

## Market Expectations in the Cross-Section of Present Values

BRYAN KELLY and SETH PRUITT\*

### ABSTRACT

Returns and cash flow growth for the aggregate U.S. stock market are highly and robustly predictable. Using a single factor extracted from the cross-section of book-to-market ratios, we find an *out-of-sample* return forecasting  $R^2$  of 13% at the annual frequency (0.9% monthly). We document similar out-of-sample predictability for returns on value, size, momentum, and industry portfolios. We present a model linking aggregate market expectations to disaggregated valuation ratios in a latent factor system. Spreads in value portfolios' exposures to economic shocks are key to identifying predictability and are consistent with duration-based theories of the value premium.

THE MOST COMMON APPROACH to measuring aggregate return and cash flow expectations is predictive regression. As suggested by the present value relationship between prices, discount rates, and future cash flows, research shows that the aggregate price-dividend ratio is among the most informative predictive variables. Typical in-sample estimates find that about 10% of annual return variation can be accounted for by forecasts based on the aggregate book-to-market ratio, but find little or no out-of-sample predictive power.<sup>1</sup> In this paper we show that reliance on aggregate quantities drastically understates the degree of value ratios' predictive content for both returns and cash flow growth, and hence understates the volatility of investor expectations. Our estimates suggest that as much as 13% of the out-of-sample variation in annual market returns (as much as 12% for dividend growth), and somewhat more of the in-sample variation, can be explained by the cross-section of past disaggregated value ratios.

To harness disaggregated information we represent the cross-section of asset-specific book-to-market ratios as a dynamic latent factor model. We relate disaggregated value ratios to aggregate expected market returns and cash flow growth. Our model is based on the idea that the same dynamic state variables driving aggregate expectations also govern the dynamics of the entire panel

\*Kelly is with Booth School of Business, University of Chicago, and Pruitt is with the Board of Governors of the Federal Reserve System. The view expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve System or its staff.

<sup>1</sup> See Cochrane (2005) and Kojen and Van Nieuwerburgh (2011) for surveys of return and cash flow predictability evidence using the aggregate price-dividend ratio. Similar results obtain from forecasts based on the aggregate book-to-market ratio.

DOI: 10.1111/jofi.12060

of asset-specific valuation ratios. This representation allows us to exploit rich cross-sectional information to extract precise estimates of market expectations.

Our approach attacks a challenging problem in empirical asset pricing: how does one exploit a wealth of predictors in relatively short time series? If the predictors number near or more than the number of observations, the standard ordinary least squares (OLS) forecaster will be poorly behaved or nonexistent (see Huber (1973)). Our solution is to use partial least squares (PLS; Wold (1975)), which is a simple regression-based procedure designed to parsimoniously forecast a single time series using a large panel of predictors. We use it to construct a *univariate* forecaster for market returns (or dividend growth) that is a linear combination of assets' valuation ratios. The weight of each asset in this linear combination is based on the covariance of its value ratio with the forecast target. Much of our analysis relies on results from Kelly and Pruitt (2012), who derive properties of PLS in a factor model setting that apply directly to the asset pricing model considered here.

Using data from 1930 to 2010, PLS forecasts based on the cross-section of portfolio-level book-to-market ratios achieve an out-of-sample predictive  $R^2$  as high as 13.1% for annual market returns and 0.9% for monthly returns (in-sample  $R^2$  of 18.1% and 2.4%, respectively). Since we construct a single factor from the cross-section, our results can be directly compared to univariate forecasts from the many alternative predictors considered in the literature. In contrast to our results, previously studied predictors typically perform well in-sample but become insignificant out-of-sample, often performing worse than forecasts based on the historical mean return (Goyal and Welch (2008)).

Our estimates shed new light on the dynamic processes for expected returns. Cross-section-based estimates suggest that discount rates are less than half as persistent, and the volatility of discount rate shocks is more than twice as large, as estimates based on the aggregate book-to-market ratio. This degree of variability in return expectations is difficult to reconcile with state-of-the-art structural asset pricing models, which are calibrated to produce persistent expected market returns with mild shock volatility (e.g., Campbell and Cochrane (1999) and Bansal and Yaron (2004)).

We establish the robustness of our main return prediction results in a number of ways. First, we evaluate various degrees of portfolio aggregation and find similar results whether we use 6, 25, or 100 Fama-French portfolios. Second, we find similarly strong out-of-sample forecast power for annual returns on value-, size-, momentum-, and industry-sorted portfolios. Third, we find consistent results when we forecast with the cross-section of individual stocks or consider different out-of-sample test windows. Fourth, we find a similar degree of predictability in an international sample. Finally, we conduct placebo tests that help rule out spurious significance of our results due to statistical biases.

Why do disaggregated prices produce such accurate forecasts? To illustrate the advantages of cross-section information, consider a simple CAPM example.<sup>2</sup>

<sup>2</sup> 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  $\mu_t$  and  $g_t$  each follow an AR(1), individual expected returns

In particular, suppose one-period expected market returns  $\mu_t$  and expected return on equity (ROE)  $g_t$  are the two common factors in the economy, and the book-to-market ratio of any asset  $i$  is

$$v_{i,t} = a_i - b_{i,\mu}\mu_t + b_{i,g}g_t + e_{i,t}, \quad (1)$$

while the aggregate book-to-market ratio is

$$v_t = a - b_\mu\mu_t + b_gg_t. \quad (2)$$

Equation (2) highlights the predictive relationship between  $v_t$ , realized market returns ( $r_{t+1} = \mu_t + \epsilon_{t+1}^r$ ), and ROE ( $\Delta cf_{t+1} = g_t + \epsilon_{t+1}^d$ ).<sup>3</sup> However, it also illustrates the limitations of the aggregate system. Predictive regressions of  $r_{t+1}$  on  $v_t$  take the form

$$\mathbb{E}_t[r_{t+1}|v_t] = \hat{a} + \hat{b}v_t = \hat{a} + \hat{b}(b_\mu\mu_t + b_gg_t) \quad (3)$$

and thus are unduly influenced by information about expected ROE. The reciprocal problem arises in forecasting  $\Delta cf_{t+1}$ . To overcome this difficulty, researchers have taken present value approaches that account for the joint relationship among  $v_t$ ,  $\mu_t$ , and  $g_t$  (see Cochrane (2008), 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  $\mu_t$  and  $g_t$  are latent, each adds noise to the signal extraction problem of the other.<sup>4</sup> If there exist other signals for  $\mu_t$  and  $g_t$  in the economy, incorporating them will improve estimates of the latent expectations. This is how disaggregated valuation ratios in equation (1) become valuable information as long as each  $v_{i,t}$  provides a nonredundant signal for  $\mu_t$  and  $g_t$ . PLS conveniently reduces the many available signals to an optimal forecast with a series of OLS regressions.

The economics literature mainly relies on principal components (PCs) to condense information from a large cross-section into a small number of predictive factors *before* estimating a linear forecast, an approach exemplified in the macroforecasting literature by Stock and Watson (2002). PC forecasts based on macroeconomic indicators have recently been applied in the context of stock return prediction by Ludvigson and Ng (2007). The key difference between PC and PLS is their method of dimension reduction. PLS condenses the cross-section according to *covariance with the forecast target* and chooses a linear combination of predictors that is optimal for forecasting. In contrast,

obey an exact one-factor model as in the CAPM,  $\mu_{i,t} = \mu_{i,0} + c_{i,\mu}\mu_t$ , and individual expected ROE obeys a one-factor model,  $g_{i,t} = g_{i,0} + c_{i,g}g_t + e_{i,t}$ . This special case is similar to the formulation of Polk, Thompson, and Vuolteenaho (2006).

<sup>3</sup> In Vuolteenaho's (2002) book-to-market identity, cash flow growth enters as ROE. We later discuss how this system relates to the Campbell and Shiller (1988) price-dividend identity, where cash flows enter in the form of dividend growth.

<sup>4</sup> This remains true despite the absence of measurement error in the aggregate book-to-market expression, as pointed out by Fama and French (1988).

the PC approach condenses the cross-section according to *covariance within the predictors*. The components that best describe predictor variation are not necessarily the factors most useful for forecasting, and therefore PCs can produce suboptimal forecasts (see Kelly and Pruitt (2012) for a detailed discussion). As we show, PCs have little forecasting success in our present value setting.

In Section I we present an economic framework for the cross-section of present values. Section II introduces PLS and relevant results from Kelly and Pruitt (2012). In Section III we present empirical findings, compare alternative methodologies, and discuss our results. We discuss the economic implications of our findings in Section IV. Technical assumptions and other details are relegated to the Internet Appendix.<sup>5</sup>

## I. The Cross-Sectional Present Value System

We assume that one-period expected log returns and log cash flow growth rates across assets and over all horizons are linear in a set of common factors<sup>6</sup>:

$$\begin{aligned}\mu_{i,t} &= \mathbb{E}_t[r_{i,t+1}] = \gamma_{i,0} + \gamma'_i \mathbf{F}_t \\ g_{i,t} &= \mathbb{E}_t[\Delta c f_{i,t+1}] = \delta_{i,0} + \delta'_i \mathbf{F}_t + \varepsilon_{i,t}.\end{aligned}\quad (4)$$

Equation (4) states that, conditional on time  $t$  information, expected one-period returns and growth rates are driven solely by the  $(K_F \times 1)$  vector of factors  $\mathbf{F}_t$  that are common across valuation ratios.

We assume that assets' expected returns are determined by systematic factors and possess no idiosyncratic behavior.<sup>7</sup> This restriction is not imposed for assets' expected dividend growth, which may possess an idiosyncratic component,  $\varepsilon_{i,t}$ . The aggregate market obeys the same structure, with no  $i$  subscripts and no idiosyncrasies:

$$\begin{aligned}\mu_t &= \gamma_0 + \boldsymbol{\gamma}' \mathbf{F}_t \\ g_t &= \delta_0 + \boldsymbol{\delta}' \mathbf{F}_t.\end{aligned}\quad (5)$$

<sup>5</sup> The Internet Appendix may be found in the online version of this article.

<sup>6</sup> 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 and French (1993) models, or indirectly via common state variables, as in Merton's (1973) ICAPM or the APT (Ross (1976), Roll and Ross (1980)). 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).

<sup>7</sup> 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.

The factor loading vectors  $\gamma$  and  $\delta$  summarize how market expectations respond to movements in the underlying economic factors.<sup>8</sup> For example, in the case of returns we assume that at least  $K_f \leq K_F$  of the factors receive nonzero loadings (i.e., have nonzero elements in  $\gamma$ ). If  $K_f < K_F$ , the remaining  $K_F - K_f$  factors are *irrelevant* for explaining market return expectations. The case of cash flows (and  $\delta$ ) is analogous. We discuss later how PLS successfully filters out the impact of these irrelevant factors, while the PC method cannot be guaranteed to do so.

To emphasize the parsimony of our approach we focus on  $K_f = 1$ , the case in which a single factor drives expected returns or cash flow growth (though our approach generalizes to multiple factors). We do not need to make any assumptions about the total number of common factors among value ratios ( $K_F$ ), since target-irrelevant factors will be filtered out by PLS.

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

$$\begin{aligned} r_{i,t+1} &= \mu_{i,t} + \eta_{i,t+1}^r \\ \Delta cf_{i,t+1} &= g_{i,t} + \eta_{i,t+1}^c. \end{aligned}$$

Finally, we assume that the factor vector evolves as a first-order vector autoregression<sup>9</sup>

$$\mathbf{F}_{t+1} = \mathbf{\Lambda}_1 \mathbf{F}_t + \boldsymbol{\xi}_{t+1}. \quad (6)$$

The above structure may be embedded in the linearized present value formula of Vuolteenaho (2002). This accounting-based identity relates an asset's log book-to-market ratio to future discount rates and earnings growth

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

where  $v_{i,t}$  is the log book-to-market ratio of stock  $i$ ,  $r_{i,t+j}$  is its log return,  $\Delta cf_{i,t+j}$  is its ROE, and  $\kappa_i$  and  $\rho_i$  are linearization constants. ROE is defined as<sup>10</sup>

$$ROE_{t+j} = \log \left( 1 + \frac{\text{earnings}_{t+j}}{\text{book equity}_{t+j-1}} \right).$$

<sup>8</sup> We assume that all model parameters are constant. When we estimate the model using Fama-French size- and value-sorted portfolios, we find an impressive degree of stability in estimated parameters between the first and second half of the sample. This is not the case when the cross-section involves individual stocks. Our framework may be generalized to incorporate time-varying parameters, or may be implemented using rolling window parameter estimates, as in our stock-level analysis in Section III.A.3.

<sup>9</sup> That  $\mathbf{F}_t$  is a first-order process is without loss of generality since any higher order vector autoregression can be written as a VAR(1).

<sup>10</sup> Vuolteenaho (2002) represents this identity in terms of excess returns. We use his identity exactly, though we represent it in terms of returns rather than excess returns.

This weighted sum of expected one-period returns and growth rates over all future horizons, combined with (4) and (6), reduces to the following expression for the valuation ratio:

$$v_{i,t} = \phi_{i,0} + \phi_i' \mathbf{F}_t + \varepsilon_{i,t}, \quad (7)$$

where formulas for  $\phi_{i,0}$  and  $\phi_i$  are provided in the Internet Appendix. Equations (5) and (7) unify disaggregated valuation ratios and aggregate expectations via a common factor model. They also provide a framework for using cross-section information to achieve our ultimate goal of precisely estimating conditional expected market returns and cash flow growth.

An alternative to Vuolteenaho's (2002) present value system is the well-known Campbell and Shiller (1988) present value identity, which relates the log price-dividend ratio of an asset to its future discount rates and dividend growth. The Campbell–Shiller identity also falls into the framework of equations (4) to (7) when  $v_{i,t}$  is the log price-dividend ratio,  $r_{i,t+j}$  is the log return, and  $\Delta cf_{i,t+j}$  is log dividend growth. However, working with individual assets' price-dividend ratios can be problematic. Fama and French (2001) document a steep downward trend in the fraction of U.S. firms paying dividends, with only 20.8% of firms classified as cash dividend payers in 1999. Because price-dividend ratios are undefined for the majority of stocks, our analysis focuses on the cross-section of book-to-market ratios. However, we do consider the performance of certain price-dividend ratios in a robustness check.

## II. Estimation

In this section we outline our empirical methodology, which is based on Kelly and Pruitt's (2012) generalization of PLS. Interested readers can refer to that paper's discussion of the three-pass regression filter for detailed econometric development and proofs. Assumptions underlying the stated results are explained here and made precise in the Internet Appendix.

### A. Setup

To ease the algebraic development we first establish notation. PLS uses two sets of inputs. The first input is the forecasting target, which in general takes the form  $y_{t+h} = \beta_0 + \beta' F_t + \eta_{t+h}$ . We focus primarily on two targets—aggregate market returns  $r_m$  and cash flow growth  $\Delta cf_m$ :

$$y_{t+h} = \begin{cases} \gamma_0 + \gamma' F_t + \eta_{t+h}^r & \text{if } y_{t+h} = r_{m,t+h} \\ \delta_0 + \delta' F_t + \eta_{t+h}^c & \text{if } y_{t+h} = \Delta cf_{m,t+h} \end{cases} \quad (8)$$

Defining  $F_{(T \times K_F)} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_T]'$ , the matrix representation of  $y_{t+h}$  is

$$\begin{aligned} \underset{(T \times 1)}{\mathbf{y}} &= [y_{h+1}, y_{h+2}, \dots, y_{h+T}]' \\ &= \iota \beta_0 + \mathbf{F} \beta + \eta, \end{aligned} \quad (9)$$

where  $\beta_0, \beta$  are defined in the obvious way for either  $r_{m,t+1}$  or  $\Delta cf_{m,t+1}$ .

The second input to PLS is the cross-section of book-to-market ratios  $v_{i,t} = \phi_{i,0} + \phi_i' F_t + \varepsilon_{i,t}$  ( $i = 1, \dots, N$ ). These are arranged into the vector  $x_t = (v_{1,t}, \dots, v_{N,t})'$  and stacked as

$$\begin{aligned} X_{(T \times N)} &= [x_1, x_2, \dots, x_T]' \\ &= \iota \phi_0' + F \Phi' + \varepsilon, \end{aligned} \quad (10)$$

with  $\phi_{0(N \times 1)} = [\phi_{1,0}, \phi_{2,0}, \dots, \phi_{N,0}]'$  and  $\Phi_{(N \times K_F)} = [\phi_1, \phi_2, \dots, \phi_N]'$ .

### B. The Estimator

Single-factor PLS can be implemented by the following series of OLS regressions. In the first stage, for each asset  $i$  we run a time-series regression of its book-to-market ratio on the forecast target

$$v_{i,t} = \hat{\phi}_{i,0} + \hat{\phi}_i y_{t+h} + e_{i,t}. \quad (11)$$

The loading estimate  $\hat{\phi}_i$  describes the sensitivity of each  $v_{i,t}$  to the latent factor driving the forecast target. Each  $v_{i,t}$  is a function of only the *expected* portion of future returns and cash flows and is uncorrelated with their unanticipated future shocks. Therefore, first-stage time-series regression coefficients of  $v_{i,t}$  on  $r_{t+1}$  and  $\Delta cf_{t+1}$  describe how each asset's valuation ratio depends on the true expectations  $\mu_t$  and  $g_t$ .

In the second stage, for each period  $t$ , we run a cross-sectional regression of assets' book-to-market ratios on their loadings estimated in the first stage

$$v_{i,t} = \hat{c}_t + \hat{F}_t \hat{\phi}_i + w_{i,t}. \quad (12)$$

Here, the first-stage loadings become the independent variables, and the latent factors  $F_t$  are the coefficients to be estimated. The first two stages exploit the factor nature of the system to draw inferences about the underlying factors. As the factors fluctuate over time, the cross-section of valuation ratios fans out or compresses. If the true factor loadings  $\phi_i$  were known, we could consistently estimate the latent factor time series by simply running cross-section regressions of  $v_{i,t}$  on  $\phi_i$  period by period. Since  $\phi_i$  is unknown, the first-stage regression coefficients provide a preliminary description of how each  $v_{i,t}$  depends on  $F_t$ . This first-stage regression sketches a map from the cross-sectional distribution of value ratios to the latent factors. Second-stage cross-section regressions of  $v_{i,t}$  on first-stage coefficients use this map to produce estimates of the factors at each point in time.

The third step in the filter runs a predictive regression of realized returns or cash flow growth rates on the lagged factors estimated in the second stage. This final 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 or cash flow growth. The ultimate predictor,  $\hat{F}_t$ , is a discerningly constructed linear combination of disaggregated valuation ratios that collapses the cross-section system to its



fundamental driving factors. The  $R^2$  from the final step regression summarizes the predictive power embodied in the cross-section of valuation ratios.<sup>11</sup>

Because the first-stage regression takes an errors-in-variables form, second-stage estimates of the latent expectations  $(\mu_t, g_t)'$  have a multiplicative bias. Since OLS forecasts are invariant to affine transformations of regressors, the third-stage regression of realized returns or growth on the estimated factors delivers consistent estimates of  $\mu_t$  and  $g_t$ .

Kelly and Pruitt (2012) prove that this procedure is consistent: it asymptotically recovers the latent expectations of aggregate market returns and cash flow growth as the number of predictors and time-series observations both become large. Furthermore, they provide asymptotic distribution theory for the third-stage predictive coefficient under weak conditions. The key assumption is that log book-to-market ratios obey a linear factor structure, which is consistent with a range of theories for conditional expected returns (assuming that ROE is also linear in its factors). The remaining assumptions are largely technical, and impose that second moments are finite and probability limits well behaved, that there is limited time-series and cross-sectional autocorrelation among the residuals  $\eta$  and  $\varepsilon$ , and that unanticipated shocks to returns and cash flow growth are asymptotically orthogonal to past expectations.

The general version of our theory accommodates multiple factors in both returns and cash flow growth. In the interest of parsimony, and to highlight the power of our approach compared to the large set of alternative univariate predictors, we assume that returns and cash flow expectations are each driven by a single factor (though the return factor may be different from the growth rate factor). Extending our analysis to extract additional factors from the cross-section of valuation ratios transforms our third step forecasts into multivariate predictive regressions.

Importantly, Kelly and Pruitt (2012) prove that this procedure remains consistent even if there are additional factors that drive the cross-section of value ratios but do not impact expected market returns or dividend growth. They also show that consistency obtains regardless of where the target-relevant factor stands in the PC ordering for the value ratio cross-section.<sup>12</sup> Kelly and Pruitt (2012) demonstrate precisely how PLS avoids this problem by extracting only the forecast-relevant factors while discarding any irrelevant factors.

<sup>11</sup> Kelly and Pruitt (2012) provide a one-step representation of this algorithm:

$$\hat{y} = \bar{y} + \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} (\mathbf{y}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y})^{-1} \mathbf{y}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y}, \quad (13)$$

where  $\mathbf{J}_L \equiv \mathbf{I}_L - L^{-1} \mathbf{1}_L \mathbf{1}_L'$ ,  $\mathbf{I}_L$  is the  $L$ -dimensional identity matrix, and  $\mathbf{1}_L$  is an  $L$ -vector of ones. The  $\mathbf{J}$  matrices are present since each regression step is run including a constant regressor.

<sup>12</sup> Note that this is not generally the case for the PC method. For example, the first PC explains the most covariance among the value ratios, regardless of that component's relationship  $\mu_t$  or  $g_t$ . Only if this component happens to also be the factor driving variation in the forecast target will the first component produce a consistent forecast.



### C. In-Sample versus Out-of-Sample Implementation

Throughout our empirical analysis we consider both in-sample and out-of-sample approaches to implementing our forecasts.

The basic implementation, which uses all available information, is a purely in-sample estimation. First-stage regressions use the full time series of data to estimate factor loadings. Second-stage regressions construct the predictive factor at time  $t$  as a weighted sum of time  $t$  book-to-market ratios, where the weights are based on first-stage regression coefficients. Third-stage predictive regressions are also run in-sample.

In our in-sample analysis, it is possible that first-stage regressions introduce small-sample bias in our predictors since first-stage factor coefficients are based on the full time series. This is analogous to small-sample bias in standard OLS predictive regressions (e.g., Stambaugh (1986) and Nelson and Kim (1993)), which enters into forecasts via estimated predictive coefficients.<sup>13</sup> Consider, for instance, OLS forecasts of  $r_{t+1}$  on some predictor  $z_t$ , where both  $r$  and  $z$  are mean zero. The in-sample estimated coefficient is  $\hat{b}_T = (\frac{1}{T} \sum_{t=0}^{T-1} r_{t+1} z_t) / (\frac{1}{T} \sum_{t=0}^{T-1} z_t^2)$ . The forecast for  $r_{t+1}$  is given by  $\hat{b}_T z_t$ , and thus the targeted observation directly enters into the parameter estimate in its own forecast. For small  $T$ , this can favor false detection of predictability. However, this is not a look-ahead bias because it vanishes as  $T$  becomes large.<sup>14</sup> Instead it is strictly a small-sample phenomenon. Our annual forecasts consist of overlapping monthly observations that span 81 years, while our monthly forecasts use 972 nonoverlapping time-series observations. While neither of these sample sizes is particularly small, we nonetheless take the possibility of small-sample bias very seriously.

Accordingly, the focus of our empirical analysis is a recursive out-of-sample forecast implementation. This procedure has been well studied in the literature (e.g., Clark and McCracken (2001) and Goyal and Welch (2008)). The main idea is to run all regressions on training samples that exclude the return or cash flow observation that is ultimately forecasted. Because our out-of-sample results are important to the overall contribution of the paper, we describe this procedure in detail.

Let time indices represent months. Consider a forecast for the return  $r_{\tau+12}$  that is realized over the 12-month period  $t+1$  to  $t+12$ .<sup>15</sup> First-stage regressions have annual returns on the right-hand side, so the regression takes the form

$$v_{i,\tau} = \hat{\phi}_{i,0} + \hat{\phi}_i r_{\tau+12} + e_{i,\tau}.$$

<sup>13</sup> Small-sample bias is not unique to PLS, but arises in many forecast methodologies, including OLS, PC, or maximum likelihood estimation of a Kalman-filtered state space.

<sup>14</sup> An example of look-ahead bias is using a predictor that is constructed as a centered moving average of the forecast target. In this case, forward-looking information in the predictor favors false detection of predictability, and this bias does not attenuate even in very large samples.

<sup>15</sup> One-month return forecasts are constructed analogously. The 1-month case is straightforward since it involves no data overlap.

A properly constructed forecast can only use information known through month  $t$ . The latest return that may be used on the right-hand side is that from  $t - 11$  to  $t$ . The latest value ratio used on the left-hand side will therefore be  $v_{i,t-12}$ . Thus, first-stage coefficient estimates  $\hat{\phi}_i$  are in the time  $t$  information set since they use monthly returns  $\{r_{13}, \dots, r_t\}$  and monthly value ratios  $\{v_{i,1}, \dots, v_{i,t-12}\}$ ,  $i = 1, \dots, N$ .

The second-stage cross-section regressions are run for months  $1, \dots, t - 12$ , and  $t$ . The data for these regressions are value ratios up to date  $t$ , and  $\hat{\phi}_i$ , which is  $t$ -measurable. The factor estimates  $\hat{F}_1, \dots, \hat{F}_{t-12}$  are used for the third-stage regression, and the factor estimate for month  $t$  at  $\hat{F}_t$  is used to construct the out-of-sample forecast (not as an observation in the third-stage regression).

The third-stage regression takes the form

$$r_{t+12} = \beta_0 + \beta \hat{F}_t + u_{t+12}.$$

The latest return entering this regression is from  $t - 11$  to  $t$ , and the last predictor observation is  $\hat{F}_{t-12}$ . The resulting third-stage coefficient estimates  $\hat{\beta}_0$  and  $\hat{\beta}$  are therefore also month  $t$ -measurable. Finally, we construct our time  $t$  forecast of the  $t + 1$  to  $t + 12$  return as  $\hat{\beta}_0 + \hat{\beta} \hat{F}_t$ . In summary, all inputs to this forecast are constructed using data that are observable no later than time  $t$ , so this forecast is genuinely out-of-sample.

#### D. Inference

Inference for in-sample 1-month forecasts is based on the asymptotic distributions for PLS estimates derived in Kelly and Pruitt (2012), or White (1980) standard errors in the case of the alternative predictors we consider. Because 12-month forecasts use overlapping monthly data we must adjust our standard errors to reflect the dependence that this introduces into forecast errors. We do this in three ways. First, we calculate Kelly and Pruitt (2012) standard errors using only the nonoverlapping forecast errors.<sup>16</sup> As a second alternative, we calculate Hodrick (1992) standard errors, which explicitly account for the moving average structure that overlap introduces into residuals. Third, we report Newey and West (1987) standard errors with 12 lags to account for overlap-induced serial correlation among residuals. In our empirical analysis, results for all of these test statistics are similar.

We conduct out-of-sample inference with the “encompassing” forecast test ENC-NEW derived by Clark and McCracken (2001). This statistic has become widely used in the forecasting literature, and tests the null hypothesis that two predictors provide the same out-of-sample forecasting performance. When we report this statistic, we are testing the denoted predictor versus the historical mean of the target series. We report significance levels as found from Clark

<sup>16</sup> This can be done using annual observations with 12 possible year-end months. We calculate Kelly and Pruitt (2012) test statistics using nonoverlapping annual data for each year-end (January to December, February to January, etc.) and report the  $p$ -value for the median of 12 test statistics. In practice, any choice of year-end month yields very similar statistics.

and McCracken's (2001) appendix tables, where critical values for the 0.10, 0.05, and 0.01 levels are provided. The notation " $< x$ " represents the smallest significance level  $x$  for which the encompassing test statistic exceeds the critical value. When evaluating overlapping forecast errors (as we do for 12-month returns, dividend growth, and earnings growth forecasts) we use Newey-West standard errors with 12 lags to consistently estimate the appropriate asymptotic variance in the denominator of ENC-NEW, as suggested in Clark and McCracken (2005). Out-of-sample results reported in tables are based on a 1980 sample split (except for international data, which are split at 1995 owing to its more recent start date). We later report out-of-sample results over a wide range of alternative split dates to demonstrate the robustness of our findings to different test samples.

To evaluate forecasting fit, we calculate the predictive  $R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}$ , which for our PLS forecasts is equal to the  $R^2$  of the third-stage univariate regression. The out-of-sample  $R^2$  lies in the range  $(-\infty, 1]$ , where a negative number means that a predictor provides a less accurate forecast than the target's historical mean.

### E. Data

Our central empirical analysis examines market return and cash flow growth predictability by applying PLS to different cross-sections of valuation ratios. We use book-to-market ratios for Fama and French's (1993) size- and value-sorted portfolios (in which U.S. stocks are divided into 6, 25, or 100 portfolios). Data files from Ken French's website report monthly portfolio-level market equity value and annual portfolio book equity value, which we use to construct a monthly panel of portfolio book-to-market ratios. A book-to-market ratio in month  $t$  uses a portfolio's total market capitalization at the end of month  $t$  and the latest observable annual book equity total for the portfolio. We assume that portfolio book equity in calendar year  $Y$  becomes observable after June of year  $Y + 1$  following Fama and French (1993).

We also consider a variety of alternative value ratio panels. First, we explore the usefulness of individual stock-level value ratios data for predicting future market returns. We also consider price-dividend ratios for size- and value-sorted portfolios in place of book-to-market ratios. Finally, we take our analysis to international data, using the country-level portfolio valuation ratios of Fama and French (1998).

Our focus is on the 1930 to 2010 sample for U.S. data.<sup>17</sup> The international sample is available from 1975 to 2010. U.S. market returns and dividend growth are for the CRSP value-weighted index. Individual stock data are from CRSP and Compustat. U.S. and international portfolio data are from Ken French's website. Alternative predictors are from Amit Goyal's website.

<sup>17</sup> The last 12-month return in our sample is realized in December 2011.

**Table I**  
**Market Return Predictions (1930 to 2010)**

We report in-sample and out-of-sample percentage  $R^2$  for PLS forecasts of 1-year and 1-month market returns from 1930 to 2010. The sets of predictor variables are 6, 25, and 100 book-to-market ratios of size- and value-sorted portfolios. Our out-of-sample procedure splits the sample in 1980, uses the pre-1980 period as a training window, and recursively forecasts returns beginning in January 1980 (results for a wide range of alternative sample splits are shown in Figures 1 and 2). We also report  $p$ -values of three different in-sample test statistics. The first is the asymptotic predictive loading  $t$ -statistic from Kelly and Pruitt (2012), denoted “KP” in the table. For annual returns, this is calculated on every nonoverlapping set of residuals as described in the text. For annual returns we also report Hodrick (1992) and Newey and West (1987)  $p$ -values. For out-of-sample tests we report  $p$ -values for Clark and McCracken’s (2001; denoted “CM” in the table) ENC-NEW encompassing test statistic. This tests the null hypothesis of no forecast improvement over the historical mean. For annual returns we follow Clark and McCracken (2005) and use Newey-West standard errors with 12 lags.

	Panel A: 1 Year				Panel B: 1 Month	
	$R^2$ (%)	$p$ (KP/CM)	$p$ (Hodrick)	$p$ (NW)	$R^2$ (%)	$p$ (KP/CM)
6 Portfolios						
In-sample	7.72	0.015	0.012	0.004	0.60	0.031
Out-of-sample	5.81	<0.010			0.65	<0.050
25 Portfolios						
In-sample	13.50	0.004	0.004	0.001	1.12	0.003
Out-of-sample	3.49	<0.010			0.77	<0.010
100 Portfolios						
In-sample	18.05	0.003	0.010	<0.001	2.38	0.005
Out-of-sample	13.07	<0.010			0.93	<0.050

### III. Empirical Results

#### A. Market Return Predictability

Our main empirical analysis evaluates the predictability of aggregate market returns using the cross-section of book-to-market ratios. We directly estimate our model of the cross-section system described in Section I via PLS. Table I presents return forecasting results based on 6, 25, and 100 book-to-market ratios of size- and value-sorted portfolios of U.S. stocks (Fama and French (1993)). We consider two different forecasting horizons—1 month and 1 year—and report findings for both in-sample and out-of-sample forecasts.

Panel A of Table I shows that a single factor extracted via PLS demonstrates a striking degree of predictability for 1-year returns. The in-sample implementation generates a predictive  $R^2$  reaching 7.7%, 13.5%, and 18.1% based on 6, 25, and 100 portfolio book-to-market ratios, respectively ( $p$ -values below 0.015 in all cases).<sup>18</sup> Out-of-sample PLS forecasts are similarly powerful, delivering

<sup>18</sup> The third-stage coefficients are 0.08, 0.13, and 0.17 for the 6, 25, and 100 portfolio cases, respectively, when the first-stage coefficients are as plotted in Figure 6. The  $p$ -values for these estimates are shown in Table I and derived from Kelly and Pruitt’s (2012) asymptotic  $t$ -statistics. Note that these coefficients are identified only up to a normalization (see Stock and Watson (2002)).

an  $R^2$  of 5.8%, 3.5%, and 13.1% for 6, 25, and 100 portfolios. These are large economic magnitudes for out-of-sample prediction, comparable to in-sample results from commonly studied predictors such as the aggregate book-to-market ratio or Lettau and Ludvigson's (2001) consumption-wealth ratio. Each of these out-of-sample results is statistically significant at the 0.01 level or better based on Clark-McCracken tests.

Panel B reports forecasting results for 1-month returns. The monthly in-sample  $R^2$  is 0.6%, 1.1%, and 2.4% based on a single linear combination of 6, 25, or 100 portfolio book-to-market ratios, respectively, all of which are significant at the 0.05 level.<sup>19</sup> Out-of-sample 1-month return forecasts are also significant at the 0.05 level or better, with an  $R^2$  of 0.9% based on the 100-portfolio cross-section. At the monthly frequency, an out-of-sample  $R^2$  of 0.9% has large economic significance. A heuristic calculation suggested by Cochrane (1999) shows that the Sharpe ratio ( $s^*$ ) earned by an active investor exploiting predictive information (summarized by the regression  $R^2$ ) and the Sharpe ratio ( $s_0$ ) earned by a buy-and-hold investor are related by  $s^* = \sqrt{\frac{s_0^2 + R^2}{1 - R^2}}$ . Campbell and Thompson (2008) estimate a monthly equity buy-and-hold Sharpe ratio of 0.108 using data back to 1871. Therefore, an out-of-sample predictive  $R^2$  of 0.9% implies that an active investor exploiting our approach could achieve a Sharpe ratio improvement of roughly 33% over a buy-and-hold investor, using real-time information in portfolio book-to-market ratios.

How do our market return forecasts compare with predictors proposed in earlier literature? Table II compares the predictive accuracy of our approach with an extensive collection of alternative predictors considered in the literature. In particular, we explore forecasts from 16 predictors studied in a recent return predictability survey by Goyal and Welch (2008). The table considers both in-sample and out-of-sample forecasts of market returns over horizons of 1 year and 1 month from each regressor individually. Among the alternatives, the best univariate forecasts at the annual horizon (Panel A) are achieved by the consumption-wealth ratio (cay; Lettau and Ludvigson (2001)), which delivers an in-sample  $R^2$  of 14.4% and an out-of-sample  $R^2$  of 2.7% (statistically significant at the 0.01 level).<sup>20</sup> Other successful out-of-sample predictors include the cross-section premium (csp) of Polk, Thompson, and Vuolteenaho (2006), the term spread (tms), and the long-term government bond return (ltr). All of these are dominated by the single PLS factor extracted from portfolio-level book-to-market ratios.<sup>21</sup>

<sup>19</sup> The third-stage coefficients are 0.006, 0.011, and 0.030 for the 6, 25, and 100 portfolio cases, respectively.

<sup>20</sup> Specifically, this is the "cayp" variable that Goyal and Welch (2008) discuss at some length, noting that its construction uses information from the full sample. The highest frequency at which cay is available is quarterly. We therefore take each observation to represent the quarter's last monthly observation and treat the other months as missing.

<sup>21</sup> Harvey (1989) constructs expected returns from several predictors and finds strong in-sample and out-of-sample predictive performance for the NYSE value-weighted return. Most of his predictors are included in the Goyal and Welch (2008) data set: the junk bond premium (dfy), the

**Table II**  
**Market Return Predictions: Comparison with Common Alternative Predictors**

We report in-sample and out-of-sample percentage  $R^2$  for forecasts of 1-year and 1-month market returns from 1930 to 2010. PLS forecasts use 100 book-to-market ratios of size- and value-sorted portfolios. Results for alternative predictors use data from Goyal and Welch (2008) and include the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-section premium (csp), 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 return (ltr), the earnings-to-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 consumption-wealth-income ratio (cay). Additionally we consider the first or the first three PCs extracted from the 100 portfolio book-to-market ratios (in this case in-sample statistical significance is from an  $F$ -test). Not all alternative predictors are available for the full 1930 to 2010 sample. See Table I for details about test statistics and the out-of-sample procedure.

	Panel A: 1 Year					Panel B: 1 Month			
	In-Sample			Out-of-Sample		In-Sample		Out-of-Sample	
	$R^2$ (%)	$p$ (Hodrick)	$p$ (NW)	$R^2$ (%)	$p$ -value	$R^2$ (%)	$p$ -value	$R^2$ (%)	$p$ -value
100 Portfolios	18.05	0.010	<0.001	13.07	<0.010	2.38	0.005	0.93	<0.050
dfy	0.49	0.593	0.569	0.35	–	0.05	0.776	–0.39	–
infl	0.00	0.872	0.972	–0.20	–	0.04	0.636	–0.15	–
svar	0.02	0.855	0.883	–0.92	–	0.09	0.648	–0.03	–
csp	0.36	0.400	0.492	1.30	–	0.61	0.043	–0.11	–
de	0.16	0.833	0.752	–10.23	–	0.10	0.540	–1.40	–
lty	0.68	0.383	0.309	–1.08	–	0.01	0.706	–0.83	–
tms	1.47	0.141	0.083	1.81	<0.050	0.12	0.262	–0.22	–
tbl	0.06	0.782	0.774	–3.75	–	0.00	0.924	–1.00	–
dfr	0.00	0.674	0.931	–0.28	–	0.13	0.514	0.76	<0.010
dp	3.16	0.108	0.076	–3.82	–	0.18	0.325	–7.03	–
dy	3.40	0.098	0.060	–7.30	–	0.29	0.217	–0.31	–
ltr	0.77	0.005	<0.001	0.32	<0.050	0.40	0.055	–4.81	–
ep	4.81	0.017	0.016	–10.39	–	0.48	0.041	–3.34	–
bm	8.83	0.003	<0.001	–20.29	–	0.71	0.117	–3.18	–
ntis	8.68	0.081	0.055	–52.73	–	0.56	0.097	–2.34	–
cay	14.42	0.004	<0.001	2.65	<0.010	1.31	0.016	–1.97	–
pc1	1.98	0.300	0.277	–2.56	–	0.10	0.599	–2.71	–
pc123	11.31	<0.001	<0.001	–2.50	–	0.64	0.097	–2.69	–

Table II also reports forecasting results using the first three PCs extracted from the cross-section of 100 portfolio book-to-market ratios. PCs fail to demonstrate any significant return forecasting power in-sample or out-of-sample.

Predictions of 1-month returns (Table II, Panel B) tell the same story as annual forecasts. Our procedure is the dominant in-sample univariate predictor ( $R^2 = 2.4\%$ ,  $p = 0.005$ ), with only cay and the log earnings-price ratio (ep) as

dividend yield (dy), the Treasury bill rate (tbl), and a term premium (tms). As Table III reports, individually these variables have mixed out-of-sample performance. We find that the other Harvey variables not included in Goyal–Welch, namely, a January dummy and the lagged excess market return, have no additional out-of-sample forecast power.



**Table III**  
**Characteristic-Sorted Portfolio Returns Out-of-Sample  $R^2$**

We forecast monthly and annual returns on four sets of characteristic-sorted portfolios using our PLS methodology based on the cross-section of book-to-market ratios for 100 Fama-French portfolios. Data are for 1930 to 2010 with a 1980 out-of-sample split date. We report out-of-sample forecasting percentage  $R^2$  and  $p$ -values for Clark and McCracken's (2001) ENC-NEW encompassing test statistic (\*if  $p < 0.100$ , \*\* if  $p < 0.050$ , and \*\*\* if  $p < 0.010$ ).

Portfolio	1 Year	1 Month	Portfolio	1 Year	1 Month
Panel A: Value Portfolios			Panel B: Size Portfolios		
1 Growth	9.08***	0.10*	1 Small	-17.66	-5.76
2	12.11***	0.36**	2	-0.62	-1.75
3	6.80***	-0.86	3	5.96***	-0.67
4	3.93***	-1.92	4	6.95***	0.20**
5 Value	4.56***	-1.98	5 Big	11.94***	0.16*
Panel C: Momentum Portfolios			Panel D: Industry Portfolios		
1 Past losers	0.16	-0.96	Cons. nondur.	17.07***	1.35***
2	2.25***	-2.04	Cons. dur.	2.31**	-2.64
3	11.39***	-0.22	Manufacturing	7.46***	-1.12
4	9.28***	-1.26	Energy	0.68**	-0.34
5	7.50***	-0.95	Technology	2.42*	-0.32
6	12.21***	0.13**	Telecom	9.31***	0.61**
7	14.56***	0.03*	Retail	11.11***	0.15
8	9.81***	0.11**	Healthcare	6.56***	-0.05
9	15.22***	1.17***	Utilities	2.76***	-1.18
10 Past winners	4.40**	-0.32	Other	8.59***	-0.74

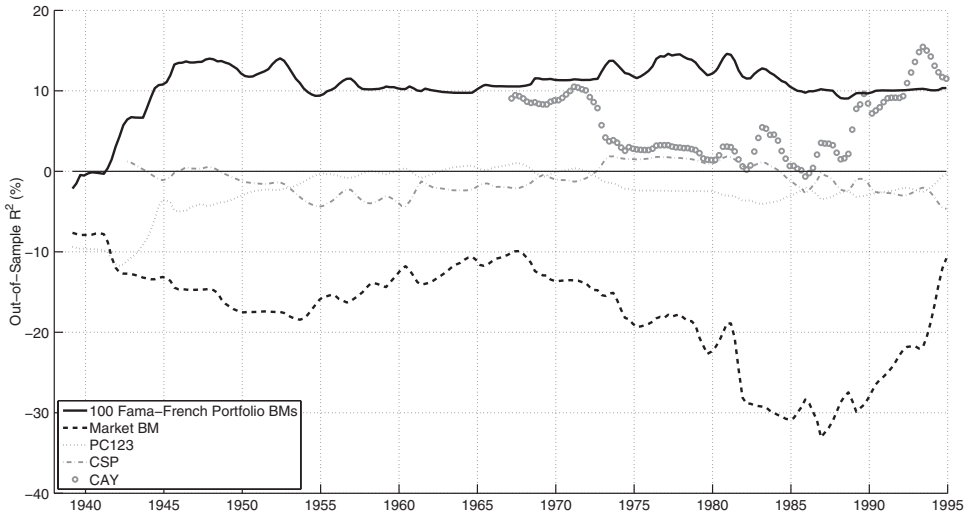
the other predictors with significant in-sample forecasting at the 0.05 level. Out-of-sample, only our procedure ( $R^2 = 0.9\%$ ,  $p < 0.050$ ) and the default rate (dfr) provide significant positive results (in-sample predictive power of dfr is small and insignificant). In summary, our single PLS factor derived from the cross-section of book-to-market ratios is the only predictor to exhibit significant performance both in-sample and out-of-sample for 1-month returns.

### A.1. Predicting Returns on Characteristic Portfolios

According to our model, the cross-section of value ratios should also predict performance of other equity portfolios. To test this, we forecast returns of portfolios that have been formed on the basis of various stock characteristics. In particular, we look at quintile portfolios formed on stock book-to-market ratios, market capitalization quintile portfolios, momentum decile portfolios, and 10 industry portfolios. Return data for each of these series are from Ken French's website.

Table III reports the out-of-sample  $R^2$  from forecasts of monthly and annual portfolio returns using book-to-market ratios of the 100 Fama-French portfolios. We draw three conclusions from these results. First, there is positive





**Figure 1. Out-of-sample  $R^2$  by sample split date, 1-year returns.** Forecasts are based on a single PLS factor from 100 book-to-market ratios of size- and value-sorted portfolios, the aggregate book-to-market ratio, the cross-section premium of Polk, Thompson, and Vuolteenaho (2006), the consumption-wealth ratio of Lettau and Ludvigson (2001), and the first three PCs of the 100 book-to-market ratio cross-section.

out-of-sample predictability of portfolio returns at the 1-year horizon based on 100 value ratios. The only exceptions are returns on the two smallest size quintile portfolios and the first momentum decile (biggest past losers). If a low-dimension factor model describes asset returns, it is perhaps unsurprising that the portfolios that deviate most from such a model would be small stocks whose mispricings are difficult to arbitrage. Second, returns of growth stocks are more predictable than value stocks. Finally, evidence of return predictability is weaker at the monthly horizon across portfolios, though we continue to find significant 1-month out-of-sample predictability for growth stocks and large stocks.

### A.2. Varying Out-of-Sample Sample Splits

Above we report out-of-sample forecasting tests based on a 1980 sample split date, but recent forecast literature suggests that sample splits themselves can be data-mined (see Hansen and Timmermann (2011) and Inoue and Rossi (2011)). To demonstrate the robustness of out-of-sample forecasts to alternative sample splits, Figure 1 plots out-of-sample annual return predictive  $R^2$  as a function of the sample split for a variety of predictors. We consider a sample split as early as 1940, which uses only 10 years of data as a training sample. The latest split we consider is 1995, which uses a 65-year training sample. The figure shows that our procedure consistently outperforms alternative predictors across sample splits. The aggregate book-to-market ratio consistently achieves

an out-of-sample  $R^2$  below  $-10\%$ . Forecasts using cays are competitive in only a small subset of the sample splits. The first three PCs jointly have poor out-of-sample performance, as do forecasts that use only the first PC (not shown due to close similarity with results from first three PCs). The remaining predictors fail to consistently demonstrate out-of-sample predictability across various split dates. The cross-section premium of Polk, Thompson, and Vuolteenaho (2006) hovers around zero for all sample splits. Our procedure begins with weak forecasting power when it has very few observations for training, but overcomes this challenge once it sees more than 10 years of data: across sample splits from 1945 to 1995 our method consistently achieves an out-of-sample  $R^2$  that exceeds  $10\%$ .<sup>22</sup>

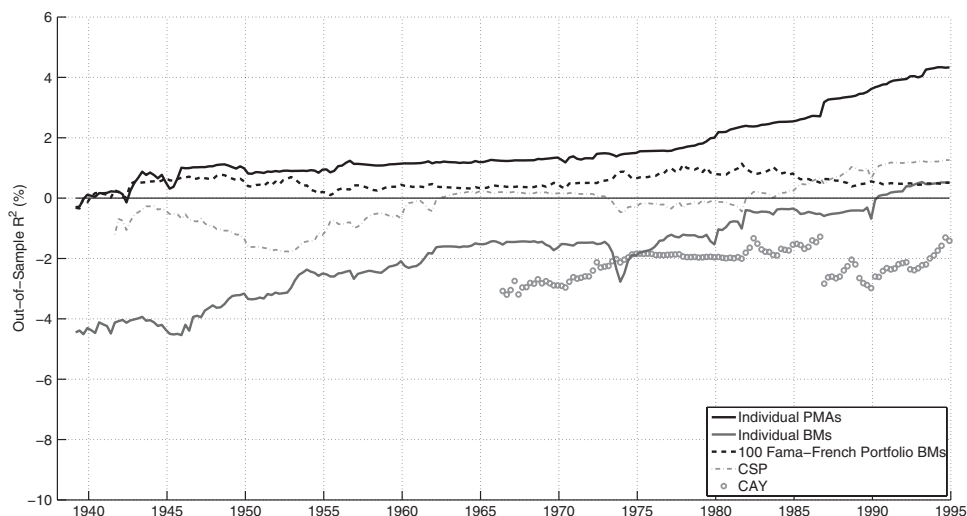
### A.3. Forecasts from Individual Stock Valuation Ratios

We next investigate the usefulness of information in individual stock valuation ratios for predicting market returns. Polk, Thompson, and Vuolteenaho (2006) is the benchmark study for combining individual firm data into a market return prediction. Their paper emphasizes the difficulty in working with noisy firm-level data, such as book value or cash flows, which may distort valuation ratios and complicate extraction of market expectations from the cross-section. To address this challenge they use a series of prefiltering adjustments to firms' valuation ratios and robust statistics. These adjustments include relying on ordinal ranks rather than cardinal values, value-weighting observations, and censoring extreme observations.

One alternative solution to this problem, suggested by Fama and MacBeth (1973) and used in our main analysis earlier, is to combine individual stocks into portfolios. In this section, we are interested in drilling deeper to investigate how much we can learn about aggregate market expectations directly from individual stock data.

Miller and Scholes (1982) and Fama and French (1988) propose a different approach to dealing with noise in individual stock valuation ratios. They suggest that, rather than using infrequent and potentially mismeasured balance sheet data, it may be beneficial to omit fundamentals information entirely and focus only on the price portion of the valuation ratio. In addition to applying our PLS approach to raw firm-level book-to-market ratios (which will be excessively noisy according to the arguments in Polk, Thompson, and Vuolteenaho (2006) and others), we apply our method to a value ratio that omits balance sheet data altogether as recommended by Miller and Scholes and Fama and French. This ratio, called the price-to-moving-average (PMA) ratio, divides a firm's month-end share price by the moving average of its monthly prices over the previous 3 years. We then use the log of this ratio as  $v_{i,t}$  in system (7).

<sup>22</sup> We also consider out-of-sample cross-validation tests of our method's forecasting power. These results are reported in the Internet Appendix.



**Figure 2. Out-of-sample  $R^2$  by sample split date, 1-month returns.** Forecasts are based on a single PLS factor from the entire cross-section of U.S. stocks' PMA ratios or book-to-market ratios, a single PLS factor from 100 book-to-market ratios of size- and value-sorted portfolios, the cross-section premium of Polk, Thompson, and Vuolteenaho (2006), and the consumption-wealth ratio of Lettau and Ludvigson (2001).

Individual firm characteristics change over time. To address stock-level parameter instability, we estimate the model using a rolling 5-year estimation window, and only include stocks that have no missing observations in the estimation window.<sup>23</sup>

We focus on monthly returns to directly compare with the benchmark of Polk, Thompson, and Vuolteenaho (2006), who study monthly forecasts. Figure 2 reports the out-of-sample monthly return forecasting  $R^2$  from individual stock PMA ratios across a range of sample splits. It also plots the  $R^2$  for Polk, Thompson, and Vuolteenaho's csp variable. Our single PLS factor extracted from the cross-section of individual stock PMA ratios consistently produces a positive out-of-sample  $R^2$ , rising above 1% in the mid 1960s and exceeding 2% per month by the mid 1980s. It uniformly outpredicts csp and cay. For comparison, we also plot results from the PLS factor extracted from 100 portfolio book-to-market ratios.

$R^2$  for a single PLS factor extracted from individual stock book-to-market ratios is also plotted in Figure 2. For small training samples these forecasts produce a negative  $R^2$ , presumably due to the noisiness of firm-level balance sheet data. In early sample splits the book-to-market ratios are not only dominated by forecasts based solely on firm-level prices, but also by csp, whose clever modifications mitigate the influence of noise in balance sheet data. The

<sup>23</sup> Results are unchanged when we loosen this requirement and use stocks that have at least 36 nonmissing observations in the most recent 60 months.

individual stock book-to-market ratio  $R^2$  series trends upwards as the training window expands. This happens because the number of individual stocks is also increasing (from a few hundred in the early sample to several thousand at the end of the sample; see the Internet Appendix). The  $R^2$  from firm-level book-to-market ratios increases more rapidly than other forecasters as the cross-section expands, suggesting that PLS forecasts learn relatively quickly as more data become available to overcome the initial noise-induced  $R^2$  deficit. By the mid 1980s the raw book-to-market  $R^2$  series intersects then exceeds that of csp. However, it never reaches the high degree of predictability demonstrated by our PLS approach applied to the cross-section of price-only valuation ratios.

We conclude from this analysis that our forecasting approach continues to demonstrate strong predictive power when we rely on the cross-section of individual stock data.

#### A.4. Forecasting with Price-Dividend Ratios

As another robustness check, we consider the ability of a cross-section of alternative valuation ratios to forecast market returns. In Section I we note that the Campbell and Shiller (1988) present value identity produces a factor model for the cross-section of log price-dividend ratios in direct analogy with equation (7) under similar assumptions. Thus far, our analysis has focused on book-to-market ratios to avoid the lack of dividend payments (and hence undefined price-dividend ratios) for a substantial fraction of U.S. firms.<sup>24</sup> While concerns about declining numbers of dividend-paying firms are partly mitigated by portfolio aggregation, in many cases portfolios can be dominated by dividend nonpayers, resulting in an erratic and highly inflated price-dividend ratio for that portfolio. In order to develop well-behaved portfolio price-dividend ratios, we form our own sets of 6, 25, and 100 portfolios on the basis of underlying firms' market equity and book-to-market ratios, with the key difference that we exclude dividend nonpayers. When forming portfolios, we only assign a stock to a portfolio in month  $t$  if it paid positive dividends in the 12 months prior to  $t$ .<sup>25</sup> This greatly increases 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. Dividend-paying behavior is highly persistent among U.S. firms, so that a firm having paid dividends in the past 12 months strongly predicts that it will pay dividends in the subsequent 12 months.

Market return forecasts based on a single PLS factor extracted from the cross-section of portfolio price-dividend ratios demonstrate high predictive accuracy, on par with our results using book-to-market ratios. In-sample annual return

<sup>24</sup> The fraction of firms that paid dividends in 1946, 1980, and 2008 is 86%, 64%, and 36%, though these fractions are substantially higher, 97%, 93%, and 76%, when weighted by market capitalization.

<sup>25</sup> We use simultaneous two-way sorts, rebalance portfolios monthly, and, most importantly, strictly preclude look-ahead information in portfolio construction, as is the case in the original Fama and French (1993) portfolios, using data for 1930 to 2009.

$R^2$ s for 6, 25, and 100 portfolios are 8.3%, 10.6%, and 33.9%, and Kelly-Pruitt, Hodrick, and Newey-West  $t$ -statistics are all significant at least at the 0.001 level. Out-of-sample  $R^2$ s are 13.7%, 4.8%, and 9.6%, respectively, all significant at the 0.010 level or better according to the Clark-McCracken test.

#### A.5. Forecasting Outside the United States

Next, we examine whether our return predictability findings hold internationally. To do so, we forecast returns on the value-weighted aggregate world portfolio (excluding the United States) by applying PLS to an international cross-section of non-U.S., country-level valuation ratios. Monthly data for country-level portfolios are available beginning in 1975 and follow the construction described in Fama and French (1998). This sample, based on data from MSCI, sorts equities from each country into high- and low-value portfolios. Countries covered in the sample are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Italy, Japan, Malaysia, the Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, Switzerland, and the United Kingdom. We use cum- and ex-dividend returns on these portfolios to calculate price-dividend ratios of the two portfolios in each country, resulting in a cross-section of 42 portfolio price-dividend ratios.<sup>26</sup> This cross-section is used to forecast the return on an international equity index (excluding the United States), which is a portfolio of country-level index returns in these 21 countries (portfolio weights are determined by a country's weight in the MSCI EAFE index).

We find that the world equity index return is highly predictable by country-level value ratios. The monthly out-of-sample  $R^2$  is 1.5% for a 1995 sample split, for which Clark and McCracken's (2001) test statistic is significant at the 0.01 level.<sup>27</sup> Figure 3 shows that this strong out-of-sample performance is robust to a wide range of sample splits. Because the data begin in 1975, we focus on monthly returns and evaluate sample splits from 1985 to 2000. As in the U.S. sample, the out-of-sample  $R^2$  is consistently positive once the procedure is given a training sample of at least 10 years, and increases as the training sample expands. The success of a single PLS factor drawn from the cross-section of value ratios in an international sample lends further confidence to the robustness of our findings.

<sup>26</sup> French's data include price and dividend data at the monthly frequency, and aggregate country book-to-market ratios at the annual frequency. Because the sample begins, at the earliest, in 1975 (some countries have an even shorter sample), for meaningful out-of-sample analysis we conduct our forecasting analysis at the monthly frequency and therefore rely on price-dividend ratios rather than book-to-market ratios as our predictors. A comparison of *annual* price-dividend and book-to-market ratios suggests that the two series are highly similar at the country level. The median correlation between the two ratios is 86% (mean of 78%) across the 21 countries we study.

<sup>27</sup> The in-sample  $R^2$  is 5.3% with a Kelly-Pruitt  $p$ -value below 0.001.

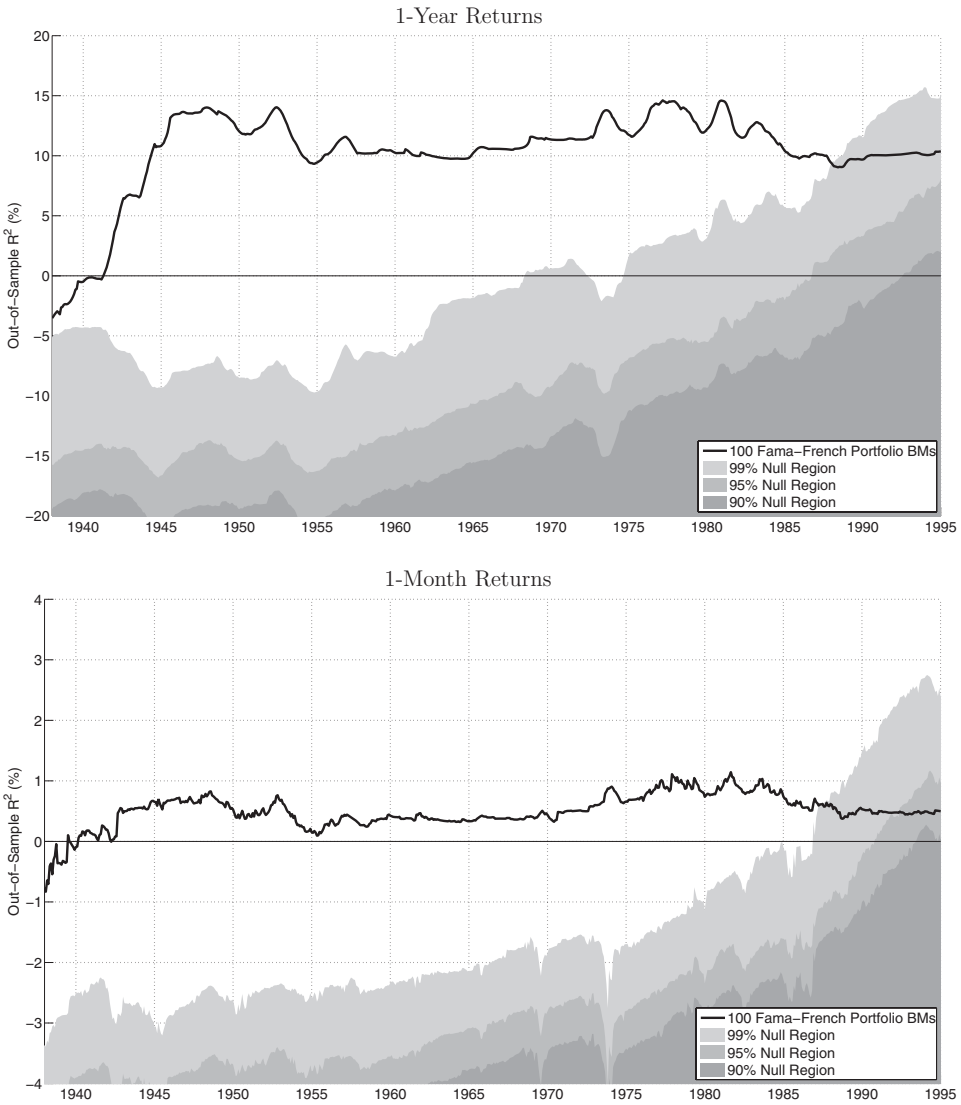


**Figure 3. Out-of-sample  $R^2$  by sample split date, ex-U.S. 1-month returns.** Forecasts of 1-month ex-U.S. stock returns are based on a single PLS factor from 42 price-dividend ratios of high value and low value portfolios across 21 countries (Fama and French (1998)). See Section III.A.6 for the list of countries.

#### A.6. Placebo Tests

To demonstrate that our procedure does not mechanically generate return predictability, we run a series of placebo tests. If our method is subject to a mechanical bias, this would imply that we successfully forecast returns even if the predictors have zero true forecasting power. We can test for this possibility by attempting to predict actual market returns using simulated predictor variables known to have no true return forecasting ability. We conduct our placebo test as follows. In each simulation we generate 100 AR(1) processes that are specified to have the same mean, variance, and autocorrelation as the Fama-French 100 portfolio book-to-market ratios. Innovations in the simulated data are produced by a random number generator so they are independent of the true U.S. market return data. We run our recursive out-of-sample forecast for actual returns using the simulated predictor data and record the out-of-sample  $R^2$  across sample splits. We repeat this test for 1,000 simulated predictor panels. Figure 4 plots the 90%, 95%, and 99% confidence intervals for the out-of-sample  $R^2$  at each sample split, as well as the actual out-of-sample  $R^2$  using the Fama-French portfolios.

The figure shows that placebo predictors broadly generate negative out-of-sample  $R^2$ . An  $R^2$  of zero would be the expected asymptotic result in the absence of look-ahead bias. Because our placebo tests use finite samples, estimated predictive parameters are generally nonzero and thus can perform even worse



**Figure 4. Placebo test out-of-sample  $R^2$  by sample split.** The figure shows confidence intervals from 1,000 predictor panels simulated to have no true predictive power for returns. We report these intervals for each out-of-sample split date considered in Figure 1. Also shown are the out-of-sample  $R^2$  by sample split from 1-year and 1-month forecasts using actual data as plotted in Figures 1 and 2, respectively.

than a naive historical mean forecast. The figures show that as the estimation window grows, the confidence intervals begin to widen. This is because the test sample becomes smaller as the estimation sample grows. For this reason, the latest sample split we consider is 1995, which leaves a 15-year test window.



Not a single sample is generated in which there is consistent positive  $R^2$  across sample splits like we find in the data. Our conclusion from the placebo analysis is that our results are unlikely to be driven by a mechanical bias.<sup>28</sup>

### B. Cash Flow Growth Predictions

Thus far we have focused on forecasts of aggregate market returns. Asset prices depend not only on discount rates, but also on expectations about assets' future cash flows. Hence, it is important for our understanding of asset pricing to also investigate how much information valuation ratios contain about the market's expectations of future cash flow growth. Our analysis focuses on forecasting dividend growth, since this quantity has been at the center of growth forecasting in the asset pricing literature (see Ball and Watts (1972), Campbell and Shiller (1988), Cochrane (1992), Fama and French (2000), Lettau and Ludvigson (2005), Kojien and Van Nieuwerburgh (2011), and Lacerda and Santa-Clara (2010)).

Aggregate dividend growth is calculated from the universe of CRSP stocks. Our analysis focuses on annual growth data in order to avoid spurious predictability arising from within-year seasonality. Table IV, Panel A reports results from our PLS approach to forecasting annual aggregate U.S. dividend growth based on 6, 25, and 100 Fama-French portfolio book-to-market ratios in the 1930 to 2010 sample. These results show little evidence of out-of-sample predictability for dividend growth. However, recent research by Chen (2009) and Kojien and Van Nieuwerburgh (2011) documents that both return and dividend growth predictability results tend to be significantly influenced by Depression-era observations, which are far more volatile than the rest of the CRSP sample (even more so than the 2008 to 2009 financial crisis).<sup>29</sup> When evaluating predictability with a least-squares criterion (including OLS and PLS), it is useful to check that predictability is not being driven by a few high-volatility observations.

Panel B shows that, when we exclude these Depression-era observations, out-of-sample dividend growth predictability improves dramatically. Post-Depression results are also stable across many remaining subsamples, as shown in the out-of-sample  $R^2$  plot of Figure 5.

This instability with respect to the Depression is unique to dividend growth forecasts. Table V shows that omitting Depression-era observations has little effect on our return predictability results. The 100 portfolio out-of-sample  $R^2$  is 8.3% in 1940 to 2010, and significant at the 0.01 level or better. Furthermore, in-sample versus out-of-sample results remain highly consistent with each other.

<sup>28</sup> The Internet Appendix reports an additional robustness test designed to detect mechanical look-ahead bias. We introduce 1-, 2-, and 3-year gaps between training and test windows in our out-of-sample forecasts and find that results are nearly identical to our main findings.

<sup>29</sup> Many return and dividend growth forecasting studies exclude the Great Depression from their forecast samples. Recent examples include van Binsbergen and Kojien (2010), Lacerda and Santa-Clara (2010), and Cochrane (2011).

**Table IV**  
**Market Cash Flow Predictions**

We report in-sample and out-of-sample percentage  $R^2$  for forecasts of 1-year market dividend growth. PLS forecasts use 6, 25, and 100 book-to-market ratios of size- and value-sorted portfolios. The out-of-sample procedure splits the sample in 1980 (results from a wide range of alternative sample splits are shown in Figure 5). See Table I for details about test statistics and the out-of-sample procedure.

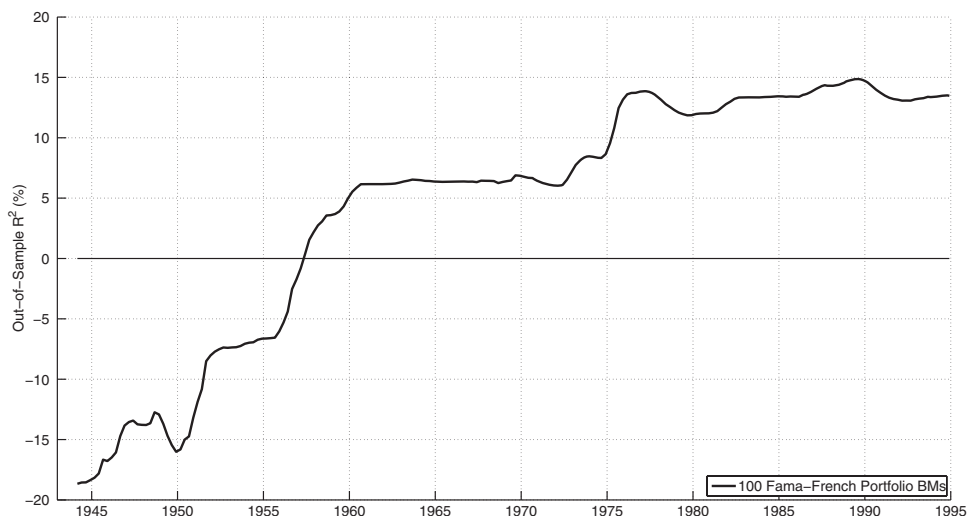
	$R^2$ (%)	$p$ (KP/CM)	$p$ (Hodrick)	$p$ (NW)
Panel A: 1930 to 2010				
6 Portfolios				
In-sample	15.35	0.009	0.002	0.002
Out-of-sample	4.17	< 0.010		
25 Portfolios				
In-sample	23.67	< 0.001	0.002	< 0.001
Out-of-sample	−7.20	—		
100 Portfolios				
In-sample	31.78	< 0.001	< 0.001	< 0.001
Out-of-sample	−8.96	—		
Panel B: 1940 to 2010				
6 Portfolios				
In-sample	11.96	0.003	< 0.001	< 0.001
Out-of-sample	11.83	< 0.010		
25 Portfolios				
In-sample	13.93	0.004	< 0.001	< 0.001
Out-of-sample	11.59	< 0.010		
100 Portfolios				
In-sample	16.72	0.006	< 0.001	< 0.001
Out-of-sample	12.14	< 0.010		

## IV. Economic Implications

### A. Sources of Improved Predictability

Why is it that the cross-section of valuation ratios possesses more powerful and more robust predictive ability than a single aggregate ratio? An answer lies in the first-stage regression loadings, which effectively become the weights used to linearly combine value ratios when we form our single predictor. Figure 6 plots these loadings for the return and dividend growth forecasting problems.<sup>30</sup> The three figures on the left plot the sensitivities of portfolio book-to-market ratios to future market returns. All portfolios have positive return exposures, with growth stocks and large caps being especially sensitive. The right column shows that growth stocks have small positive sensitivities to future cash flow growth, while value stocks have large negative exposures.

<sup>30</sup> The figure shows estimates from the 1940 to 2010 sample, as this is where we detect significant dividend growth predictability.



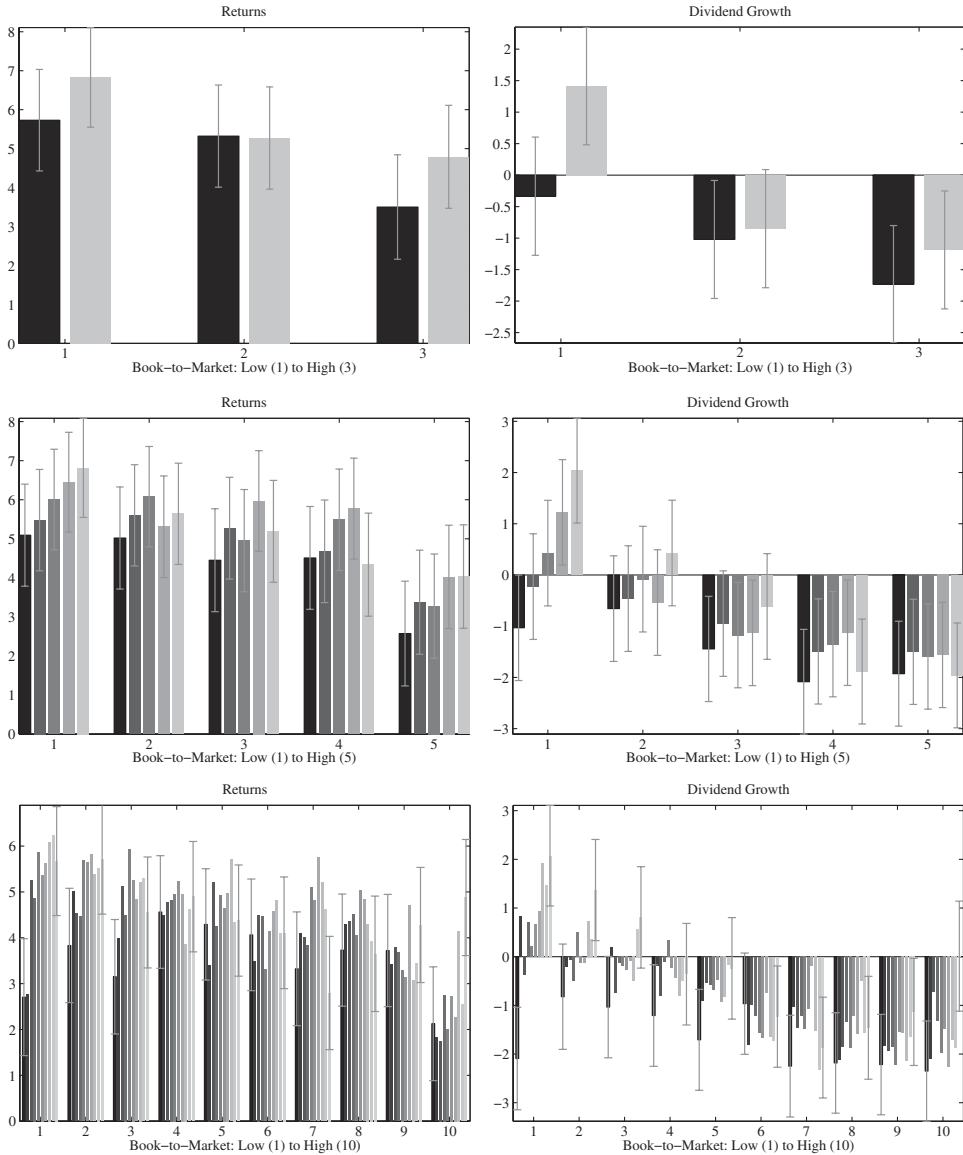
**Figure 5. Out-of-sample  $R^2$  by sample split date, annual dividend growth.** Forecasts of annual aggregate dividend growth (1940–2010) are based on a single PLS factor from 100 book-to-market ratios of size- and value-sorted portfolios of U.S. stocks from Fama and French (1993).

**Table V**  
**Market Return Predictions (1940 to 2010)**

The table reports results of PLS forecasts of 1-year and 1-month market returns, where data are for the period 1940 to 2010. See Table I for additional details.

	Panel A: 1 Year				Panel B: 1 Month	
	$R^2$ (%)	$p$ (KP/CM)	$p$ (Hodrick)	$p$ (NW)	$R^2$ (%)	$p$ (KP/CM)
6 Portfolios						
In-sample	5.48	0.042	0.043	0.007	0.61	0.072
Out-of-sample	3.88	<0.050			0.30	<0.050
25 Portfolios						
In-sample	7.10	0.021	0.027	0.002	0.95	0.019
Out-of-sample	6.76	<0.050			0.80	<0.050
100 Portfolios						
In-sample	8.77	0.017	0.015	0.002	2.34	<0.001
Out-of-sample	8.28	<0.010			1.10	<0.010

Now consider the optimal PLS return forecaster, which combines predictors according to their first-stage loading estimates. It places a positive weight on all book-to-market ratios, but overweights particularly strong predictors (growth stocks and large stocks) and underweights weaker predictors (value stocks and small stocks). However, value and growth have opposite signs in their cash flow growth loadings. This implies that the optimal return forecasting combination will have low overall exposure to cash flow growth. This is because any combination of portfolios that has all positive weights essentially nets out



**Figure 6. First-stage return and cash flow growth sensitivity estimates.** The figures show estimated first-stage regression loadings from the PLS procedure. Six, 25, or 100 Fama-French book-to-market ratios are used to forecast annual market returns and market dividend growth in the 1940 to 2010 sample. Bar shading represents portfolios' size quantiles, with the smallest market capitalization portfolios in black and larger sizes in lighter shades. Two standard error confidence intervals for estimates are also shown (to avoid crowding in the 100-portfolio case, these are only shown for the largest and smallest portfolios in each value group).

the influence of cash flow growth. Recall from the introduction's CAPM example that cash flow growth expectations act as a source of noise when predicting returns with a single valuation ratio. So, by combining the cross-section of ratios, we not only overweight the especially strong return predictors, but we also isolate discount rate sensitivity by canceling out noise associated with cash flow dynamics.

This feature also holds for the optimal dividend forecaster. It has a positive weight in growth stocks and a negative weight in value. Because all return sensitivities are positive, this predictor has a low overall sensitivity to return dynamics while having high sensitivity to cash flow growth.

These first-stage loading estimates have interesting economic interpretations. Dechow, Sloan, and Soliman (2004) show empirically that growth stocks have higher cash flow duration than value stocks. Lettau and Wachter (2007, 2011) and Santos and Veronesi (2004) provide theoretical links between cash flow duration and an asset's price sensitivity to variation in discount rates. The evidence in Figure 6 that growth stocks are particularly informative for forecasting returns conforms to the theoretical prediction that high duration assets have more exposure to discount rate shocks. When expected future returns rise, book-to-market ratios rise (prices fall) for all assets, but more so for growth stocks since their cash flows materialize further in the future, thus are more heavily discounted. Also, the relatively high expected return sensitivity of large stocks is consistent with the rationale that larger firms have a relatively large fraction of systematically risky cash flows, as in Berk, Green, and Naik (1999). Because systematic cash flows are discounted with a risk premium, large stocks become more informative about variation in expected returns.

Our results are also consistent with the findings of Campbell and Vuolteenaho (2004), who use a decomposition of returns into cash flow and discount rate news to show that growth stocks have relatively high betas on news about discount rates. Finally, our finding that value stocks are the strongest predictors of cash flows (they have the largest first-stage coefficients in absolute value) is further congruent with the Campbell and Vuolteenaho result that value stocks have high cash flow news betas.

### *B. Expected Returns and Macroeconomic Variables*

Using our discount rate estimates, we next investigate the time-series relation between expected returns and the macroeconomy. Table VI reports the association between our estimated time series of 1-month and 1-year expected market returns with measures of macroeconomic activity, uncertainty, sentiment, and credit risk. Our analysis of macroeconomic drivers of discount rates is motivated in part by Ferson and Harvey (1991).

First, we find that discount rates are broadly countercyclical. They are significantly lower in periods of high industrial production growth, low unemployment, and low aggregate book-to-market ratio. They are also lower amid high GDP growth (though insignificantly so) and high levels of the Chicago Fed National Activity Index (significant only for 1-month expected returns).

**Table VI**  
**Correlation of Estimated Expected Returns with Macro Variables**

We report percentage correlations and associated *t*-statistics between our estimated 1-year and 1-month expected return series and various macroeconomic time series. These include industrial production growth, GDP growth, Chicago Fed National Activity Index, unemployment rate, aggregate book-to-market ratio of aggregate U.S. stock market, surplus consumption ratio, recession probability estimates of Chauvet and Piger (2008), Anxious Index from the Survey of Professional Forecasters, GDP growth volatility, dispersion in growth forecasts in the Survey of Professional Forecasters, macroeconomic uncertainty index from Bloom et al. (2012), consumption growth volatility, sentiment index (both raw and orthogonalized against a set of macro series) of Baker and Wurgler (2007), IPO volume, new equity issues and closed-end fund discount also from Baker and Wurgler (2007), Baa–Aaa yield spread, and 10-year minus 30-day government debt yield spread. Correlations are calculated using all available post–World War II data for each series.

	1 Year		1 Month	
	Corr.	<i>t</i>	Corr.	<i>t</i>
Macroeconomic activity				
IP growth	−10.3	−2.6	−11.5	−2.9
GDP growth	−7.8	−1.1	−7.6	−1.1
CFNAI	−3.6	−0.8	−11.2	−2.6
Unemployment	41.2	11.2	32.1	8.4
Agg. book-to-market	65.7	21.5	55.6	16.5
Surp. Cons. Ratio	−8.3	−2.1	−6.8	−1.7
CP recession	8.9	2.0	16.5	3.8
SPF recession	19.9	2.6	21.4	2.8
Macroeconomic uncertainty				
GDP Gr. volatility	28.1	4.1	32.5	4.8
SPF uncertainty	43.2	6.1	36.5	5.0
Uncertainty index	17.2	2.4	23.6	3.4
Cons. Gr. volatility	28.5	7.4	18.5	4.7
Sentiment				
Index	−26.8	−6.5	−21.4	−5.1
Index (orth.)	−22.7	−5.4	−18.1	−4.3
IPO volume	−17.9	−4.5	−21.1	−5.3
Equity new issues	−32.3	−8.4	−33.6	−8.8
Closed-end discount	31.7	7.8	27.0	6.5
Credit				
Baa–Aaa spread	31.2	8.1	21.9	5.5
Term spread	−0.8	−0.2	−2.9	−0.7

In the habit persistence model of Campbell and Cochrane (1999), the surplus consumption ratio is the state variable that drives discount rates. A low surplus consumption ratio is associated with bad times and high discount rates. Our 1-year expected returns series shares a significant −8.3% correlation with the surplus consumption ratio, consistent with the prediction of habit theory.<sup>31</sup>

We also consider two recession probability measures. The first is a real-time recession probability calculated by Chauvet and Piger (2008). The second is the “Anxious Index” from the Survey of Professional Forecasters, which asks

<sup>31</sup> We use the surplus consumption ratio calculated by Dew-Becker (2012).

survey panelists “to estimate the probability that real GDP will decline in the quarter in which the survey is taken and in each of the following four quarters.” In both cases we find that expected returns are significantly higher when there is a higher likelihood of economic contraction.

Next, we find an important association between macroeconomic uncertainty and our estimated expected returns. Three of the four uncertainty measures that we use are from Bloom et al. (2012). These include GDP volatility estimated with a GARCH model, the cross-section standard deviation of professional forecasts of industrial production growth, and an uncertainty index that is a composite of seven macroeconomic uncertainty variables. We find a strong correlation between these measures and discount rates, ranging between 17% and 43% depending on the uncertainty measure and investment horizon.

We also consider consumption uncertainty measured from a GARCH model for year-on-year consumption growth rates (real nondurables and services). In the long-run risks theory of Bansal and Yaron (2004), consumption growth volatility is the key state variable, generating high discount rates in highly volatile periods. We find strong support for this prediction in the data, with a significant 28.5% correlation between consumption volatility and our estimated 1-year expected returns.<sup>32</sup>

Third, we find that discount rates fall in periods of high investor sentiment. We proxy for investor sentiment using five variables from Baker and Wurgler (2007). The first is a sentiment index that is the first PC of six individual sentiment proxies. We also use a version of this index that is orthogonalized with respect to macroeconomic conditions (see their paper for details), as well as three individual sentiment proxies including IPO volume, the share of equity in new security issues, and the closed-end fund discount (which is an inverse measure of sentiment). For all measures, sentiment is significantly negatively correlated with our estimated discount rate series. That is, high sentiment is associated with high prices and low expected returns.

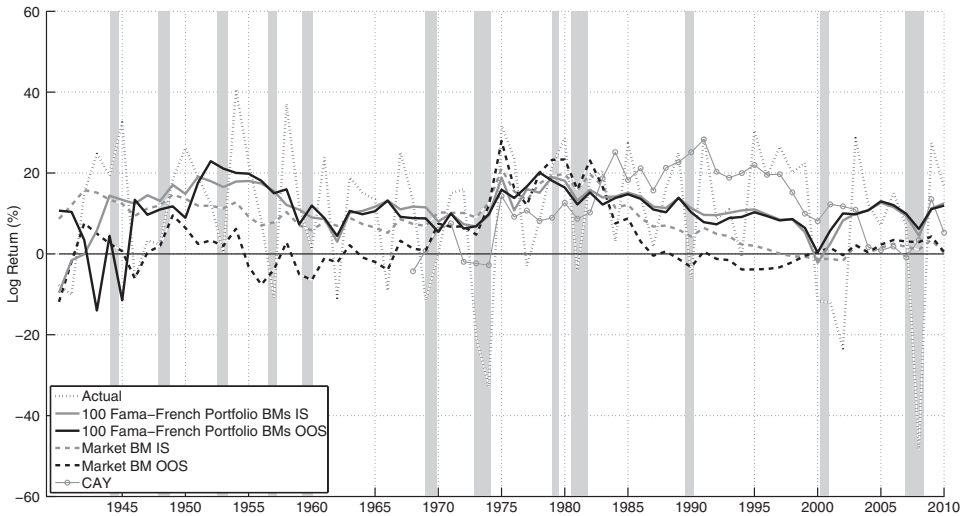
Lastly, we find that discount rates are significantly positively associated with credit risk as measured by the difference between yields on Baa- and Aaa-rated bonds. We find no significant relation with the term spread between long- and short-maturity U.S. government debt yields.

### C. Time-Series Properties of Discount Rates

Estimates of the expected annual return based on our cross-sectional approach are plotted in Figure 7, and are compared against fitted values from regressions on aggregate valuation ratios. The first conclusion we draw

<sup>32</sup> Using a cross-section of portfolio price-dividend ratios, Jagannathan and Marakani (2011) estimate two long-run risk factors. One factor captures expected consumption and dividend growth and another captures consumption growth volatility. Consistent with other uncertainty measures, we find that our estimated expected annual return series is significantly positively correlated (68%) with their consumption volatility factor. We find no significant correlation with their consumption/dividend growth factor.





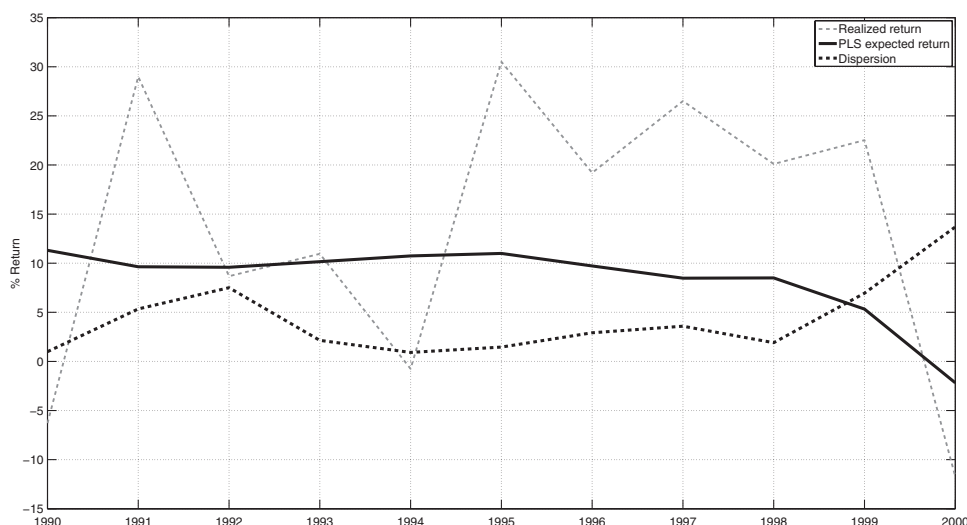
**Figure 7. Market 1-year return predictions.** The figure shows annual realized returns for the aggregate U.S. stock market, in-sample and out-of-sample forecasts from our PLS factor extracted from 100 book-to-market ratios of size- and value-sorted portfolios, and in-sample forecasts from predictive regressions on the aggregate book-to-market ratio, aggregate price-dividend ratio, and consumption-wealth ratio of Lettau and Ludvigson (2001). NBER recession dates are represented by the shaded areas.

from this figure is that in-sample versus out-of-sample discount rate estimates from our procedure are highly consistent with each other.

The figure also shows that our estimated 1-year-ahead expected return process differs from estimates based on the aggregate book-to-market ratio in several qualitatively important ways. We can understand the source of these differences by studying the cross-sectional properties of value ratios.

In the data, portfolio book-to-market ratios fan out and compress over time. Section IV.A shows that  $v_{i,t}$  and  $\hat{\phi}_i$  are negatively correlated in the cross-section—growth stocks have comparatively low book-to-market ratios and high discount rate sensitivity. Thus, fanning out corresponds to book-to-market ratios for value stocks (the highest book-to-market stocks) becoming even higher, and those of growth stocks becoming even lower. In other words, value stocks become especially cheap and growth stocks especially expensive. According to the cash flow duration rationale discussed earlier, this would happen when there is a decrease in expected future returns. Because growth stocks have longer duration, a fall in discount rates causes their prices to rise more sharply than those of value stocks.

This theoretical implication manifests itself quite strongly in our data. When the cross-section of book-to-market ratios fans out we estimate low expected future returns, and when it compresses we estimate high expected returns. The



**Figure 8. Expected returns and book-to-market dispersion.** The figure plots realized and predicted future 12-month returns for 1990 to 2000 estimated in-sample with PLS using 100 Fama-French portfolio book-to-market ratios. It also plots the cross-sectional standard deviation of the 100 Fama-French book-to-market ratios each period. To facilitate comparison, the dispersion series is scaled to have the same mean and variance as the PLS expected return series.

correlation between our fitted return series and the month-by-month standard deviation in book-to-market ratios for 100 Fama-French portfolios is  $-67\%$ .<sup>33</sup>

This idea helps us understand episodes in which aggregate book-to-market forecasts diverge from cross-section-based forecasts. One important example is the technology boom of the late 1990s. Our cross-section-based expected return estimate for 1999 is  $5\%$ , far above the  $-0.6\%$  estimate from the aggregate book-to-market ratio. Figure 8 plots our in-sample fitted expected return alongside the cross-sectional standard deviation of the 100 Fama-French portfolios.

<sup>33</sup> Despite the strong correlation between dispersion and our PLS predictor, dispersion substantially underperforms our approach in terms of predictability. It produces an in-sample annual  $R^2$  of only  $1.3\%$ , with an insignificant  $t$ -statistic of  $0.8$ . The out-of-sample  $R^2$  for a 1980 sample split is  $3.0\%$ , versus  $13.1\%$  for PLS. We find a very similar result if we instead measure dispersion as the value spread (tenth Fama-French value decile minus first) as in Liu and Zhang (2008). The cross-sectional standard deviation of 100 Fama-French portfolios and the 10-minus-1 value spread share a correlation of  $94\%$ . PLS captures information that the dispersion and value spread miss. Second-stage PLS regressions that produce the time  $t$  observation of our predictor variable rely on the cross-section covariance between value ratios  $v_{i,t}$  and first-stage  $\phi_i$  estimates. The dispersion in  $v_{i,t}$ s is one input to this calculation, but this statistic is too coarse on its own. The cross-sectional covariance is the proper statistical measure since it incorporates not only information about differences in value ratio levels, but also information about how these levels relate to portfolios' discount rate sensitivities, which is information ignored by dispersion.

**Table VII**  
**Persistence and Volatility of Expected Returns**

For each series we report the unconditional mean and volatility (reported as “Vol (level)”). We also estimate AR(1) model  $x_t = c + \rho x_{t-1} + u_t$  and report the AR(1) coefficient and volatility of  $u_t$  innovations (reported as “Vol (shock)”). This is done for realized aggregate U.S. stock market returns, expected returns as estimated using the PLS factor from 100 portfolios either in-sample or out-of-sample, and in-sample estimates of expected returns from the aggregate book-to-market ratio and aggregate price-dividend ratio. Twelve-month series use nonoverlapping July to June observations. Data are for the 1946 to 2010 period to allow for calculation of out-of-sample forecasts.

	Mean	AR(1)	Vol (level)	Vol (shock)
Panel A: 1 Year				
Realized Returns	10.01	−0.08	16.45	16.14
100 Portfolios (in-sample)	11.19	0.26	5.41	5.24
100 Portfolios (out-of-sample)	11.23	0.16	5.80	5.76
Aggregate Book-to-Market	8.24	0.91	5.83	2.44
Aggregate price-dividend	8.19	0.93	3.51	1.32
Panel B: 1 Month				
Realized Returns	0.84	0.09	4.35	4.33
100 Portfolios (in-sample)	1.02	0.78	0.39	0.24
100 Portfolios (out-of-sample)	0.97	0.69	0.35	0.25
Aggregate Book-to-Market	0.67	0.99	0.43	0.05
Aggregate price-dividend	0.69	0.99	0.22	0.02

From 1995 to 1999, estimated expected returns fell gradually from roughly 12% to 5%. Then, in 2000, expected returns fell just below zero. The 1995 to 2000 period witnessed a slow upward trend in the dispersion of book-to-market ratios for the Fama-French portfolios, punctuated by a rapid widening in 2000. Figure 8 reveals how information in the cross-sectional distribution of value ratios provides useful economic information that is missed when working with the aggregate book-to-market ratio.

Table VII presents the mean, persistence, and volatility of various estimated 1-year expected return processes. It also reports the volatility of innovations to expected returns.

First, we find that the persistence of expected returns is far lower than previous research suggests. The AR(1) coefficient on annual returns implied by aggregate book-to-market or price-dividend ratio regressions is 0.91 to 0.93, while the AR(1) based on our arguably improved estimates is below 0.3. The volatility of discount rate *levels* is similar across estimation approaches, but this masks extreme differences in the volatility of discount rate *shocks*. For example, if expected returns follow an AR(1), then the volatility of shocks is scaled up to an unconditional expected return volatility according to

$$\text{Unconditional Vol} = \frac{\text{Shock Vol}}{\sqrt{1 - \text{Persistence}^2}}.$$

Table VII reports that the estimated volatility of expected return AR(1) shocks is over 5.2% based on our methodology, versus a shock volatility of 2.4% based on aggregate book-to-market forecasts.

Traditional in-sample return predictions using aggregate value ratios have presented a puzzle in that measured volatility is too large to be reconciled with standard consumption-based asset pricing theories. Some estimates suggest that expected return volatility exceeds the level of the already high equity premium (Cochrane (2011)). Recently, the profession has converged on a new set of models to, among other things, help explain this feature of the data (Campbell and Cochrane (1999), Bansal and Yaron (2004)). These models agree on time-varying returns that are highly persistent, and thus they can match discount rate volatility with relatively little volatility of discount rate shocks. For instance, in the long-run risks model of Bansal and Yaron, expected excess market returns inherit the persistence of the conditional variance of consumption growth. In their calibration, this conditional variance has a monthly AR(1) coefficient of 0.979 (or  $0.979^{12} = 0.775$  annually). If the variability of expected returns estimated in earlier literature was an indication of a puzzle, our findings substantially deepen it by suggesting that we would need to double the magnitude of shocks needed to match discount rate volatility given the low persistence of our estimated series.

## V. Conclusion

We derive a dynamic latent factor model representation for the cross-section of asset valuation ratios. The same factors that drive the panel of present values also determine aggregate expectations of market returns and cash flow growth, enabling us to use rich cross-sectional information in constructing forecasts. To analyze these latent processes, we use recent econometric results for PLS in a factor model setting. By extracting information from disaggregate valuation ratios we are able to construct remarkably accurate forecasts of returns and cash flow growth rates both in-sample and out-of-sample. The resulting estimates imply that discount rates are much less persistent, and their shocks more volatile, than previous literature suggests. These facts stand in contrast to the far more persistent conditional expectations implied by standard models of asset prices. Our results are robust to a variety of cross-sections and out-of-sample procedures, and they hold in both U.S. and international data. The cross-section of valuation ratios, as present value identities imply, hold a wealth of information about investor expectations. Better identifying the economic fundamentals that drive these valuation ratios is a promising avenue for future research.

## REFERENCES

- Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–151.
- Ball, Ray, and Ross Watts, 1972, Some time series properties of accounting income, *Journal of Finance* 27, 663–681.
- Bansal, Ravi, Robert Dittmar, and Christian Lundblad, 2005, Consumption, dividends, and the cross section of equity returns, *Journal of Finance* 60, 1639–1672.
- Bansal, Ravi, and S. Viswanathan, 1993, No arbitrage and arbitrage pricing: A new approach, *Journal of Finance* 48, 1231–1262.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Berk, Jonathan, Richard Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.
- Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry, 2012, Really uncertain business cycles, Working paper, NBER.
- Campbell, John Y., and James H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.
- Campbell, John Y., and Robert Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.
- Campbell, John Y., and S. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.
- Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.
- Chauvet, Marcelle, and Jeremy Piger, 2008, A comparison of the real-time performance of business cycle dating methods, *Journal of Business and Economic Statistics* 26, 42–49.
- Chen, Long, 2009, On the reversal of return and dividend growth predictability: A tale of two periods, *Journal of Financial Economics* 92, 128–151.
- Clark, Todd, and Michael McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85–110.
- Clark, Todd, and Michael McCracken, 2005, Evaluating direct multistep forecasts, *Econometric Reviews* 24, 369–404.
- Cochrane, James H., 1992, Explaining the variance of price-dividend ratios, *Review of Financial Studies* 5, 243–280.
- Cochrane, John H., 1999, New facts in finance, *Economic Perspectives* 23, 36–58.
- Cochrane, John H., 2005, *Asset Pricing* (Princeton University Press, Princeton, NJ).
- Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.
- Cochrane, John H., 2011, Discount rates, *Journal of Finance* 66, 1047–1108.
- Connor, Gregory, 1984, A unified beta pricing theory, *Journal of Economic Theory* 34, 13–31.
- Dechow, Patricia M., Richard G. Sloan, and Mark T. Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197–228.
- Dew-Becker, Ian, 2012, A model of time-varying risk premia with habits and production, Working paper, Harvard University.
- Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 1998, Value versus growth: The international evidence, *Journal of Finance* 53, 1975–1999.
- Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.
- Fama, Eugene F., and Kenneth R. French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics* 60, 3–43.

- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Ferson, Wayne, and Campbell Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.
- Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.
- Hansen, Peter R., and Allan Timmermann, 2011, Choice of sample split in out-of-sample forecast evaluation, Working paper, Stanford University.
- Harvey, Campbell, 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.
- Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.
- Huber, Peter J., 1973, Robust regression: Asymptotics, conjectures and Monte Carlo, *Annals of Statistics* 1, 799–821.
- Inoue, Atsushi, and Barbara Rossi, 2011, Out-of-sample forecast tests robust to the choice of window size, Working papers 8542, C.E.P.R.
- Jagannathan, Ravi, and Srikant Marakani, 2011, Long run risks and price/dividend ratio factors, Working paper, Northwestern University.
- Kelly, Bryan T., and Seth Pruitt, 2012, The three-pass regression filter: A new approach to forecasting with many predictors, Working paper, Chicago Booth.
- Kojan, Ralph S.J., and Stijn Van Nieuwerburgh, 2011, Predictability of returns and cash flows, *Annual Review of Financial Economics* 3, 467–491.
- Lacerda, Felipe, and Pedro Santa-Clara, 2010, Forecasting dividend growth to better predict returns, Working paper, Chicago Booth.
- Lettau, Martin, and Sydney C. Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.
- Lettau, Martin, and Sydney C. Ludvigson, 2005, Expected returns and expected dividend growth, *Journal of Financial Economics* 76, 583–626.
- Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the return predictability evidence, *Review of Financial Studies* 21, 1607–1652.
- Lettau, Martin, and Jessica A. Wachter, 2007, Why is long-horizon equity less risky? A duration-based explanation of the value premium, *Journal of Finance* 62, 55–92.
- Lettau, Martin, and Jessica A. Wachter, 2011, The term structures of equity and interest rates, *Journal of Financial Economics* 101, 90–113.
- Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- Liu, Naiping, and Lu Zhang, 2008, Is the value spread a useful predictor of returns? *Journal of Financial Markets* 11, 199–227.
- Ludvigson, Sydney C., and Serena Ng, 2007, The empirical risk-return relation: A factor analysis approach, *Journal of Financial Economics* 83, 171–222.
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Miller, Merton H., and Myron S. Scholes, 1982, Dividends and taxes: Some empirical evidence, *Journal of Political Economy* 90, 1118–1141.
- Nelson, Charles R., and Myung J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.
- Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2006, Cross-sectional forecasts of the equity premium, *Journal of Financial Economics* 81, 101–141.
- Roll, Richard, and Stephen A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, *Journal of Finance* 35, 1073–1103.
- Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Santos, Tano, and Pietro Veronesi, 2004, Conditional betas, Working paper, NBER.

- Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.
- Stambaugh, Robert, 1986, Bias in regressions with lagged stochastic regressors, Working paper, University of Chicago.
- 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, 1167–1179.
- Treynor, Jack L., 1961, Toward a theory of market value of risky assets, Working paper, Harvard Business School.
- van Binsbergen, Jules, and Ralph Koijen, 2010, Predictive regressions: A present-value approach, *Journal of Finance* 65, 1439–1471.
- Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns?, *Journal of Finance* 57, 233–264.
- White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.
- Wold, Herman, 1975, Estimation of principal components and related models by iterative least squares, in P.R. Krishnaiah, ed.: *Multivariate Analysis* (Academic Press, New York).