

NBER WORKING PAPER SERIES

ON THE PREDICTABILITY OF  
STOCK RETURNS: AN ASSET-  
ALLOCATION PERSPECTIVE

Shmuel Kandel  
Robert F. Stambaugh

Working Paper No. 4997

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 1995

The authors are grateful for comments by workshop participants at the Ohio State University, Rutgers University, the University of Pennsylvania, the University of Rochester, the University of Washington, and Washington University in Saint Louis. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1994 by Shmuel Kandel and Robert F. Stambaugh. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

NBER Working Paper #4997  
January 1995

ON THE PREDICTABILITY OF  
STOCK RETURNS: AN ASSET-  
ALLOCATION PERSPECTIVE

ABSTRACT

The predictability of monthly stock returns is investigated from the perspective of a risk-averse investor who uses the data to update initially vague beliefs about the conditional distribution of returns. The optimal stocks-versus-cash allocation of the investor can depend importantly on the current value of a predictive variable, such as dividend yield, even though a null hypothesis of no predictability might not be rejected at conventional significance levels. When viewed in this economic context, the empirical evidence indicates a strong degree of predictability in monthly stock returns.

Shmuel Kandel  
Finance Department  
The Wharton School  
University of Pennsylvania  
Philadelphia, PA 19104-6367

Robert F. Stambaugh  
Finance Department  
The Wharton School  
University of Pennsylvania  
Philadelphia, PA 19104-6367  
and NBER

## 1. Introduction

Let  $r_t$  denote the continuously compounded return on the value-weighted portfolio of the New York Stock Exchange (NYSE) in month  $t$ , and let  $d_t$  denote the dividend yield on that portfolio at the end of month  $t$  (the sum of the dividends paid in months  $t - 11$  through  $t$  divided by the value of the portfolio at the end of month  $t$ ). Regressing the return on the lagged dividend yield, using the 804 returns from January 1927 through December 1993, produces the results,

$$r_t = \begin{array}{r} -0.0008 \\ (-0.1262) \end{array} + \begin{array}{r} 0.2016 \\ (1.3832) \end{array} d_{t-1} + \epsilon_t \quad \begin{array}{l} R^2 = 0.0024 \\ \hat{\sigma}(\epsilon) = 0.0559, \end{array} \quad (1)$$

where  $t$ -statistics appear in parentheses. The regression explains only 0.24% of the variance in monthly stock returns, and the  $t$  statistic for the slope coefficient is only 1.38.<sup>1</sup> Many might interpret these results as suggesting that the ability of dividend yields to forecast monthly stock returns is at best weak, but is such an interpretation of the evidence justified on economic grounds?

Consider a risk-averse investor with a one-month investment horizon who, on December 31, 1993, must decide how to allocate funds between common stocks and riskless cash. The dividend yield on that date is equal to 2.71%, which when substituted for  $d_{t-1}$  into (1), gives a fitted conditional mean of the continuously compounded stock return equal to 47 basis points. In comparison, because a dividend yield of 2.71% is less than the average yield of 4.35% for the 1927–93 period, the sample mean of the continuously compounded return over that period is 79 basis points. To what extent, if any, should the investor making the asset-allocation decision incorporate this estimated difference of 32 basis points between the unconditional and conditional means? On one hand, the modest  $t$  statistic implies a substantial probability of observing a slope coefficient greater than the estimate of 0.2 even if stock returns are completely unpredictable. If returns are unpredictable, then an investor who uses the estimated 32-basis-point difference in expected monthly returns would probably tend to *under*-allocate to stocks. On the other hand, stock returns could be predictable in precisely the manner estimated, in which case an investor who simply ignores the estimated 32-basis-point difference would probably tend to *over*-allocate to stocks. How should the investor's asset-allocation decision depend on the current dividend yield?

We suggest that the answer to the above asset-allocation problem provides an economi-

---

<sup>1</sup>Fama and French (1988) present the results of similar regressions and conclude that dividend yields account for only small fractions (typically less than 5%) of the variance of monthly returns. To our knowledge, the earliest study that reports a time-series regression of stock returns on dividend yields is Rozeff (1984).

cally relevant metric by which to characterize the strength of the regression evidence on the predictability of stock returns. Our framework assumes a risk-averse investor with initially vague beliefs about the distribution of stock returns. The investor uses the above regression evidence to update those beliefs, and these revised beliefs are then used by the investor to compute the optimal asset allocation. We find that the asset allocation chosen by the investor depends importantly on the level of the current dividend yield. For example, an investor with relative risk aversion equal to 2 would allocate about 60% to stocks with a current yield of 2.71%, whereas the same investor would allocate nearly 100% to stocks if the current yield were instead equal to its long-run average of 4.35%.

The importance of dividend yield in the allocation decision also appears to be fairly robust to two additional considerations. The first consideration, known commonly as “data mining,” is that the forecasting variable, dividend yield in this case, might simply have been selected *ex post* as the best predictor within a larger set of regressors. As argued by Keim and Stambaugh (1986), given that an asset’s current price is inversely related to the discount rate applied to expected future cash flows, variables that are inversely related to price levels are reasonable candidates *ex ante* as predictors of returns. At high frequencies, dividends are smooth relative to stock prices, so the dividend yield exhibits a strong inverse association with the level of equity prices and thereby arises as a plausible predictive variable. Dividend yield can also be identified as a predictive variable in more formal equilibrium models with time-varying moments of consumption growth, such as in Kandel and Stambaugh (1991). Thus, although we select dividend yield in part because it occupies a central role in much of the recent empirical literature on the predictability of stock returns, we would suggest that the data-mining critique is not well placed here. Nevertheless, given that data-mining is often raised as a concern in discussions of return predictability, some consideration of the issue seems appropriate.<sup>2</sup> We construct a data-mining scenario by allowing there to be a given number of additional regressors, all of which, when included in a multiple regression along with dividend yield, receive zero estimated slopes in the 1927–93 sample. The investor’s prior beliefs must admit the possibility, however, that all of the variables might predict returns.

The second additional consideration is the assumption made regarding the investor’s prior beliefs. Rather than assuming that the investor has vague beliefs about the distribution of stock returns, we also examine cases where the investor possesses informed beliefs about

---

<sup>2</sup>Foster and Smith (1994) address the data-mining problem in the context of testing the point null hypothesis of no predictability using a frequentist approach. They obtain cutoff  $R^2$  values that are adjusted to account for data mining, and they conclude that some of the reported magnitudes of  $R^2$  in models for predicting monthly stock returns could be achieved with reasonably high probability by searching across a relatively small number of regressors.

the stock return's mean and variance, and these informed beliefs are weighted against the predictability of stock returns. We represent such an "informative no-predictability" prior as being equivalent to having observed fifty years of hypothetical data in which stock returns have the same sample mean and variance as in the actual sample but where the estimated slope coefficients on all predictive variables are *exactly* zero.<sup>3</sup>

These additional considerations must reduce the importance of dividend yield in the allocation decision, but a substantial role for that variable can still remain. For example, with the informative no-predictability prior and with dividend yield having been mined from a data set containing 100 additional variables, the investor with relative risk aversion of 2 would allocate 70% to stocks at the current yield of 2.71%, but the same investor would allocate 94% to stocks if the yield were instead equal to its long-run average of 4.35%. We conclude that, when judged by this economic metric, the sample evidence reveals an important degree of predictability in monthly stock returns.

The remainder of the paper proceeds as follows. Section 2 presents the methodological framework in which the investor's optimal asset allocation is computed. This framework is then used in section 3 to compute optimal allocations over a range of values for the current dividend yield. The roles of risk aversion, prior beliefs, sample size, and data mining are explored, and the value to the investor of entertaining return predictability is investigated by comparing "certainty-equivalent" rates of return. Section 4 shows that the  $R^2$  of the regression in (1) plays the key role in determining the sensitivity of the optimal stock allocation to the predictive variable, and additional insight into that role is provided in an infinite-sample setting. Our investigation of return predictability in this study reveals that statistical significance, as judged by standard frequentist methods, can contrast sharply with economic significance. In order to demonstrate that such a result does not hinge on statistical assumptions specific to our regression framework, we present in section 5 a simple two-state, two-outcome example that illustrates a similar contrast between economic and statistical significance. Section 6 concludes the paper and suggests directions for future research.

---

<sup>3</sup>In fact, Goetzmann and Jorion (1994) report estimates of the regression in (1) using pre-1927 data that differ from the post-1927 results. Using Cowles monthly U.S. indexes for NYSE stocks for the period from 1872-1926, they obtain an  $R^2$  of 0.003, a slope of  $-0.142$ , and a  $t$ -statistic of  $-1.41$ . They find, however, a somewhat stronger relation for the post-1927 period than we report in (1), apparently because they use returns and dividends for the S&P Composite Index instead of the value-weighted NYSE portfolio. As a consequence, their overall-period results actually correspond more closely to our post-1927 results: for the entire 1872-1992 period they report an  $R^2$  of 0.002, a slope of 0.159, and a  $t$  statistic of 1.79.

## 2. The Allocation Decision

### 2.1. The Investor's Objective

Let  $\omega$  denote the fraction of the investor's portfolio allocated to stocks, where  $0 \leq \omega \leq 1$ . For an allocation of  $\omega$  in stocks at the end of month  $T$ , the investor's wealth at the end of month  $T + 1$ , per \$1 invested at time  $T$ , is

$$W_{T+1} = \omega \exp\{r_{T+1}\} + (1 - \omega) \exp\{i_T\}, \quad (2)$$

where  $i_T$  is the continuously compounded riskless rate on cash for month  $T + 1$ , observed at the end of month  $T$ . The investor chooses  $\omega$  so as to maximize the expected value of the utility function

$$v(W) = \begin{cases} \frac{1}{1-A} W^{1-A} & \text{for } A > 0 \text{ and } A \neq 1, \\ \ln W & \text{for } A = 1. \end{cases} \quad (3)$$

That is, if  $\Phi_T$  denotes the data set observed by the investor through the end of month  $T$ , then the investor is assumed to solve

$$\max_{\omega} \int v(W_{T+1}) p(r_{T+1} | \Phi_T) dr_{T+1}, \quad (4)$$

where  $p(r_{T+1} | \Phi_T)$  denotes the density of  $r_{T+1}$  conditional on  $\Phi_T$ .

The parameter  $A$  in the iso-elastic utility function in (3) is commonly referred to as the investor's coefficient of relative risk aversion, and we will follow that conventional terminology. When returns are not independently and identically distributed, however, and when an investor chooses consumption and asset allocations in a dynamic multiperiod setting, the value of  $A$  defined with respect to the investor's derived utility of wealth can differ dramatically from that investor's relative risk aversion defined with respect to utility for consumption in any single period. Black (1990) presents a continuous-time model in which both the direct utility of consumption,  $u(C)$ , and the derived utility of wealth,  $v(W)$ , are of the same iso-elastic form, except that the risk aversion parameter in  $u(C)$  can be much higher than the corresponding parameter in  $v(W)$ . In essence, the investor's ability to smooth consumption translates into a greater tolerance for risk when making the asset-allocation decision.<sup>4</sup> We will entertain values of  $A$  from 1 to 20, often concentrating on the lower end of that range, but the reader should bear in mind that such values can be consistent with higher values for risk aversion defined with respect to the direct utility of consumption.

<sup>4</sup>In a similar vein, Kandel and Stambaugh (1991) demonstrate that, in a multiperiod setting, the "price of risk" (the ratio of the mean excess stock return to variance of the return), often used to infer values of risk aversion, can be much lower than the coefficient of relative risk aversion defined with respect to the direct utility of consumption.

## 2.2. The Conditional Distribution of the Stock Return

The investor assesses a conditional distribution of the stock return by assuming a likelihood function with unknown parameters and then updating initially vague beliefs about those parameters using the data. We model this process using a Bayesian regression model with diffuse priors. The continuously compounded stock return  $r_t$  is the dependent variable in the regression

$$r_t = x'_{t-1}\beta + \epsilon_t, \quad (5)$$

where the  $K$ -element vector of independent variables is given by  $x'_{t-1} = (1 \ d_{t-1} \ z_{t-1})$ ,  $z_{t-1}$  contains  $N$  variables that are observed at the end of month  $t - 1$ , and  $K = N + 2$ . When  $N = 0$ , then  $d_{t-1}$  is the only independent variable (other than the constant). The disturbances  $\epsilon_t$ ,  $t = 1, 2, \dots, T$ , are assumed to be independent mean-zero draws from a normal distribution with variance  $\sigma_\epsilon^2$ . Although we assume  $E\{\epsilon_t|x_{t-1}\} = 0$ , the vector  $x_{t-1}$  is in general stochastic, and some elements of  $x_{t-1}$  can be correlated with past disturbances.<sup>5</sup> For example, the dividend yield at the end of month  $t - 1$  is likely to be negatively correlated with the unexpected return in that month,  $\epsilon_{t-1}$ . Thus, the regression in (5) departs somewhat from the standard Bayesian regression framework, in which  $x_{t-1}$  is assumed to be either nonstochastic or stochastic but distributed independently of the disturbances with a distribution involving neither  $\beta$  nor  $\sigma_\epsilon$ .<sup>6</sup>

Our approach to modeling the stochastic properties of the regressors is to include them as dependent variables in a set of  $N + 1$  additional regression equations. That is, let  $y'_t = (r_t \ d_t \ z_t)$ , and define the multivariate regression

$$y'_t = x'_{t-1}B + u'_t, \quad (6)$$

where  $B$  is a  $K \times K$  matrix of regression coefficients whose first column is  $\beta$ , and  $u_t$  is a  $K$ -vector of disturbances whose first element is  $\epsilon_t$ . In other words, the elements of  $x_{t-1}$  are modeled as a first-order vector autoregression (VAR). We assume that the vectors  $u_t$ ,  $t = 1, \dots, T$ , are independent mean-zero draws from a multivariate normal distribution with a covariance matrix equal to  $\Sigma$ . The  $T$  observations for this multivariate regression model are represented in the matrix notation,

$$Y = XB + U, \quad (7)$$

<sup>5</sup>See Stambaugh (1986) and Nelson and Kim (1993) for treatments of this problem in a frequentist setting.

<sup>6</sup>See, for example, page 59 of Zellner (1971).

where

$$Y = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_T \end{bmatrix}, X = \begin{bmatrix} x'_0 \\ x'_1 \\ \vdots \\ x'_{T-1} \end{bmatrix}, \text{ and } U = \begin{bmatrix} u'_1 \\ u'_2 \\ \vdots \\ u'_T \end{bmatrix}. \quad (8)$$

Note that rows 1 through  $(T - 1)$  in  $Y$  are simply rows 2 through  $T$  in  $X$ , excluding the first column of ones in the latter matrix. Following an approach common to many Bayesian time-series models, the analysis will proceed conditional on the “pre-sample” observation  $x_0$ . The joint probability density function for the elements of  $U$  is given by

$$p(U|\Sigma) \propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}\text{tr}U'U\Sigma^{-1}\right], \quad (9)$$

where “tr” denotes the trace operator. The Jacobian of the transformation from  $U$  to  $Y$  is equal to unity, so the likelihood function is given by

$$P(Y|B, \Sigma, x_0) \propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}\text{tr}(Y - XB)'(Y - XB)\Sigma^{-1}\right]. \quad (10)$$

In our initial investigation, we assume that the investor’s prior beliefs about the unknown parameters, before observing the data, are “diffuse” and represented as

$$p(B, \Sigma) \propto |\Sigma|^{-(K+1)/2}. \quad (11)$$

This prior density is intended to represent vague initial beliefs about the parameters of the regression model. As will be explained later, we also consider an alternative set of prior beliefs that are weighted against the predictability of stock returns.

The investor’s data set  $\Phi_T$  contains  $Y$  and  $x_0$ . Note that the vector of independent variables at the end of month  $T$  is  $x'_T = (1 \ y'_T)$ , where  $y'_T$  is simply the last row of  $Y$ . The distribution of  $r_{T+1}$  conditional on the data is known in Bayesian analysis as the “predictive pdf.” With the prior distribution as in (11) and the likelihood function as in (10), the predictive pdf for  $r_{T+1}$  is given by a Student  $t$  distribution:

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma(1/2)\Gamma(\nu/2)} \left(\frac{h}{\nu}\right)^{1/2} \left[1 + \frac{h}{\nu}(r_{T+1} - x'_T\hat{\beta})^2\right]^{-(\nu+1)/2}, \quad (12)$$

where

$$\hat{\beta} = (X'X)^{-1}X'y, \quad (13)$$

$$h = \frac{1}{s^2}[1 - x'_T(X'X + x_Tx'_T)^{-1}x_T], \quad (14)$$

$$s^2 = \frac{1}{\nu}(y - X\hat{\beta})'(y - X\hat{\beta}), \quad (15)$$



$$\nu = T - 2N - 3, \quad (16)$$

and  $y$  is the first column of  $Y$ . When  $N = 0$ , so that  $d_T$  is the only variable used in predicting returns, then (14) can be rewritten as

$$h = \frac{1}{s^2} \left[ 1 + \frac{1}{T} \left( 1 + \frac{(d_T - \bar{d})^2}{\hat{\sigma}_d^2} \right) \right]^{-1}, \quad (17)$$

where

$$\bar{d} = \frac{1}{T} \sum_{t=0}^{T-1} d_t, \text{ and} \quad (18)$$

$$\hat{\sigma}_d^2 = \frac{1}{T} \sum_{t=0}^{T-1} (d_t - \bar{d})^2. \quad (19)$$

Additional details are provided in the Appendix.

The conditional distribution in (12) is used in the maximization in (4), with the integration from  $-\infty$  to  $\infty$ . When  $A > 1$ , expected utility is equal to  $-\infty$  when  $\omega = 1$ , although the optimal  $\omega$  can be very close 1. We simply restrict  $\omega \leq .99$  throughout. The maximization problem, a special case of the Bayesian one-period control problem discussed by Zellner (1971, pp. 320-327), is solved numerically, since we are unaware of an exact analytic solution.<sup>7</sup> In particular, an analytic solution cannot consist of a simple mean-variance-based allocation formula, because the mean and the variance of the simple rate of return,  $R_{T+1} = \exp(r_{T+1})$ , do not exist when  $T$  is finite.<sup>8</sup> The first and second conditional moments of the continuously compounded stock return are given by

$$E\{r_{T+1}|\Phi_T\} = x'_T \hat{\beta}, \text{ and} \quad (20)$$

$$\text{var}\{r_{T+1}|\Phi_T\} = \frac{\nu}{h(\nu - 2)}, \quad (21)$$

where we impose  $\nu > 2$  throughout the analysis.

<sup>7</sup>The numerical solution is obtained using Brent's method with parabolic interpolation for the maximization and an adaptive recursive Newton-Cotes eight-panel rule to evaluate the integral. See Brent (1973), Forsythe, Malcolm, and Moler (1977), and Press et. al. (1986).

<sup>8</sup>This follows from a similar observation about the moment-generating function of the Student  $t$  distribution, as noted in Kendall and Stuart (1977, p. 63). A mean-variance-based solution to the one-period Bayesian portfolio problem with i.i.d. returns is obtained by Bawa, Brown, and Klein (1979). Their assumptions imply a Student  $t$  predictive distribution for the *simple* rate of return (permitting infinitely negative wealth), and they obtain a mean-variance solution to the expected-utility maximization by approximating that distribution as being Normal. A number of studies have investigated the use of Bayesian estimation techniques in asset allocation where a mean-variance optimization is specified *a priori* rather than derived as the outcome of maximizing expected utility under a Bayesian predictive distribution. A partial list of such studies includes Jobson, Korkie, and Ratti (1979), Jobson and Korkie (1980), Jorion (1985, 1986, 1991), and Frost and Savarino (1986). Another approach is explored by Grauer and Hakansson (1992), who maximize expected utility using a historical series of returns as the possible outcomes in a discrete predictive distribution, where each historical outcome is mean-adjusted using a Bayesian estimator of expected returns.

### 2.3. Data Mining

The regression presented in (1) has only one independent variable,  $d_{t-1}$ . Suppose it were the case, however, that this choice is made by observing that, if dividend yield is included along with the  $N$  additional independent variables in  $z_{t-1}$ , then only the ordinary-least-squares coefficient on  $d_{t-1}$  is nonzero. Such a scenario offers a version of “data mining,” in which  $d_{t-1}$  is presented as the independent variable in the simple regression model only because none of the other  $N$  variables receive non-zero coefficients in the multiple regression.

Even when the sample slope coefficients on the  $N$  additional variables in  $z_{t-1}$  are zero, the conditional distribution of  $r_{T+1}$  is different from that obtained when  $N = 0$ , where the additional variables are excluded *a priori*. The conditional mean of  $r_{T+1}$  is unaffected, but, from (16), increasing  $N$  reduces the degrees of freedom parameter  $\nu$ , which, given (21), increases the conditional variance of  $r_{T+1}$ . Moreover, the value of  $h$  in (14) is made smaller by the inclusion of the additional variables, which, again using (21), also increases the conditional variance of  $r_{T+1}$ .

We can analyze the effect of the additional variables on the quantity  $h$  in the following manner. Define

$$\bar{z} = \frac{1}{T} \sum_{t=0}^{T-1} z_t, \text{ and} \quad (22)$$

$$\hat{\Sigma}_z = \frac{1}{T} \sum_{t=0}^{T-1} (z_t - \bar{z})(z_t - \bar{z})'. \quad (23)$$

Assume also that, in the sample, the  $N$  additional variables are uncorrelated with dividend yield, i.e.,

$$\frac{1}{T} \sum_{t=0}^{T-1} (z_t - \bar{z})(d_t - \bar{d}) = 0. \quad (24)$$

Then, as shown in the Appendix,  $h$  in (14) can be rewritten as

$$h = \frac{1}{s^2} \left[ 1 + \frac{1}{T} \left( 1 + \frac{(d_T - \bar{d})^2}{\hat{\sigma}_d^2} + (z_T - \bar{z})' \hat{\Sigma}_z^{-1} (z_T - \bar{z}) \right) \right]^{-1}, \quad (25)$$

which is decreasing in the quantity,

$$q = (z_T - \bar{z})' \hat{\Sigma}_z^{-1} (z_T - \bar{z}). \quad (26)$$

Providing reasonable hypothetical values for the sample quantity  $q$  is difficult in general, but a simple distributional result obtains for a special case. If  $z_0, z_1, \dots, z_T$  are independent

draws from a multivariate normal distribution, then<sup>9</sup>

$$\frac{T - N}{N(T + 1)}q \text{ is distributed as } F_{N, T-N}. \quad (27)$$

Although successive realizations of  $z_t$  are likely to be correlated in general, with nonzero coefficients appearing for those variables in the VAR defined within (6), we use (27) in later computations designed to gauge, at least roughly, the potential impact of data mining on the asset-allocation decision. The value of  $q$  is set using the 99th percentile of the  $F$  distribution in (27). This extreme value is chosen as a partial attempt to recognize that typical values of  $q$  are likely to be larger if the elements of  $z_t$  exhibit positive serial dependence.

#### 2.4. Prior Beliefs Weighted Against Predictability

With the diffuse prior in (11), the density for the regression coefficients is uniform over the real line. We also examine the asset-allocation decision using a prior density that assigns greater probability to regression slope coefficients in the neighborhood of zero. We construct this “informative no-predictability” prior by first combining the diffuse prior in (11) with a hypothetical sample of size  $T_0$  in which the sample regression coefficients on dividend yield and the  $N$  additional predictive variables are exactly zero. The posterior distribution for the regression parameters  $B$  and  $\Sigma$  obtained from that hypothetical sample is then used as the prior in computing the predictive distribution for  $r_{T+1}$  based on the actual sample of size  $T$ . This predictive distribution must, by basic principles of Bayesian analysis, be the same as that obtained when the diffuse prior is combined with one sample of size  $T^* = T_0 + T$  that simply combines the hypothetical and actual samples.

We assume for simplicity that, if  $X_0$  denotes the  $T_0 \times K$  matrix containing the observations of  $x_t$  in the hypothetical sample, then

$$X_0'X_0 = \frac{T_0}{T}X'X, \quad (28)$$

and we also assume that the sample means and variances of  $r_t$  are the same in the actual and hypothetical samples. Define

$$\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t, \quad (29)$$

$$\hat{\sigma}_r^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2, \quad (30)$$

---

<sup>9</sup>See, for example, chapters 5 and 7 of Anderson (1984).

$$\beta_0 = [\bar{r} \ 0 \ \dots \ 0]', \quad (31)$$

$$\beta^* = \left(\frac{T_0}{T^*}\right) \beta_0 + \left(\frac{T}{T^*}\right) \hat{\beta}, \quad (32)$$

$$s^{*2} = \frac{1}{\nu^*} \left[ T_0 \hat{\sigma}_r^2 + (y - X\hat{\beta})'(y - X\hat{\beta}) + \left(\frac{T_0}{T^*}\right) (\hat{\beta} - \beta_0)' X' X (\hat{\beta} - \beta_0) \right], \quad (33)$$

$$h^* = \frac{1}{s^{*2}} \left( 1 - x_T' \left[ \left(1 + \frac{T_0}{T}\right) X' X + x_T x_T' \right]^{-1} x_T \right), \text{ and} \quad (34)$$

$$\nu^* = T^* - 2N - 3. \quad (35)$$

The predictive distribution of  $r_{T+1}$  is of the same Student- $t$  form as in (12):

$$p(r_{T+1} | \Phi_T) = \frac{\Gamma((\nu^* + 1)/2)}{\Gamma(1/2)\Gamma(\nu^*/2)} \left(\frac{h^*}{\nu^*}\right)^{1/2} \left[ 1 + \frac{h^*}{\nu^*} (r_{T+1} - x_T' \beta^*)^2 \right]^{-(\nu^* + 1)/2}. \quad (36)$$

When  $N = 0$ , or under the data-mining scenario described previously (where the last  $N$  elements of  $\hat{\beta}$  equal zero), equations (33) and (34) can be simplified as

$$s^{*2} = \frac{1}{\nu^*} \left[ T_0 \hat{\sigma}_r^2 + (y - X\hat{\beta})'(y - X\hat{\beta}) + \left(\frac{T_0 T}{T^*}\right) \hat{\beta}_{2,1}^2 \hat{\sigma}_d^2 \right], \quad (37)$$

and

$$h^* = \frac{1}{s^{*2}} \left[ 1 + \frac{1}{T^*} \left( 1 + \frac{(d_T - \bar{d})^2}{\hat{\sigma}_d^2} + (z_T - \bar{z})' \hat{\Sigma}_s^{-1} (z_T - \bar{z}) \right) \right]^{-1}, \quad (38)$$

where  $\hat{\beta}_{2,1}$  is the second element of  $\hat{\beta}$ . The value of  $q$  in (26) is constructed using the 99th percentile of the  $F$  distribution in (27), except that  $T^*$  replaces  $T$ .

### 3. Asset-Allocation Results

In this section we use the regression estimates reported in (1) to compute the optimal allocation in stocks,  $\omega$ , conditional on different hypothetical values of the most recent dividend yield,  $d_T$ . The objective is to explore the economic significance of stock-return predictability by analyzing the sensitivity of  $\omega$  to this single predictive variable. We first present results based on the diffuse prior in (11), and then we present results based on a no-predictability informative prior with a 50-year hypothetical prior sample.

#### 3.1. Allocations Based on a Diffuse Prior

Figure 1 presents optimal allocations based on returns and dividends for the 804-month period from 1927–93. The riskless rate  $r_T$  is held constant at 0.235%, the continuously

compounded monthly yield on the Treasury Bill with 27 days to maturity as of 12/31/93.<sup>10</sup> In computing these results we set  $N = 0$ , so there is no allowance for additional independent variables. The figure displays, for five values of relative risk aversion ( $A$ ), the optimal allocation in stocks as a function of current dividend yield. The range for the dividend yield is centered at the average yield for the sample, 4.35%, and extends two sample standard deviations ( $2 \times 1.36\%$ ) above and below that value. For logarithmic utility ( $A = 1$ ), the optimal stock allocation is only 0.35 at the lowest dividend yield of 1.6%, but the optimal allocation rises to 1.0 for yields of about 3% or greater. For an investor with  $A = 2$ , the optimal allocation in stocks begins at only 0.18 for the lowest yield but rises to an allocation of essentially all-stocks at yields of about 4% or greater.<sup>11</sup> At higher levels of risk aversion ( $A = 5, 10, \text{ and } 20$ ), the optimal stock allocation remains less than 1.0 for the range of yields considered, but the differences in  $\omega$  across the range of yields are substantial. Even when  $A = 20$ , for example, the allocation to stocks increases by roughly a factor of 10 (.02 versus .2) as the dividend yield increases from 1.6% to 7.1%.

Figure 2 presents the same analysis as figure 1, except that the number of additional independent variables ( $N$ ) is set to three hundred instead of zero. Even with this very large number of additional independent variables, the current dividend yield still plays an important role in the asset-allocation decision, particularly at the lowest levels of risk aversion. With logarithmic utility, for example, the optimal allocation to stocks is about 0.50 when the dividend yield is only 1.6% but increases to 1.00 when the dividend yield is 7.1%. An investor with  $A = 2$  doubles the allocation to stocks, from 0.25 to 0.50, as the yield moves from 1.6% to 7.1%. Although the fraction allocated to stocks is low in any event for the higher levels of risk aversion, substantial *relative* increases in the stock allocation occur in those cases as well. For example, an investor with  $A = 10$  also doubles the stock allocation (from 0.05 to 0.10) over the range of dividend yield considered. The general effect of raising  $N$  is both to flatten the relation between the optimal stock allocation and dividend yield and to lower the allocation at the average yield.

The basic message delivered by figures 1 and 2 is that the evidence on the predictability of stock returns is strong enough to exert a substantial influence on an investor's allocation of funds between stocks and cash. That is, if the investor initially possesses vague beliefs, as represented by a diffuse prior, then the sample contains information sufficient to cause the investor to allocate funds in a manner that depends importantly on the current dividend

<sup>10</sup>This value is obtained from the CRSP Government Bond Files.

<sup>11</sup>Recall that, when  $A > 1$ , expected utility falls to  $-\infty$  as  $\omega$  approaches 1.0, so the optimal allocation in these cases is close but not equal to 1.0.

yield. Obviously, this description of the data contrasts sharply with the message delivered by the standard approach to judging “significance” of a regression.

Table 1 reports additional results that illustrate the separate roles of sample size ( $T$ ) and the number of additional variables ( $N$ ) in the asset-allocation decision. In order to isolate  $T$  from other quantities that would change across samples of different sizes, the values of  $\hat{\beta}$ ,  $\bar{d}$ ,  $\hat{\sigma}_d^2$ , and  $(\nu/T)s^2$  are held constant at the values obtained from the actual 1927–93 sample. That is, we change only the size of the samples that generate a common set of sufficient statistics. Optimal stock allocations are computed under various combinations of  $T$  and  $N$  for relative risk aversion equal to both two and ten.

When  $\nu$  ( $= T - 2N - 3$ ) is small, which occurs with either a small value of  $T$  or a sufficiently large value of  $N$ , then the optimal stock allocations increase only slightly as dividend yield moves from low to high. In these cases, such as ( $T = 6, N = 0$ ) or ( $T = 804, N = 350$ ), the evidence of predictability is sufficiently weak so as to render the current dividend yield relatively unimportant in the allocation decision. Note however that, when  $N = 0$ , the optimal allocations reach their large-sample values at rather modest sample sizes: the allocations at  $T = 60$  differ only slightly from those at  $T = 804$ . (Recall that  $T = 804$  corresponds to the actual 1927–93 sample.) In other words, when there is known to be at most one predictive variable, estimation risk quickly becomes unimportant to the stock-versus-cash allocation decision. A rather quick disappearance of estimation risk is also observed in earlier studies that assume returns to be unpredictable, so that the unconditional mean and variance are the only unknown parameters. For example, Brown (1979) reports that the optimal stock allocation in such a problem is typically at least 95% of the infinite-sample allocation with samples as small as 50 months.<sup>12</sup> In general, the results in table 1 indicate that, even when the number of independent variables is a substantial fraction of the sample size, the optimal allocation depends importantly on the predictive variable.

Table 2 reports the conditional standard deviations of the continuously compounded monthly stock return  $r_{T+1}$ , and these values provide some intuition for the allocations in table 1. When  $T$  is large and  $N$  is small, then the conditional standard deviation of the monthly stock return is about 5.6%, which is no doubt familiar to most readers as a typical magnitude for the monthly stock volatility. As is apparent from table 2, however, when  $T$  is small or when  $N$  is large relative to  $T$ , then the conditional volatility of the monthly stock

<sup>12</sup>See his table 8.1, page 115. Brown assumes diffuse priors for the unconditional mean and variance, a likelihood function in which the *simple* rate of return (as opposed to the continuously compounded rate) is normally distributed, and a utility function that is either quadratic or negative exponential.

return can be several times greater, reaching 20% or more.<sup>13</sup> In those cases, the “estimation risk” arising from uncertainty about the parameters of the return distribution accounts for much of the total conditional volatility. This higher estimation risk reduces the importance of the predictive variable  $d_T$  in the allocation decision, as discussed previously, but this higher risk also lowers the allocation to stocks at the average level of  $d_T$  (4.35%). The allocations at the average level of  $d_T$  in table 1 are virtually identical to the allocations that would be chosen if the sample regression coefficient on  $d_{t-1}$  in (5) were zero (along with the coefficients on  $z_{t-1}$ ).<sup>14</sup> In that complete absence of sample evidence of predictability, the investor would essentially allocate the same amount to stocks for all levels of  $d_T$ , and this allocation is lower in the cases where the conditional volatility of stocks is high.

From the perspective of the investor, how important is it that predictability, or at least its possibility, be taken into account in making the allocation decision? We pursue this question by comparing the “certainty equivalents” for two different allocations, where a certainty equivalent is defined as the riskless rate that would provide utility equal to the expected utility of the chosen allocation. The first allocation,  $\omega_1$ , is taken from table 1. The second allocation,  $\omega_2$ , uses data only on historical returns, and the investor’s prior beliefs rule out predictability. That is, the likelihood function assumes that returns are drawn independently from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and the investor has diffuse prior beliefs about  $\mu$  and  $\sigma$ . For a given  $T$ , the second allocation is very close to that from table 1 for the case when  $N = 0$  and  $d_T$  is at its average level (4.35%). The certainty equivalents for both allocations,  $c(\omega_1)$  and  $c(\omega_2)$ , are then computed with respect to the predictive distribution used to compute the first allocation, which uses data for returns as well as the predictive variables and where the investor’s prior beliefs admit the possibility that returns are predictable. In essence, an investor who perceives this predictive distribution but is forced to allocate  $\omega_2$  to stocks values this allocation with a certainty equivalent of  $c(\omega_2)$ , which is necessarily less than the certainty equivalent of  $c(\omega_1)$  for the optimal allocation.

The difference in certainty equivalents,  $c(\omega_1) - c(\omega_2)$ , is reported in table 3, where the values of  $T$ ,  $N$ , and relative risk aversion ( $A$ ) are the same as in table 1. One source of large

---

<sup>13</sup>Much of the increase in the conditional variance that occurs as  $N$  increases relative to  $T$  can be attributed to the fact that the quantity  $\nu = T - 2N - 3$  is used in (15) as the denominator for computing  $s^2$ , which enters the conditional variance according to (21) and (25). That is, for a given sum of  $T$  squared fitted residuals, the conditional volatility increases in the number of variables used to fit the regression. This effect is analogous to that in the standard frequentist setting for a multiple regression with  $K$  regressors (including the intercept), where the unbiased estimate of the residual variance is obtained by dividing the sum of squared fitted residuals by  $T - K$ , whatever the estimated values of the  $K - 1$  slope coefficients.

<sup>14</sup>For a given  $T$  and  $N$ , the allocation would be identical to that in table 1 at the average level of  $d_T$  and slightly less than that allocation at the more extreme values of  $d_T$ .

differences is an over-allocation to stock when the variance of its predictive distribution is high. When  $T$  is sufficiently large, the variance of the pdf of  $r_{T+1}$  can be much lower when one assumes i.i.d. behavior than when one entertains the possibility that a large number of variables might predict returns. Consider, for example, the case where  $T = 804$  and  $N = 300$ . An investor with relative risk aversion equal to 2 would, under the i.i.d. assumption, assess the monthly standard deviation of stocks to be about 5.6% and would allocate nearly 100% of his portfolio to stocks ( $\omega_2$ ). In contrast, the optimal allocation when predictability is entertained ( $\omega_1$ ) ranges from 0.25 to 0.50 across the various levels of  $d_T$ . Thus,  $\omega_2$  exceeds  $\omega_1$  by between 0.75 and 0.50, and this over-allocation to stocks occurs when the standard deviation of the predictive distribution for the stock return is nearly 15% per month (table 2). Not surprisingly, the certainty-equivalent losses in this case range between 6.8% and 15.8%, expressed as differences in annualized rates.

Another source of differences in certainty equivalents is an over-allocation to stock when its conditional expected return is low or an under-allocation when its conditional expected return is high. When  $T = 804$  and  $N = 0$ , the variance of the predictive distribution used to compute the optimal allocation  $\omega_1$  is essentially identical to the variance of the predictive distribution under the i.i.d. assumption used to compute  $\omega_2$ . Thus, in this case, the difference between  $\omega_1$  and  $\omega_2$  essentially reflects only the difference between the conditional mean of the predictive pdf for  $r_{T+1}$ , given  $d_T$ , and the “unconditional” mean in the pdf obtained under the i.i.d. assumption. An investor with relative risk aversion equal to 10 would, under the i.i.d. assumption, choose a stock allocation of about 0.23 ( $\omega_2$ ). The optimal allocation  $\omega_1$  is 0.05 at  $d_T = 1.64\%$  but rises to 0.40 at  $d_T = 7.05\%$ . Thus,  $\omega_2$  exceeds  $\omega_1$  by 0.18 at the low dividend yield but is less than  $\omega_1$  by about that amount at the high yield. In both cases, the certainty-equivalent loss associated with  $\omega_2$  is about 60 basis points. For the investor with risk aversion equal to 2, the certainty-equivalent loss associated with  $\omega_2$  is 225 basis points when  $d_T = 1.64\%$ , resulting from an over-allocation to stocks of about 0.73, whereas the loss in the high-yield state is negligible, since  $\omega_1$  and  $\omega_2$  are then both equal to essentially 100%.

### 3.2. Allocations with an Informative Prior

When the investor’s prior beliefs are weighted against predictability, then the optimal allocation to stock should, one might think, be less sensitive to the current value of the predictive variable than when prior beliefs are diffuse. We explore this intuition by recomputing all of the above results using the informative no-predictability prior described earlier. The size of the hypothetical prior sample is 50 years (or  $T_0 = 600$ , since we are dealing with monthly



returns). Figures 3 and 4 and tables 4 through 6 display the results for the same cases analyzed above for the diffuse prior.

The allocations displayed in figure 3 are based on the actual 804-month sample for the case where there are no additional independent variables ( $N = 0$ ). These allocations conform to the simple intuition suggested above: they rise less steeply as a function of dividend yield than do the allocations with a diffuse prior (figure 1). With the informative prior, the log-utility investor ( $A = 1$ ) allocates 100% to stocks throughout the range of  $d_T$  considered, so that, for example, a dividend yield of 1.6% no longer causes such an investor to back away from an all-stock position. At higher levels of risk aversion, however, the allocations still display considerable sensitivity to dividend yield. The investor with  $A = 2$  allocates essentially 100% to stocks for yields greater than about 3.5%, but the stock allocation drops to 65% at a yield of 1.6%. For the higher levels of risk aversion, the allocations at least double as dividend yield moves from the bottom to the top of the range considered. Thus, even with prior beliefs equivalent to 50 years of hypothetical data in which the sample regression coefficient of returns on dividend yield is *exactly* zero, the evidence of predictability in the actual 804-month sample is still sufficient to give the predictive variable an important role in the asset-allocation decision.

Perhaps more interesting are the allocations in figure 4, which are also based on the actual 804-month sample but with 300 additional independent variables ( $N = 300$ ). In this case, the allocations that use the no-predictability informative prior exhibit *greater* sensitivity to dividend yield than do those using the diffuse prior (figure 2). This somewhat counter-intuitive result can be understood by comparing the volatilities of the predictive distributions obtained using the two priors. Recall that, with the diffuse prior, the conditional standard deviation of the continuously compounded monthly stock return ( $r_{T+1}$ ) is nearly 15% when  $T = 804$  and  $N = 300$  (table 2). With the informative prior, however, this standard deviation, from table 5, is only about 8.5%. In essence, the informative prior lowers estimation risk. Even though the conditional mean of  $r_{T+1}$  is less sensitive to  $d_T$  with the informative prior, this effect is outweighed in the allocation decision by the lower predictive variance. Thus, dividend yield retains an important role in the allocation decision even when prior beliefs are weighted against predictability and there are many additional independent variables that receive zero estimated regression slopes (in both the hypothetical and actual samples).

Tables 4 and 6 report, for the informative no-predictability prior, the allocations and certainty-equivalent comparisons for the same sets of  $T$ ,  $N$ , and  $A$  as reported in tables 1

and 3 for the diffuse prior. Comparing the allocations in table 4 to those in table 1 further illustrates the phenomenon that moving from the diffuse to the informative prior reduces both estimation risk and the sensitivity of  $E\{r_{T+1}|\Phi_T\}$  to  $d_T$ . In cases where estimation risk is already low using the diffuse prior, the reduced sensitivity of  $E\{r_{T+1}|\Phi_T\}$  to  $d_T$  exerts the more important effect on the allocation decision. Thus, when  $T$  is sufficiently large relative to  $N$ , the allocations are less sensitive to  $d_T$  using the informative prior. This result occurs, for example, in most of the cases where  $N = 0$ . In cases where  $T$  is small or  $N$  is large, so that there is substantial estimation risk using the diffuse prior, then lowering that estimation risk exerts the more important effect, so the allocation then becomes more sensitive to  $d_T$ . In these cases, the allocation to stocks at the average level of  $d_T$  also increases, again due to the reduced stock volatility. For example, when  $T = 360$  and  $N = 150$ , an investor with risk aversion of 2 and the informative prior chooses stock allocations ranging from 0.56 when  $d_T = 1.64\%$  to 0.92 when  $d_T = 7.05\%$ , and the average allocation, chosen at the average yield of 4.35%, is equal to 0.74. If the same investor instead has the diffuse prior, then those allocations range from 0.25 to 0.34 and average only 0.30.

Moving from the diffuse to the informative no-predictability prior generally reduces the certainty-equivalent loss associated with the suboptimal allocation chosen using the i.i.d. assumption. In cases where the estimation risk is already low using a diffuse prior, the reduced sensitivity of  $E\{r_{T+1}|\Phi_T\}$  to  $d_T$  simply reduces the deviation between the optimal allocation ( $\omega_1$ ) at an extreme value of  $d_T$  and the suboptimal “average” allocation ( $\omega_2$ ) based on the i.i.d. assumption. For example, when  $T = 804$ , an investor with risk aversion equal 2 allocates essentially 100% to stock under the i.i.d. assumption, whether the prior is diffuse or informative. (In the i.i.d. case, the informative prior simply provides the equivalent of 50 years of information about the unconditional mean and variance of  $r_t$ .) If  $N = 0$  and this investor has a diffuse prior, then the optimal allocation at  $d_T = 1.64\%$  is equal to 0.27 using a diffuse prior (table 1) as compared to 0.64 using the informative prior (table 4). The conditional standard deviation of  $r_{T+1}$  is about the same using either prior, about 5.6%, so the certainty-equivalent loss simply drops from 225 basis points (table 3) to 51 basis points (table 6).

When the estimation risk is reduced significantly in moving to the informative prior, so that the allocations actually become more sensitive to  $d_T$ , the certainty-equivalent loss associated with  $\omega_2$  still tends to be lower with the informative prior. There are at least two effects here, both of which often work in the same direction. First,  $|\omega_1 - \omega_2|$  is often lower using the informative prior than using the diffuse prior, even though  $\omega_1$  is more sensitive to  $d_T$  using the informative prior. This result occurs because the lower estimation risk obtained

with the informative prior also tends to raise the stock allocations at all levels of  $d_T$ , as noted earlier. If  $T = 804$  and  $N = 300$ , for example, then the allocation chosen at  $d_T = 1.64\%$  by the investor with risk aversion of 2 is 0.25 using the diffuse prior as compared to 0.42 using the informative prior, which is closer to  $\omega_2$  (essentially 100% with either prior, as noted above). A second effect, which tends to reinforce the first effect in the example just cited, is that the lower estimation risk obtained using the informative prior also tends to reduce the certainty-equivalent loss associated with a given over-allocation to stocks. That is, even if  $\omega_1$  where the same when  $d_T = 1.64\%$  using both priors, the lower conditional volatility of  $r_{T+1}$  using the informative prior—8.56% (table 5) versus 14.89% (table 2)—would most likely result in a lower certainty-equivalent loss.

## 4. The Role of $R^2$

### 4.1. Rewriting the Conditional Distribution

Although we use dividend yield as the predictive variable, identical results would obtain for any predictive variable that produces the same in-sample goodness-of-fit as reported for the regression in (1). That is, the sample statistics used to construct the various conditional distributions for  $r_{T+1}$  can be rewritten in terms of  $\bar{r}$  and  $\hat{\sigma}_r^2$ , the unconditional sample moments of returns defined in (29) and (30), and

$$R^2 = 1 - \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T\hat{\sigma}_r^2}, \quad (39)$$

the regression's unadjusted sample R-squared.

To see this, first express the most recent observation of the predictive variable as being  $\delta$  sample standard deviations away from its sample mean. In the case of dividend yield,

$$d_T = \bar{d} + \delta\hat{\sigma}_d, \quad (40)$$

where  $\bar{d}$  and  $\hat{\sigma}_d^2$  are defined in (18) and (19). Then it is easily verified that, for example, the predictive pdf in (12) when  $N = 0$  can be rewritten as

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)} \left(\frac{h}{\nu}\right)^{1/2} \left[1 + \frac{h}{\nu}(r_{T+1} - \bar{r} - \delta\sqrt{R^2}\hat{\sigma}_r)^2\right]^{-(\nu+1)/2}, \quad (41)$$

where, rewriting (15) and (17),

$$s^2 = \frac{T}{\nu}(1 - R^2)\hat{\sigma}_r^2, \quad (42)$$

$$h = \frac{1}{s^2} \left[ 1 + \frac{1}{T}(1 + \delta^2) \right], \quad (43)$$

and  $\nu = T - 3$ . A similar rewriting is also straightforward in the data-mining scenario and with the informative no-predictability prior.

As reported in (1), the regression of monthly stock returns on dividend yields produces an  $R^2$  equal to 0.0024. Recall also that, in the previous section, we report optimal stock allocations for values of the current dividend yield between 1.64% and 7.05%, or for  $\delta$  ranging between -2 and 2. Given (41) through (43), the allocations reported in the previous section would apply over a range of  $\pm 2\delta$  for any other predictive variable that produces an in-sample  $R^2$  equal to 0.0024.

#### 4.2. Known Parameters (Infinite Sample)

As the sample size  $T$  approaches infinity, the investor learns precisely the values of the parameters of the conditional distribution of the stock return. In all of the cases considered above, the stock return's conditional distribution in finite samples is of the Student- $t$  form, so the conditional distribution converges to a normal density as  $T$  approaches infinity.

Assume the investor knows that the continuously compounded stock return  $r_{t+1}$  over the discrete period from  $t$  to  $t + 1$  has mean  $\mu_t$ , variance  $\sigma_t^2$ , and is an infinitely divisible normal random variable.<sup>15</sup> If the investor's portfolio of stocks and cash is rebalanced continuously to maintain a constant fraction  $\omega_t$  in stocks, then the continuously compounded return on the portfolio from  $t$  to  $t + 1$  will also be normally distributed. In this case, it is easily verified that the optimal value of  $\omega_t$  for an investor whose utility for wealth at time  $t + 1$  is of the form in (3) is given by<sup>16</sup>

$$\omega_t = \frac{\mu_t - i_t}{A\sigma_t^2} + \frac{1}{2A}. \quad (44)$$

In the previous section, we solve for the optimal buy-and-hold allocation at the beginning of a month, so that no rebalancing occurs during the month. Nevertheless, the continuous-rebalancing solution in equation (44) provides a close approximation to the infinite-sample allocations reported in table 1, where  $\mu_t$  is set equal to the fitted value from (1) and  $\sigma_t$  is set equal to  $\hat{\sigma}_t$  reported with that regression. (Recall that the "infinite-sample" allocations reported earlier simply treat the sample statistics obtained from the actual 804-month sample

<sup>15</sup>See Ingersoll (1987, chapter 12), for a discussion of infinite divisibility.

<sup>16</sup>The derivation is a straightforward application of results contained in Merton (1969), and an expression equivalent to (44) also arises as a solution to a special case of the dynamic consumption-investment problem analyzed in that study.

as if they were generated by an infinite sample.) In fact, when rounded to the nearest 0.01, the optimal allocations computed using (44) agree exactly with all of the allocations reported in table 1 for the infinite-sample case (where the upper bound of 0.99 is again binding when  $A = 2$  and  $d_T$  equals the three largest values).

Equation (44) can be used to provide additional insight into the role of  $R^2$  in the allocation decision. If the interest rate ( $i_t$ ) and the conditional variance ( $\sigma_t$ ) are held constant, then we can simply consider the dispersion in  $\omega_t$  across realizations of  $\mu_t$ , i.e. across realizations of the predictive variable. Following the definition of  $R^2$  in (39), we write

$$R^2 = \text{var}\{\mu_t\}/\sigma_r^2,$$

and

$$\sigma_t^2 = (1 - R^2)\sigma_r^2,$$

where  $\sigma_r^2$  denotes the unconditional variance of  $r_t$ . Then (44) can be rewritten as

$$\omega_t = \frac{\mu_t - i_t}{A(1 - R^2)\sigma_r^2} + \frac{1}{2A}, \quad (45)$$

which implies

$$\text{std. dev.}\{\omega_t\} = \frac{1}{A\sigma_r} \frac{\sqrt{R^2}}{(1 - R^2)}. \quad (46)$$

If we evaluate (46) at  $A = 10$ ,  $R^2 = 0.0024$ , and  $\sigma_r = 0.056$ , where the last two quantities reflect sample estimates, then  $\text{std. dev.}\{\omega_t\} = 0.0877$ , which is approximately the difference between the successive values of  $\omega$  reported for the infinite-sample case in table 1. Similarly, when  $A = 2$  then (46) gives  $\text{std. dev.}\{\omega_t\} = 0.439$ , which is approximately the difference between the infinite-sample allocations in table 1 for the two lowest dividend yields (where the upper bound of 0.99 on the stock allocation is not binding).

Recall from the previous section that, when  $N$  is low, estimation risk becomes relatively unimportant in samples as small as sixty months. That is, in such cases, the variance of the conditional distribution is close to that obtained when the statistics of the actual sample are assumed to have been generated instead by an infinite sample. Equation (46) provides a simple explanation of why, in the cases that are approximated reasonably well by the infinite-sample result, an  $R^2$  equal to 0.0024 produces substantial differences in allocations across alternative values of the predictive variable.<sup>17</sup>

---

<sup>17</sup>Solnik (1993) explores the economic significance of asset-allocation strategies based on time-varying conditional expected returns that are taken as fitted values from a forecasting regression, so that estimation risk is not included.

## 5. Economic versus Statistical Significance: A Simple Example

The analysis in the previous sections illustrates the potentially sharp contrast between economic significance and statistical significance, when the latter is judged by standard frequentist methods. Since the forms of the prior distributions and the likelihood functions used in that analysis are, to some degree, specific to the regression setting employed, we construct here a much simpler example to demonstrate that the potentially sharp differences between the economic and statistical significance of stock-return predictability are not confined to our regression framework.

Assume that, in any given period, the rate of return on common stock is either 40% or -40%. In addition to past stock returns, a sample contains past realizations of a random state variable, assumed to be identically and independently distributed over time and drawn from a Binomial distribution. We label the possible realizations of this state variable "state 1" and "state 2." In our example, the sample includes  $M = 8$  periods for each state. The sample distribution of stock returns observed following each state can be reported in a  $2 \times 2$  contingency table:

	state 1	state 2
r= 40%	6	4
r=-40%	2	4

Let  $\theta_i$  and  $(1 - \theta_i)$  be the probabilities that, conditional on observing state  $i$  at the beginning of a period, the subsequently observed stock return during that period will be 40% and -40%.

First consider a frequentist test of whether returns are predictable. The null hypothesis of independence between returns and the state variable,  $\theta_1 = \theta_2$ , can be tested against the alternative hypothesis,  $\theta_1 \neq \theta_2$ , using Fisher's exact test for the  $2 \times 2$  contingency table.<sup>18</sup> The test is based on the conditional distribution of  $M_1$  (the number of periods with a 40% return following state 1) given the two-way table's row and column sums. In our example, the probability of getting  $M_1 \geq 6$  equals 0.304. Hence, the null hypothesis of independence between returns and the state variable (or the absence of predictability) cannot be rejected at conventional significance levels.

Now consider an investor with logarithmic utility who has to allocate his wealth between the stock and cash, which we assume earns a zero riskless rate of return. The investor

---

<sup>18</sup>See Kendall and Stuart (1979).

assumes that the probabilities  $\theta_1$  and  $\theta_2$  are constant over time and understands that  $\theta_1$  can differ from  $\theta_2$ , although he has the same prior distribution about  $\theta_i$  for both states:

$$p(\theta_i) = \frac{[\theta_i(1 - \theta_i)]^{a-1}}{B(a, a)}, \quad (47)$$

where  $a \geq 1$  and  $B(\cdot)$  is the "beta" function. We consider two values of  $a$  for the prior distribution in (47). When  $a = 1$ , the prior distribution for each of the  $\theta_i$ 's is the Bayes-Laplace uniform prior on  $(0, 1)$ . When  $a = 6$ , the prior distribution has greater mass around  $\theta_i = 0.5$ . These two prior distributions are displayed in Figure 5.

If state  $i$  is observed at the beginning of each of  $M$  periods, the probability that the 40% stock return will be realized in  $M_i$  of those periods is

$$p(M_i | \theta_i, M) = \binom{M}{M_i} \theta_i^{M_i} (1 - \theta_i)^{M - M_i}. \quad (48)$$

Combining the prior distribution in (47) with the likelihood function in (48) yields the posterior distribution of  $\theta_i$ , which is also a beta distribution:<sup>19</sup>

$$p(\theta_i | M_i, M) = \frac{\theta_i^{M_i + a - 1} (1 - \theta_i)^{M - M_i + a - 1}}{B((M_i + a), (M - M_i + a))}. \quad (49)$$

Let  $\hat{\theta}_i$  denote the predictive probability that, conditional on observing the above sample evidence and then observing state  $i$  at time  $T$ , the stock return at time  $(T + 1)$  will be 40%. Then,

$$\hat{\theta}_i = E(\theta_i | M_i, M) = \frac{(M_i + a)}{(M + 2a)}. \quad (50)$$

Let  $\omega$  ( $0 \leq \omega \leq 1$ ) denote the fraction of the investor's portfolio allocated to stocks at the beginning of period  $T + 1$ . Conditional on observing state  $i$  at time  $T$ , the investor's optimization problem is

$$\max_{0 \leq \omega \leq 1} [\hat{\theta}_i \ln(1 + 0.4\omega) + (1 - \hat{\theta}_i) \ln(1 - 0.4\omega)], \quad (51)$$

the solution to which is

$$\omega = \begin{cases} 0 & \text{if } (2\hat{\theta}_i - 1) \leq 0 \\ \left(\frac{2\hat{\theta}_i - 1}{0.4}\right) & \text{if } 0 < (2\hat{\theta}_i - 1) < 0.4 \\ 1 & \text{if } (2\hat{\theta}_i - 1) \geq 0.4. \end{cases} \quad (52)$$

The number of sample periods where the 40% stock return follows state 2 is equal to four, which is exactly half of the total sample size for that state. Hence, the posterior distribution

<sup>19</sup>See Zellner, 1971, p. 39

of  $\theta_2$  is also centered around 0.5 and  $\hat{\theta}_2 = 0.5$  for both values of  $a$  in the prior. Since the distribution of the stock return in this case is symmetric around zero, any risk-averse investor will refrain from investing any money in stock. It can be verified easily from (52) that  $\omega = 0$  at  $\hat{\theta}_2 = 0.5$ . If state 1 is observed, however, then the predictive probability of the 40% stock return will be greater than 0.5. For the prior with greater mass around  $\theta_1 = 0.5$  ( $a = 6$ ), equations (50) and (52) imply  $\hat{\theta}_1 = 0.6$  and  $\omega = 0.5$ . For the uniform prior ( $a = 1$ ),  $\hat{\theta}_1 = 0.7$  and  $\omega = 1$ , meaning that the portfolio will consist solely of the stock. This example demonstrates a potential economic importance of stock return predictability similar to that found in the previous sections. Although the investor's prior beliefs are the same for the two states, and although the sample contains only eight observations for each state, the investor's optimal portfolio differs significantly across the two states. These results contrast sharply with the classical frequentist test of predictability.

## 6. Conclusions

The sample evidence on stock-return predictability can be sufficient to cause asset allocations to depend importantly on the current value of a predictive variable, even though a null hypothesis of no predictability might not be rejected at conventional levels of statistical significance. An example is provided in this study using the dividend yield on the value-weighted NYSE portfolio to predict that portfolio's monthly rate of return. The predictive variable's importance in the allocation decision also survives, to a considerable degree, an allowance for data mining or a specification of prior beliefs weighted against predictability.

The research conducted here can be extended along a number of dimensions. We have confined our analysis to one predictive variable, dividend yield, in order to present the basic ideas of the paper as simply as possible. The analysis could be extended to include other variables that have also demonstrated some ability to predict stock returns, such as long-term and short-term interest rates.<sup>20</sup> Similarly, we confine our attention to a single stock portfolio, but additional risky assets could be introduced into the allocation decision as well.<sup>21</sup>

The regression framework we employ assumes that the regression disturbances are homoskedastic. A number of studies have concluded, however, that stock returns exhibit

<sup>20</sup>See, for example, Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), and Fama and French (1989).

<sup>21</sup>The effects of estimation risk on asset allocation have been analyzed empirically for multiple risky assets in an i.i.d. setting [e.g., Bawa, Brown, and Klein (1979)], but we are unaware of empirical studies that extend the problem to consider predictable returns.



conditional heteroskedasticity [e.g., French, Schwert, and Stambaugh (1987)]. An interesting extension of our framework would be to analyze the asset-allocation problem when conditional heteroskedasticity is included in the investor's prior distribution and assumed likelihood function. If, for example, the conditional volatility of the stock return is positively related to the current dividend yield, the sensitivity of the optimal stock allocation to that yield could be reduced.

A number of possible extensions involve the length of the investment horizon. We essentially assume that, given a current amount to be invested, the investor maximizes an iso-elastic derived utility of wealth over the next month. Although this maximization can be consistent with a longer-horizon objective, the role of the investment horizon could be analyzed directly. We know, for example, that in regressions of stock returns on dividend yields, the R-squared tends to rise with the return horizon. This has been demonstrated empirically in long-horizon regressions [e.g., Fama and French (1988)], it arises as an implication of the joint time-series properties of the monthly return and dividend-yield series when estimated in a VAR [e.g., Kandel and Stambaugh (1987)], and it arises as a theoretical implication in equilibrium models with time-varying moments of consumption growth [e.g., Kandel and Stambaugh (1991)]. It would be interesting to explore, possibly by lengthening the investment horizon in the asset-allocation problem, whether the higher R-squared values associated with longer horizons correspond to a greater importance of the predictive variable in the asset-allocation problem.

Related to the issue of the investment horizon is the role of dynamic rebalancing. One might, for example, allow the portfolio to be rebalanced each month but assume that the utility function in (3) applies instead to wealth realized at the end of twenty years. This type of problem has been investigated empirically by Brennan, Schwartz, and Lagnado (1993), using a monthly approximation to an analytic solution for continuous rebalancing, but they do not include estimation risk in their analysis.<sup>22</sup> Addressing the latter would require that each month the investor not only incorporate a new observation of the predictive variables into the conditional mean but also update his beliefs about all parameters of the predictive distribution. Of course, transaction costs would present an additional challenge to such a modeling effort.

---

<sup>22</sup>Estimation risk in the case of continuous rebalancing has been addressed in a number of theoretical studies. See, for example, Dothan and Feldman (1986), Gennotte (1986), Detemple (1986), Feldman (1989, 1992), and Karatzas and Xue (1991).

## Appendix

Zellner (1971, pp. 233-236) analyzes the predictive distribution for the traditional multivariate regression model. Including lagged dependent variables as independent variables in this model does not change the analysis, provided the pre-sample observation  $x_0$  is treated as deterministic.<sup>23</sup> Hence, this analysis applies to the multivariate regression model (7)-(11), and the predictive pdf is in the multivariate Student  $t$  form,

$$p(y_{T+1}|\Phi_T) \propto [1 + g(y'_{T+1} - x'_T \hat{B})S^{-1}(y'_{T+1} - x'_T \hat{B})]^{-(T-N-1)/2}, \quad (\text{A.1})$$

where

$$\hat{B} = (X'X)^{-1}X'Y, \quad (\text{A.2})$$

$$S = (Y - X\hat{B})'(Y - X\hat{B}), \quad (\text{A.3})$$

and

$$g = 1 - x'_T(X'X + x_T x'_T)^{-1}x_T. \quad (\text{A.4})$$

The predictive pdf in (A.1) can be rewritten as

$$p(y_{T+1}|\Phi_T) = \frac{\nu^{(\nu/2)}\Gamma[(\nu + K)/2]|V|^{(1/2)}}{(\Gamma(1/2))^K \Gamma(\nu/2)} [\nu + y'_{T+1} - x'_T \hat{B})V(y'_{T+1} - x'_T \hat{B})]^{-(\nu+K)/2}, \quad (\text{A.5})$$

where  $V = g\nu S^{-1}$ . The predictive distribution of  $r_{T+1}$  (which is the first element of  $y_{T+1}$ ) is a univariate Student  $t$  distribution:

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma(1/2)\Gamma(\nu/2)} \left(\frac{g}{S_{11}}\right)^{1/2} \left[1 + \frac{a}{S_{11}}(r_{T+1} - x'_T \hat{\beta})^2\right]^{-(\nu+1)/2}, \quad (\text{A.6})$$

where  $S_{11}$  is the (1, 1) element of  $S$ .<sup>24</sup> Observe that, using the definitions of  $h$  and  $s^2$  in (14) and (15),  $S_{11} = \nu s^2$  and  $a = h s^2$ . Hence,

$$\frac{g}{S_{11}} = \frac{h}{\nu}. \quad (\text{A.7})$$

The predictive distribution (12) is obtained by substituting (A.7) into (A.6).

To obtain the expressions for  $h$  in (17) and (25), first observe that<sup>25</sup>

$$1 - x'_T(X'X + x_T x'_T)^{-1}x_T = \frac{1}{1 + x'_T(X'X)^{-1}x_T}. \quad (\text{A.8})$$

<sup>23</sup>See, for example, Hamilton (1994, p. 358).

<sup>24</sup>See Zellner (1971, page 387).

<sup>25</sup>See, for example, Zellner (1971, page 73).

Substituting (A.8) into (14) yields

$$h = \left(\frac{1}{s^2}\right) \left(\frac{1}{1 + x_T'(X'X)^{-1}x_T}\right). \quad (\text{A.9})$$

Under our specification of “data mining,” the  $N$  additional variables are uncorrelated in the sample with the dividend yield (equation (24)). Hence,

$$\left(\frac{1}{T}\right) X'X = \begin{bmatrix} P & \begin{pmatrix} 1 \\ \bar{d} \end{pmatrix} \bar{z}' \\ \bar{z} \begin{pmatrix} 1 & \bar{d} \end{pmatrix} & (\bar{z}\bar{z}' + \hat{\Sigma}_z) \end{bmatrix}, \quad (\text{A.10})$$

where

$$P = \begin{bmatrix} 1 & \bar{d} \\ \bar{d} & (\bar{d}^2 + \hat{\sigma}_d^2) \end{bmatrix}. \quad (\text{A.11})$$

Note that

$$P^{-1} \begin{pmatrix} 1 \\ \bar{d} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (\text{A.12})$$

$$P^{-1} \begin{pmatrix} 1 \\ \bar{d} \end{pmatrix} \begin{pmatrix} 1 & \bar{d} \end{pmatrix} P^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad (\text{A.13})$$

and

$$\begin{pmatrix} 1 & d_T \end{pmatrix} P^{-1} \begin{pmatrix} 1 \\ d_T \end{pmatrix} = \left(1 + \frac{(d_T - \bar{d})^2}{\hat{\sigma}_d^2}\right). \quad (\text{A.14})$$

By inverting (A.10) and employing (A.11) through (A.13) we get

$$(X'X)^{-1} = \frac{1}{T} \begin{bmatrix} P^{-1} + (\bar{z}'\hat{\Sigma}_z^{-1}\bar{z}) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \bar{z}'\hat{\Sigma}_z^{-1} \\ -\hat{\Sigma}_z^{-1}\bar{z} \begin{pmatrix} 1 & 0 \end{pmatrix} & \hat{\Sigma}_z^{-1} \end{bmatrix}, \quad (\text{A.15})$$

which, together with (A.14), gives

$$x_T'(X'X)^{-1}x_T = \frac{1}{T} \left(1 + \frac{(d_T - \bar{d})^2}{\hat{\sigma}_d^2} + (z_T - \bar{z})'\hat{\Sigma}_z^{-1}(z_T - \bar{z})\right). \quad (\text{A.16})$$

Substituting (A.16) into (A.9) yields (25). When  $N = 0$ , the expression for  $h$  in (25) simplifies to (17).

Table 1  
Optimal Stock Allocation with a Diffuse Prior

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
Relative risk aversion = 2						
6	0	0.24	0.31	0.39	0.44	0.43
60		0.27	0.66	0.99	0.99	0.99
120		0.27	0.68	0.99	0.99	0.99
360		0.27	0.70	0.99	0.99	0.99
804		0.27	0.70	0.99	0.99	0.99
60	10	0.26	0.42	0.58	0.74	0.88
	25	0.24	0.26	0.27	0.29	0.30
120	25	0.26	0.41	0.57	0.72	0.86
	50	0.25	0.27	0.30	0.32	0.34
360	50	0.26	0.51	0.75	0.99	0.99
	150	0.25	0.28	0.32	0.35	0.38
804	100	0.26	0.53	0.81	0.99	0.99
	200	0.26	0.41	0.56	0.71	0.86
	300	0.25	0.31	0.38	0.44	0.50
	350	0.25	0.28	0.30	0.33	0.36
$\infty$	-	0.27	0.71	0.99	0.99	0.99
Relative risk aversion = 10						
6	0	0.05	0.06	0.08	0.09	0.09
60		0.05	0.13	0.21	0.29	0.35
120		0.05	0.14	0.22	0.30	0.38
360		0.05	0.14	0.23	0.31	0.39
804		0.05	0.14	0.23	0.31	0.40
60	10	0.05	0.08	0.12	0.15	0.18
	25	0.05	0.05	0.05	0.06	0.06
120	25	0.05	0.08	0.11	0.14	0.17
	50	0.05	0.05	0.06	0.06	0.07
360	50	0.05	0.10	0.15	0.20	0.25
	150	0.05	0.06	0.06	0.07	0.08
804	100	0.05	0.11	0.16	0.22	0.27
	200	0.05	0.08	0.11	0.14	0.17
	300	0.05	0.06	0.07	0.09	0.10
	350	0.05	0.05	0.06	0.07	0.07
$\infty$	-	0.05	0.14	0.23	0.32	0.40

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.

Table 2  
 Conditional Standard Deviation of the Stock Return  
 with a Diffuse Prior  
 (Percent per Month)

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
6	0	18.51	15.78	14.76	15.78	18.51
60		6.07	5.92	5.88	5.92	6.07
120		5.82	5.75	5.72	5.75	5.82
360		5.66	5.63	5.63	5.63	5.66
804		5.62	5.60	5.60	5.60	5.62
60	10	9.33	9.19	9.14	9.19	9.33
	25	32.29	32.00	31.90	32.00	32.29
120	25	9.49	9.42	9.39	9.42	9.49
	50	24.28	24.15	24.11	24.15	24.28
360	50	7.48	7.45	7.44	7.45	7.48
	150	20.35	20.30	20.29	20.30	20.35
804	100	7.09	7.08	7.08	7.08	7.09
	200	9.50	9.48	9.48	9.48	9.50
	300	14.89	14.88	14.87	14.88	14.89
	350	22.37	22.35	22.34	22.35	22.37
$\infty$	-	5.58	5.58	5.58	5.58	5.58

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.

Table 3

Differences in Certainly Equivalents with a Diffuse Prior: Optimal Allocation  
versus Allocation Based Only on Historical Returns  
(Percent per Year)

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
Relative risk aversion = 2						
6	0	10.77	6.08	3.67	3.27	4.53
60		2.64	0.53	0.00	0.01	0.01
120		2.42	0.43	0.00	0.01	0.02
360		2.29	0.37	0.00	0.01	0.02
804		2.25	0.35	0.00	0.01	0.02
60	10	6.34	3.69	1.83	0.69	0.13
	25	61.59	57.43	54.35	52.30	51.21
120	25	6.55	3.98	2.10	0.86	0.19
	50	38.47	35.28	32.55	30.26	28.37
360	50	4.05	1.75	0.41	0.00	0.01
	150	28.13	25.30	22.73	20.41	18.32
804	100	3.64	1.41	0.23	0.00	0.01
	200	6.56	4.08	2.20	0.92	0.19
	300	15.76	13.08	10.70	8.60	6.77
	350	33.32	30.53	27.95	25.55	23.33
$\infty$	-	2.22	0.34	0.00	0.02	0.03
Relative risk aversion = 10						
6	0	1.98	1.13	0.68	0.59	0.81
60		0.66	0.18	0.00	0.09	0.37
120		0.63	0.17	0.00	0.12	0.48
360		0.61	0.15	0.00	0.14	0.56
804		0.60	0.15	0.00	0.15	0.59
60	10	1.59	1.01	0.58	0.29	0.11
	25	15.78	15.00	14.42	14.01	13.80
120	25	1.70	1.13	0.69	0.37	0.16
	50	10.18	9.53	8.97	8.49	8.10
360	50	1.08	0.55	0.21	0.03	0.01
	150	7.57	6.98	6.43	5.93	5.46
804	100	0.97	0.46	0.14	0.01	0.05
	200	1.75	1.20	0.75	0.41	0.18
	300	4.23	3.65	3.12	2.64	2.21
	350	9.06	8.49	7.94	7.43	6.95
$\infty$	-	0.60	0.15	0.00	0.15	0.61

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.

Table 4  
Optimal Stock Allocation with A No-Predictability  
Informative Prior

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
Relative risk aversion = 2						
6		0.99	0.99	0.99	0.99	0.99
60		0.99	0.99	0.99	0.99	0.99
120	0	0.99	0.99	0.99	0.99	0.99
360		0.81	0.98	0.99	0.99	0.99
804		0.64	0.89	0.99	0.99	0.99
60	10	0.99	0.99	0.99	0.99	0.99
	25	0.94	0.98	0.99	0.99	0.99
120	25	0.89	0.96	0.99	0.99	0.99
	50	0.82	0.88	0.93	0.99	0.99
360	50	0.71	0.85	0.98	0.99	0.99
	150	0.56	0.65	0.74	0.83	0.92
804	100	0.55	0.75	0.94	0.99	0.99
	200	0.48	0.63	0.78	0.92	0.99
	300	0.42	0.52	0.63	0.74	0.84
	350	0.39	0.48	0.56	0.65	0.74
$\infty$	-	0.27	0.71	0.99	0.99	0.99
Relative risk aversion = 10						
6		0.22	0.23	0.23	0.23	0.23
60		0.21	0.22	0.23	0.23	0.24
120	0	0.20	0.21	0.23	0.24	0.25
360		0.16	0.19	0.23	0.26	0.29
804		0.13	0.18	0.23	0.28	0.33
60	10	0.20	0.21	0.22	0.22	0.23
	25	0.19	0.20	0.20	0.21	0.22
120	25	0.18	0.19	0.20	0.22	0.23
	50	0.16	0.18	0.19	0.20	0.21
360	50	0.14	0.17	0.20	0.22	0.25
	150	0.11	0.13	0.15	0.17	0.18
804	100	0.11	0.15	0.19	0.23	0.26
	200	0.10	0.13	0.15	0.18	0.21
	300	0.08	0.10	0.13	0.15	0.17
	350	0.08	0.09	0.11	0.13	0.15
$\infty$	-	0.05	0.14	0.23	0.32	0.40

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.

Table 5  
**Conditional Standard Deviation of the Stock Return  
with a No-Predictability Informative Prior  
(Percent per Month)**

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
6	0	5.63	5.62	5.61	5.62	5.63
60		5.63	5.62	5.61	5.62	5.63
120		5.63	5.61	5.61	5.61	5.63
360		5.62	5.61	5.60	5.61	5.62
804		5.60	5.60	5.60	5.60	5.60
$\infty$		5.58	5.58	5.58	5.58	5.58
60	10	5.82	5.81	5.80	5.81	5.82
	25	6.06	6.05	6.04	6.05	6.06
120	25	6.02	6.01	6.00	6.01	6.02
	50	6.41	6.39	6.39	6.39	6.41
360	50	6.18	6.17	6.17	6.17	6.18
	150	7.56	7.55	7.55	7.55	7.56
804	100	6.37	6.36	6.36	6.36	6.37
	200	7.30	7.29	7.29	7.29	7.30
	300	8.56	8.56	8.55	8.56	8.56
	350	9.43	9.42	9.42	9.42	9.43
$\infty$	-	5.58	5.58	5.58	5.58	5.58

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.

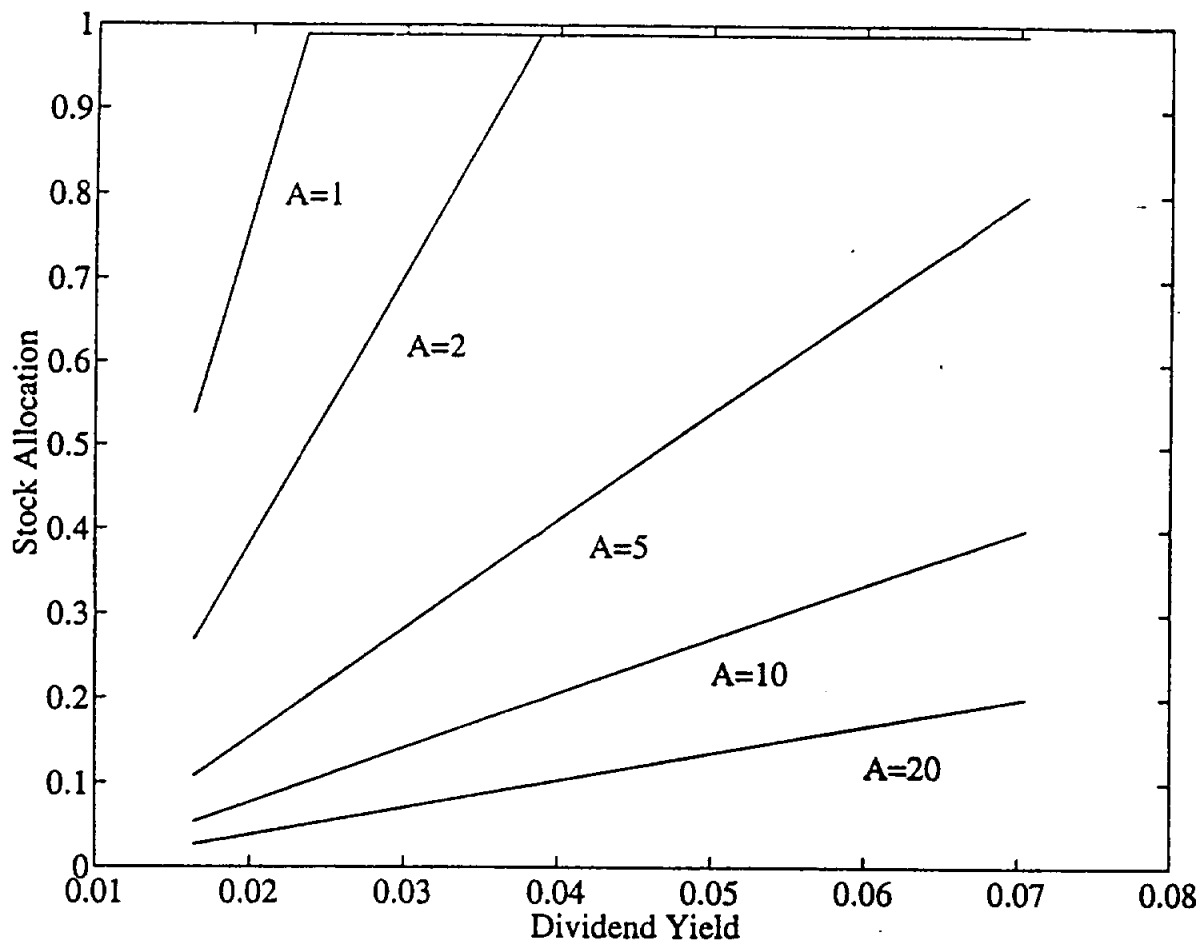


Table 6

Differences in Certainly Equivalents with a No-Predictability Informative Prior:  
Optimal Allocation versus Allocation Based Only on Historical Returns  
(Percent per Year)

$T$	$N$	$d = 1.64\%$	$d = 2.99\%$	$d = 4.35\%$	$d = 5.70\%$	$d = 7.05\%$
Relative risk aversion = 2						
6		0.00	0.00	0.00	0.00	0.00
60		0.00	0.00	0.00	0.00	0.00
120	0	0.00	0.00	0.00	0.00	0.00
360		0.13	0.00	0.00	0.00	0.01
804		0.51	0.04	0.00	0.01	0.01
60	10	0.00	0.00	0.00	0.00	0.00
	25	0.01	0.00	0.00	0.00	0.00
120	25	0.04	0.00	0.00	0.00	0.00
	50	0.15	0.07	0.02	-0.00	0.00
360	50	0.38	0.10	0.00	0.00	0.01
	150	1.38	0.86	0.46	0.19	0.04
804	100	1.03	0.31	0.01	0.00	0.01
	200	1.82	0.91	0.32	0.03	0.00
	300	3.16	2.07	1.22	0.60	0.20
	350	4.21	3.05	2.08	1.30	0.71
$\infty$	-	2.22	0.34	0.00	0.02	0.03
Relative risk aversion = 10						
6		0.00	0.00	0.00	0.00	0.00
60		0.01	0.00	0.00	0.00	0.00
120	0	0.02	0.00	0.00	0.00	0.01
360		0.09	0.02	0.00	0.02	0.08
804		0.20	0.05	0.00	0.05	0.19
60	10	0.02	0.01	0.00	0.00	0.00
	25	0.04	0.02	0.01	0.01	0.00
120	25	0.05	0.03	0.01	0.00	0.00
	50	0.11	0.07	0.04	0.02	0.01
360	50	0.17	0.08	0.02	0.00	0.01
	150	0.48	0.34	0.23	0.14	0.07
804	100	0.35	0.16	0.04	0.00	0.03
	200	0.58	0.35	0.18	0.06	0.01
	300	0.95	0.69	0.47	0.30	0.16
	350	1.25	0.97	0.73	0.52	0.36
$\infty$	-	0.60	0.15	0.00	0.15	0.61

Note:  $d$  denotes the current dividend yield,  $T$  is the number of months in the sample, and  $N$  is the number of independent variables in addition to  $d$ . The sample statistics are based on the period 1/1927-12/1993.



**Figure 1. Asset allocations with a diffuse prior and no additional independent variables ( $N = 0$ ). The plot displays, for five values of relative risk aversion ( $A$ ), the fraction of the optimal portfolio allocated to stocks at various levels of the current dividend yield.**

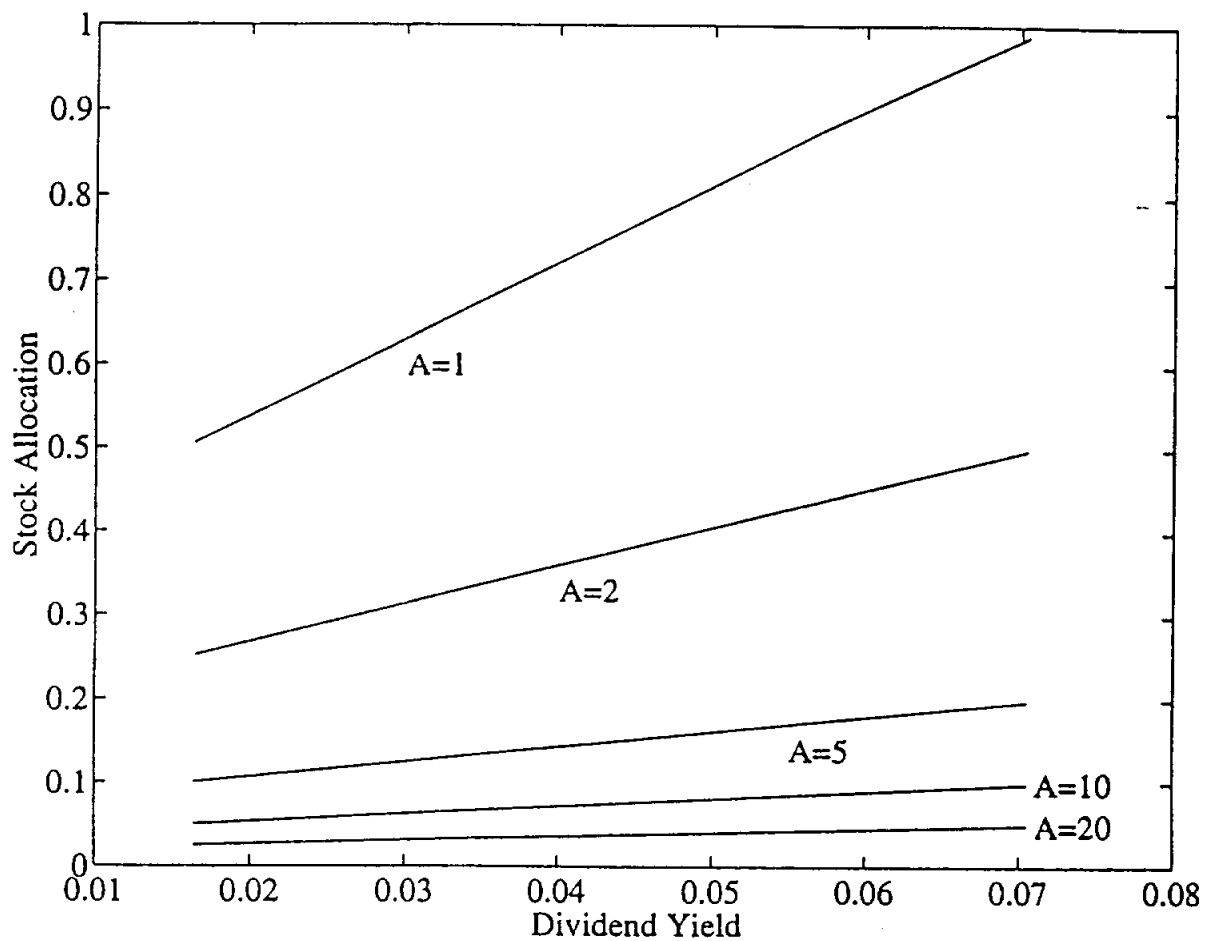


Figure 2. Asset allocations with a diffuse prior and 300 additional independent variables ( $N = 300$ ). The plot displays, for five values of relative risk aversion ( $A$ ), the fraction of the optimal portfolio allocated to stocks at various levels of the current dividend yield.

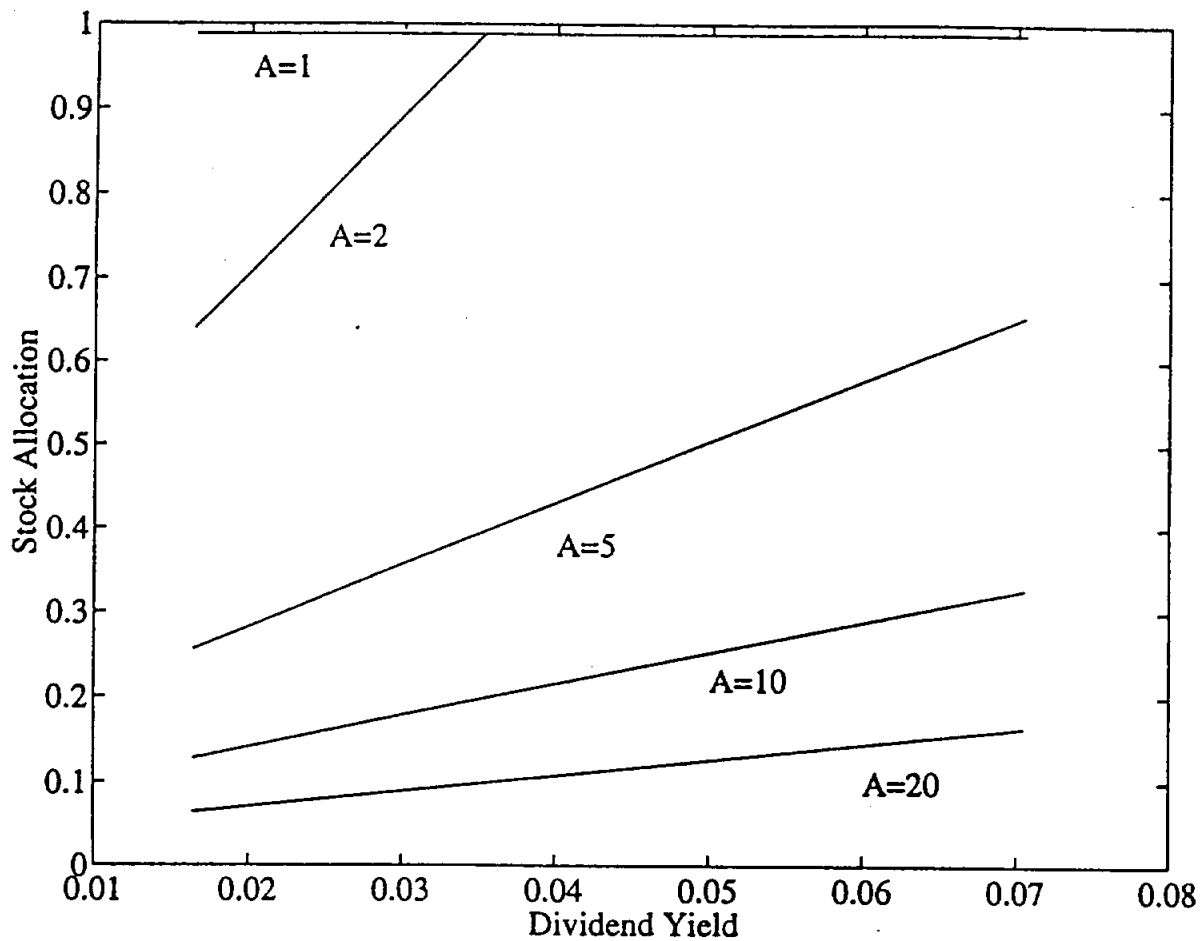


Figure 3. Asset allocations with an informative no-predictability prior and no additional independent variables ( $N = 0$ ). The plot displays, for five values of relative risk aversion ( $A$ ), the fraction of the optimal portfolio allocated to stocks at various levels of the current dividend yield.

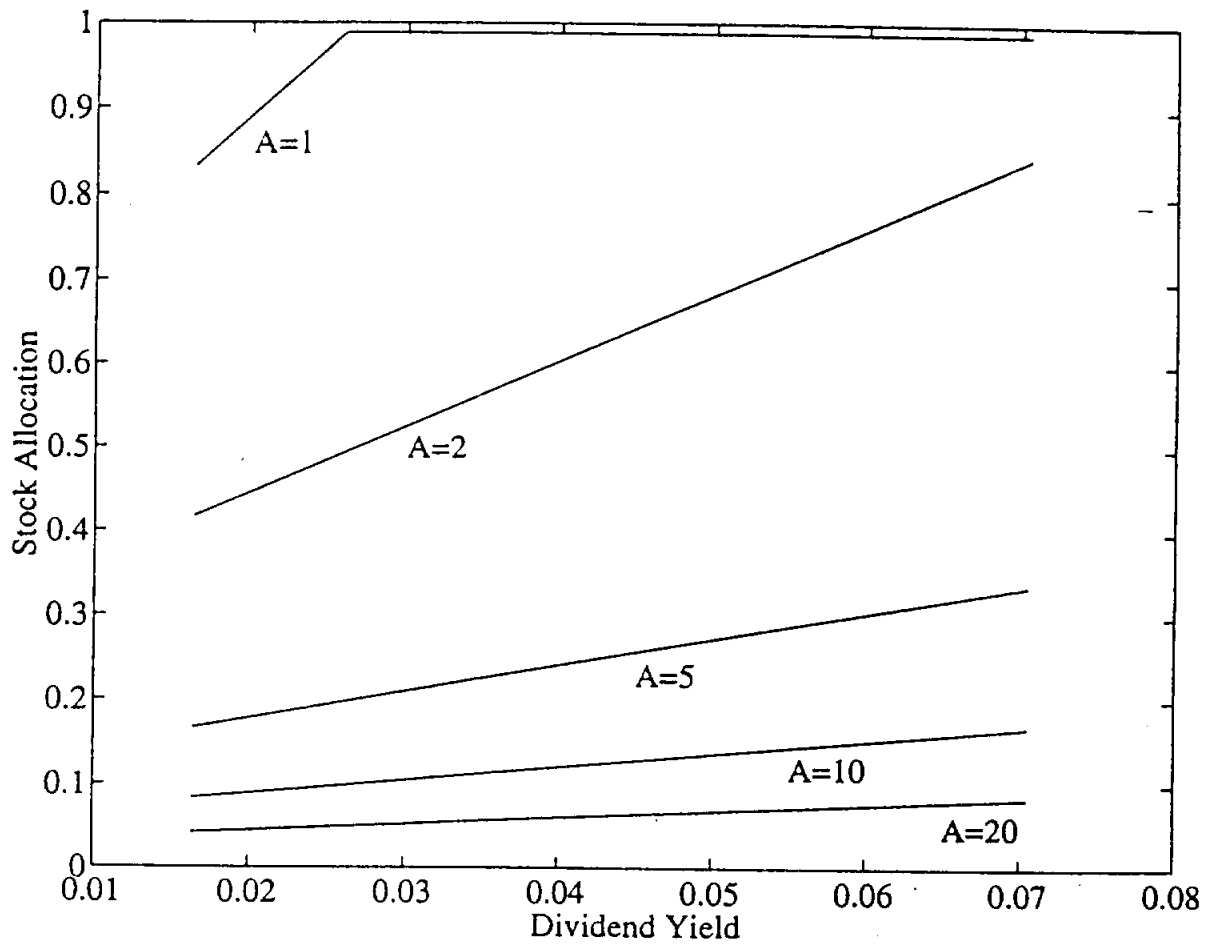


Figure 4. Asset allocations with an informative no-predictability prior and 300 additional independent variables ( $N = 300$ ). The plot displays, for five values of relative risk aversion ( $A$ ), the fraction of the optimal portfolio allocated to stocks at various levels of the current dividend yield.

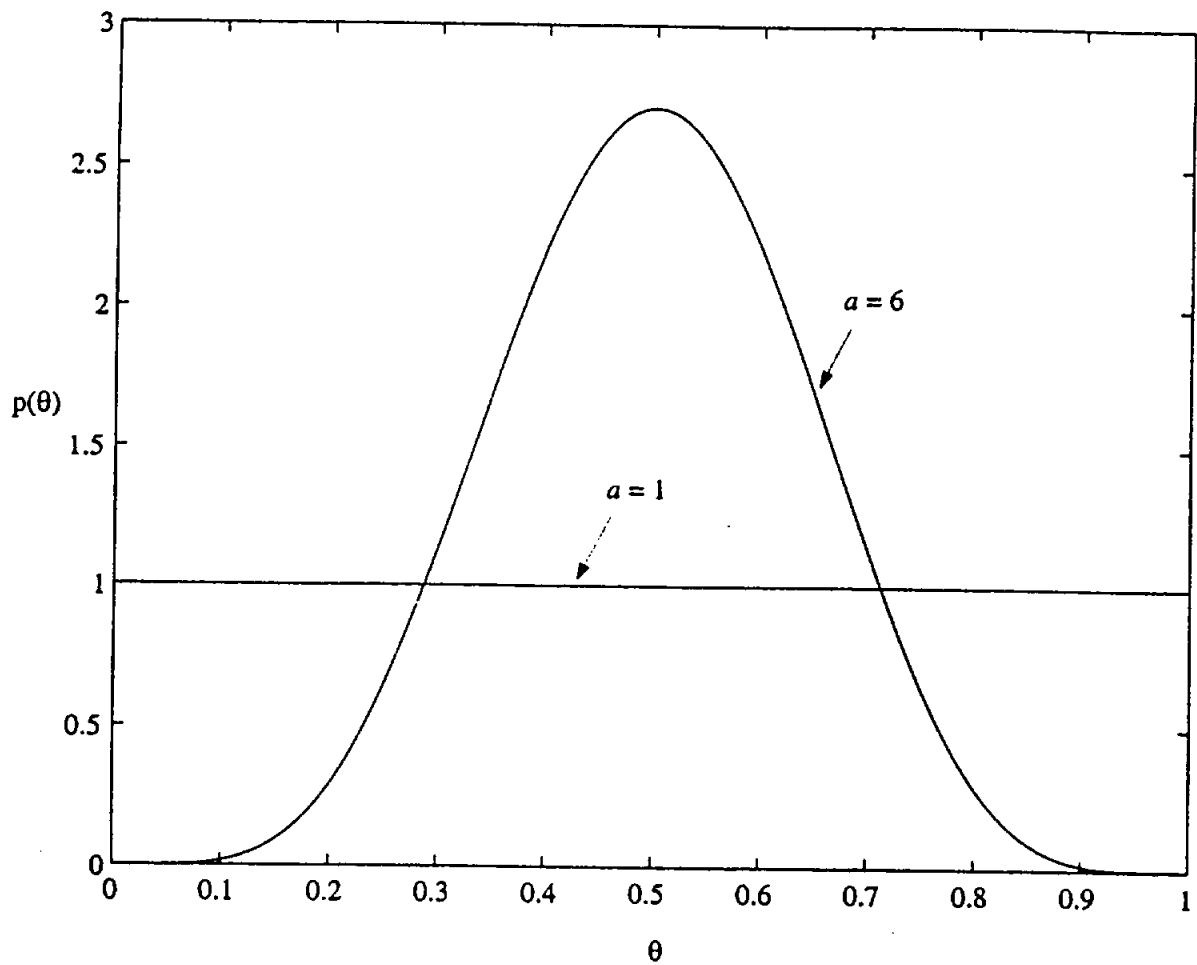


Figure 5. Prior distributions for  $\theta$ . The plot displays, for  $a = 1$  and  $a = 6$ , the prior distribution

$$p(\theta) = \frac{[\theta(1 - \theta)]^{a-1}}{B(a, a)},$$

where  $B(\cdot)$  is the "beta" function.

## References

- Anderson, T.W., 1984, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.
- Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.
- Black, Fischer, 1990, "Mean Reversion and Consumption Smoothing," *The Review of Financial Studies* 3, 107-114.
- Brennan, Michael J., Eduardo S. Schwartz, and Ron Lagnado, 1993, "Strategic Asset Allocation," Working paper no. 11-93, Anderson Graduate School of Management, UCLA, Los Angeles, CA.
- Brent, Richard P., 1973, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, N.J.
- Brown, Stephen J., 1979, "Optimal Portfolio Choice Under Uncertainty: A Bayesian Approach," chapter 8 in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.
- Campbell, John Y., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373-399.
- Detemple, Jerome, 1986, "Asset Pricing in a Production Economy with Incomplete Information," *Journal of Finance* 41, 383-391.
- Dothan, Michael U. and David Feldman, 1986, "Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy," *Journal of Finance* 41, 369-382.
- 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.
- Fama, Eugene F. and G. William Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics* 5, 115-146.
- Feldman, David, 1989, "The Term Structure of Interest Rates in a Partially Observable Economy," *Journal of Finance* 44, 789-812.
- Feldman, David, 1992, "Logarithmic Preferences, Myopic Decisions, and Incomplete Information," *Journal of Financial and Quantitative Analysis* 27, 619-629.
- Forsythe, G.E., M.A. Malcolm, and C.B. Moler, 1977, *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, N.J.
- Foster, F. Douglas, and Tom Smith, 1994, "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal  $R^2$ ," working paper, University of Iowa, Iowa City, IA and Duke University, Durham, NC.
- French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, "Expected Stock

- Returns and Volatility," *Journal of Financial Economics* 19, 3-29.
- Frost, Peter A. and James E. Savarino, 1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis* 21, 293-305.
- Gennotte, Gerard, 1986, "Optimal Portfolio Choice Under Incomplete Information," *Journal of Finance*, 41, 733-746.
- Goetzmann, William N. and Philippe Jorion, 1994, "A Longer Look at Dividend Yields," working paper, Yale University, New Haven, CT and University of California at Irvine, Irvine, CA.
- Grauer, Robert R. and Nils H. Hakansson, 1992, "Stein and CAPM Estimators of the Means in Asset Allocation: A Case of Mixed Success," working paper, Simon Fraser University, Burnaby, BC, and University of California, Berkeley, CA.
- Hamilton, James D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.
- Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making*, Rowman and Littlefield, Savage, MD.
- Jobson, J.D. and Bob Korkie, 1980, "Estimation for Markowitz Efficient Portfolios," *Journal of the American Statistical Association* 75, 544-554.
- Jobson, J.D., Bob Korkie, and V. Ratti, 1979, "Improved Estimation for Markowitz Efficient Portfolios Using James-Stein Type Estimators," *Proceedings of the American Statistical Association*, Business and Economics Statistics Section, 41, 279-284.
- Jorion, Philippe, 1985, "International Portfolio Diversification with Estimation Risk," *Journal of Business* 58, 259-278.
- Jorion, Philippe, 1986, "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis* 21, 279-292.
- Jorion, Philippe, 1991, "Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection," *Journal of Banking and Finance* 15, 717-728.
- Kandel, Shmuel and Robert F. Stambaugh, 1987, "Long-Horizon Returns and Short-Horizon Models," CRSP working paper no. 222, University of Chicago, Chicago, IL.
- Kandel, Shmuel and Robert F. Stambaugh, 1991, "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics* 27, 39-71.
- Karatzas, Ionnis and Xing-Xiong Xue, 1991, "A Note on Utility Maximization Under Partial Observations," *Mathematical Finance*, 1, 57-70.
- Keim, Donald B. and Robert F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics* 17, 357-390.
- Kendall, Sir Maurice and Alan Stuart, 1977, *The Advanced Theory of Statistics, Volume 1* Macmillan, New York.
- Kendall, Sir Maurice and Alan Stuart, 1979, *The Advanced Theory of Statistics, Volume 2* Macmillan, New York.
- Merton, Robert C., 1969, "Lifetime Portfolio Selection Under Uncertainty: The Continuous-



- Time Case," *Review of Economics and Statistics* 51, 247-257.
- Nelson, Charles R. and Myung J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641-661.
- Press, William A., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, 1986, *Numerical Recipes*, Cambridge University Press, Cambridge.
- Rozeff, Michael, 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management*, Fall, 68-75.
- Solnik, Bruno, 1993, "The Performance of International Asset Allocation Strategies Using Conditioning Information," *Journal of Empirical Finance* 1, 33-55.
- Stambaugh, Robert F., 1986, "Bias in Regressions with Lagged Stochastic Regressors," CRSP working paper no 156, University of Chicago, Chicago, IL.
- Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.