USC FBE FINANCE SEMINAR
presented by Jay Shanken
FRIDAY, March 5, 2004
10:30 am - 12:00 pm, Room: **JKP-202**

Dividend Yield, Risk, and Mispricing: A Bayesian Analysis*

Jay Shanken

Goizueta Business School
Emory University
1300 Clifton Road
Atlanta, GA 30322
(404) 727-4772
jay_shanken@bus.emory.edu

Ane Tamayo

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

March 2, 2004

Preliminary and incomplete draft. For USC internal use only.

*This is a completely revamped version of an earlier paper entitled "Risk, Mispricing, and Asset Allocation: Conditioning on Dividend Yield." A broader set of issues is now considered in the context of a richer model specification. 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.

Dividend Yield, Risk, and Mispricing: A Bayesian Analysis

Abstract

In the asset pricing literature, time-variation in market expected excess return tracked by financial ratios like dividend yield is typically attributed either to 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. Our investor is also uncertain about the tradeoff between risk and the market risk premium and uses Merton's (1980) proportionality condition as a prior reference point. We find that predictability in monthly market risk is statistically and economically important, but most of the yield related-variation in expected return is unrelated to risk. Differences in beliefs about mispricing and the risk-return tradeoff can have smaller but still "significant" effects on the utility of an investor allocating funds across a market index and riskless T-bills.

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). In addition, conventional inference for predictive regressions can be misleading due to small-sample biases (Stambaugh, 1999), 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 (2003) has recently developed an innovative 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. Furthermore, Whitelaw (1994) alludes to unreported evidence of "strongly significant" *simple* relations between default spreads, dividend yields, and stock volatility for the period 1953-89.

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 expected return and volatility should be positively related. For example, Merton (1980) considers a model in which expected excess returns on the market are proportional to volatility.

---

[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).

However, a positive relation between variance and the market risk premium is not a theoretical necessity, even in a rational market with perfect information, as argued by Merton (1980) and 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.

Stock return predictability has more recently 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] As KS recognize, if market risk increases with yield, the optimal allocation to stock might be less aggressive than they observe, or even decrease with yield.

---

[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).

Our paper extends this Bayesian literature by allowing for time variation in both market risk and expected return.[5] This poses substantial computational challenges that are met through the use of a statistical technique called importance sampling. To incorporate the well-known persistence in volatility in a parsimonious manner, a lagged measure of realized within-month daily volatility, similar to that used by French, Schwert and Stambaugh (1987), is included as a predictive variable. Time-varing market variance is then modeled as a function of lagged volatility and dividend yield. Lagged volatility tracks persistence similar to that often modeled through ARCH terms, while the more highly autocorrelated yield might track variation akin to a GARCH term.[6]

Another contribution of our paper is a careful modeling of *informative* priors in a manner that goes a long way toward reflecting some of the important economic uncertainties that confront an investor or a researcher when thinking about predictability issues. The KS paper focuses on diffuse priors, a convenient starting point for such research in that it permits them to draw upon standard analytical results for the homoskedastic vector-auto-regression model.[7] Again, the use of importance sampling makes it possible for us to accommodate more detailed priors in a heteroskedastic environment. A related novel feature of our study is a comparison of predictive return distributions and asset allocations based on posterior beliefs with those based on the prior, allowing us to identify the incremental significance of the data in the presence of informative priors.

Perhaps the most important consideration in evaluating the implications of predictability for asset allocation is the economic source of that predictability. Classical models attribute variation in expected return to changes in risk. That variation could be tracked by financial ratios like dividend yield if increases in risk simultaneously lower the current price and raise the yield through a discount rate effect (e.g., Fama and French,

---

[5] A recent paper by Pastor and Stambaugh (2001) allows for changes in volatility across different regimes, rather than as a function of observable predictive variables. 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.

[6] Wang (2002) models conditional variance as an ARCH(1) process while Johannes, Polson, and Stroud (2002) model expected return and variance as mean-reverting processes, but do not condition on observable predictive variables. Neither paper includes variance in the expected return relation.

1988). However, a central premise of behavioral finance is that expected returns on financial securities can also be influenced by irrational mispricing. In particular, price might be *temporarily* depressed in relation to expected future dividends (proxied by recent dividends) due to negative irrational investor sentiment. Expected future return would then be high if the sentiment is likely to mean-revert over the given horizon. It is evident from the literature that there is a wide range of beliefs about the relative importance of these factors 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? To address these issues, we specify a model in which expected excess market return is a linear function of the market's conditional variance and dividend yield. Thus, expected return can vary with yield directly, or indirectly, through the relation between yield and market risk. Insofar as our control for "risk" is adequate, the coefficient on dividend yield is an indication of mispricing. We consider several priors on the magnitude of the yield-related mispricing effect, linking it to beliefs about behavioral hypotheses like overreaction or underreaction to information.

As noted above, even in the absence of mispricing effects, financial theory leaves some doubt about the relation between conditional volatility and expected return. It is also important, therefore, to incorporate uncertainty about this relation in the Bayesian analysis. Here, the Merton proportionality condition would seem to be a natural reference point *a priori*, as in the early papers by Merton (1980) and French, Schwert and Stambaugh (1987). Able (1988) notes that in the context of his model, "for some allowable, though implausible, parameter values, the risk premium is a decreasing function of risk." In this spirit, we consider a prior centered on the Merton restriction, but with a standard deviation such that the prior probability of a negative relation between variance and expected excess return is fairly low. We also entertain dogmatic

---

[7] They also provide an initial exploration of informative priors in the following manner. They take as their informative prior, a posterior distribution obtained by updating a diffuse prior with hypothetical data of a given length.

priors in which the Merton condition is believed to hold exactly, or, at the other extreme, the link between volatility and expected return is ignored entirely.

The paper is organized as follows. First, descriptive statistics and preliminary regression evidence are presented in Section I. Section II describes the specification of our model of risk and return. Section III provides an overview of the Bayesian methodology while Section IV discusses the choice of prior distributions. Empirical evidence on our risk/return parameters is analyzed in Section V and the impact of different prior beliefs on predictive return moments and asset allocation is explored in Section VI. Section VII summarizes our findings and considers directions for future research.

I. Preliminary Evidence

Before developing the formal econometric specification and conducting the Bayesian analysis, we look at some descriptive statistics and regression evidence that summarizes the empirical relations among returns, risk, and dividend yield. This will give us an initial feel for the data before exploring the impact of prior beliefs in a richer context.

A. Descriptive statistics

We begin by looking at the means and standard deviations of continuously compounded excess returns, dividend yields, and the within-month sample standard deviation of daily returns over the period 1926-02 and various subperiods. The standard deviation is scaled by the square root of 22 (trading days) so as to be a measure of risk over a monthly horizon.[8] 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.[9] As in Fama and French (1988, 1989), the dividend yield is

---

[8] The measure is also scaled up by a factor of 1.13 to account for positive autocorrelation in daily returns due to nonsynchroneities in individual stock prices. This factor is the square root of the ratio of the autocorrelation-adjusted mean monthly variance estimate to the unadjusted mean estimate. The adjustment follows French, Schwert, and Stambaugh (1987) and the means are computed over the period of July 1962 through December 2002.

[9] Since daily data for the NYSE index are not available prior to July 1962, we do the following. First, we regress daily NYSE index returns on those of the S&P 500 index over the period July 1962 through

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 return, r_t, dividend yield dp_t, and scaled standard deviation, sd_t, are provided in Table I. Figures 1a and 1b (next draft) plot the monthly series of excess returns and standard deviations over the entire period.[10] At nearly 30% per month, the ex post risk in October 1987 stands out as very extreme by historical standards. When using sd as a predictive variable, we replace this observation by 20%, a sort of Winsorization that implicitly assumes investors expected more mean reversion in volatility than our simple model would otherwise imply for that month.

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 16 bps per month for the 1927-1939 period was far below the mean of 48 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 1940-2002 period lie within these ranges.

Our empirical analysis will focus on the relatively stationary period since 1940, with some 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 (2003) have argued that the predictive power of dividend yield has declined over time, as has the level of yield itself, which is apparent

---

December 2002. The series are nearly perfectly correlated over this period (correlation equals 0.996). The regression coefficients are then used to impute daily NYSE returns from the S&P 500 returns over the earlier period. We are grateful to Bill Schwert for providing the daily S&P 500 data.

[10] The averages of the monthly standard deviation estimates are lower than the standard deviations of monthly returns for several reasons. First, there is a Jensen's inequality effect (the square root function is concave). Second, the monthly estimates correspond to conditional moments while the standard deviation of monthly returns estimates the unconditional moment; the unconditional variance is the expected conditional variance plus the variance of the conditional expected returns.

in Table I. A formal modeling of changing parameters, while of considerable interest, is beyond the scope of this paper.

B. Preliminary regression evidence

We now look at empirical results for a linear predictive regression,

$$r_{t+1} = \beta_0 + \beta_1 dp_t + \beta_2 sd_t^2 + \text{error}, \qquad (1)$$

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, dp_t is the dividend yield at time t, sd_t^2 is the variance estimate for month t, and ε_{t+1} is a disturbance term.[11] The slope parameters β_1 and β_2 measure the extent to which expected return varies with dividend yield or lagged variance, respectively. Henceforth, "return" always refers to continuously compounded excess return.

Regression results for the period 1940-2002 are presented in the first row of panel A of Table II. The estimate of β_1, the coefficient on yield is positive and appears to be statistically significant based on the heteroskedasticity-adjusted standard error. Given the small-sample problems documented by Stambaugh (1999), however, we view these statistics as merely suggestive. The coefficient on the risk measure is negative, but small in relation to the standard error.

In order to gain some insight into the time-series properties of our risk measure, we also consider the regression

$$\log(sd_{t+1}) = \delta_0 + \delta_1 \log(sd_t) + \delta_2 dp_t + \text{error}. \qquad (2)$$

Taking logs makes the distribution of the risk variable more symmetric and is consistent with the formal model of the next section. Results are given in panel B of Table II. Yield has no explanatory power. Strong persistence in volatility is apparent in the highly significant positive slope coefficient on the lagged value, and the adjusted R^2 is 42%. The coefficient is well below one, reflecting mean-reversion in risk. Thus, sd_t by itself is

[11] Taking logs makes the risk estimates less skewed and is also convenient for reasons that will be apparent later on.

a crude proxy for ex ante risk. A more appealing specification is developed in the next section.

Finally, in panel C, we report the simple contemporaneous "auxiliary" regression of risk on yield,

$$\log(sd_t) = \delta_3 + \delta_4 dp_t + \text{error}. \qquad (3)$$

There is effectively no relation over the period 1940-2002. We have also examined subperiods and find that the relation between yield and risk is positive and significant up through the mid-1990s. In the Bayesian framework, it will be interesting to explore the extent to which this link would have affected investors' perceptions about the conditional distribution of returns at that point in time. Therefore, we also present results for the period 1960-1994, over which risk and and yield were particularly strongly positively associated, as seen in the second row of Panel C. In practice, investors may discount data in the more distant past due to concerns about temporal instability of the various relations. Examining subperiod results may also provide some insight into the potential importance of such considerations.

Returning to Panel A, we see that for 1960-1994, the slope coefficient on yield in the expected return relation is more than twice that for the full period, though the standard error increases more than proportionally.[12] The coefficient on risk is more negative now, but just 1.6 standard errors below zero. Finally, Panel B shows that our risk estimates display persistence similar to that over the full period, though yield now appears to have some positive incremental value in forecasting risk.

II. The Model of Risk and Return

As in Merton (1980) and French, Schwert, and Stambaugh (1987), we allow the market risk premium to vary directly with the ex ante variance.[13] In addition, we let the expected return depend separately on dividend yield:

$$r_{t+1} = k_0 + k_1\sigma_{\varepsilon t}^2 + \beta_m dp_t + \varepsilon_{t+1}, \qquad (4)$$

[12] This is consistent with Lewellen's (2003) observations about the effect of the low-yield, high-return years of the late 90's, which are excluded.

[13] Future research might consider measures of covariance risk with respect to aggregate consumption.

where ε_{t+1} is the unexpected return and $\sigma_{\varepsilon t}$ is its standard deviation conditional on information at time t. The coefficient k_1 measures the extent to which changes in risk affect expected return, while β_m reflects any additional yield-related variation in expected return that is not captured by the risk measure.[14]

The subscript on β_m indicates that, relative to a benchmark in which the expected market return is driven (linearly) by risk, the remaining component would be attributed to "mispricing," an alternative that is often considered in the literature. The need for a joint hypothesis of this sort can be likened to Fama's (1970) well-known dictum that tests of market efficiency always entail a joint hypothesis about the equilibrium pricing process. The point here is not to take the mispricing interpretation too literally, but rather to provide a simple benchmark for thinking about what might constitute yield-related mispricing. In a similar spirit, MacKinlay (1995) considers the plausibility of estimated Sharpe ratios in assessing the rationality of cross-sectional size and book-to-market expected return effects.[15] The exact benchmark for what a Sharpe ratio should be in a rational world is hardly clear, but the exercise is informative nonetheless.

In Section I, we considered a regression of our risk estimate on it's lagged value and dividend yield. Here we do something similar, relating the conditional *ex ante* standard deviation to the predictive variables through a nonlinear relation. The model for conditional standard deviation at time t is:

$$\sigma_{\varepsilon t} = \text{stdev}(\varepsilon_{t+1} \mid x_t) = c \cdot \exp[\lambda_1 \log(sd_t) + \lambda_2 dp_t], \qquad (5)$$

where $x_t \equiv [\log(sd_t),\ dp_t]'$. This exponential specification ensures that the standard deviation is always positive and amounts to assuming that $\log(\sigma_{\varepsilon t})$ is linear in dp_t and $\log(sd_t)$.[16] Inclusion of the lagged ex post volatility estimate, sd, is intended to capture the well-known short-term persistence in volatility observed in the estimation of ARCH models. Dividend yield is more highly autocorrelated than sd and may track slowly changing components of risk as well as leverage effects impounded in market price.

---

[14] Related work by Attanasio (1991) considers the relation between several predictive variables and expected return after controlling for the effect on return volatility.
[15] See related discussion of Sharpe ratios and approximate arbitrage in Shanken (1992).
[16] Previous linear specifications for squared residuals (Shanken, 1990 and Campbell, 1987?) or absolute residuals (Schwert, 1989) do not impose positivity.

Since dividend yield and variance are always positive, it is hard to develop intuition about the intercept k_0 in (4), or c in (5). This will be important later when specifying prior beliefs and interpreting estimates. If we work with the independent variables measured as deviations from some fixed reference points, then k_0 and c can be viewed as the conditional moments when yield and risk are at these levels. In our empirical analysis, dp will be the deviation of yield from it's sample mean and log(sd) the deviation from the log of the average standard deviation estimate.[17] Henceforth, we use the same variable names to refer to the transformed deviations, with x_t in deviation form as well.

The parameter c in (5) can now be interpreted as the "long-run" standard deviation of return or, more specifically, the value of $\sigma_{\varepsilon t}$ conditional on $x_t = 0$. Time variation in risk depends on the parameters $\lambda \equiv (\lambda_1, \lambda_2)$. If $\lambda = 0$, the error term is homoskedastic. Note that if dp_t is related to $\log(sd_t)$, as we saw in Table II for the 1960-1994 subperiod, the *total* effect of yield on risk depends not only on λ_2, but also on λ_1 and the auxiliary regression coefficient δ_4 in equation (3). Thus, yield may covary with risk even if λ_2 is zero. In such a case, the comovement is already reflected in the lagged volatility measure and there is no *incremental* effect of yield.

By (4), the "long-run" expected market return $E(r_{t+1} \mid x_t=0)$, denoted by α, can be expressed as $k_0 + k$, where $k \equiv k_1c^2$. For ease of interpretation, we focus on k, which is measured in units of expected return, rather than k_1. Equation (4) can then be rewritten as

$$r_{t+1} = (\alpha - k) + k(\sigma_{\varepsilon t}^2/c^2) + \beta_m dp_t + \varepsilon_{t+1}. \tag{6}$$

If expected return is proportional to variance then, in the absence of mispricing, $k = \alpha$. If there is no relation between expected return and risk, $k = 0$. We note that, in general, the impact of behavioral mispricing need not be confined to β_m. For example, expected return might be weakly related to risk ($k << \alpha$) because uninformed investors sometimes fail to perceive an actual increase in risk.

[17] There is flexibility in the choice of the reference point. We use this, rather than the average of the log(sd), because the resulting "long-run" standard deviation is closer to the unconditional estimate.

Harvey (1989) rejects the hypothesis that the market price of risk and the Sharpe ratio are constant. We let γ denote the long-run Sharpe ratio γ/c. Using (4) and (5), in the absence of mispricing ($\beta_m=0$), the market's Sharpe ratio is $k_0/\sigma_{\varepsilon t} + k_1\sigma_{\varepsilon t}$. If the proportionality condition holds ($k_0 = 0$) and $\alpha = k > 0$, then k_1 is also positive and the ratio increases with the conditional standard deviation. In general, the relation need not be monotonic and it can be decreasing in $\sigma_{\varepsilon t}$ if $k_0 > 0$ and $k_1 < 0$. There will be additional variation in the Sharpe ratio that is unrelated to risk, however, if β_m is nonzero. For example, a positive β_m implies that expected return increases with yield, beyond any effect of changing risk, tending to increase the Sharpe ratio as well.

III. Overview of the Bayesian Framework

In the standard Bayesian regression framework, the regressors are distributed independently of the disturbances, with a distribution that does not depend on the regression parameters. These conditions hold, in particular, if the regressors are nonstochastic. As KS note, however, independence is violated in a predictive regression on yield because the return surprise at time t impacts all values of dividend yield from time t forward. Correlation between the return surprise and future volatility estimates might be expected as well due to leverage or other effects. Hence, one cannot simply work with the joint density of returns conditional on the time series x_t. KS accommodate stochastic regressors by modeling return and yield as elements of a vector autoregression with homoskedastic errors, conditioning on the initial sample value of yield, and imposing a noninformative prior on the VAR parameters.[18] 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, as well as conditional heteroskedasticity in returns and yields. We also develop a method for dealing with

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

issues that arise in conditioning on lagged estimates of volatility. A general overview is provided below. Additional details are given in the appendices.[19]

A. Equations for the stochastic regressors, yield and standard deviation

A distinguishing feature of our approach to the stochastic regressor problem is that we work with the following representation of the likelihood function or joint density of returns and predictive variables:

$$f_t(r_{t+1}, x_{t+1}) = f(r_{t+1} \mid x_t)\, f(x_{t+1} \mid r_{t+1,} x_t), \tag{7}$$

where $f(r_{t+1} \mid x_t)$ is the density for the return model in (4). Consider the joint distribution of the predictive variables. Given the rapid reaction of prices and slow adjustment of subsequent dividends to new information, *percentage* changes in yield and price will tend to be similar, but of the opposite sign. Thus, changes in log(yield) should be strongly negatively related to contemporaneous returns, with a coefficient near minus one.

The density $f(dp_{t+1} \mid r_{t+1,} x_t)$ is characterized by the following regression equation:

$$\log(\text{yield}_{t+1}) = \varphi + \rho \cdot \log(\text{yield}_t) + \phi \cdot r_{t+1} + w_{t+1}, \tag{8}$$

Estimates of the model are reported in Table III. As expected, the estimate of ρ is close to one and the estimate of ϕ is close to minus one. Moreover, the adjusted R^2 is 0.999 for the full period and 0.997 for the subperiod. Insofar as the relation is nearly a mechanical identity, it is not likely to contain much information, if any, about the parameters of primary interest in (5) and (6).[20] To simplify the analysis, we formally assume that the prior for the parameters in (8) is independent of the earlier priors.

B. The ex post volatility equation

We now turn to the conditional density $f(sd_{t+1} \mid r_{t+1,} x_t)$. It is tempting to simply introduce one more regression equation for sd and assume prior independence between

---

[19] In an earlier version of the paper we noted that, under our assumptions, if risk were constant, as in KS, analysis of the return equation would proceed as if the regressor, dividend yield, were nonstochastic. This continues to be true even if risk is linear in yield, though analysis for the exponential specification is a bit more complicated.

[20] The only parameter in (7) that is not completely dominated by the data in this equation is the intercept. Modeling some sort of prior dependence for this parameter might be of interest.

the additional parameters and the primary parameters, as we did for dividend yield. However, this mechanical approach would ignore the fact that the density of this sample statistic for month t+1 depends on the ex ante level of risk at the beginning of the month. Thus, sd contains important information about the parameters of the conditional risk relation in (5). In general, characterizing this density could be quite complicated. To obtain a tractable solution to the stochastic regressor problem in this context, we model the daily returns *within* month t+1 as independent and identically normally distributed *conditional* on x_t, the state of the world at the beginning of month t+1. Although a strong assumption, the distribution *is* permitted to vary from month to month. Thus, the unconditional distribution of monthly returns, a mixture of heteroskedastic normals in our framework, should display the fat tails observed empirically [Blattberg and Gonedes (19??)].

For simplicity, assume there are 22 trading days in each month. Recalling that returns are continuously compounded, the sum of daily returns is just the monthly return r_{t+1}. Given our i.i.d. assumption conditional on x_t, the mean and variance of the daily returns in month t+1 equal the corresponding quantities for monthly returns, in (4) and (5), divided by 22. Moreover, the normality assumption implies that the sum and sample variance of daily returns are (conditionally) independent. Therefore, $f(sd_{t+1} \mid r_{t+1}, x_t)$ reduces to $f(sd_{t+1} \mid x_t)$.[21] It is convenient now to work with the transformed variable, $v_{t+1} \equiv 22sd_{t+1}^2$, and apply standard results under (conditional) normality. Thus, $f(v_{t+1} \mid x_t)$ is the density for a variable that is distributed as $\sigma_{\varepsilon t}^2/22$ (the daily variance) times a chi-square variate with 21 degrees of freedom. This provides the final element needed to derive the posterior distributions via the likelihood function in (7).

Note that the density for v_{t+1} depends on the parameters c and λ through $\sigma_{\varepsilon t}^2$, which changes over time with x_t. Stepping back from the technical details, the substantive effect of incorporating the ex post volatility equation is to provide more precise estimates of the parameters in the variance relation (5). The increased precision is achieved by exploiting the within-month variation in returns rather than simply relying on squared monthly residual returns to identify the conditional variance relation.

C. 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 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 expectation under the prior of $\theta f(D \mid \theta)$, divided by $p(D)$. Here, f is the likelihood function or density for the data D and $p(D)$ is the probability of the data under the prior, i.e., the expectation under the prior of the likelihood function evaluated at the given data. The expectation in the numerator can be obtained, via simulation, by repeatedly drawing values of θ from the prior density and averaging the products, $\theta f(D \mid \theta)$. 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 products will be extremely volatile and often close to zero. This leads to slow convergence and other computational problems.

The idea behind importance sampling is ingenious in its simplicity. Random draws are made, not from the prior, but from an *importance density*, $i(\theta)$. Almost any density can serve as $i(\theta)$, but the idea is to pick a density that might roughly approximate the unknown posterior distribution. If one then computes the *weighting function*, $w(\theta) \equiv f(y \mid \theta)p(\theta)/i(\theta)$, at each iteration, the average of $\theta w(\theta)$ converges to the required expectation of $\theta f(D \mid \theta)$ under the prior. This amusing conclusion is derived as follows:

---

[21] A more elaborate model would incorporate the fact that monthly returns and unexpected volatility have a negative contemporaneous relation. See French, Schwert and Stambaugh (1987).

$$\int \theta\, w(\theta)\, i(\theta)\, d\theta \;=\; \int [\theta f(D \mid \theta) p(\theta)/i(\theta)]\, i(\theta)\, d\theta \;=\; \int \theta f(D \mid \theta) p(\theta)\, d\theta. \qquad (9)$$

The advantage of drawing from $i(\theta)$, rather than the prior, is that, insofar as $i(\theta)$ does succeed in approximating the posterior, the $w(\theta)$'s will tend to be far more stable, leading to much faster convergence. The improved stability is a consequence of the fact that the numerator of the weighting function is proportional to the posterior density.[22] By an argument similar to that in (9), the average of the weights $w(\theta)$ converges to $p(D)$ as the number of random draws approaches infinity. The posterior mean for θ can then be obtained by taking the ratio.

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 rough posterior mean and standard deviation in place of the prior moments. After several rounds, each with a modest number of draws of θ, variability in the importance weights is reduced substantially. At that point, one more importance sampling procedure is run with a large number of draws. Additional details are given in the appendix.[23]

While we have discussed the computation of the posterior mean, the same ideas apply if the expectation of a more complicated function $\pi(\theta)$ is desired. To compute expected utility, returns are randomly drawn from their conditional distribution, given the fixed yield and draw of θ. These returns, $R(\theta)$, then play the role of $\pi(\theta)$.

IV. Prior beliefs about the risk/return parameters

Having looked at the preliminary regression evidence, we now examine the data from a Bayesian perspective. We ask how the evidence would affect the perceptions of individuals with a range prior beliefs. Thus, understanding the *mapping* between prior and posterior beliefs is a central theme of the analysis. Our continuous priors embrace

[22] In fact, if $i(\theta)$ were equal to the posterior, the weights would all be equal to $p(D)$.
[23] This method of computing $p(D)$ is also used in the evaluation of Bayes factors which are discussed later in the paper.

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 ($\beta_m = 0$) [To be included in next draft]. We also consider diffuse priors for each parameter, which allow the data to dominate posterior beliefs. These are 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.

A detailed description of the intuitive economic motivation behind the choice of parameter values for our balanced priors is given below. Some of these choices will inevitably seem a bit arbitrary, but we think they provide a range of beliefs that serve as reasonable points of departure for thinking about these issues.[24] Though one might argue for some interaction between mispricing and risk, for simplicity we assume the priors for ($c, \gamma, \lambda_1, \lambda_2, \beta_m$) are independent. Since $\alpha = \gamma c$, its prior will be induced by the priors for γ and c. Apart from the prior for c, we employ the same priors on the basic parameters for the full period and the subperiod. By a standard result, if a fixed model is assumed, then updating the initial prior using the pre-1960 data and employing this prior in the subperiod analysis is equivalent to directly combining all of the data with the initial prior.[25]

A. 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.[26] This prior distribution is taken to be lognormal to ensure positivity. Based on the Schwert (1990) data and descriptive statistics above, we let μ_c = 4.5% and σ_c = 50 bps as of the beginning of 1960 and μ_c = 6%, σ_c = 100 bps for 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

---

[24] Readers wishing to provide a description of their personal priors can e-mail us. Anonymity will be protected.
[25] This won't be quite true in our application since we delete the last 6 years of the sample when studying the subperiod.
[26] 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.

experienced in the 30s, but with considerably more uncertainty than that of the pre-1927 period. Later, we see that the data totally dominate beliefs about c.

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 1940-2002 period, the average excess return on the NYSE value-weighted stock index was 54 bps per month with standard deviation 4.2%, similar to the sample moments for 1841-1920. The corresponding Sharpe ratio is about 0.13. 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 lower prior mean of $\mu_\gamma = 0.111$, so that $\mu_\alpha = 50$ bps when $\mu_c = 4.5\%$ and 67 bps when $\mu_c = 6\%$. 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. We let $\sigma_\gamma = 0.05$ ($\mu_\gamma = 0.111$), 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.

B. Priors for the time-varying risk parameters λ

We want our prior for λ_1 to reflect a belief that risk will display positive persistence from one month to the next, though with some mean-reversion. Given the great precision with which this coeficient can be estimated, the exact prior specification is not important. We take it to be normal with $\mu_{\lambda 1} = \sigma_{\lambda 1} = 0.5$.

In order to get some feel for the economic significance of different values of λ, we note that, as a rough approximation when λx is small, $\sigma_\varepsilon \approx c(1 + \lambda x)$. So, a value of λ_2 equal to 10, say, implies (roughly) that a 100 bps change in yield will be associated with a proportionate change in σ_ε of about 10% [10% of $\mu_c = (0.10)(0.045) = 45$ bps] when log(sd) (deviation) is zero. We might expect the direct relation between yield and risk to be positive through a leverage effect (Black, 1976 and Christie, 1982); i.e., leverage and stock risk both increase when price declines, with the decline also increasing the yield if dividends change slowly over time. However, λ_2 measures the *partial* effect of dp on risk controlling for sd, so the expected relation is less clear. Insofar as the estimate sd_t captures the information about ex ante risk at time t with error, however, we might expect the partial effect of dp to still be positive, though smaller than the direct effect. We take the prior for λ_2 to be normal with $\mu_{\lambda 2} = 10$ and $\sigma_{\lambda 2} = 10$. Thus, a positive relation is expected under the prior, but some negative mass, about 16%, is placed on negative values, reflecting our uncertainty about the parameter. We will see that the data tend to dominate beliefs about λ_2 with this level of prior uncertainty.

C. Priors for the time-varying expected return parameters, β_m and k.

The slope coefficient on yield in (6), β_m, reflects expected return variation that related to yield but is unrelated to risk. A common contrarian view is that an overpriced (underpriced) market, in which anticipated dividend growth is unrealistically high (low), can be reflected in a low (high) dividend yield. This situation might arise, for example, if investors tend to extrapolate recent economic strength too far into the future (Lakonishok, Shleifer, and Vishny, 1994), a sort of overreaction scenario. If this irrational sentiment is mean-reverting, the implied “correction” can lead to a positive relation between yield and actual (as opposed to perceived) expected return, implying a positive value for β_m.

The influential paper by DeBondt and Thaler (1985) viewed overreaction as the fundamental behavioral hypothesis emerging from the work of Kahneman and Tversky. To capture this perspective, we consider an "overreaction prior" (OP) that is normally distributed with $\mu_{\beta m} = 0.20$, implying that a 100 bps change in yield is associated with a substantial change of 20 bps in expected return, or 2.4% annualized. In this case, a two-percentage-point drop in yield below the mean, a large decline in historical terms, leaves the (prior) expected return at 10 bps (50 – 40), or just 1.2% *per annum*. We let $\sigma_{\beta m} = 0.10$, indicating a high degree of confidence that β_m is positive. Note that at relatively high values of β_m entertained by this prior, expected returns will sometimes be negative, an implausible conclusion from the risk-based perspective.[27]

Even if overreaction is a fundamental behavioral bias, the leap from experimental psychology to implications for financial market equilibrium entails additional uncertainty. Moreover, an underreaction scenario might also be viewed as plausible. Suppose, for example, that economic news has been good and continued solid growth is expected in the future, resulting in a low yield. However, say that investors underreact to the current information and underprice the market in this sort of environment. A future correction will now lead to a higher than average subsequent return when yield is low, resulting in a negative value of β_m.

If the overreaction and underreaction scenarios are considered equally likely a priori, then the prior mean for the mispricing effect might be near zero, though the variance could still be large. In this view, the literature's emphasis on the overreaction scenario is driven more by the data than compelling theoretical arguments. To accommodate this perspective, we also consider a prior that is "neutral" (NP) about the direction of the mispricing effect with $\mu_{\beta m} = 0.0$ and $\sigma_{\beta m} = .10$.[28] Later in the paper, our OP and NP results will be compared to those for a dogmatic "no-mispricing prior" that assumes β_m is zero [next draft].

---

[27] Kothari and Shanken (1997) exploit this observation in their interpretation of predictability results.

[28] "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). 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.

Financial theory provides a natural reference point for thinking about the relation between market expected return and risk. Merton (1980, p.329) emphasizes that if state variables other than aggregate wealth have a relatively small effect on consumption, or if the variance of changes in wealth is much larger than the variance of changes in the state variables, then the market risk premium can be reasonably approximated as proportional to market variance. If the representative investor has a utility function with constant relative risk aversion, then the constant of proportionality will equal the relative risk aversion coefficient. This relation motivates the early empirical studies by Merton (1980) and French, Schwert and Stambaugh (1987).

While Merton himself clearly recognizes (p. 328) the *possibility* that an increase in market risk will not be accompanied by an increase in the risk premium, he refers to a positive relation as a "generally-reasonable assumption." Similarly, Abel identifies an equilibrium context in which the market risk premium is a decreasing function of (dividend) risk, but states (pp. 391-392) "I do not claim that such a negative relation holds for empirically relevant parameter values." We take these comments as motivation for a prior that imposes a positive relation with high probability. Recall that in our parametrization, the proportionality condition corresponds to $\alpha = k$, where $\alpha = \gamma c$. We specify a prior for k, conditional on α, that is normal with mean and standard deviation both equal to α. Thus, our "best guess" is that k will equal α and, more generally, k will be positive with probability 0.84, given a positive value of α.[29] We also entertain dogmatic priors that assume either $k = \alpha$ or $k = 0$.

V. Empirical Evidence: Bayesian Posterior Distributions

We begin by examining the full 1940-2002 period and then turn to the subperiod 1960-1994 considered earlier. Posterior beliefs about the model parameters are discussed in this section and implications for asset allocation are explored in Section VI.

A. Posterior distributions based on diffuse priors

[29] The prior probability that α is positive is 0.99 under OP and NP. See Tables IV and V.

Results for the 1940-2002 period based on diffuse priors are reported in the first panel of Table IV. Posterior means (the Bayesian "estimates") are given first, with posterior standard deviations (similar to conventional standard errors) below in parentheses. We often refer to the moments based on diffuse priors as "data-based" estimates or standard deviations. The numbers in brackets are posterior probabilities that the given parameter exceeds zero [next draft]. The long-run expected return parameter, α, is almost certainly positive with a posterior mean of 56 bps per month. The estimate of the long-run standard deviation, c, is about four and a half percent and, as expected, it is estimated with great precision. The implied Sharpe ratio, γ, is significantly positive as well.

The estimated coefficient on dp, β_m, is similar to β_1 in Table II, and is about two and a half standard deviations above zero. Consequently, the relation between yield and expected return, controlling for risk, is postive with (posterior) probability nearly equal to one. The estimated effect is economically substantial - a one percentage point increase in yield implies an increase in expected return of 27 bps per month or a little more than 3% per year, holding risk constant.

The estimated relation between risk and expected return is also positive. The mean of k is less than one standard deviation above zero, however, with the probability of a positive relation about three fourths. The estimate of 0.20 is substantially below the 0.56 estimate of α, indicating that expected return varies less with risk than would be the case under the proportionality condition ($\alpha = k$). [compute prob $k<\alpha$ an in next draft]

We now turn to the parameters that describe time variation in risk. As in Table II, there is a strong relation between the lagged risk measure sd_t and ex ante market risk, as measured by λ_1. The evidence suggests that the incremental effect of yield is negative, but the estimate of λ_2 implies that risk declines by less than 1% (of it's value) when yield increases by one percentage point.

The parameter β_{tot} measures the *total* effect of yield on expected return in our nonlinear model and is analogous to a simple predictive regression coefficient in a linear model.[30] The estimates of β_m and β_{tot} are approximately equal, a consequence of the weak relation between yield and risk (Table II) and the relatively low value of k in this period. Thus, all of the predictability in yield is attributed to mispricing in this case.

As noted earlier, the relation between yield and risk was quite different in the 1960-1994 period. Based on this data, an investor with diffuse priors would conclude that the incremental effect of yield on risk is positive with posterior probability one (see Table V). The economic effect is modest, with risk increasing by about 3% of it's initial value when yield increases by one percentage point. The total effect of yield on risk is greater, since dp and log(sd) are themselves positively related in this period (see Table II). However, (normalized) risk has even less impact on expected return in this period, with mean k equal to 0.14. As a result, the difference between β_{tot} and β_m is still small, about 0.04. Thus, even in a period selected to highlight the relation between risk and yield, most of the yield-related predictability in expected return is unrelated to the variance measure of total risk.

The means of both β and β_m are somewhat higher in the 1960-1994 period, but the posterior standard deviations more than double, due in part to the smaller sample. Still, the probabilities for positive predictability (total and partial) are quite high, 0.97 and 0.94, respectively. Recall that the mean market return was only 31 bps over this period, as compared to 54 bps for the 1940-2002 period. This explains the lower mean of α in Table V. Changes in the other parameters are of less interest and comparisons across periods are subject to the caveat that the notion of "long-run" corresponds to different values of the mean yield in the two tables.

B. Posterior results with informative prior beliefs that focus on mispricing

We now examine the informative prior results in the second and third panels of Tables IV and V for the full period and subperiod respectively. Informative priors, as specified here, have minimal effects on the variance parameters since the prior precisions are substantially lower than that of the data-based estimates. This is particularly true for

---

[30] It is defined as the change in expected return as yield varies from one percentage point below to one point above the mean, divided by .02, with the values of the risk inputs determined by the auxiliary regression of log(sd) on dp.

c and λ_1 which have extremely precise estimates. Although our prior for the risk-return tradeoff parameter was intended to be somewhat informative, in that a fairly low probability is assigned to negative values of k, such a prior again turns out to have low precision relative to the data-based estimate. Thus, the posterior means are close to the diffuse mean of 0.20.

Informative prior beliefs about mispricing have a much greater effect since the prior standard deviations that we judge to be reasonable for β_m are similar to the data-based standard deviations for the full period and substantially lower for the 1960-1994 subperiod. In a standard regression context, posterior precision (reciprocal of variance) is the sum of prior and sample precisions, while the posterior mean is a precision-weighted average of the prior mean and the regression estimate. This need not be true for all parameters in our nonlinear specification, but it holds approximately for β_m.[31] For example, the posterior mean β_m of 0.13 under the neutral prior is about halfway between the prior mean of zero and the data-based estimate of 0.27 in Table IV. The associated posterior probability that β_m is positive increases to 0.96 from the prior probability of 0.50. Overreaction prior results are much closer to the diffuse prior results since the prior mean is notfar from the data-based estimate.

Estimates are shrunk more toward the prior means using the samller 1960-1994 data set (Table V). As a result, the NP investor is now less convinced (posterior probability 0.73) that β_m is positive, despite the higher data-based estimate of 0.37. The gap between the OP and NP posterior means for β_m rises to 0.17 for this period. This translates into an annualized difference in expected returns of more than 2% per annum when dividend yield increases by one percentage point.

We observe some interaction between posterior beliefs about mispricing and beliefs about the risk/return tradeoff. This occurs in the 1960-1994 period, in which posterior means for k are inversely related to the means for β_m. We offer the following possible explanation for this phenomenon. The data for this period are consistent with a

---

[31] Recall that α is the product of γ and c. The informative posterior means for α are below the prior means because of the large prior mean for c relative to the data. They are a bit lower than the diffuse mean over 1940-2002 because the prior mean for γ is lower than the diffuse mean. The similar behavior of k may arise from its link to α in the conditional prior for k.

positive relation between expected return and yield, as evidenced by the estimate of β_{tot}. Since yield and variance are also positively related, the model must determine how much of this predictability is attributed to yield via β_m and how much to risk via k. Consider an extreme scenario in which the true value of β_m is positive and k is zero, but the prior constrains β_m to equal zero. In this case, we might expect the model to produce a positive value of k to proxy for the positive yield/return relation found in the data. By this line of reasoning, the neutral prior, which shrinks β_m toward zero, would have the lowest value of k, as observed in Table V.[32]

We have also examined posteriors (not shown) under the "dogmatic" prior assumption that the proportionality condition, $k = \alpha$, holds with probability one. This tends to lower the estimates of β_m (0.37 to 0.30, diffuse) when yield is positively related to risk (1960-1994) since more of the total yield-related predictability is attributed to risk in that case and, hence, less to mispricing. Although we don't report posteriors under the assumption that $k = 0$, i.e., no relation between risk and expected return, in that case β_m would equal β and the distributions would presumably be similar to those for β in Tables IV and V.

VI. Asset Allocation Analysis

In general, given probability beliefs about model parameters and specific levels of the predictive variables, a "predictive" or perceived return distribution can be computed. An optimal portfolio and corresponding certainty equivalent return (CER) can then be determined based on an assumed utility function. Recall that KS consider a simple regression of returns on dividend yield. Suppose that the "current" level of yield is above it's mean. To assess the "economic significance" of their slope parameter estimate, they determine the portfolio that would be optimal if yield were at it's mean, i.e., ignoring the predictability evidence. Alternatively, with yield in deviation form, this amounts to plugging in a zero for the slope parameter. Of course, this portfolio will be not be

[32] Insofar as risk is not highly correlated with yield, this proxy effect would be far from perfect. We do not consider this argument a formal proof , but hopefully it hints at some of the conditions that might give rise to the observed patterns for k. Note that the argument does not apply to the 1940-2002 period since yield and risk were not positively related.

optimal with respect to the actual predictive distribution and utility with respect to that distribution will be lowered accordingly. The CER loss incurred using the "suboptimal portfolio" in this manner is the metric KS use to capture the notion of economic significance.

In our model, expected return varies with both risk and yield, and risk varies with the lagged estimate of risk as well as yield. With this many parameters and predictive inputs, each for a range of priors, the number of permutations is large and evaluating significance becomes far more complicated. To make the analysis more manageable we have, therefore, chosen to focus on what seem to us the most interesting issues for representative sets of predictive inputs. Another novel feature of our analysis is the exploration of predictive distributions based on prior as well as posterior beliefs. We formally describe the optimization framework in the next section and then present the evidence.

A. The optimization framework

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

$$U(W) = W^{1-A}/(1-A), \tag{10}$$

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,

$$\begin{aligned} W_{T+1} &= W_T[\omega \exp(r_{T+1} + i_{T+1}) + (1-\omega) \exp(i_{T+1})] \\ &= W_T \exp(i_{T+1})[\omega \exp(r_{T+1}) + (1-\omega)] \end{aligned} \tag{11}$$

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 (10) and (11) 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 D_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 probability distribution of the parameters:

$$p(r_{T+1} \mid D_T) = \int f(r_{T+1} \mid \theta_1)\, p(\theta_1 \mid D_T)\, d\theta_1, \tag{12}$$

where $\theta_1 \equiv (\gamma, c, \lambda, \beta_m)$ and $p(\theta_1 \mid D_T)$ 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 (12), the *prior-predictive distribution*:

$$p(r_{T+1}) = \int f(r_{T+1} \mid \theta_1)\, p(\theta_1)\, d\theta_1, \tag{13}$$

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

$$\omega^* \approx \mu_T/(A\sigma^2_T) + ½A, \tag{14}$$

where μ_T is the predictive mean and σ^2_T the predictive variance. We use this to simplify the computations. In particular, if there is no mispricing ($\beta_m = 0$) and the Merton condition ($k = \alpha$) holds then, based on (6), 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 dp and sd vary must be driven by a belief in mispricing.

B. Optimal weight and CER comparisons at different levels of dividend yield

First, for each of our three priors, we examine the significance of the mispricing evidence, i.e., the influence of posterior beliefs about β_m. This is similar to the analysis in KS in that β_m is set to zero in determining the suboptimal portfolio, while the actual predictive distribution is based on a hypothetical value for yield. Here, the risk input is set to zero as well.[33] Results for the 1940-2002 period are given in Table VI.

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. The next two columns give the predictive return moments, first based on the prior (when informative) and then based on the posterior. Let's start with the diffuse prior. When yield is at its mean, the predictive expected return is just the posterior mean of α, 56 bps per month (see Table IV). When yield is high, the mean increases to 110 bps and at the low value of yield it is just 1 bp. Predictive risk declines slightly with yield. As a result, the optimal portfolio weights rise sharply, from 11% to 100%, as yield increases. Thus, due to the strong evidence of mispricing, predictability has a large impact on asset allocation. The corresponding optimal CERs increase from 1 to 71 bps.

Now, suppose we were to ignore predictability related to mispricing, while continuing to incorporate expected return variation induced by the (minimal here) relation between yield and risk. The "suboptimal" weights derived in this scenario are given under the weight column β_m, the parameter we are ignoring. The weights are nearly constant as yield increases. Consider the high-yield case. The loss of 13 bps shown under the last β_m loss column tells us that CER drops from the optimal 71 bps to 58 bps if we invest 66% of our portfolio in the market when the portfolio should be fully invested based on the true predictive distribution. The loss is even higher in the low yield scenario, almost 2 percent annualized. Such numbers are not dramatic by cross-sectional asset pricing anomaly standards, but they are substantial in relation to the long-run market (excess) return between 6 and 7 percent per annum for this period, a more

[33] Although there is a significant correlation between dp and log(sd), there is substantial unexplained variation. Therefore, fixing risk while varying yield is a reasonable thought experiment and simplifies the analysis.

appropriate reference point.[34] The CER losses are a little smaller for OP investors and considerably smaller in the NP case. The NP investors have a much lower perception of β_m, and so ignoring the mispricing effect has less impact on expected returns and optimal portfolio weights.

The CER and optimal weight comparisons discussed above, like those in earlier literature, are properly interpreted as an assessment of the economic significance of *posterior* beliefs. In the diffuse case, significance can reasonably be attributed to the impact of the data. 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 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 when informative priors are employed. One natural approach is to compare the respective optimal portfolios under the prior and posterior-based predictive distributions.

With NP or OP, the biggest difference between prior and posterior weights occurs when yield is high. In this case, predictive risk is much higher before looking at the data than after. Prior expected return is higher as well because of the higher risk and higher belief about k, but there is an offsetting effect of the lower prior belief about β_m and apparently the risk effect dominates. This results in less aggressive investment in stock under the prior. The resulting CER loss if the OP investor were to ignore the data and invest 59% in the market, rather than being fully invested, is 13 bps or a bit more than 1.5% per annum. The loss is 9 bps for the NP investor. We conclude the data do play a "significant" role in determining posterior beliefs over the 1940-2002 period when yield is high.[35] Note that the NP prior-predictive weights are constant at 46% as yield varies in Table VI. This is apparently a consequence of the fact that the most likely scenario under the prior is no mispricing and expected return proportionality in variance, the situation discussed at the end of part A.

---

[34] Losses are a few bps smaller in the 1960-1994 period.
[35] Corresponding losses for the 1960-1994 period are quite small.

Another interesting question is whether it matters much which prior an investor starts with. Naturally, the prior will matter more when there is less data, as in the 1960-1994 subperiod. In Table VII, we show the losses at each level of yield when an investor with a given prior (under "optimal") is forced to hold a suboptimal portfolio determined by another prior. The optimal CERs are given in the first column and the losses in the next two columns of each panel. The biggest loss of 11 bps occurs when yield is low and the diffuse investor (weight = 0) is forced to hold the NP investor's portfolio (weight = 43%). This makes sense since differences in posterior beliefs about mispricing are greatest for those investors. The corresponding loss for the 1940-2002 period (not shown) is just 4 bps.

C. Optimal weight and CER comparisons at different levels of log(sd)

Now we examine the economic significance of predictability related to the lagged risk measure, log(sd), while setting dp to zero. This is less straightforward than our analysis of β_m since movements in risk will, to some degree, be accompanied by changes in expected return. As noted earlier, the effects on asset allocation would be offsetting under the proportionality condition. We study each effect separately, as well as the joint effect. Since beliefs about mispricing are of secondary interest in this context, we report diffuse prior results only.

To focus on significance of the risk effect, we force our investor to hold a suboptimal portfolio that is based on the correct expected returns, given the model, but with risk at the long-run level c. The associated CER loss is then a measure of the economic significance of volatility persistence, as measured by λ_2. Results for the 1940-2002 period are given in the first panel of Table VIII. Although the optimal and suboptimal allocations are quite different at both low and high levels of yield, the largest CER loss of 25 bps (3 percent per annum) occurs when yield is high and the increased risk is ignored, resulting in overinvestment in the market. In contrast, there is barely any loss associated with ignoring expected return movements related to risk (second panel) given the relatively low estimate of k in Table IV. Ignoring all predictability related to log(sd) leaves the market weight unchanged as sd varies, with more moderate losses of 5 bps in the high and low risk states.

D. Alternative prior beliefs about the risk/return tradeoff

Finally, we examine different prior beliefs about the risk/return tradeoff parameter k. The N(α, α) conditional prior used above is centered on the proportionality condition, with the probability of a positive relation between risk and expected return equal to 0.84. Now we consider two very different dogmatic alternatives. One prior assumes that the proportionality condition, k = α, holds exactly. The other reflects certainty that k = 0 and there is no relation between market risk and return. We present Bayes factors as well as CER loss analysis in Table IX for the 1940-2002 period.

The first column gives the priors that will be compared. The average likelihood associated with the first prior in each pair is in the numerator of the Bayes factor calculation and that same prior serves as the belief under which the optimal portfolio is computed. The second prior in each pair is used to compute the average likelihood in the denominator and determines the "suboptimal" portfolio for the CER analysis. The Bayes factor can be interpreted as the posterior odds in favor of the first "hypothesis" relative to the second when the prior odds equal one. We assume diffuse priors for parameters other than k and set dp to zero while the level of log(sd) varies.

Recall from Table IV that the data-based estimate of k for the full period is 0.20 with posterior standard deviation 0.27, while the estimate of α is 0.56 with standard deviation 0.16. The continuous N(α,α) prior assigns some probability to values of k near 0.20, though it also assigns substantial probability to values of k much larger than 0.20. As a result, the associated average likelihood is 140% of the average likelihood when k is constrained to equal α, but only about 40% of the value under k = 0. The latter odds ratio translates into a posterior probability of 2.5/(1+2.5) = 71% in favor of the "null hypothesis" that k=0. A dogmatic belief in the proportionality condition fares even worse, with a Bayes factor of about 30% relative to k = 0.

Different beliefs about k give rise to different levels of predictive expected return. Naturally, this matters most when posterior predictive risk is high. Thus, the CER loss for each pair of priors in Table IX is greatest when sd is high. The biggest loss of 15 bps (1.8 percent annualized) is observed when an investor who believes that k = α is forced to hold the portfolio of one who is certain there is no connection between risk and return. This makes sense since the most extreme Bayes factor is found in this case as well.

Interestingly, although the odds in favor of k = 0 are about 2.5 to 1 compared to the N(α, α) alternative, the resulting portfolios are not very different, with stock weights of 33% and 45%, respectively, and an implied loss of just 1 bp. Which metric is the right one? We think of the Bayes factor or odds ratio approach as relevant when a researcher has a purely "scientific" curiosity about which parameter value is the correct one. In contrast, the CER approach assesses whether differences in parameter values "matter" from a very specific perspective – that of asset allocation with a particular utility function. Each is informative in its own way.

VII. Summary and Conclusions

[Next draft]

Appendix A. Obtaining the posterior moments

Let $D \equiv (D_1, D_2, \ldots, D_T)$, where $D_t \equiv (r_t, x_t)$ is the data for the problem. Recall that $x_t \equiv (dp_t, v_t)'$, where dp_t is the dividend yield and v_t is the within-month sum of squared deviations from the sample mean of daily returns for month t. The joint density (conditioned on x_0) for the data is $f(D) = f(D_1)f(D_2 \mid D_1) \ldots f(D_T \mid D_1, D_2, \ldots, D_{T-1})$, where $f(D_{t+1} \mid D_1, D_2, \ldots, D_t) = f(r_{t+1} \mid D_1, D_2, \ldots, D_t)f(x_{t+1} \mid r_{t+1}, D_1, D_2, \ldots, D_t)$ for t = 1, …, T-1. We assume that x_t captures the state of the world at time t in the sense that $f(r_{t+1} \mid D_1, D_2, \ldots, D_t) = f(r_{t+1} \mid x_t)$ and $f(x_{t+1} \mid r_{t+1}, D_1, D_2, \ldots, D_t) = f(x_{t+1} \mid r_{t+1}, x_t)$. With v_t omitted, this representation of the joint density of returns and yields includes the restricted VAR of Stambaugh (1999) with homoskedastic errors as a special case, with yield following an AR(1) process.

We assume that the (negligible) residual, w_{t+1}, in the yield equation (8) is independent (conditional on x_t) of the daily returns within month t+1 and hence independent of their sum, ε_{t+1}, and v_{t+1} as well. In this case,

$$f(x_{t+1} \mid r_{t+1}, x_t) = f(dp_{t+1} \mid r_{t+1}, x_t)\, f(v_{t+1} \mid x_t). \qquad \text{(A.1)}$$

where the density for dp_{t+1} is based on (8) and the density for v_{t+1} is that of $\sigma_{\varepsilon t}^2/22$ times a chi-square variate with 21 degrees of freedom, as discussed in Section III.B.

Let θ be the parameter vector for the joint distribution of $(r_{t+1}, dp_{t+1}, v_{t+1})$ and partition θ as (θ_1, θ_2), where $\theta_1 \equiv (\gamma, \beta_m, c, \lambda, k)$ and $\theta_2 \equiv (\varphi, \rho, \phi, \sigma_w)$, where θ_1 is the vector of return-risk parameters in (5) and (6) and θ_2 the additional parameters in (8).[36] Note that the density for v_{t+1} does not depend on θ_2. Given the discussion above, the joint density of the data or, equivalently, the likelihood function can be written as

$$f(D \mid \theta) = h(D, \theta_1)g(D, \theta)k(D, \theta_1). \qquad \text{(A.2)}$$

Here, $h(D, \theta_1)$ is the product of the conditional densities for r_{t+1}, $g(D, \theta)$ the product of the densities for dp_{t+1} and $k(D, \theta_1)$ the product for v_{t+1}. Note that h is identical to what the conditional density of r, given x, would be if x were nonstochastic.

[36] Recall that $\alpha = \gamma c$.

Since g does not depend on θ_1, it will act as a constant of proportionality in determining the posterior density for θ_1 and can be ignored. Likewise, the prior for θ_2 can be ignored, given our assumption that it is independent of $p(\theta_1)$, the prior for θ_1. Therefore, the posterior moments for θ_1 can be obtained using the importance sampling approach outlined in Section III.C with $h(D, \theta_1) \cdot k(D, \theta_1)$ as the likelihood function and the prior $p(\theta_1)$.

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 (14) 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.

Backus, David and Allan Gregory, 1993, Theoretical Relations Between Risk Premiums and Conditional Variances, *Journal of Business and Economic Statistics* 11, 177-185.

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, A., and I. Welch (2003): “Predicting the Equity Premium with Dividend Ratios,” *Management Science* 49(5), 639–654.

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.

Judge, G., W. Griffiths, R. Hill, Lutkepohl, and T. Lee, 1985, *The Theory and Practice of Economerics*, John Wiley, New York.

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, 2003, Anomalies and Asset Allocation, monograph published by 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, Dividend Yield and Risk for 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, D/P, 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 monthly standard deviation, SD, is sample standard deviation of daily returns multiplied by the square root of 22 tading days.*

	Monthly Excess Return			D/P		SD	
	Mean	Standard Deviation	Sharpe Ratio	Mean	Standard Deviation	Mean	Standard Deviation
1927-2002	0.48	5.44	0.09	4.08	1.49	4.58	3.13
1927-1939	0.16	9.38	0.02	4.92	1.82	8.16	4.77
1940-2002	0.54	4.20	0.13	3.91	1.35	3.84	1.98
1960-1994	0.31	4.31	0.07	3.69	0.81	3.64	1.95

*S&P 500 daily returns used before July 1962 with adjustments described in footnote 8.

Table II: Regression Evidence

Panel A

$$r_{t+1} = \beta_0 + \beta_1 dp_t + \beta_2 sd_t^2 + \text{error}$$

	β_0 %	β_1	$\beta_2 \times 10^4$	Adj. R^2 (%)
1940-2002	-0.69 (0.48)	0.32* (0.11)	-0.09 (0.94)	0.7
1960-1994	-1.90 (1.03)	0.66* (0.29)	-1.44 (0.90)	1.2

* Indicates estimate more than 2.0 standard errors from zero.

Panel B

$$\log(sd_{t+1}) = \delta_0 + \delta_1 \log(sd_t) + \delta_2 dp_t + \text{error}$$

	δ_0	δ_1	δ_2	Adj. R^2 (%)
1940-2002	0.48* (0.05)	0.65* (0.03)	-1.05 (0.87)	42.1
1960-1994	0.20* (0.06)	0.58* (0.04)	8.22* (2.14)	46.2

* Indicates estimate more than 2.0 standard errors from zero.

Panel C

$$\log(sd_t) = \delta_3 + \delta_4 dp_t + \text{error}$$

	δ_4	δ_5	Adj. R^2 (%)
1940-2002	1.25* (0.11)	0.08 (2.44)	-0.1
1960-1994	0.33* (0.13)	23.5* (3.7)	23.4

* Indicates estimate more than 2.0 standard errors from zero.

Table III: Log-Dividend Yield Equation

This table reports estimates for the log dividend yield regression model with White adjusted standard errors are given in parentheses.

$$\log(yield_{t+1}) = \varphi + \rho \cdot \log(yield_t) + \phi \cdot r_{t+1} + w_{t+1}.$$

Table IV: Bayesian Estimates: 1940 – 2002

This table reports Bayesian estimates for the regression

$$r_{t+1} = (\alpha - k) + k(\sigma_{\varepsilon t}^2/c^2) + \beta_m dp_t + \varepsilon_{t+1},$$

where r_{t+1} is the continuously compounded excess return on the NYSE value-weighted stock index and dp_t is the NYSE dividend yield. The unexpected return, ε_{t+1}, is normally distributed with mean 0 and conditional standard deviation $\sigma_{\varepsilon t} = c \cdot \exp[\lambda_1 \log(sd_t) + \lambda_2 dp_t]$, where sd_t is the monthly standard deviation estimate based on daily data within month t. Both dp and sd are measured as deviations from long-run values.

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 [dp = log(sd) = 0] Sharpe ratio and c is the long-run standard deviation of return; $\gamma c = \alpha$ is the long-run expected excess return.

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 under the given prior.

	α (%)	β_{tot}	c (%)	λ_1	λ_2	γ	β_M	k (%)
Diffuse								
Posterior								
Mean	0.56	0.27	4.48	0.67	-0.88	0.124	0.27	0.20
(Stdev)	(0.16)	(0.11)	(0.03)	(0.01)	(0.39)	(0.036)	(0.11)	(0.27)
[Prob(*>0)]	[1.00]	[0.99]	[1.00]	[1.00]	[0.01]	[1.00]	[1.00]	[0.76]
Neutral								
Prior								
Mean	0.66	0.12	6.00	0.50	10.0	0.111	0.00	0.67
(Stdev)	(0.32)	(0.26)	(1.00)	(0.50)	(10.0)	(0.050)	(0.10)	(0.81)
[Prob(*>0)]	[0.99]	[0.68]	[1.00]	[0.84]	[0.84]	[0.99]	[0.50]	[0.83]
Posterior								
Mean	0.53	0.13	4.49	0.67	-0.85	0.119	0.13	0.18
(Stdev)	(0.13)	(0.07)	(0.03)	(0.01)	(0.39)	(0.029)	(0.07)	(0.26)
[Prob(*>0)]	[1.00]	[0.96]	[1.00]	[1.00]	[0.01]	[1.00]	[0.96]	[0.75]
Bayes Factor								
Overreaction								
Prior								
Mean	0.66	0.32	6.00	0.50	10.0	0.111	0.20	0.67
(Stdev)	(0.32)	(0.26)	(1.00)	(0.50)	(10.0)	(0.050)	(0.10)	(0.81)
[Prob(*>0)]	[0.99]	[0.95]	[1.00]	[0.84]	[0.84]	[0.99]	[0.98]	[0.83]
Posterior								
Mean	0.53	0.23	4.49	0.67	-0.86	0.119	0.23	0.17
(Stdev)	(0.13)	(0.07)	(0.03)	(0.01)	(0.39)	(0.029)	(0.07)	(0.26)
[Prob(*>0)]	[1.00]	[1.00]	[1.00]	[1.00]	[0.01]	[1.00]	[1.00]	[0.75]
Bayes Factor								

Table V: Bayesian Estimates: 1960 – 1994

This table reports Bayesian estimates for the regression

$$r_{t+1} = (\alpha - k) + k(\sigma_{\varepsilon t}^2/c^2) + \beta_m dp_t + \varepsilon_{t+1},$$

where r_{t+1} is the continuously compounded excess return on the NYSE value-weighted stock index and dp_t is the NYSE dividend yield. The unexpected return, ε_{t+1}, is normally distributed with mean 0 and conditional standard deviation $\sigma_{\varepsilon t} = c \cdot \exp[\lambda_1 \log(sd_t) + \lambda_2 dp_t]$, where sd_t is the monthly standard deviation estimate based on daily data within month t. Both dp and sd are measured as deviations from long-run values.

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 [dp = log(sd) = 0] Sharpe ratio and c is the long-run standard deviation of return; $\gamma c = \alpha$ is the long-run expected excess return.

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 under the given prior.

	α (%)	β_{tot}	c (%)	λ_1	λ_2	γ	β_M	k (%)
Diffuse								
Posterior								
Mean	0.30	0.41	4.21	0.68	3.05	0.072	0.37	0.14
(Stdev)	(0.20)	(0.23)	(0.03)	(0.02)	(0.97)	(0.048)	(0.24)	(0.27)
[Prob(*>0)]	[0.93]	[0.97]	[1.00]	[1.00]	[1.00]	[0.93]	[0.94]	[0.70]
Neutral								
Prior								
Mean	0.50	0.18	4.50	0.50	10.0	0.11	0.00	0.50
(Stdev)	(0.23)	(0.31)	(0.50)	(0.50)	(10.0)	(0.05)	(0.10)	(0.60)
[Prob(*>0)]	[0.99]	[0.74]	[1.00]	[0.84]	[0.84]	[0.99]	[0.50]	[0.83]
Posterior								
Mean	0.37	0.16	4.21	0.68	3.13	0.087	0.06	0.33
(Stdev)	(0.15)	(0.12)	(0.03)	(0.02)	(0.97)	(0.035)	(0.09)	(0.28)
[Prob(*>0)]	[0.99]	[0.92]	[1.00]	[1.00]	[1.00]	[0.99]	[0.73]	[0.90]
Bayes Factor								
Overreaction								
Prior								
Mean	0.50	0.38	4.50	0.50	10.0	0.11	0.20	0.50
(Stdev)	(0.23)	(0.31)	(0.50)	(0.50)	(10.0)	(0.05)	(0.10)	(0.60)
[Prob(*>0)]	[0.99]	[0.95]	[1.00]	[0.84]	[0.84]	[0.99]	[0.98]	[0.83]
Posterior								
Mean	0.37	0.30	4.21	0.68	3.11	0.087	0.23	0.26
(Stdev)	(0.15)	(0.11)	(0.03)	(0.02)	(0.97)	(0.035)	(0.09)	(0.27)
[Prob(*>0)]	[0.99]	[1.00]	[1.00]	[1.00]	[1.00]	[0.99]	[0.99]	[0.84]
Bayes Factor								

Table VI: Predictive Analysis with Changing Yield: 1940 – 2002

This table reports predictive return moments for the market index, the optimal allocations to stock (weight), and certainty equivalent (excess) returns (CER) or losses associated with investment in optimal or suboptimal portfolios, respectively. These are given for each prior and 3 levels of yield, shown as sample standard deviations from the mean. The log(sd) deviation is set to zero.

The β_m columns assume that the mispricing effect of dividend yield is ignored in computing expected returns, resulting in "suboptimal" portfolio allocations. The associated loss is the difference between the optimal CER computed under the correct posterior-predictive distribution and the suboptimal CER. The prior weight and CER columns ignore the data entirely – the resulting loss measures the *incremental* significance of the data, given the informative prior.

PRIOR	Mean (Stdev) (%)		WEIGHT (%)			CER (bps)		
				Suboptimal			Loss	
Yield	Prior	Posterior	Optimal	Prior	β_m	Optimal	Prior	β_m
Diffuse								
-1.5	-	0.01 (4.57)	11	-	64	1	-	16
0	-	0.56 (4.49)	65	-	65	21	-	0
1.5	-	1.10 (4.41)	100	-	66	71	-	13
Neutral								
-1.5	0.48 (5.19)	0.28 (4.57)	36	46	62	7	1	4
0	0.66 (6.08)	0.53 (4.49)	63	46	63	20	2	0
1.5	1.09 (7.83)	0.79 (4.41)	91	46	64	40	9	3
Overreact								
-1.5	0.07 (5.19)	0.06 (4.57)	16	15	62	1	0	11
0	0.66 (6.08)	0.53 (4.49)	63	50	63	20	1	0
1.5	1.49 (7.83)	1.00 (4.41)	100	59	64	61	13	10

Table VII

Influence of Different Prior Beliefs about Mispricing: 1960-1994

Certainty equivalent losses in basis points per month when the investor is forced to hold a "suboptimal" portfolio based on a prior that is different from the actual one. The optimal portfolio is computed under the actual prior. [not quite as described in text – next draft]

Optimal: Neutral

Diffuse	Overreaction
8	3
1	0
4	2

Optimal: Overreaction

Diffuse	Neutral
2	3
0	0
0	2

Optimal: Diffuse

Neutral	Overreaction
11	3
1	1
4	0

Table VIII: Predictive Analysis of Changing Risk: 1940 – 2002

This table reports predictive return moments for the market index, the optimal allocations to stock (weight), and certainty equivalent returns (CER) or losses associated with investment in optimal or suboptimal portfolios, respectively. These are given for 3 levels of the predictive variable log(sd), measured as sample standard deviations from the mean. The yield deviation is set to zero. Priors are diffuse.

The "suboptimal" portfolio allocations are computed under a different assumption in each panel. In the first, the impact of log(sd) on risk is ignored; in the second, the impact of log(sd) on expected return through the changing ex ante variance is ignored; in the third, both of these effects are ignored. The associated loss is the difference between the optimal CER computed under the correct posterior-predictive distribution and the suboptimal CER.

COMPARISON λ_1	**WEIGHT (%)**		**CER (bps)**	
	Optimal	**Suboptimal**	**Optimal**	**Loss**
Ignore impact on risk				
-1.5	100	54	27	7
0	65	65	21	0
1.5	45	90	23	25
Ignore impact on E(r)				
-1.5	100	100	27	0
0	65	65	21	0
1.5	45	34	23	1
Ignore impact on risk and E(r)				
-1.5	100	65	27	5
0	65	65	21	0
1.5	45	65	23	5

Table IX

Influence of Different Prior Beliefs about the Risk-Return Tradeoff: 1940-2002

The average likelihood associated with the first prior in each pair is in the numerator of the Bayes factor and that same prior serves as the belief under which the optimal portfolio is computed. The second prior in each pair is used to compute the average likelihood in the denominator of the Bayes factor and determines the "suboptimal" portfolio for the CER analysis. Comparisons are made for 3 levels of the predictive variable log(sd), measured as sample standard deviations from the mean. The yield deviation is set to zero. Priors on parameters other than k are diffuse.

COMPARISON	Bayes	WEIGHT (%)		CER (bps)	
log(sd)	Factor	Optimal	Suboptimal	Optimal	Loss
k ~ N(α,α), k = α					
-1.5		100	70	27	4
0	1.40	65	70	21	0
1.5		45	70	23	8
k ~ N(α,α), k = 0					
-1.5		100	100	27	0
0	0.41	65	62	21	0
1.5		45	33	23	1
k = α, k = 0					
-1.5		70	100	11	2
0	0.29	70	62	25	0
1.5		70	33	56	15