

Predictive Systems: Living with Imperfect Predictors

by*

Ľuboš Pástor

and

Robert F. Stambaugh

First draft: September 15, 2006

This version: January 29, 2008

Abstract

We develop a framework for estimating expected returns—a *predictive system*—that allows predictors to be imperfectly correlated with the conditional expected return. When predictors are imperfect, the estimated expected return depends on past returns in a manner that hinges on the correlation between unexpected returns and innovations in expected returns. We find empirically that prior beliefs about this correlation, which is most likely negative, substantially affect estimates of expected returns as well as various inferences about predictability, including assessments of a predictor’s usefulness. Compared to standard predictive regressions, predictive systems deliver different and more precise estimates of expected returns.

*Graduate School of Business, University of Chicago, NBER, and CEPR (Pástor) and the Wharton School, University of Pennsylvania, and NBER (Stambaugh). Helpful comments were received from the audiences at the Fall 2006 NBER Asset Pricing Meeting, 2006 Wharton Frontiers of Investing conference, 2007 Western Finance Association conference, 2007 European Finance Association conference, 2007 ESSFM at Gerzensee, 2007 Vienna Symposia in Asset Management, Tel Aviv University Conference in Honor of Shmuel Kandel, Boston College, Goldman Sachs, Hong Kong University of Science and Technology, National University of Singapore, New York University, Norwegian School of Economics and Business Administration (Bergen), Norwegian School of Management (Oslo), Singapore Management University, University of California at San Diego, University of Chicago, University of Iowa, University of Michigan, University of Pennsylvania, University of Texas at Austin, and University of Texas at Dallas. We also thank Jacob Boudoukh, Ken French, Cam Harvey, Jesper Rangvid, Cesare Robotti, Ross Valkanov, Pietro Veronesi, and especially Jonathan Lewellen, John Cochrane, and two anonymous referees for many helpful suggestions.

1. Introduction

Many studies in finance analyze comovement between expected asset returns and various observable quantities, or “predictors.” A question of frequent interest is how x_t , a vector of predictors observed at time t , is related to μ_t , the conditional expected return defined in the equation

$$r_{t+1} = \mu_t + u_{t+1}, \quad (1)$$

where r_{t+1} denotes the stock return from time t to time $t + 1$, and the unexpected return u_{t+1} has mean zero conditional on information available at time t . One approach to modeling expected returns is to use a “predictive regression” in which r_{t+1} is regressed on x_t and the expected return is given by $\mu_t = a + b'x_t$, where a and b denote the regression’s intercept and slope coefficients.¹ This approach seems too restrictive in modeling expected return as an exact linear function of the observed predictors. It seems more likely that the predictors are imperfect, in that they are correlated with μ_t but cannot deliver it perfectly.

At the same time, the predictive regression approach seems too lax in ignoring a likely economic property of the unexpected return—its negative correlation with the innovation in the expected return. For example, if the expected return obeys the first-order autoregressive process,

$$\mu_{t+1} = \alpha + \beta\mu_t + w_{t+1}, \quad (2)$$

then it seems likely that the correlation between the unexpected return and the innovation in the expected return is negative, or that $\rho_{uw} \equiv \rho(u_{t+1}, w_{t+1}) < 0$. That is, unanticipated increases in expected future returns (or discount rates) should be accompanied by unexpected negative returns. While it is possible for $\rho_{uw} < 0$ to be violated, we argue that such violations are unlikely. The likely negativity of ρ_{uw} , which is not exploited in estimating the predictive regression, emerges as an important consideration in estimating expected returns when predictors are imperfect.

We develop an approach to estimating expected returns that generalizes the standard predictive regression approach. The framework we propose, which we term a *predictive system*, allows the predictors in x_t to be imperfect, in that $\mu_t \neq a + b'x_t$. The predictive system also allows us to explore roles for a variety of prior beliefs about the behavior of expected returns, chief among which is the belief that unexpected returns are negatively correlated with innovations in expected

¹Of the many studies that estimate predictive regressions for stock returns, some early examples include Fama and Schwert (1977), Rozeff (1984), Keim and Stambaugh (1986), Campbell (1987), and Fama and French (1988). There is also a substantial literature analyzing econometric issues associated with predictive regressions, including Mankiw and Shapiro (1986), Stambaugh (1986, 1999), Nelson and Kim (1993), Elliott and Stock (1994), Cavanagh, Elliott, and Stock (1995), Ferson, Sarkissian, and Simin (2003), Lewellen (2004), Campbell and Yogo (2006), Jansson and Moreira (2006), and Lettau and van Nieuwerburgh (2007).

returns ($\rho_{uw} < 0$). We find that, compared to predictive regressions, predictive systems deliver different and more precise estimates of expected returns. When predictors are imperfect, their predictive ability is supplemented by information in lagged returns as well as lags of the predictors, and the predictive system delivers that information via a parsimonious model. The correlation ρ_{uw} plays a key role in determining how that additional sample information is used as well as how important that information is in explaining variation in expected returns.

The additional information in lagged returns is used in an interesting way. Suppose that recent returns have been unusually low. On one hand, one might think that the expected return has declined, since a low mean is more likely to generate low realized returns, and the conditional mean is likely to be persistent. On the other hand, one might think that the expected return has increased, since increases in expected future returns tend to produce low realized returns. When ρ_{uw} is sufficiently negative, the latter effect outweighs the former and recent returns enter negatively when estimating the current expected return. At the same time, more distant past returns enter positively because they are more informative about the level of the unconditional expected return than about recent changes in the conditional expected return.

We illustrate the role of lagged returns in a simplified setting where historical returns are the only available sample information (D_t). Suppose, for example, that an investor in January 2000 is forming an expectation of the stock market return over the following quarter based on the post-war history of realized market returns. Does the dramatic rise in stock prices in the 1990s increase or decrease the investor's expectation of future return? The answer depends on the extent to which the 1990s' bull market was caused by unexpected declines in expected returns. The conditional expected stock return in this simplified setting is just a weighted average of all past realized returns,

$$E(r_{t+1}|D_t) = \sum_{s=0}^{t-1} \kappa_s r_{t-s},$$

and the weights κ_s depend on ρ_{uw} . For example, if this investor believes that $\rho_{uw} = -0.85$, so that 72% of the variance in unexpected returns is due to changes in expected returns (the estimates of Campbell (1991) are in that neighborhood), then returns realized during the most recent decade receive negative weights, while the returns from the previous four decades receive positive weights. The investor in this example views the 1990s' bull market as a bearish indicator.

Imperfection in predictors complicates inference about their relations to expected return. We show that if predictors are imperfect, the residuals in the predictive regression of r_{t+1} on x_t are serially correlated. This correlation is often ignored when computing standard errors in predictive regressions. The serial correlation in residuals joins other features of predictive regressions that are already well known to complicate inferences, especially in finite samples, such as persistence in the

predictors and correlation between the residuals and innovations in the predictors (e.g., Stambaugh, 1999). Using our alternative framework—the predictive system—we develop a Bayesian approach that allows us to conduct clean finite-sample inference about various properties of the expected return. This approach also allows us to incorporate prior beliefs about ρ_{uw} .

A striking example of the importance of such prior beliefs is provided by regressing post-war U.S. stock market returns on the “bond yield,” defined as minus the yield on the 30-year Treasury bond in excess of its most recent 12-month moving average. That variable receives a highly significant positive slope (with a p -value of 0.001) in the predictive regression, but its AR(1) innovations are positively correlated with the residuals in that regression. The latter correlation, opposite in sign to what one would anticipate for ρ_{uw} , suggests that the bond yield is a rather imperfect predictor of stock returns. When judged in a predictive system, the bond yield’s importance as a predictor depends heavily on prior beliefs about ρ_{uw} . With noninformative beliefs about ρ_{uw} , the bond yield appears to be a very useful predictor; for example, the posterior mode of its conditional correlation with μ_t is 0.9. However, with a more informative belief that innovations in expected returns are negatively correlated with unexpected returns and explain at least half of their variance, the bond yield’s conditional correlation with μ_t drops to 0.2. With the more informative belief, the current value of the bond yield explains only 3% of the variance of μ_t . Adding lagged unexpected returns allows the system to explain 86% of this variance, and further adding lagged predictor innovations increases the fraction of explained variance of μ_t to 95%.

Prior beliefs also affect the predictive system’s advantage in explanatory power over the predictive regression. In the same bond yield example, with noninformative prior beliefs, the predictive system produces an estimate of μ_t that is 1.4 times more precise than the estimate from the predictive regression. With the more informative beliefs, though, the system’s estimate is 12.5 times more precise. We measure improvements in precision by improvements in explanatory power. Specifically, we compute the posterior mean of the ratio of the R^2 from the predictive regression to the R^2 from the regression of r_{t+1} on the return forecast from the predictive system.

We also include as predictors two more familiar choices, the market’s dividend yield and the consumption-wealth variable “CAY” proposed by Lettau and Ludvigson (2001). Prior beliefs about ρ_{uw} play a less dramatic role with these predictors than with the bond yield, but different prior beliefs can nevertheless produce substantial differences in estimated expected returns. We assess the economic significance of these expected return differences by comparing average certainty equivalents for mean-variance investors whose risk aversion would dictate an all-equity portfolio (i.e., no cash or borrowing) when expected return and volatility equal their long-run sample values. When all three predictors are included, an investor with the more informative belief

mentioned above would suffer an average quarterly loss of 1.5% if forced to hold the portfolio selected each quarter by an investor who estimates expected return by the maximum likelihood procedure (which reflects noninformative views about all parameters, including ρ_{uw}).

The predictive system, along with our Bayesian approach, also allows us to compute multi-period return variances that incorporate imperfect predictors and parameter uncertainty. We find that prior beliefs about ρ_{uw} play an important role here as well. In our example, using the dividend yield as the predictor, the five-year return variance is about 33 percent higher with a non-informative prior for ρ_{uw} than with the more informative prior discussed above. More negative values of ρ_{uw} imply stronger mean reversion in stock returns, resulting in lower long-horizon variances. We also find that the prior for ρ_{uw} has a much larger effect at longer investment horizons.

Person, Sarkissian, and Simin (2003) show that persistent predictors may exhibit spurious predictive power in finite samples even if they have no such power in population (e.g., if they have been data-mined). We provide tools that can be helpful in avoiding the spurious regression problem. A spurious predictor is unlikely to produce expected return estimates whose innovations are substantially negatively correlated with unexpected returns. Therefore, under an informative prior about this correlation, a predictive system would likely find the spurious predictor to be almost uncorrelated with μ_t . The basic intuition holds also outside the predictive system framework: if a predictor does not generate a negative correlation between expected and unexpected returns, it is unlikely to be highly correlated with the true conditional expected return.

This study is clearly related to an extensive literature on return predictability, but it also contributes to a broader agenda of incorporating economically motivated informative prior beliefs in inference and decision making in finance. Studies in the latter vein include Pástor and Stambaugh (1999, 2000, 2001, 2002ab), Pástor (2000), Baks, Metrick, and Wachter (2001), and Jones and Shanken (2005). Studies that employ informative priors in the context of return predictability include Kandel and Stambaugh (1996), Avramov (2002, 2004), Cremers (2002), Avramov and Wermers (2006), and Wachter and Warusawitharana (2006).

The paper is organized as follows. Section 2 introduces the predictive system and discusses its properties. Section 3 explains why ρ_{uw} is likely to be negative and how it affects expected returns. Section 4 presents our empirical work, in which we estimate the predictive system with a Bayesian approach. We compare the explanatory powers of the predictive system and predictive regression and quantify the differences in expected return estimates. We also assess the degree to which various predictors are correlated with the expected return, decompose the variation in the expected return into three components, and analyze the variance of the multiperiod returns. Section 5 reviews our conclusions. Many technical aspects of our analysis are presented in the Appendix.

2. Predictive System

In the predictive regression approach, the expected return is modeled as a linear combination of the predictors in x_t . This modeling assumption is unlikely to be exact, in that no linear combination of the predictors is likely to capture perfectly the true unobserved expected return, μ_t . We relax this assumption and develop an alternative predictive framework, which we call a *predictive system*. The predictive regression is a special case of the predictive system, as we show in Section 2.3.

We define the predictive system in its most general form as a vector autoregression for (μ_t, x_t) , with an arbitrary number of lags. We do not analyze this general form here. Given our objective to provide an initial exploration of predictive systems, simplicity is a virtue. We examine a simple version of the predictive system, in which μ_t and x_t follow AR(1) processes:

$$r_{t+1} = \mu_t + u_{t+1} \quad (3)$$

$$x_{t+1} = \theta + Ax_t + v_{t+1} \quad (4)$$

$$\mu_{t+1} = \alpha + \beta\mu_t + w_{t+1}. \quad (5)$$

The residuals in the system are assumed to be distributed identically and independently across t as

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix} \right). \quad (6)$$

We assume throughout that $0 < \beta < 1$ and that the eigenvalues of A lie inside the unit circle.

The predictive system is a version of a state space model.² Equation (3) defines the unobserved conditional expected return μ_t . Equation (4) is a standard assumption in the predictability literature. A special case of the predictive system arises when there are no predictors, in which case equation (4) is absent and the data include only returns. Equation (5) postulates a simple persistent process for μ_t . This reduced-form model could be consistent with a variety of economic models, rational or behavioral, in which the expected return varies over time in a persistent fashion.

2.1. Conditional Expected Return

The value of μ_t is unobservable, but the predictive system implies a value for $E(\mu_t|D_t) = E(r_{t+1}|D_t)$, where D_t denotes the history of returns and predictors observed through time t . Using

²Harvey (1989) provides a textbook treatment of state-space models, including a brief discussion of the case with non-zero correlations among all of the model's disturbances, which is the case here. In the Appendix, we provide an independent treatment specific to the system in (3) through (6). Studies that analyze return predictability using state space models include Conrad and Kaul (1988), Lamoureux and Zhou (1996), Johannes, Polson, and Stroud (2002), Ang and Piazzesi (2003), Brandt and Kang (2004), Dangi and Halling (2006), Duffee (2006), and Rytchkov (2007).

the Kalman filter, we find that this conditional expected return can be written as the unconditional expected return plus linear combinations of past return forecast errors and innovations in the predictors. Specifically, if we define the forecast error for the return in each period t as

$$\epsilon_t = r_t - E(r_t | D_{t-1}), \quad (7)$$

then the expected return conditional on the history of returns and predictors is given by

$$E(r_{t+1} | D_t) = E(r) + \sum_{s=0}^{\infty} (\lambda_s \epsilon_{t-s} + \phi'_s v_{t-s}), \quad (8)$$

where the unconditional mean return $E(r) = \alpha / (1 - \beta)$ and, in steady state,

$$\lambda_s = m\beta^s \quad (9)$$

$$\phi_s = n\beta^s, \quad (10)$$

where m and n are functions of the parameters in equations (3) through (6).³ The conditional expected return thus depends on the full history of returns and predictor realizations. We analyze this dependence in more detail in Section 3.2., where we plot λ_s as a function of s and ρ_{uw} .

Since the forecast errors ϵ_t in equation (8) are defined relative to conditional expectations that are updated through time based on the available return histories, part of the effects of past return realizations are impounded in those earlier conditional expectations. To isolate the full effect of each past period's total return, we can subtract the unconditional mean from each return, defining $\epsilon_t^U = r_t - E(r)$, and then rewrite the conditional expected return in equation (8) as

$$E(r_{t+1} | D_t) = E(r) + \sum_{s=0}^{\infty} (\omega_s \epsilon_{t-s}^U + \delta'_s v_{t-s}), \quad (11)$$

where, again in steady state,

$$\omega_s = m(\beta - m)^s \quad (12)$$

$$\delta_s = n(\beta - m)^s. \quad (13)$$

It can be verified that the rate of decay in ω_s and δ_s , $\beta - m$, is nonnegative. This alternative representation of the conditional expected return will also be useful in Section 3.2.

³In general, m and n are also functions of time, but as the length of the history in D_t grows long, they converge to steady-state values that do not depend on t . That convergence is reached fairly quickly in the settings we consider. We first present the steady-state expressions, for simplicity, but later employ the finite-sample Kalman filter as well. The Appendix derives the functions m and n in finite samples as well as in steady state. The Appendix also shows (in equation A41) that the finite-sample versions of m and n can be interpreted as the slope coefficients from the regression of μ_t on r_t and x_t , respectively, conditional on the sample information at time $t - 1$.

2.2. Temporal Dependence in Returns

Returns in the predictive system exhibit interesting temporal dependence. Given the AR(1) process for μ_t in equation (5), we can rewrite μ_t as an MA(∞) process (the Wold representation):

$$\mu_t = E(r) + \sum_{i=0}^{\infty} \beta^i w_{t-i}. \quad (14)$$

Using equations (3) and (14), the return k periods ahead can be written as

$$r_{t+k} = E(r) + \sum_{i=0}^{\infty} \beta^i w_{t+k-1-i} + u_{t+k} \quad (15)$$

$$= (1 - \beta^{k-1})E(r) + \beta^{k-1}\mu_t + \sum_{i=1}^{k-1} \beta^{k-1-i} w_{t+i} + u_{t+k}. \quad (16)$$

This equation implies that the autocovariance of returns is equal to

$$\text{Cov}(r_t, r_{t-k}) = \beta^{k-1} (\beta\sigma_{\mu}^2 + \sigma_{uw}), \quad (17)$$

where $\sigma_{\mu}^2 = \sigma_w^2/(1 - \beta^2)$ is the unconditional variance of μ_t . As a result, the serial correlation in returns can be positive or negative, depending on the parameter values. The positive component of (17) is due to persistence in μ_t ; the negative component is due to $\rho_{uw} < 0$, or mean reversion in stock returns. The knife-edge case of zero autocorrelation obtains for $\rho_{uw} = -\beta\sigma_w/(\sigma_u(1 - \beta^2))$.

The AR(1) process for μ_t also implies that returns follow an ARMA(1,1) process,

$$r_{t+1} = (1 - \beta)E(r) + \beta r_t + \epsilon_{t+1}^* - \gamma \epsilon_t^*. \quad (18)$$

When we implement the Kalman filter without using any predictor information, so that D_t includes only the return history, we obtain equation (8) without the last term ($\sum_{s=0}^{\infty} \phi_s' v_{t-s}$). That equation implies a specification of (18) with $\epsilon_t^* = \epsilon_t$ and $\gamma = \beta - m$.

2.3. Predictive Regression

The traditional approach to modeling return predictability, a predictive regression,

$$r_{t+1} = a + b'x_t + e_{t+1}, \quad (19)$$

arises as a special case of the predictive system if the predictors in x_t are “perfect” in that $\mu_t = a + b'x_t$. The predictors are perfect if there exists a b such that $w_t = b'v_t$ and $A'b = \beta b$.⁴

⁴ $A'b = \beta b$ means that β is an eigenvalue of A' corresponding to the eigenvector b ; one example is $A = \beta I$.

For example, if x_t contains one predictor, this predictor is perfect if its innovations are perfectly correlated with the innovations in μ_t (i.e., $\rho_{vw} = \pm 1$) and if its autocorrelation is the same as that of μ_t (i.e., $A = \beta$). In general, though, the predictors in x_t are imperfect in that $\mu_t \neq a + b'x_t$.

When the predictors approach perfection, then $m \rightarrow 0$, $n \rightarrow b$, and equation (8) becomes

$$E(r_{t+1}|D_t) = E(r) + b' \sum_{s=0}^{\infty} A^s v_{t-s} = E(r) + b'[x_t - E(x)] = a + b'x_t, \quad (20)$$

where $E(x) = (I - A)^{-1}\theta$ is the unconditional mean of x_t . That is, when the predictors approach perfection, the system-based conditional expected return approaches the regression-based conditional mean, $a + b'x_t$. When the predictors are imperfect, however, their entire history enters the conditional expected return, since the weighted sum of their past innovations in equation (8) does not then reduce to a function of just x_t . Moreover, when the predictors are imperfect, the expected return depends also on the full history of returns in addition to the history of the predictors.

Predictive systems have interesting implications for predictive regressions. All parameters of the predictive regression in equation (19) can be computed from the parameters of the predictive system in equations (3) through (6).⁵ Most interesting, the residual autocovariance is given by

$$\begin{aligned} \text{Cov}(e_t, e_{t+1}) &= \beta(\sigma_\mu^2 - V'_{x\mu} V_{xx}^{-1} V_{x\mu}) + \sigma_{uw} - V'_{x\mu} V_{xx}^{-1} \sigma_{vu} \\ &= \beta \text{Var}(\mu_t|x_t) + \text{Cov}(u_t, w_t - b'v_t). \end{aligned} \quad (21)$$

If the predictors are perfect, $\text{Var}(\mu_t|x_t) = 0$ and $w_t = b'v_t$, so $\text{Cov}(e_t, e_{t+1})$ is zero. With imperfect predictors, though, $\text{Var}(\mu_t|x_t) > 0$, $w_t \neq b'v_t$, and $\text{Cov}(e_t, e_{t+1})$ is generally non-zero. This observation suggests a simple diagnostic for predictor imperfection: if the residuals from the predictive regression exhibit non-zero autocorrelation, then the predictors x_t are imperfect.

The serial correlation in the residuals complicates the calculation of standard errors in the predictive regression approach. Since the first term in equation (21) is nonnegative, $\text{Cov}(e_t, e_{t+1})$ is often positive, in which case assuming uncorrelated predictive regression residuals leads to understated standard errors. Ferson, Sarkissian, and Simin (2003) make a similar point when the predictor is “spurious,” or uncorrelated with expected return. Their setting is a special case of (3)–(6) with one predictor and a diagonal covariance matrix in (6).⁶ In specifying a diagonal covariance matrix for the disturbances, they assume not only that the predictor is spurious but also that the innovations in expected return are uncorrelated with unexpected returns (i.e., $\rho_{uw} = 0$). In this

⁵For example, the regression slope b can be computed from the system’s parameters as $b = V_{xx}^{-1} V_{x\mu}$, where V_{xx} is given in the Appendix in equation (A10), $V_{x\mu} = (I_K - \beta A)^{-1} \sigma_{vw}$, and I_K is a $K \times K$ identity matrix.

⁶The objectives of Ferson et al. differ from ours. For example, they do not use this multiple-equation setting to estimate expected return or to examine its dependence on lagged returns and predictors.

special case, we see from (21) that $\text{Cov}(e_t, e_{t+1}) = \beta\sigma_\mu^2$. Ferson et al. do not report this expression but do find, using simulations, that the positive residual serial correlation can substantially affect inference in predictive regressions. Duffee (2006) also uses simulations to make a related point in the context of bond predictability.

3. Correlation between Expected and Unexpected Returns

A key quantity in this paper is ρ_{uw} , the correlation between the unexpected return, u_t , and the innovation in the expected return, w_t . Henceforth, we refer to ρ_{uw} simply as the “correlation between expected and unexpected returns,” a slightly inaccurate but much shorter description. This correlation is important for return predictability in several ways. First, it determines how past returns affect the forecasts of future returns, as we show in Section 3.2. Second, economically motivated prior beliefs about ρ_{uw} play an important role in various inferences about predictability, as we show in Section 4. The ability to incorporate prior beliefs about ρ_{uw} is a key feature of the predictive system. We begin this section by discussing some theoretical properties of ρ_{uw} .

3.1. Why ρ_{uw} Is Likely to Be Negative

The basic motivation behind the belief that $\rho_{uw} < 0$ is that asset prices tend to fall when discount rates rise. More precisely, $\rho_{uw} < 0$ means that unanticipated increases in expected returns tend to be accompanied by unexpected negative returns. This intuition holds perfectly for nominal returns on Treasury bonds. Since the nominal cash flows of Treasury bonds are fixed, the bond price variation is driven only by discount rate shocks, and $\rho_{uw} = -1$. Stock returns, however, are driven also by cash flow shocks. It is possible, at least in principle, that positive discount rate shocks could be accompanied by such large positive cash flow shocks that stock prices rise rather than fall as a result. In this section, we derive conditions under which $\rho_{uw} < 0$, and argue that these conditions are likely to be satisfied for the aggregate stock market.

Following Campbell (1991), the unexpected return can be decomposed approximately as

$$u_{t+1} = \eta_{C,t+1} - \eta_{E,t+1}, \quad (22)$$

where $\eta_{C,t+1}$ represents the unanticipated revisions in expected future cash flows and $\eta_{E,t+1}$ captures the revisions in expected future returns. If the expected return follows the process in equation (2) with $0 < \beta < 1$, then $\eta_{E,t+1}$ is equal to w_{t+1} multiplied by a positive constant, so that

$$\rho_{uw} = \rho(u_{t+1}, \eta_{E,t+1}). \quad (23)$$

Since the partial correlation between u_{t+1} and $\eta_{E,t+1}$ is minus one (equation (22)), it seems natural to believe a priori that the simple correlation between u_{t+1} and $\eta_{E,t+1}$, or ρ_{uw} , is negative. Before seeing any data, it is not obvious why $\eta_{C,t+1}$ and $\eta_{E,t+1}$ should be correlated, and a belief that $\eta_{C,t+1}$ and $\eta_{E,t+1}$ are uncorrelated translates into a belief that ρ_{uw} is negative. More precisely, it follows directly from equations (22) and (23) that $\rho_{uw} < 0$ if and only if

$$\rho(\eta_{C,t+1}, \eta_{E,t+1}) < \frac{\sigma(\eta_{E,t+1})}{\sigma(\eta_{C,t+1})}, \quad (24)$$

where the σ 's denote standard deviations. In order for $\rho_{uw} < 0$ to be violated, cash flow shocks would have to be more important than discount rate shocks in explaining the variance of stock returns, i.e., $\sigma(\eta_{C,t+1}) > \sigma(\eta_{E,t+1})$, and the correlation between those shocks, $\rho(\eta_{C,t+1}, \eta_{E,t+1})$, would have to be positive and sufficiently high. It seems difficult to argue that one could expect such a high correlation a priori. In fact, such a high correlation between the shocks to cash flows and discount rates seems unlikely because it would make stock returns unrealistically smooth. It is easy to see from equation (22) that a violation of the condition in (24) would require that

$$\text{Var}(u_{t+1}) < \text{Var}(\eta_{C,t+1}) - \text{Cov}(\eta_{C,t+1}, \eta_{E,t+1}) < \text{Var}(\eta_{C,t+1}) - \text{Var}(\eta_{E,t+1}) < \text{Var}(\eta_{C,t+1}). \quad (25)$$

That is, for $\rho_{uw} < 0$ to be violated, stock returns would have to be less volatile than when the expected return is constant (i.e., when $\text{Var}(\eta_{E,t+1}) = 0$). In reality, though, stock returns appear to be more volatile than when the expected return is constant. Shiller (1981) and LeRoy and Porter (1981) find that stock returns are much more volatile than the present value of future dividends discounted at constant rates. To explain this “excess volatility puzzle,” discount rates must vary over time in a way that increases stock volatility. But if ρ_{uw} were positive, discount rate variation would reduce stock volatility, thereby deepening the puzzle, which seems unappealing a priori.⁷

The analysis of ρ_{uw} presented above is somewhat theoretical, but there is also more direct empirical evidence suggesting that ρ_{uw} is negative. Campbell (1991) uses a vector-autoregressive approach to decompose unexpected stock market returns into components due to cash flow shocks, $\eta_{C,t+1}$, and discount rate shocks, $\eta_{E,t+1}$, as in our equation (22). He considers two subperiods, 1927–1951 and 1952–1988. Since our empirical sample in Section 4. begins in 1952, we can use Campbell’s results from the 1927–1951 period as one source of prior information about ρ_{uw} . Based on quarterly data (which we also use in our empirical work), Campbell estimates in his Table 2

⁷Our argument is not fully precise because Shiller and LeRoy and Porter analyze the unconditional variance of stock returns, whereas $\text{Var}(u_{t+1})$ in equation (25) represents the conditional variance. But the difference between the two variances is relatively small because the variance of μ_t is generally agreed to be much smaller than the variance of u_{t+1} in equation (1). Moreover, we can allow plenty of margin for error on both sides of our argument. On one side, $\rho_{uw} > 0$ implies $\text{Var}(u_{t+1}) < \text{Var}(\eta_{C,t+1})$ with as many as three inequality signs in equation (25). On the other side, the evidence of Shiller and others that $\text{Var}(u_{t+1}) > \text{Var}(\eta_{C,t+1})$ seems quite strong (see Shiller’s Figure 1).

that $\sigma(\eta_{E,t+1}) > \sigma(\eta_{C,t+1})$ in 1927–1951, meaning that discount rate news is more important than cash flow news in explaining the variance of stock market returns. This result makes the condition (24) hold trivially, since $\rho(\eta_{C,t+1}, \eta_{E,t+1}) < 1$. Moreover, Campbell obtains negative estimates of $\rho(\eta_{C,t+1}, \eta_{E,t+1})$ for the 1927–1951 period, which again makes (24) hold trivially independent of $\sigma(\eta_{E,t+1})/\sigma(\eta_{C,t+1})$. Campbell’s empirical results from a sample period that predates our sample therefore provide further support to the prior belief that $\rho_{uw} < 0$.

In fact, prior empirical evidence points to large negative estimates of ρ_{uw} . Table 2 of Campbell (1991) reports estimates of the variance of $\eta_{E,t+1}$, $\sigma_{\eta_E}^2$, and its covariance with $\eta_{C,t+1}$, $\sigma(\eta_C, \eta_E)$, both as fractions of σ_u^2 , from which the implied estimates of $\rho(u_{t+1}, \eta_{E,t+1})$ can be computed as $[\sigma(\eta_C, \eta_E)/\sigma_u^2 - \sigma_{\eta_E}^2/\sigma_u^2] / [\sigma_{\eta_E}/\sigma_u]$. Given equation (23), we interpret these values as implied estimates of ρ_{uw} . For the 1927–1951 period, Campbell’s results imply values of ρ_{uw} ranging from -0.67 to -0.87 across three different specifications. In 1952–1988, the implied estimates of ρ_{uw} range from -0.92 to -0.94, and in the full sample, 1927–1988, they range from -0.71 to -0.86.

Campbell’s evidence on ρ_{uw} is not definitive because it treats VAR-based estimates of expected returns as the true expected returns. Campbell uses three predictors: the dividend-price ratio (D/P), the lagged stock return, and the (relative) one-month T-bill rate. He reports that his results are sensitive to the exclusion of D/P, but they are robust as long as D/P is included among the predictors. Campbell and Ammer (1993) use seven predictors, including D/P, and report estimates in their Table III that imply even more negative estimates of ρ_{uw} , ranging from -0.93 to -0.95 in postwar data. Van Binsbergen and Koijen (2007) estimate ρ_{uw} ranging from -0.67 to -0.45, with an average of -0.60, in postwar data. They do not rely on prespecified predictors but instead use data on dividend growth and returns in the context of a present-value model. They also find a positive correlation between shocks to expected return and dividend growth, similar to Menzly, Santos, and Veronesi (2004), Lettau and Ludvigson (2005), and Kothari, Lewellen, and Warner (2006). Note that $\rho(\eta_{C,t+1}, \eta_{E,t+1})$ can be positive without violating the condition in equation (24).

While we believe that $\rho_{uw} < 0$ is a sensible prior belief, we entertain three different priors on ρ_{uw} , including a noninformative prior, in our empirical work in Section 4.

3.2. The Role of ρ_{uw} in Determining Expected Returns

The correlation between expected and unexpected returns, ρ_{uw} , plays a critical role in determining the conditional expected return. To illustrate this role, we consider a special case of the predictive system in which there are no predictors. With no predictors, D_t includes only the return history, and the conditional expected return in equation (8) is simply $E(r_{t+1}|D_t) =$

$E(r) + \sum_{s=0}^{\infty} \lambda_s \epsilon_{t-s}$, a weighted sum of past forecast errors in returns (the Wold representation). Panel A of Figure 1 plots the values of λ_s in an example with the predictive R-squared (the fraction of the variance in r_{t+1} explained by μ_t) equal to 0.05, β equal to 0.9, and four different values of ρ_{uw} ranging from -0.99 to 0. The figure shows that different values of ρ_{uw} produce different values of m , and hence also different behaviors for $\lambda_s (= m\beta^s)$.

The results in Figure 1 can be understood by noting that there are essentially two effects of the return history on the current expected return. The first might be termed the “level” effect. Observing recent realized returns that were higher than expected suggests that they were generated from a distribution with a higher mean. If the expected return is persistent, as it is in this example with $\beta = 0.9$, then that recent history suggests that the current mean is higher as well. So the level effect positively associates past forecast errors in returns with expected future returns. The second effect, which might be termed the “change” effect, operates via the correlation between expected and unexpected returns. In particular, suppose ρ_{uw} is negative, as we suggest is reasonable. Then observing recent realized returns that were higher than expected suggests that expected returns fell in those periods. That is, part of the reason that realized returns were higher than expected is that there were price increases associated with negative shocks to expected future returns and thus to discount rates applied to expected future cash flows. So the change effect negatively associates past forecast errors in returns with expected future returns. Overall, the net impact of the return history on the current return depends on the relative strengths of the level and change effects.

The level and change effects can be mapped into the return autocovariance in (17). When ρ_{uw} is sufficiently negative, then $\beta\sigma_{\mu}^2 < -\sigma_{uw}$, returns are negatively autocorrelated, and the change effect prevails. Also, $m < 0$ in that case, so the λ_s 's in (9) are negative. When $\beta\sigma_{\mu}^2 > -\sigma_{uw}$, returns are positively autocorrelated, the λ_s 's are positive, and the level effect prevails.

When $\rho_{uw} = 0$, there is no change effect and only the level effect is present. For that case, the λ_s 's in Figure 1 start at a positive value for the first lag, about 0.04, and then decay toward zero. The level and change effects offset each other when $\rho_{uw} = -0.47$ (this is the knife-edge case of zero autocorrelation in equation (17)), or when the fraction of the variance in unexpected returns explained by expected-return shocks, ρ_{uw}^2 , is about 22%. In that case, the λ_s 's plot as a flat line at zero. This result is worth emphasizing: for $\rho_{uw} = -0.47$, rational investors who know the unconditional expected return do not update their beliefs about the conditional expected return, regardless of what realized returns they observe. The change effect dominates when $\rho_{uw} = -0.85$, where the λ_s 's start around -0.04 at the first lag, and it is even stronger when $\rho_{uw} = -0.99$, where the λ_s 's start around -0.08. Clearly, the correlation between expected and unexpected returns is a critical determinant of the relation between the return history and the current expected return.

The conditional expected return depends on the true unconditional mean, $E(r)$, which must be estimated in practice. A natural estimator is the sample mean. Consider again the no-predictor case where the summation on the right-hand side of equation (11) is truncated at $s = t - 1$ and $E(r)$ is replaced by the sample mean, $(1/t) \sum_{l=1}^t r_l$. The estimated conditional expected return then becomes a weighted average of past returns,

$$E(r_{t+1}|D_t) = \sum_{s=0}^{t-1} \kappa_s r_{t-s}, \quad (26)$$

where

$$\kappa_s = \frac{1}{t} \left(1 - \sum_{l=1}^t \omega_l \right) + \omega_s, \quad (27)$$

and $\sum_{s=0}^{t-1} \kappa_s = 1$. The weights (κ_s 's) are plotted in Panel B of Figure 1 for $t = 208$, corresponding to the number of quarters used in our empirical analysis. When $\rho_{uw} = 0$, all past returns enter positively but recent returns are weighted more heavily. In the $\rho_{uw} = -0.47$ case, where the level and change effects exactly offset each other, all of the weights equal $1/t$, so the conditional expected return is then just the historical sample average. For the larger negative ρ_{uw} values, where the change effect is stronger, the weights switch from negative at more recent lags to positive at more distant lags (as the weights must sum to one). For example, when changes in expected returns explain about 72% of the variance in unexpected returns ($\rho_{uw} = -0.85$), the returns from the most recent 10 years (40 quarters) contribute negatively to the estimated current expected return, while the returns from the earlier 42 years contribute positively.

An additional perspective on the role of ρ_{uw} is provided by the time series of conditional expected returns plotted in Figure 2. In constructing these series, we maintain the same setting and same parameter values as in Figure 1. The unconditional mean return $E(r)$ is set equal to the sample average for our 208-quarter sample period, and then, starting from the first quarter in the sample, the conditional mean is updated through time using the finite-sample Kalman filter applied to the realized returns data. As before, the level and change effects exactly offset each other when $\rho_{uw} = -0.47$, so the conditional expected return in that case is simply a flat line at the sample average for the period. A striking feature of the plot is that the expected return series for $\rho_{uw} = 0$ is virtually the mirror image of the series for $\rho_{uw} = -0.85$. Moreover, the differences among the various series of conditional expected returns are large in economic terms, often several percent per quarter. As before, we see that ρ_{uw} plays a key role in estimating expected returns.

Figure 3 compares the R^2 's from three approaches to predicting r_{t+1} using some or all of the information observable at time t , which includes the history of returns and a single predictor x_t . The first approach is the predictive system, which uses all of that available history. The

second is the predictive regression, which uses only the current value x_t . The third approach is the ARMA(1,1) model in equation (18), which uses only past returns. The four panels correspond to the values $\{0, 0.3, 0.6, 0.9\}$ for ρ_{vw} , the conditional correlation between μ_t and the single predictor x_t . In all four panels, $\beta = A = 0.9$, and the true predictive R^2 (from the regression of r_{t+1} on μ_t) is 0.05. We consider two values of ρ_{uv} , “high” and “low”, which correspond to partial correlations between u_t and v_t given w_t of $\rho_{uv|w} = 0.9$ and $\rho_{uv|w} = -0.9$, respectively.⁸

Since all three approaches compared in Figure 3 use only information observable at time t , they all produce R^2 's smaller than 0.05. The R^2 from the predictive regression rises from 0 to 0.04 as ρ_{vw} rises from 0 to 0.9 across the four panels. This increase is intuitive: as μ_t and x_t become more highly correlated, the predictive regression becomes more useful in predicting returns. The predictive regression R^2 is invariant to ρ_{uw} . In contrast, the R^2 from the ARMA(1,1) model, which summarizes the usefulness of past returns in predicting future returns, is heavily influenced by ρ_{uw} . When $\rho_{uw} = -0.47$, this R^2 is zero: past returns contain no information about future returns because the level and change effects cancel out. For $\rho_{uw} \neq -0.47$, stock returns are serially correlated and the ARMA(1,1) R^2 is positive; in fact, it can be higher than the predictive regression R^2 . For example, when $\rho_{vw} = 0.3$ and $\rho_{uw} \notin (-0.74, -0.13)$, past returns are more useful than x_t in predicting r_{t+1} . The highest R^2 's are invariably achieved by the predictive system, which uses more information to predict future returns than do the other two approaches.

4. Empirical Analysis

In this section we use the predictive system to conduct an empirical analysis of return predictability. We first present evidence from predictive regressions, for benchmark purposes. Then we discuss identification issues and use the system to estimate expected returns via maximum likelihood. Finally, we turn to the main analysis, which takes a Bayesian approach.

4.1. Evidence from Predictive Regressions

We begin by estimating predictive regressions on quarterly data in 1952–2003 for three predictors. The first predictor is the market-wide dividend yield, which is equal to total dividends paid over the previous 12 months divided by the current total market capitalization. We compute the dividend yield from the with-dividend and without-dividend monthly returns on the value-weighted

⁸We specify the partial correlation $\rho_{uv|w}$ instead of the simple correlation ρ_{uv} because our control over ρ_{uv} is limited. The permissible range of values for ρ_{uv} depends on ρ_{vw} and ρ_{uw} (see equation (32)) and we vary both ρ_{vw} and ρ_{uw} . By choosing $\rho_{uv|w}$ of 0.9 and -0.9, we are specifying ρ_{uv} close to the boundaries of its permissible range.

portfolio of all NYSE, Amex, and Nasdaq stocks, which we obtain from the Center for Research in Security Prices (CRSP) at the University of Chicago. The second predictor is CAY from Lettau and Ludvigson (2001), whose updated quarterly data we obtain from Martin Lettau’s website. The third predictor is the “bond yield,” which we define as minus the yield on the 30-year Treasury bond in excess of its most recent 12-month moving average. The bond yield data are from the Fixed Term Indices in the CRSP Monthly Treasury file. The three predictors are used to predict quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a one-month T-bill, which is also obtained from CRSP.

Whereas the first two predictors have been used extensively, the third predictor appears to be new. It seems plausible for the long-term T-bond yield to be related to future stock returns since expected returns on stocks and T-bonds may comove due to discount-rate-related factors. Subtracting the 12-month average yield is an adjustment that is commonly applied to the short-term risk-free rate (e.g., Campbell, 1991, and Campbell and Ammer, 1993).

Table I reports, for various predictive regressions, the estimated slope coefficient vector \hat{b} , the R^2 , and the estimated correlation between unexpected returns and the innovations in expected returns. This correlation, which represents the regression-based counterpart of ρ_{uw} , is computed as $\text{Corr}(e_t, b'v_t)$, following equations (4) and (19). Table I also reports the OLS t -statistics and the bootstrapped p -values associated with these t -statistics as well as with the R^2 .⁹ Panel A reports the full-sample results covering 1952 Q1 – 2003 Q4. Panels B and C report sub-sample results.¹⁰

The results suggest that all three predictors have some forecasting ability. The dividend yield produces the weakest evidence (highest p -values, lowest R^2 s) in all three sample periods. When included as the single predictor, the dividend yield is marginally significant in the full sample (p -value of 5.7%). It is significant in the first subperiod ($p = 1.4\%$) but not in the second subperiod ($p = 40.9\%$). The significance of the dividend yield weakens further when the other two predictors are included in the predictive regression.

In contrast, both the bond yield and CAY are highly significant predictors. When used alone,

⁹In the bootstrap, we repeat the following procedure 20,000 times: (i) Resample T pairs of (\hat{v}_t, \hat{e}_t) , with replacement, from the set of OLS residuals from regressions (4) and (19); (ii) Build up the time series of x_t , starting from the unconditional mean and iterating forward on equation (4), using the OLS estimates $(\hat{\theta}, \hat{A})$ and the resampled values of \hat{v}_t ; (iii) Construct the time series of returns, r_t , by adding the resampled values of \hat{e}_t to the sample mean (i.e., under the null that returns are not predictable); (iv) Use the resulting series of x_t and r_t to estimate regressions (4) and (19) by OLS. The bootstrapped p -value associated with the reported t -statistic (or R^2) is the relative frequency with which the reported quantity is smaller than its 20,000 counterparts bootstrapped under the null of no predictability.

¹⁰Since we use the T-bond and T-bill yields in our analysis, we begin our sample in 1952, after the 1951 Treasury-Fed accord that made possible the independent conduct of monetary policy. Campbell and Ammer (1993), Campbell and Yogo (2006), and others also begin their samples in 1952 for this reason.

both predictors exhibit p -values of 0.1% or less in the full sample, and they are also significant in both subperiods. If judged by the p -values, CAY is the stronger predictor in the first subperiod but the bond yield is stronger in the second subperiod. When all three predictors are used together, both CAY and the bond yield are highly and about equally significant in the full sample.

In addition to the p -values and R^2 s, it is also informative to examine the correlations between the expected and unexpected returns, $\text{Corr}(b'v_t, e_t)$, shown in the fourth column of Table I. When the single predictor is either the dividend yield or CAY, these correlations are negative and highly significant: -91.9% for the dividend yield and -53.6% for CAY in the full sample. These negative correlations are not surprising since both predictors are negatively related to stock prices, by construction. For the bond yield, however, this correlation is positive and highly significant in all three sample periods, ranging from 21.7% to 25.1%. This positive correlation makes it unlikely that the bond yield is perfectly correlated with the true conditional expected return.

The correlation between expected and unexpected returns is a useful diagnostic that should be considered when examining the output of a predictive regression. Since this correlation is likely to be negative, predictive models in which this correlation is positive seem less plausible.¹¹ The model in which the bond yield is the single predictor is a good example. Based on the predictive-regression p -value, the bond yield would appear to be a highly successful predictor whose forecasting ability is better than that of the dividend yield and comparable to that of CAY. However, the bond yield produces expected return estimates whose innovations are positively correlated with unexpected returns, suggesting that this predictor is imperfect. We suspect that the same statement can be made about many macroeconomic variables that the literature has related to expected returns. In the rest of the paper, we develop a predictive framework that allows us to incorporate the prior belief that the correlation between expected and unexpected returns is negative.

4.2. Identification and Maximum Likelihood Estimation

In the absence of any priors or parameter restrictions, not all of the parameters in equations (3) through (6) are identified. We can nevertheless obtain estimates of conditional expected returns using equation (4) and the recursive representation for returns,

$$r_{t+1} = (1 - \beta)E(r) + \beta r_t + n'v_t - (\beta - m)\epsilon_t + \epsilon_{t+1}, \quad (28)$$

¹¹Strictly speaking, the arguments based on equations (22) and (24) apply when r_{t+1} denotes the total stock return, but they should hold to a close approximation also when r_{t+1} denotes the excess stock return, as used here. For excess returns, Campbell (1991) shows that equation (22) has an additional term representing news about future interest rates, and he estimates the variance of that term to be an order of magnitude smaller than the variances of $\eta_{C,t+1}$ and $\eta_{E,t+1}$.

which follows directly from the steady-state representation of the conditional expected return in (8). The parameters in (4) and (28) are identified and can be estimated using maximum likelihood, by representing those two equations as a state-space system and applying standard methodology (e.g., Hamilton, 1994, section 13.4).¹² The parameters in those equations, along with the covariance matrix of $[\epsilon \ v_t']$, identify the parameters appearing in equations (3) through (5) but not all of the parameters in the covariance matrix in (6). Only Σ_{vv} is identified just by the data. Identifying the remaining elements of Σ requires additional information about at least one of them.¹³

Figure 4 plots the time series of expected returns obtained via maximum likelihood estimation as well as the expected-return estimates obtained from OLS estimation of the predictive regression. Panels A and B display results with a single predictor, either the dividend yield or CAY. In Panel C, those variables are combined with the bond-yield variable in the three-predictor case. First, observe that the fluctuation of the expected return estimates seems too large to be plausible. In Panel B, for example, expected returns range from -5% to 8% per quarter, and the range is even wider in Panel C. Later on, we obtain smoother time series of μ_t by specifying informative prior beliefs. Second, observe that although the series of estimated expected returns exhibit marked differences across the three sets of predictors, the differences between the predictive-regression estimates and the predictive-system estimates for a given set of predictors are much smaller.¹⁴

As the length of the sample grows, posterior beliefs about the parameters in (28) and thus (8) will converge to values that do not depend on prior beliefs about the parameters in the predictive system (as long as those priors do not strictly preclude such values). Therefore, after observing a sufficiently long sample, prior beliefs about ρ_{uw} , for example, will not impact forecasts of future returns. (Our actual sample is evidently not long in that sense, as prior beliefs about ρ_{uw} exert a substantial effect on estimates of expected returns.) On the other hand, given the lack of full identification of Σ , prior beliefs about ρ_{uw} will matter even in large samples when making inferences about the correlation between the predictors and the true unobservable expected return μ_t .

4.3. Bayesian Approach

We develop a Bayesian approach for estimating the predictive system. This approach has several advantages over frequentist alternatives such as the maximum likelihood approach. First, the

¹²The likelihood function is detailed in the Appendix.

¹³Rytchkov (2007) discusses identification issues in a similar setting.

¹⁴We also estimate expected returns from the predictive system under diffuse priors (the discussion of prior beliefs follows later in the text). We find that the resulting estimates (not plotted here) behave similarly to both the OLS estimates from the predictive regression and the maximum likelihood estimates from the predictive system.

Bayesian approach allows us to specify economically motivated prior distributions for the parameters of interest. Second, it produces posterior distributions that deliver finite-sample inferences about relatively complicated functions of the underlying parameters, such as the correlations between μ_t and x_t and the R^2 s from the regression of r_{t+1} on μ_t . Finally, it incorporates parameter uncertainty as well as uncertainty about the path of the unobservable expected return μ_t .

We obtain posterior distributions using *Gibbs sampling*, a Markov Chain Monte Carlo (MCMC) technique (e.g., Casella and George, 1992). In each step of the MCMC chain, we first draw the parameters $(\theta, A, \alpha, \beta, \Sigma)$ conditional on the current draw of $\{\mu_t\}$, and then we use the *forward filtering, backward sampling* algorithm developed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994) to draw the time series of $\{\mu_t\}$ conditional on the current draw of $(\theta, A, \alpha, \beta, \Sigma)$.

We impose informative prior distributions on three quantities: the correlation ρ_{uw} between expected and unexpected returns, the persistence β of the true expected return μ_t , and the predictive R^2 from the regression of r_{t+1} on μ_t . These prior distributions are plotted in Figure 5.

The key prior distribution is the one on ρ_{uw} . We consider three priors on ρ_{uw} , all of which are plotted in Panel A of Figure 5. The “noninformative” prior is flat on most of the $(-1, 1)$ range, with prior mass tailing off near ± 1 to avoid potential singularity problems. The “less informative” prior imposes $\rho_{uw} < 0$ in that 99.9% of the prior mass of ρ_{uw} is below zero. As shown in Panel B, this prior implies a relatively noninformative prior on ρ_{uw}^2 , with most prior mass between 0 and 0.8. Finally, the “more informative” prior on ρ_{uw} is specified such that the implied prior on ρ_{uw}^2 has 99.9% of its mass above 0.5, with a mean of about 0.77. Since ρ_{uw}^2 is the R^2 from the regression of unexpected returns on shocks to expected returns, it represents the fraction of market variance that is due to news about discount rates. Therefore, the more informative prior reflects the belief that at least half of the variance of market returns is due to discount rate news. This belief is motivated by empirical evidence. For example, the evidence of Campbell (1991) implies estimates of ρ_{uw}^2 ranging from 0.50 to 0.74 across three different specifications in his full sample period 1927–1988. Campbell’s evidence from 1927–1951, a sample period that predates ours, implies estimates of ρ_{uw}^2 ranging from 0.44 to 0.76 (ρ_{uw} ranging from -0.67 to -0.87). The estimates of ρ_{uw}^2 implied by the postwar evidence are even larger; they range from 0.84 to 0.88 in 1952–1988, and the estimates of Campbell and Ammer (1993) in their Table III range from 0.86 to 0.91. All of these estimates are in line with the more informative prior.

Putting a prior on ρ_{uw} presents a technical challenge. We do not impose the standard inverted Wishart prior on the covariance matrix Σ because such a prior would be informative about all elements of Σ , not only about ρ_{uw} , and we do not wish to be informative about the elements that involve v_t . Instead, we build on Stambaugh (1997) and form the prior on Σ as the posterior from

a hypothetical sample that contains more information about u_t and w_t than about v_t . In addition, we develop a hyperparameter approach that allows us to change the prior on ρ_{uw} without changing the priors on any other parameters. The details are in the Appendix.

In addition to putting a prior on ρ_{uw} , we also impose a prior belief that the conditional expected return μ_t is stable and persistent. To capture the belief that μ_t is stable, we impose a prior that the predictive R^2 from the regression of r_{t+1} on μ_t is not very large, which is equivalent to the belief that the total variance of μ_t is not very large. The prior on the R^2 , which is plotted in Panel C of Figure 5, has a mode close to 1%, most of its mass is below 5%, and there is very little prior mass above 10%. To capture the belief that μ_t is persistent, we impose a prior that β , the slope of the AR(1) process for μ_t , is smaller than one but not by much.¹⁵ The prior on β , which is plotted in Panel D of Figure 5, has most of its mass above 0.7 and there is virtually no prior mass below 0.5. We do not impose a prior belief that $\mu_t > 0$. Although such a belief is reasonable under a fully rational view, we do not wish to preclude the possibility that some of the variation in μ_t is driven by investor sentiment. The prior distributions on all other parameters (θ , A , α , and most elements of Σ) are noninformative. Separately, we also consider a “diffuse” prior, which is completely noninformative about all model parameters, including ρ_{uw} , β , and R^2 .

4.4. Predictive System vs. Predictive Regression

In contrast to a predictive regression, the predictive system allows us to conduct finite-sample inferences that explicitly incorporate predictor imperfection. The predictive system also produces more precise inferences about expected returns. To demonstrate this, we compare the explanatory powers of the system and the regression for a broad range of parameter values. Specifically, we compare the R^2 in the regression of r_{t+1} on x_t for the predictive regression with the R^2 in the regression of r_{t+1} on $E(r_{t+1}|D_t) \equiv E(\mu_t|D_t)$ for the predictive system. The ratio of these R^2 values when r_{t+1} is the dependent variable is the same as when μ_t is the dependent variable,

$$\frac{R^2(r_{t+1} \text{ on } x_t)}{R^2(r_{t+1} \text{ on } E(\mu_t|D_t))} = \frac{R^2(\mu_t \text{ on } x_t)}{R^2(\mu_t \text{ on } E(\mu_t|D_t))}, \quad (29)$$

since each of the R^2 values in the latter ratio is equal to its corresponding value in the first ratio multiplied by $\text{Var}(r_{t+1})/\text{Var}(\mu_t)$. The parameters in equations (3) through (6) can be used to obtain the covariance matrix of μ_t and x_t and thereby the R^2 in the regression of μ_t on x_t ,

$$R^2(eg) = \frac{\text{Var}[E(\mu_t|x_t)]}{\text{Var}(\mu_t)}. \quad (30)$$

¹⁵Ferson, Sarkissian, and Simin (2003, footnote 2) discuss several reasons to believe expected return is persistent.

As shown in the Appendix, we can solve analytically for the steady-state value of $\text{Var}(\mu_t|D_t)$, which allows us to compute the R^2 in the regression of μ_t on $E(\mu_t|D_t)$ as

$$R^2(sys) = \frac{\text{Var}[E(\mu_t|D_t)]}{\text{Var}(\mu_t)} = 1 - \frac{\text{Var}(\mu_t|D_t)}{\text{Var}(\mu_t)}. \quad (31)$$

The ratio in equation (29) is computed as $R^2(reg)/R^2(sys)$. Note that this R^2 ratio cannot exceed 1 because $x_t \in D_t$. In other words, the estimates of μ_t from the predictive system are at least as precise as the estimates from the predictive regression, simply because the system uses more information. The smaller the R^2 ratio, the larger the advantage of using the predictive system.

We use the R^2 ratios to quantify the explanatory advantage of the predictive system, using the same sample as in Section 4.1. Panel A of Table II shows the posterior means and standard deviations of the R^2 ratios for four different priors and four different sets of predictors. First, observe that the posterior means of the R^2 ratios are all comfortably lower than one, ranging from 0.08 to 0.86 across the 16 cases, and from 0.46 to 0.70 when all three predictors are used jointly. Second, the R^2 ratios are sensitive to the prior on ρ_{uw} . For example, with the bond yield as the single predictor, the R^2 ratio is estimated to be 0.73 under the diffuse prior. When we impose the prior belief that ρ_{uw} is negative, the R^2 ratio declines to 0.34 under the less informative prior and then further to 0.08 under the more informative prior. In other words, under the prior that more than half of the market variance is due to discount rate news, the expected return estimates from the predictive system are about 12.5 times more precise than those from the predictive regression. For the dividend yield, we observe the opposite pattern—the R^2 ratio increases from 0.28 to 0.59 to 0.81 for the same priors. The opposite patterns result from the opposite effects that the prior on ρ_{uw} has on the adequacy of x_t as a predictor in the two cases, as we will see later.

Panel B of Table II shows the posterior means and standard deviations of one minus the ratio of the mean squared errors from the predictive system and the predictive regression. The mean squared error is defined as $MSE = E\{(r_{t+1} - f_t)^2\}$, where f_t is a return forecast. All posterior means in Panel B are positive, ranging from 0.01 to 0.17, confirming that the predictive system forecasts returns more precisely than the predictive regression does.

Another way of comparing the predictive system with the predictive regression is to compare their estimates of the slope coefficient b from the predictive regression. Figure 6 plots the posterior distributions of b computed under three scenarios. The dashed line is the posterior of b computed from the predictive regression under no prior information. This posterior has a Student t distribution whose mean is equal to the maximum likelihood estimate (MLE) of b (Zellner, 1971, pp. 65–67). The dashed line thus represents “conventional inference” on predictability. The other

two lines in Figure 6 plot the implied posteriors of b computed from the predictive system.¹⁶ The dotted line corresponds to the prior that is noninformative about ρ_{uw} but informative about β and R^2 . In all three panels of Figure 6, the dotted line is substantially different from the dashed line, which means that imposing the prior that μ_t is stable and persistent significantly affects the inference about predictability. In addition, the dotted line is shifted toward zero compared to the dashed line, which means that the prior belief that μ_t is stable and persistent weakens the evidence of predictability. Finally, the solid line corresponds to the prior that is informative not only about β and R^2 but also about ρ_{uw} . The prior on ρ_{uw} clearly affects the inference about predictability. Consider Panel A, in which the single predictor is the bond yield. Whereas the traditional inference (dashed line) would conclude with almost 100% certainty that the bond yield is a useful predictor ($b > 0$), the system-based inference with the more informative prior on ρ_{uw} (solid line) concludes no such thing because almost half of the posterior mass of b is below zero. This prior also slightly weakens the predictive power of CAY but it strengthens the predictive power of the dividend yield.

4.5. How Imperfect Are Predictors?

The predictive system also allows us to learn about the correlation between the expected return μ_t and the predictors. Since μ_t is not observed, the manner in which one learns about such correlations merits some discussion. Consider, for simplicity, the case of a single predictor x_t whose autocorrelation A is equal to β . The unconditional correlation between the expected return and the predictor, $\rho_{x\mu}$, is then equal to ρ_{vw} , the conditional correlation.¹⁷ By virtue of the fact that the correlation matrix for $(u_t \ v_t \ w_t)$ must be non-negative definite, it is readily verified that

$$\rho_{vw} = \rho_{uv}\rho_{uw} + \rho_{\Delta}, \quad \text{where} \quad \rho_{\Delta}^2 \leq (1 - \rho_{uv}^2)(1 - \rho_{uw}^2). \quad (32)$$

In other words, even though correlations are not transitive (two correlations don't imply the third), they become nearly transitive when at least one of them approaches ± 1 .

We specify noninformative priors for ρ_{vw} and ρ_{uv} . Doing otherwise would most likely involve priors about each predictor's usefulness—directly through ρ_{vw} but indirectly through ρ_{uv} as well, given informative priors about ρ_{uw} . While such an approach could be reasonable, especially in a forecasting setting, we wish to illustrate here how our framework can deliver inferences about each

¹⁶Although b does not appear explicitly in the predictive system, its value can be computed from the system's parameters (see footnote 5), so its posterior draws can be constructed from the draws of the system's parameters.

¹⁷More generally,

$$\rho_{x\mu} = \rho_{vw} \left[\frac{(1 - \beta^2)(1 - A^2)}{(1 - \beta A)^2} \right]^{\frac{1}{2}},$$

so that $\rho_{x\mu}^2 \leq \rho_{vw}^2$.

predictor's usefulness without making prior judgments about that property. Moreover, we prefer not to add such complexity to this initial exploration of predictive systems. The data are quite informative about ρ_{uv} anyway, in that with only modest predictability in returns, the value of ρ_{uv} is close to that of ρ_{rv} , which can be estimated from the series of r_t and x_t . (When the predictive R^2 is low, $\rho_{ru}^2 = (1 - R^2)$ is close to one, and equation (32) then implies that ρ_{uv} is well approximated by ρ_{rv} .) Information about ρ_{uw} enters largely through the prior. When the prior is concentrated on large negative values, then the likely values of ρ_{Δ} in (32) are small, so the prior information about ρ_{uw} and the sample information about ρ_{uv} get combined to provide information about ρ_{vw} . Alternatively, if the data indicate that ρ_{uv} is close to ± 1 (e.g., in Table I, $\rho_{uv} \approx -0.9$ when the predictor is the dividend yield), then again ρ_{Δ} is likely to be small, so that ρ_{uv} and ρ_{uw} are again jointly informative about ρ_{vw} .

Figures 7 and 8 analyze the degree to which two of our predictors can capture the unobservable true expected return μ_t . We report results for two predictive systems, in which the predictors are the dividend yield alone (Figure 7) and the bond yield alone (Figure 8). Panel A of each figure plots the posterior distribution of the R^2 from the regression of μ_t on x_t . This R^2 is assumed to be one in a predictive regression, but its posterior in the predictive system has very little mass at values close to one. In both figures, the R^2 s larger than 0.8 receive very little posterior probability and the values larger than 0.9 are deemed almost impossible, regardless of the prior. This evidence suggests that neither of the two predictors is likely to be perfectly correlated with μ_t .¹⁸

The R^2 depends on the prior for ρ_{uw} in an interesting way. In Panel A of Figure 7, becoming increasingly informative about ρ_{uw} shifts the posterior of the R^2 to the right, with the mode shifting from about 0.3 under the noninformative prior to about 0.6 under the more informative prior. This makes sense: since the dividend yield exhibits a highly negative contemporaneous correlation with stock returns, imposing a prior that μ_t also possesses such negative correlation makes the dividend yield more closely related to μ_t . Exactly the opposite happens in Panel A of Figure 8, where becoming increasingly informative about ρ_{uw} shifts the posterior of the R^2 to the left so that its mode is close to zero under the more informative prior. This makes sense as well because the bond yield is positively correlated with stock returns (Table I).

Panel B of both figures plots the posterior of the predictive R^2 from the regression of r_{t+1} on μ_t . Putting a more informative prior on ρ_{uw} increases the R^2 in both figures, but these effects are relatively small. Since we put a fairly informative prior on the predictive R^2 (see Panel C of Figure

¹⁸Note that even if x_t were perfectly correlated with μ_t in population, the posterior of their correlation would have nontrivial mass below one in any finite sample. Since we always observe finite samples, we always perceive imperfect correlation between x_t and μ_t . Also note that in the NBER version of this article, we report results analogous to those in Figures 8 and 9 for a predictive system that uses three predictors: the dividend yield, bond yield, and CAY.

5), the posterior is not dramatically different from the prior in either figure.

Panels C and D of both figures plot the posteriors of the correlations between each predictor and μ_t , both conditional (ρ_{vw}) and unconditional ($\rho_{x\mu}$). These correlations are all well below 1 and they are quite sensitive to the prior on ρ_{uw} . As we become increasingly informative about ρ_{uw} , we perceive the dividend yield to be more highly correlated with μ_t but the bond yield to be less highly correlated with μ_t . The effect for the bond yield is dramatic, judging by the posterior modes in Panel C of Figure 8. Under the noninformative prior, the bond yield has 90% conditional correlation with μ_t , but under the more informative prior, this correlation drops to 20%.

Overall, Figures 8 and 9 show that our predictors are imperfectly correlated with μ_t and that the inference about this correlation is substantially affected by the prior beliefs about ρ_{uw} . Prior beliefs informed by economic principles strengthen the predictive appeal of the dividend yield but they weaken the predictive appeal of the bond yield.

4.6. Estimates of Expected Return

Figure 9 plots the time series of expected returns estimated by three different approaches. The dashed line plots the fitted values from the predictive regression. These traditional expected return estimates seem too volatile to be plausible, as we also observed in Figure 4. For example, in Panel C, which includes all three predictors, expected returns range from -6% to 9% per quarter. Not surprisingly, imposing the prior that μ_t is stable and persistent (dotted line) produces smoother expected return estimates. Adding the more informative prior on ρ_{uw} (solid line) further smoothes the expected return estimates: in Panel C, they range from -1.5% to 3.5% per quarter. The informative priors have substantial effects on expected returns not only in Panel C but also in Panel B in which CAY is the single predictor: while the regression-fitted values range from -5.5% to 7.5% per quarter, the solid line ranges from -1.5% to 2.5%. Only in Panel A, in which the dividend yield is the single predictor, the effect of the prior is relatively mild. The reason is that the regression-fitted values in Panel A are already fairly smooth and negatively correlated with stock returns.

While eyeballing the expected return estimates seems informative, we also compute measures summarizing their differences. Table III compares five different series of expected return estimates. The first is the series of fitted values from the predictive regression, and the others are produced by four different approaches to estimating the predictive system. One of the latter approaches estimates the predictive system by MLE, while the other three impose the prior that μ_t is stable and persistent but differ in their prior on ρ_{uw} . We compare the five series of expected return estimates in three different ways: pairwise correlations, mean absolute differences, and average utility losses.

The utility losses are computed for a mean-variance investor allocating between the market and the T-bill who knows the variance of market returns but must estimate the market's expected return. The investor's risk aversion, 2.54, is such that the optimal portfolio is fully invested in the market, on average. We compute the investor's certainty equivalent loss resulting from holding a portfolio that is optimal under a different approach for estimating expected returns. For example, the 0.11% per quarter average utility loss in the first row of Panel A is suffered by an investor who wants to estimate expected return in the predictive system by MLE but is forced to use the fitted values from the predictive regression. Finally, the three panels consider three different sets of predictors: the dividend yield, CAY, and the two predictors combined with the bond yield.

Panel A of Table III shows that when the dividend yield is the single predictor, the expected return estimates are fairly similar across the five estimation approaches, confirming the evidence from Panel A of Figure 9. No average utility loss exceeds 0.20% per quarter, no mean absolute difference is larger than 0.65% per quarter, and all correlations exceed 81%. We also observe that imposing informative priors makes the system-based estimates closer to the regression-based estimates. For example, the utility losses fall monotonically from 0.11% to 0.03% as move from column two to column five in the first row of Panel A.

The differences across the five approaches are substantially larger in Panel B where we use CAY to predict returns. For example, compare the system-based estimates obtained by MLE versus the more informative prior. The mean absolute difference in expected returns is 1.65% per quarter and the average certainty equivalent loss from using one estimate in place of the other is 1.38% per quarter. Both quantities are highly economically significant. In Panel C, where we use all three predictors, the differences across the five approaches are also large and similar in magnitude.

In all three panels, the smallest differences are obtained for the noninformative versus the less informative prior on ρ_{uw} . No average utility loss exceeds 0.06% per quarter, no mean absolute difference is larger than 0.37% per quarter, and all correlations exceed 95.4%. However, moving from the less informative to the more informative prior on ρ_{uw} can produce sizeable differences in expected returns. For example, the mean absolute difference in Panel C is 1.46% per quarter and the average utility loss is 0.84% per quarter.

To sum up, when we use the dividend yield as the single predictor, the system-based expected return estimates are close to the regression-based estimates. In all other cases, the system and the regression generate substantially different expected returns, and the system-based estimates are significantly affected by the prior on ρ_{uw} .

4.7. Variance Decomposition of Expected Return

In the predictive regression approach, expected return μ_t is modeled as an exact linear function of the predictors in x_t . In a predictive system, however, the data provide additional information about μ_t because the lagged values of unexpected returns and predictor innovations also enter the expected return estimates (see Section 3.2.). In this section, we decompose the variance of μ_t to assess the relative importance of the various sources of information in a predictive system.

We can rewrite the AR(1) process for x_t as an MA(∞) process, as we did for μ_t in equation (14): $x_t = E_x + \sum_{i=0}^{\infty} A^i v_{t-i}$. Then we project w_t linearly on u_t and v_t :

$$w_t = [\sigma_{wu} \ \sigma_{wv}] \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}^{-1} \begin{bmatrix} u_t \\ v_t \end{bmatrix} + \chi_t = \psi_u u_t + \psi_v v_t + \chi_t. \quad (33)$$

Substituting for w_t from equation (33) into equation (14), we obtain

$$\mu_t = (E(r) - \psi_v E(x)) + \psi_v x_t + \psi_u \sum_{i=0}^{\infty} \beta^i u_{t-i} + \psi_v \sum_{i=0}^{\infty} (\beta^i I_K - A^i) v_{t-i} + \sum_{i=0}^{\infty} \beta^i \chi_{t-i}, \quad (34)$$

where K is the number of predictors and I_K is a $K \times K$ identity matrix. Equation (34) shows how the lagged values of unexpected returns u_{t-i} and predictor innovations v_{t-i} affect μ_t in the presence of the current predictor values in x_t . Based on this equation, we can decompose the variance of μ_t into the components due to x_t , $\{u_s\}_{s \leq t}$, and $\{v_s\}_{s \leq t}$. See the Appendix for details.

Table IV reports the posterior means and standard deviations of the R^2 s from the regressions of μ_t on x_t (column 1), μ_t on x_t and $\{u_s\}_{s \leq t}$ (column 2), and μ_t on x_t and $\{u_s, v_s\}_{s \leq t}$ (column 3). We consider four sets of predictors x_t : the dividend yield, bond yield, CAY, and the combination of all three predictors. For each set of predictors, we estimate the predictive system under three different priors. All three priors assume that μ_t is stable and persistent but they differ in their degree of informativeness about ρ_{uw} .

First, note that x_t never accounts for more than 63% of the variance of μ_t and that it can account for as little as 3% of this variance. In contrast, x_t combined with $\{u_s, v_s\}_{s \leq t}$ can account for as much as 95% of the variance of μ_t , and those components account for more than 80% of the variance in 10 of the 12 cases in Table IV. The most striking effect obtains for the bond yield, for which adding $\{u_s, v_s\}_{s \leq t}$ to x_t increases the R^2 from 0.03 to 0.95. It seems clear that a predictive regression, which uses only x_t to predict returns, does not use the data as effectively as a predictive system, which also uses $\{u_s, v_s\}_{s \leq t}$ in addition to x_t .

The R^2 's in Table IV are substantially affected by the prior on ρ_{uw} . For example, consider the first columns of Panels A and B. Under the noninformative prior on ρ_{uw} , both the dividend yield

and the bond yield explain about a third of the variance of μ_t . As we become more informative about ρ_{uw} , this fraction increases from 0.34 to 0.40 to 0.57 for the dividend yield, but it decreases from 0.33 to 0.24 to 0.03 for the bond yield. These opposite patterns reflect the opposite signs of the correlations between stock returns and the two predictors, as explained earlier.

The lagged unexpected returns $\{u_s\}_{s \leq t}$ contain a significant amount of information about μ_t beyond that included in x_t . When $\{u_s\}_{s \leq t}$ is added to x_t in estimating μ_t , the R^2 's increase by anywhere between 7% and 83%. For example, under the more informative prior on ρ_{uw} , the R^2 increases from 0.03 to 0.86 for the bond yield, from 0.53 to 0.87 for CAY, and from 0.63 to 0.85 when $\{u_s\}_{s \leq t}$ is added to all three predictors. The lagged predictor innovations $\{v_s\}_{s \leq t}$ also contain useful information about μ_t . When $\{v_s\}_{s \leq t}$ is added to x_t and $\{u_s\}_{s \leq t}$, the R^2 's increase by between 1% and 41%. The smallest increases, of 1% to 5%, obtain for the dividend yield, while the largest increases, of 9% to 41%, obtain for all three predictors combined.

To summarize, the past values of unexpected returns and predictor innovations contain useful incremental information about the current expected return. This information is used by the predictive system but not by the standard predictive regression.

4.8. Variance of Multiperiod Returns

In this final section, we analyze the effect of prior beliefs about ρ_{uw} on the variance of multiperiod returns. Let $r_{T,T+k}$ denote the multiperiod log return in periods T through $T+k$. Consider the problem of assessing the variance of $r_{T,T+k}$ conditional on the sample information D_T . Stambaugh (1999) and Barberis (2000) analyze this “predictive” variance in a VAR setting in which expected return is given by a predictive regression. We examine this variance in our setting with imperfect predictors, and find that it is substantially affected by prior beliefs about ρ_{uw} .

From the definition $r_{T,T+k} = r_{T+1} + r_{T+2} + \dots + r_{T+k}$, the conditional moments are

$$E(r_{T,T+k}|\Theta_T) = kE(r) + \frac{1-\beta^k}{1-\beta}(\mu_T - E(r)) \quad (35)$$

$$\begin{aligned} \text{Var}(r_{T,T+k}|\Theta_T) &= k\sigma_u^2 + \frac{\sigma_w^2}{(1-\beta)^2} \left[k-1 - 2\beta \frac{1-\beta^{k-1}}{1-\beta} + \beta^2 \frac{1-\beta^{2(k-1)}}{1-\beta^2} \right] \\ &\quad + \frac{2\sigma_{uw}}{1-\beta} \left[k-1 - \beta \frac{1-\beta^{k-1}}{1-\beta} \right], \end{aligned} \quad (36)$$

where Θ_T comprises μ_T and the parameters of the predictive system. The predictive variance is

$$\text{Var}(r_{T,T+k}|D_T) = E[\text{Var}(r_{T,T+k}|\Theta_T)|D_T] + \text{Var}[E(r_{T,T+k}|\Theta_T)|D_T], \quad (37)$$

by variance decomposition. We use repeated draws from the posterior distribution of Θ_T to compute the two posterior moments on the right-hand side of equation (37).

Figure 10 plots the predictive k -period variance on a per-period basis (i.e., divided by k), obtained from the predictive system with the dividend yield as the predictor. The figure shows that the prior for ρ_{uw} has a large effect on the long-run variance. For example, at the five-year investment horizon, the variance is 0.0051 under the more informative prior, but it is 0.0068, or 33% larger, under the noninformative prior. In addition, the effect of ρ_{uw} increases with the investment horizon. At the one-year horizon, the variance under the noninformative prior exceeds its counterpart under the more informative prior by 9%, but this difference grows to 33% at the five-year horizon.

The prior for ρ_{uw} affects the long-run variance by affecting the perceived mean reversion in stock returns. When ρ_{uw} is sufficiently negative, high (low) realized returns tend to mean-revert because they tend to be accompanied by decreases (increases) in expected future returns. The effect of mean reversion is stronger when ρ_{uw} takes larger negative values. As a result, the long-horizon variance in Figure 10 is at its lowest under the more informative prior for ρ_{uw} .

Figure 10 also shows that the predictive variance is a U-shaped function of the investment horizon. This pattern is an outcome of several forces, including mean reversion in stock returns, uncertainty about future values of μ_t , and parameter uncertainty. Analyzing these forces in detail is beyond the scope of this study, but we are presently exploring them in separate work.

5. Conclusions

Predictive systems allow predictors to be imperfectly correlated with the conditional expected return. When predictors are imperfect, expected returns conditional on available data depend not only on the most recent values of those predictors but also on lagged returns and lags of the predictors. Recent returns receive negative weights when a significant portion of the variance in unexpected returns is due to changes in expected returns. The lags of returns and predictors often account for a large fraction of the variation in estimates of conditional expected returns.

Predictive systems also allow one to incorporate a prior belief that expected and unexpected returns are negatively correlated. We find that such a belief has an important impact on estimates of expected returns and various inferences about predictability, including a predictor's correlation with expected return as well as the variance of returns over longer investment horizons.

Although our focus is on predictive systems, we also find two implications for predictive re-

gressions. First, we show that if predictors are imperfect, the predictive regression residuals are autocorrelated. This autocorrelation should be incorporated when computing standard errors in predictive regressions. In addition, this autocorrelation provides a simple diagnostic for predictor imperfection: non-zero autocorrelation indicates imperfect predictors. Second, we argue that researchers running predictive regressions should examine the regression-implied correlation between expected and unexpected returns. Predictive regressions in which this correlation is positive are unlikely to perfectly capture time variation in expected stock market returns.

Our initial exploration of predictive systems can be extended in many directions. First, we are intentionally noninformative about the degree of imperfection in a predictor, but one could instead incorporate an informative prior belief about a predictor's correlation with expected return. The latter approach is likely to be preferable when inference is less the objective than is producing the best forecast given one's own prior judgment. Along these lines, one could study the implications of predictive systems for asset allocation. Second, we assume that the conditional mean return follows an AR(1) process, but it would also make sense to consider more complicated processes. For example, if the mean were allowed to have not only a slow-moving persistent component but also a higher-frequency transient component, the bond yield, which is not very persistent, might be inferred to be more highly correlated with the conditional mean. Third, we assume that the return variance is constant, but one could allow it to be time-varying, potentially in a manner correlated with expected return (e.g., Brandt and Kang, 2004). Fourth, we consider three predictors but it would also be interesting to examine the degrees of imperfection in various other predictors that have been proposed in the literature (e.g., Ferson and Harvey (1991), Lamont (1998), Lewellen (1999), Ang and Bekaert (2007), Santos and Veronesi (2006), etc.). Fifth, we analyze predictability in U.S. stock market returns, but it would also be interesting to apply predictive systems to international markets (e.g., Ferson and Harvey, 1993).

It could also be useful to expand the predictive system to incorporate cash flow news. We have argued that the innovation in the expected return should be negatively correlated with the unexpected return, but if one could account for the portion of the latter that is correlated with cash flow news, the remaining portion would be driven entirely by news about expected return. These issues are beyond the scope of this paper but they merit more attention. See Cochrane (2008), Rytchkov (2007), and van Binsbergen and Koijen (2007) for recent analyses of the interaction between return predictability and cash flow predictability. Cash flow forecasts also enter the calculations of the implied cost of capital, which is used by Pástor, Sinha, and Swaminathan (2008) to proxy for the conditional expected market return.

One might ask whether the predictive system produces out-of-sample forecasts with lower

mean squared error (MSE) than a simpler approach such as a predictive regression or the sample average.¹⁹ A Bayesian investor with a quadratic (MSE) loss function would prefer a forecast that combines his priors and the data to estimate the conditional expected return based on the correct model. The correct model, when estimated using a finite sample, tends to produce out-of-sample MSEs higher than those from estimates of simpler models when the true degree of predictability is sufficiently small, as discussed by Clark and West (2006, 2007) and Hjalmarsson (2006). Thus, a simple comparison of out-of-sample MSEs would not speak directly to the question of whether the predictive system is the right model from the investor's perspective. That question, one of model selection, is beyond the scope of this study but could be an interesting area for future research.

¹⁹Goyal and Welch (2003, 2006), Campbell and Thompson (2008), and Rapach, Strauss and Zhou (2007), among others, investigate the abilities of predictive regressions and sample averages to forecast stock returns out of sample.

Appendix.

We begin working with a generalized version of the predictive system with more than one asset, so that r_t , x_t , and μ_t are all vectors. To maintain the usual convention that matrices are denoted by uppercase letters, we replace β by B , the σ 's by the corresponding Σ 's, etc.

We restate the predictive system from equations (3) through (5) here in the multi-asset case:

$$r_{t+1} = \mu_t + u_{t+1} \quad (\text{A1})$$

$$x_{t+1} = \theta + Ax_t + v_{t+1} \quad (\text{A2})$$

$$\mu_{t+1} = \alpha + B\mu_t + w_{t+1}, \quad (\text{A3})$$

with the disturbances distributed identically and independently across t as

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix} \right). \quad (\text{A4})$$

Let D_0 denote the null information set, so that the unconditional moments are given as

$$\begin{bmatrix} r_t \\ x_t \\ \mu_t \end{bmatrix} | D_0 \sim N \left(\begin{bmatrix} E_r \\ E_x \\ E_r \end{bmatrix}, \begin{bmatrix} V_{rr} & V_{rx} & V_{r\mu} \\ V_{xr} & V_{xx} & V_{x\mu} \\ V_{\mu r} & V_{\mu x} & V_{\mu\mu} \end{bmatrix} \right). \quad (\text{A5})$$

Let z_t denote the vector of the observed data at time t , $z_t = (r_t' x_t')'$. Denote the data we observe through time t as $D_t = (z_1, \dots, z_t)$, and note that our complete data consist of D_T . Also define

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\mu} = \begin{bmatrix} V_{r\mu} \\ V_{x\mu} \end{bmatrix}. \quad (\text{A6})$$

From the above we obtain

$$E_r = (I - B)^{-1}\alpha, \quad V_{rr} = V_{\mu\mu} + \Sigma_{uu}, \quad V_{rx} = V_{\mu x}A' + \Sigma_{uv}, \quad (\text{A7})$$

$$E_x = (I - A)^{-1}\theta, \quad V_{xx} = AV_{xx}A' + \Sigma_{vv}, \quad V_{\mu x} = BV_{\mu x}A' + \Sigma_{wv}, \quad (\text{A8})$$

$$V_{\mu\mu} = BV_{\mu\mu}B' + \Sigma_{ww}, \quad V_{\mu r} = BV_{\mu\mu} + \Sigma_{wu}. \quad (\text{A9})$$

Given the well known identity $\text{vec}(DFG) = (G' \otimes D)\text{vec}(F)$, we can write

$$\text{vec}(V_{xx}) = [I - (A \otimes A)]^{-1}\text{vec}(\Sigma_{vv}) \quad (\text{A10})$$

$$\text{vec}(V_{\mu\mu}) = [I - (B \otimes B)]^{-1}\text{vec}(\Sigma_{ww}) \quad (\text{A11})$$

$$\text{vec}(V_{\mu x}) = [I - (A \otimes B)]^{-1}\text{vec}(\Sigma_{wv}). \quad (\text{A12})$$

Drawing the time series of μ_t

To draw the time series of the unobservable values of μ_t conditional on the current parameter draws, we apply the *forward filtering, backward sampling* (FFBS) approach developed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). See also West and Harrison (1997, chapter 15).

Filtering

The first stage follows the standard methodology of Kalman filtering. Define

$$a_t = E(\mu_t | D_{t-1}) \quad b_t = E(\mu_t | D_t) \quad e_t = E(z_t | \mu_t, D_{t-1}) \quad (\text{A13})$$

$$f_t = E(z_t | D_{t-1}) \quad P_t = \text{Var}(\mu_t | D_{t-1}) \quad Q_t = \text{Var}(\mu_t | D_t) \quad (\text{A14})$$

$$R_t = \text{Var}(z_t | \mu_t, D_{t-1}) \quad S_t = \text{Var}(z_t | D_{t-1}) \quad G_t = \text{Cov}(z_t, \mu_t' | D_{t-1}) \quad (\text{A15})$$

Conditioning on the (unknown) parameters of the model is assumed throughout but suppressed in the notation for convenience. First note that

$$\mu_0 | D_0 \sim N(b_0, Q_0), \quad (\text{A16})$$

where $b_0 = E_r$ and $Q_0 = V_{\mu\mu}$,

$$\mu_1 | D_0 \sim N(a_1, P_1), \quad (\text{A17})$$

where $a_1 = E_r$ and $P_1 = V_{\mu\mu}$, and

$$z_1 | D_0 \sim N(f_1, S_1), \quad (\text{A18})$$

where $f_1 = E_z$ and $S_1 = V_{zz}$. Note that

$$G_1 = V_{z\mu} \quad (\text{A19})$$

and that

$$z_1 | \mu_1, D_0 \sim N(e_1, R_1), \quad (\text{A20})$$

where

$$e_1 = f_1 + G_1 P_1^{-1} (\mu_1 - a_1) \quad (\text{A21})$$

$$R_1 = S_1 - G_1 P_1^{-1} G_1'. \quad (\text{A22})$$

Combining this density with equation (A17) using Bayes rule gives

$$\mu_1 | D_1 \sim N(b_1, Q_1), \quad (\text{A23})$$

where

$$b_1 = a_1 + P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1} (z_1 - f_1) \quad (\text{A24})$$

$$Q_1 = P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} P_1. \quad (\text{A25})$$

Continuing in this fashion, we find that all conditional densities are normally distributed, and we obtain all the required moments for $t = 2, \dots, T$:

$$a_t = \alpha + B b_{t-1} \quad (\text{A26})$$

$$P_t = B Q_{t-1} B' + \Sigma_{ww} \quad (\text{A27})$$

$$f_t = \begin{bmatrix} b_{t-1} \\ \theta + A x_{t-1} \end{bmatrix} \quad (\text{A28})$$

$$S_t = \begin{bmatrix} Q_{t-1} + \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} \quad (\text{A29})$$

$$G_t = \begin{bmatrix} Q_{t-1} B' + \Sigma_{uw} \\ \Sigma_{vw} \end{bmatrix} \quad (\text{A30})$$

$$e_t = f_t + G_t P_t^{-1} (\mu_t - a_t) \quad (\text{A31})$$

$$R_t = S_t - G_t P_t^{-1} G_t' \quad (\text{A32})$$

$$b_t = a_t + P_t (P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1} (z_t - f_t) \quad (\text{A33})$$

$$= a_t + G_t' S_t^{-1} (z_t - f_t) \quad (\text{A34})$$

$$Q_t = P_t (P_t + G_t' R_t^{-1} G_t)^{-1} P_t. \quad (\text{A35})$$

The values of $\{a_t, b_t, Q_t, P_t\}$ for $t = 1, \dots, T$ are retained for the next stage.

Sampling

Let $\zeta_t = [r_t \ x_t \ \mu_t]'$. We wish to draw $(\mu_0, \mu_1, \dots, \mu_T)$ conditional on D_T . The backward-sampling approach relies on the Markov property of the evolution of ζ_t and the resulting identity,

$$p(\zeta_0, \zeta_1, \dots, \zeta_T | D_T) = p(\zeta_T | D_T) p(\zeta_{T-1} | \zeta_T, D_{T-1}) \cdots p(\zeta_1 | \zeta_2, D_1) p(\zeta_0 | \zeta_1, D_0). \quad (\text{A36})$$

We first sample μ_T from $p(\mu_T | D_T)$, the normal density obtained in the last step of the filtering. Then, for $t = T-1, T-2, \dots, 1, 0$, we sample μ_t from the conditional density $p(\zeta_t | \zeta_{t+1}, D_t)$. (Note that the first two subvectors of ζ_t are already observed and thus need not be sampled.) To obtain that conditional density, first note that

$$\zeta_{t+1} | D_t \sim N \left(\begin{bmatrix} b_t \\ \theta + Ax_t \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix} \right), \quad (\text{A37})$$

$$\zeta_t | D_t \sim N \left(\begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} \right), \quad (\text{A38})$$

and

$$\text{Cov}(\zeta_t, \zeta_{t+1}' | D_t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix}. \quad (\text{A39})$$

Therefore,

$$\zeta_t | \zeta_{t+1}, D_t \sim N(h_t, H_t), \quad (\text{A40})$$

where

$$h_t = \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix} \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} r_{t+1} - b_t \\ x_{t+1} - \theta - Ax_t \\ \mu_{t+1} - a_{t+1} \end{bmatrix}$$

and

$$H_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix} \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix}'$$

The mean and covariance matrix of μ_t are taken as the relevant elements of h_t and H_t .

Expected returns and past values

In this section, we derive the equations (8), (11), and (26). We still work in the general case in which r_t is a vector of returns rather than a scalar. Therefore, to continue denoting matrices by uppercase letters, we replace m by M , n by N , λ by Λ , ϕ by Φ , δ by Δ , ω by Ω , and κ by \mathcal{K} .

Below, we express the vector of conditional expected returns, $b_t = E(r_{t+1}|D_t)$, as a function of past returns and predictors. Denote

$$[M_t \ N_t] \equiv P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1} = G_t' S_t^{-1}, \quad (\text{A41})$$

so that, from equation (A33), for $t > 1$,

$$\begin{aligned} b_t &= a_t + [M_t \ N_t](z_t - f_t) \\ &= \alpha + Bb_{t-1} + [M_t \ N_t] \begin{bmatrix} r_t - b_{t-1} \\ x_t - \theta - Ax_{t-1} \end{bmatrix} \\ &= (I - B)E_r + (B - M_t)b_{t-1} + M_t r_t + N_t v_t, \end{aligned} \quad (\text{A42})$$

or

$$b_t - E_r = B(b_{t-1} - E_r) + M_t(r_t - b_{t-1}) + N_t v_t. \quad (\text{A43})$$

For $t = 1$, we obtain

$$b_1 - E_r = M_1(r_1 - b_0) + N_1 v_1,$$

where v_1 denotes $x_1 - E_x$. Repeated substitution for the lagged values of $(b_t - E_r)$ gives

$$b_t = E_r + \sum_{s=1}^t \Lambda_s(r_s - b_{s-1}) + \sum_{s=1}^t \Phi_s v_s, \quad (\text{A44})$$

where

$$\Lambda_s = B^{t-s} M_s \quad (\text{A45})$$

$$\Phi_s = B^{t-s} N_s. \quad (\text{A46})$$

That is, the expected return conditional on data observed through period t can be written as the unconditional mean E_r plus a linear combination of past return forecast errors, $\epsilon_s = r_s - b_{s-1}$, plus a linear combination of past innovations in the predictors. This is equation (8) in the text.

The conditional expected return b_t can be rewritten so that past forecast errors are replaced by returns in excess of the unconditional mean E_r . To do so, modify equation (A42) as

$$b_t - E_r = (B - M_t)(b_{t-1} - E_r) + M_t(r_t - E_r) + N_t v_t \quad (\text{A47})$$

so that repeated substitution for the lagged values of $(b_t - E_r)$ then yields

$$b_t = E_r + \sum_{s=1}^t \Omega_s (r_s - E_r) + \sum_{s=1}^t \Delta_s v_s \quad (\text{A48})$$

where

$$\Omega_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) M_s & \text{for } s < t \\ M_s & \text{for } s = t \end{cases} \quad (\text{A49})$$

$$\Delta_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) N_s & \text{for } s < t \\ N_s & \text{for } s = t \end{cases} \quad (\text{A50})$$

That is, b_t is then equal to the unconditional mean return E_r plus linear combinations of past returns in excess of E_r and past innovations in the predictors. This is equation (11) in the text.

If E_r is replaced by the sample mean in equation (11), then the estimate of b_t becomes

$$\hat{b}_t = \sum_{s=1}^t \mathcal{K}_s r_s + \sum_{s=1}^t \Delta_s v_s, \quad (\text{A51})$$

where

$$\mathcal{K}_s = \frac{1}{t} \left(I - \sum_{l=1}^t \Omega_l \right) + \Omega_s, \quad (\text{A52})$$

and $\sum_{s=1}^t \mathcal{K}_s = I$. This is a generalized version of equation (26) in the text.

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which r_t is a scalar. This simplification turns μ_t , α , and B into scalars as well. Therefore, we now turn back to the notation from the text in which B is replaced by β and the relevant Σ 's by σ 's.

Drawing the parameters

This section describes how we obtain the posterior draws of all parameters conditional on the current draw of the time series of μ_t .

Prior distributions

First, we discuss the prior on $(\theta, A, \alpha, \beta)$. We require both x_t and μ_t to be stationary, so that all eigenvalues of A must lie inside the unit circle and $\beta \in (-1, 1)$. Apart from this restriction, our prior is noninformative about A but informative about β , $\beta \sim N(0.99, 0.15^2)$ (see Figure 5). We reparameterize the model to replace the intercepts θ and α by the unconditional means of μ_t and x_t , which we denote by E_μ and E_x , respectively. The equations (4) and (5) then read $x_{t+1} = E_x + A(x_t - E_x) + v_{t+1}$ and $\mu_{t+1} = E_\mu + \beta(\mu_t - E_\mu) + w_{t+1}$. This reparameterization

allows us to increase the speed of convergence of our MCMC chain by putting a mildly informative prior on E_μ , $E_\mu \sim N(\bar{\mu}, \sigma_{E_\mu}^2)$, centered at the sample mean return with a large prior standard deviation of 1% per quarter. We use a noninformative prior for E_x , $E_x \sim N(0, \sigma_{E_x}^2 I_K)$ with a large σ_{E_x} . All four parameters, A , β , E_μ , and E_x , are independent a priori.

The prior on Σ is more complicated. We divide the elements of Σ into two subsets: first, the 2×2 submatrix $\Sigma_{11} \equiv [\sigma_u^2 \ \sigma_{uw}; \sigma_{wu} \ \sigma_w^2]$, and second, the elements of Σ that involve v : $\Sigma_{(v)} \equiv (\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$. We choose a prior that is informative about Σ_{11} but noninformative about $\Sigma_{(v)}$. Such a prior is obtained as a posterior of Σ when a noninformative prior is updated with a hypothetical sample in which there are T_0 observations of (u, w) but only $S_0 \ll T_0$ observations of v (see Stambaugh, 1997). We choose T_0 equal to one fifth of the sample size, which makes the prior on Σ_{11} informative (five times less informative than the actual sample). We choose $S_0 = K + 3$, where K is the number of predictors, which makes the prior on $\Sigma_{(v)}$ virtually noninformative (as informative as a sample of only $K + 3$ observations, where $K = 1$ or 3).

The prior on Σ_{11} is inverted Wishart, $\Sigma_{11} \sim IW(T_0 \hat{\Sigma}_{11,0}, T_0 - K - 1)$, so the prior mean is $E(\Sigma_{11}) = \hat{\Sigma}_{11,0} (T_0 / (T_0 - K - 4))$. Denote the (i, j) element of $\hat{\Sigma}_{11,0}$ by M_{ij} , for $i = 1, 2$ and $j = 1, 2$. The value of M_{11} is chosen such that the prior mean of σ_u^2 is equal to 95% of the sample variance of market returns. The value of M_{22} is chosen to deliver the prior mean of σ_w^2 which, combined with β of 0.97, sets the variance of μ_t equal to 5% of the sample variance of market returns. These values of M_{11} and M_{22} lead to a prior for the R^2 from the regression of r_{t+1} on μ_t that we find plausible (see Figure 5). To be able to put different priors on ρ_{uw} while keeping the same prior on σ_u^2 and σ_w^2 , we adopt a hyperparameter approach. We assume that M_{12} is an unknown hyperparameter with a uniform prior distribution on the interval $(-\underline{c} \sqrt{M_{11} M_{22}}, \bar{c} \sqrt{M_{11} M_{22}})$. Since the prior mean of ρ_{uw} is approximately equal to $M_{12} / \sqrt{M_{11} M_{22}}$, this prior mean is approximately uniformly distributed as $U(-\underline{c}, \bar{c})$. For all three priors on ρ_{uw} , we specify $\underline{c} = -0.90$ and we vary \bar{c} as follows: 0.9 for the noninformative prior, -0.35 for the less informative prior, and -0.87 for the more informative prior. These choices produce the priors on ρ_{uw} plotted in Figure 5.

The prior on $\Sigma_{(v)}$ is obtained by changing variables from $(\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$ to the slope C ($3 \times K$) and the residual covariance matrix Ω ($K \times K$) from the regression of v_t on (u_t, w_t) : $C = (0 \ C_2)'$ (zero intercept), $C_2 = [\sigma_{vu} \ \sigma_{vw}] \Sigma_{11}^{-1}$, and $\Omega = \Sigma_{vv} - C_2 \Sigma_{11} C_2'$. We put a normal-inverted-Wishart prior on C and Ω : $\Omega \sim IW(S_0 \hat{\Omega}_0, S_0 - 1)$ and $\text{vec}(C) | \Omega \sim N(\hat{c}_0, \Omega \otimes (X_0' X_0)^{-1})$, where $\hat{\Omega}_0$, \hat{c}_0 , and $X_0' X_0$ represent estimates from the S_0 periods in the hypothetical sample in which both v_t and (u_t, w_t) are available. The choices of $\hat{\Omega}_0$ and \hat{c}_0 are inconsequential because they represent means of distributions with large variances. The prior variances are large, for two reasons. First, we choose a very small value for S_0 , as explained above. Second, we choose $X_0' X_0$ equal to S_0 times a diagonal matrix whose $(1, 1)$ element is one and the remaining diagonal elements are tiny positive numbers (so $(X_0' X_0)^{-1}$ is large). That is, we choose a hypothetical sample in which v is much more volatile than (u, w) in the short overlapping period of S_0 observations. As a result, the priors on C and Ω are noninformative.

As mentioned above, these priors on Σ_{11} , C , and Ω can be thought of as posteriors. After changing variables from Σ to (Σ_{11}, C, Ω) , the diffuse prior on Σ , $p(\Sigma) \propto |\Sigma|^{-(K+3)/2}$, translates into $p(\Sigma_{11}, C, \Omega) \propto |\Sigma_{11}|^{(K-3)/2} |\Omega|^{-(K+3)/2}$. When this noninformative prior is updated with

the hypothetical sample of T_0 observations of (u, w) and S_0 observations of v , the posteriors of Σ_{11} , C , and Ω are exactly the same as the priors described above. See Stambaugh (1997).

Posterior distributions

Drawing $(\theta, A, \alpha, \beta)$ given Σ

After changing variables from (θ, α) into (E_x, E_μ) , the equations (4) and (5) can be written as

$$\underbrace{\begin{pmatrix} x_{t+1} \\ \mu_{t+1} \end{pmatrix}}_{q_{t+1}} - \underbrace{\begin{pmatrix} A & 0 \\ 0 & \beta \end{pmatrix}}_{L_1} \underbrace{\begin{pmatrix} x_t \\ \mu_t \end{pmatrix}}_{q_t} - \underbrace{\begin{pmatrix} I_K - A & 0 \\ 0 & 1 - \beta \end{pmatrix}}_{L_2} \underbrace{\begin{pmatrix} E_x \\ E_\mu \end{pmatrix}}_{E_{x\mu}} = \begin{pmatrix} v_{t+1} \\ w_{t+1} \end{pmatrix},$$

where the covariance matrix of the residuals is $\Sigma_{(vw)} \equiv [\Sigma_{vv} \ \sigma_{vw}; \sigma_{vw} \ \sigma_w^2]$. The prior for $E_{x\mu}$ is

$$E_{x\mu} \sim N(E_{x\mu_0}, V_{x\mu_0}),$$

where $E_{x\mu_0} \equiv (0 \ \bar{\mu})'$ and $V_{x\mu_0} \equiv [\sigma_{E_x}^2 I_K \ 0; 0 \ \sigma_{E_\mu}^2]$. Since both the prior and the likelihood are normally distributed, the full conditional posterior distribution of $E_{x\mu}$ is also normal,

$$E_{x\mu} | \cdot \sim N(\tilde{E}_{x\mu}, \tilde{V}_{x\mu}), \quad (\text{A53})$$

where $\tilde{V}_{x\mu} = (V_{x\mu_0}^{-1} + TL_2' \Sigma_{(vw)}^{-1} L_2)^{-1}$ and $\tilde{E}_{x\mu} = \tilde{V}_{x\mu} [V_{x\mu_0}^{-1} E_{x\mu_0} + L_2' \Sigma_{(vw)}^{-1} \sum_{t=1}^T (q_{t+1} - L_1 q_t)]$.

Let $x^k \equiv (x_2^k, \dots, x_T^k)'$ denote the $(T-1) \times 1$ vector of realizations of predictor k in periods $2, \dots, T$, for $k = 1, \dots, K$. Also, let $x^{(l)}$ denote the $(T-1) \times K$ vectors of realizations of all K predictors in periods $1, \dots, T-1$. Similarly, let $\mu \equiv (\mu_2, \dots, \mu_T)'$ and $\mu^{(l)} \equiv (\mu_1, \dots, \mu_{T-1})'$, and let E_{x^k} be the k -th element of E_x . Denote

$$z = \begin{pmatrix} x^1 - \iota_{T-1} E_{x^1} \\ \vdots \\ x^K - \iota_{T-1} E_{x^K} \\ \mu - \iota_{T-1} E_\mu \end{pmatrix}, \quad Z = \begin{pmatrix} x^{(l)} - \iota_{T-1} E_x' & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & x^{(l)} - \iota_{T-1} E_x' & 0 \\ 0 & 0 & 0 & \mu^{(l)} - \iota_{T-1} E_\mu \end{pmatrix},$$

where ι_{T-1} is a $(T-1) \times 1$ vector of ones, the dimensions of z are $[(T-1)(K+1)] \times 1$, and the dimensions of Z are $[(T-1)(K+1)] \times (K^2+1)$. Then we can write the equations (4) and (5) as

$$z = Zb + \text{errors},$$

where $b = (\text{vec}(A)' \ \beta)'$ and the covariance matrix of the error terms is $\Sigma_{(vw)} \otimes I_{T-1}$. The prior distribution on b is given by

$$b \sim N(b_0, V_{b_0}) \times 1_{b \in \mathcal{S}},$$

where b_0 and V_{b_0} are chosen as explained earlier and $1_{b \in \mathcal{S}}$ is equal to one when x_t and μ_t are stationary and zero otherwise. Let $\hat{V}_b = [Z'(\Sigma_{(vw)}^{-1} \otimes I_{T-1})Z]^{-1}$ and $\hat{b} = \hat{V}_b Z'(\Sigma_{(vw)}^{-1} \otimes I_{T-1})z$. The full conditional posterior distribution of b is then given by

$$b | \cdot \sim N(\tilde{b}, \tilde{V}_b) \times 1_{b \in \mathcal{S}}, \quad (\text{A54})$$

where $\tilde{V}_b = (V_{b_0}^{-1} + \hat{V}_b^{-1})^{-1}$ and $\tilde{b} = \tilde{V}_b (V_{b_0}^{-1} b_0 + \hat{V}_b^{-1} \hat{b})$. We obtain the posterior draws of b by making draws from $N(\tilde{b}, \tilde{V}_b)$ and retaining only draws that satisfy $b \in S$. The posterior draws of A and β are constructed from the posterior draws of b from the definition $b = (\text{vec}(A)'\beta)'$.

Drawing Σ given $(\theta, A, \alpha, \beta)$

Recall that we change variables from $\Sigma = [\sigma_u^2 \sigma_{uv} \sigma_{uw}; \sigma_{vu} \Sigma_{vv} \sigma_{vw}; \sigma_{wu} \sigma_{wv} \sigma_w^2]$ to the set of (Σ_{11}, C, Ω) , where $\Sigma_{11} \equiv [\sigma_u^2 \sigma_{uw}; \sigma_{wu} \sigma_w^2]$, and C and Ω are the slope and the residual covariance matrix from the regression of v on (u, w) .

The prior for Σ_{11} is conditional on the hyperparameter M_{12} . This hyperparameter can be drawn from its full conditional posterior density, $p(M_{12}|\cdot, D_t)$, which is given by

$$p(M_{12}|\Sigma_{11}) \propto |\hat{\Sigma}_{11,0}|^{\frac{T_0-K-1}{2}} \exp\left\{-\frac{T_0}{2}\text{tr}(\Sigma_{11}^{-1}\hat{\Sigma}_{11,0})\right\}, \quad M_{12} \in (-\underline{c}\sqrt{M_{11}M_{22}}, \bar{c}\sqrt{M_{11}M_{22}}), \quad (\text{A55})$$

where M_{12} is the (1, 2) element of $\hat{\Sigma}_{11,0}$. Although this is not a density of a well known distribution, we can make posterior draws of M_{12} easily. We approximate this density by a piecewise linear function, using a fine (250-point) grid on the interval $(-\underline{c}\sqrt{M_{11}M_{22}}, \bar{c}\sqrt{M_{11}M_{22}})$. For a random draw $z \sim U(0, 1)$, we find the points on the grid whose cumulative probability densities are immediately above and below z , and we compute the value of M_{12} by linear interpolation.

Conditional on M_{12} , we have the matrix $\hat{\Sigma}_{11,0}$ in the prior distribution for Σ_{11} . In addition, conditional on $(\theta, A, \alpha, \beta)$, we have the sample of the residuals $(u_t, v_t, w_t), t = 1, \dots, T$. Let $Y_{1,T}$ denote the $T \times 2$ matrix of $[u_t \ w_t]$, let $Y_{2,T}$ denote the $T \times K$ matrix of v_t , and let $X = [1_T \ Y_{1,T}]$. The sample estimates from the regression of $Y_{2,T}$ on $Y_{1,T}$ are given by $\hat{C} = (X'X)^{-1}X'Y_{2,T}$, $\hat{\Omega} = (Y_{2,T} - X\hat{C})'(Y_{2,T} - X\hat{C})/T$, and $\hat{\Sigma}_{11} = Y_{1,T}'Y_{1,T}/T$. The posterior of Σ_{11} has an inverted Wishart distribution:

$$\Sigma_{11}|\cdot \sim IW(T_0\hat{\Sigma}_{11,0} + T\hat{\Sigma}_{11}, T + T_0 - K - 1). \quad (\text{A56})$$

In addition, let $V_C = (X_0'X_0 + X'X)^{-1}$, $\tilde{C} = V_C [(X_0'X_0)\hat{C}_0 + (X'X)\hat{C}]$, $\tilde{c} = \text{vec}(\tilde{C})$, and $D = \hat{C}_0'X_0'X_0\hat{C}_0 + \hat{C}'X'X\hat{C} - \tilde{C}'V_C^{-1}\tilde{C}$. The posterior of Ω has an inverted Wishart distribution:

$$\Omega|\cdot \sim IW(S_0\hat{\Omega}_0 + T\hat{\Omega} + D, T + S_0 - 1), \quad (\text{A57})$$

and the conditional posterior of $c = \text{vec}(C)$ is normal:

$$c|\Omega, \cdot \sim N(\tilde{c}, \Omega \otimes V_C). \quad (\text{A58})$$

Given the posterior draws of (Σ_{11}, C, Ω) , we construct the remaining (non- Σ_{11}) elements of Σ as follows: $[\sigma_{vu} \ \sigma_{vw}] = C_2 \Sigma_{11}$ and $\Sigma_{vv} = \Omega + C_2 \Sigma_{11} C_2'$, where $C = (C_1 \ C_2)'$.

Our inference is based on 25,000 draws from the posterior distribution. First, we generate a sequence of 76,000 draws. We discard the first 1,000 draws as a ‘‘burn-in’’ and take every third draw from the rest to obtain a series of 25,000 draws that exhibit little serial correlation. The posterior draws of the relevant quantities such as ρ_{uw} , $\rho_{x\mu}$, $R^2(\mu_t \text{ on } x_t)$, $R^2(r_{t+1} \text{ on } \mu_t)$, etc. are constructed easily from the posterior draws of the basic parameters in the model.

Maximum likelihood estimation

Denote the variance-covariance matrix of the disturbances in equations (4) and (28) as

$$\text{Cov}\left(\begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix}, \begin{bmatrix} \epsilon_t & v_t \end{bmatrix}\right) = \Sigma^* = \begin{bmatrix} \sigma_\epsilon^2 & \sigma'_{v\epsilon} \\ \sigma_{v\epsilon} & \Sigma_{vv} \end{bmatrix}. \quad (\text{A59})$$

Maximum likelihood estimates are computed as the values of E_z , β , m , n , A , σ_ϵ^2 , $\sigma_{v\epsilon}$, and Σ_{vv} that minimize

$$-2 \ln L = \sum_{t=1}^T \left[|V_{t|t-1}| + (z_t - \hat{z}_{t|t-1})' V_{t|t-1}^{-1} (z_t - \hat{z}_{t|t-1}) \right], \quad (\text{A60})$$

where $\hat{z}_{1|0} = E_z$,

$$V_{1|0} = \begin{bmatrix} \sigma_r^2 & \sigma'_{xr} \\ \sigma_{xr} & V_{xx} \end{bmatrix},$$

$$\begin{aligned} \sigma_r^2 &= (1 - \beta^2)^{-1} [n' \Sigma_{vv} n + (1 - \beta^2 + m^2) \sigma_\epsilon^2 + 2m \sigma'_{v\epsilon} n], \\ \sigma_{xr} &= (I - \beta A)^{-1} [A \Sigma_{vv} n + [I - (\beta - m)A] \sigma_{v\epsilon}], \\ \hat{z}_{t|t-1} &= E_z + F_{11}(z_{t-1} - E_z) + F_{12} \Sigma^* V_{t-1|t-2}^{-1} (z_{t-1} - \hat{z}_{t-1|t-2}), \quad t = 2, \dots, T, \\ V_{t|t-1} &= F_{12} (\Sigma^* - \Sigma^* V_{t-1|t-2}^{-1} \Sigma^*) F'_{12} + \Sigma^*, \quad t = 2, \dots, T, \\ F_{11} &= \begin{bmatrix} \beta & 0 \\ 0 & A \end{bmatrix}, \quad F_{12} = \begin{bmatrix} -(\beta - m) & n' \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

and V_{xx} is given by (A10).

The R^2 ratios.

The numerator of the R^2 ratio in equation (29) is computed as

$$R^2(\mu_t \text{ on } x_t) = \frac{\text{Var}(E(\mu_t | x_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(E(\mu_t) + V_{\mu x} V_{xx}^{-1} (x_t - E(x_t)))}{\text{Var}(\mu_t)} = \frac{V_{\mu x} V_{xx}^{-1} V'_{\mu x}}{V_{\mu \mu}}, \quad (\text{A61})$$

where V_{xx} , $V_{\mu\mu}$, and $V_{x\mu}$ are given in equations (A10), (A11), and (A12), respectively.

The denominator of the R^2 ratio in equation (29) is computed as

$$R^2(\mu_t \text{ on } D_t) = \frac{\text{Var}(E(\mu_t | D_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(\mu_t) - \text{Var}(\mu_t | D_t)}{\text{Var}(\mu_t)} = 1 - \frac{Q_t}{V_{\mu \mu}}, \quad (\text{A62})$$

where Q_t is given in equation (A35). We replace Q_t by its steady-state value, Q , which can be shown to be equal to a solution of a quadratic equation:

$$\begin{aligned}
Q &= \frac{\sqrt{\xi_1^2 - 4\xi_2} - \xi_1}{2}, & (A63) \\
\xi_1 &= (1 - \beta^2)(\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu}) + 2\beta(\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu}) - (\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw}) \\
&= (1 - \beta^2)\text{Var}(u|v) + 2\beta\text{Cov}(u, w|v) - \text{Var}(w|v) \\
\xi_2 &= (\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu})^2 - (\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu})(\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw}) \\
&= \text{Cov}(u, w|v)^2 - \text{Var}(u|v)\text{Var}(w|v) < 0
\end{aligned}$$

The value of Q is also used in computing the steady-state values of M_t and N_t from equation (A41), denoted by m_t and n_t in the scalar case:

$$m = (\beta Q + \text{Cov}(u, w|v))(Q + \text{Var}(u|v))^{-1} \quad (A64)$$

$$n = (\sigma_{wv} - m\sigma_{uv})\Sigma_{vv}^{-1}. \quad (A65)$$

Variance decomposition of expected return.

In equation (34), the conditional expected return μ_t depends on three time-varying variables:

1. C1 = x_t , the current predictor values
2. C2 = $\sum_{i=0}^{\infty} \beta^i u_{t-i}$, an infinite sum of current and lagged unexpected returns
3. C3 = $\sum_{i=0}^{\infty} (\beta^i I_K - A^i) v_{t-i}$, an infinite sum of current and lagged predictor innovations, plus an error term. In the variance decomposition in Table IV, we consider regressions of μ_t on various subsets of (C1, C2, C3). Let C denote a given subset of (C1, C2, C3). The R^2 from the regression of μ_t on C is equal to

$$R^2(\mu_t \text{ on } C) = \frac{V'_{\mu C} V_C^{-1} V_{\mu C}}{V_{\mu\mu}}. \quad (A66)$$

The matrix V_C , the covariance matrix of C , is pieced together from

$$\begin{aligned}
\text{Var}(C1) &= V_{xx} \\
\text{Var}(C2) &= \sigma_u^2(1 - \beta^2)^{-1} \\
\text{vec}(\text{Var}(C3)) &= [(1 - \beta^2)^{-1}I_{K^2} - (I_K - \beta A)^{-1} \otimes I_K - I_K \otimes (I_K - \beta A)^{-1} + \\
&\quad + (I_{K^2} - A \otimes A)^{-1}] \text{vec}(\Sigma_{vv}) \\
\text{Cov}(C1, C2) &= (I_K - \beta A)^{-1}\sigma_{vu} \\
\text{Cov}(C2, C3) &= [(1 - \beta^2)^{-1}I_K - (I_K - \beta A)^{-1}] \sigma_{vu} \\
\text{vec}(\text{Cov}(C1, C3')) &= [I_K \otimes (I_K - \beta A)^{-1} + (I_{K^2} - A \otimes A)^{-1}] \text{vec}(\Sigma_{vv}),
\end{aligned}$$

and $V_{\mu C}$, the vector of covariances between μ_t and C , is built from

$$\begin{aligned}
\text{Cov}(\mu_t, C1') &= \Psi_v \text{Var}(C1) + \Psi_u \text{Cov}(C1, C2)' + \Psi_v \text{Cov}(C1, C3')' \\
\text{Cov}(\mu_t, C2) &= \Psi_u \text{Var}(C2) + \Psi_v \text{Cov}(C1, C2) + \Psi_v \text{Cov}(C2, C3) \\
\text{Cov}(\mu_t, C3') &= \Psi_v \text{Var}(C3) + \Psi_v \text{Cov}(C1, C3') + \Psi_u \text{Cov}(C2, C3)'.
\end{aligned}$$

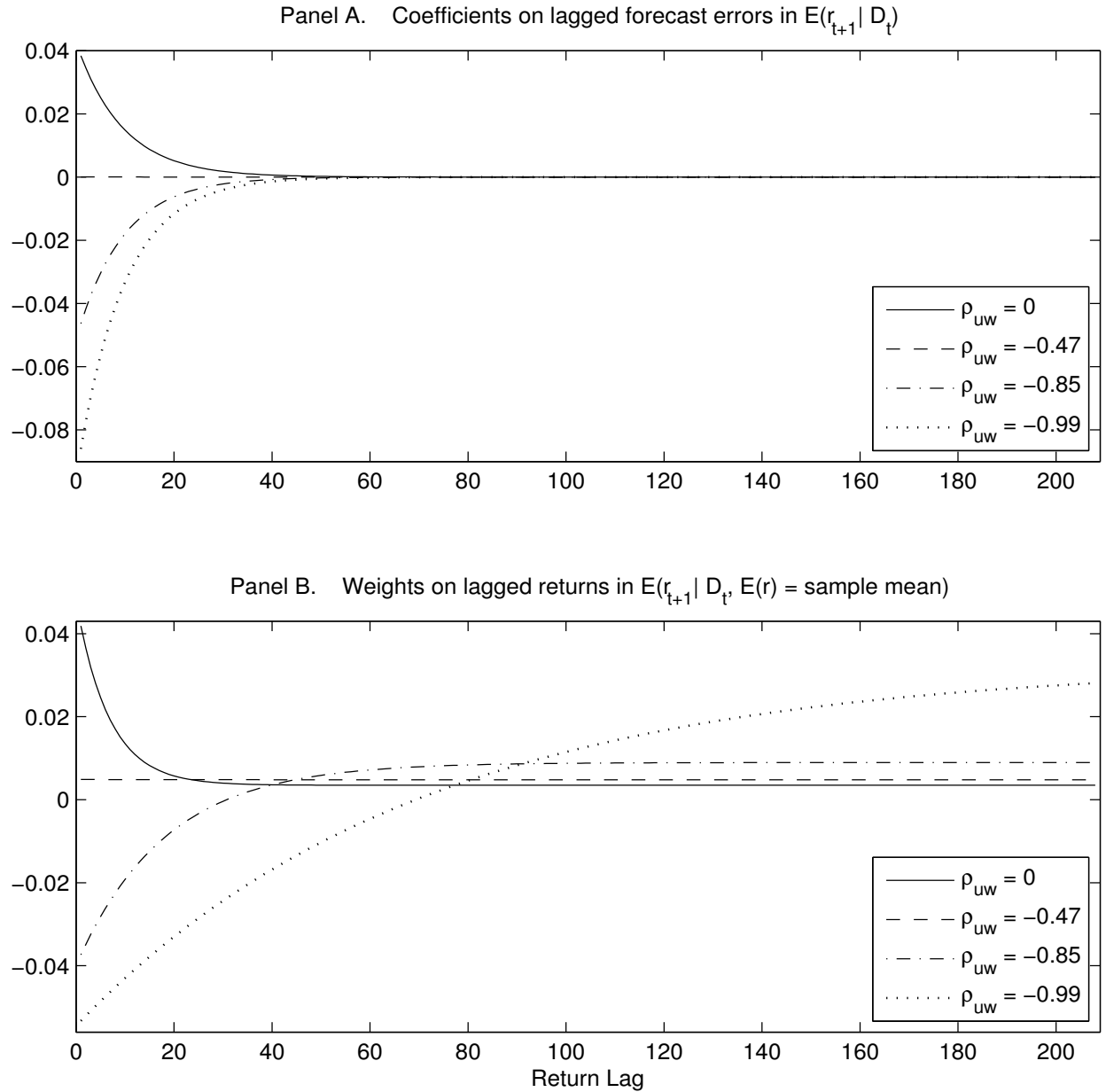


Figure 1. The effect of lagged returns on $E(r_{t+1} | D_t)$ when no predictors are used. Panel A plots λ_s , the coefficients on lagged forecast errors ($\epsilon_{t-s} = r_{t-s} - E(r_{t-s} | D_{t-s-1})$) in $E(r_{t+1} | D_t)$. Panel B plots κ_s , the weights on lagged total returns in $E(r_{t+1} | D_t)$ when the unconditional mean return is estimated by the sample mean over the previous 208 quarters (which is the length of the sample used in subsequent analysis). No predictors are used in the predictive system. The steady-state values of all coefficients are plotted. The different lines correspond to different values of ρ_{uw} , the correlation between expected and unexpected returns. The mean reversion coefficient in the AR(1) process for the conditional expected return μ_t is set equal to $\beta = 0.9$. The predictive R^2 —the fraction of variation in r_{t+1} than can be explained by μ_t —is set equal to $R^2 = 0.05$.

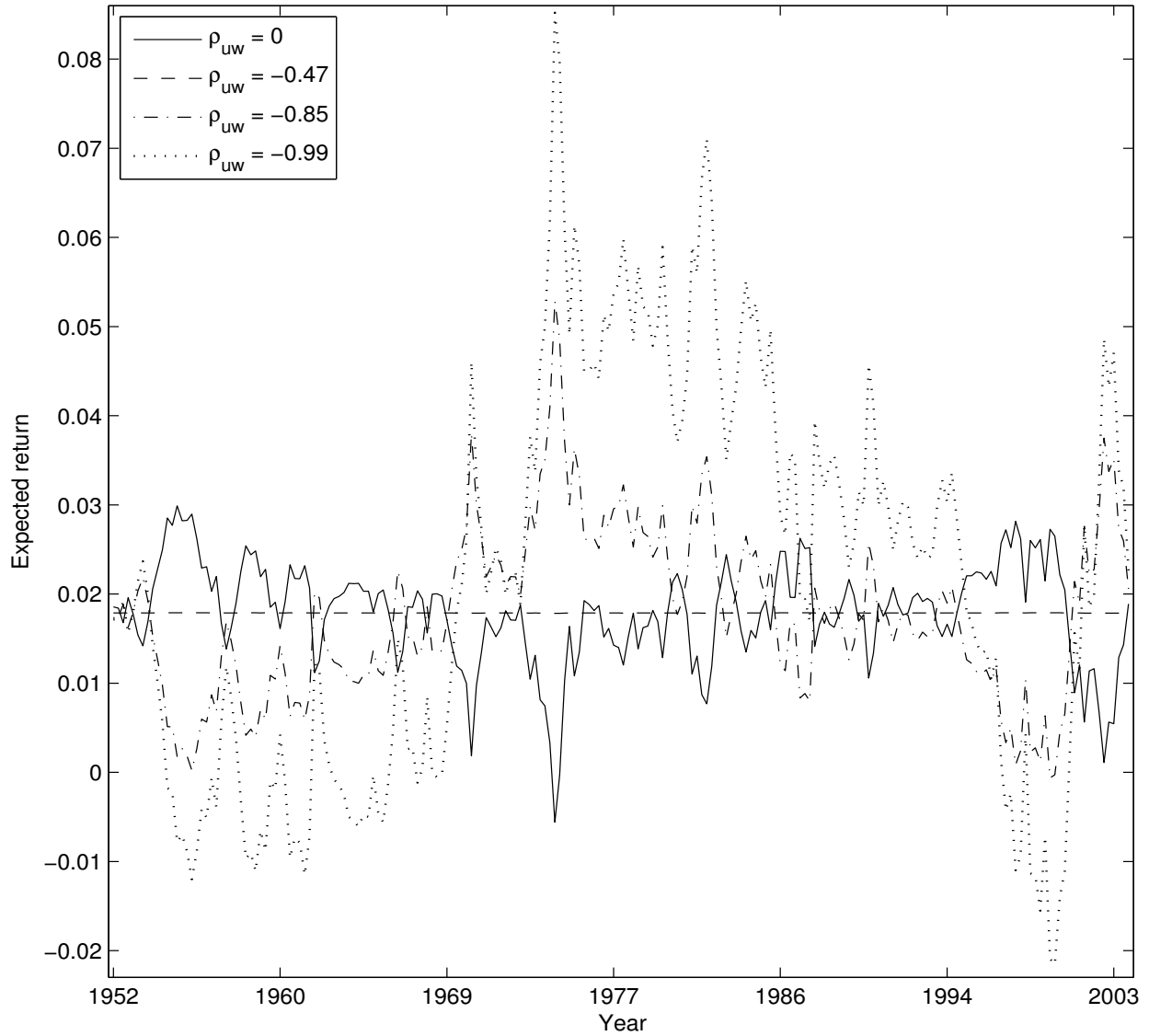


Figure 2. The equity premium $E(r_{t+1}|D_t)$ from the predictive system with no predictors. This figure plots the time series of the quarterly equity premium estimated for four different values of ρ_{uw} , the correlation between expected and unexpected returns. The mean reversion coefficient in the AR(1) process for the conditional expected return μ_t is set equal to $\beta = 0.9$. The predictive R^2 —the fraction of variation in r_{t+1} than can be explained by μ_t —is set equal to $R^2 = 0.05$. The parameters represent quarterly values.

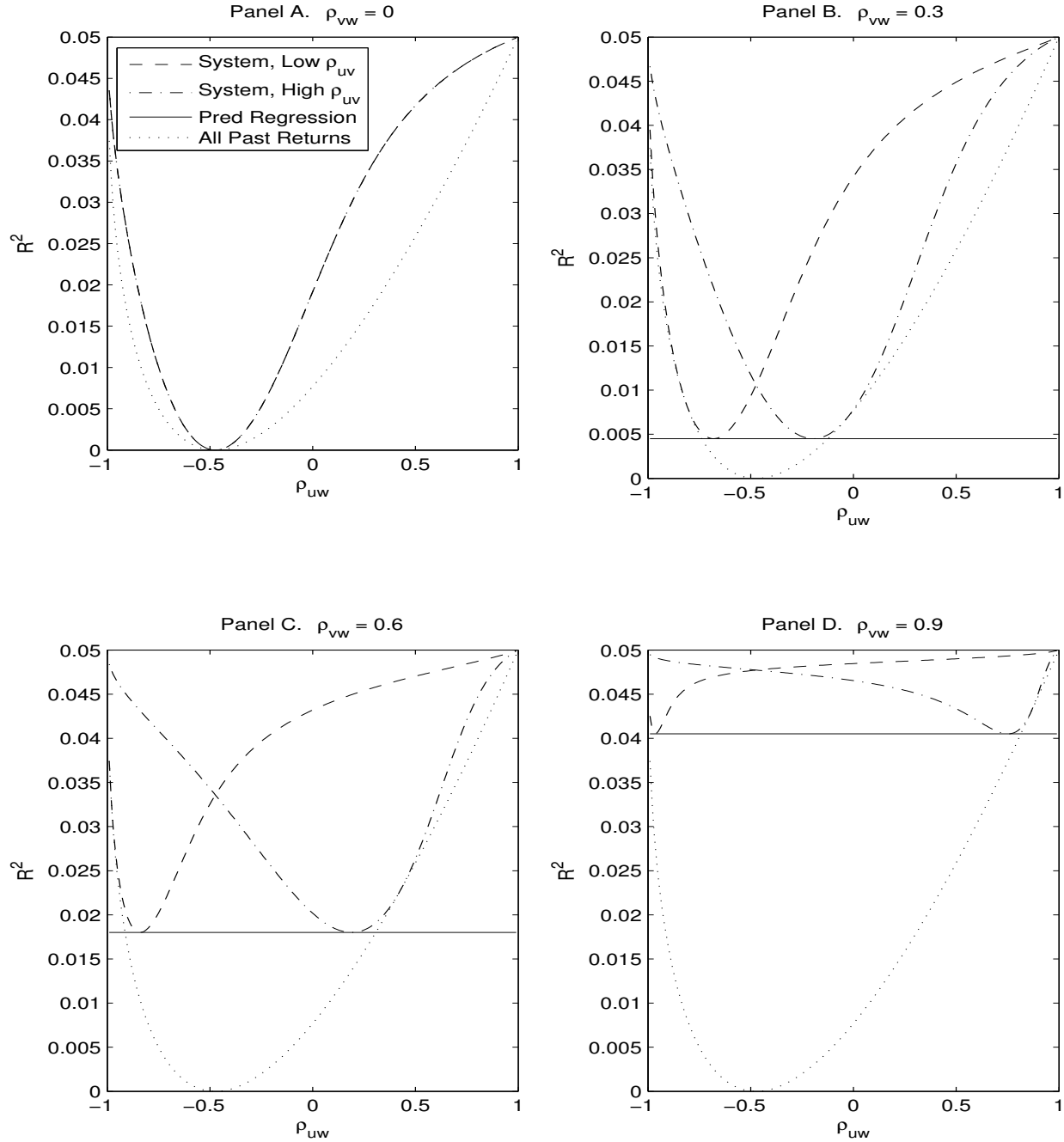


Figure 3. Predictive R^2 's. Each panel plots the R^2 's from three approaches to predicting stock returns r_{t+1} using information observable at time t . The approaches are: the predictive regression of r_{t+1} on a single predictor x_t (solid line), the ARMA(1,1) model that uses the full history of past returns but no predictor data (dotted line), and the predictive system, which uses the full history of returns and predictor realizations (dashed and dash-dot lines). The dashed (dash-dot) line corresponds to a “low” (“high”) value of ρ_{uv} , which represents the value obtained when the partial correlation between u_t and v_t given w_t equals $\rho_{uv|w} = -0.9$ (0.9). The conditional correlation between μ_t and x_t , ρ_{vw} , ranges from 0 in Panel A to 0.9 in Panel D. In all four panels, $\beta = A = 0.9$, and the true predictive R^2 (from the regression of r_{t+1} on μ_t) is 0.05. In Panel A, the solid line coincides with the x axis, and the dashed and dash-dot lines overlap.

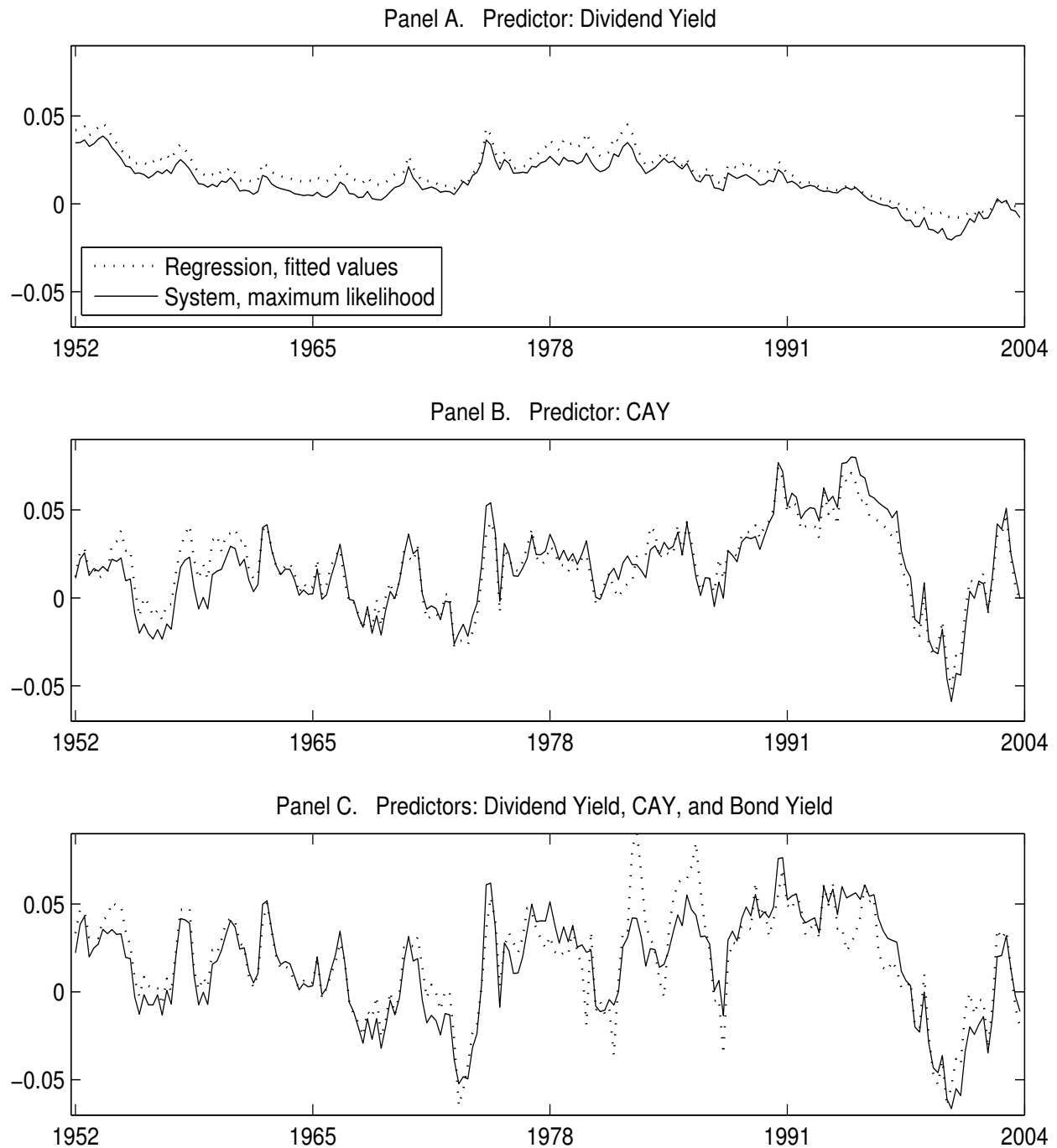


Figure 4. The equity premium: Regression vs. system with no prior information. This figure plots the time series of the quarterly equity premium estimated in two different environments. The *dotted* line plots the OLS fitted values from the predictive regression of r_{t+1} on the given predictor(s). The *solid* line plots the maximum likelihood estimates of $E(r_{t+1}|D_t)$ from the predictive system. In Panel A, the estimation uses one predictor, dividend yield. In Panel B, the single predictor is CAY. In Panel C, three predictors are used: dividend yield, CAY, and the bond yield. The sample period is 1952Q1–2003Q4.

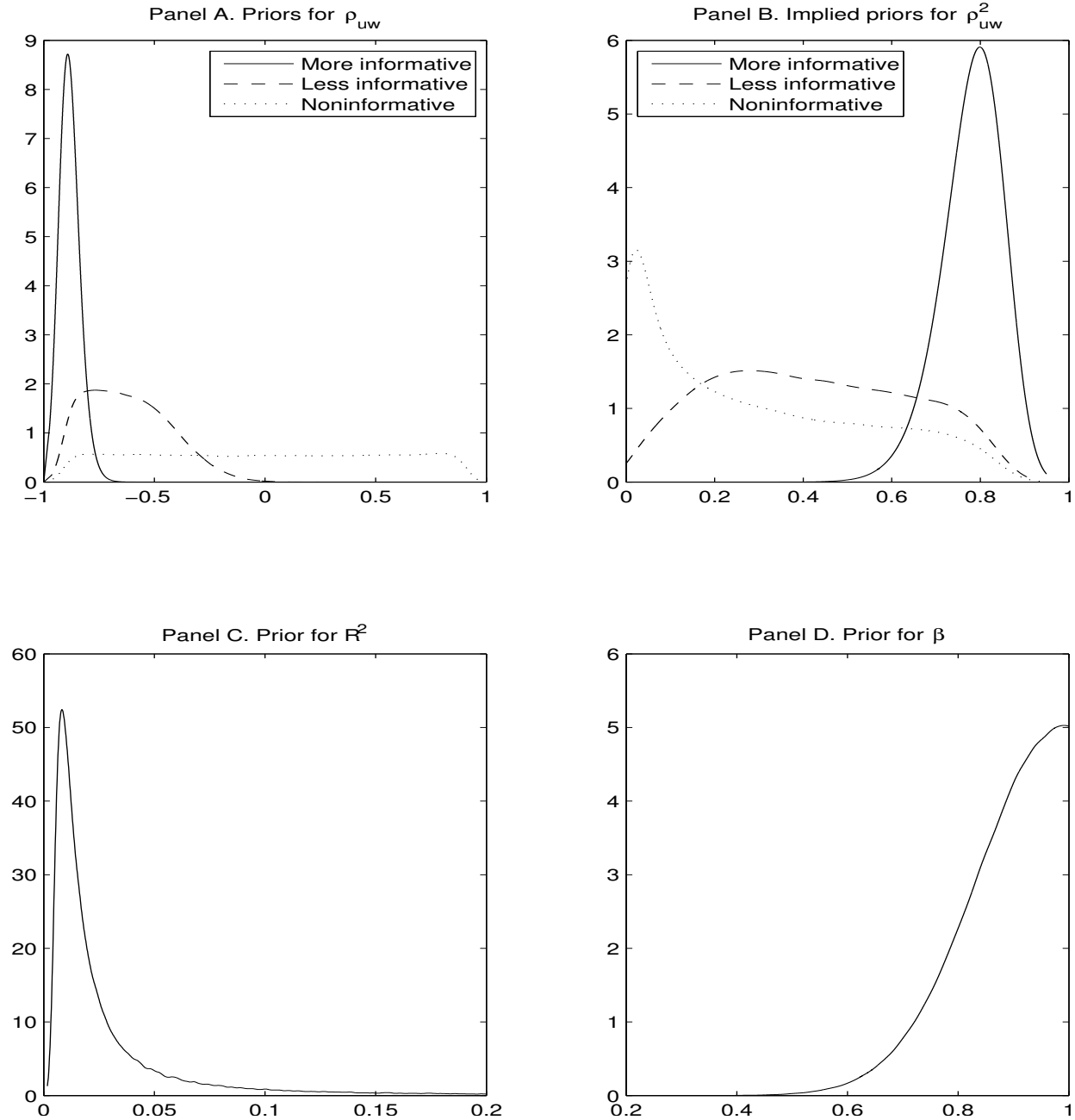


Figure 5. Prior distributions. Panel A plots three prior distributions for the correlation between expected and unexpected returns, ρ_{uw} . The noninformative prior (dotted line) is flat between -0.9 and 0.9, with tails fading away as ρ_{uw} approaches ± 1 . The less informative prior (dashed line) has 99.9% of its mass below zero ($\rho_{uw} < 0$). The more informative prior (solid line) has 99.9% of its mass below -0.71, so that $\rho_{uw}^2 > 0.5$ (i.e., unexpected changes in the discount rate explain over half of the variance of unexpected market returns). Panel B plots the corresponding implied priors on ρ_{uw}^2 . Panel C plots the prior on the predictive R^2 from the regression of returns r_{t+1} on expected returns μ_t . Panel D plots the prior on the slope coefficient β in the AR(1) process for μ_t . All parameters correspond to quarterly data.

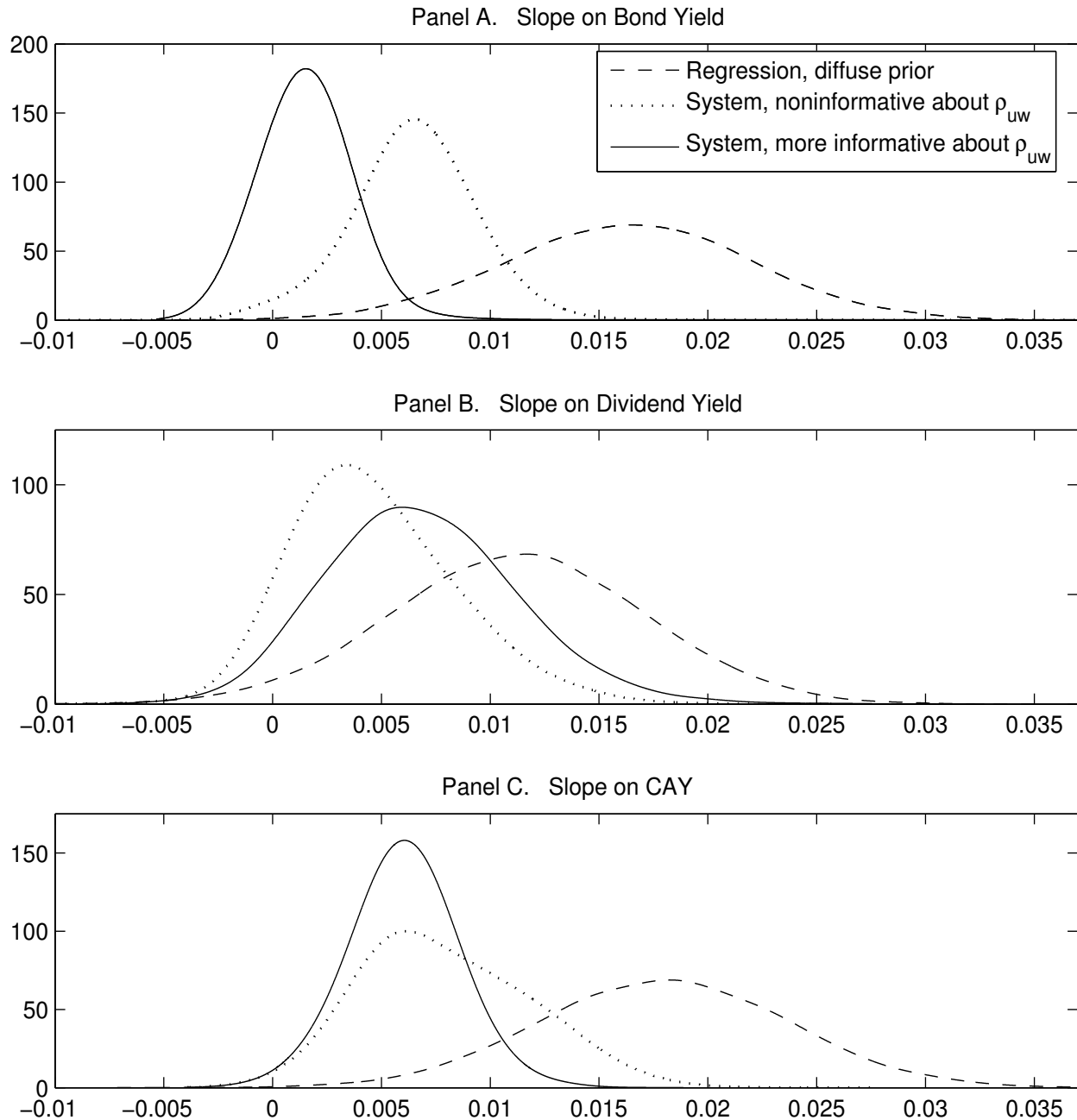


Figure 6. Posterior distributions of slope coefficients from predictive regressions. We estimate both the predictive system and the predictive regression with three predictors: the bond yield, dividend yield, and CAY. The *dashed* line plots the posteriors from the standard predictive regression of r_{t+1} on x_t under the diffuse prior. The *dotted* line plots the implied posteriors constructed from the results of the predictive system under the “noninformative” prior on ρ_{uw} . The *solid* line plots the implied posteriors from the predictive system under the “more informative” prior on ρ_{uw} ($\rho_{uw} < -0.71$). To facilitate comparisons across panels, all predictors are scaled to have unit variance. The sample period is 1952Q1–2003Q4.

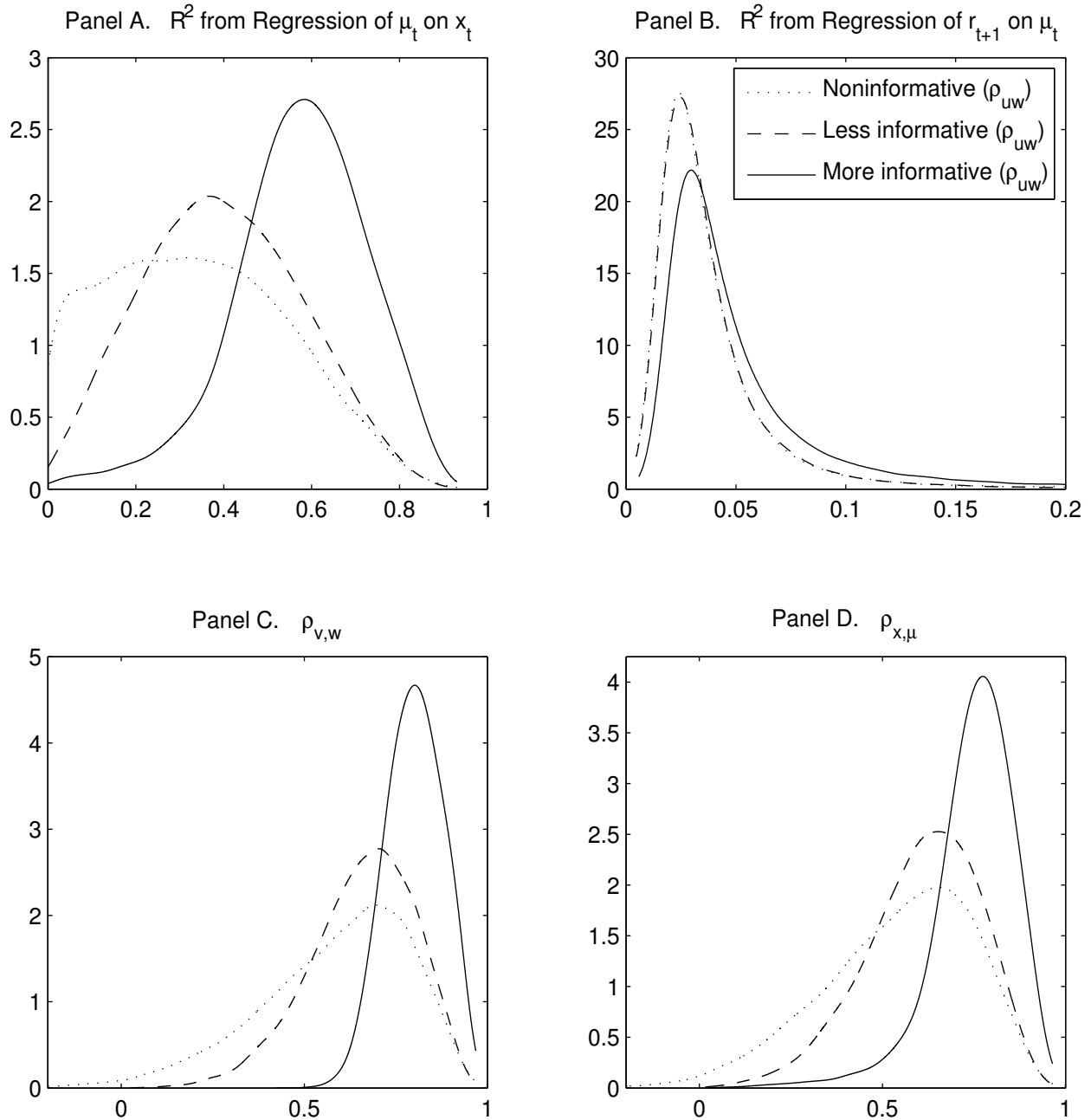


Figure 7. Posterior distributions for the relations between the dividend yield and expected return. Panel A plots the posterior of the fraction of variation in the expected return μ_t that can be explained by the predictor x_t , which is the dividend yield. Panel B plots the posterior of the predictive R^2 . Panel C plots the posterior of the *conditional* correlation ρ_{vw} between the dividend yield and μ_t . Panel D plots the posterior of the *unconditional* correlation $\rho_{x\mu}$ between the dividend yield and μ_t . The three lines in each panel represent three different prior distributions. The *solid* line represents the “more informative” prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the “less informative” prior on ρ_{uw} ($\rho_{uw} < 0$), and the *dotted* line is the “noninformative” prior on ρ_{uw} . The sample period is 1952Q1–2003Q4.

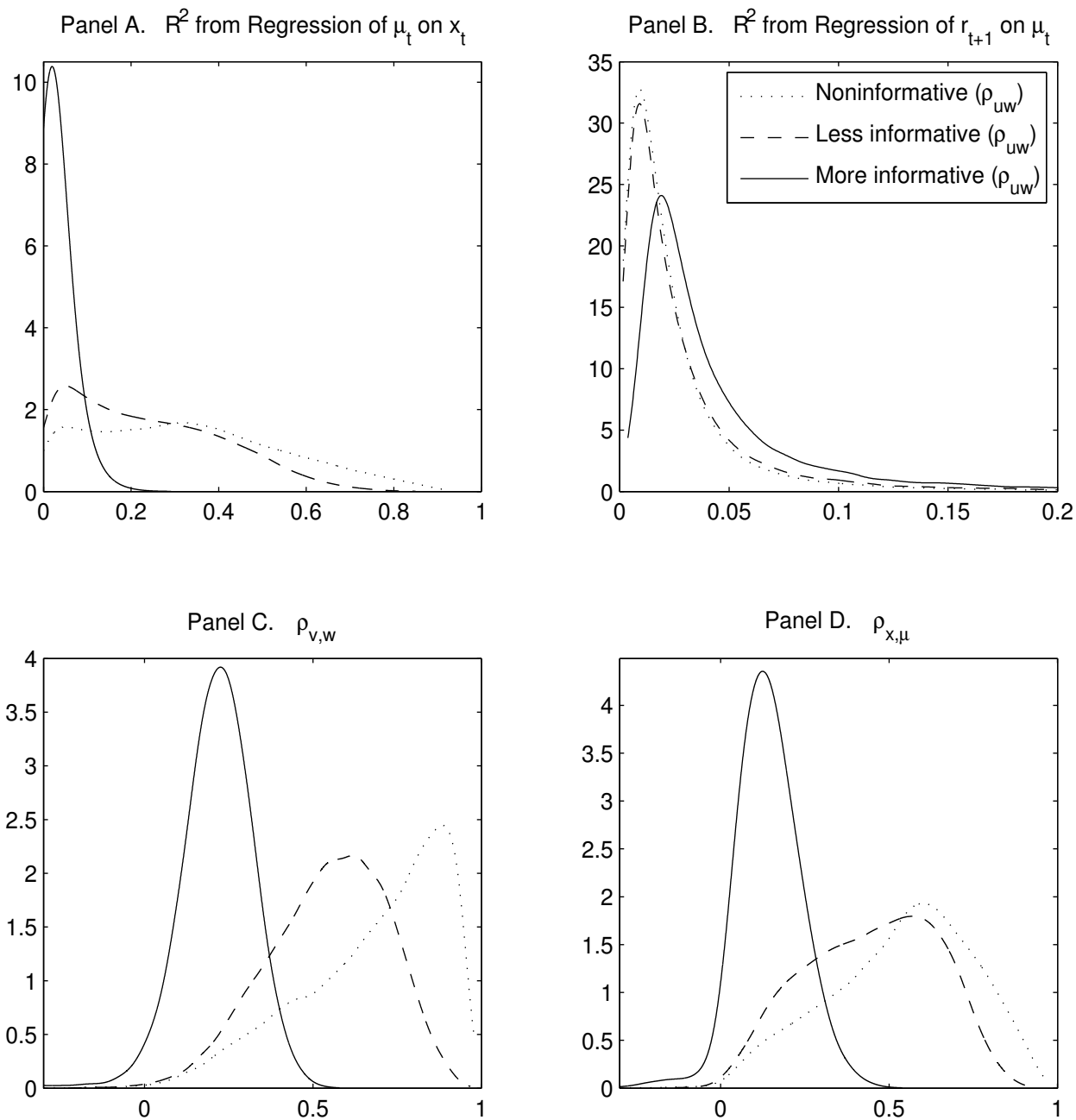


Figure 8. Posterior distributions for the relations between the bond yield and expected return. Panel A plots the posterior of the fraction of variation in the expected return μ_t that can be explained by the predictor x_t , which is the bond yield. Panel B plots the posterior of the predictive R^2 . Panel C plots the posterior of the *conditional* correlation ρ_{vw} between the bond yield and μ_t . Panel D plots the posterior of the *unconditional* correlation $\rho_{x\mu}$ between the bond yield and μ_t . The three lines in each panel represent three different prior distributions. The *solid* line represents the “more informative” prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the “less informative” prior on ρ_{uw} ($\rho_{uw} < 0$), and the *dotted* line is the “noninformative” prior on ρ_{uw} . The sample period is 1952Q1–2003Q4.

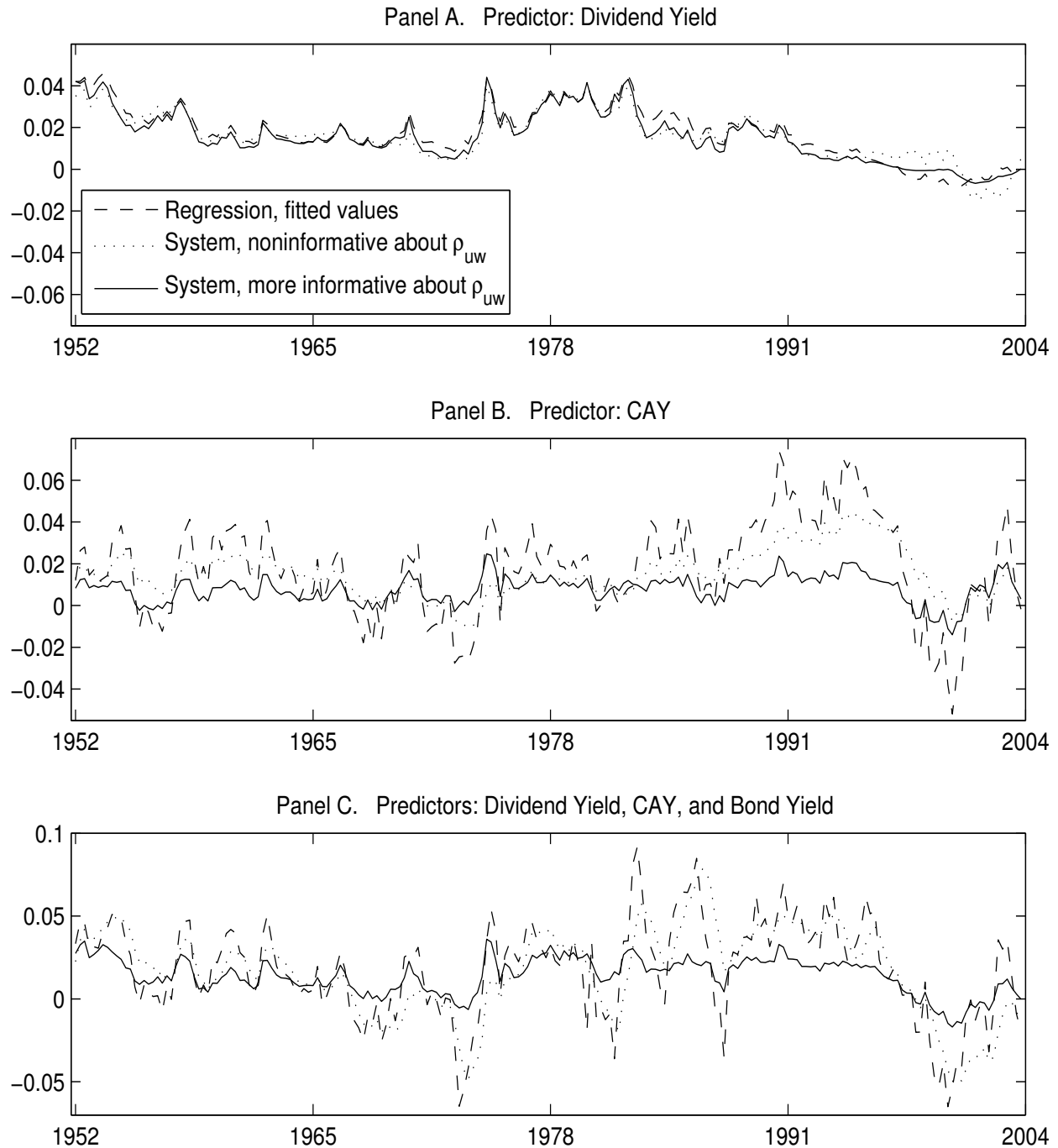


Figure 9. The equity premium: Regression vs. system with prior information. This figure plots the time series of the quarterly equity premium estimated in three different environments. The *dashed* line plots the OLS fitted values from the predictive regression of r_{t+1} on the given predictor(s). The *dotted* line plots the posterior means of $E(r_{t+1}|D_t)$ from the predictive system under the “noninformative” prior on ρ_{uw} . The *solid* line plots the posterior means of $E(r_{t+1}|D_t)$ from the predictive system under the “more informative” prior on ρ_{uw} ($\rho_{uw} < -0.71$). In Panel A, the estimation uses one predictor, dividend yield. In Panel B, the single predictor is CAY. In Panel C, three predictors are used: dividend yield, CAY, and bond yield. The sample period is 1952Q1–2003Q4.

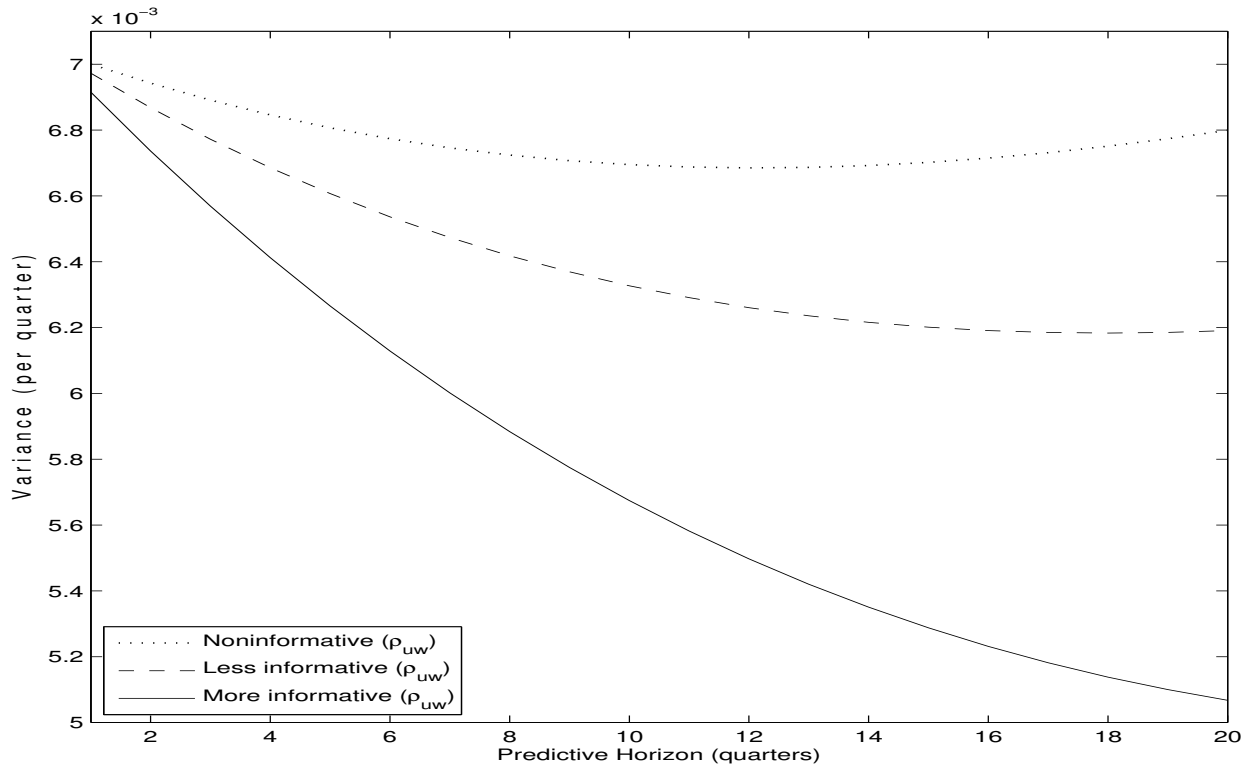


Figure 10. Variance of multiperiod returns. This figure plots the per-quarter variance of the predictive distribution of k -period stock returns. The investment horizon k ranges from 1 to 20 quarters. The variances are estimated in the predictive system using the dividend yield as an imperfect predictor. The *solid* line represents the “more informative” prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the “less informative” prior on ρ_{uw} ($\rho_{uw} < 0$), and the *dotted* line is the “noninformative” prior on ρ_{uw} . The sample period is 1952Q1–2003Q4.

Table I
Predictive Regressions

This table summarizes the results from predictive regressions $r_t = a + b'x_{t-1} + e_t$, where $x_t = \theta + Ax_{t-1} + v_t$. r_t denotes quarterly excess stock market return and x_{t-1} denotes the predictors (listed in the column headings) lagged by one quarter. The table reports the estimated slope coefficients \hat{b} , the correlation $\text{Corr}(e_t, b'v_t)$ between unexpected returns and shocks to expected returns, and the (unadjusted) R^2 from the predictive regression. The correlations and R^2 s are reported in percent (i.e., $\times 100$). The OLS t -statistics are given in parentheses “()”. The t -statistic of $\text{Corr}(e_t, b'v_t)$ is computed as the t -statistic of the slope from the regression of the sample residuals \hat{e}_t on $\hat{b}'\hat{v}_t$. The p -values associated with all t -statistics and R^2 s are computed by bootstrapping and reported in brackets “[]”.

Bond Yield	Dividend Yield	CAY	$\text{Corr}(e_t, b'v_t)$	R^2
<u>Panel A. 1952 Q1 – 2003 Q4</u>				
2.716 (3.024) [0.001]			21.735 (3.204) [0.001]	4.231 [0.002]
	1.153 (2.184) [0.057]		-91.887 (-33.506) [1.000]	2.252 [0.059]
		1.704 (4.035) [0.000]	-53.556 (-9.124) [1.000]	7.292 [0.000]
2.573 (2.902) [0.003]	1.028 (1.966) [0.058]	1.346 (3.139) [0.003]	-35.635 (-5.487) [1.000]	11.777 [0.000]
<u>Panel B. 1952 Q1 – 1977 Q4</u>				
6.385 (2.801) [0.004]			25.079 (2.629) [0.008]	7.080 [0.007]
	2.658 (2.785) [0.014]		-96.531 (-37.522) [1.000]	7.003 [0.015]
		3.028 (4.267) [0.000]	-47.663 (-5.503) [1.000]	15.024 [0.000]
3.489 (1.490) [0.090]	1.345 (1.349) [0.177]	2.129 (2.534) [0.012]	-53.153 (-6.369) [1.000]	17.975 [0.000]
<u>Panel C. 1978 Q1 – 2003 Q4</u>				
2.073 (2.053) [0.020]			22.624 (2.357) [0.011]	3.931 [0.047]
	0.784 (1.152) [0.409]		-88.194 (-18.989) [1.000]	1.273 [0.423]
		1.165 (2.104) [0.037]	-56.949 (-7.031) [1.000]	4.122 [0.045]
2.203 (2.197) [0.023]	0.755 (1.101) [0.313]	0.968 (1.734) [0.118]	-18.619 (-1.923) [0.967]	8.828 [0.053]

Table II
Explanatory Power of the Predictive Regression Relative to the Predictive System:
Empirical Results

Panel A shows the posterior means and standard deviations (the latter in parentheses) of the ratios of two R-squareds, $R^2(reg)/R^2(sys)$. A ratio smaller than one indicates that the predictive system estimates μ_t more precisely than the predictive regression does. The smaller the ratio, the larger the advantage of using the predictive system. $R^2(reg)$, computed as the R-squared from the regression of μ_t on the given predictors, summarizes the usefulness of the predictive regression in estimating μ_t . $R^2(sys)$, computed as $1 - \text{Var}(\mu_t|D_t)/\text{Var}(\mu_t)$ where D_t contains all historical market returns and predictor realizations, summarizes the usefulness of the predictive system in estimating μ_t . Panel B shows the posterior means and standard deviations of $1 - \text{MSE}(sys)/\text{MSE}(reg)$. $\text{MSE}(reg)$ is the mean squared error from the predictive regression of r_{t+1} on the given predictors. $\text{MSE}(sys)$ is the mean squared error from the predictive system. Positive values of one minus the MSE ratio indicate that the predictive system forecasts returns more precisely than the predictive regression does. The results are reported for four different prior distributions on ρ_{uw} , the correlation between expected and unexpected returns. Four sets of predictors are considered: dividend yield, bond yield, CAY, and all three predictors combined. The sample period is 1952Q1–2003Q4.

	Predictors			
	Dividend Yield	Bond Yield	CAY	All 3 Predictors
<u>Panel A. The R-squared Ratios, $R^2(reg)/R^2(sys)$.</u>				
Diffuse Prior	0.28 (0.17)	0.73 (0.23)	0.86 (0.16)	0.59 (0.30)
Noninformative Prior on ρ_{uw}	0.50 (0.27)	0.44 (0.25)	0.61 (0.27)	0.46 (0.22)
Less Informative Prior on ρ_{uw}	0.59 (0.22)	0.34 (0.20)	0.73 (0.23)	0.50 (0.22)
More Informative Prior on ρ_{uw}	0.81 (0.19)	0.08 (0.08)	0.64 (0.22)	0.70 (0.19)
<u>Panel B. The Mean Squared Error Ratios, $1 - \text{MSE}(sys)/\text{MSE}(reg)$.</u>				
Diffuse Prior	0.04 (0.07)	0.06 (0.15)	0.03 (0.09)	0.17 (0.22)
Noninformative Prior on ρ_{uw}	0.02 (0.04)	0.02 (0.05)	0.02 (0.04)	0.03 (0.04)
Less Informative Prior on ρ_{uw}	0.02 (0.04)	0.02 (0.06)	0.01 (0.04)	0.04 (0.05)
More Informative Prior on ρ_{uw}	0.02 (0.07)	0.03 (0.07)	0.02 (0.06)	0.02 (0.07)

Table III
Comparing Estimates of Expected Return.

This table compares the time series of the posterior means of $E(r_{t+1}|D_t)$ obtained in five different environments:

- (1) Predictive regression: OLS fitted values
- (2) Predictive system: Maximum likelihood estimates
- (3) Predictive system: Noninformative prior about ρ_{uw}
- (4) Predictive system: Less informative prior about ρ_{uw}
- (5) Predictive system: More informative prior about ρ_{uw}

The priors in (3)-(5) are informative about the persistence and volatility of μ_t . The correlations between the quarterly series of the posterior means of $E(r_{t+1}|D_t)$ are reported in italics below the main diagonal of each left-panel 5×5 matrix. Above the main diagonal of the same matrix are the mean absolute differences between the posterior means of $E(r_{t+1}|D_t)$ in percent per quarter. Each right-panel 5×5 matrix reports the average utility losses, in percent per quarter, of a mean-variance investor who is forced to hold a suboptimal portfolio of the stock market and a risk-free T-bill: a portfolio that is optimal under the beliefs in the given row when the true beliefs are in the given column. (For example, the (2,5) cell of the 5×5 matrix reports the certainty equivalent loss of an investor who has the more informative prior but is forced to hold the portfolio that is optimal under the maximum likelihood estimates.) The risk aversion is chosen such that there is no borrowing or lending given the sample mean and variance of market returns. The sample period is 1952Q1-2003Q4.

<i>Correlation (%) \ Mean Abs Diff (%)</i>					Average Utility Loss (%)				
(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
<u>Panel A. Predictor: Dividend Yield</u>									
(1)	0.57	0.44	0.41	0.29	0	0.11	0.09	0.07	0.03
(2)	<i>97.49</i>	0.65	0.62	0.49	0.11	0	0.20	0.18	0.10
(3)	<i>90.47</i>	<i>81.11</i>	0.06	0.27	0.09	0.19	0	0.00	0.03
(4)	<i>92.42</i>	<i>83.49</i>	<i>99.78</i>	0.22	0.07	0.17	0.00	0	0.02
(5)	<i>97.67</i>	<i>91.32</i>	<i>95.81</i>	<i>97.41</i>	0.03	0.10	0.03	0.02	0
<u>Panel B. Predictor: CAY</u>									
(1)	0.66	1.13	1.22	1.60	0	0.19	0.57	0.65	1.13
(2)	<i>94.82</i>	1.36	1.39	1.65	0.19	0	0.88	0.95	1.38
(3)	<i>86.28</i>	<i>82.02</i>	0.37	0.96	0.57	0.85	0	0.06	0.39
(4)	<i>96.88</i>	<i>93.26</i>	<i>95.43</i>	0.63	0.64	0.92	0.06	0	0.16
(5)	<i>89.71</i>	<i>92.03</i>	<i>59.38</i>	<i>80.11</i>	1.09	1.32	0.38	0.16	0
<u>Panel C. Predictors: Dividend Yield, CAY, Bond Yield</u>									
(1)	0.92	1.33	1.27	1.60	0	0.40	0.82	0.74	1.19
(2)	<i>91.06</i>	1.27	1.22	1.62	0.42	0	0.76	0.68	1.16
(3)	<i>80.38</i>	<i>82.36</i>	0.14	1.51	0.80	0.72	0	0.01	0.94
(4)	<i>82.30</i>	<i>84.11</i>	<i>99.79</i>	1.46	0.72	0.64	0.01	0	0.84
(5)	<i>83.42</i>	<i>89.93</i>	<i>80.75</i>	<i>84.00</i>	1.19	1.13	0.96	0.87	0

Table IV
Variance Decomposition of Expected Return.

This table reports the posterior means and standard deviations (the latter in parentheses) of the R^2 s from the regressions of the market's expected excess return μ_t on its selected components. The first column of each panel, labeled x_t , shows the fraction of variance of μ_t that can be explained by the set of predictors listed in the panel heading. Four sets of predictors are considered: the dividend yield, bond yield, CAY, and the combination of all three of these predictors. The second column of each panel, labeled $x_t, \{u_s\}_{s \leq t}$, shows the fraction of variance of μ_t that can be explained jointly by the predictors and by the innovations to stock market returns $u_t, u_{t-1}, u_{t-2}, \dots$. The third column, labeled $x_t, \{u_s, v_s\}_{s \leq t}$, shows the fraction of variance of μ_t that can be explained jointly by the predictors, by the innovations to stock market returns $u_t, u_{t-1}, u_{t-2}, \dots$, and by the innovations to the predictors $v_t, v_{t-1}, v_{t-2}, \dots$. For each set of predictors, the predictive system is estimated under three different priors, which are described in the row labels. The sample period is 1952Q1-2003Q4.

	Components of Expected Return			Components of Expected Return		
	x_t	$x_t, \{u_s\}_{s \leq t}$	$x_t, \{u_s, v_s\}_{s \leq t}$	x_t	$x_t, \{u_s\}_{s \leq t}$	$x_t, \{u_s, v_s\}_{s \leq t}$
	<u>Panel A. Dividend Yield</u>			<u>Panel B. Bond Yield</u>		
Noninformative	0.34	0.43	0.48	0.33	0.64	0.83
Prior on ρ_{uw}	(0.20)	(0.21)	(0.21)	(0.21)	(0.21)	(0.13)
Less Informative	0.40	0.49	0.53	0.24	0.73	0.86
Prior on ρ_{uw}	(0.18)	(0.18)	(0.18)	(0.17)	(0.16)	(0.11)
More Informative	0.57	0.80	0.81	0.03	0.86	0.95
Prior on ρ_{uw}	(0.15)	(0.06)	(0.06)	(0.03)	(0.05)	(0.04)
	<u>Panel C. CAY</u>			<u>Panel D. All Three Predictors</u>		
Noninformative	0.50	0.59	0.81	0.42	0.49	0.90
Prior on ρ_{uw}	(0.22)	(0.22)	(0.12)	(0.20)	(0.22)	(0.07)
Less Informative	0.60	0.70	0.83	0.46	0.55	0.90
Prior on ρ_{uw}	(0.20)	(0.17)	(0.10)	(0.20)	(0.22)	(0.07)
More Informative	0.53	0.87	0.92	0.63	0.85	0.94
Prior on ρ_{uw}	(0.18)	(0.07)	(0.05)	(0.17)	(0.12)	(0.04)

References

- Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.
- Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745–787.
- Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.
- Avramov, Doron, 2004, Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.
- Avramov, Doron, and Russ Wermers, 2006, Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339–377.
- Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45–85.
- Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.
- van Binsbergen, Jules H., and Ralph S. J. Koijen, 2007, Predictive Regressions: A Present-Value Approach, Working paper, Duke University.
- Brandt, Michael W., and Qiang Kang, 2004, On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach, *Journal of Financial Economics* 72, 217–257.
- Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.
- Campbell, John Y., 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157–179.
- Campbell, John Y., and John Ammer, 1993, What moves the stock and bond markets? A variance decomposition for long-term asset returns, *Journal of Finance* 48, 3–37.
- Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.
- Campbell, John Y., and Samuel B. Thompson, 2008, Predicting the equity premium out of sample: Can anything beat the historical average?, *Review of Financial Studies*, forthcoming.
- Carter, Chris K., and Robert Kohn, 1994, On Gibbs sampling for state space models, *Biometrika* 81, 541–553.
- Casella, G., and E.I. George, 1992, Explaining the Gibbs sampler, *The American Statistician* 46, 167–174.

- Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.
- Clark, Todd E., and Kenneth D. West, 2006, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* 135, 155–186.
- Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.
- Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies*, forthcoming.
- Conrad, Jennifer, and Gautam Kaul, 1988, Time-variation in expected returns, *Journal of Business* 61, 409–425.
- Cremers, K.J. Martijn, 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.
- Dangl, Thomas, and Michael Halling, 2006, Equity return prediction: Are coefficients time-varying?, Working paper, University of Vienna.
- Duffee, Gregory R., 2006, Are variations in term premia related to the macroeconomy?, Working paper, University of California, Berkeley.
- Elliott, Graham, and James H. Stock, 1994, Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory* 10, 672–700.
- Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–26.
- Fama, Eugene F., and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.
- Ferson, Wayne E., and Campbell R. Harvey, 1991, “The Variation of Economic Risk Premiums,” *Journal of Political Economy* 99, 385–415.
- Ferson, Wayne E., and Campbell R. Harvey, 1993, “The Risk and Predictability of International Equity Returns,” *Review of Financial Studies* 6, 527–566.
- Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics?, *Journal of Finance* 58, 1393–1413.
- Frühwirth-Schnatter, Sylvia, 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.
- Goyal, Amit and Ivo Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.
- Goyal, Amit and Ivo Welch, 2006, A Comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies*, forthcoming.
- Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press, Princeton, NJ).
- Harvey, Andrew C., 1989, *Forecasting, structural time series models and the Kalman filter* (Cambridge University Press, Cambridge, UK).

- Hjalmarsson, Erik, 2006, Should we expect significant out-of-sample results when predicting stock returns?, Working paper, Board of Governors of the Federal Reserve System.
- Jansson, Michael, and Marcelo J. Moreira, 2006, Optimal inference in regressions with nearly integrated regressors, *Econometrica* 74, 681–714.
- Johannes, Michael, Nicholas Polson, and Jon Stroud, 2002, Sequential optimal portfolio performance: Market and volatility timing, Working paper, Columbia University.
- Jones, Christopher S., and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.
- Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.
- Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.
- Kothari, S.P., Jonathan Lewellen, and Jerold B. Warner, 2006, Stock returns, aggregate earnings surprises, and behavioral finance, *Journal of Financial Economics* 79, 537–568.
- Lamont, Owen, 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.
- Lamoureux, Christopher G., and Guofu Zhou, 1996, Temporary components of stock returns: What do the data tell us?, *Review of Financial Studies* 9, 1033–1059.
- 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 dividend growth, *Journal of Financial Economics* 76, 583–626.
- Lettau, Martin, and Stijn Van Nieuwerburgh, 2007, Reconciling the Return Predictability Evidence, *Review of Financial Studies*, forthcoming.
- Leroy, Stephen F. and Richard D. Porter, 1981, The present-value relation: Tests based on implied variance bounds, *Econometrica* 49, 555–574
- Lewellen, Jonathan, 1999, The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* 54, 5–43.
- Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.
- Mankiw, N. Gregory, and Matthew D. Shapiro, 1986, Do we reject too often?: Small sample properties of tests of rational expectations models, *Economics Letters* 20, 139–145.
- Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.
- Nelson, Charles R., Myung J. Kim, 1993, Predictable stock returns: the role of small sample bias, *Journal of Finance* 48, 641–661.
- Pástor, Ľuboš, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.

- Pástor, Ľuboš, Meenakshi Sinha, and Bhaskaran Swaminathan, 2008, Estimating the intertemporal risk-return tradeoff using the implied cost of capital, *Journal of Finance*, forthcoming.
- Pástor, Ľuboš, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67–121.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2001, The equity premium and structural breaks, *Journal of Finance* 56, 1207–1239.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2002a, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2002b, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2007, Out-of-sample equity premium prediction: Consistently beating the historical average, Working paper, Washington University.
- Rozeff, Michael S., 1984. Dividend yields are equity risk premiums, *Journal of Portfolio Management* 68–75.
- Rytchkov, Oleg, 2007, Filtering out expected dividends and expected returns, Working paper, MIT.
- Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.
- Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.
- Stambaugh, Robert F., 1986, Bias in regressions with lagged stochastic regressors, Working paper, University of Chicago.
- Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics*, 45, 285–331.
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Wachter, Jessica, and Missaka Warusawitharana, 2006, Predictable returns and asset allocation: Should a skeptical investor time the market?, Working paper, University of Pennsylvania.
- West, Mike, and Jeff Harrison, 1997, *Bayesian Forecasting and Dynamic Models* (Springer-Verlag, New York, NY).
- Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley and Sons, New York, NY).