NBER WORKING PAPER SERIES

RISK, MISPRICING, AND ASSET ALLOCATION: CONDITIONING ON DIVIDEND YIELD

Jay Shanken
Ane Tamayo

Working Paper 8666
http://www.nber.org/papers/w8666

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2001

Thanks to Greg Bauer, Chris Jones, Lance Young, Bill Schwert, Rob Stambaugh, seminar participants at the University of Rochester, the University of Alberta, the University of Pennsylvania, Emory University and, especially, Jon Lewellen for helpful comments and discussions. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Risk, Mispricing, and Asset Allocation: Conditioning on Dividend Yield
Jay Shanken and Ane Tamayo
NBER Working Paper No. 8666
December 2001
JEL No. C11, G11, G12, G14

ABSTRACT

In the asset pricing literature, time-variation in market expected excess return captured by financial ratios like dividend yield is typically viewed as a reflection of either changing risk, related to the business cycle, or irrational mispricing. Extending the work on asset allocation and dividend yield by Kandel and Stambaugh (1996) to accommodate variation in risk as well as expected return, we develop Bayesian methods to examine the interaction between the data and an investor's initial beliefs about the sources of return predictability. Although results vary with the subperiod examined, different views on the relative importance of these factors can have important implications for asset allocation between a stock index and a riskless asset. In general, however, the simple risk/return model of Merton (1980) explains very little of the yield-related return predictability observed.

Jay Shanken
William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, NY 14627
and NBER
(716) 275-4896
shanken@ssb.rochester.edu
http://www.simon.rochester.edu/fac/shanken/shanken.html

Ane Tamayo
London Business School
Regent's Park
London NW1 4SA
United Kingdom
020 7262 5050
atamayo@london.edu

1. Introduction

The evidence on time series return predictability goes back at least to the late 70s (e.g., Fama and Schwert, 1977). Financial and macroeconomic variables such as interest rates, bond yield spreads, dividend yields, etc. seem to have predictive power for returns (e.g., , Rozeff, 1984; Keim and Stambaugh, 1986; Fama and French, 1988, 1989; and Campbell and Shiller, 1988). Although the predictability appears to be related to the business cycle (e.g., Fama and French, 1989), the statistical significance of the evidence has been questioned due to data mining concerns (e.g., Lo and Mackinlay, 1990; Foster, Smith and Whaley, 1997 and Bossaerts and Hillion, 1999). Conventional inference for predictive regressions can be misleading due to small-sample biases (Stambaugh, 1999), as well, raising doubts about earlier results that rely on standard asymptotics. A few studies have, therefore, resorted to simulation analysis and found some, albeit weaker, evidence of statistical significance.[1] On the other hand, Lewellen (2001) has recently developed an alternative approach to assessing statistical significance in predictive regressions and finds strong evidence of predictability for dividend yield and other financial ratios.

Empirical research on time variation in return volatility is also extensive. Return variance is highly persistent, particularly at high frequencies. It is negatively related to return surprises and positively related to leverage as well as the level of interest rates and default spreads.[2] Attanasio (1991) finds a positive dividend-yield effect after controlling for 9 monthly ARCH terms. More relevant to our study, Whitelaw (1994) alludes to unreported evidence of "strongly significant" *simple* relations between default spreads, dividend yields, and stock volatility for the period 1953-89.

Given these empirical findings, several studies have examined the intertemporal relation between the conditional mean and volatility of returns (e.g., Merton, 1980; Campbell, 1987; French, Schwert and Stambaugh, 1987; Attanasio, 1991; Glosten, Jaganathan and Runkle, 1993; Whitelaw, 1994). At the market level, intuition and theoretical models, suggest that conditional expected return and volatility should be

---

[1] E.g. Hodrick (1992), Goetzmann and Jorion (1993), Nelson and Kim (1993), and Kothari and Shanken (1997).

[2] E..g., Campbell (1987), French, Schwert and Stambaugh (1987), Schwert (1989), Glosten, Jaganathan and Runkle (1993), Whitelaw (1994), and Henstchel (1995).

positively related. For example, in Merton's (1980) model conditional expected returns on the market are proportional to conditional volatility. However, a positive relation between variance and expected market return is not a theoretical necessity, even in a rational market with perfect information, as shown in Abel (1988) and Backus and Gregory (1993).

Empirical evidence on the covariation between expected return and risk is mixed. For example, using a GARCH in mean model, French, Schwert and Stambaugh (1987) document a statistically significant positive relation between expected returns and conditional volatility. However, for other models, they find that this relation is not statistically significant. Studies generalizing the GARCH in mean model to allow for leverage effects (Black, 1976, and Christie, 1982) have found that expected returns are negatively related to conditional volatility (e.g. Glosten, Jaganathan and Runkle, 1993). Finally, modeling both conditional expected returns and volatility as functions of several financial variables, Whitelaw (1994) also uncovers a negative relation.

More recently, stock return predictability has been evaluated from a Bayesian perspective that links beliefs about predictability to the asset allocation decision of an individual investor. In their pioneering work, Kandel and Stambaugh (1996), henceforth KS, show that predictive regression results that would typically be dismissed as "statistically insignificant" by conventional standards, can have important implications for static asset allocation between a riskless asset and a stock market index.[3] The stock position becomes more aggressive as the current dividend yield increases, reflecting the higher expected excess return. All of the analysis is conducted under the assumption that market risk is constant over time, however. Multiperiod extensions by Stambaugh (1999) and Barberis (2000) have also assumed constant market risk in examining asset allocation based on dividend yield.[4]

Our paper extends this literature by analyzing the asset allocation decision of an investor who conditions on dividend yield and allows for time variation in both market risk and expected return. As indicated above, the simple relation between yield and

---

[3] Similarly, the limitations of p-value analysis in hypothesis testing are illustrated for the Gibbons, Ross and Shanken (1989) test of portfolio efficiency in Shanken (1987).
[4] The decision-theoretic approach of KS has been used to explore other predictability issues in papers by Bauer (2000), Cremers (2000) and Tamayo (2000) and Avramov (2001).

market risk has not received much attention in past work. Therefore, before moving on to the Bayesian analysis, we provide descriptive evidence on this relation, as well as on expected return predictability.

A common theme in the literature is that yield-related time-variation in expected returns is driven either by changing risk or behavioral "mispricing" (or both). As KS recognize, if market risk increases with yield, the optimal allocation to stock might be less aggressive or even decrease with yield. Similarly, we note that if mispricing is the dominant factor, then the optimum could be close to that obtained assuming constant risk. Clearly, there is a wide range of beliefs about the relative importance of risk and mispricing in the determination of time-varying expected returns. In a Bayesian statistical framework, differences in *prior* beliefs about these issues can be formally incorporated and the impact of historical data on these beliefs analyzed. How, we might ask, would the beliefs and investment decisions of Eugene Fama and Richard Thaler differ (assuming rationality for both) after observing the same data? We can also evaluate the impact of the low dividend yields and high returns of the late 1990s on beliefs about predictability.

To address these issues, we employ a simple model that decomposes expected return variation into a component related to risk and and a residual that we refer to as the "mispricing" component. Naturally, mispricing can only be defined in relation to a benchmark that describes how expected returns "should" vary over time. Following Merton (1980), the hypothesis imposed here as a starting point for such analysis, is that the market risk premium is, *apart from mispricing*, proportional to the conditional variance of returns. The need for a joint hypothesis of this sort can be likened, in some respects, to Fama's (1970) well-known dictum that tests of market efficiency always entail a joint hypothesis about the equilibrium pricing process.

The Merton condition is obviously a strong joint hypothesis for the purpose of making positive economic assertions, in that it ignores market frictions and the potential impact of hedging demands on expected returns in an intertemporal equilibrium. The extent to which empirical deviations from this condition are interpreted as evidence of mispricing, rather than deviations from the simple risk-return relation due to misspecification, will depend on one's judgment as to the likely magnitude of these other effects and their connection to dividend yield. From a normative perspective, however, the proportionality assumption for expected return and variance can be viewed as a pragmatic device for formulating prior beliefs about predictability parameters in the asset allocation decision. Clearly, it would be of interest to consider more sophisticated risk models in future work.

We emphasize that our paper has little to say about the moments of market return conditioned on *all* available information. In particular, short-run autoregressive dynamics for risk are not modeled. Although the Bayesian simulation techniques that we have adapted to the present problem should be useful in addressing broader predictability issues as well, our goal in this paper is to provide further insight into predictability based on dividend yield. In particular, we i) assess the extent to which the yield-related expected return variation documented in many past studies can be attributed to a tendency for market risk to be high when *yield* is high and, ii) explore the implications of these observations for an asset allocation decision that is conditioned on the current level of dividend yield.

Methodologically, in addition to allowing for variation in risk as well as expected return, we explore the implications of informative prior beliefs in more detail than past studies of this sort.[5] We also provide sufficient conditions under which Bayesian analysis of the predictive return regression reduces to the simpler nonstochastic regressor case. Unlike KS, we analyze the posterior distribution of the model parameters as well as the predictive distribution of returns (see related work by Avramov (2001)). These features are accommodated through the application of *importance sampling* simulation techniques that, to our knowledge, have not previously been used in the predictability literature. Another novel feature of the study is our comparison of predictive return distributions and asset allocations based on prior and posterior beliefs, which allows us to identify the incremental significance of the data in the presence of informative priors.

The paper is organized as follows. First, descriptive statistics on dividend yield and returns are presented in Section 2. Our regression specification with changing risk is

---

[5] A recent paper by Pastor and Stambaugh (2001) allows for changes in volatility across different regimes, rather than as a function of observable state variables like dividend yield. They also work with informative priors that incorporate a belief about the proportional relation between the market risk premium and variance, but do not explore asset allocation issues.

described in Section 3 and Generalized Method of Moments estimates are reported. Section 4 discusses the specification of prior distributions for our Bayesian analysis and Section 5 provides an overview of the Bayesian methodology. The main results are presented in Section 6 for the 1940-99 period and in Section 7 for the more recent data since 1960. Section 8 summarizes our findings and considers directions for future research.

2. Descriptive statistics

We begin our analysis by looking at the means and standard deviations of continuously compounded excess returns and dividend yields over the period 1926-99 and various subperiods. The value-weighted NYSE return is obtained from the CRSP data file and the NYSE dividend yield is backed out as a function of the returns with and without distributions. As in Fama and French (1988, 1989), the dividend yield is computed as the sum of the dividends paid over the prior twelve months divided by the price at the beginning of the forecasting period.

The statistics for the NYSE value-weighted excess returns and dividend yields are provided in Table 1. Figures 1a and 1b plot the excess returns and squared excess (deviation from mean) returns over the entire period. As has often been noted, the period before 1940 was extremely volatile and perhaps could be viewed as a different "regime." In addition, the average excess return of 17 bps per month for the 1927-1939 period was far below the mean of 54 bps for the full sample. Schwert (1990), in studying aggregate stock returns from 1802 to 1987, emphasizes that, apart from this unusual period, the properties of stock returns have been remarkably homogeneous. For example, monthly volatility in consecutive 20-year periods from 1841 to 1920 ranged from 4.14% to 4.79%. Mean (total) returns were more variable in the 20-year periods, ranging from 44 bps per month to 102 bps, with an average of 64 bps. The sample moments for the 1940-99 period lie within these ranges.

Our empirical analysis will focus on the relatively stationary period since 1940, with emphasis on the data since 1960. We consider the 1926-39 data and the earlier data of Schwert when specifying prior beliefs, however. If an investor has doubts about the constancy of parameters in the model, the more recent data might be viewed as more relevant. Recently, Goyal and Welsch (1999) have argued that the predictive power of dividend yield has declined over time, as has the level of yield itself, which is apparent in Table 1. This provides additional motivation for our focus on the more recent period. A formal modeling of changing parameters, while of considerable interest, is beyond the scope of this paper.

It is well known that the historically low yields of the late 1990's were accompanied by extraordinarily high returns, thus weakening the yield/return relation. It is interesting to look at the shift in beliefs that resulted from the last five years of data and so we examine the 1960-94 period as well as the 1960-99. As we will see, this provides a broad range of scenarios in which to study predictability in market risk and expected return.

3. The regression model for returns

In this section, we introduce the regression model with heteroskedasticity and present generalized method of moment (GMM) estimates of the expected return and risk parameters. We then reformulate the model in a manner that splits yield-related expected return variation into a component proportional to market risk and a residual "mispricing" component. Estimates of this model are presented as well. Bayesian estimation results are presented in the following sections.

3.1. The basic regression model

We begin by considering empirical results for the standard linear predictive regression:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the difference between the continuously compounded returns realized at t+1 on the stock index and a one-month riskless T-bill, x_t is the dividend yield at time t, and ε_{t+1} is a homoskedastic disturbance, which is normally distributed independently over time with mean zero and standard deviation σ_ε. The slope parameter β measures the extent to which expected return varies with dividend yield. Henceforth, "return" always refers to continuously compounded excess return.

Since dividend yield is always positive, it is hard to develop intuition about expected return conditional on $x_t = 0$. This is important when specifying informative prior beliefs or, more generally, in interpreting statistical estimates of α. If we measure x_t as the deviation of yield from some fixed level, then α can be viewed as the expected return when yield is at this level. In our empirical analysis, we work with deviations from the sample mean.

Columns 2-3 of Table 2 contain OLS estimates of α and β for various subperiods.[6] Given the small-sample problems documented by Stambaugh (1999), however, we view the numbers as merely suggestive. The estimates of α (mean excess returns) are all positive and, except for 1960-94, are more than two standard errors above zero. The beta estimates are also positive, but only the 1940-99 and 1960-94 estimates are more than 1.5 standard errors from zero. As Lewellen (2001) notes, inclusion of the low-yield, high-return years of the late 90's drastically lowers the slope, cutting the value by nearly two-thirds in our sample.[7]

3.2. The model with changing market risk and a risk/mispricing decomposition

In addition to time variation in expected returns, we allow for variation in risk associated with changes in dividend yield. The conditional standard deviation at time t is modeled as

$$\sigma_{\varepsilon t} = \text{stdev}(\varepsilon_{t+1} \mid x_t) = c \cdot \exp(\lambda x_t) . \tag{2}$$

This exponential specification ensures that the standard deviation is always positive and amounts to assuming that $\log(\sigma_{\varepsilon t})$ is linear in x_t.[8] Given that yield is measured as a deviation from the mean, c is the standard deviation of return at the average level of yield. Yield-related variation in risk depends on the parameter λ. For example, $\lambda = 10$ means that a one percentage point increase in yield is associated with an increase in

6 Anticipating the results below, heteroskedasticity-consistent standard errors are reported in parentheses. The usual OLS standard errors are not very different except for the 1927-99 period in which the OLS standard error is about 40% lower.

7 Using an innovative approach that conditions on the value of the dividend yield autocorrelation estimate, he finds, surprisingly, that statistical rejections of a zero slope are similar and sometimes stronger, depending on the index.

8 Previous linear specifications for squared residuals (Shanken, 1990 and Campbell, 1987?) or absolute residuals (Schwert, 1989) do not impose positivity.

standard deviation approximately equal to 10% of c. If $\lambda = 0$, the error term is homoskedastic.

The moment condition that we use later to estimate c and λ is based on the following nonlinear regression

$$\varepsilon_{t+1}^{2} = c^{2}\exp(2\lambda x_{t}) + u_{t+1}, \tag{3}$$

where ε_{t+1} is the disturbance in (1). Assuming $E(u_{t+1} \mid x_{t}) = 0$, we have

$$E(\varepsilon_{t+1}^{2} \mid x_{t}) = c^{2} \cdot \exp(2\lambda x_{t}) = \sigma_{\varepsilon t}^{2}.$$

Now consider the impact of changing risk on the market risk premium. As in Merton (1980) and French, Schwert, and Stambaugh (1987), we allow the market risk premium to vary directly with the ex ante variance. Unlike those studies, our risk measure is conditioned on the current level of dividend yield, the variable of interest in this paper. To identify the component of return variation due to risk, we make the strong assumption that the "price of risk," i.e., the ratio of expected return to variance, is constant over time. Given the considerable interest in the sources of return predictability and the paucity of direct evidence on this question, analysis with this basic measure will serve as a useful starting point for thinking about these issues.[9]

Let γ denote the long-run (continuously compounded) Sharpe ratio, defined as the ratio of expected return to standard deviation when yield is at its mean (x=0). Assuming that $E(r_{t+1} \mid x_t) = k\sigma_{\varepsilon t}^2$, it follows that $kc^2 = \gamma c$ or $k = \gamma/c$. Thus, we consider the equation

$$\begin{aligned} r_{t+1} &= (\gamma/c)\sigma_{\varepsilon t}^{2} + \beta_m x_t + \varepsilon_{t+1} \\ &= \gamma c \cdot \exp(2\lambda x_t) + \beta_m x_t + \varepsilon_{t+1}. \end{aligned} \tag{4}$$

where $\sigma_{\varepsilon t}$ is given in (2) and β_m reflects the remaining yield-related variation in expected return. When yield is at its mean, $\sigma_{\varepsilon t} = c$ and the expected return given by (4) is just γc, which plays the role of α in (1). The subscript m indicates that, under the assumptions of

[9] In a similar spirit, Attanasio (1991) considers the relation between several predictive variables and expected return after controlling for the effect on return volatility.

the model, the remaining component would be attributable to "mispricing," an alternative that is often considered in the literature.

By (4), if there is no mispricing (β_m=0) then the Sharpe ratio equals $(\gamma/c)\sigma_{\varepsilon t}$. For $\gamma > 0$, the ratio increases with the conditional standard deviation and equals γ when x = 0. There will be additional variation in the Sharpe ratio that is unrelated to risk if β_m is nonzero. For example, a positive mispricing beta implies that expected return goes up when yield increases, beyond what can be attributed to the simultaneous increase (if $\lambda > 0$) in risk. Thus, the Sharpe ratio increases as well. Harvey (1989) rejects the hypothesis that the ratio of market risk premium to variance (or standard deviation) is, in fact, constant. Our model permits the ratio to vary, but attributes all such variation to mispricing effects.

One can, of course, take a more agnostic view of (4). It simply decomposes the total expected return variation captured by yield, in equation (1), into a component that is proportional to total risk and a residual component that is left over. Breaking down predictability in this manner and thinking in terms of implied Sharpe ratios aids in the formulation of prior beliefs about predictability. MacKinlay (1995) makes a similar appeal to the plausibility of estimated Sharpe ratios in assessing the rationality of cross-sectional size and book-to-market expected return effects.[10]

Note that if we linearize the exponential term, (4) can be written as

$$r_{t+1} \approx \gamma c(1 + 2\lambda x_t) + \beta_m x_t + \varepsilon_{t+1} \quad (5a)$$

$$= \gamma c + (\beta_m + 2\gamma c\lambda)x_t + \varepsilon_{t+1}. \quad (5b)$$

The mispricing and risk components in (5a) are, naturally, perfectly correlated, both being functions of dividend yield. In conjunction with the variance relation, the parameters are, nonetheless, identified since there is no separate constant in that representation.

It is apparent from (5b) that, apart from the nonlinearity, (4) is simply a reparametrization of (1) and (2) with intercept and slope

$$\alpha = \gamma c \text{ and } \beta = \beta_m + 2\gamma c\lambda, \quad (6)$$

[10] See related discussion of Sharpe ratios and approximate arbitrage in Shanken (1992).

respectively. The risk component of β is $2\gamma c\lambda$ and the mispricing component is β_m.[11] This relation does not entail any restrictions on α or β. Rather, given the risk parameters c and λ, the long-run Sharpe ratio γ and the mispricing slope β_m are essentially defined by (6).

Of course, if the Merton proportionality condition is misspecified, then the interpretation of β_m will be affected. For example, if expected return is *linear* in market variance, with intercept k_0, then a little algebra shows that $\beta_m \approx -2k_0\lambda$ in the absence of other "mispricing" effects. If, say, $\lambda = 10$ and expected return is 10 bps per month when risk is zero, then β_m is "biased" downward by 0.02, a relatively small amount in the context of our analysis below. Although prior beliefs about this effect could be incorporated as well, we opt for simplicity and assume that k_0 is zero in our discussion of results.

3.3 GMM estimation of the risk/mispricing model

Before going on to the Bayesian analysis, we estimate (3) and (4) as a system using the generalized method of moments (Hansen, 1982).[12] Simultaneous GMM estimation of the system accounts for the fact that the disturbances in (1) must be estimated and, hence, provides asymptotically correct heteroskedasticity-consistent standard errors. The system is likely to be subject to small-sample problems similar to those analyzed by Stambaugh (1999), however. The moment conditions for estimation are

$$E\begin{bmatrix} \left(r_{t+1} - \gamma c \exp(2\lambda x_t) - \beta_m x_t\right) \\ \left(r_{t+1} - \gamma c \exp(2\lambda x_t) - \beta_m x_t\right)x_t \\ \left(\varepsilon_{t+1}{}^2 - c^2 \exp(2\lambda x_t)\right) \\ \left(\varepsilon_{t+1}{}^2 - c^2 \exp(2\lambda x_t)\right)x_t \end{bmatrix} = 0. \qquad (7)$$

[11] If the risk-related component of expected return is assumed to be a constant multiple of the conditional standard deviation, $2\gamma c\lambda$ is replaced by $\gamma c\lambda$ in (6) and $\exp(2x\lambda)$ by $\exp(x\lambda)$ in (4). Thus, the variation in expected return attributed to risk is lower under this assumption. Clearly, other specifications of the dynamic behavior of the Sharpe ratio could also be explored.

[12] We've also estimated the system using a two-step weighted least squares procedure with similar results.

Estimates of the risk parameters c and λ are given in columns 4 and 5 of Table 2. The average level of risk is near 4.2% for each subperiod since 1940 and, consistent with observations in Merton (1980), is estimated with much greater precision than the other parameters. The estimate of c is 5% when the data before 1940 are included. All estimates of λ are positive, several standard errors from zero in the full period, and close to two standard errors or more in the post-1960 periods.

Thus, based on the evidence discussed so far, we have one period (1960-94) in which estimates of the yield-related risk and expected return parameters are both large, economically, and others in which only the changes in risk (1960-99, 1927-99), or the changes in expected return (1940-99) are estimated to be considerable. This provides a convenient "laboratory" in which to study the effect of time-varying risk on expected returns.

Estimates of γ and β_m are given in columns 6 and 7 of Table 2. Since $\gamma \approx \alpha/c$, it is not surprising that the estimated prices of risk are all positive with "significance" paralleling that of α. The estimates of β_m are positive, except for the 1927-99 period, and are more than 1.5 standard errors from zero for 1940-99 and 1960-94. The estimates of β and β_m are close in the 1940-99 period since the estimate of λ, the parameter for changing risk, is small. Thus, essentially all of the evidence for yield-related expected return variation comes from the mispricing component in this period. A similar conclusion applies for the 1960-94 period in which the estimates of β and β_m are 0.56 and 0.46, respectively. In this case, the difference of 0.10 is larger due to the higher value of λ, though dampened by the relatively low estimated Sharpe ratio, γ. The strongest evidence of changing risk is found in the full period 1927-99, though the process does not appear to be stationary over this period.[13]

The last column of Table 2 presents estimates of the approximate change in the Sharpe ratio for a one-percentage point change in yield. More precisely, δ is the *average* change in Sharpe ratio as yield increases from one point below to one point above its mean and is given by

---

[13] The standard error for λ is even smaller when adjusted for residual autocorrelation.

$$\delta = \frac{\beta_m}{2c}[\exp(-0.01\lambda) + \exp(0.01\lambda)] + \gamma[\exp(0.01\lambda) - \exp(-0.01\lambda)]/0.02. \qquad (8)$$

A δ of 5, for example, means that the ratio increases by .05 when yield increases by 100 bps. Since the estimates of γ range from 0.07 to 0.15, this represents a substantial increase relative to the long-run ratio. Given the link between mispricing and the Sharpe ratio discussed earlier, it is not surprising that the estimates and standard errors for β_m and δ are highly correlated across subperiods. For example, δ equals 12.1 for the 1960-1994 period in which β_m attains its highest value.

4. Prior beliefs about the risk/return parameters

Having looked at the coventional regression evidence, we now examine the data from a Bayesian perspective. We ask how the evidence would affect individuals with a range prior beliefs. Thus, understanding the *mapping* between prior and posterior beliefs is a central theme of the analysis. Our "balanced priors" are continuous and embrace the possibility that mispricing and risk effects are present simultaneously. In contrast, "skeptical" versions of these priors entertain the possibility that there is no mispricing effect ($\beta_m = 0$) or that conditional risk does not vary with yield ($\lambda = 0$), each with positive prior probability. The skeptical priors accommodate initial beliefs that are oriented more toward the classical risk paradigm or beliefs that give behavioral biases more credibility than changing risk.

The next subsection is an overview of factors that affect the relation between yield and expected return. A detailed description of the intuitive economic motivation behind the choice of parameter values for our balanced priors is then given in Sections 4.2 and 4.3. Some of these choices will inevitably seem a bit arbitrary, but we think they provide reasonable points of departure for our analysis.[14]

4.1. Dividend yield and expected return

Before discussing the prior distributions, it will be helpful to impose a little formal structure. Let g* be the expected growth rate of aggregate dividends and,

[14] Readers wishing to provide a description of their personal priors can e-mail us. Anonymity will be protected.

allowing for deviations from pure rational expectations, let investors' forecasted growth g = g* + m, where m is a stationary mean-zero mispricing component. By the constant growth formula,[15]

$$D/P = r - g^* - m, \quad (9)$$

where r is the rate "used" by investors to discount expected future cash flows. This discount rate may vary over time. We assume that higher risk implies a larger risk premium. Other things equal, more risk and a higher r lower the price and increase the yield. On the other hand, correctly perceived shifts in expected dividend growth g* should be rationally impounded in prices and yields, but need not affect expected returns, holding the other components of yield constant.

A positive value of m reflects excessive optimism. This increases price and, holding D constant, depresses the dividend yield. The actual expected return is lower than r, given the likely "correction" of mispricing implied by our mean-reversion assumption. Similarly, if m is negative then yield and expected return both increase with D constant.

The discussion above summarizes the *partial* effects of m, r, and g* on yield and expected return. Since predictive regressions like (1) and (4) do not control for expected growth, correlation between g* and the other components of yield must be considered.[16] First, regarding m, if investors extrapolate recent growth too far into the future (Lakonishok, Shleifer, and Vishny, 1994), the market will tend to be underpriced when expected g* is low. Thus, g* and m are positively correlated in this sort of overreaction scenario. In this case, both yield and expected return are high and the yield/return relation is positive. A positive relation can still be observed even if mispricing is unrelated to expected growth, though a weaker relation would be expected without the mutual reinforcement of both factors.

Alternatively, in an underreaction scenario, g* and m could be negatively correlated. The market might be underpriced when economic conditions are good, with solid growth expected in the future. If g* is the dominant component of yield, a low

---

[15] The more general framework of Campbell and Shiller (1988) could be used, but the constant growth model suffices for our purposes.
[16] We are still discussing partial effects of m and r, but now without g* held constant.

yield (high g*) is associated with higher expected return (negative m*). The recent behavioral literature notwithstanding, the early paper by DeBondt and Thaler (1985) viewed overreaction as the fundamental behavioral hypothesis emerging from the work of Kahneman and Tversky. "Mispricing" effects need not be entirely irrational, however, and can arise through a rational process of learning about the economy in a world of parameter uncertainty (Lewellen and Shanken, 2001).[17] We will consider a prior that is strongly oriented toward overreaction (OP) as well as a prior that is more neutral (NP) about the direction of the mispricing effect.

Now consider the impact of g* on the relation between r and dividend yield. Suppose there is a decrease in expected future dividend growth. This will lower price and increase yield as well as leverage. The higher yield is associated with increases in both expected return and risk induced by the leverage effect (Black, 1976 and Christie, 1982). This reinforces the positive partial relation between yield and movements in r. In the absence of a compelling story to the contrary, we focus on priors that emphasize positive relations between risk, expected return, and dividend yield, apart from mispricing.

4.2. Priors for the long-run levels of risk and the Sharpe ratio, c and γ

We start by specifying a prior for the standard deviation c.[18] This prior distribution is taken to be lognormal to ensure positivity. Based on the Schwert (1990) data and descriptive statistics above, we let $\mu_c = 4.5\%$ and $\sigma_c = 50$ bps as of the beginning of 1960 and $\mu_c = 6\%$, $\sigma_c = 100$ bps prior to 1940. Thus, beliefs about risk are quite informative, in keeping with the historical homogeneity discussed earlier. The 1940 investor is characterized as expecting somewhat greater economic stability than was experienced in the 30s, but with considerably more uncertainty than that of the pre-1927 period.

---

[17] If the researcher's model mimics the learning process of *investors*, the implied predictability will not show up in the predictive distribution but could be observed in conventional regression analysis.

[18] Given the specification of (1) in terms of deviations from the mean yield, our priors are conditioned on this level of yield. Allowing certain sample characteristics to enter into the prior is often referred to as an "empirical Bayes" approach. A more sophisticated approach might incorporate perceived trends in yield related, for example, to the increasing use by firms of stock repurchases.

Now we turn to the Sharpe ratio γ. The prior is taken to be normal, independent of c. Therefore, the prior mean for long-run expected return α is simply $E(\gamma)E(c)$. Over the 1940-99 period, the average excess return on the NYSE value-weighted stock index is 62 bps per month with standard deviation 4.2%, similar to the sample moments for 1841-1920. The corresponding Sharpe ratio is about 0.15. In the spirit of Jorion and Goetzmann (1999), who argue that ex post performance in the U.S. market has likely exceeded ex ante expectations due to "survivor biases," we specify a prior mean of μ_γ = 1/9, so that μ_α = 50 bps when μ_c = 4.5% and 67 bps when μ_c = 6%. Since μ_γ is positive, the conditional mean of α increases with c.

Although the variance of returns can be estimated fairly precisely with relatively little data, this is not true of the mean. For example, with a standard deviation of 4.5% and i.i.d. returns over 40 years, the standard error of the mean monthly return is still about 20 bps. Therefore, we need to allow for relatively greater uncertainty about the Sharpe ratio, as compared to c. Although a prior standard deviation for γ as low as 0.03 might be justified based on past data, we let $\sigma_\gamma = 0.05$ ($\mu_\gamma = 1/9$), reflecting uncertainty about the influence of survivorship. This is still low enough to reflect a strong belief that γ and the corresponding risk premium are well above zero.

A few comments on the long-run impact of mispricing before we turn to the other model parameters. While irrational mispricing can induce variation in expected returns, if the mispricing reverts to a mean of zero, it should not have a direct effect on the long-run level of expected return. Uncertainty about *future* mispricing can affect risk, however, and thus have an indirect impact on expected return. For example, De Long, Shleifer, Summers and Waldmann (1990) analyze a model in which the random perceptions of irrational investors induce additional return variability and an associated price discount (higher expected return), as both rational and irrational investors demand compensation for the additional "sentiment risk." This would be reflected in the parameters γ and c of our model. Thus, risk-related variation in expected return need not be unambiguously rational, as it can arise from the interplay between rational and irrational investors.

4.3. Priors for the time-varying risk/return parameters, λ and β_m

The prior for λ is assumed to be normal, independent of c and γ. As discussed in section 4.1, a positive relation between risk and yield is reasonable a priori. Therefore, λ is expected to be positive. To a first-order approximation, $\sigma_\varepsilon \approx c(1 + \lambda x)$. As the specification of a prior is far from an exact science, we will be content to rely on this approximation to guide our choice of "reasonable" parameter values. In particular, if λ= 20, a 100 bps change in yield will be associated with a proportionate change in σ_ε of about 20% [20% of μ_c = (0.20)(0.045) = 90 bps]. By (6), with γ = .111, this translates into an expected return shift of about 20 bps per month. We let $\mu_\lambda = 20$ and $\sigma_\lambda = 15$ in our "balanced priors," reflecting a strong belief in a positive risk relation ($\lambda = 0$ is a two standard deviation "event"). Later, we will see that the data tend to dominate beliefs about λ with this level of prior uncertainty.

The slope coefficient on yield in (4), β_m, reflects expected return variation due to mispricing. As discussed in section 4.1, β_m could be negative or positive depending on the nature of the mispricing. In either case, if β_m is sufficiently large in magnitude, implied expected returns can be negative at times, an implausible conclusion from the risk-based perspective. Kothari and Shanken (1997) exploit this observation in their interpretation of predictability results. Risk and mispricing effects need not be mutually exclusive, though one might argue for some relation between them. For simplicity, the priors for β_m and λ are taken to be independent.[19]

If the overreaction and underreaction scenarios are considered equally likely a priori, then the prior mean for the mispricing effect might be zero, though the variance could be large. In this view, the literature's emphasis on the overreaction scenario is driven more by the data than compelling theoretical arguments. Suppose, instead, that overreaction is taken more seriously, a priori. We let $\mu_{\beta m} = 0.20$ in our "overreaction prior" (OP), implying that a 100 bps change in yield is associated with a change of 20 bps in expected return associated with mispricing. In this case, a two-percentage-point drop

---

[19] One might argue that greater return volatility should be associated with greater variation in expected returns through λ and β_m in the prior. As modeled here, variation due to changing risk (roughly $\gamma c \lambda$) increases with the level of c. However, variation related to mispricing is not correlated with c as currently modeled. Incorporating correlation of this sort and allowing for negative correlation between λ and β_m,

in yield below the mean, a large decline in historical terms, leaves the expected return at 10 bps, or just 1.2% *per annum*.

Even if overreaction is viewed as the fundamental behavioral bias, the leap from experimental psychology to financial market equilibrium entails additional uncertainty. To reflect this, we let $\sigma_{\beta m} = 0.15$. The prior probability that β_m is positive is then about 90% - high, but not a sure bet. For those who would not rule out behavioral influences or model mispecification, but are more uncertain about the likely effect, we let $\mu_{\beta m} = 0.05$ and $\sigma_{\beta m} = .20$ in our "neutral prior" (NP). A small positive value for the mean reflects the earlier observation that even random mispricing effects can induce a positive coefficient. Later in the paper, we consider "skeptical priors" that rule out mispricing with some positive probability.

"Diffuse" or "noninformative" priors are often used in Bayesian analysis in an attempt to represent prior "ignorance" or, more pragmatically, to let the data dominate the resulting posterior beliefs. This is not always appropriate, insofar as one often does have some belief about parameter values based on theory or earlier evidence. Still, an examination of noninformative priors can provide a useful reference point for further analysis. We employ flat priors (identically equal to one) for all parameters except c. In that case, we follow the usual approach of taking the prior for ln(c) to be flat.

5. Overview of the Bayesian methodology

In the standard Bayesian regression framework, the regressors are distributed independently of the disturbances, with a distribution that does not depend on the parameters α, β, and σ_ε. These conditions hold, in particular, if the regressors are nonstochastic. As KS have noted, however, independence is violated in (1) because the return surprise at time t impacts all values of dividend yield from time t forward. As a result, one cannot simply work with the joint density of returns conditional on the time series of yields. KS accommodate stochastic regressors by modeling return and yield as elements of a homoskedastic vector autoregression, conditioning on the initial sample

---

conditional on c, would seem resonable. Since the data dominate the posterior for c, this would essentially amount to using conditional priors for β_m and λ, given the posterior mean for c.

value of yield, and imposing a noninformative prior on the VAR parameters.[20] In this case, results from Zellner (1971) can then be applied.

An alternative approach developed in this paper permits quite general informative prior beliefs for the parameters in the return equation. It is based on a noninformative prior assumption for the additional yield-related parameters, as discussed below. In the special case of constant risk, as in KS, analysis of the return equation under our assumptions would proceed as if the regressor, dividend yield, were nonstochastic. This continues to be true if conditonal risk is linear in yield, though analysis with the exponential specification is a bit more complicated. A general overview is provided below. Additional details are given in the Appendix.

5.1. The dividend yield equation

As in the VAR, the conditional joint distribution of return and yield at time t is assumed to depend only on x_t. This density is expressed as

$$f_t(r_{t+1}, x_{t+1}) = f(r_{t+1} \mid x_t)\, f(x_{t+1} \mid r_{t+1}, x_t), \tag{10}$$

where $f(r_{t+1} \mid x_t)$ is the density associated with the regression (4) and $f(x_{t+1} \mid r_{t+1}, x_t)$ is the density based on[21]

$$\log(\text{yield}_{t+1}) = \varphi + \rho \cdot \log(\text{yield}_t) + \phi \cdot \varepsilon_{t+1} + v_{t+1}. \tag{11}$$

The disturbance in (11) is assumed to be i.i.d. normal, independent of the ε's, with mean zero and constant standard deviation σ_v conditional on x_t. We characterize the conditional density for yields in terms of logs to reflect the fact that *percentage* changes in yield and price will tend to be similar, but of the opposite sign. Thus, the parameter ϕ is expected to be close to –1.0.

Equation (11) can also be viewed as an AR(1) model with composite error term $\phi\varepsilon_{t+1} + v_{t+1}$. Since (unexpected) return is modeled as heteroskedastic in (2), this autoregressive error will generally be conditionally heteroskedastic as well. Table 3 presents estimates for (11). As in previous research, the autocorrelation parameter, ρ, is

---

[20] The resulting posterior distribution is identical to that for the nonstochastic regressor case apart from a degrees of freedom adjustment.

always close to one, reflecting the strong persistence in yields. The return parameter ϕ is close to –1.0, as anticipated. Since the return equation is our primary interest, it is convenient to employ a noninformative prior for the parameters of (11). More specifically, conditioning on (γ, c, λ, β_m), so that ε_{t+1} can be treated as observable, the prior for (φ, ρ, ϕ, $\log(\sigma_v)$) is taken to be flat.[22]

5.2. The simulation methodology

The complexity of our regression specification is such that simple analytical formulas for posterior moments are not readily available. Therefore, we make use of a simulation technique known as "importance sampling," an alternative to the Gibbs sampling method that is often used in Bayesian applications (see Bauwens, Lubrano, and Richard (1999), pp.76-83, and Geweke (1989)). Importance sampling can be used to approximate the expectation of any function of the model parameters. Moreover, since it entails i.i.d. sampling, the standard central limit theorem can be invoked to obtain direct measures of precision for the simulation estimates.

Consider, for example, the computation of the posterior mean for θ, the sole parameter in some model. By standard Bayesian analysis, the posterior is proportional to the product of the prior and the likelihood function. Hence, the posterior mean is the integral, over θ, of the product $\theta\ f(y \mid \theta)\ p(\theta)$, divided by p(y). Here, f is the likelihood function or density for the data y, $p(\theta)$ is the prior density, and p(y) is the probability of the data under the prior. Since the required integral can be viewed as an expectation under the prior, it can be evaluated, via simulation, by repeatedly drawing values of θ from the prior density and averaging the products, $\theta\ f(y \mid \theta)$. If the data and the prior "disagree," however, i.e., if the likelihood function and the prior density are concentrated in different regions of the parameter space, the convergence can be quite slow.

The simple idea behind importance sampling is ingenious. Random draws are made, not from the prior, but from an *importance density*, $i(\theta)$, that is a better approximation to the desired posterior distribution. If one now computes the *weighting*

---

[21] Note that conditioning on x_t and r_{t+1} is equivalent to conditioning on x_t and ε_{t+1}.
[22] It might prove worthwhile to explore informative priors for the yield equation as well, but using noninformative priors seems like a reasonable simplifying assumption for now.

function, $w(\theta) \equiv f(y \mid \theta)p(\theta)/i(\theta)$, at each iteration, the average of $\theta w(\theta)$ will converge to the expectation under the importance density or

$$\int \theta\, w(\theta)\, i(\theta)\, d\theta = \int [\theta\, f(y \mid \theta)p(\theta)/i(\theta)]\, i(\theta)\, d\theta = \int \theta\, f(y \mid \theta)p(\theta)\, d\theta. \qquad (12)$$

The required integral is again obtained since $i(\theta)$ cancels out. By a similar argument, the average of the weights $w(\theta)$ converges to $p(y)$ as the number of random draws approaches infinity. The posterior mean is then obtained by taking the appropriate ratio.

Although the convergence results above are quite general, reducing the variability of the weights improves the speed of convergence. Since the numerator of the weighting function is proportional to the posterior density for θ, the weights tend to be fairly equal if the importance density is a good approximation to the posterior. A variety of methods have been proposed to specify importance densities. We have had success with the following approach. Initially, the prior is taken as $i(\theta)$ and "rough" estimates of the posterior moments are obtained through simulation. A second round importance density is then specified using the estimated posterior means and standard deviations in place of the prior moments. After several rounds, each with a modest number of iterations, variability in the importance weights is reduced substantially. At that point, one more simulation is run with a large number of iterations to obtain the desired degree of approximation. Additional details are given in the appendix.[23]

Although we have discussed the computation of the mean for simplicity, the ideas are unchanged if the function of interest is the expectation of a more complicated function $\pi(\theta)$. To compute expected utility, returns are randomly drawn from the conditional distribution of returns, given the fixed yield and θ. These returns, $R(\theta)$, then play the role of $\pi(\theta)$.

6. Bayesian evidence: 1940-99

We begin by examining the longer 1940-99 period and later turn to the more recent experience since 1960. Posterior beliefs about the model parameters are discussed first, and the implications for asset allocation afterwards.

6.1. Posterior distributions for diffuse and balanced priors: 1940-99

The diffuse results in the first panel of Table 4 are similar to the 1940-99 row of Table 2. The posterior means for β, β_m, and δ are lower than the GMM estimates, however.[24] The lack of equality is due to the nonlinear heteroskedastic feature of our model. The posterior standard deviations in parentheses are mostly smaller than the GMM standard errors, and the standard deviations for c and λ are much lower.

Continuing with the diffuse panel, the numbers in brackets are the posterior probabilities that the given parameter exceeds zero. The probability is 86% for λ (changing risk) and it is 95% or higher for all other parameters. Since the mean for λ is fairly small (a 2% increase for a 100 bps increase in yield), most of the return predictability is attributed to the mispricing component. The mean for β_m is 0.18 and it is 0.21 for the total predictability parameter β. As a result, the expected return from mispricing increases by 18 bps (per month) and the expected Sharpe ratio increases by a third from its long run value ($\gamma = 0.15$) when yield rises by 100 bps.

We now turn to the informative priors in the second and third panels of Table 4. Obviously, a strong prior belief about the value of a parameter will carry over to the posterior. Less clear is the *extent* to which the data will impact posterior beliefs in light of the different priors. In a standard regression context, posterior precision (reciprocal of standard deviation) is the sum of prior and sample precisions, while the posterior means are precision-weighted averages of the prior means and the regression estimates. We find that these relations are approximately valid for γ and β_m, and not too bad for the heteroskedasticity parameter λ, with the diffuse posterior standard deviations interpreted as indications of data precision.[25]

As might be expected, posterior beliefs about parameters other than the betas and δ are not very sensitive to the choice of NP or OP. Since the informative prior and

---

[23] This method of computing p(y) is also used in the evaluation of Bayes factors which are discussed later in the paper.

[24] Since expected return is a nonlinear function of yield through the impact of variance, β is defined here as the average change in expected return for a 100 bps increase or decrease in yield from its mean.

diffuse posterior standard deviations for γ are nearly equal, the posterior mean, .13, is halfway between the corresponding means, 0.11 and 0.15. The data are much more informative (lower diffuse standard deviations) about λ and, to a lesser extent β_m, so the NP/OP posterior means are closer to the diffuse means for these parameters. The data are so informative about the standard deviation of returns that the posterior for c is barely affected by the high prior mean of 6% and standard deviation of 1% under NP/OP.

Recall that the neutral prior (NP) is relatively more agnostic about the sign of potential mispricing effects, with a 60% probability for a positive β_m, as compared to the 91% probability under the overreaction prior (OP). The diffuse/OP posterior mean of 0.18 for β_m is fairly large in economic terms, amounting to an increase in the annualized market risk premium of more than 2 percentage points for a 100 bps increase in yield. The posterior mean based on NP is naturally a bit lower, reflecting the lower prior mean. In this case, the probability that β_m is positive increases from 60% to 93% after learning from the data. For the already confident OP investor, the probability increases from 90% to 98%.

6.2 Bayes factors and skeptical priors: 1940-99

The priors considered thus far have implicitly reflected a belief that risk and "mispricing" effects exist with probability one. This continuous approach is typical of most applications in the finance literature, including the no-predictability prior of KS, which combines hypothetical past sample evidence of no predictability with a diffuse prior.[26] In this context, beliefs about the sign and magnitude of the effect can be analyzed, as above, but the question, "is there predictability in risk or return that is related to dividend yield?" is not of interest – it is already presumed.

Here, our focus is on the mispricing component of expected return variation. Intuitively, we would like to examine whether no-mispricing or mispricing is more likely in light of the given data. In the case of a simple alternative, say $\beta_m = 0.20$, with no other

---

[25] The posterior mean for β under NP is lower than both the prior and diffuse posterior means. This occurs because the NP posterior mean for β_m is lower than the diffuse mean and the posterior risk effect through λ is small, while the prior risk effect is large.

[26] An exception is the Bayesian test of portfolio efficiency in Shanken (1987). An interesting example in the context of mutual fund performance is Baks, Metrick, and Wachter (2001).

parameters to consider, we would just consider the likelihood ratio for the hypotheses, $L(\beta_m = 0)/L(\beta_m = 0.20)$, a function of the data alone. In our broader context, the likelihood that β_m=0 must be evaluated jointly with the other model parameters, γ, c, and λ, and averaged according to our prior for those parameters. Moreover, with a composite hypothesis like $\beta_m \neq 0$, the likelihoods must also be averaged over all nonzero values of β_m, again weighted by prior beliefs.

To formalize these ideas, supposethat one assigns positive prior probability to the composite hypothesis that "very little" of the variation in yield is related to mispricing. It is convenient to model this in terms of the simple point "null" hypothesis that β_m equals zero. In such situations, Bayesian analysis commonly proceeds with a "mixed prior" that assigns some positive probability to the null, with the remaining mass spread continuously over the rest of the parameter space.[27] This probability, call it π, would be higher for strong advocates of traditional market efficiency, but could still be positive for individuals inclined toward behavioral models or explanations for return phenomena that involve learning.[28] As long as the likelihood function is fairly flat in the neighborhood of zero considered relevant, results obtained this way will approximate those for more detailed priors that focus on the composite "null" hypothesis.

Let $NP_{\beta m=0}$ and $OP_{\beta m=0}$ denote the conditional versions of our balanced priors, neutral and overreaction, given $\beta_m = 0$. Formally, we consider a mixed prior, $p(\gamma, c, \beta_m, \lambda)$, that equals $NP_{\beta m=0}$ with probability π and equals NP with probability $1-\pi$. This prior assigns probability π to the hypothesis, $\beta_m = 0$, with the remaining mass spread out according to NP. Similarly for OP. In this context, the *Bayes factor* for $\beta_m = 0$, BF, refers to the expected likelihood under $NP_{\beta m=0}$ divided by the expected likelihood under NP. In the simple hypotheses case mentioned above, BF is just the likelihood ratio.

In general, the *posterior odds ratio* for a hypothesis is defined as the posterior probability that the hypothesis is true divided by the posterior probability for the alternative. This ratio serves as the basis for a sort of Bayesian hypothesis test or, more

---

[27]Recent papers by Avramov (2001) and Cremers (2000) consider this possibility within a broader analysis of "model uncertainty." The predictive variables in their analyses are treated as nonstochastic, however. See Chapter 4 of Leamer (1978) for a good discussion of these issues.

accurately, hypothesis comparison. The posterior odds ratio in favor of $\beta_m = 0$ can be written as

$$P(\beta_m = 0)/ P(\beta_m \neq 0) = BF^*\pi/(1-\pi), \qquad (13)$$

the Bayes factor times the prior odds ratio. Intuitively, if zero is more likely than other values of β_m, given the observed data, then BF is greater than zero and the odds in favor of $\beta_m = 0$ increase. This approach contrasts with conventional p-value analysis, which assesses whether an observed statistic is "unusual" solely from the perspective of the *null hypothesis* –alternative hypotheses play no role in the p-value computation.

As reported in the second panel of Table 4, BF for $\beta_m = 0$ equals 0.72 for the neutral prior and 0.51 for the overreaction prior. Thus, the odds move away from $\beta_m = 0$ somewhat with NP, but are cut in half with OP. Let us consider the reason for this difference. As we have seen, the sample evidence favors relatively large positive values of β_m, consistent with expections under OP. On average, these positive values are given greater weight under OP than under NP. Therefore, the expected likelihood ratio for $\beta \neq 0$ is higher, and the Bayes factor lower under OP. The bottom line is that a skeptical investor who believes a priori that mispricing, if any, will tend to induce reversals (OP), finds this belief confirmed in the data and becomes relatively more convinced that mispricing effects do indeed exist. For example, with $\pi = 0.50$, the skeptical OP probability for no-mispricing decreases from 0.50 to 0.34 = BF/(1+BF), while the skeptical NP probability declines to 0.42.

Turning to the risk parameter λ, recall that the balanced priors for this parameter reflect an expectation that risk will increase if yield rises. The Bayes factors for $\lambda = 0$ in Table 4 are quite high, indicating a significant shift in beliefs toward a constant risk model. For example, an NP investor skeptical about the possibility of risk changing with dividend yield ($\pi = 0.50$) comes away from the data with a posterior probability of 88% that risk is constant conditional on dividend yield. It is important to appreciate that such conclusions will, in general, vary with the investors' prior beliefs.

---

[28] Of course, this presumes a high degree of confidence in the *approximate* validity of the simple mapping of risk into expected return.

6.3 The optimization framework

As in KS, we consider an individual investor with a single-period investment horizon and an iso-elastic utility function,

$$U(W) = W^{1-A}/(1-A), \tag{14}$$

where A is the coefficient of relative risk aversion. At the end of month T, the investor chooses a weight ω* on the market index, so as to maximize the expected utility of end-of-period wealth,

$$\begin{aligned} W_{T+1} &= W_T[\omega \exp(r_{T+1} + i_{T+1}) + (1-\omega) \exp(i_{T+1})] \\ &= W_T \exp(i_{T+1})[\omega \exp(r_{T+1}) + (1-\omega)] \end{aligned} \tag{15}$$

where $0 \le \omega \le 1$, r_{T+1} is the continuously compounded excess stock return, as earlier, and i_{T+1} is the continuously compounded riskless interest rate for month T+1. In general, the certainty equivalent return premium (CER) is the excess return that, if known for sure, would provide the same utility as the optimal portfolio.

A is taken to be 5 in our illustrations, similar to values used in previous studies - low enough to generate substantial allocations to stock, but high enough to avoid too many corner solutions. The riskless rate is 40 bps per month throughout. It follows from (14) and (15) that ω* does not depend on the level of the riskless rate, though the optimal level of utility is affected. Our investor is not assumed to be a representative investor for the economy. In fact, since we entertain the possibility of behavioral biases and associated mispricing effects in equilibrium, our fully rational Bayesian investor cannot, in general, be representative. The investor maximizes expected utility with respect to his *predictive* probability distribution, which conditions on past empirical evidence, yields and returns (the vector y_T), as of the end of month T, and prior beliefs.

The predictive distribution can be viewed as a mixture of distributions, each conditioned on a set of parameter values, and averaged according to the posterior distribution of the parameters:

$$p(r_{T+1} \mid y_T) = \int f(r_{T+1} \mid \theta_1)\, p(\theta_1 \mid y_T)\, d\theta_1, \tag{16}$$

where $\theta_1 \equiv (\gamma, c, \lambda, \beta_m)$ and $p(\theta_1 \mid y_T)$ is the posterior density derived from a prior density and the data. We sometimes refer to this posterior-based distribution simply as "the predictive distribution." Also of interest is the prior-based analog of (16), the *prior-predictive distribution*:

$$p(r_{T+1}) = \int f(r_{T+1} \mid \theta_1)\, p(\theta_1)\, d\theta_1, \tag{17}$$

where $p(\theta_1)$ is a proper (informative) prior density for θ_1.

KS note, and we confirm for our model, that the following approximation to the optimal weight works quite well

$$\omega^* \approx \mu_T/(A\sigma^2_T) + \tfrac{1}{2}A, \tag{18}$$

where μ_T is the predictive mean and σ^2_T the predictive variance. If there is no mispricing, ($\beta_m = 0$) then, based on (4), the ratio of expected return to variance remains constant as yield changes. Insofar as this holds approximately for the predictive mean and variance as well, ω^* will also be constant. Then, any shifts observed in the optimal allocation to stock as yield varies must be driven by a belief in mispricing.

6.4. Optimal weight and CER comparisons: 1940-99

Predictive results for the period 1940-99 are given in Table 5. The first column lists the scenarios examined. For each prior, values of dividend yield at the mean and 1.5 sample standard deviations above or below the mean are considered. Let's start with the diffuse prior. When yield is at its mean, the predictive expected return is just the posterior mean of α, 61 bps per month (Table 4). When yield is high, the mean increases to 102 bps and at the low yield it is 21 bps. Predictive risk also increases with yield, but not as dramatically. As a result, the optimal weights (under "True") rise sharply, from 37% to 100%, as yield increases. The corresponding CERs, in the next-to-last column (under "True"), increase from 5 to 64 bps. Thus, due to the strong evidence of mispricing, predictability has a large impact on asset allocation.

In the spirit of the Kandel and Stambaugh's analysis of economic significance, consider the effect of ignoring predictability in both expected return and risk, and allocating 81% (the optimal allocation when yield is at the mean) to stocks, regardless of the actual level of yield. The numbers under $\mu\sigma$ in the CER-posterior portion of the table are the certainty-equivalent losses implied by deviating from the true optimal allocations, where "true" refers to analysis based on the predictive distribution perceived as correct by the investor. For example, when yield is low and the (true) predictive expected return is just 21 bps, it is optimal to invest 37% of the portfolio in stock. Investing 81%, under these circumstances, results in a CER loss of 7 bps. On the other hand, when yield is high, the loss from underinvestment is 5 bps.

The numbers just reported illustrate one method for evaluating the joint significance of predictability in both expected return and risk. Suppose, instead, that we want to evaluate the *partial* effect of ignoring expected return variation, holding risk constant. To get at this, we again optimize using the expected return for yield at its mean (x=0), but this time we use the correct predictive variance for the given level of yield. The corresponding "optimal" weights (but suboptimal with respect to the correct predictive distribution) under μ in Table 5 display the anticipated properties. Ignoring the low expected return when yield is low makes the investor overly aggressive (87% vs. 37% in stock), resulting in a CER loss of 10 bps. Similarly, there is underinvestment at the high yield.[29]

To evaluate the partial effect of ignoring variation in risk, holding expected return constant, we optimize using the predictive variance for yield at its mean, together with the true predictive expected return for the given level of yield.[30] The weights are close to those under the true predictive distribution and the CER losses are thus essentially zero because of the weak relation between risk and dividend yield over this period.

Given the risk/mispricing theme of our paper, it is also of interest to evaluate the separate economic significance of the mispricing component of expected return variation. As noted above, shifts in asset allocation are, to a close approximation, driven solely by the mispricing component. For example, with β_m equal to zero, the optimal weight is

---

[29] The optimal weights decline as we go down the μ column, because risk increases while expected return is held constant. The weights increase in the σ column because expected return increases with risk is held constant.

[30] We are not suggesting that an investor is likely to incorporate risk variation and ignore expected return variation, any more than we expect an investor to correctly assess a predictive distribution but then plug in the wrong value of yield, as in KS. These are just ways of evaluating the economic "relevance" of predictability.

81% under the diffuse prior at each level of yield, even though risk and expected return change with yield. It follows that the CER measure of significance is essentially the same whether we ignore all predictability or just the mispricing component of expected return variation. Thus, the $\mu\sigma$ column in our tables should be interpreted accordingly.

All of the CER and optimal weight comparisons discussed thus far, like those in earlier literature, are properly interpreted as assessing the economic significance of *posterior* beliefs. Insofar as the prior is truly "uninformative," a slippery notion, significance may reasonably be attributed to the data in this case. With an informative prior, however, the posterior reflects the prior as well as the data. For example, suppose that predictability is expected under the prior, but not under the posterior. Ignoring predictability in the posterior, as earlier, would have no effect on asset allocation, yet the data have indeed had a significant impact on the investor's beliefs. Therefore, an alternative metric is needed to assess the *incremental* economic significance of the data or empirical evidence. One natural approach is to simply compare optimal weights under the prior and posterior-predictive distributions.

The prior-predictive weights are lower than the posterior-based weights in Table 5 because of the high (compared to the data) prior values of the risk parameter, c, and the resulting high predictive risk. Given the large prior mean for λ, the largest differences in asset allocation are observed when yield is high. The CER loss implied by following the roughly 50-50 prior allocations, rather than being fully invested in stock at the high yield, is 15 bps for OP and 14 bps for NP. The losses represent about 25% of the optimal CERs. By these measures, the *data* have had an economically important influence from the perspective of the NP/OP investors. Also, note that the CER losses in this comparison are larger than those based on ignoring predictability (the $\mu\sigma$ numbers). Again, those numbers tell us something different - about the significance of the (final) posterior predictability beliefs.

To summarize, we find substantial evidence of expected return predictability that is unrelated to risk in the 1940-99 period and this evidence has important implications for asset allocation. The reduction in perceived risk relative to the highly volatile experience of the 1920s and 30s is also economically significant. On the other hand, risk itself does not appear to vary much with dividend yield. We now consider beliefs based on the more recent period since the 1960.

7. Bayesian evidence: 1960-94 and 1960-99

In this section, we examine the beliefs of investors who form posteriors based on the data from 1960 onward. First, we look at the period 1960-94 and then we consider the impact of the last five years of the 1990s by examining the period 1960-99.

7.1. Posterior distributions: 1960-94

As with the 1940-99 period, the diffuse posterior means in Table 6 are similar to the GMM estimates in Table 2, but the differences between GMM and the posterior means for β_m, β, and δ are greater now. The means for α and γ are a little lower than GMM as well, while the mean for λ is larger than the GMM estimate. Again, the posterior standard deviations are lower than the GMM standard errors, particularly for c and λ.

The low mean for γ, relative to the long-run Sharpe ratio for 1940-99, is a reflection of the low average returns on the market over the 1960-94 period. The diffuse posterior mean of 16.4 for λ, based on the 1960-94 data, is quite high in economic terms, with risk increasing by roughly 16% for a 100 bps increase in yield. The balanced-prior investors become virtually certain that risk increases with yield, as indicated by the posterior probability of 1.00 in brackets. Moreover, the very low Bayes factors for λ in Table 6 tell us that even a "skeptical" investor with, say, prior probability $\pi = 0.9$ for constant risk, will see the odds shift overwhelmingly toward the conclusion that risk does increase with yield after looking at the data.

The diffuse mean for β_m is very high as well, implying an increase in expected return of 37 bps for a 100 bps increase in yield. As a result, the expected Sharpe ratio more than doubles (increases by 0.10) from its long run value of 0.06 when yield rises by 100 bps. Since the mean for β is 0.47, again most of the return predictability is attributed to the mispricing component. There is considerable uncertainty around these expected effects, however, given the large posterior standard deviations for β_m and δ. Posterior means for β_m and β, based on the informative priors, are much lower than the diffuse mean. This high degree of shrinkage toward the prior means reflects the lower precision of the data in this shorter estimation period. The posterior probabilities for $\beta_m > 0$ are similar to those observed in the longer period and the Bayes factors are nearly equal.

7.2. Optimal weight and CER comparisons: 1960-94

Predictive results for the period 1960-94 are given in Table 7. Predictive expected return increases greatly with yield, with the mean under diffuse/OP actually negative now when yield is low.[31] Predictive risk also increases with yield, from 3.5% to 5.2%. The expected return effect dominates, however, as the optimal weights on stock all rise sharply with yield, from zero to 73% in the case of the diffuse prior. These patterns reflect the large positive posterior values of β_m and λ. As earlier, the partial effect of expected return variation is large, but now the partial effect for risk is significant as well. When yield is high, ignoring variation in risk increases the allocation to stock by nearly 30 percentage points, with the diffuse/OP investors fully invested in stock. The associated CER losses are 5-6 bps.

Large differences between prior and posterior-predictive weights are only observed in Table 7 when yield is low. In particular, with the neutral prior, the optimal prior weight is 49% while the posterior-based weight is just 15% - a reflection of the low prior mean for β_m and high mean for α, relative to the data. Notice that the optimal weights under OP are equal when yield is high, despite the nontrival differences between prior and posterior predictive moments. This is because the drop in utility due to the decline in expected return from 111 bps to 84 bps is just offset by the increase in utility due to the decline in risk from 6% to 5.2%. In other words, the prior and posterior-based optimal portfolios lie roughly on the same indifference curve.

This example shows that when predictability in both moments is admitted, focusing exclusively on the "bottom line" from an investor's perspective can mask shifts in beliefs that would be judged "economically significant" from the broader perspective of a researcher. Therefore, we advocate the reporting of predictive moments, as well as

[31] Since these are moments of continuously compounded returns, it is possible for the allocation of stock to be positive when the predictive mean is slightly negative, as with OP and low yield. The actual expected (excess) return is positive.

utility-based measures of economic significance, in such contexts. The variation in predictive moments in Table 7 is quite significant in our view.

To summarize, the data for the 1960-94 period point to substantial predictability in both return moments, with the evidence for risk variation particularly strong. Most of the predictability in expected return is not proportional to risk, however. While the partial effects of expected return and risk variation on asset allocation are both large, the former effect is dominant.

7.3. Predictive risk and parameter uncertainty: 1960-94

In general, the predictive variance reflects the underlying "fundamental" uncertainty in returns, σ_ε, as well as uncertainty about the true values of the model parameters and the implied expected return. For example, the NP/OP prior-predictive standard deviation when yield is at the mean is 4.53%, a bit larger than the prior mean of 4.5% for c. This small increase mainly reflects uncertainty about c, as the estimation risk (parameter uncertainty) effects for α and γ can be shown to be negligible.

When yield deviates from the mean, parameter uncertainty should become more important since estimation errors in β_m and λ are then also relevant. It is somewhat surprising, therefore, that the (posterior) predictive risk levels for NP and OP are similar to, and actually larger than those for the diffuse case in Table 7. Surprising, because parameter uncertainty should be maximized using the diffuse prior, as is confirmed by the evidence (posterior standard deviations) in Table 6. However, valid comparisons must take into account differences in the posterior means of c and especially λ across priors.

One way of assessing the importance of parameter uncertainty is to compare the predictive standard deviation to the standard deviation of returns under the assumption that the posterior means equal the true parameters. The latter standard deviation is just $E(c)\cdot\exp(x\cdot E(\lambda))$. When we do this for the diffuse predictive distribution (not in tables), the risk attributable to parameter uncertainty is just 1 bp for yield at the mean, 2 bps at the low yield and 3 bps at the high yield. For the informative priors, the numbers are lower by about 1 bp. The parameter uncertainty effect should be greater for the *prior-predictive* risk measures since uncertainty is higher before observing the data. For NP, the numbers are 3 bps when yield is at the mean, 15 bps at low yield and 26 bps at high yield, with virtually all of the parameter uncertainty coming through the risk parameters, c and λ. Similarly for OP.

To summarize, a small amount of prior-predictive risk comes from parameter uncertainty when yield is at its mean. Prior parameter uncertainty becomes more important as yield deviates from the mean and this effect is due almost entirely to the changing risk feature of our model. With regard to posterior-predictive risk, most of the parameter uncertainty effect has been eliminated through the learning that occurs from observing the data. Avramov (2001) performs a similar risk decomposition in a broader context, allowing for uncertainty about the correct set of predictive variables as well. As also noted by Kothari and Shanken (2001) in the context of (unconditional) asset allocation with a market index and anomaly-based tilt portfolios, the inherent variability of stock returns dominates predictive risk.

7.4. Influence of the late 1990s

As noted earlier in discussing the OLS/GMM estimates in Table 2, the high returns and low dividend yields of the late 1990s were unprecedented. It is interesting, therefore, to consider the impact of this data on the beliefs of Bayesian investors with the priors discussed above. Analyzing these priors in light of the data for 1960-99 is equivalent to letting the posterior for 1960-94 serve as the prior at the beginning of 1995 and then adding five more years of data. The results are given in Table 8.

It is striking that the belief in mispricing is completely eliminated for the diffuse investor, with the posterior mean for β_m now virtually zero (from 0.37 earlier) and the probability of a positive mispricing effect dropping from 93% to 48%. Similar remarks apply to the NP investor. For the OP investor, the posterior mean drops from 0.24 to 0.11 and the probability of a positive effect declines from 97% to 83%. The impact for this investor is substantial, but naturally less dramatic, given the strong prior belief in a positive effect.

The Bayes factors provide additional perspective on the shifts in beliefs, with the odds now favoring no-mispricing. In particular, the BF of 2.0 for the OP prior implies that the posterior probability of no-mispricing is two-thirds for the skeptical OP investor (π=0.5), a complete reversal from the two-thirds probability *for* mispricing at the end of

1994. Although the 1960-99 posterior means for λ are quite a bit lower, the posterior probabilities and Bayes factors still strongly support a belief in changing risk.

Now we consider the impact of these dramatic shifts in beliefs on optimal asset allocations. Predictive results for the 1960-99 period are reported in Table 9. Consistent with the reversal of belief in mispricing, the optimal asset allocations barely change with yield ("true" columns) for the diffuse and NP priors, quite a contrast from Table 7. Since predictability now comes mainly through changing variance (mean $\beta_m \approx$ zero), the effects of risk and expected return changes on utility are roughly offsetting, as discussed in Section 7.2. With OP, the data have not completely offset the strong prior belief in a positive mispricing effect, so there is still substantial variation in optimal allocations, from 41% to 70%, as yield goes from low to high.

Although the overall effect of predictability on asset allocation is now negligible for the diffuse/NP priors, the partial effects on optimal weights are more substantial, though much smaller than in Table 7. For example, when yield is high the NP optimal allocation drops from 59% to 48% if predictability in expected return is ignored (μ column), and increases to 73% if variation in risk is ignored (σ column). There is little impact on CERs, however. Larger effects are observed for OP.

We conclude this section by taking a more historical view of the belief shifts of the late 1990s. The dividend yield at the end of 1999 was 1.57%. Based on posterior-predictive beliefs at this point in time, the optimal allocations to stock (not in tables) would have been 59% with the diffuse prior, 54% with NP, and 22% with OP. Again, the OP investor's stock allocation differs because he still believes that expected returns are low when yield is low, beyond what is implied by the decline in risk. Additional perspective on the economic significance of the late 90s data is obtained by posing the following question. How would the allocations at the end of 1999 have differed if they were based on the beliefs held at the end of 1994? The answer is easy to describe. Each of our balanced prior investors would have been completely out of the market because of the very low perceived expected stock returns!

Finally, what about our other investors, the skeptical cousins of OP and NP who find it credible a priori that there might not be a mispricing effect? It can be shown that the posterior distribution based on the skeptical NP prior is a weighted average of the posteriors corresponding to NP and $NP_{\beta m=0}$ (see Section 6.2). Likewise for OP. As a result, their posterior means for β_m are closer to zero and thus the skeptical investors should be less influenced by the low dividend yield. Based on predictive beliefs using the 1960-99 data, the skeptical OP investor would put 46% in stock at the 1.57% yield, while the skeptical NP investor would allocate 56%. Using the 1960-94 data instead, even the *skeptical* OP investor's belief is strong enough to keep him out of the market and fully invested in "cash" at the very low yield. As we saw earlier, it is harder to persuade the NP investor that mispricing effects exist. As a result, the skeptical NP investor would put some money in stock based on the 1960-94 beliefs, but only 3% of her portfolio. Thus, the actions of our skeptical investors are also greatly influenced by the data for 1995-99.

8. Summary and Conclusions

We have analyzed the distribution of returns, conditioned on dividend yield, of the CRSP value-weighted stock market index. This is the first study to look at yield-related predictability of both market risk and expected return in a Bayesian framework. Three overlapping subperiods, 1940-99, 1960-94, and 1960-99, have been examined in detail. The evidence of variation in expected return is strong in the 1940-99 and 1960-94 periods, with (diffuse) posterior probabilities of 0.97 and 0.96, respectively, that expected return increases with yield. The high returns and low dividend yields of the late 1990s, when added to the 1960-94 data, completely wipe out the evidence of predictability, however, with the posterior probability for a positive effect dropping to 0.48.

A distinguishing feature of our study is that expected return variation is decomposed into a component proportional to conditional variance, as in Merton (1980), and a residual that we loosely refer to as the "mispricing" component. The evidence that market risk increases with yield is very strong (posterior probabilities of 1.00) in the periods since 1960. Even an investor who initially assigns substantial probability to a constant risk model (a "skeptical" investor) becomes quite convinced of risk variation after looking at the data. The results are not robust, however. That same skeptical investor would be strongly persuaded by the 1940-99 data that risk is *constant*.

In general, there is not much support for the hypothesis that expected-return predictability related to yield is explained by corresponding changes in the conditional variance of returns. Even over the 1960-94 period, in which the evidence of risk variation is strongest, the implied variation in expected returns based on our simple model is small compared to the total variation observed. The mean slope on dividend yield is 0.47, while the mispricing component has a mean of 0.37 under the diffuse prior. A different model of the relation between risk and return might fair better, but it will have a lot to explain.

While positive statements about the "true" return process must be tempered by the lack of robustness of the predictability evidence, interesting observations can still be made about the implications of the data for asset allocation decisions. First, the partial effects on the optimal allocation to stock (versus cash) of ignoring either mispricing-related predictability or risk variation can both be substantial. For example, with predictive beliefs based on the 1960-94 data and a diffuse prior, if the dividend yield is 1.5 sample standard deviations above the mean, the optimal weight on stock is 73%. If one ignores the mispricing component of expected return variation, the optimal allocation is 39% while it is 100% if risk variation is ignored.

For beliefs based on the 1960-99 data, asset allocations barely change with yield when investors have fairly diffuse priors about the likely sign of potential mispricing effects. Predictive risk and expected return both vary with yield, but with no evidence of mispricing over this period, their effects on utility, and hence portfolio choice, are offsetting. Only an investor with a strong prior belief in a positive yield effect still sees a significant role for yield in the allocation decision at the end of 1999. Given these observations, we suggest the reporting of predictive moments as one means of assessing economic significance from the research perspective, rather than relying solely on utility-based measures.

Comparing the optimal portfolios that would be chosen before (prior) and after (posterior) observing the data can also be an interesting exercise. We assume that prior expected risk as of the beginning of 1940, near the end of the Great Depression, would have been quite high by current standards. Since the data tend to be especially informative about the risk parameter, this does not have much of an influence on posterior beliefs based on the 1940-99 data. The risk level drops from a prior mean of 6% per month to a posterior mean of 4.2%, with associated optimal allocations to stock of 46% (prior) and 73% (posterior).

The late 1990s shifts in beliefs are also examined from an asset allocation perspective. Standing there at the end of 1999, with dividend yield at 1.57% and beliefs based on the 1960-99 data, optimal asset allocations vary from 22% to 59% for our "balanced prior" investors. One way of highlighting the importance of the last five years of data is to consider the changes in allocations that would occur if that data were ignored. We find that the same investors would no longer invest *any* money in stock, in this case.

We conclude with a few comments about potential extensions of this work. Certainly, including multiple predictors of risk and expected return simultaneously would be of interest. The implications of imposing stationarity on the yield process (see Stambaugh, 1999) or, more generally, specifying informative priors for the parameters of the yield equation, could be explored. Perhaps most interesting would be an analysis, extending the work of Stambaugh (1999) and Barberis (2000), of the impact of changing expected return *and risk* on the multiperiod investment decision.

Appendix A. Obtaining the posterior moments

Let $y \equiv (y_1, y_2, \dots, y_T)$, where $y_t \equiv (r_t, x_t)$. The joint density (conditioned on x_0) for y is $f(y) = f(y_1)f(y_2 \mid y_1) \dots f(y_T \mid y_1, y_2, \dots, y_{T-1})$, where $f(y_{t+1} \mid y_1, y_2, \dots, y_t) = f(r_{t+1} \mid y_1, y_2, \dots, y_t)f(x_{t+1} \mid r_{t+1}, y_1, y_2, \dots, y_t)$ for t = 1, …,T-1. We assume that x_t captures the state of the world at time t in the sense that $f(r_{t+1} \mid y_1, y_2, \dots, y_t) = f(r_{t+1} \mid x_t)$ and $f(x_{t+1} \mid r_{t+1}, y_1, y_2, \dots, y_t) = f(x_{t+1} \mid r_{t+1}, x_t)$. This representation of the joint density of returns and yields includes the restricted VAR of Stambaugh (1999) as a special case, with return modeled as in (1) and yield following a homoskedastic AR(1) process. The vector (r_{t+1}, x_{t+1}) would be jointly normally distributed conditional on x_t, implying that the conditional density $f(x_{t+1} \mid r_{t+1}, x_t)$ is a linear regression in x_t and r_{t+1}.

Let θ be the parameter vector for our two-equation system, (4) and (11), and partition θ as (θ_1, θ_2), where $\theta_1 \equiv (\gamma, \beta_m, c, \lambda)$ and $\theta_2 \equiv (\varphi, \rho, \phi, \sigma_v)$. Given the discussion above, the joint density of the data or, equivalently, the likelihood function can be written as

$$f(y \mid \theta) = h(y, \theta_1)g(y, \theta). \tag{A.1}$$

Here, $h(y, \theta_1)$ is the product of the conditional densities for r and $g(y, \theta)$ is the product of the conditional densities for x. Note that h is identical to what the conditional density of r, given x, would be if x were nonstochastic. In our application, the vector θ_1 also enters g, however, since $\varepsilon_{t+1} = r_{t+1} - (\gamma/c)\cdot\exp(2x_t\lambda) - \beta_m x_t$.

In general, using $p(\theta) = p(\theta_1)\, p(\theta_2 \mid \theta_1)$ and integrating the posterior $p(\theta \mid y)$ over θ_2, the marginal posterior for θ_1 can be written as

$$p(\theta_1 \mid y) = \int p(\theta_1)\, p(\theta_2|\theta_1)\, f(y \mid \theta_1, \theta_2)\, d\theta_2 \,/\, p(y)$$

or

$$p(\theta_1 \mid y) = p(\theta_1)p(y \mid \theta_1)/p(y), \tag{A.2}$$

where $p(y \mid \theta_1) = E_{p(\theta_2|\theta_1)}[f(y \mid \theta)]$ is the marginalized likelihood of θ_1, defined in terms of the conditional prior $p(\theta_2 \mid \theta_1)$, and $p(y) = E_{p(\theta)}[f(y \mid \theta)] = E_{p(\theta_1)}[p(y \mid \theta_1)]$ is the probability of the data.[32] Using the factorization of the likelihood function in (A.1),

[32] With a diffuse prior for θ_2, these are not really expectations, but the associated integrals are well-defined.

$$p(y \mid \theta_1) = h(y, \theta_1)E_{p(\theta 2|\theta 1)}g(y, \theta). \qquad \text{(A.3)}$$

In order to compute the expectation in (A.3) with θ_1 fixed we observe that, in our application, it is identical to the normalizing constant in a standard linear regression model with the same form as (11) and density $g(y, \theta)$. Since the associated integral is taken over θ_2, with the data treated as constants, the stochastic nature of the regressors in (11) is irrelevant in this context. With a noninformative prior for the regression parameters θ_2, i.e., $p(\theta_2 \mid \theta_1) \propto 1/\sigma_v$, the standard (conditional) posterior for θ_2 would be of the normal-inverted gamma form.[33] It is easily shown in this case that the normalizing constant, the product of prior and likelihood divided by the posterior, satisfies

$$E_{p(\theta 2|\theta 1)}g(y, \theta) \propto 1/\{|(Z(\theta_1)'Z(\theta_1)|^{.5}RSS(\theta_1)^{.5(T-4)}\}. \qquad \text{(A.4)}$$

Here, $Z(\theta_1)$ is the (T-1)x3 matrix of independent variables in (11) and $RSS(\theta_1)$ is the residual sum of squares, both of which depend on θ_1 through ε_t.[34]

Our application focuses on the posterior moments of θ_1 and the predictive moments of returns. In order to evaluate these quantities, it suffices to consider functions $\pi(\theta_1)$. By the argument of section 5.2, with $p(y \mid \theta_1)$ playing the role of $f(y \mid \theta)$, if θ_1 is repeatedly drawn from $p(\theta_1)$ then the average value of $\pi(\theta_1)p(y \mid \theta_1)/p(y)$ will converge to $E_{p(\theta 1|y)}[\pi(\theta_1)] = E_{p(\theta|y)}[\pi(\theta_1)]$. Again, an importance density for θ_1 is incorporated to improve computational accuracy.

If the expectation in (A.3) were a constant function of θ_1, the marginalized likelihood would equal $h(y, \theta_1)$ and the marginal posterior for θ_1 would reduce to the usual posterior for the nonstochastic regressor case. This occurs if expected return is linear in x_t and if equation (11) is modeled in terms of x as well (not a nonlinear transformation of x). Given (4), the linearity of expected return holds if and only if the variance relation is also linear in x_t; in particular, if variance is constant, as in most of the

---

[33]Although we do not make use of this fact, the conditional posterior for θ_2 is of this standard form. This follows from the observation that the factor $h(y, \theta_1)$ is a proportionality constant in the actual likelihood (A.1).

[34]T-1, rather than T, because there is no time T value of yield in the data y.

literature. Under these assumptions, it is easy to see that RSS(θ_1) is independent of θ_1 since the span of x_t, ε_{t+1}, and a constant is independent of θ_1. Moreover, it can be shown that the determinant in (A.4) is also constant in this case.

Although we have chosen to work with the conditionally diffuse prior for θ_2 in order to focus on the return equation, the basic methodology can be modified to accommodate draws of θ_2 from an informative prior. Also, it is not essential that priors and returns be normally distributed in order to apply the general procedure. This flexibility will undoubtedly be of use in future extensions of this work.

Appendix B: Obtaining the predictive moments and optimal weights

In the case of expected predictive utility, E[U(R)], we would ideally want to compute $\pi(\theta_1) \equiv E[U(R) \mid \theta_1)]$. Since this is not readily obtainable, for each draw of θ_1, we randomly draw $R(\theta_1)$ from the normal density $f(R \mid \theta_1)$ and compute $U(R(\theta_1))$. The utilities generated in this manner are i.i.d. draws. The expectation of $U(R(\theta_1))$, conditional on θ_1, is $E[U(R) \mid \theta_1]$ and the unconditional expectation is, by the law of iterated expectations, E[U(R)]. Thus, $U(R(\theta_1))$ plays the role of $\pi(\theta_1)$ in this context.

We also modify the procedure just described to incorporate antithetic sampling (see Bauwens, Lubrano, and Richard (1999), pp. 75-76), which increases computational efficiency. Given θ_1 and the fixed value of dividend yield, the mean and variance of the conditional normal distribution of returns is known. The unexpected return, ε, is drawn from a normal distribution with mean zero and the given variance, and is then added to the expected return. Then, an antithetic return is obtained by repeating the computation with $-\varepsilon$ in place of ε. Utility is computed for each pair of returns and the average serves as $U(R(\theta_1))$. The expectation of $U(R(\theta_1))$ is unchanged using this modification, but variability is reduced, improving the computations.

To compute the optimal weight, the returns from 100,000 iterations are saved and an optimization routine is used, again incorporating the antithetic perspective. In order to get a measure of the precision (and bias) for the optimal weight, a bootstrap approach is used. That is, we sample returns from the original 100,000 with replacement and compute a new optimal weight. This is done 100 times, generating a series of 100 i.i.d.

optimal weights. Bias is assessed by comparing the original optimum to the bootstrap average and precision estimated by the mean-squared errors of the 100 weights around the original optimum. The difference between the approximation in (18) and the "exact" optimal weight was less than 0.01 in a range of cases. The bootstrap routine should prove useful in more complicated situations in which simple approximations are not available.

References

Abel, Andrew, 1988, Stock prices under the time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model, *Journal of Monetary Economics* 22, 375-393.

Attanasio, Orasio, 1991, Risk, time-varying second moments and market efficiency, *Review of Economic Studies* 58, 479-94.

Avramov, Doron, 2001, Stock return predictability and model uncertainty, *Journal of Financial Economics*, forthcoming.

Baks, Klaus, Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, Journal of Finance 56, 45-85.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225-264.

Black, Fischer, 1976, Studies of stock price volatility changes, in: *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section*, American Statistical Association, 177-181.

Bauwens Luc, Michel Lubran and Jean-Francois Richard, 1999, *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

Bossaerts, P. and P. Hillion, 1999, Implementing statistical criteria to select return forecasting models: what do we learn?, *Review of Financial Studies* 12, 405-428.

Campbell, John, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.

Campbell, J. and R. Shiller, 1988, The dividend price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-228.

Christie, Andrew, 1982, The stochastic behavior of common stock variances: value, leverage and interest rate effects, *Journal of Financial Economics* 10, 407-432.

DeBondt, Werner and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793-805.

De Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Fama, Eugene, 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383-417.

Fama, Eugene F. and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115-146.

Fama Eugene F. and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3-25.

Fama Eugene F. and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23-49.

Foster, F. D., T. Smith, and R. E. Whaley, 1997, Assessing goodness-of-fit of asset pricing models: the distribution of maximal R^2, *Journal of Finance* 52, 591-607.

French, Kenneth, G.W. Schwert, and R. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics*, 19, 3-29.

Geweke, John, 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317-1339.

Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.

Glosten, Lawrence , Ravi Jagannathan and David Runkle, 1993, On the relation between the expected value and volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1179-1801.

Goetzmann, William N. and Philippe Jorion, 1993, Testing the predictive power of dividend yields, *Journal of Finance* 48, 663-679.

Goyal, Amit and Ivo Welch, 1999, Predicting the equity premium, working paper, Yale University.

Harvey, Campbell, 1989, Time-varying conditional covariances in tests of asset-pricing models, *Journal of Financial Economics* 24, 289-317.

Hentschel, Ludger, 1995, All in the family: Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics* 39, 71-104.

Hodrick, R., 1992, Dividend yields and expected stock returns: alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357-386.

Jorion, Philippe and William Goetzmann, 1999, Global stock markets in the twentieth century, *Journal of Finance* 54, 953-980.

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. and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169-203.

Kothari, S.P. and Jay Shanken, 2001, Anomalies and Asset Allocation, Draft of monograph prepared for the Association for Investment Management and Research (AIMR).

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541-1578.

Leamer, E., 1978, *Specification Searches: Ad-Hoc Inference with Non-experimental Data*, John Wiley, New York.

Lewellen, Jonathan, 2001, Predicting returns with financial ratios, Working Paper, MIT.

Lewellen, Jonathan and Jay Shanken, 2001, Learning, asset-pricing tests, and market efficiency, *Journal of Finance*, forthcoming.

Lo, Andrew W. and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studie*s 3, 431-467.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3-28.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics*, 8, 323-361.

Nelson, Charles R. and Myung J. Kim, 1993, Predictable stock returns: the role of small sample bias, Journal of Finance 48, 641-661.

Rozeff, M., 1984, Dividend yields are equity risk premiums, *Journal of Portfolio Management* 11, 68-75.

Schwert, G. William, 1990, Indexes of U.S. Stock Prices from 1802-1987, *Journal of Business* 63, 399-442.

Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance*, 44 1115-1153.

Shanken, Jay, 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics* 19, 195-215.

Shanken, Jay, 1990, Intertemporal asset pricing: an empirical investigation, *Journal of Econometrics* 45, 99-120.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1-33.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375-421.

Tamayo, Ane, 2001, Stock return predictability, conditional asset pricing models and portfolio selection, Working Paper, University of Rochester.

Whitelaw, Robert, 1994, Time variations and covariations in the expectation and volatility of stock market returns, *Journal of Finance* 49, 515-541.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics,* John Wiley and Sons, New York, NY.

Table I: Descriptive Statistics

Excess Return and Dividend Yield on the NYSE value-weighted index

This table presents the descriptive statistics for the continuously compounded excess return and dividend yield on the NYSE value-weighted index. The dividend yield is computed as the sum of the dividends paid over the prior twelve months divided by the price at the beginning of the forecasting period.

	Monthly Excess Return			D/P	
	Mean	**Standard Deviation**	**Sharpe Ratio**	**Mean**	**Standard Deviation**
1927-1999	0.54	5.46	0.10	4.18	1.43
1927-1939	0.17	9.37	0.02	4.91	1.81
1940-1999	0.62	4.17	0.15	4.02	1.29
1960-1999	0.43	4.27	0.10	3.48	0.95
1960-1994	0.31	4.31	0.07	3.69	0.81

Table 2: OLS and GMM Estimates

This table reports the Ordinary Least Squares and Generalized Method of Moments parameter estimates from regressions that examine the predictive ability of the dividend yield. r_{t+1} is the (excess) return on the CRSP value-weighted stock index and x_t is the NYSE dividend yield, measured as deviation from its mean. White adjusted standard errors in parentheses.

Columns 2-3 contain OLS estimates of the simple relation between expected return and yield. α is the expected return when yield is at its mean and β measures the (total) impact of yield on expected return:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1} . \quad (1)$$

Columns 4-8 contain GMM estimates for the regression equation

$$r_{t+1} = \frac{\gamma}{c} \sigma_{\varepsilon t}{}^2 + \beta_m x_t + \varepsilon_{t+1} , \quad (2)$$

where $\varepsilon_{t+1} \mid x_t \sim N(0, \sigma_{\varepsilon t})$ and $\sigma_{\varepsilon t} = c * \exp(\lambda x_t)$. The slope β_m measures the impact of yield on expected excess return, net of its impact through $\sigma_{\varepsilon t}$. In this specification, c is the standard deviation, γ the Sharpe ratio, and γc the expected excess return, all when yield is at its mean.

δ measures the impact of yield on the Sharpe ratio as the average change in the ratio for a 1% increase or decrease in yield from its mean.

	α (%)	β	c (%)	λ	γ	β_M	δ
1927-1999	0.54**	0.16	5.02**	17.1**	0.09**	-0.14	-1.2
	(0.18)	(0.24)	(0.22)	(2.2)	(0.03)	(0.22)	(4.4)
1940-1999	0.62**	0.25**	4.15**	2.1	0.15**	0.23*	5.8*
	(0.15)	(0.12)	(0.19)	(2.9)	(0.04)	(0.14)	(3.0)
1960-1999	0.43**	0.19	4.21**	11.0*	0.10**	0.10	3.4
	(0.19)	(0.23)	(0.23)	(5.6)	(0.05)	(0.25)	(5.6)
1960-1994	0.31	0.56*	4.20**	15.8**	0.07	0.46*	12.1*
	(0.21)	(0.28)	(0.25)	(6.1)	(0.05)	(0.29)	(6.6)

* and ** indicate estimates more than 1.5 or 2.0 standard errors from zero, respectively.

Table 3: Log-Dividend Yield Equation

This table reports GMM parameter estimates for the log dividend yield model,

$$\log(yield_{t+1}) = \varphi + \rho \cdot \log(yield_t) + \phi \cdot \varepsilon_{t+1} + v_{t+1},$$

where $v_{t+1} | x_t \sim N(0, \sigma_v)$, and ε_{t+1} is the unexpected return. The GMM estimation is carried out jointly with equations (1) and (2) of Table 2. White adjusted standard errors in parentheses.

	φ	**ρ**	**ϕ**	**Adj(R^2)**
1927-1999	-0.031 (0.032)	0.991 (0.010)	-0.987 (0.018)	0.997
1940-1999	-0.017 (0.016)	0.995 (0.005)	-1.010 (0.013)	0.998
1960-1999	-0.022 (0.027)	0.994 (0.008)	-0.998 (0.011)	0.998
1960-1994	-0.073 (0.038)	0.978 (0.011)	-0.995 (0.012)	0.997

Table 4: Bayesian Estimates: 1940 – 1999

This table reports Bayesian estimates for the regression

$$r_{t+1} = \frac{\gamma}{c}\sigma_{\varepsilon t}^{2} + \beta_m x_t + \varepsilon_{t+1},$$

where r_{t+1} is the excess return on the CRSP value-weighted stock index and x_t is the NYSE dividend yield, measured as deviation from its mean. The conditional standard deviation of return is $\sigma_{\varepsilon t} = c * \exp(\lambda x_t)$ and unexpected return is distributed as $\varepsilon_{t+1} | x_t \sim N(0, \sigma_{\varepsilon t})$.

β_m measures the impact of yield on expected excess return, net of its impact through $\sigma_{\varepsilon t}$; the total impact of yield on expected return is given by β; γ is the long-run (when yield is at its mean) Sharpe ratio and c is the long-run standard deviation of return; $\gamma c = \alpha$ is the long-run expected excess return. Finally, δ measures the impact of yield on the Sharpe ratio.

Means and standard deviations (in parentheses) of the posterior distributions of the parameters are reported. The probability that the given parameter is larger than zero is given in brackets. The Bayes factor is the expected likelihood conditional on the parameter of interest equaling zero divided by the unconditional expected likelihood for the given prior.

	α (%)	β	c (%)	λ	γ	β_M	δ
Diffuse							
Posterior							
Mean	0.61	0.21	4.16	2.3	0.15	0.18	4.7
(Stdev)	(0.16)	(0.11)	(0.11)	(2.1)	(0.04)	(0.11)	(2.7)
[Prob(*>0)]	[1.00]	[0.97]	[1.00]	[0.86]	[1.00]	[0.95]	[0.96]
Neutral							
Prior							
Mean	0.67	0.34	6.00	20.0	0.11	0.05	3.1
(Stdev)	(0.32)	(0.35)	(1.00)	(15.0)	(0.05)	(0.20)	(4.2)
[Prob(*>0)]	[0.99]	[0.85]	[1.00]	[0.91]	[0.99]	[0.60]	[0.78]
Posterior							
Mean	0.56	0.18	4.20	2.8	0.13	0.14	3.8
(Stdev)	(0.13)	(0.10)	(0.11)	(2.1)	(0.03)	(0.10)	(2.3)
[Prob(*>0)]	[1.00]	[0.96]	[1.00]	[0.91]	[1.00]	[0.93]	[0.95]
Bayes Factor				7.05		0.72	
Overreaction							
Prior							
Mean	0.67	0.49	6.00	20.0	0.11	0.20	5.8
(Stdev)	(0.32)	(0.32)	(1.00)	(15.0)	(0.05)	(0.15)	(3.5)
[Prob(*>0)]	[0.99]	[0.96]	[1.00]	[0.91]	[0.99]	[0.91]	[0.96]
Posterior							
Mean	0.56	0.21	4.20	2.6	0.13	0.18	4.8
(Stdev)	(0.13)	(0.09)	(0.11)	(2.1)	(0.03)	(0.09)	(2.1)
[Prob(*>0)]	[1.00]	[0.99]	[1.00]	[0.90]	[1.00]	[0.98]	[0.99]
Bayes Factor				8.09		0.51	

Table 5: Predictive Analysis: 1940 – 1999

This table reports predictive return moments for the market index, the optimal allocations to stock (weight), and the associated certainty equivalent return premia (CER) for each prior and level of yield, the latter in sample standard deviations from the mean.

"True" refers to the predictive distribution for the model in Table 6. In contrast, "μ" and "σ," refer to "optimizations" incorrectly based on predictive distributions that ignore variation in expected return or risk, respectively (μσ ignores both). "Ignore" means that the predictive mean or variance is taken to be the value for yield = 0. Prior weights are computed from optimization based on the prior predictive distribution.

The CERs in columns 8-11 are computed under the true (posterior) predictive distribution and are based on the weights from 4-7. Actual CERs are reported for the true posterior, whereas the certainty equivalent *loss* from the use of suboptimal weights is reported for the other scenarios. For μ, σ, and μσ, the change in weights and CER loss are measures of the economic significance of expected return and/or risk predictability and reflect the combination of prior and data. For "prior," these measures reflect the *incremental* significance of the data.

PRIOR / **Yield**	**Mean (Stdev) (%)**		**WEIGHT (%)**				**CER (bps)**			
	Prior	**Posterior**	**Prior**	**Posterior** μ	**True**	σ	**Prior**	**Posterior** μ	**μσ (True)**	σ
Diffuse										
-1.5	-	0.21 (4.00)	-	87	37	35	-	10	7 (5)	0
0	-	0.61 (4.16)	-	81	81	81	-	0	0 (28)	0
1.5	-	1.02 (4.36)	-	74	100	100	-	7	5 (64)	0
Neutral										
-1.5	0.27 (4.51)	0.23 (3.99)	36	80	38	36	0	7	5 (6)	0
0	0.66 (6.08)	0.56 (4.20)	46	73	73	73	3	0	0 (24)	0
1.5	1.80 (9.84)	0.90 (4.45)	47	67	100	100	14	6	4 (51)	0
Overreact										
-1.5	-0.02 (4.51)	0.15 (4.01)	8	80	29	27	1	10	8 (3)	0
0	0.66 (6.08)	0.56 (4.20)	46	73	73	73	3	0	0 (24)	0
1.5	2.09 (9.83)	0.98 (4.43)	53	67	100	100	15	9	6 (59)	0

Table 6: Bayesian Estimates: 1960 – 1994

This table reports Bayesian estimates for the regression

$$r_{t+1} = \frac{\gamma}{c}\sigma_{\varepsilon t}{}^{2} + \beta_m x_t + \varepsilon_{t+1},$$

where r_{t+1} is the excess return on the CRSP value-weighted stock index and x_t is the NYSE dividend yield, measured as deviation from its mean. The conditional standard deviation of return is $\sigma_{\varepsilon t} = c * \exp(\lambda x_t)$ and unexpected return is distributed as $\varepsilon_{t+1} | x_t \sim N(0, \sigma_{\varepsilon t})$.

β_m measures the impact of yield on expected excess return, net of its impact through $\sigma_{\varepsilon t}$; the total impact of yield on expected return is given by β; γ is the long-run (when yield is at its mean) Sharpe ratio and c is the long-run standard deviation of return; $\gamma c = \alpha$ is the long-run expected excess return. Finally, δ measures the impact of yield on the Sharpe ratio.

Means and standard deviations (in parentheses) of the posterior distributions of the parameters are reported. The probability that the given parameter is larger than zero is given in brackets. The Bayes factor is the expected likelihood conditional on the parameter of interest equaling zero divided by the unconditional expected likelihood for the given prior.

	α (%)	β	c (%)	λ	γ	β_M	δ
Diffuse							
Posterior							
Mean	0.26	0.47	4.22	16.4	0.06	0.37	10.3
(Stdev)	(0.20)	(0.27)	(0.15)	(4.3)	(0.05)	(0.26)	(6.3)
[Prob(*>0)]	[0.90]	[0.96]	[1.00]	[1.00]	[0.90]	[0.93]	[0.95]
Neutral							
Prior							
Mean	0.50	0.26	4.50	20.0	0.11	0.05	3.4
(Stdev)	(0.23)	(0.29)	(0.50)	(15.0)	(0.05)	(0.20)	(5.1)
[Prob(*>0)]	[0.99]	[0.83]	[1.00]	[0.91]	[0.99]	[0.60]	[0.75]
Posterior							
Mean	0.36	0.29	4.24	16.8	0.08	0.17	5.5
(Stdev)	(0.15)	(0.17)	(0.14)	(4.1)	(0.03)	(0.16)	(3.9)
[Prob(*>0)]	[0.99]	[0.96]	[1.00]	[1.00]	[0.99]	[0.86]	[0.92]
Bayes Factor				0.001		0.730	
Overreaction							
Prior							
Mean	0.50	0.41	4.50	20.0	0.11	0.20	6.9
(Stdev)	(0.23)	(0.26)	(0.50)	(15.0)	(0.05)	(0.15)	(4.2)
[Prob(*>0)]	[0.99]	[0.96]	[1.00]	[0.91]	[0.99]	[0.91]	[0.96]
Posterior							
Mean	0.36	0.36	4.24	16.7	0.08	0.24	7.3
(Stdev)	(0.15)	(0.14)	(0.14)	(4.1)	(0.03)	(0.13)	(3.2)
[Prob(*>0)]	[0.99]	[1.00]	[1.00]	[1.00]	[0.99]	[0.97]	[0.99]
Bayes Factor				0.001		0.486	

Table 7: Predictive Analysis: 1960 – 1994

This table reports predictive return moments for the market index, the optimal allocations to stock (weight), and the associated certainty equivalent return premia (CER) for each prior and level of yield, the latter in sample standard deviations from the mean.

"True" refers to the predictive distribution for the model in Table 4. In contrast, "μ" and "σ," refer to "optimizations" incorrectly based on predictive distributions that ignore variation in expected return or risk, respectively (μσ ignores both). "Ignore" means that the predictive mean or variance is taken to be the value for yield = 0. Prior weights are computed from optimization based on the prior predictive distribution.

The CERs in columns 8-11 are computed under the true (posterior) predictive distribution and are based on the weights from 4-7. Actual CERs are reported for the true posterior, whereas the certainty equivalent *loss* from the use of suboptimal weights is reported for the other scenarios. For μ, σ, and μσ, the change in weights and CER loss are measures of the economic significance of expected return and/or risk predictability and reflect the combination of prior and data. For "prior," these measures reflect the *incremental* significance of the data.

PRIOR	**Mean (Stdev) (%)**		**WEIGHT (%)**				**CER (bps)**			
Yield	**Prior**	**Posterior**	**Prior**	**Posterior** μ	**True**	σ	**Prior**	**Posterior** μ	μσ **(True)**	σ
Diffuse										
-1.5	-	-0.30 (3.48)	-	53	0	0	-	21	14 (0)	0
0	-	0.26 (4.23)	-	39	39	39	-	0	0 (7)	0
1.5	-	0.86 (5.19)	-	29	73	100	-	14	8 (37)	5
Neutral										
-1.5	0.27 (3.68)	0.03 (3.47)	49	70	15	14	3	9	4 (1)	0
0	0.50 (4.53)	0.36 (4.24)	59	50	50	50	0	0	0 (11)	0
1.5	0.93 (6.00)	0.75 (5.23)	62	36	65	93	0	6	2 (29)	6
Overreact										
-1.5	0.08 (3.67)	-0.06 (3.47)	22	70	1	4	1	14	7 (0)	0
0	0.50 (4.53)	0.36 (4.24)	59	50	50	50	0	0	0 (11)	0
1.5	1.11 (6.00)	0.84 (5.21)	72	36	72	100	0	8	3 (35)	6

Table 8: Bayesian Estimates: 1960 – 1999

This table reports Bayesian estimates for the regression

$$r_{t+1} = \frac{\gamma}{c}\sigma_{\varepsilon t}{}^2 + \beta_m x_t + \varepsilon_{t+1},$$

where r_{t+1} is the excess return on the CRSP value-weighted stock index and x_t is the NYSE dividend yield, measured as deviation from its mean. The conditional standard deviation of return is $\sigma_{\varepsilon t} = c * \exp(\lambda x_t)$ and unexpected return is distributed as $\varepsilon_{t+1} | x_t \sim N(0, \sigma_{\varepsilon t})$.

β_m measures the impact of yield on expected excess return, net of its impact through $\sigma_{\varepsilon t}$; the total impact of yield on expected return is given by β; γ is the long-run (when yield is at its mean) Sharpe ratio and c is the long-run standard deviation of return; $\gamma c = \alpha$ is the long-run expected excess return. Finally, δ measures the impact of yield on the Sharpe ratio.

Means and standard deviations (in parentheses) of the posterior distributions of the parameters are reported. The probability that the given parameter is larger than zero is given in brackets. The Bayes factor is the expected likelihood conditional on the parameter of interest equaling zero divided by the unconditional expected likelihood for the given prior.

	α (%)	β	c (%)	λ	γ	β_M	δ
Diffuse							
Posterior							
Mean	0.43	0.07	4.32	9.4	0.10	-0.01	0.7
(Stdev)	(0.20)	(0.18)	(0.14)	(3.1)	(0.05)	(0.18)	(4.2)
[Prob(*>0)]	[0.98]	[0.65]	[1.00]	[1.00]	[0.98]	[0.48]	[0.56]
Neutral							
Prior							
Mean	0.50	0.26	4.50	20.0	0.11	0.05	3.4
(Stdev)	(0.23)	(0.29)	(0.50)	(15.0)	(0.05)	(0.20)	(5.1)
[Prob(*>0)]	[0.99]	[0.83]	[1.00]	[0.91]	[0.99]	[0.60]	[0.75]
Posterior							
Mean	0.45	0.10	4.33	9.7	0.10	0.01	1.3
(Stdev)	(0.15)	(0.14)	(0.14)	(3.0)	(0.03)	(0.13)	(3.1)
[Prob(*>0)]	[1.00]	[0.77]	[1.00]	[1.00]	[1.00]	[0.54]	[0.66]
Bayes Factor				0.06		1.54	
Overreaction							
Prior							
Mean	0.50	0.41	4.50	20.0	0.11	0.20	6.9
(Stdev)	(0.23)	(0.26)	(0.50)	(15.0)	(0.05)	(0.15)	(4.2)
[Prob(*>0)]	[0.99]	[0.96]	[1.00]	[0.91]	[0.99]	[0.91]	[0.96]
Posterior							
Mean	0.46	0.20	4.33	9.3	0.11	0.11	3.6
(Stdev)	(0.15)	(0.12)	(0.14)	(2.9)	(0.03)	(0.12)	2.7
[Prob(*>0)]	[1.00]	[0.95]	[1.00]	[1.00]	[1.00]	[0.83]	[0.9]
Bayes Factor				0.09		2.00	

Table 9: Predictive Analysis: 1960 – 1999

This table reports predictive return moments for the market index, the optimal allocations to stock (weight), and the associated certainty equivalent return premia (CER) for each prior and level of yield, the latter in sample standard deviations from the mean.

"True" refers to the predictive distribution for the model in Table 5. In contrast, "μ" and "σ," refer to "optimizations" incorrectly based on predictive distributions that ignore variation in expected return or risk, respectively (μσ ignores both). "Ignore" means that the predictive mean or variance is taken to be the value for yield = 0. Prior weights are computed from optimization based on the prior predictive distribution.

The CERs in columns 8-11 are computed under the true (posterior) predictive distribution and are based on the weights from 4-7. Actual CERs are reported for the true posterior, whereas the certainty equivalent *loss* from the use of suboptimal weights is reported for the other scenarios. For μ, σ, and μσ, the change in weights and CER loss are measures of the economic significance of expected return and/or risk predictability and reflect the combination of prior and data. For "prior," these measures reflect the *incremental* significance of the data.

PRIOR	**Mean (Stdev) (%)**		**WEIGHT (%)**				**CER (bps)**			
Yield	**Prior**	**Posterior**	**Prior**	**Posterior** μ	**True**	σ	**Prior**	**Posterior** μ	μσ **(True)**	σ
Diffuse										
-1.5	-	0.35 (3.86)	-	67	57	48	-	0	0 (12)	0
0	-	0.43 (4.33)	-	56	56	56	-	0	0 (15)	0
1.5	-	0.52 (4.87)	-	46	54	66	-	0	0 (17)	0
Neutral										
-1.5	0.27 (3.68)	0.34 (3.86)	49	71	56	46	0	1	0 (12)	1
0	0.50 (4.53)	0.45 (4.33)	59	58	58	58	0	0	0 (16)	0
1.5	0.93 (6.00)	0.59 (4.89)	62	48	59	73	0	1	0 (21)	1
Overreact										
-1.5	0.08 (3.67)	0.23 (3.87)	22	71	41	35	1	3	1 (6)	0
0	0.50 (4.53)	0.46 (4.33)	59	59	59	59	0	0	0 (16)	0
1.5	1.11 (6.00)	0.71 (4.86)	72	49	70	86	0	2	0 (29)	1

Figure 1a: NYSE Value-Weighted Excess Returns

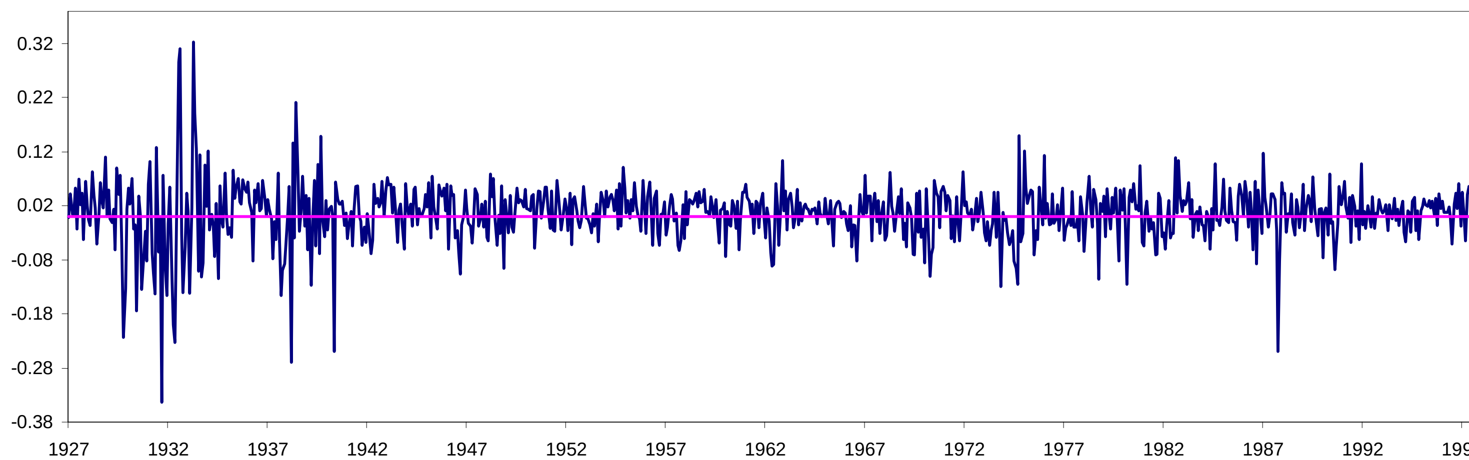

Figure 1b: NYSE Value-Weighted Squared (Deviation from the Mean) Excess Returns