Predictive Systems: Living with Imperfect Predictors

by*

Ľuboš Pástor

and

Robert F. Stambaugh

First draft: September 15, 2006
This version: October 24, 2006

Abstract

The standard regression approach to investigating return predictability seems too restrictive in one way but too lax in another. A predictive regression assumes that expected returns are captured exactly by a set of given predictors but does not exploit the likely economic property that innovations in expected returns are negatively correlated with unexpected returns. We develop an alternative framework—a predictive system—that accommodates imperfect predictors and beliefs about that negative correlation. In this framework, the predictive ability of imperfect predictors is supplemented by information in lagged returns as well as lags of the predictors. Compared to predictive regressions, predictive systems deliver different and substantially more precise estimates of expected returns as well as different assessments of a given predictor's usefulness.

---

*Graduate School of Business, University of Chicago, NBER, and CEPR (Pástor) and the Wharton School, University of Pennsylvania, and NBER (Stambaugh). Helpful comments were gratefully received from seminar participants at the Hong Kong University of Science and Technology, National University of Singapore, Singapore Management University, University of Iowa, University of Pennsylvania, University of Texas at Austin, and University of Texas at Dallas.

1. Introduction

Many studies in finance analyze comovement between the expected return on stocks 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

$$r_{t+1} = \mu_t + u_{t+1}, \tag{1}$$

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. Typically, such investigations rely on a "predictive regression" in which r_{t+1} is regressed on x_t and the expected return is modeled as

$$\mu_t = a + b'x_t, \tag{2}$$

where a and b denote the regression's intercept and slope coefficients.[1] However, equation (2) seems too restrictive in assuming that the true conditional expected return is explained perfectly by the observed predictors. It seems more likely that the predictors are imperfect, in that they might be 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,

$$\mu_{t+1} = \alpha + \beta\mu_t + w_{t+1}, \tag{3}$$

then it seems likely that the correlation between the unexpected return and the expected-return innovation is negative, or that $\rho(u_{t+1}, w_{t+1}) < 0$. That is, an unanticipated increase in expected future returns (or discount rates) should be accompanied by an unexpected negative return (or price drop). The likely negative correlation between expected and unexpected returns, which is not exploited in estimating the predictive regression, emerges as an important consideration in estimating expected returns when predictors are imperfect. For example, the inference about a given predictor's relation to μ_t should depend on whether innovations in that predictor exhibit a negative correlation with unexpected returns.

Let $\rho_{uw} \equiv \rho(u_{t+1}, w_{t+1})$ denote the correlation between expected and unexpected returns. Although ρ_{uw} need not be negative, we suggest it is reasonable to believe that it is. To understand

[1] Of 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, Elliot, and Stock (1995), Lewellen (2004), Campbell and Yogo (2006), and Jansson and Moreira (2006).

this viewpoint, first note that the unexpected return can be represented approximately as

$$u_{t+1} = \eta_{C,t+1} - \eta_{E,t+1}, \tag{4}$$

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 (Campbell, 1991). If expected return follows the process in (3) with $0 < \beta < 1$, then $\eta_{E,t+1} = g w_{t+1}$, where $g > 0$ is a constant, so

$$\rho_{uw} = \rho(u_{t+1}, \eta_{E,t+1}). \tag{5}$$

It follows directly from equations (4) and (5) that $\rho_{uw} < 0$ if and only if

$$\rho(\eta_{C,t+1}, \eta_{E,t+1}) < \frac{\sigma(\eta_{E,t+1})}{\sigma(\eta_{C,t+1})}, \tag{6}$$

where the σ's denote standard deviations. It is easy to see from equation (4) that a violation of the condition in (6) would require that $\sigma(u_{t+1}) < \sigma(\eta_{C,t+1})$, or that returns be less volatile than when the expected return is constant. Since actual return volatility instead appears to be higher than what a constant expected return can accommodate, as observed by Shiller (1981) and LeRoy and Porter (1981), it seems likely that the condition in (6) is satisfied and therefore that $\rho_{uw} < 0$.

A prior belief that $\rho_{uw} < 0$ is also supported by empirical estimates based on market-level data preceding our sample period, which begins in 1952. Using a vector-autoregressive approach, Campbell (1991) obtains negative estimates of $\rho(\eta_{C,t+1}, \eta_{E,t+1})$ for the 1927–1951 period in three different specifications (see his Table 2), making the condition in (6) hold trivially.[2] Based on quarterly data (which we use in our empirical work), Campbell's Table 2 also reports $\sigma(\eta_{E,t+1}) > \sigma(\eta_{C,t+1})$ in 1927–1951, which again makes the condition in (6) hold trivially independent of $\rho(\eta_{C,t+1}, \eta_{E,t+1})$. The same table also reports the variance of $\eta_{E,t+1}$ and its covariance with $\eta_{C,t+1}$ (as fractions of $\sigma^2(u_{t+1})$), from which the implied value of $\rho(u_{t+1}, \eta_{E,t+1})$ can be computed. For the 1927–1951 period, Campbell's estimates imply values of $\rho(u_{t+1}, \eta_{E,t+1})$ ranging from -0.67 to -0.87 across three different specifications.[3] These implied values can be interpreted as estimates of ρ_{uw}, given equation (5).

This study develops and applies 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 they are not assumed to deliver perfectly

[2] On the other hand, Menzly, Santos, and Veronesi (2004), Lettau and Ludvigson (2005), and Kothari, Lewellen, and Warner (2005) find a positive correlation between shocks to expected return and dividend growth.

[3] In the second half of his sample, 1952–1988, the estimates of $\rho(u_{t+1}, \eta_{E,t+1})$ range from -0.92 to -0.94, and in the full sample, 1927–1988, they range from -0.71 to -0.86. These values are computed from Campbell's Table 2 as $\rho(u_{t+1}, \eta_{E,t+1}) = \left[\sigma(\eta_C, \eta_E)/\sigma_u^2 - \sigma_{\eta_E}^2/\sigma_u^2\right] / \left[\sigma_{\eta_E}/\sigma_u\right]$. Note from equation (5) that ρ_{uw}^2 can be interpreted as the fraction of the variance in unexpected returns that is explained by news about expected returns.

the true expected return μ_t as in equation (2). The predictive system also allows us to explore roles for a variety of prior beliefs about the behavior of expected returns. Chief among these is the belief that unexpected returns are likely to be negatively correlated with expected returns ($\rho_{uw} < 0$), but we also include beliefs that the degree of true predictability in equation (1) is relatively modest and that the expected return μ_t is fairly persistent.

We find that, compared to predictive regressions, predictive systems deliver different and substantially more precise estimates of expected return. When predictors are imperfect, their predictive ability can generally be supplemented by information in lagged returns as well as lags of the predictors, and the predictive system delivers that information via a parsimonious model. Prior beliefs play a key role in determining how that additional sample information is used as well as its relative importance in explaining variation in expected return.

The additional information in lagged returns is used in an interesting way. Suppose one believes that the conditional expected return is fairly persistent and then observes that recent returns have been unusually low. On one hand, since a low mean is more likely to generate low realized returns, one might think that the expected return has declined. On the other hand, since increases in expected future returns tend to accompany price drops and thus low returns, one might think that the expected return has increased. When the correlation between expected and unexpected returns 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. 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, and the weights depend on the fraction of the variance in unexpected returns that is explained by changes in expected returns. For example, if this investor believes that fraction is about 72% (the values of ρ_{uw}^2 implied by Campbell (1991) are in that neighborhood), then returns realized during the most recent decade receive negative weights in the current expected return, while the returns from the previous four decades receive positive weights. In other words, the investor in this example would view the 1990s' bull market as a bearish indicator.

A striking example of the importance of prior beliefs about the correlation between expected and unexpected returns is provided by regressing post-war U.S. stock market returns on what we call 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 innovations are positively correlated with the residuals in that regression. The latter correlation, opposite in sign to what one would anticipate for the correlation between expected and unexpected returns, 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 the correlation between expected and unexpected returns. With noninformative beliefs about that correlation, 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. Prior beliefs also affect the predictive system's advantage in explanatory power over the predictive regression. 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 in terms of its posterior variance, but with the more informative beliefs, the system's estimate is 12.5 times more precise. Moreover, under the more informative beliefs, 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%.

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 the correlation between expected and unexpected returns 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 the correlation between expected and unexpected returns).

Ferson, 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). Our paper provides tools that can be helpful in avoiding the spurious regression problem. A spurious predictor is unlikely to produce a substantially negative correlation between expected and unexpected returns. Therefore, under our 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 remainder of the paper proceeds as follows. In section 2, we first implement the traditional predictive regression approach to modeling expected stock returns and examine the estimated correlations between expected and unexpected returns. We then present our alternative approach, the predictive system, and discuss its implications for expected stock returns. Section 3 presents our empirical results. We first outline our Bayesian approach to implementing the predictive system and discuss the specifications of prior beliefs that we entertain. We then compare the explanatory powers of the predictive system and predictive regression, assess the degree to which various predictors are correlated with expected return, and analyze the behavior of estimated expected returns. Finally, we decompose the variance of expected return into components due to the current predictor values, lagged unexpected returns, and lagged predictor innovations. Section 4 briefly reviews the paper's conclusions and suggests directions for future research. Many technical aspects of our analysis are presented in the Appendix.

2. Modeling Expected Returns

2.1. Traditional Approach: Predictive Regression

We begin by estimating predictive regressions on quarterly data 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 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 and Ammer, 1993).

Table I reports the estimated slope coefficients $\hat{b}$ and the R^2's from the predictive regressions, as well as the estimated correlations between unexpected returns and the innovations in expected returns. To obtain the innovations in expected return, we make the common assumption that the vector of predictors follows a first-order autoregressive process,

$$x_t = \theta + Ax_{t-1} + v_t, \tag{7}$$

where v_t is distributed independently through time. The correlation between expected and unexpected returns is then simply $\text{Corr}(b'v_t, e_t)$, where the predictive regression disturbance is $e_t = r_t - a - b'x_{t-1}$. Table I also reports the OLS t-statistics and the bootstrapped p-values associated with these t-statistics as well as with the R^2s. Panel A reports the full-sample results covering 1952 Q1 – 2003 Q4. Panels B and C report sub-sample results.[4]

The results suggest that all three predictors have some forecasting ability. The dividend yield produces the weakest evidence (highest p-values, lowest R^2s) 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.

[4]Since 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.

In contrast, both the bond yield and CAY are highly significant predictors. When used alone, 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^2s, it is also informative to examine the correlations between expected and unexpected returns, shown in the fourth column of Table I. When the single predictor is either the dividend yield or CAY, this correlation is negative and highly significant (-91.9% for the dividend yield and -53.6% for CAY in the full sample), which seems plausible. 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. Basic economic principles suggest that this correlation is likely to be negative, so predictive models in which this correlation is positive seem less plausible.[5] 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. 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.

2.2. Predictive System

The predictive regression approach assumes that the expected return is perfectly correlated with a linear combination of the predictors in x_t. We generalize this approach to recognize that no combination of those predictors need capture perfectly the true unobserved expected return, μ_t. Our alternative framework, which we call a *predictive system*, combines the three equations in (1),

[5] Strictly speaking, the arguments based on equations (4) and (6) apply when r_{t+1} denotes the total (real) 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 (4) has an additional term representing news about future interest rates, and he estimates the variance of that term to be typically an order of magnitude smaller than the variances of $\eta_{C,t+1}$ and $\eta_{E,t+1}$.

(3), and (9):

$$r_{t+1} = \mu_t + u_{t+1} \tag{8}$$

$$x_{t+1} = \theta + Ax_t + v_{t+1} \tag{9}$$

$$\mu_{t+1} = \alpha + \beta\mu_t + w_{t+1}, \tag{10}$$

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

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}\right). \tag{11}$$

We assume throughout that $|\beta| < 1$ and that the eigenvalues of A lie inside the unit circle. The above predictive system nests the predictive regression model discussed earlier when μ_t is perfectly correlated with $b'x_t$ for some b, requiring $w_t = b'v_t$ and $A = \beta I$. In general, though, the predictors in x_t are correlated with μ_t but do not capture it perfectly. An extreme version of imperfect predictors occurs when there are no predictors, so that equation (9) is absent from the system and the data include only returns. In fact, we will use that simplified setting initially to illustrate some properties of the predictive system before moving on to our principal setting in which the predictors are the same as those used above in estimating the predictive regressions.

The predictive system is a version of a state-space model. In the standard form for such models, the disturbances of the "state equation" that governs the dynamics of the state variables are uncorrelated with those of the "observation equation" that relates the observed quantities to the state variables. In our case, σ_{uw} and the elements of σ_{vw} are generally non-zero, so representing the predictive system in the standard state-space form would require the state vector to include all variables (r_t, x_t, and μ_t) and thus the observation equation to have disturbances that are identically zero.[6] We do not proceed in that fashion but nevertheless rely in part on methods and insights developed in the state-space literature.[7]

[6]Specifically, to put equations (8) through (10) in standard form as in, for example, Hamilton (1994, p. 372-3), let $[r_t\ x_t'\ \mu_t]'$ minus its unconditional mean be the state vector ξ_t, so that the state equation becomes

$$\xi_{t+1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & A & 0 \\ 0 & 0 & \beta \end{bmatrix}\xi_t + \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ w_{t+1} \end{bmatrix}, \tag{12}$$

and the observation equation becomes

$$\begin{bmatrix} r_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mathrm{E}(r) \\ \mathrm{E}(x) \end{bmatrix}[1] + \begin{bmatrix} 1 & 0 & 0 \\ 0 & I & 0 \end{bmatrix}\xi_t + \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{13}$$

[7]Studies that analyze the predictability of stock and bond returns using state space models include Ang and Piazzesi (2003), Dangl and Halling (2006), Duffee (2006), and others.

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

$$\epsilon_t = r_t - \mathrm{E}(r_t|D_{t-1}), \tag{14}$$

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

$$\mathrm{E}(r_{t+1}|D_t) = \mathrm{E}(r) + \sum_{s=0}^{\infty} \left(\lambda_s \epsilon_{t-s} + \phi_s' v_{t-s}\right), \tag{15}$$

where, in steady state,

$$\lambda_s = m\beta^s, \tag{16}$$

$$\phi_s = n\beta^s, \tag{17}$$

and m and n are functions of the parameters in equations (8) through (11).[8] When the predictors approach perfection, where $w_t = b'v_t$ and $A = \beta I$, then m approaches zero and n approaches b. At those limiting values equation (15) becomes

$$\begin{aligned} \mathrm{E}(r_{t+1}|D_t) &= \mathrm{E}(r) + b' \sum_{s=0}^{\infty} A^s v_{t-s} \\ &= \mathrm{E}(r) + b'[x_t - \mathrm{E}(x)], \end{aligned} \tag{18}$$

which is identical to $a + b'x_t$, the conditional mean given by the predictive regression. When the predictors are imperfect, however, their entire history enters the conditional expected return, since the weighted sum of their past innovations in equation (15) 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.

2.3. Expected Return: The Role of ρ_{uw}

A key feature of the predictive system, in addition to accomodating imperfect predictors, is the ability to incorporate economically motivated prior beliefs about the correlation between the

[8] In 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.

unexpected return, u_t, and the innovation in the expected return, w_t. We now simply denote that correlation as ρ_{uw}. As discussed earlier, it seems likely that $\rho_{uw} < 0$. We find that incorporating such beliefs about ρ_{uw} plays a key role in computing expected returns and assessing the usefulness of various predictors.

As mentioned earlier, an extreme version of imperfect predictors occurs when there are no predictors, so that D_t includes only the return history. This special case provides a simplified setting in which to illustrate the critical role ρ_{uw} can play in the relation between D_t and the conditional expected return. With no predictors, the summation on the right-hand side of equation (15) includes only the first term, so the conditional expected return is simply a weighted sum of past forecast errors in returns (thereby giving the Wold representation). We consider here 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. Panel A of Figure 1 plots the values of $\lambda_s (= m\beta^s)$, the coefficient in equation (15) that multiplies the forecast error ϵ_{t-s}. Not surprisingly, with $\beta = 0.9$, the geometric rate of decay in the coefficients makes them relatively small by lags of around 40 periods. More interesting is the role of ρ_{uw} in determining m. Differences in m across the values for ρ_{uw} produce strikingly different behaviors for the λ_s'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 change effect prevails when ρ_{uw} is sufficiently negative.

When $\rho_{uw} = 0$, there is no change effect and only the level effect is present. For that case, the λ_s's start at a positive value for the first lag, about 0.04, and then decay downward toward zero. The level and change effects can perfectly offset each other, as occurs in this example when

$\rho_{uw} = -0.47$, 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 do not update their beliefs about expected return at all, regardless of what realized returns they observe. The change effect dominates when $\rho = -0.85$, where the λ_s's start around -0.04 at the first lag, and it is even stronger when $\rho = -0.99$, 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.

Since the forecast errors (ϵ_t's) in the above analysis 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 $\epsilon_t^U = r_t - \mathrm{E}(r)$, and then rewrite the conditional expected return in equation (15) as

$$\mathrm{E}(r_{t+1}|D_t) = \mathrm{E}(r) + \sum_{s=0}^{\infty} \left(\omega_s \epsilon_{t-s}^U + \delta_s' v_{t-s}\right), \tag{19}$$

where, again in steady state,

$$\omega_s = m(\beta - m)^s \tag{20}$$

$$\delta_s = n(\beta - m)^s. \tag{21}$$

It can be verified that $\beta - m \geq 0$ for $\beta > 0$. Panel B of Figure 1 plots the values of ω_s in the same no-predictor example discussed above. The patterns are qualitatively similar to those in Panel A, in that the ω_s's are again positive and declining for $\rho_{uw} = 0$, flat at zero for $\rho_{uw} = -0.47$, and negative and increasing for $\rho = -0.85$ and $\rho = -0.99$. In this representation, though, the rates of geometric decay differ, since they depend on m, and returns at longer lags exert a greater relative impact as ρ_{uw} takes larger negative values.

In practice, the true unconditional mean $\mathrm{E}(r)$ must be estimated. Consider again the no-predictor case where, in equation (19), the summation on the right-hand side is truncated at $s = t - 1$ and $\mathrm{E}(r)$ is replaced by the sample mean, $(1/t)\sum_{l=1}^{t} r_l$. Then, given β and m (essentially second-moment quantities), the estimated conditional expected return becomes a weighted average of past returns,

$$\hat{\mathrm{E}}(r_{t+1}|D_t) = \sum_{s=0}^{t-1} \kappa_s r_{t-s}, \tag{22}$$

where

$$\kappa_s = \frac{1}{t}\left(1 - \sum_{l=1}^{t} \omega_l\right) + \omega_s, \tag{23}$$

and $\sum_{s=0}^{t-1} \kappa_s = 1$. The weights ($\kappa_s$'s) are plotted in Panel C 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 no-predictor setting presented above, with the same parameter values as in Figure 1. The unconditional mean return $\mathrm{E}(r)$ 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 the flat (dashed) line at the sample average for the period. The most striking feature of the plot is that the expected return series for $\rho_{uw} = 0$ (solid line) is virtually the mirror image of the series for $\rho_{uw} = -0.85$ (dash-dot line). For example, when $\rho_{uw} = -0.85$, the conditional expected return plots above the unconditional mean during much of the 1970's and early 1980's by amounts that, quarter by quarter, correspond closely to the amounts by which the conditional expected return plots below the unconditional mean when $\rho_{uw} = 0$. 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.

2.4. Predictive System vs. Predictive Regression

We have discussed how an important feature of the predictive system is its ability to incorporate economically motivated prior beliefs about parameters such as ρ_{uw}. We can also compare the predictive regression to the predictive system without incorporating such priors. In the absence of any priors or parameter restrictions, not all of the parameters in equations (8) through (11) are identified, but we can nevertheless obtain estimates of conditional expected returns by joint maximum likelihood estimation of equation (9) and the equation

$$r_{t+1} = (1-\beta)\mathrm{E}(r) + \beta r_t + n' v_t - (\beta - m)\epsilon_t + \epsilon_{t+1}, \tag{24}$$

whose parameters are identified. It is easily verified that equation (24) follows directly from the steady-state representation of the predictive system's conditional expected return in equation (15).

Figure 3 plots the time series of expected returns obtained via maximum-likelihood estimation of the predictive system (equations (9) and (24)) 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.[9]

The above analysis compares the predictive system and the predictive regression in the absence of any information about the parameters beyond what our data provide. At the other extreme is a comparison in which values of the parameters are simply specified. In the latter setting we can compare explanatory power, measured as the R-squared (R^2) in the regression of r_{t+1} on x_t for the predictive regression and as the R^2 in the regression of r_{t+1} on $\mathrm{E}(r_{t+1}|D_t) \equiv \mathrm{E}(\mu_t|D_t)$ for the predictive system. First note that 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,

$$\frac{R^2(r_{t+1} \text{ on } x_t)}{R^2(r_{t+1} \text{ on } \mathrm{E}(\mu_t|D_t))} = \frac{R^2(\mu_t \text{ on } x_t)}{R^2(\mu_t \text{ on } \mathrm{E}(\mu_t|D_t))}, \tag{25}$$

since each of the R^2 values in the latter ratio is equal to its corresponding value in the first ratio multiplied by $\mathrm{Var}(r_{t+1})/\mathrm{Var}(\mu_t)$. The parameters in equations (8) through (11) can be used to obtain the unconditional covariance matrix of μ_t and x_t and thereby the R^2 in the regression of μ_t on x_t,

$$R_1^2 = \frac{\mathrm{Var}[\mathrm{E}(\mu_t|x_t)]}{\mathrm{Var}(\mu_t)}. \tag{26}$$

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

$$R_2^2 = \frac{\mathrm{Var}[\mathrm{E}(\mu_t|D_t)]}{\mathrm{Var}(\mu_t)} = 1 - \frac{\mathrm{Var}(\mu_t|D_t)}{\mathrm{Var}(\mu_t)}. \tag{27}$$

[9]We 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.

The ratio on the right-hand side of equation (25) is then computed as R_1^2/R_2^2. Note that this R^2 ratio cannot exceed one because $x_t \in D_t$. In other words, the predictive system always produces a more precise estimate of μ_t than the predictive regression, simply because it uses more information. The smaller the R^2 ratio, the larger the advantage of using the predictive system.

Table II reports R_1^2/R_2^2 under various possible combinations of parameters in the single-predictor case. We use the same values for the true predictive R^2 and ρ_{uw} as before but we now let β take not only the value of 0.9 but also 0.97, and we do the same for A. Note that 0.97 is closer to the quarterly sample autocorrelations of predictors such as the dividend yield. Finally, we let ρ_{vw}, the correlation between v_t and w_t, range from 0.1 to 0.9. The parameter combinations given do not uniquely determine R_1^2/R_2^2, so for each combination Table II reports that ratio's minimum and maximum values as well as its mean, computed as the equally weighted average across values from -1 to 1 for the partial correlation of u_t and v_t given w_t.

Table II shows that the R^2 ratio can take essentially any value in its admissible range of $(0, 1)$, but some interesting patterns emerge. The degree of imperfection in the predictor is low when ρ_{vw} is high and when μ_t and x_t have similar autocorrelations ($\beta \approx A$). The relative explanatory power of the predictive regression should be the highest in those cases and, indeed, when $\rho_{vw} = 0.9$ and $\beta = A$ $(= 0.9)$, R_1^2/R_2^2 ranges from 0.81 to 1.0 and is relatively insensitive to ρ_{uw}. With more imperfection in the predictor, the relative performance of the predictive regression can fall substantially. Even maintaining $\rho_{vw} = 0.9$ but letting β and A assume different values, 0.9 versus 0.97, results in R_1^2/R_2^2 dropping below 0.70, sometimes considerably so. In other words, simply having a predictor whose persistence departs from that of the true expected return by what might seem a rather modest degree is sufficient to place the predictive regression at a distinctly greater disadvantage in terms of explanatory power.

The comparison of the predictive regression to the predictive system has thus far been conducted at two opposite ends of the spectrum in terms of information about parameters. At one end, where the only information comes from our data, we see that the predictive regression and the predictive system deliver estimates of conditional expected returns that are reasonably similar for a given set of predictors. At the other end of the spectrum, with parameter values taken as given, we see that the conditional expected returns delivered by the predictive system can be considerably more precise than those from the predictive regression when the predictors are imperfect. In the next section, we revisit the comparison of explanatory powers of the two approaches at intermediate points on the spectrum, where we neither take parameters as given nor rely solely on the data but instead we estimate the parameters under economically motivated priors.

3. Empirical Analysis

3.1. Estimation Approach

We develop a Bayesian approach for estimating the predictive system. This approach has several advantages over the frequentist alternatives such as the maximum likelihood approach. First, the 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^2s 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 $(\theta, A, \alpha, \beta, \Sigma)$. The details are in the Appendix.

3.2. Prior Beliefs

We impose informative prior distributions on three quantities:

1. The correlation ρ_{uw} between unexpected returns and innovations in expected returns,
2. The persistence β of the true expected return μ_t,
3. The predictive R^2 from the regression of r_{t+1} on μ_t.

These prior distributions are plotted in Figure 4.

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 4. 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. 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 (see equation (5)). 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 the evidence of Campbell (1991), Campbell and Ammer (1993), and others who show that aggregate market returns are driven mostly by discount rate news. Our more informative prior implies a prior on ρ_{uw}^2 whose mean is about 0.77, which fits the existing evidence quite well.[10]

Note that 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 see no economic reason to be informative about the variance of v_t or about its covariances with the other error terms. Instead, we build on Stambaugh (1997) and form the prior on Σ as the posterior from a hypothetical sample that contains more information about the covariance between u_t and w_t than about the other covariance elements of Σ. The details are in the Appendix.

In addition to the 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 4, 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. The prior on β, which is plotted in Panel D of Figure 4, has most of its mass above 0.7 and there is virtually no prior mass below 0.5. The prior distributions on all other parameters (θ, A, α, and most elements of Σ) are noninformative.

3.3. Explanatory Advantage of the Predictive System

In Section 2.4., we show in a theoretical setting that a predictive system produces more precise estimates of expected return than a predictive regression. In this section, we quantify the advantage of the predictive system empirically. The sample period is 1952Q1–2003Q4, as before.

Recall that our theoretical comparison of the explanatory powers of the predictive system and the predictive regression is based on the ratio of two R^2's in equation (25), and that the smaller the R^2 ratio, the larger the advantage of using the predictive system. Table III shows the posterior

[10]Following up on footnote 3, the implied estimates of ρ_{uw}^2 from Campbell (1991) range from 0.50 to 0.74 in 1927–1988, from 0.44 to 0.76 in 1927–1951, and from 0.84 to 0.88 in 1952–1988 across three different specifications. The estimates of ρ_{uw}^2 implied by the evidence of Campbell and Ammer (1993) in their Table III range from 0.86 to 0.91.

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. This result shows that the theoretical explanatory advantage of the predictive system extends to our empirical setting. 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.

The predictive system produces more precise expected return estimates because it uses more information, not only the most recent predictor values but also their lags and the full history of asset returns. One way to analyze this additional information is to examine the coefficients λ_s and ϕ_s from equation (15), which capture the influence of past unexpected returns and predictor innovations on the estimate of expected return. Figure 5 plots the first 30 lags of λ_s and ϕ_s. Both coefficients decay as the number of lags increases, by construction, but they are mostly nontrivially different from zero at the first 10-20 quarterly lags. Both coefficients also depend on the prior for ρ_{uw}, as expected.

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.[11] The dotted line corresponds to the prior that is noninformative about ρ_{uw} but informative about the process for μ_t (i.e., β and R^2). In all three panels of Figure 6, the dotted line is substantially

[11] Although b does not play an explicit role in the predictive system, its value can be computed from the system's parameters as $b = V_{\mu x} V_{xx}^{-1}$, or the covariance between μ_t and x_t multiplied by the inverse of the covariance matrix of x_t. Since both $V_{\mu,x}$ and V_{xx} can be computed from the system's parameters, posterior draws of b can be constructed from the posterior draws of the underlying parameters. We construct the posteriors of several other quantities such as λ_s and ϕ_s in an analogous manner.

different from the dashed line, which means that imposing the prior that μ_t is stable and persistent significantly affects the inference about predictability. Put differently, in the standard predictive regression approach, which does not impose such a prior, the estimates of μ_t are more variable or less persistent or both (recall Figure 3). 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 the process for μ_t but also about ρ_{uw}. The prior on ρ_{uw} clearly affects the inference about predictability. For example, 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.

3.4. How Imperfect Are Predictors?

In the traditional predictive regression framework, expected return is assumed to be perfectly correlated with a linear combination of a given set of predictors. However, this assumption is almost surely too strong. The predictive system relaxes this assumption. Instead of assuming that the correlation between μ_t and x_t is one, the predictive system allows us to estimate this correlation. In this section, we use the predictive system to analyze the degree to which a given set of predictors can capture the unobservable true expected return μ_t.

Figures 7 through 9 summarize the evidence for three predictive systems, in which the predictors are the dividend yield alone (Figure 7), the bond yield alone (Figure 8), and the dividend yield, bond yield, and CAY together (Figure 9).

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 all three figures, the R^2s 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 none of the three sets of predictors are likely to be perfectly correlated with μ_t.[12]

[12]Note 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.

The R^2 also 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). In Panel A of Figure 9, becoming increasingly informative about ρ_{uw} shifts the posterior to the right again. The reason is that the set of predictors includes the dividend yield and CAY, both of which are negatively correlated with contemporaneous returns.

Panel B of each figure 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 Figures 7 and 8 and decreases it in Figure 9, but these effects are relatively small. Since we put a fairly informative prior on the predictive R^2 (see Panel C of Figure 4), the posterior is not dramatically different from the prior in any of the three figures.

The remaining panels of each figure plot the posteriors of the partial correlations between each predictor and μ_t, both conditional ($\rho_{v,w}$) and unconditional ($\rho_{x,\mu}$). (Partial correlations are correlations that control for the presence of other predictors.) These correlations are all well below one 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 (Figure 7) and the bond yield to be less highly correlated with μ_t (Figure 8). For CAY in Figure 9, the prior does not affect $\rho_{v,w}$ much but it increases $\rho_{x,\mu}$. Among the three predictors in Figure 9, the bond yield exhibits the lowest partial correlations with μ_t under the more informative prior. For example, almost a third of the posterior distribution for $\rho_{x,\mu}$ is below zero and the posterior mode is only about 0.2. The dividend yield and CAY exhibit substantially higher $\rho_{x,\mu}$'s, with posterior modes of about 0.7, and their conditional correlations $\rho_{v,w}$ are even slightly higher.

Overall, Figures 7 through 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 power of the dividend yield and CAY but they weaken the predictive power of the bond yield.

3.5. Estimates of Expected Return

Figure 10 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 3. 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 IV 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} (noninformative, less informative, more informative). 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 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 0.15% 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 IV 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 10. No average utility loss exceeds 0.15% per quarter, no mean absolute difference is larger than 0.55% per quarter, and all correlations exceed 84.5%. 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 0.15% 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 1.40% per quarter. Both quantities are highly economically significant. In Panel C, where we use all three predictors, the differences across the five approaches are even larger. For the same comparison as earlier in this paragraph, the mean absolute difference in expected returns is 1.80% per quarter and the average certainty equivalent loss is 1.49% per quarter.

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}.

3.6. Variance Decomposition of Expected Return

In a predictive regression, expected return μ_t is perfectly correlated with a 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 2.2.). 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.

First, we rewrite the AR(1) processes for x_t and μ_t in equations (9) and (10), respectively, as moving average processes with an infinite number of lags:

$$x_t = E_x + \sum_{i=0}^{\infty} A^i v_{t-i} \tag{28}$$

$$\mu_t \;=\; E_r + \sum_{i=0}^{\infty} \beta^i w_{t-i}, \tag{29}$$

where $E_r \equiv \mathrm{E}(r_t)$ and $E_x \equiv \mathrm{E}(x_t)$. Then we project w_t linearly on u_t and v_t:

$$w_t \;=\; [\sigma_{wu}\ \sigma_{wv}] \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}^{-1} \begin{bmatrix} u_t \\ v_t \end{bmatrix} + \upsilon_t = \psi_u u_t + \psi_v v_t + \upsilon_t. \tag{30}$$

Substituting for w_t from equation (30) into equation (29), we obtain

$$\mu_t \;=\; (E_r - \psi_v E_x) + \psi_v x_t + \psi_u \sum_{i=0}^{\infty} \beta^i u_{t-i} + \psi_v \sum_{i=0}^{\infty} \left(\beta^i I_K - A^i\right) v_{t-i} + \sum_{i=0}^{\infty} \beta^i \upsilon_{t-i}, \tag{31}$$

where K is the number of predictors and I_K is a $K \times K$ identity matrix. Equation (31) 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.[13] Based on this equation, we can decompose the variance of μ_t into the components due to x_t, $\{u_s\}_{s\le t}$, and $\{v_s\}_{s\le t}$. See the Appendix for details.

Table V reports the posterior means and standard deviations of the R^2s from the regressions of μ_t on x_t (column 1), μ_t on x_t and $\{u_s\}_{s\le t}$ (column 2), and μ_t on x_t and $\{u_s, v_s\}_{s\le t}$ (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 $\{u_s, v_s\}_{s\le t}$ 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 V. The most striking effect obtains for the bond yield, for which adding $\{u_s, v_s\}_{s\le t}$ 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 $\{u_s, v_s\}_{s\le t}$ in addition to x_t.

The R^2's in Table V 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.

[13] If the predictors mean-revert at the same rate as μ_t and possess no lagged cross-correlations ($\beta^i I_K = A^i$), then the term involving v_{t-i} drops out of the equation.

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. Conclusions

Unlike a predictive regression, a predictive system accommodates imperfect predictors as well as the prior belief that expected and unexpected returns are negatively correlated. 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 lagged returns receive a negative weight when prior beliefs attribute a significant portion of the variance in unexpected returns to changes in expected returns. The prior for the correlation between expected and unexpected returns can also have a substantial effect on inferences about a predictor's correlation with expected return. The lags of returns and predictors often account for a large fraction of the variation in conditional expected returns when the predictors are imperfect. We observe economically significant differences across estimates of conditional expected returns, not only for predictive regressions versus predictive systems but across different specifications of priors within predictive systems as well.

Our initial exploration of predictive systems could be extended in various directions. 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. The predictive system is formulated as a one-period-ahead model, but it seems clear that it can deliver conditional expected returns for longer horizons as well. It could be interesting to investigate whether, when predictors are imperfect, observations of long-horizon returns can provide additional insight into the properties of expected returns, such

as their persistence. 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.[14]

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 we are exploring them in a separate work in progress. See Cochrane (2006) for a recent analysis of the interaction between return predictability and cash flow predictability.

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 just the sample average.[15] A Bayesian investor with a quadratic (MSE) loss function would prefer a forecast that combines his priors and the available 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 not sufficiently high, as discussed by Clark and West (2004, 2005) 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. Clark and West (2004, 2005) develop frequentist tests based on out-of-sample statistics, and it could be interesting to pursue model selection issues for predictive systems.

[14] See, for example, Lamont (1998), Lewellen (1999), Ang and Bekaert (2006), Santos and Veronesi (2006), etc.

[15] Goyal and Welch (2003, 2005) and Campbell and Thompson (2005), among others, investigate the abilities of predictive regressions and sample averages to forecast stock returns out of sample.

Appendix.

In parts of the Appendix, we work 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. In those parts, we maintain the usual convention that matrices are denoted by uppercase letters, so we replace β by B, the σ's by the corresponding Σ's, etc.

We restate the predictive system from equations (8) through (10) here in the multi-asset case:

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} && \text{(A1)} \\ x_{t+1} &= \theta + Ax_t + v_{t+1} && \text{(A2)} \\ \mu_{t+1} &= \alpha + B\mu_t + w_{t+1}, && \text{(A3)} \end{aligned}$$

with the disturbances distributed identically and independently across t as

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

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

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

Let z_t denote the vector of the observed data at time t,

$$z_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}. \tag{A6}$$

Denote the data we observe through time t as $D_t = (z_1, \ldots, z_t)$, and note that our complete data consist of D_T. Also define

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\mu} = \begin{bmatrix} V_{r\mu} \\ V_{x\mu} \end{bmatrix}. \tag{A7}$$

From the above we obtain

$$\begin{aligned} E_r &= (I - B)^{-1}\alpha && \text{(A8)} \\ E_x &= (I - A)^{-1}\theta && \text{(A9)} \\ V_{rr} &= V_{\mu\mu} + \Sigma_{uu} && \text{(A10)} \\ V_{xx} &= AV_{xx}A' + \Sigma_{vv} && \text{(A11)} \\ V_{\mu\mu} &= BV_{\mu\mu}B' + \Sigma_{ww} && \text{(A12)} \\ V_{rx} &= V_{\mu x}A' + \Sigma_{uv} && \text{(A13)} \\ V_{\mu x} &= BV_{\mu x}A' + \Sigma_{wv} && \text{(A14)} \\ V_{\mu r} &= BV_{\mu\mu} + \Sigma_{wu} && \text{(A15)} \end{aligned}$$

and equations (A11), (A12), and (A14) can be written in explicit form as[16]

$$\begin{aligned}
\text{vec}\,(V_{xx}) &= [I-(A\otimes A)]^{-1}\text{vec}\,(\Sigma_{vv}) && \text{(A16)}\\
\text{vec}\,(V_{\mu\mu}) &= [I-(B\otimes B)]^{-1}\text{vec}\,(\Sigma_{ww}) && \text{(A17)}\\
\text{vec}\,(V_{\mu x}) &= [I-(A\otimes B)]^{-1}\text{vec}\,(\Sigma_{wv}). && \text{(A18)}
\end{aligned}$$

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, originally 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

$$\begin{aligned}
a_t &= \text{E}(\mu_t|D_{t-1}) && \text{(A19)}\\
b_t &= \text{E}(\mu_t|D_t) && \text{(A20)}\\
e_t &= \text{E}(z_t|\mu_t, D_{t-1}) && \text{(A21)}\\
f_t &= \text{E}(z_t|D_{t-1}) && \text{(A22)}\\
P_t &= \text{Var}(\mu_t|D_{t-1}) && \text{(A23)}\\
Q_t &= \text{Var}(\mu_t|D_t) && \text{(A24)}\\
R_t &= \text{Var}(z_t|\mu_t, D_{t-1}) && \text{(A25)}\\
S_t &= \text{Var}(z_t|D_{t-1}) && \text{(A26)}\\
G_t &= \text{Cov}(z_t, \mu_t'|D_{t-1}) && \text{(A27)}
\end{aligned}$$

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

$$\mu_0|D_0 \sim N(b_0, Q_0), \tag{A28}$$

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

$$\mu_1|D_0 \sim N(a_1, P_1), \tag{A29}$$

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

$$z_1|D_0 \sim N(f_1, S_1), \tag{A30}$$

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

$$G_1 = V_{z\mu} \tag{A31}$$

[16]The solutions employ the identity $\text{vec}\,(DFG) = (G'\otimes D)\text{vec}\,(F)$ (e.g., Hamilton, 1994, p. 265).

and that

$$z_1|\mu_1, D_0 \sim N(e_1, R_1), \tag{A32}$$

where

$$\begin{aligned} e_1 &= f_1 + G_1 P_1^{-1}(\mu_1 - a_1) && \text{(A33)} \\ R_1 &= S_1 - G_1 P_1^{-1} G_1'. && \text{(A34)} \end{aligned}$$

Combining this density with equation (A29) using Bayes rule gives

$$\mu_1|D_1 \sim N(b_1, Q_1), \tag{A35}$$

where

$$\begin{aligned} b_1 &= a_1 + P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1}(z_1 - f_1) && \text{(A36)} \\ Q_1 &= P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} P_1. && \text{(A37)} \end{aligned}$$

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

$$\begin{aligned} a_t &= \alpha + B b_{t-1} && \text{(A38)} \\ P_t &= B Q_{t-1} B' + \Sigma_{ww} && \text{(A39)} \\ f_t &= \begin{bmatrix} b_{t-1} \\ \theta + A x_{t-1} \end{bmatrix} && \text{(A40)} \\ S_t &= \begin{bmatrix} Q_{t-1} + \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} && \text{(A41)} \\ G_t &= \begin{bmatrix} Q_{t-1} B' + \Sigma_{uw} \\ \Sigma_{vw} \end{bmatrix} && \text{(A42)} \\ e_t &= f_t + G_t P_t^{-1}(\mu_t - a_t) && \text{(A43)} \\ R_t &= S_t - G_t P_t^{-1} G_t' && \text{(A44)} \\ b_t &= a_t + P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1}(z_t - f_t) && \text{(A45)} \\ Q_t &= P_t(P_t + G_t' R_t^{-1} G_t)^{-1} P_t. && \text{(A46)} \end{aligned}$$

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

Sampling

Let

$$\zeta_t = \begin{bmatrix} r_t \\ x_t \\ \mu_t \end{bmatrix}. \tag{A47}$$

We wish to draw $(\mu_0, \mu_1, \ldots, \mu_T)$ conditional on D_T. The backward-sampling approach relies on the Markov property of the evolution of ζ_t and the resulting identity,

$$p(\zeta_0, \zeta_1, \ldots, \zeta_T|D_T) = p(\zeta_T|D_T)\, p(\zeta_{T-1}|\zeta_T, D_{T-1}) \cdots p(\zeta_1|\zeta_2, D_1)\, p(\zeta_0|\zeta_1, D_0). \tag{A48}$$

We first sample μ_T from $p(\mu_T|D_T)$, the normal density obtained in the last step of the filtering. Then, for $t = T-1, T-2, \ldots, 1, 0$, 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

$$\zeta_{t+1}|D_t \sim N\left(\begin{bmatrix} b_t \\ \theta + Ax_t \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}\right), \tag{A49}$$

$$\zeta_t|D_t \sim N\left(\begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix}\right), \tag{A50}$$

and

$$\mathrm{Cov}(\zeta_t, \zeta'_{t+1}|D_t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix}. \tag{A51}$$

Therefore,

$$\zeta_t|\zeta_{t+1}, D_t \sim N(h_t, H_t), \tag{A52}$$

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 (15), (19), and (22). 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, $b_t = \mathrm{E}(r_{t+1}|D_t)$, as a function of past returns and predictors. Denote

$$[M_t \; N_t] \equiv P_t(P_t + G'_t R_t^{-1} G_t)^{-1} G'_t R_t^{-1}, \tag{A53}$$

so that, from equation (A45), for $t > 1$,

$$\begin{aligned} b_t &= a_t + [M_t\ N_t](z_t - f_t) \\ &= \alpha + Bb_{t-1} + [M_t\ N_t]\begin{bmatrix} r_t - b_{t-1} \\ x_t - \theta - Ax_{t-1} \end{bmatrix} \\ &= (I - B)E_r + (B - M_t)b_{t-1} + M_t r_t + N_t v_t, \end{aligned} \tag{A54}$$

or

$$b_t - E_r = B(b_{t-1} - E_r) + M_t(r_t - b_{t-1}) + N_t v_t. \tag{A55}$$

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

$$b_t = E_r + \sum_{s=1}^{t} \Lambda_s(r_s - b_{s-1}) + \sum_{s=1}^{t} \Phi_s v_s, \tag{A56}$$

where

$$\Lambda_s = B^{t-s} M_s \tag{A57}$$

$$\Phi_s = B^{t-s} N_s. \tag{A58}$$

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 (15) in the text.

The current 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 (A54) slightly as

$$b_t - E_r = (B - M_t)(b_{t-1} - E_r) + M_t(r_t - E_r) + N_t v_t \tag{A59}$$

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

$$b_t = E_r + \sum_{s=1}^{t} \Omega_s(r_s - E_r) + \sum_{s=1}^{t} \Delta_s v_s \tag{A60}$$

where

$$\Omega_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) M_s & \text{for } s < t \\ M_s & \text{for } s = t \end{cases} \tag{A61}$$

$$\Delta_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) N_s & \text{for } s < t \\ N_s & \text{for } s = t \end{cases} \tag{A62}$$

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 (19) in the text.

If E_r is replaced by the sample mean, $(1/t)\sum_{l=1}^{t} r_l$, in equation (19), then the estimate of b_t becomes

$$\hat{b}_t = \sum_{s=1}^{t} \mathcal{K}_s r_s + \sum_{s=1}^{t} \Delta_s v_s, \tag{A63}$$

where

$$\mathcal{K}_s = \frac{1}{t}\left(I - \sum_{l=1}^{t} \Omega_l\right) + \Omega_s, \tag{A64}$$

and $\sum_{s=1}^{t} \mathcal{K}_s = I$. This is a generalized version of equation (22) in the text.

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which we predict the return on a single asset, so that 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 4). 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 (9) and (10) 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, $E_\mu \sim N(\bar{\mu}, \sigma^2_{E_\mu})$, centered at the sample mean return with a large prior standard deviation of 1% per quarter. We use a noninformative prior for E_x, $E_x \sim N(0, \sigma^2_{E_x} I_K)$ 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 informative about the 2×2 submatrix $\Sigma_{11} \equiv [\sigma_u^2 \ \sigma_{uw}; \sigma_{wu} \ \sigma_w^2]$ but noninformative about the elements of Σ that involve v (i.e., $\Sigma_{vv}, \sigma_{vu}, \sigma_{vw}$). Such a prior on Σ is obtained as a posterior 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. We choose T_0 equal to one fifth of the sample size, which makes the prior on Σ_{11} quite informative (five times less informative than the actual sample). We choose $S_0 = K + 3$, where K is the number of predictors used ($K = 1$ or 3), which makes the prior on the elements of Σ that involve v virtually noninformative (as informative as a sample of only $K + 3$ observations).

We obtain this latter prior by changing variables from $(\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$ to the slope C and the residual covariance matrix Ω from the regression of v_t on (u_t, w_t). 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 the estimates from a hypothetical sample of S_0 observations. We choose a very small value for S_0, as explained above. In addition, we choose a hypothetical sample in which the right-hand side variable is much less volatile than the residuals. Both choices help make the prior on C and Ω noninformative.

The prior on Σ_{11} is inverted Wishart, $\Sigma_{11} \sim IW(T_0\hat{\Sigma}_{11,0}, T_0 - K - 1)$. The prior mean of σ_u^2 is set equal to 95% of the sample variance of market returns, and the prior mean of σ_w^2 is obtained from the total variance of μ_t under the assumption that $\beta = 0.97$. These assumptions lead to a prior for the R^2 from the regression of r_{t+1} on μ_t that we find plausible (see Figure 4). 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 in which the off-diagonal element of $\hat{\Sigma}_{11,0}$ is unknown. Specifically, denoting the (i, j) element of $\hat{\Sigma}_{11,0}$ by M_{ij}, for $i = 1, 2$ and $j = 1, 2$, we assume that M_{11} and M_{22} are known but M_{12} is an unknown hyperparameter with a uniform prior distribution on the interval $(-\underline{c}\sqrt{M_{11}M_{22}}, \overline{c}\sqrt{M_{11}M_{22}})$. Under this specification, the prior mean of ρ_{uw} is approximately uniformly distributed as $U(-\underline{c}, \overline{c})$. For all three priors on ρ_{uw}, we specify $\underline{c} = -0.90$ and we vary $\overline{c}$ 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 4.

Posterior distributions

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

After changing variables from (θ, α) into (E_x, E_μ), the equations (9) and (10) 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 $\Sigma_{(vw)} \equiv \left[\Sigma_{vv}\ \sigma_{vw}; \sigma_{wv}\ \sigma_w^2\right]$. The prior for $E_{x\mu}$ is

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

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

$$E_{x\mu}|\cdot \sim N\left(\tilde{E}_{x\mu}, \tilde{V}_{x\mu}\right), \tag{A65}$$

where $\tilde{V}_{x\mu} = (V_{x\mu_0}^{-1} + TL_2'\Sigma_{(vw)}^{-1}L_2)^{-1}$ and $\tilde{E}_{x\mu} = \tilde{V}_{x\mu}\left[V_{x\mu_0}^{-1}E_{x\mu_0} + L_2'\Sigma_{(vw)}^{-1}\sum_{t=1}^{T}(q_{t+1} - L_1q_t)\right]$.

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

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 (9) and (10) 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

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

where b_0 and V_{b_0} are chosen as explained earlier and $1_{b \in S}$ is equal to one when x_t and μ_t are stationary and zero otherwise. Let $\hat{V}_b = \left[Z'(\Sigma_{(vw)}^{-1} \otimes I_{T-1}) Z \right]^{-1}$ 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

$$b|\cdot \sim N\left(\tilde{b}, \tilde{V}_b\right) \times 1_{b \in S}, \tag{A66}$$

where $\tilde{V}_b = (V_{b_0}^{-1} + \hat{V}_b^{-1})^{-1}$ and $\tilde{b} = \tilde{V}_b \left(V_{b_0}^{-1} b_0 + \hat{V}_b^{-1} \hat{b} \right)$. We obtain the posterior draws of b by making draws from $N\left(\tilde{b}, \tilde{V}_b\right)$ 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 $\Sigma = \left[\sigma_u^2 \ \sigma_{uv} \ \sigma_{uw}; \sigma_{vu} \ \Sigma_{vv} \ \sigma_{vw}; \sigma_{wu} \ \sigma_{wv} \ \sigma_w^2\right]$ 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

$$p(M_{12}|\Sigma_{11}) \propto |\hat{\Sigma}_{11,0}|^{\frac{T_0-K-1}{2}} \exp\left\{ -\frac{T_0}{2} \text{tr}(\Sigma_{11}^{-1} \hat{\Sigma}_{11,0}) \right\}, \ M_{12} \in (-\underline{c}\sqrt{M_{11} M_{22}}, \overline{c}\sqrt{M_{11} M_{22}}), \tag{A67}$$

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 $(-\underline{c}\sqrt{M_{11} M_{22}}, \overline{c}\sqrt{M_{11} M_{22}})$. 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 $(u_t, v_t, w_t), t = 1, \ldots, T$. 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 $X = [\iota_T\ Y_{1,T}]$. 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:

$$\Sigma_{11}|\cdot \ \sim \ IW(T_0\hat{\Sigma}_{11,0} + T\hat{\Sigma}_{11}, T + T_0 - K - 1). \tag{A68}$$

In addition, let $V_C = (X'_0X_0 + X'X)^{-1}$, $\tilde{C} = V_C\left[(X'_0X_0)\hat{C}_0 + (X'X)\hat{C}\right]$, $\tilde{c} = \text{vec}\,(\tilde{C})$, and $D = \hat{C}'_0X'_0X_0\hat{C}_0 + \hat{C}'X'X\hat{C} - \tilde{C}'V_C^{-1}\tilde{C}$. The posterior of Ω has an inverted Wishart distribution:

$$\Omega|\cdot \ \sim \ IW(S_0\hat{\Omega}_0 + T\hat{\Omega} + D, T + S_0 - 1), \tag{A69}$$

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

$$c|\Omega, \cdot \ \sim \ N(\tilde{c}, \Omega \otimes V_C). \tag{A70}$$

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.

The R^2 ratios.

The numerator of the R^2 ratio in equation (25) is computed as

$$R^2(\mu_t \text{ on } x_t) = \frac{\text{Var}(\text{E}(\mu_t|x_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(\text{E}(\mu_t) + V_{\mu x}V_{xx}^{-1}(x_t - \text{E}(x_t)))}{\text{Var}(\mu_t)} = \frac{V_{\mu x}V_{xx}^{-1}V'_{\mu x}}{V_{\mu\mu}}, \tag{A71}$$

where V_{xx}, $V_{\mu\mu}$, and $V_{x\mu}$ are given in equations (A16), (A17), and (A18), respectively.

The denominator of the R^2 ratio in equation (25) is computed as

$$R^2(\mu_t \text{ on } D_t) = \frac{\text{Var}(\text{E}(\mu_t|D_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(\mu_t) - \text{Var}(\mu_t|D_t))}{\text{Var}(\mu_t)} = 1 - \frac{Q_t}{V_{\mu\mu}}, \tag{A72}$$

where Q_t is given in equation (A46). We replace Q_t by its steady-state value, Q, which can be shown to be equal to a solution of a quadratic equation:

$$Q \ = \ \frac{\sqrt{\xi_1^2 - 4\xi_2} - \xi_1}{2}, \tag{A73}$$

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

The value of Q is also used in computing the steady-state values of M_t and N_t from equation (A53), denoted by m_t and n_t in the scalar case:

$$
\begin{aligned}
m &= (\beta Q + \mathrm{Cov}(u,w|v))(Q + \mathrm{Var}(u|v))^{-1} && \text{(A74)} \\
n &= (\sigma_{wv} - m\sigma_{uv})\Sigma_{vv}^{-1}. && \text{(A75)}
\end{aligned}
$$

Variance decomposition of expected return.

In equation (31), 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 $= \sum_{i=0}^{\infty} \left(\beta^i I_K - A^i\right) v_{t-i}$, an infinite sum of current and lagged predictor innovations ,

plus an error term. In the variance decomposition in Table V, 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

$$
R^2(\mu_t \text{ on } C) = \frac{V_{\mu C}' V_C^{-1} V_{\mu C}}{V_{\mu\mu}}. \qquad \text{(A76)}
$$

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

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

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

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

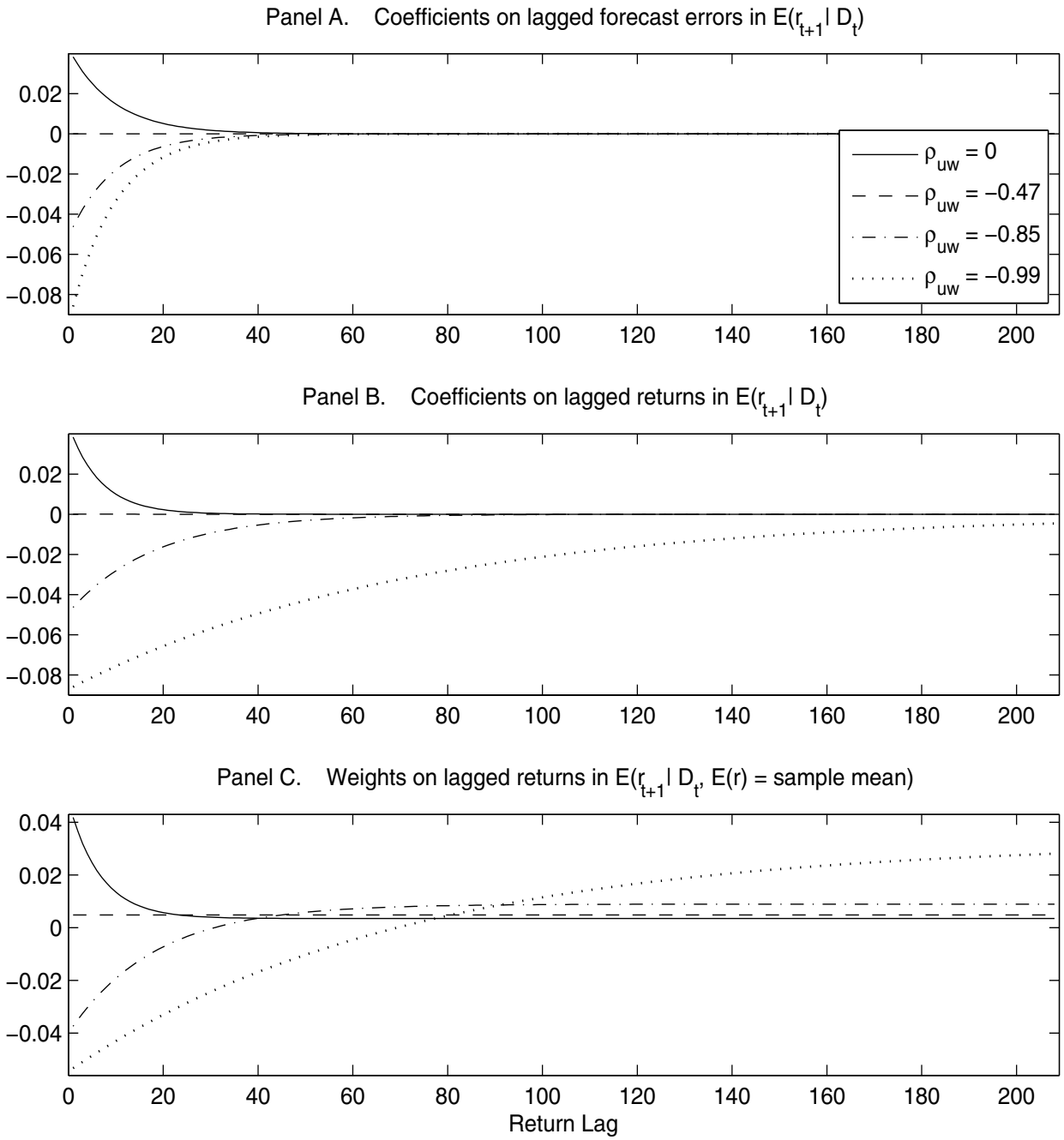


Figure 1. The effect of lagged returns on $\mathrm{E}(r_{t+1}|D_t)$ when no predictors are used. Panel A plots λ_s, the coefficients on lagged forecast errors ($\epsilon_{t-s} = r_{t-s} - \mathrm{E}(r_{t-s}|D_{t-s-1})$) in $\mathrm{E}(r_{t+1}|D_t)$. Panel B plots ω_s, the coefficients on lagged total returns in $\mathrm{E}(r_{t+1}|D_t)$. Panel C plots κ_s, the weights on lagged total returns in $\mathrm{E}(r_{t+1}|D_t)$ 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 $R^2 = 0.05$.

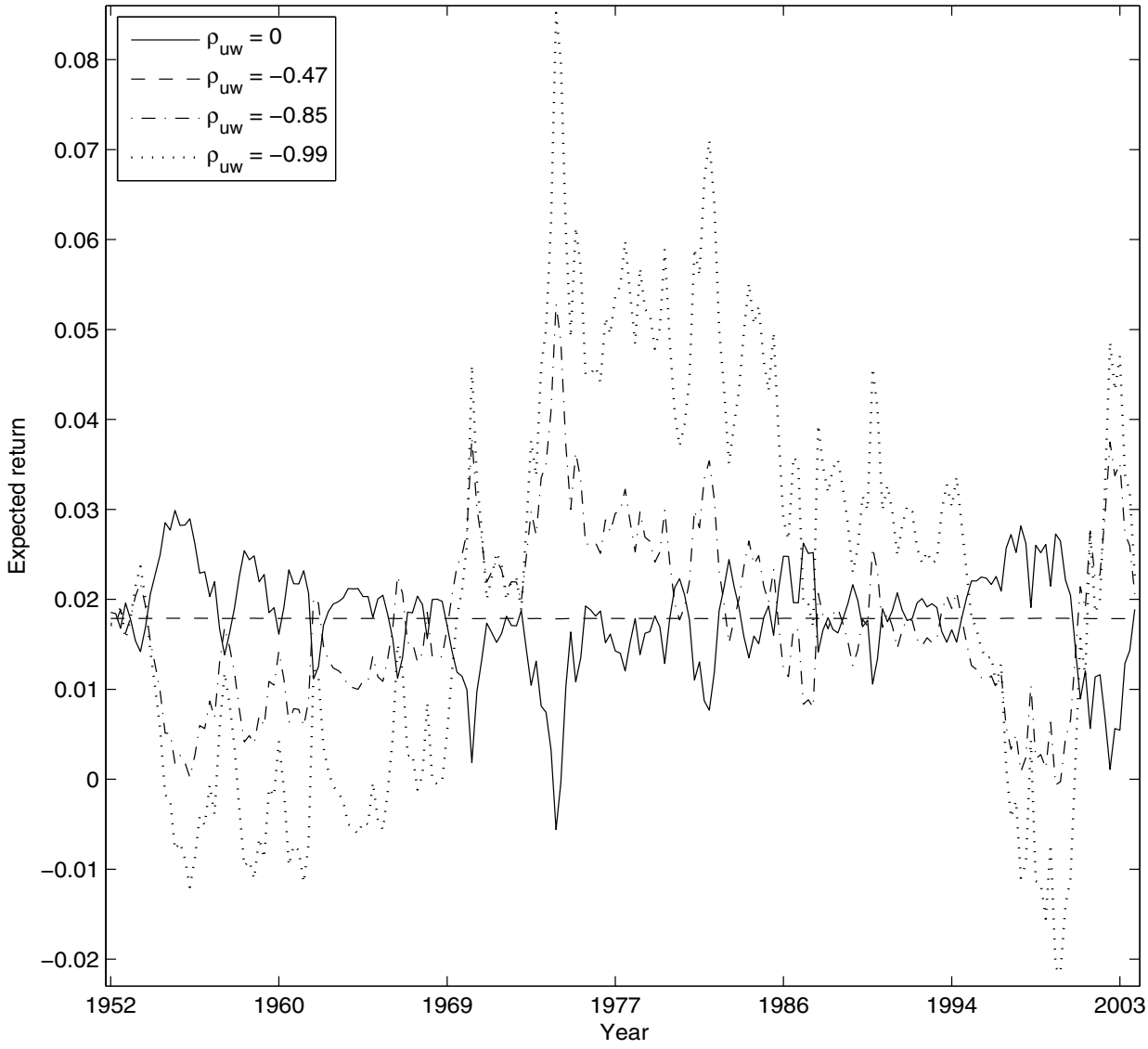


Figure 2. The equity premium $\mathrm{E}(r_{t+1}|D_t)$ **from the predictive system with no predictors.** This figure plots the time series of the quarterly equity premium estimated for four different values of ρ_{uw}, the correlation between expected and unexpected returns. The mean reversion coefficient 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.

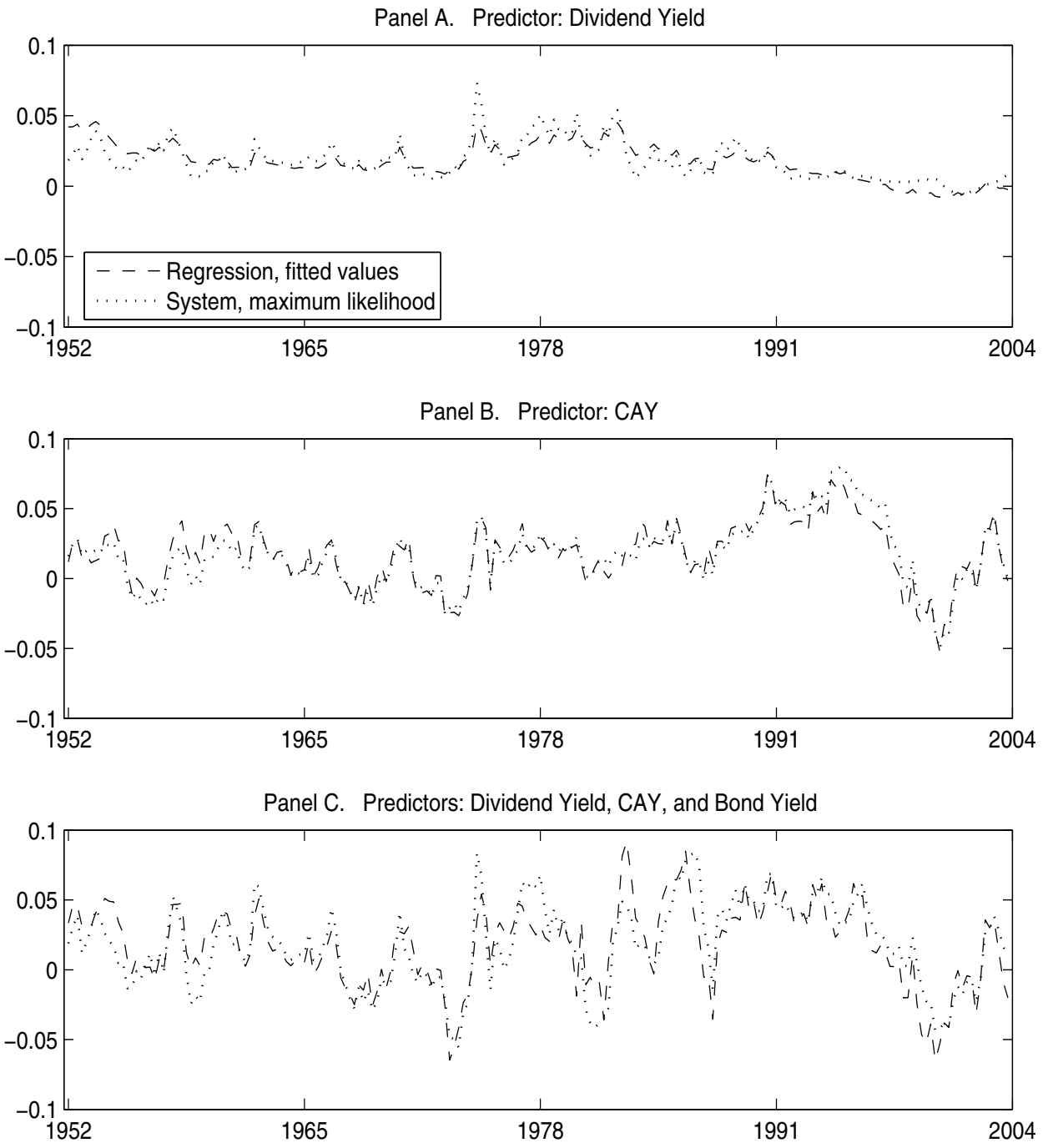


Figure 3. The equity premium: Regression vs. system with no prior information. This figure plots the time series of the quarterly equity premium estimated in two different environments. The *dashed* line plots the OLS fitted values from the predictive regression of r_{t+1} on the given predictor(s). The *dotted* line plots the maximum likelihood estimates of $\mathrm{E}(r_{t+1}|D_t)$ 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.

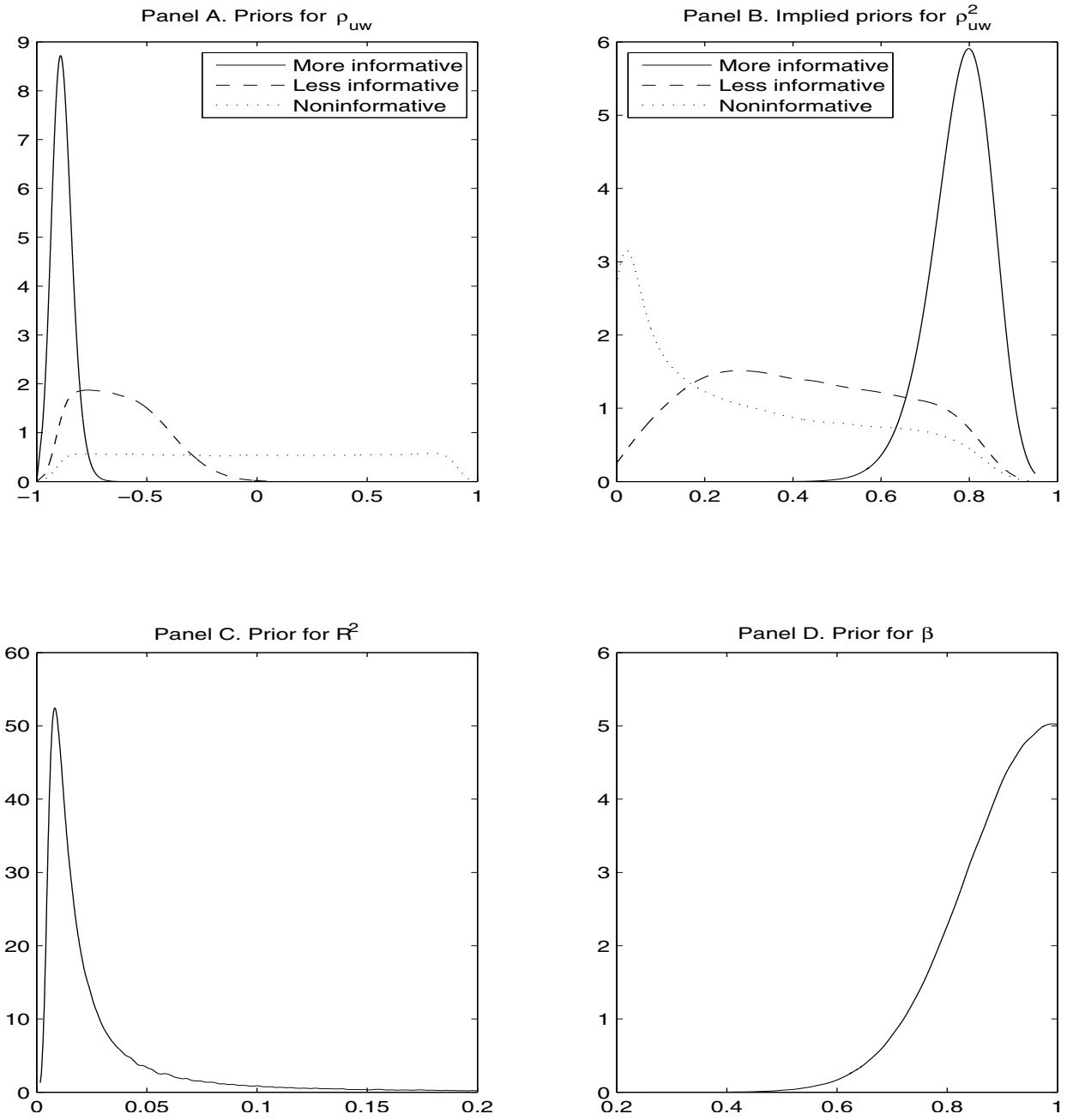


Figure 4. Prior distributions. Panel A plots three prior distributions for the correlation between expected and unexpected returns, ρ_{uw}. The noninformative prior (dotted line) is flat between -0.9 and 0.9, with tails fading away as ρ_{uw} approaches ± 1. The less informative prior (dashed line) has 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.

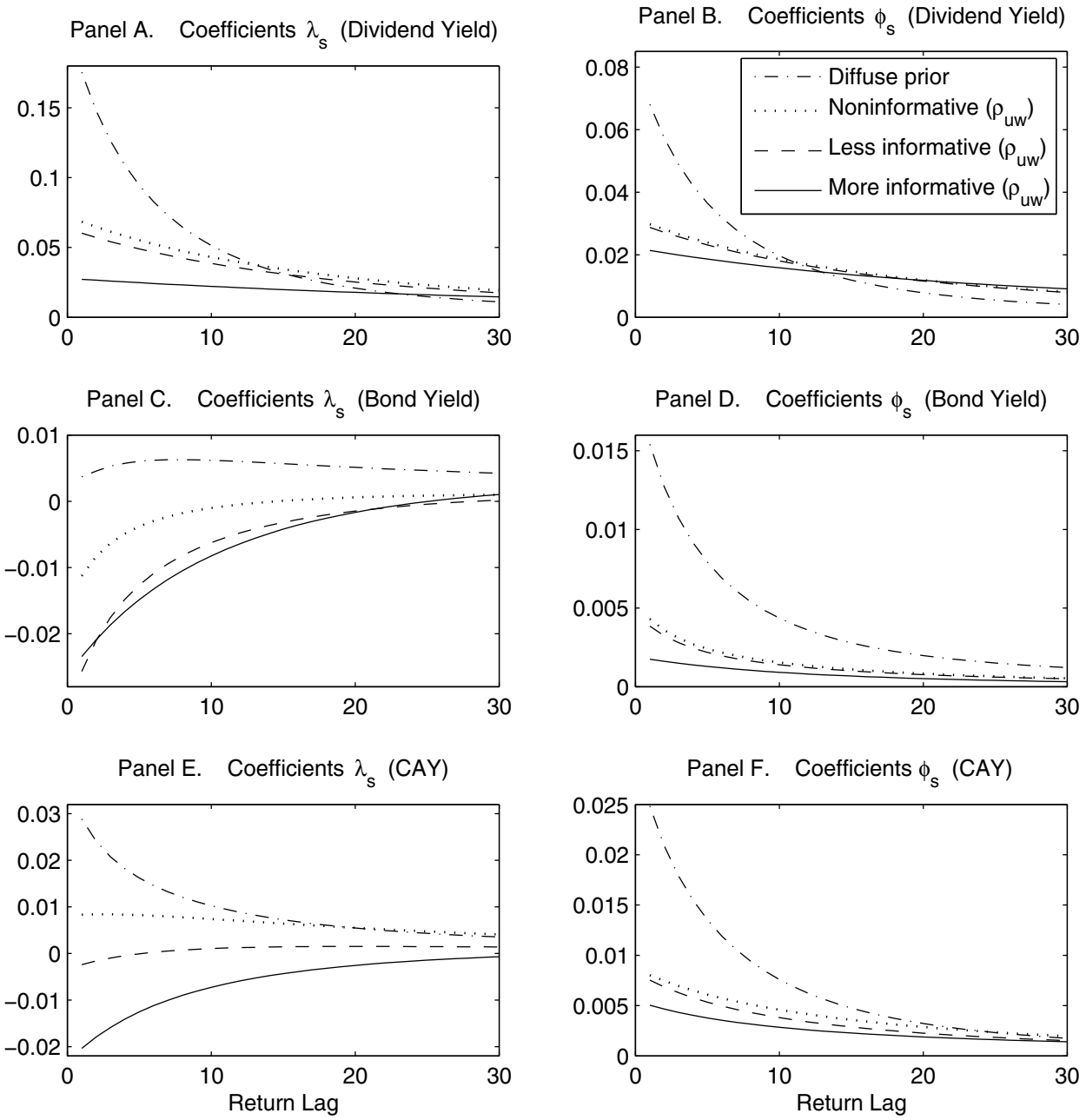


Figure 5. Coefficients on lagged forecast errors and predictor innovations in $\mathrm{E}(r_{t+1}|D_t)$**.** Panels A, C, and E plot steady-state values of λ_s, the coefficients on lagged forecast errors in the expression for the conditional expected return. Panels B, D, and F plot steady-state values of ϕ_s, the coefficients on lagged predictor innovations. Panel headings indicate which predictors are used in the predictive system. The four lines in each panel represent four different prior distributions on ρ_{uw}, the correlation between expected and unexpected returns. 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} ($\rho_{uw} < 0$), the *dotted* line is the "noninformative" prior on ρ_{uw}; and the *dash-dot* line is the "diffuse" prior that is noninformative about all parameters in the model (not only about ρ_{uw}). The sample period is 1952Q1–2003Q4.

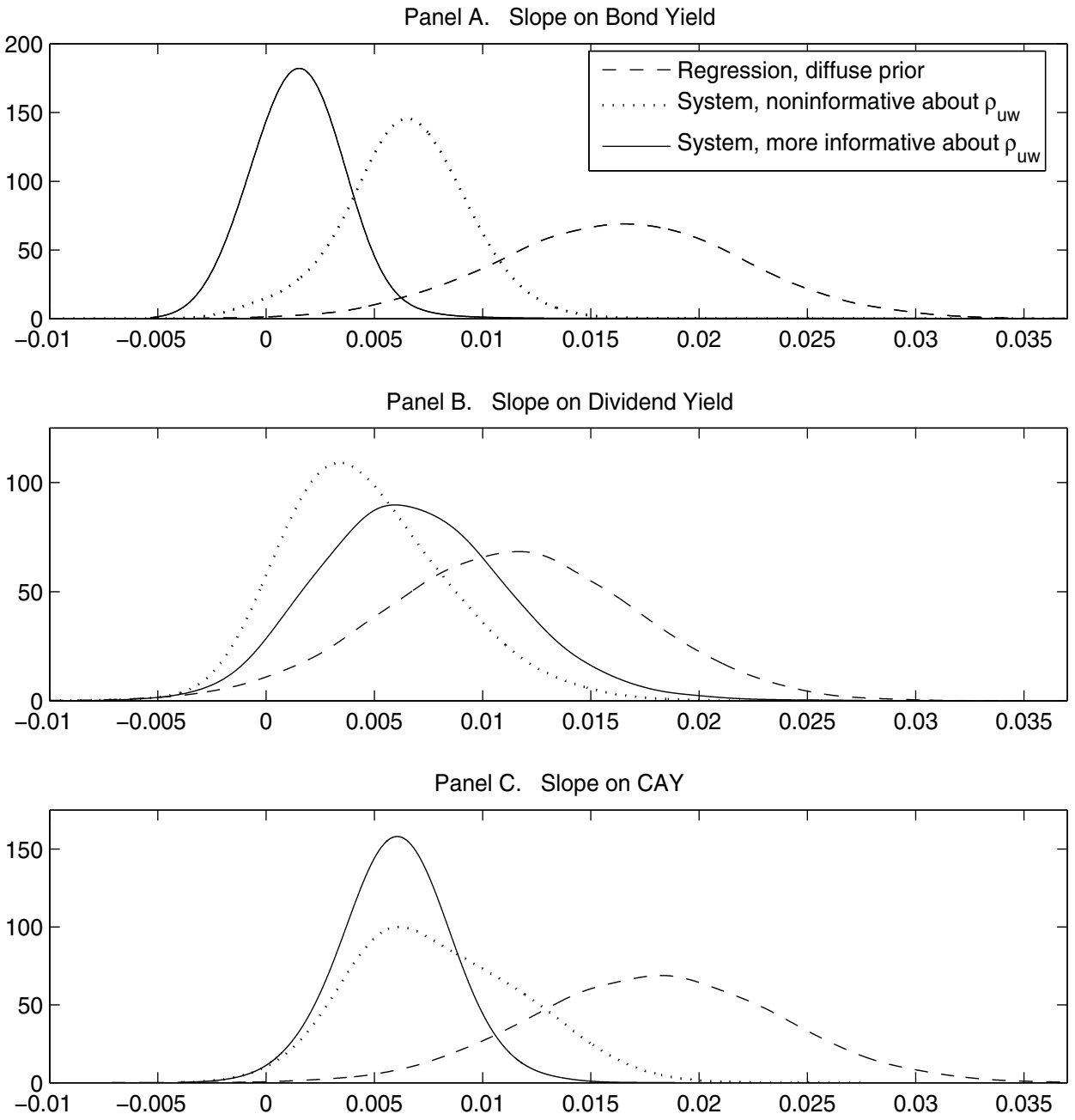


Figure 6. Posterior distributions of slope coefficients from predictive regressions. We estimate both the predictive system and the predictive regression with three predictors: the bond yield, dividend yield, and CAY. The *dashed* line plots the posteriors from the standard predictive regression of r_{t+1} 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.

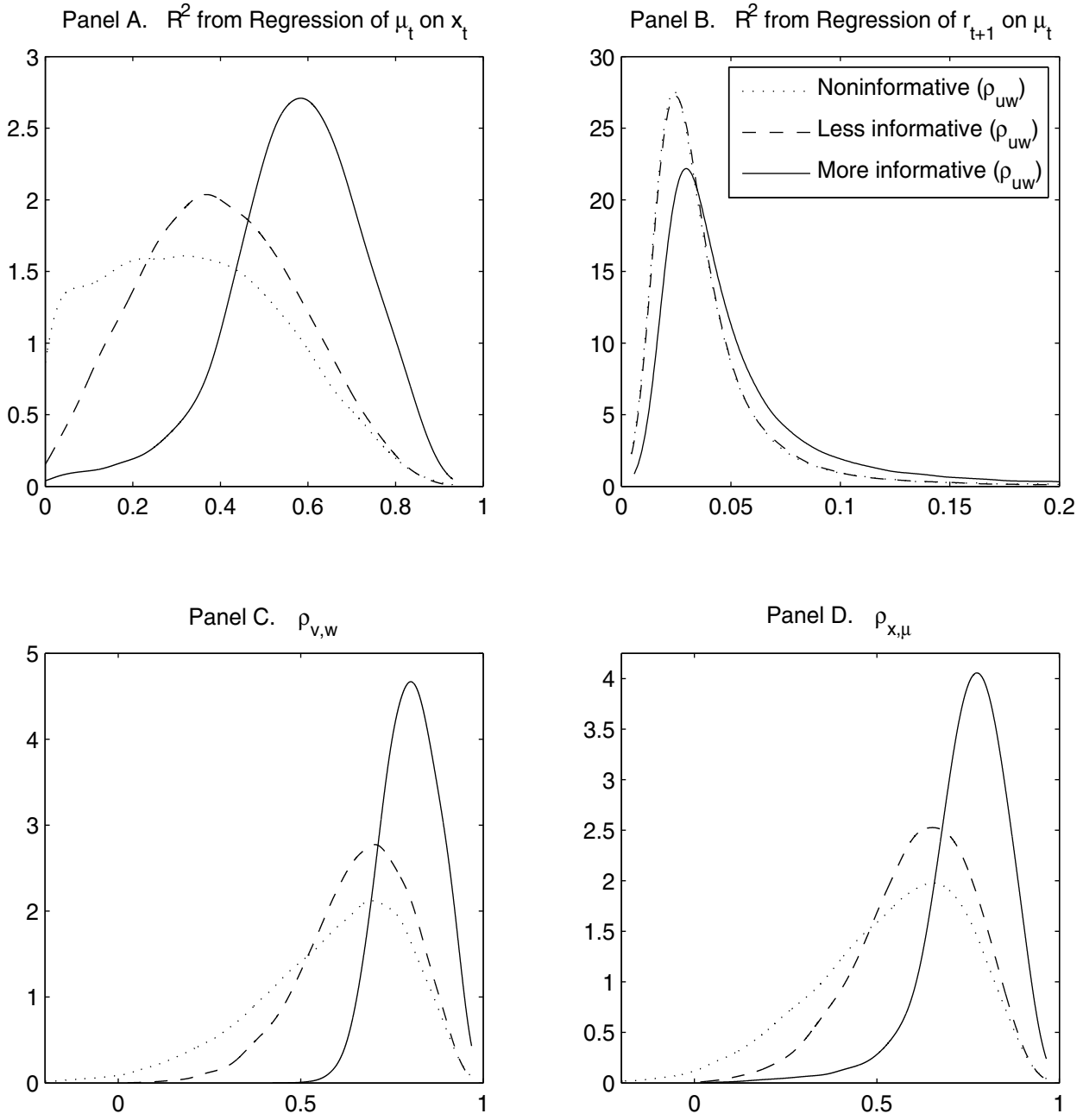


Figure 7. Posterior distributions for one predictor: Dividend yield. The three lines in each panel represent three different prior distributions. The *solid* line represents the "more informative" prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the "less informative" prior on ρ_{uw} ($\rho_{uw}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

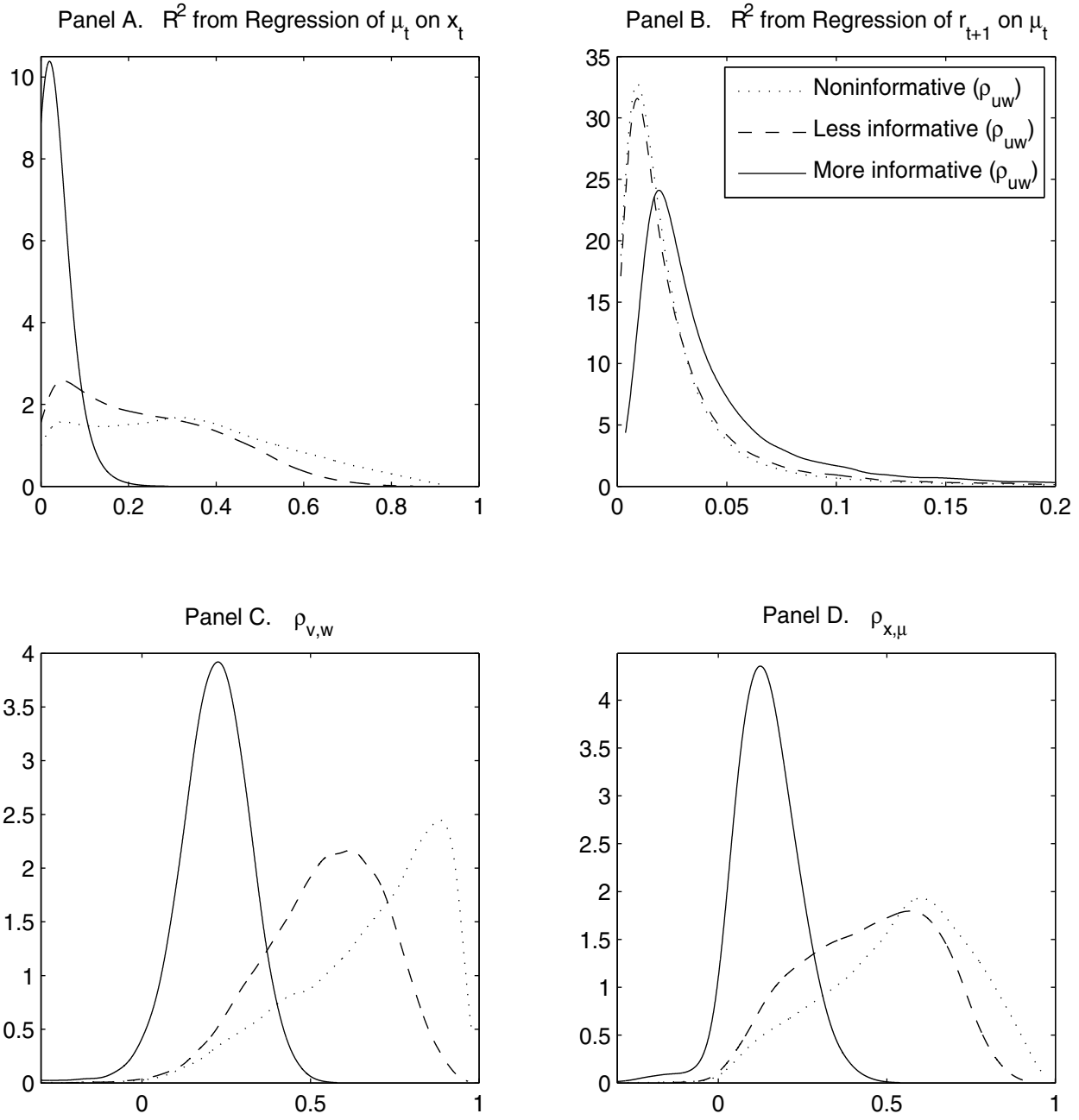


Figure 8. Posterior distributions for one predictor: Bond yield. The three lines in each panel represent three different prior distributions. The *solid* line represents the "more informative" prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the "less informative" prior on ρ_{uw} ($\rho_{uw}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

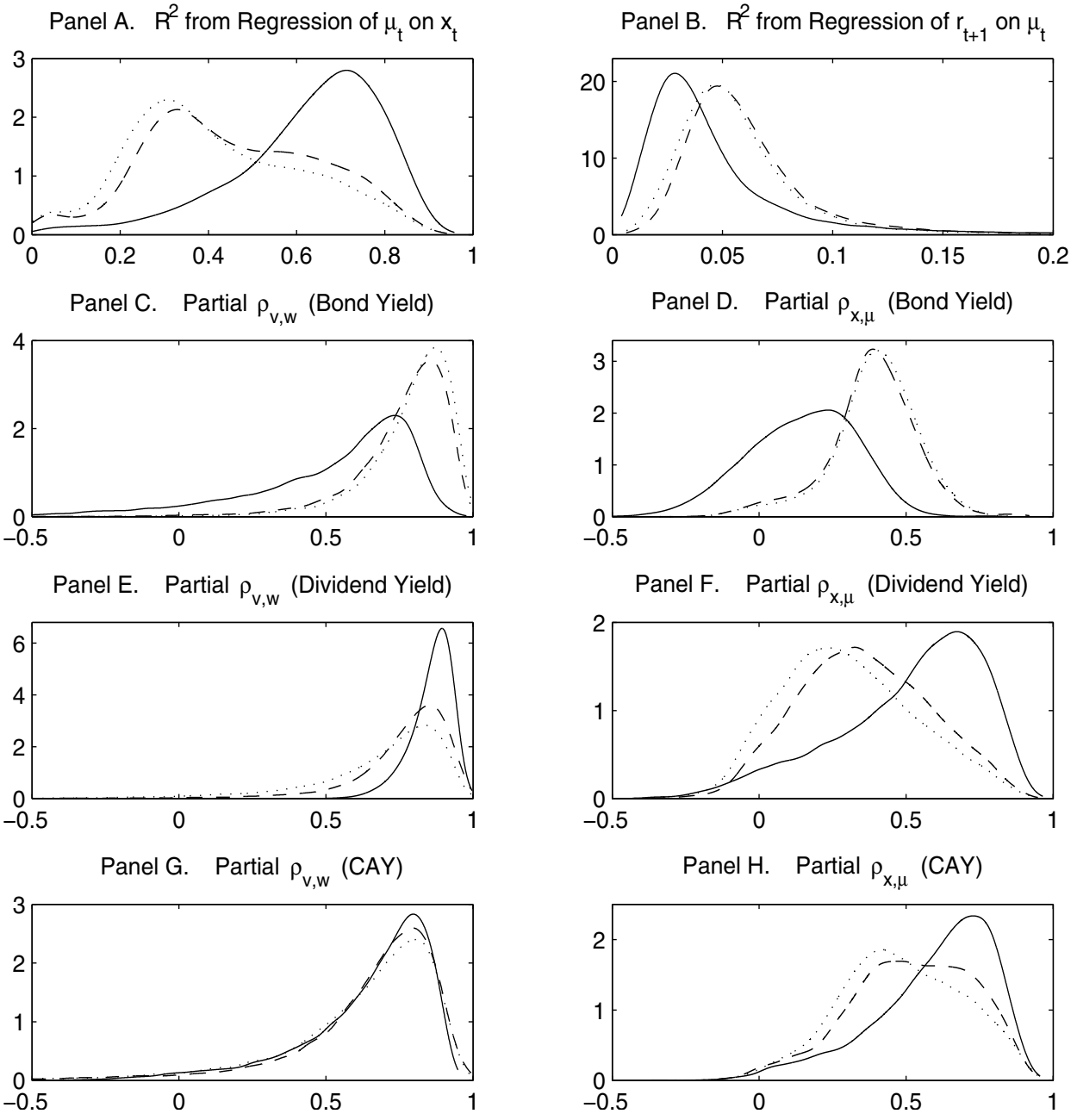


Figure 9. Posterior distributions for three predictors: Bond yield, dividend yield, and CAY. The three lines in each panel represent three different prior distributions. The *solid* line represents the "more informative" prior on ρ_{uw} ($\rho_{uw} < -0.71$), the *dashed* line is the "less informative" prior on ρ_{uw} ($\rho_{uw} < 0$), and the *dotted* line is the "noninformative" prior on ρ_{uw}. Panel A plots the posterior of the fraction of variation in the expected return μ_t that can be explained by the predictors x_t. Panel B plots the posterior of the predictive R^2. Panels C, E, and G plot the posteriors of the *conditional* partial correlation $\rho_{v,w}$ between the given predictor and μ_t. Panels D, F, and H plot the posteriors of the *unconditional* partial correlation $\rho_{x,\mu}$ between the given predictor and μ_t. The sample period is 1952Q1–2003Q4.

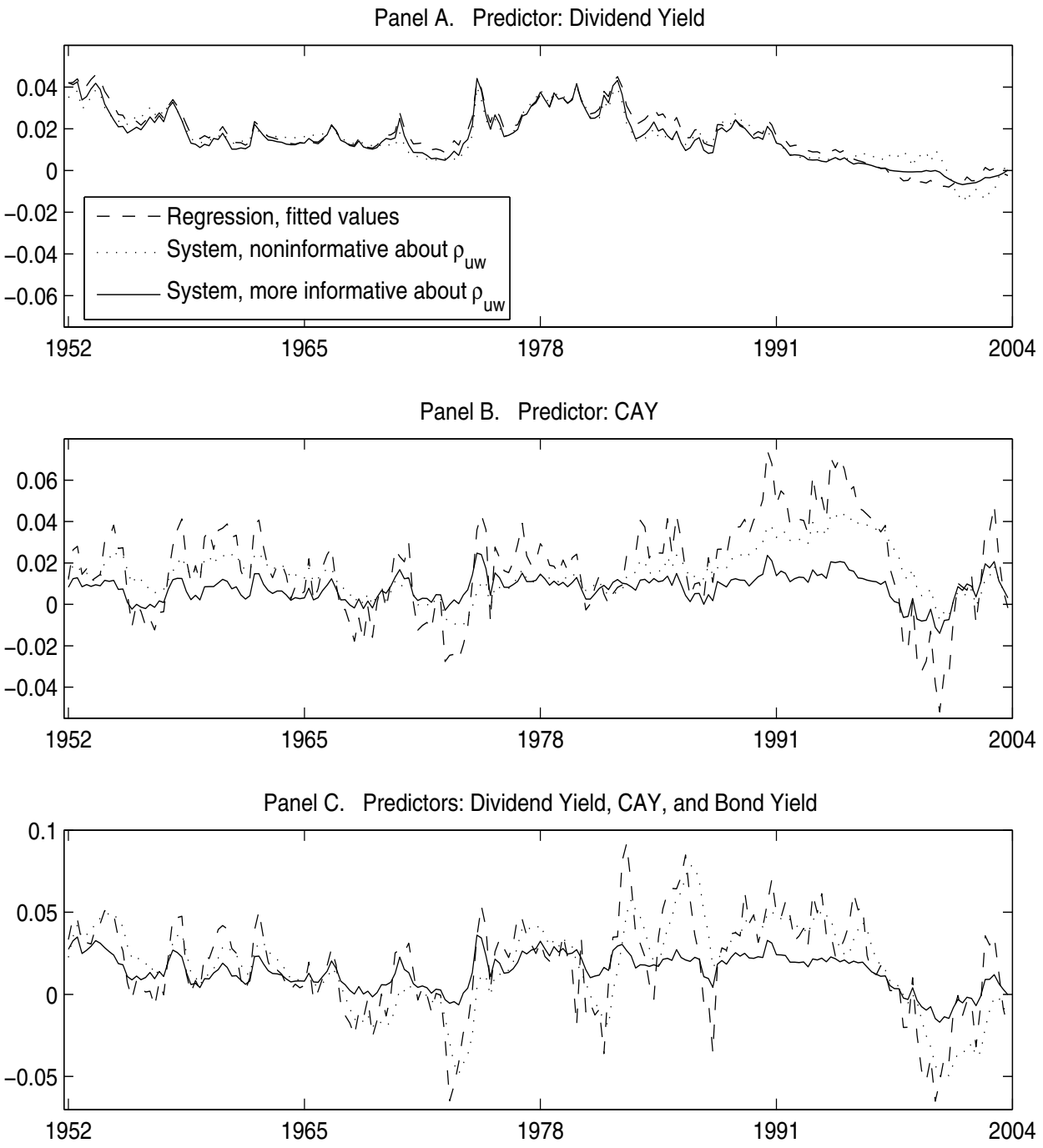


Figure 10. The equity premium: Regression vs. system with prior information. This figure plots the time series of the quarterly equity premium estimated in three different environments. The *dashed* line plots the OLS fitted values from the predictive regression of r_{t+1} on the given predictor(s). The *dotted* line plots the posterior means of $\mathrm{E}(r_{t+1}|D_t)$ from the predictive system under the "noninformative" prior on ρ_{uw}. The *solid* line plots the posterior means of $\mathrm{E}(r_{t+1}|D_t)$ 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.

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 returns 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 R^2 from the predictive regression. 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 coefficient 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^2s are computed by bootstrapping and reported in brackets "[ ]".

Bond Yield	Dividend Yield	CAY	$\text{Corr}(e_t, b'v_t) \times 100$	$R^2 \times 100$
		Panel A. 1952 Q1 – 2003 Q4		
2.716			21.735	4.231
(3.024)			(3.204)	[0.002]
[0.001]			[0.001]	
	1.153		-91.887	2.252
	(2.184)		(-33.506)	[0.059]
	[0.057]		[1.000]	
		1.704	-53.556	7.292
		(4.035)	(-9.124)	[0.000]
		[0.000]	[1.000]	
2.573	1.028	1.346	-35.635	11.777
(2.902)	(1.966)	(3.139)	(-5.487)	[0.000]
[0.003]	[0.058]	[0.003]	[1.000]	
		Panel B. 1952 Q1 – 1977 Q4		
6.385			25.079	7.080
(2.801)			(2.629)	[0.007]
[0.004]			[0.008]	
	2.658		-96.531	7.003
	(2.785)		(-37.522)	[0.015]
	[0.014]		[1.000]	
		3.028	-47.663	15.024
		(4.267)	(-5.503)	[0.000]
		[0.000]	[1.000]	
3.489	1.345	2.129	-53.153	17.975
(1.490)	(1.349)	(2.534)	(-6.369)	[0.000]
[0.090]	[0.177]	[0.012]	[1.000]	
		Panel C. 1978 Q1 – 2003 Q4		
2.073			22.624	3.931
(2.053)			(2.357)	[0.047]
[0.020]			[0.011]	
	0.784		-88.194	1.273
	(1.152)		(-18.989)	[0.423]
	[0.409]		[1.000]	
		1.165	-56.949	4.122
		(2.104)	(-7.031)	[0.045]
		[0.037]	[1.000]	
2.203	0.755	0.968	-18.619	8.828
(2.197)	(1.101)	(1.734)	(-1.923)	[0.053]
[0.023]	[0.313]	[0.118]	[0.967]	

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

This table shows the ratios of two R-squareds, R_1^2/R_2^2. 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_1^2, computed as the R-squared from the regression of the true expected return μ_t on a given predictor, summarizes the usefulness of the predictive regression in estimating μ_t. R_2^2, computed as $1-\mathrm{Var}(\mu_t|D_t)/\mathrm{Var}(\mu_t)$ where D_t contains all historical returns and predictor realizations, summarizes the usefulness of the predictive system in estimating μ_t. The table reports the mean, minimum, and maximum of the possible values of R_1^2/R_2^2 under the model parameters specified. ρ_{uw} is the correlation between expected and unexpected returns, ρ_{vw} is the correlation between the residuals of the AR(1) processes for μ_t and the predictor, β and A are the first-order autocorrelations of μ_t and the predictor, respectively. The values are computed under the assumption that μ_t explains 5% of the variance in realized returns.

	$\rho_{uw}=0$			$\rho_{uw}=-0.473$			$\rho_{uw}=-0.85$			$\rho_{uw}=-0.99$		
ρ_{vw}	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
						Panel A. $\beta=0.9,\ A=0.97$						
0.1	0.034	0.007	0.045	0.524	0.182	0.696	0.026	0.010	0.031	0.009	0.007	0.009
0.2	0.131	0.028	0.181	0.527	0.183	0.696	0.099	0.033	0.125	0.035	0.028	0.037
0.3	0.271	0.063	0.407	0.531	0.184	0.696	0.208	0.064	0.281	0.079	0.063	0.084
0.4	0.404	0.112	0.678	0.538	0.185	0.696	0.329	0.112	0.495	0.139	0.111	0.149
0.5	0.487	0.174	0.696	0.549	0.191	0.696	0.436	0.175	0.659	0.215	0.174	0.233
0.6	0.537	0.251	0.696	0.565	0.253	0.696	0.516	0.251	0.696	0.305	0.251	0.335
0.7	0.574	0.342	0.696	0.588	0.342	0.696	0.573	0.342	0.696	0.406	0.341	0.453
0.8	0.612	0.446	0.696	0.617	0.446	0.696	0.616	0.446	0.696	0.513	0.446	0.571
0.9	0.652	0.564	0.696	0.653	0.564	0.696	0.653	0.564	0.696	0.618	0.564	0.668
						Panel B. $\beta=0.9,\ A=0.9$						
0.1	0.049	0.010	0.065	0.753	0.262	1.000	0.037	0.015	0.045	0.013	0.010	0.013
0.2	0.188	0.040	0.260	0.757	0.263	1.000	0.142	0.048	0.179	0.050	0.040	0.053
0.3	0.389	0.091	0.585	0.763	0.264	1.000	0.298	0.092	0.403	0.113	0.090	0.120
0.4	0.580	0.161	0.974	0.773	0.265	1.000	0.472	0.161	0.711	0.200	0.160	0.214
0.5	0.700	0.251	1.000	0.788	0.275	1.000	0.626	0.251	0.946	0.309	0.250	0.334
0.6	0.771	0.361	1.000	0.812	0.363	1.000	0.741	0.361	1.000	0.438	0.360	0.481
0.7	0.825	0.491	1.000	0.844	0.492	1.000	0.823	0.491	1.000	0.583	0.490	0.651
0.8	0.878	0.640	1.000	0.886	0.641	1.000	0.885	0.640	1.000	0.736	0.640	0.820
0.9	0.936	0.810	1.000	0.938	0.810	1.000	0.938	0.810	1.000	0.888	0.810	0.959
						Panel C. $\beta=0.97,\ A=0.9$						
0.1	0.017	0.007	0.020	0.051	0.022	0.063	0.501	0.273	0.684	0.433	0.371	0.495
0.2	0.066	0.028	0.080	0.182	0.061	0.251	0.603	0.386	0.696	0.605	0.541	0.659
0.3	0.145	0.063	0.181	0.332	0.102	0.531	0.625	0.434	0.696	0.652	0.600	0.689
0.4	0.247	0.112	0.322	0.444	0.139	0.683	0.633	0.460	0.696	0.670	0.628	0.695
0.5	0.360	0.174	0.503	0.515	0.179	0.696	0.638	0.477	0.696	0.679	0.643	0.696
0.6	0.464	0.251	0.660	0.561	0.252	0.696	0.641	0.490	0.696	0.684	0.654	0.696
0.7	0.547	0.341	0.696	0.594	0.342	0.696	0.645	0.499	0.696	0.687	0.661	0.696
0.8	0.607	0.446	0.696	0.623	0.446	0.696	0.650	0.508	0.696	0.689	0.667	0.696
0.9	0.651	0.564	0.696	0.655	0.564	0.696	0.662	0.565	0.696	0.690	0.673	0.696

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

This table shows the posterior means and standard deviations (the latter in parentheses) of the ratios of two R-squareds, R_1^2/R_2^2. 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_1^2, computed as the R-squared from the regression of the true expected return μ_t on the given predictors, summarizes the usefulness of the predictive regression in estimating μ_t. R_2^2, 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. The resuts 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
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)

Table IV
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.

	Correlation (%) \ Mean Abs Diff (%)					Average Utility Loss (%)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
					Panel A. Predictor: Dividend Yield					
(1)		0.55	0.44	0.41	0.29	0	0.15	0.09	0.07	0.03
(2)	*84.55*		0.48	0.45	0.47	0.15	0	0.13	0.11	0.11
(3)	*90.47*	*87.51*		0.06	0.27	0.09	0.13	0	0.00	0.03
(4)	*92.42*	*89.36*	*99.78*		0.22	0.07	0.11	0.00	0	0.02
(5)	*97.67*	*90.47*	*95.81*	*97.41*		0.03	0.11	0.03	0.02	0
					Panel B. Predictor: CAY					
(1)		0.59	1.13	1.22	1.60	0	0.16	0.57	0.65	1.13
(2)	*95.15*		1.23	1.32	1.65	0.16	0	0.72	0.86	1.40
(3)	*86.28*	*87.33*		0.37	0.96	0.57	0.70	0	0.06	0.39
(4)	*96.88*	*95.75*	*95.43*		0.63	0.64	0.83	0.06	0	0.16
(5)	*89.71*	*88.20*	*59.38*	*80.11*		1.09	1.34	0.38	0.16	0
					Panel C. Predictors: Dividend Yield, CAY, Bond Yield					
(1)		1.27	1.33	1.27	1.60	0	0.74	0.82	0.74	1.19
(2)	*84.30*		1.33	1.33	1.80	0.75	0	0.92	0.88	1.49
(3)	*80.38*	*80.68*		0.14	1.51	0.80	0.88	0	0.01	0.94
(4)	*82.30*	*81.75*	*99.79*		1.46	0.72	0.84	0.01	0	0.84
(5)	*83.42*	*79.08*	*80.75*	*84.00*		1.19	1.45	0.96	0.87	0

Table V
Variance Decomposition of Expected Return.

This table reports the posterior means and standard deviations (the latter in parentheses) of the R^2s 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 $u_t, u_{t-1}, u_{t-2}, \ldots$. 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 $u_t, u_{t-1}, u_{t-2}, \ldots$, and by the innovations to the predictors $v_t, v_{t-1}, v_{t-2}, \ldots$. 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	$x_t, \{u_s\}_{s\leq t}$	$x_t, \{u_s, v_s\}_{s\leq t}$
	Panel A. Dividend Yield			Panel B. Bond Yield		
Noninformative Prior on ρ_{uw}	0.34 (0.20)	0.43 (0.21)	0.48 (0.21)	0.33 (0.21)	0.64 (0.21)	0.83 (0.13)
Less Informative Prior on ρ_{uw}	0.40 (0.18)	0.49 (0.18)	0.53 (0.18)	0.24 (0.17)	0.73 (0.16)	0.86 (0.11)
More Informative Prior on ρ_{uw}	0.57 (0.15)	0.80 (0.06)	0.81 (0.06)	0.03 (0.03)	0.86 (0.05)	0.95 (0.04)
	Panel C. CAY			Panel D. All Three Predictors		
Noninformative Prior on ρ_{uw}	0.50 (0.22)	0.59 (0.22)	0.81 (0.12)	0.42 (0.20)	0.49 (0.22)	0.90 (0.07)
Less Informative Prior on ρ_{uw}	0.60 (0.20)	0.70 (0.17)	0.83 (0.10)	0.46 (0.20)	0.55 (0.22)	0.90 (0.07)
More Informative Prior on ρ_{uw}	0.53 (0.18)	0.87 (0.07)	0.92 (0.05)	0.63 (0.17)	0.85 (0.12)	0.94 (0.04)

References

Ang, Andrew, and Geert Bekaert, 2006, Stock return predictability: Is it there?, *Review of Financial Studies*, forthcoming.

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, *The 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.

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, 2005, Predicting the equity premium out of sample: Can anything beat the historical average?, Working paper, Harvard University.

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, 2004, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics*, forthcoming.

Clark, Todd E., and Kenneth D. West, 2005, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics*, forthcoming.

Cochrane, John H., 2006, The dog that did not bark: A defense of return predictability, Working paper, University of Chicago.

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., 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, *Managemente Science* 49, 639–654.

Goyal, Amit and Ivo Welch, 2005, A Comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies*, forthcoming.

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.

Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press, Princeton, NJ).

Jansson, Michael, and Marcelo J. Moreira, 2006, Optimal inference in regressions with nearly integrated regressors, *Econometrica* 74, 681–714.

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, 2005, Stock returns, aggregate earnings surprises, and behavioral finance, *Journal of Financial Economics*, forthcoming.

Lamont, Owen, 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.

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.

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, 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.

Rozeff, Michael S., 1984. Dividend yields are equity risk premiums, *Journal of Portfolio Management* 68–75.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies*, forthcoming.

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).