pmatrix} \tag{A2}$$

We assume equal correlation between shocks to dividend growth and each price-dividend ratio. These are denoted by the vector $\boldsymbol{\rho}_{d,i}$, which has the same correlation in all positions. The same assumption is made for the correlation between shocks to returns and each price-dividend ratio. Finally, we assume that shocks among individual price-dividend ratios are equicorrelated, and we denote this within-firm equicorrelation matrix as $\boldsymbol{\rho}_i$, which has ones on the main diagonal and the same value ρ_i in all off-diagonal positions. The assumption of equicorrelation in each of these instances serves two purposes. First, it eases computational burden when estimating the system by decreasing its parameterization. Second, it ensures the system is identified.[34]

The Kalman Filter: The filter is initialized at the state vector's unconditional mean and variance,

$$\begin{aligned} \boldsymbol{S}_{0|0} &= \mathbf{0}_{2n_x \times 1} \\ \boldsymbol{P}_{0|0} &= \mathbb{E}[\boldsymbol{S}_t \boldsymbol{S}_t'] \end{aligned}$$

The augmented state space requires only an adjustment to the equation for $\boldsymbol{S}_t$:

$$\begin{aligned} \boldsymbol{S}_{t|t-1} &= \boldsymbol{H}\boldsymbol{S}_{t-1|t-1} \\ \boldsymbol{P}_{t|t-1} &= \boldsymbol{H}\boldsymbol{P}_{t-1|t-1}\boldsymbol{H}' + \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}' \\ \boldsymbol{\eta}_t &= \boldsymbol{Y}_t - \boldsymbol{M}_0 - \boldsymbol{M}_1\boldsymbol{S}_{t|t-1} \\ \boldsymbol{V}_t &= \boldsymbol{M}_1\boldsymbol{P}_{t|t-1}\boldsymbol{M}_1' + \boldsymbol{R} \\ \boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1}\boldsymbol{M}_1'\boldsymbol{V}_t^{-1} \\ \boldsymbol{S}_{t|t} &= \boldsymbol{S}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{\eta}_t \\ \boldsymbol{P}_{t|t} &= (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{M}_1)\boldsymbol{P}_{t|t-1}. \end{aligned}$$

The system parameter vector is

$$(\boldsymbol{\Lambda}_1', \phi_0, \boldsymbol{\phi}_1, \beta_0, \boldsymbol{\beta}_1, \boldsymbol{a}', vec(\boldsymbol{b})', \boldsymbol{\sigma}_s', \sigma_d, \sigma_r, \boldsymbol{\sigma}_i', \rho_s, \rho_{d,r}, \rho_{d,i}, \rho_{r,i}, \rho_i)',$$

which is estimated via maximum likelihood.

A.D Portfolio Construction

We construct portfolio-level log price-dividend ratios from the CRSP monthly stock file using data on prices and returns with and without dividends. Six (two-by-three sorts), nine (three-by-three sorts), twelve (three-by-four sorts), 25 (five-by-five sorts) and 36 (six-by-six sorts) portfolios are formed on the basis of underlying firms' market equity and book-to-market ratio, mimicking the methodology of Fama and French (1993). Characteristics for year t are constructed as follows. Market equity is price multiplied by common shares outstanding at the end of December. Book-to-market is the ratio of book equity in year $t-1$ to market equity at the end of year t. Book equity is calculated from the Compustat file as book value of stockholders' equity

[34] A proof that this system is identified is available to the interested reader upon request.

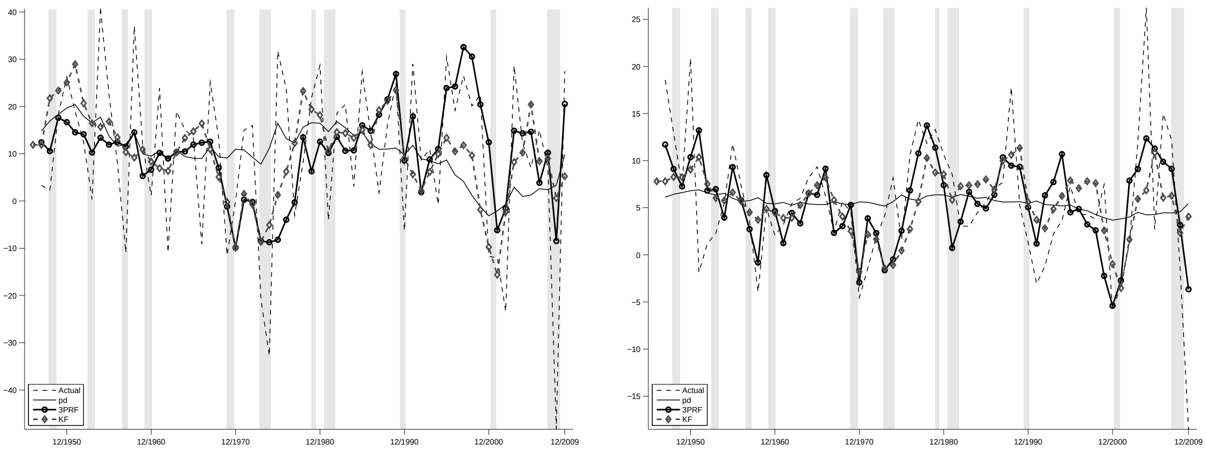
plus balance sheet deferred taxes and investment tax credit (if available) minus book value of preferred stock. Book value of preferred stock is defined as either the redemption, liquidation or par value of preferred stock (in that order). When Compustat data is unavailable, we use Moody's book equity data (if available) from Davis, Fama and French (2000). We focus on annual data to avoid seasonality in dividends, as is common in the literature. Unlike Fama and French, we rebalance the characteristic-based portfolios each month based on historical data. Using portfolio returns with and without dividends, we calculate the log price-dividend ratio for these portfolios at the end of December the following year.

For a stock to be assigned to a portfolio at time t, we require that it is classified as common equity (CRSP share codes 10 and 11) traded on NYSE, AMEX or NASDAQ, and that its $t-1$ year-end market equity value is non-missing. When forming portfolios on the basis of book-to-market we require that a firm has positive book equity at $t-1$. For the "past dividend payers" sample, a firm is included if it paid a dividend at any time in the twelve months prior to t. We perform sorts simultaneously rather than sequentially to ensure uniformity in characteristics across portfolios in both dimensions. For the past dividend payers sample, stock sorts for characteristic-based portfolio assignments are performed using equally-spaced quantiles as breakpoints to avoid excessively lop-sided allocations of firms to portfolios. That is, for a K-bin sort, portfolio breakpoints are set equal to the $\left\{\frac{100}{K}, 2\frac{100}{K}, ..., (K-1)\frac{100}{K}\right\}$ quantiles of a given characteristic. For the "all stocks" sample, NYSE breakpoints are used. As Tables A4 and A3 show, this maintains a high degree of similarity between the characteristics of portfolios in each sample.

A.E Additional Tables and Figures

Figure A1: Market Expectations: Three-Pass Regression Filter and Kalman Filter

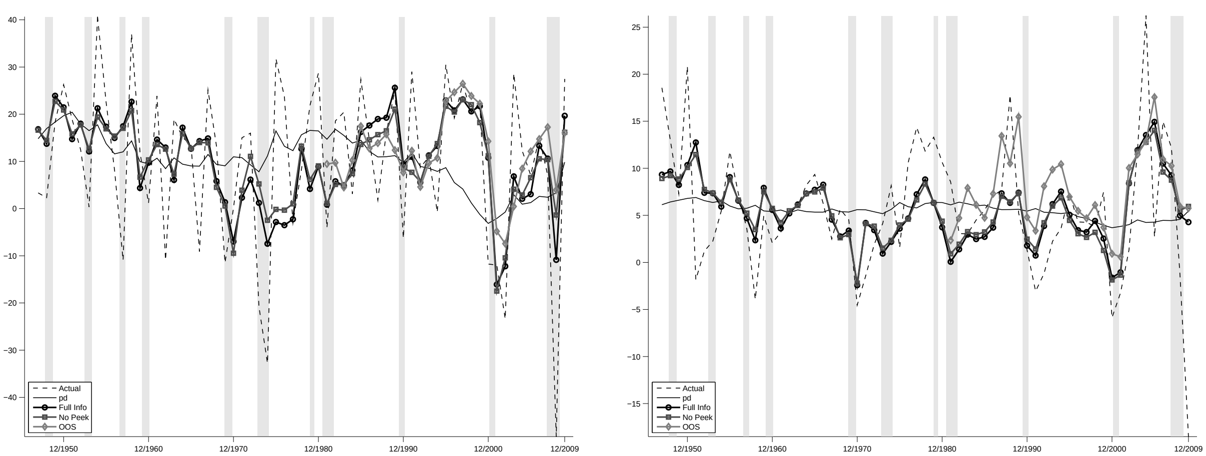
(a) Returns (b) Dividend Growth



Notes: Annual CRSP data, 1946-2009. Expectations are aligned with the realizations they forecast; that is, a forecast $\hat{\mu}_t$ of r_{t+1} is plotted at year $t+1$ in the figure. Realized market returns are labeled "Actual." Expectations given by a predictive regression on the aggregate price-dividend ratio are labeled "pd." "3PRF" labels Full Information expectation estimates from six size and book-to-market portfolios using Past Dividend Paying stocks. "KF" labels expectation estimates from cross-sectional state space model estimated by the Kalman Filter. Vertical unit is log-percent; NBER recession dates are represented by the shaded area.

Figure A2: Actual and Expected Market Returns and Dividend Growth: All Stocks

(a) Returns (b) Dividend Growth



Notes: Annual CRSP data, 1946-2009. Vertical axis is in percent. Expectations are aligned with the realizations they forecast; that is, a forecast $\hat{\mu}_t$ of r_{t+1} is plotted at year $t+1$ in the figure. Realizations are labeled "Actual." Expectations conditioned on the aggregate price-dividend ratio are labeled "pd." Expectations given by the 3PRF using six size and book-to-market portfolios of dividend payers are labeled "Full Info," "No Peek," and "OOS" for the three implementation versions, using three factors in all cases. NBER recession dates are represented by the shaded area.

Table A1: Market Return Predictions: All Stocks

	Three Factors			**One Factor**		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
			1946-2009 Sample			
			6 Portfolios			
R^2 (%)	31.35	24.30	29.58	4.83	2.18	−0.54
Test	*9.54*	*7.12*	*8.18*	*0.60*	*0.47*	–
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.442*	*0.497*	–
			9 Portfolios			
R^2 (%)	29.25	21.22	23.01	4.05	1.62	−4.19
Test	*9.75*	*6.62*	*5.23*	*0.53*	*0.32*	–
p-val	*< 0.001*	*0.001*	*< 0.01*	*0.469*	*0.574*	–
			12 Portfolios			
R^2 (%)	26.43	20.68	24.35	5.55	3.63	−3.86
Test	*9.26*	*5.73*	*5.63*	*0.89*	*0.80*	–
p-val	*< 0.001*	*0.002*	*< 0.01*	*0.350*	*0.375*	–
			25 Portfolios			
R^2 (%)	30.12	16.14	7.70	8.96	0.70	1.88
Test	*9.30*	*4.43*	*3.23*	*1.09*	*0.10*	–
p-val	*< 0.001*	*0.007*	*< 0.05*	*0.301*	*0.753*	–
			36 Portfolios			
R^2 (%)	23.80	6.16	−9.81	8.16	0.00	−6.79
Test	*6.90*	*2.19*	–	*1.35*	*0.00*	–
p-val	*< 0.001*	*0.098*	–	*0.250*	*0.989*	–
			1930-2009 Sample			
			6 Portfolios			
R^2 (%)	2.05	10.90	−7.59	1.21	2.34	−4.11
Test	*0.23*	*1.71*	–	*0.15*	*0.31*	–
p-val	*0.878*	*0.172*	–	*0.703*	*0.580*	–
			9 Portfolios			
R^2 (%)	1.95	10.27	−20.59	1.25	2.39	−3.65
Test	*0.12*	*2.15*	–	*0.19*	*0.24*	–
p-val	*0.946*	*0.101*	–	*0.666*	*0.623*	–
			12 Portfolios			
R^2 (%)	4.85	6.11	−1.85	1.52	0.12	0.16
Test	*0.62*	*0.49*	–	*0.21*	*0.05*	–
p-val	*0.607*	*0.689*	–	*0.652*	*0.828*	–
			25 Portfolios			
R^2 (%)	18.11	1.41	−1.44	16.81	2.37	−4.16
Test	*4.74*	*0.32*	–	*7.06*	*0.67*	–
p-val	*0.004*	*0.807*	–	*0.010*	*0.415*	–
			36 Portfolios			
R^2 (%)	20.47	5.36	−10.11	15.18	1.66	−7.32
Test	*5.92*	*1.84*	–	*5.80*	*0.60*	–
p-val	*0.001*	*0.146*	–	*0.018*	*0.440*	–

Notes: Annual data from CRSP, 1946-2009. The training window for out-of-sample forecasts is 34 years (out-of-sample forecasts begin in 1981). Portfolios are formed on the basis of market equity and book-to-market value using all stocks. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using the realized market return as sole proxy.

Table A2: DG Predictions: All Stocks

	Three Factors			One Factor		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Full Information	No Peek		Full Information	No Peek	
1946-2009 Sample						
6 Portfolios						
R^2 (%)	26.76	23.25	18.74	6.19	3.13	1.94
Test	*11.28*	*8.63*	*5.28*	*0.95*	*0.61*	*0.02*
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.335*	*0.437*	–
9 Portfolios						
R^2 (%)	29.27	23.39	18.03	6.21	2.55	1.87
Test	*11.62*	*8.96*	*5.46*	*0.75*	*0.48*	–
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.389*	*0.492*	–
12 Portfolios						
R^2 (%)	29.29	21.55	15.22	4.79	3.18	−1.42
Test	*13.10*	*7.82*	*4.21*	*0.89*	*0.70*	–
p-val	*< 0.001*	*< 0.001*	*< 0.01*	*0.350*	*0.405*	–
25 Portfolios						
R^2 (%)	33.13	18.49	4.11	7.39	0.46	1.24
Test	*14.22*	*6.55*	*2.21*	*1.51*	*0.07*	–
p-val	*< 0.001*	*0.001*	*< 0.10*	*0.223*	*0.787*	–
36 Portfolios						
R^2 (%)	37.97	19.51	−24.11	9.91	0.65	1.79
Test	*15.68*	*6.02*	–	*1.46*	*0.11*	–
p-val	*< 0.001*	*0.001*	–	*0.231*	*0.745*	–
1930-2009 Sample						
6 Portfolios						
R^2 (%)	9.41	15.39	−19.17	0.06	1.73	−2.93
Test	*2.58*	*1.12*	–	*0.01*	*0.25*	–
p-val	*0.060*	*0.345*	–	*0.927*	*0.615*	–
9 Portfolios						
R^2 (%)	26.50	15.99	−31.12	0.05	2.86	−2.06
Test	*5.62*	*0.94*	–	*0.01*	*0.31*	–
p-val	*0.002*	*0.426*	–	*0.938*	*0.582*	–
12 Portfolios						
R^2 (%)	23.30	10.46	−20.55	5.60	0.63	−12.81
Test	*5.63*	*0.87*	–	*0.96*	*0.17*	–
p-val	*0.002*	*0.462*	–	*0.330*	*0.677*	–
25 Portfolios						
R^2 (%)	19.44	40.75	−36.40	0.04	2.86	−5.29
Test	*5.04*	*5.53*	–	*0.02*	*0.59*	–
p-val	*0.003*	*0.002*	–	*0.891*	*0.447*	–
36 Portfolios						
R^2 (%)	31.76	10.48	−37.37	0.83	0.57	−1.11
Test	*10.51*	*1.67*	–	*0.32*	*0.21*	–
p-val	*< 0.001*	*0.180*	–	*0.574*	*0.651*	–

Notes: Annual data from CRSP. The training window for the 1946-2009 out-of-sample forecasts is 34 years (out-of-sample forecasts begin in 1981), and for the 1930-2009 sample is 40 years (out-of-sample forecasts begin in 1971). Portfolios are formed on the basis of market equity and book-to-market value using all stocks. *Test* denotes one of two test statistics. For in-sample results, we report the bootstrapped F-statistic in *italics* (and associated p-values) using the circular block bootstrap with 1000 simulations. Out-of-sample results report Clark and McCracken's (2001) encompassing test statistic in *italics* with p-values from Clark and McCracken's (2001) appendix tables. This tests the null hypothesis of no forecast improvement over the historical mean. Full information, no-peek and out-of-sample procedures are described in Section II.C. Columns under the heading "Three Factors" extract three factors using realized market return, realized market dividend growth, and market price-dividend ratio as proxies. Columns labeled "One Factor" extract a single factor using the realized market return as sole proxy.

Table A3: Size and Book-to-Market Portfolio Summary Statistics: Past Dividend Payers (1946-2009)

Portfolio	# Stocks in Portfolio Min	Max	% <5	Mean ME	Mean B/M	σ(ME)	σ(B/M)	$\sigma(r)$	$\sigma(\Delta d)$
					6 Portfolios				
1	50	344	0	99.43	0.41	0.93	0.29	0.18	0.39
2	103	430	0	83.66	0.76	0.91	0.27	0.16	0.34
3	121	626	0	59.97	1.69	0.92	0.38	0.19	0.53
4	124	618	0	4295.29	0.36	1.35	0.28	0.15	0.35
5	95	478	0	2272.71	0.73	1.36	0.27	0.14	0.32
6	53	324	0	1637.32	1.45	1.21	0.30	0.16	0.43
					9 Portfolios				
1	23	188	0	49.43	0.42	0.85	0.30	0.18	0.53
2	66	269	0	44.83	0.77	0.85	0.27	0.16	0.38
3	87	486	0	33.63	1.73	0.80	0.39	0.18	0.50
4	64	335	0	273.27	0.39	1.05	0.28	0.17	0.33
5	76	331	0	249.45	0.75	1.02	0.27	0.16	0.35
6	64	276	0	234.01	1.52	1.03	0.32	0.18	0.46
7	93	476	0	5678.47	0.36	1.35	0.29	0.15	0.36
8	54	334	0	3471.29	0.73	1.37	0.27	0.14	0.33
9	29	227	0	2526.51	1.45	1.24	0.31	0.16	0.44
					12 Portfolios				
1	11	132	0	49.45	0.36	0.85	0.31	0.19	0.74
2	34	181	0	48.15	0.62	0.86	0.27	0.17	0.41
3	49	250	0	41.08	0.92	0.83	0.27	0.16	0.40
4	66	400	0	32.50	1.91	0.79	0.41	0.19	0.56
5	46	255	0	279.30	0.34	1.06	0.29	0.18	0.38
6	41	244	0	252.42	0.61	1.01	0.27	0.16	0.44
7	53	252	0	247.70	0.90	1.03	0.27	0.16	0.33
8	42	194	0	230.07	1.73	1.03	0.35	0.19	0.54
9	71	372	0	6053.06	0.31	1.35	0.29	0.15	0.39
10	48	277	0	3787.58	0.60	1.39	0.27	0.15	0.33
11	24	226	0	3138.77	0.89	1.39	0.26	0.15	0.40
12	17	171	0	2383.17	1.64	1.19	0.34	0.18	0.56
					25 Portfolios				
1	0	56	17.06	26.68	0.32	0.78	0.34	0.23	1.36
2	6	73	0	25.76	0.54	0.79	0.28	0.19	0.75
3	16	92	0	25.74	0.76	0.81	0.27	0.17	0.52
4	29	135	0	22.66	1.03	0.77	0.27	0.16	0.43
5	34	241	0	19.24	2.12	0.73	0.46	0.18	0.73
6	9	99	0	85.69	0.31	0.90	0.32	0.19	0.77
7	15	107	0	83.30	0.54	0.90	0.28	0.17	0.39
8	23	125	0	80.34	0.75	0.88	0.27	0.17	0.45
9	24	132	0	79.24	1.02	0.89	0.27	0.18	0.42
10	22	144	0	77.62	1.94	0.89	0.37	0.21	0.62
11	19	129	0	243.19	0.30	1.03	0.29	0.18	0.64
12	17	120	0	234.68	0.54	1.03	0.28	0.17	0.55
13	23	125	0	229.80	0.74	1.02	0.27	0.16	0.42
14	24	127	0	229.88	1.01	1.02	0.27	0.17	0.38
15	15	92	0	221.04	1.87	1.01	0.38	0.21	0.64
16	29	154	0	737.71	0.30	1.11	0.30	0.17	0.45
17	25	135	0	724.20	0.53	1.11	0.28	0.16	0.54
18	22	123	0	699.27	0.74	1.10	0.27	0.16	0.46
19	12	98	0	697.77	1.01	1.10	0.27	0.16	0.43
20	10	73	0	703.94	1.88	1.10	0.35	0.20	0.48
21	36	213	0	8801.80	0.28	1.34	0.30	0.15	0.45
22	23	149	0	6373.92	0.53	1.41	0.27	0.15	0.39
23	18	127	0	5766.80	0.74	1.46	0.27	0.15	0.40
24	6	110	0	4495.82	1	1.42	0.27	0.16	0.52
25	5	78	0	3769.01	1.75	1.25	0.40	0.20	0.65

Table A3: Portfolio Summary Statistics Continued

	# Stocks in Portfolio								
Portfolio	Min	Max	% Mo. <5	Mean ME	Mean B/M	σ(ME)	σ(B/M)	$\sigma(r)$	$\sigma(\Delta d)$
				36 Portfolios					
1	0	34	35.42	23.40	0.29	0.79	0.40	0.30	1.94
2	1	50	6.51	21.44	0.49	0.77	0.28	0.21	0.98
3	4	59	0.13	21.68	0.66	0.81	0.27	0.18	0.80
4	13	78	0	20.97	0.85	0.78	0.27	0.17	0.53
5	19	111	0	18.93	1.12	0.75	0.27	0.16	0.54
6	26	192	0	16.42	2.29	0.73	0.51	0.19	0.76
7	3	61	0.91	62.94	0.29	0.87	0.34	0.22	1.27
8	8	68	0	62.55	0.49	0.86	0.28	0.18	0.55
9	13	85	0	61.43	0.66	0.86	0.27	0.17	0.51
10	15	93	0	59.63	0.84	0.84	0.27	0.17	0.52
11	14	99	0	58.75	1.11	0.84	0.28	0.18	0.48
12	10	115	0	58.47	2.07	0.86	0.37	0.21	0.69
13	5	92	0	151.29	0.28	0.96	0.32	0.19	0.67
14	7	85	0	148.38	0.49	0.95	0.28	0.17	0.55
15	9	93	0	147.57	0.65	0.95	0.27	0.17	0.51
16	17	91	0	148.21	0.84	0.96	0.27	0.17	0.47
17	15	88	0	147.37	1.10	0.98	0.27	0.18	0.59
18	15	78	0	142.73	2.09	0.97	0.44	0.22	0.74
19	14	94	0	368.40	0.28	1.07	0.30	0.19	0.61
20	11	84	0	362.77	0.48	1.07	0.28	0.17	0.36
21	14	95	0	353.74	0.65	1.05	0.27	0.16	0.66
22	11	84	0	356.56	0.84	1.06	0.27	0.16	0.47
23	8	96	0	353.48	1.09	1.06	0.27	0.17	0.48
24	8	56	0	347.26	1.95	1.06	0.36	0.21	0.67
25	21	118	0	1017.08	0.27	1.16	0.31	0.17	0.52
26	18	121	0	974.33	0.48	1.13	0.28	0.16	0.50
27	12	94	0	949.33	0.65	1.11	0.27	0.16	0.36
28	7	78	0	933.69	0.83	1.11	0.27	0.16	0.53
29	6	68	0	946.43	1.10	1.10	0.27	0.17	0.49
30	5	63	0	961.12	2.04	1.13	0.42	0.21	0.56
31	25	171	0	10255.24	0.26	1.35	0.31	0.15	0.50
32	17	119	0	7796.01	0.47	1.44	0.28	0.15	0.43
33	10	92	0	6887.15	0.65	1.42	0.27	0.15	0.50
34	8	93	0	5968.40	0.84	1.42	0.27	0.16	0.49
35	4	86	0.78	4900.96	1.10	1.37	0.28	0.16	0.57
36	3	49	4.43	4264.45	1.84	1.28	0.37	0.21	0.71

Notes: Annual CRSP data, 1946-2009. Summary statistics for portfolios of past dividend payers formed on the basis of market equity and book-to-market ratio as described in Section III.A. Reported are the time series minimum and maximum number of stocks in each portfolio, as well as the fraction of observations for which fewer than five stocks are in the portfolio. We also report the time series mean for the average market equity and book-to-market ratio for firms in the portfolio, the time series standard deviation of portfolio average market equity (divided by its time series mean market equity labeled simply $\sigma(ME)$), book-to-market ratio, log annual return and log annual dividend growth.

Table A4: Size and Book-to-Market Portfolio Summary Statistics: All Stocks (1946-2009)

Portfolio	# Stocks in Portfolio Min	Max	% <5	Mean ME	Mean B/M	σ(ME)	σ(B/M)	$\sigma(r)$	$\sigma(\Delta d)$
					6 Portfolios				
1	58	1562	0	125.98	0.36	1.18	0.37	0.21	0.51
2	122	1453	0	114.31	0.76	1.16	0.29	0.18	0.42
3	150	1986	0	71.44	1.96	1.15	0.37	0.20	0.66
4	151	769	0	4397.79	0.34	1.31	0.33	0.15	0.35
5	117	339	0	2960.16	0.73	1.37	0.29	0.14	0.31
6	56	227	0	2351.17	1.50	1.35	0.33	0.17	0.48
					9 Portfolios				
1	25	1338	0	75.08	0.37	1.21	0.38	0.23	0.62
2	75	1248	0	71.82	0.76	1.20	0.29	0.18	0.45
3	107	1882	0	49.15	2.02	1.18	0.38	0.21	0.73
4	73	557	0	480.08	0.36	1.20	0.34	0.19	0.43
5	87	368	0	472.13	0.74	1.21	0.29	0.17	0.33
6	67	208	0	468.45	1.54	1.22	0.33	0.19	0.60
7	106	514	0	6252.79	0.34	1.30	0.33	0.15	0.36
8	69	214	0	4559.88	0.72	1.39	0.29	0.14	0.32
9	30	150	0	3539.60	1.51	1.36	0.34	0.17	0.48
					12 Portfolios				
1	14	1055	0	74.60	0.31	1.21	0.39	0.24	0.83
2	37	819	0	76.02	0.61	1.21	0.30	0.19	0.43
3	58	1116	0	65.94	0.93	1.17	0.28	0.19	0.54
4	85	1470	0	46.07	2.27	1.18	0.40	0.22	0.81
5	53	460	0	478.98	0.31	1.20	0.35	0.19	0.50
6	52	319	0	479.22	0.60	1.21	0.30	0.17	0.40
7	62	228	0	467.46	0.90	1.21	0.28	0.17	0.40
8	45	145	0	468.16	1.77	1.23	0.34	0.20	0.66
9	79	454	0	6511.72	0.30	1.30	0.34	0.16	0.41
10	61	187	0	4776.61	0.59	1.37	0.29	0.14	0.33
11	30	140	0	4299.24	0.89	1.41	0.28	0.15	0.45
12	17	109	0	3399.03	1.75	1.34	0.37	0.18	0.59
					25 Portfolios				
1	2	717	8.72	46.23	0.27	1.24	0.40	0.27	1.56
2	5	481	0	47.46	0.53	1.22	0.31	0.22	0.78
3	17	626	0	45.99	0.75	1.23	0.29	0.19	0.67
4	31	803	0	40.14	1.05	1.16	0.28	0.19	0.64
5	48	1164	0	31.65	2.59	1.19	0.44	0.23	0.98
6	10	276	0	200.40	0.28	1.27	0.39	0.23	1.01
7	18	210	0	198.83	0.52	1.28	0.30	0.19	0.53
8	23	217	0	196.90	0.74	1.27	0.29	0.18	0.53
9	27	178	0	194.48	1.03	1.28	0.28	0.18	0.50
10	24	125	0	189.62	2.04	1.28	0.38	0.23	0.86
11	20	246	0	467.64	0.28	1.22	0.36	0.20	0.55
12	18	163	0	468.55	0.52	1.22	0.30	0.17	0.69
13	27	131	0	466.43	0.74	1.21	0.29	0.17	0.35
14	22	101	0	465.37	1.02	1.22	0.28	0.17	0.58
15	14	65	0	466.17	1.97	1.24	0.35	0.22	0.77
16	32	225	0	1194.87	0.28	1.21	0.37	0.19	0.62
17	32	131	0	1170.94	0.52	1.18	0.30	0.16	0.49
18	25	102	0	1144.24	0.73	1.15	0.28	0.17	0.69
19	14	81	0	1153.06	1.02	1.16	0.28	0.17	0.49
20	10	56	0	1165.76	1.99	1.19	0.40	0.21	0.80
21	42	264	0	9948.03	0.27	1.31	0.35	0.16	0.50
22	27	100	0	8066.70	0.51	1.42	0.30	0.15	0.38
23	19	72	0	7158.80	0.73	1.35	0.29	0.15	0.44
24	9	64	0	5965.24	1.02	1.39	0.28	0.15	0.59
25	4	46	0.13	5236.78	1.86	1.36	0.38	0.19	0.71

Table A4: PORTFOLIO SUMMARY STATISTICS CONTINUED

Portfolio	# Stocks in Portfolio Min	Max	% Mo. <5	Mean ME	Mean B/M	σ(ME)	σ(B/M)	$\sigma(r)$	$\sigma(\Delta d)$
					36 PORTFOLIOS				
1	0	621	21.22	40.02	0.24	1.25	0.42	0.29	2.52
2	2	363	6.90	40.70	0.48	1.21	0.32	0.23	1.17
3	4	424	0.13	40.97	0.65	1.25	0.30	0.21	0.74
4	13	554	0	38.26	0.86	1.20	0.29	0.19	0.76
5	27	738	0	34.12	1.16	1.16	0.29	0.19	0.74
6	36	980	0	27.66	2.84	1.17	0.47	0.24	1.15
7	2	218	1.30	155.90	0.26	1.29	0.42	0.25	1.29
8	8	153	0	156.67	0.47	1.30	0.31	0.21	0.73
9	14	176	0	155.74	0.65	1.29	0.29	0.19	0.52
10	17	169	0	154.80	0.85	1.28	0.28	0.18	0.64
11	17	151	0	153.68	1.13	1.29	0.29	0.20	0.64
12	16	108	0	148.82	2.28	1.28	0.42	0.24	1.02
13	9	212	0	329.88	0.26	1.24	0.37	0.22	0.61
14	12	129	0	330.34	0.47	1.24	0.31	0.19	0.84
15	15	126	0	327.74	0.64	1.24	0.29	0.18	0.75
16	18	101	0	323.62	0.84	1.24	0.28	0.17	0.59
17	16	94	0	329.40	1.12	1.24	0.28	0.19	0.70
18	13	53	0	324.79	2.18	1.25	0.39	0.24	0.89
19	15	166	0	666.87	0.26	1.20	0.37	0.20	0.84
20	13	112	0	662.21	0.47	1.20	0.31	0.17	0.68
21	13	99	0	666.14	0.64	1.20	0.29	0.17	0.38
22	15	70	0	662.58	0.84	1.19	0.28	0.17	0.41
23	10	65	0	664.76	1.12	1.20	0.28	0.18	0.64
24	6	45	0	657.77	2.05	1.19	0.37	0.22	0.99
25	21	171	0	1544.52	0.26	1.21	0.37	0.19	0.78
26	21	87	0	1512.77	0.46	1.19	0.30	0.16	0.63
27	15	84	0	1515.49	0.64	1.20	0.29	0.17	0.48
28	7	62	0	1520.66	0.84	1.19	0.28	0.17	0.67
29	6	57	0	1535.92	1.13	1.20	0.28	0.18	0.54
30	4	40	0.26	1498.46	2.23	1.20	0.47	0.22	0.78
31	28	216	0	11496.56	0.24	1.32	0.36	0.16	0.55
32	20	78	0	9693.93	0.46	1.42	0.31	0.15	0.41
33	13	56	0	8367.86	0.64	1.36	0.29	0.15	0.46
34	9	56	0	7946.66	0.84	1.41	0.28	0.16	0.56
35	4	47	0.13	6552.12	1.12	1.36	0.29	0.16	0.67
36	1	31	10.55	5882.47	2.03	1.40	0.41	0.21	0.87

Notes: Annual CRSP data, 1946-2009. Summary statistics for portfolios of all stocks formed on the basis of market equity and book-to-market ratio as described in Section III.A. Reported are the time series minimum and maximum number of stocks in each portfolio, as well as the fraction of observations for which fewer than five stocks are in the portfolio. We also report the time series mean for the average market equity and book-to-market ratio for firms in the portfolio, the time series standard deviation of portfolio average market equity (divided by its time series mean market equity labeled simply $\sigma(ME)$), book-to-market ratio, log annual return and log annual dividend growth.