THE JOURNAL OF FINANCE • VOL. LI, NO. 2 • JUNE 1996

On the Predictability of Stock Returns: An Asset-Allocation Perspective

SHMUEL KANDEL and ROBERT F. STAMBAUGH*

ABSTRACT

Sample evidence about the predictability of monthly stock returns is considered from the perspective of a risk-averse Bayesian investor who must allocate funds between stocks and cash. The investor uses the sample evidence to update prior beliefs about the parameters in a regression of stock returns on a set of predictive variables. The regression relation can seem weak when described by usual statistical measures, but the current values of the predictive variables can exert a substantial influence on the investor's portfolio decision, even when the investor's prior beliefs are weighted against predictability.

INVESTORS IN THE STOCK market are interested in predicting future stock returns, and the academic literature offers numerous empirical investigations of stock-return predictability. Many of these investigations report the results of estimating linear time-series regressions of stock returns on one or more predictive variables, and considerable effort has been devoted to assessing the strength and reliability of this regression evidence from a statistical perspective. Given that the regression coefficients are estimated with error, confronting the investor with what is commonly termed "estimation risk," to what extent might the regression evidence influence a rational, risk-averse investor's portfolio decision?

Consider an investor who, on December 31, 1993, must allocate funds between the value-weighted portfolio of the New York Stock Exchange (NYSE) and one-month Treasury bills. The investor is given the results of estimating the following regression using monthly data from January 1927 through December 1993,

$$r_t = x'_{t-1}b + \epsilon_t, \tag{1}$$

* Recanati Graduate School of Business Administration, Tel-Aviv University and The Wharton School, University of Pennsylvania (Kandel) and The Wharton School, University of Pennsylvania and National Bureau of Economic Research (Stambaugh). The authors are grateful for comments by René Stulz, three anonymous referees, and workshop participants at Baruch College (CUNY), Duke University, Northwestern University, Ohio State University, Rutgers University, the University of North Carolina, the University of Pennsylvania, the University of Rochester, the University of Washington, and Washington University in St. Louis. We also wish to thank participants in the NBER 1995 Summer Institute, especially Bill Schwert, the discussant for the paper.

where r_t is the continuously compounded NYSE return in month t, in excess of the continuously compounded T-bill rate for that month, x_{t-1} is a vector of "predictive" variables that are observed at the end of month $t-1$, b is a vector of coefficients, and ϵ_t is the regression disturbance in month t. Suppose that the unadjusted sample R-squared for the regression in (1) is equal to $R^2 = 0.025$, which is fairly typical of values reported in studies using monthly data beginning in 1927.[1] The investor is provided that value along with the vector of OLS coefficient estimates, $\hat{b}$, and other summary statistics often published in the academic literature.

In addition to the regression results, the investor is given the most recent vector of the predictive variables, x_T, where December 1993 is denoted as month T. To what extent does the investor's asset allocation decision depend on x_T? The average excess monthly return ($\bar{r}$) for the entire 804-month sample period is 49 basis points (bp). Suppose that the fitted regression prediction for the excess return in January 1994, $x_T' \hat{b}$, is equal to -40 bp, which is less than $\bar{r}$ by 89 bp. The sample standard deviation of excess returns ($\hat{\sigma}_r$) for the 804-month period is equal to 560 bp, so 89 bp represents one sample standard deviation of the regression's fitted values ($\sqrt{R^2}\hat{\sigma}_r = 89$ bp). If the investor would allocate 61 percent to stocks if $x_T' \hat{b}$ were equal to the average excess return of 49 bp, how much less does the investor allocate to stocks when $x_T' \hat{b}$ is actually 89 bp lower than that long-run average? How much does the investor value the ability to allocate less than 61 percent to stocks in this case?

Answers to these questions could provide a metric by which to assess the economic significance of the regression evidence on stock-return predictability. This study explores such questions from the perspective of a Bayesian investor who uses the sample evidence to update prior beliefs about the regression parameters. The investor then uses these revised beliefs to compute the optimal asset allocation.[2] Our analytical framework, although simplified in a number of respects, proves tractable in addressing the questions posed above and illustrates the potential insights offered by this type of approach.

We find that the economic significance of the sample evidence is not readily conveyed by standard statistical measures. For example, an investor who uses the regression results can assign an important role to the predictive variables, and yet those same regression results can produce a large p-value for the null hypothesis that the coefficients on the predictive variables are jointly equal to zero. With an unadjusted R-squared of 0.025, as given above, a large p-value can be obtained by recognizing that, even though one might compute a low p-value when the number of regressors is small, say 3 or 4, the p-value must

[1] For example, Campbell (1991) reports $R^2 = 0.024$ in a regression of the continuously compounded real return to the value-weighted NYSE on the lagged return, the dividend-price ratio, and the one-month T-bill rate minus its past twelve-month average.

[2] The revised beliefs can be examined without exploring their implications for investment decisions. For example, Lamoureux and Zhou (1995) use data on the value-weighted NYSE portfolio to compute revised beliefs (Bayesian posterior distributions) about various predictability measures, such as variance ratios and autocorrelation coefficients, but they do not explore the implications of these posterior beliefs for an asset-allocation decision.

be increased if those regressors are selected ex post from a larger number of variables, say 25.[3] If the unadjusted R-squared in the regression on all 25 variables is still only 0.025, the worst case, then the standard regression F statistic, computed based on 25 regressors, implies a p-value of about 75 percent. Even when the investor observes such a result, however, the fitted regression prediction influences the investor's asset-allocation decision. In fact, returning to the questions posed in the example, we find that an investor whose coefficient of relative risk aversion equals 2 and who possesses vague prior beliefs about the parameters in that 25-variable regression will, after updating those beliefs using the regression evidence, allocate no funds to stocks when $x_T'\hat{b}$ is equal to -40 bp, whereas the same investor would indeed allocate about 61 percent to stocks if $x_T'\hat{b}$ were instead equal to the long-run average of 49 bp. The ability to allocate 0 percent instead of 61 percent is worth about 29 bp to the investor, valued in terms of differences in a certainty-equivalent monthly return. If an investor's prior beliefs, instead of being vague, are weighted against predictability to a degree equivalent to having observed over 162 years of prior data in which the sample R-squared is *exactly* zero, then that investor still allocates about 30 percent less to stocks when the fitted value is -40 bp than when the fitted value is the long-run average of 49 bp, and the ability to do so is still worth about 4 bp in certainty-equivalent monthly return to that investor.

The article proceeds as follows. Before turning to the Bayesian regression framework used to obtain the type of results cited above, we first outline the basic principles of the conditional Bayesian decision approach that we employ, and we highlight some differences between this approach and others. Much of this discussion, contained in Section I, is organized around an example of the asset-allocation decision within a simple two-state, two-outcome setting. Section II then gives the details of our analysis of the asset-allocation problem within the regression setting. Although the specification we adopt omits some potentially important features of the data, such as heteroskedasticity, which might be interesting to include in future efforts, we find that using this fairly standard Bayesian regression model allows us to analyze the asset-allocation decision for a wide variety of sample characteristics and regression outcomes. In particular, we are able to compare the economic significance of the regression evidence with standard characterizations of the evidence based on regression statistics, and we find that the contrast is often a sharp one. Section III concludes the paper and suggests directions for future research.

I. Analyzing the Asset-Allocation Decision: General Approach

A. The Investor's Allocation Decision

We consider a risk-averse investor with a one-month investment horizon who must allocate funds between stocks and riskless cash. Let ω denote the fraction

[3] See Foster and Smith (1994).

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

$$W_{T+1} = W_T[\omega \exp\{r_{T+1} + i_{T+1}\} + (1 - \omega) \exp\{i_{T+1}\}], \tag{2}$$

where W_T is the investor's wealth at the end of month T, i_{T+1} is the continuously compounded riskless rate on cash for month $T + 1$, observed at the end of month T, and r_{T+1} is the stock's continuously compounded return in month $T + 1$ in excess of i_{T+1}. The investor chooses ω so as to maximize the expected value of the utility function

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

The parameter A in the iso-elastic utility function in (3) is commonly referred to as the investor's coefficient of relative risk aversion. We entertain three values of A—one, two, and five—which produce a wide range of optimal asset allocations in the results reported later. We wish to stress, however, that our analysis does not address issues of market equilibrium, and the investor in this asset-allocation setting should not necessarily be viewed as a representative investor.

Let Φ_T denote the data set observed by the investor through the end of month T, and let $p(r_{T+1}|\Phi_T)$ denote the density of r_{T+1} conditional on Φ_T. The investor is assumed to solve

$$\max_{0 \leq \omega \leq 1} \int v(W_{T+1}) p(r_{T+1}|\Phi_T) dr_{T+1}. \tag{4}$$

Given the form of the utility function in (3), the optimal stock allocation ω^* does not depend on the value of W_T, which we simply set to 1.0.

In assessing the conditional distribution of r_{T+1}, the investor follows principles of conditional Bayesian analysis.[4] In deriving $p(r_{T+1}|\Phi_T)$, known in this Bayesian framework as the predictive probability density function (pdf), the investor updates beliefs about a vector of parameters $\theta \in \Theta$, where θ is assumed to be random. After observing the data, the investor's beliefs about θ are summarized by the posterior pdf of θ, which can be written as[5]

$$p(\theta|\Phi_T) \propto p(\theta) p(\Phi_T|\theta), \tag{5}$$

where $p(\Phi_T|\theta)$ is the pdf for the observations given the parameters, known also as the likelihood function of θ, and $p(\theta)$ denotes the prior pdf for θ. The

[4] This conditional Bayesian decision approach is discussed further in subsection *D*. See also Berger (1985).

[5] See Zellner (1971, p. 14.)

prior pdf represents the investor's knowledge about the parameter vector θ before observing the sample information. Since it is impossible to specify one prior that would be appropriate for all investors, in this study we consider a number of prior distributions, including noninformative as well as informative priors.[6] To obtain the predictive pdf for r_{T+1}, the posterior in (5) is first multiplied by $p(r_{T+1}|\theta, \Phi_T)$ the likelihood function for the future observation, to obtain

$$p(r_{T+1},\theta|\Phi_T) = p(r_{T+1}|\theta,\Phi_T) \cdot p(\theta|\Phi_T). \tag{6}$$

Integration of this joint density in (6) with respect to θ then gives the desired predictive pdf,

$$p(r_{T+1}|\Phi_T) = \int_\Theta p(r_{T+1},\ \theta|\Phi_T)d\theta = \int_\Theta p(r_{T+1}|\theta,\ \Phi_T) \cdot p(\theta|\Phi_T)d\theta, \tag{7}$$

which does not depend on θ.

The expected-utility maximization in (4) is a version of the general Bayesian one-period control problem.[7] Beginning with Klein and Bawa (1976), a number of studies have computed optimal portfolios in a one-period conditional Bayesian framework where the investor uses a model (likelihood function) in which returns are assumed to be identically and independently distributed (i.i.d.).[8] This study analyzes a portfolio decision where the investor instead uses a model in which returns can possess predictability.

B. Economic Significance of Stock-Return Predictability

Our principal approach to assessing the economic significance of the sample evidence on predictability is to analyze the sensitivity of the optimal allocation to the value of the most recent observation of the predictive variables included in Φ_T. In other words, the optimal allocation ω^*, the solution to (4), is compared to a suboptimal allocation ω^a, the solution to (4) when the sample Φ_T is replaced by a different hypothetical sample Φ_T^a. The most recent observation of the predictive variables in Φ_T^a is different from that in Φ_T, but Φ_T^a is essentially identical to Φ_T in other respects, in the sense that

$$p(\theta|\Phi_T^a) = p(\theta|\Phi_T). \tag{8}$$

[6] For a review of noninformative and informative priors see, for example, Judge et al. (1985).

[7] See Zellner (1971, pp. 320–327).

[8] See also Brown (1979), 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.

Differences between ω^* and ω^a reveal the degree to which the sample evidence about stock-return predictability plays a role in the investor's asset-allocation decision.

Additional insight into the economic significance of the sample evidence can be obtained by comparing the investor's expected utility associated with the optimal allocation ω^* to the expected utility associated with the suboptimal allocation ω^a. The expected utilities for both allocations are computed using the single predictive pdf, $p(R_{T+1}|\Phi_T)$, and these expected utilities are compared in terms of the investor's certainty equivalent return (CER) for each allocation.[9] A previous use of certainty-equivalent comparisons to assess the economic significance of empirical evidence is provided by McCulloch and Rossi (1990), who employ a conditional Bayesian decision framework in an investigation of the Arbitrage Pricing Theory (APT) of Ross (1976). They compare an investor's CER for a portfolio that is optimal under beliefs that a linear factor-pricing model holds exactly to the CER for a portfolio that is optimal under beliefs that allow for departures from an exact linear pricing relation. In each case, they compute an optimal portfolio and CER using the predictive pdf obtained under the given set of beliefs. In other words, they compare expected utilities computed using two different probability distributions, and their approach differs from ours in that key respect. We compute certainty equivalents for the optimal and suboptimal portfolio allocations using one common probability distribution, since it is difficult to interpret differences in expected utilities (or certainty equivalents) computed under different distributions.

C. A Simple Example

In this subsection, we use a simple example to illustrate the manner in which the results of the asset-allocation decision can reveal the economic significance of sample evidence about return predictability. This example is also used in the next subsection in a discussion of the differences between the conditional Bayesian decision approach and other approaches often used to characterize the sample evidence.

Consider an investor with logarithmic utility ($A = 1$). Assume that the riskless rate is zero and that the simple rate of return on the stock in any month t, R_t, is either 40 percent or -40 percent. In addition to past stock returns, the investor's sample contains realizations of a random state variable, s_{t-1}. We label the two possible realizations of this state variable as $s_{t-1} = 1$ ("state 1") and $s_{t-1} = 2$ ("state 2"). The parameter vector is given by $\theta = (\theta_1, \theta_2)$,

[9] A CER is interpreted as the monthly rate of return on wealth that, if earned with certainty, would provide the investor with utility equal to the expected utility $\bar{v}$ for the given allocation. In general, the CER is obtained by solving the equation

$$v(W_T(1 + \text{CER})) = \bar{v},$$

where v is the utility function in (3).

where θ_i is the probability that, conditional on observing $s_{t-1} = i$ at the beginning of month t, the subsequently observed stock return in month t will be 40 percent. The investor assumes that θ_1 and θ_2 are constant over time. The state variable s_{t-1} is assumed to be identically and independently distributed over time, independent of past returns, and drawn from a binomial distribution whose parameter is independent of the θ_i's.

The investor's prior joint distribution assumes the parameters θ_1 and θ_2 are independent, $p(\theta_1, \theta_2) = p(\theta_1) \cdot p(\theta_2)$, and the marginal prior distribution for each θ_i is given by

$$p(\theta_i) = \frac{[\theta_i(1 - \theta_i)]^{c-1}}{B(c, c)}, \quad i = 1, 2, \tag{9}$$

where $c \geq 1$ and $B(\cdot)$ is the "Beta" function. The prior joint distribution for θ_1 and θ_2 implies a prior distribution for the difference $(\theta_1 - \theta_2)$, and the latter distribution reflects the investor's prior beliefs about the extent to which stock returns can be predicted using the state variable. We consider three values of c for the prior distribution in (9): $c = 1$, $c = 6$, and $c = 21$. When $c = 1$, the prior distribution for each of the θ_i's is the Bayes-Laplace uniform prior on (0, 1). As c increases, the prior distribution becomes more concentrated around 0.5. The implied prior distributions of $(\theta_1 - \theta_2)$, for the three values of c, are numerically evaluated and displayed as dashed curves in Figure 1. The larger is c, the more concentrated around zero is this prior distribution, and the more weighted against predictability are the investor's beliefs.

The investor's data set Φ_T consists of Ψ_T, a sample of T pairs of past realizations of the state variable and the subsequent stock return, and s_T, the state observed at the end of the most recent month T:

$$\Phi_T = \{\Psi_T, s_T\}. \tag{10}$$

In our example, the sample Ψ_T includes $T = 16$ pairs with $T_i = 8$ months for each state, and these sample data can be represented by a 2×2 contingency table:

	State 1	State 2
R = 40%	6	4
R = −40%	2	4

Conditional on observing state i at the beginning of each of T_i months, the probability that the 40 percent stock return will be realized in M_i of those months is given by the binomial likelihood function,

$$p(M_i|\theta_i, T_i) = \binom{T_i}{M_i} \theta_i^{M_i}(1 - \theta_i)^{T_i - M_i}, \qquad i = 1, 2. \tag{11}$$

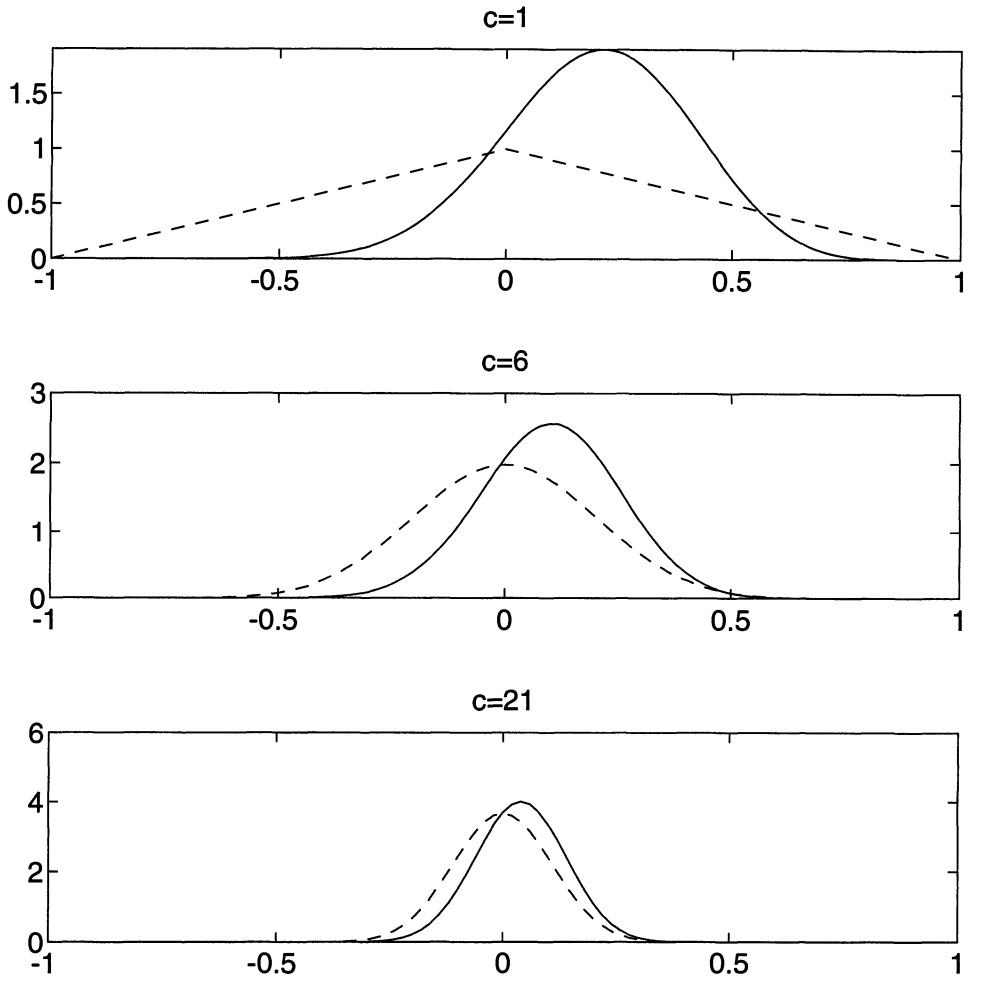

Figure 1. Prior and posterior distributions for $(\theta_1 - \theta_2)$. The plot displays, for $c = 1$, $c = 6$, and $c = 21$, the prior distribution (dashed curve) and the posterior distribution (solid curve) for the quantity $(\theta_1 - \theta_2)$. In the two-state, two-outcome example, θ_i is the probability of the high-return (40 percent) outcome, conditional on observing state i, and the parameter c determines the strength of the prior beliefs that $\theta_1 = \theta_2 = 0.5$.

Combining the prior distribution in (9) with the likelihood function in (11) yields the marginal posterior distribution for θ_i, a Beta distribution:[10]

$$p(\theta_i|\Phi_T) = p(\theta_i|\Psi_T) = p(\theta_i|M_i,\ T_i)$$
$$= \frac{\theta_i^{M_i+c-1}(1-\theta_i)^{T_i-M_i+c-1}}{B((M_i+c),\ (T_i-M_i+c))}, \quad i = 1,\ 2. \quad (12)$$

[10] See Zellner (1971, p. 39).

The investor's posterior beliefs about the predictability of stock returns are reflected in the implied posterior distributions for the difference $(\theta_1 - \theta_2)$. These distributions are numerically evaluated and displayed, for the three values of c, by the solid curves in Figure 1. All of these posterior distributions center at positive values, but, the larger is c, the closer is the posterior distribution of $(\theta_1 - \theta_2)$ to the prior distribution (which is centered at zero).

Conditional on observing state j at time T, the predictive distribution of the stock return at time $T + 1$ is a binomial distribution, where $\hat{p}$ denotes the predictive probability that the return will be 40 percent. To obtain $\hat{p}$, first note that

$$p(R_{T+1} = 40\% | \Phi_T = \{\Psi_T, s_T = j\}, \theta_j) = \theta_j, \tag{13}$$

and then substitute (13) into (7) to get

$$\begin{aligned}\hat{p} = p(R_{T+1} = 40\% | \Phi_T = \{\Psi_T, s_T = j\}) \\ = \int p(R_{T+1} = 40\% | \Phi_T, \theta_j) \cdot p(\theta_j | \Phi_T) d\theta_j \qquad (14) \\ = \int \theta_j \cdot p(\theta_j | \Phi_T) d\theta_j \\ = \frac{(M_j + c)}{(T_j + 2c)}.\end{aligned}$$

With this binomial predictive distribution for R_{T+1}, the investor's optimization problem in (4) becomes

$$\max_{0 \le \omega \le 1} [\hat{p} \ln(1 + 0.4\omega) + (1 - \hat{p}) \ln(1 - 0.4\omega)], \tag{15}$$

and its solution is

$$\omega^* = \begin{cases} 0 & \text{if } (2\hat{p} - 1) \le 0 \\ \left(\dfrac{2\hat{p} - 1}{0.4}\right) & \text{if } 0 < (2\hat{p} - 1) < 0.4 \\ 1 & \text{if } (2\hat{p} - 1) \ge 0.4. \end{cases} \tag{16}$$

In the sample Ψ_T given in our example, $M_2 = 4$ and $T_2 = 8$, and we see from (14) that, when $s_T = 2$, $\hat{p} = 0.5$ for all values of c in the prior. Since the predictive distribution of the stock return in this case is symmetric around zero, any risk-averse investor refrains from investing any money in stock when $s_T = 2$. (Recall that the riskless rate is zero.) It is easily verified from (16) that $\omega^* = 0$ at $\hat{p} = 0.5$.

If $s_T = 1$, however, then the predictive probability $\hat{p}$ is greater than 0.5, and the optimal stock allocation ω^* is positive. Specifically, since $M_1 = 6$ and $T_1 = 8$ in the sample Ψ_T, we see from (14) that $\hat{p} = (6 + c)/(8 + 2c)$. This value of $\hat{p}$ is given below for each value of c, along with the optimal allocation ω^* from (16), the expected stock return

$$\hat{R}_{T+1} = E\{R_{T+1}|\Phi_T\} = \hat{p} \cdot 40\% + (1 - \hat{p}) \cdot (-40\%), \tag{17}$$

and the corresponding expected monthly return on the optimal portfolio,

$$\hat{R}_p = \omega^* \cdot \hat{R}_{T+1}. \tag{18}$$

c	$\hat{p}$	$\hat{R}_{T+1}$	ω^*	$\hat{R}_p$	ΔCER
1	0.70	0.16	1.00	0.16	0.0858
6	0.60	0.08	0.50	0.04	0.0203
21	0.54	0.032	0.20	0.0064	0.0032

The values in the last column will be discussed later.

The above results demonstrate the potential economic significance of stock return predictability. Although the investor's prior beliefs about the θ_i's 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. For a prior with $c = 1$, the sample evidence leads the investor to choose a stock allocation of 100 percent if state 1 is observed but zero if state 2 is observed. As c increases, so that prior beliefs become weighted more heavily against predictability, the stock allocation for state 1 decreases, but it is still 20 percent for $c = 21$.

In this two-state example, the certainty-equivalent comparison discussed earlier is conducted as follows. For a given state $s_T = j$, we compare the investor's CER for the optimal allocation to the investor's CER for a suboptimal allocation, where the latter allocation would have been optimal had state $i \neq j$ occurred. The CER for both allocations is computed under the same probability distribution, the predictive pdf for $s_T = j$. We report this comparison here for $s_T = 1$, so in the notation introduced in the previous subsection, $\Phi_T = \{\Psi_T, s_T = 1\}$ and $\Phi_T^a = \{\Psi_T, s_T = 2\}$. It is easily verified from (12) that the condition in (8) is satisfied. The allocation ω^*, optimal when $s_T = 1$, is compared to a suboptimal allocation of $\omega^a = 0$ (the optimal allocation if $s_T = 2$). The difference between the CER of ω^* and the CER of ω^a is given above as ΔCER for each of the three values of c. This measure ranges from 8.58 percent when $c = 1$ to 0.32 percent when $c = 21$. In all three cases, however, ΔCER is more than half of the expected return on the optimal portfolio, providing an illustration of the potential economic significance of sample evidence on stock-return predictability.

D. Economic Significance, Statistics, and Conditional Bayesian Decisions

In academic research, empirical evidence about the predictability of stock returns is often evaluated in terms of standard test statistics.[11] In general, these test statistics are defined with respect to the point null hypothesis that returns are unpredictable, and the strength of the empirical evidence is often assessed by examining a test's p-value. Some readers of a published study might wish to make a formal accept/reject decision about a hypothesis, but we suggest that the p-value is probably more often interpreted as a continuous measure of the strength or reliability of the evidence.[12] Similarly, in our analysis, the investor's allocation decision does not involve accepting or rejecting a specific hypothesis or, more generally, selecting a model from a set of possible models. The investor's problem is to select a portfolio, not a model. Moreover, we doubt that our approach would be very helpful to a researcher whose goal is hypothesis testing or model selection, whether from a Bayesian or frequentist perspective.[13] Rather, our objective, as stated earlier, is simply to use the asset-allocation decision as a metric by which to assess the economic importance of the empirical evidence on predictability, and we find that such an assessment often contrasts with those based on p-values or other standard statistical measures.[14]

Interpretations of p-values no doubt differ across readers, with some readers attaching more importance than others to reported p-values of, say 1 percent, but we suggest that p-values of 30 percent or more, if published, would probably not be taken seriously by many readers as evidence of stock-return predictability. The potential contrast between such a p-value and an outcome of a conditional Bayesian asset-allocation decision is easily illustrated in the context of the simple example presented above. Fisher's exact test for the 2×2 contingency table is used to construct a p-value associated with the null hypothesis of no predictability, $\theta_1 = \theta_2$.[15] This test is based on the conditional distribution of M_1 (the number of periods with a 40 percent return following state 1) given the two-way table's row and column sums. A one-tailed p-value is computed as the probability of getting $M_1 \geq 6$, which equals 0.304 in the previous example. The typical interpretation of such a p-value contrasts

[11] For a recent review of the literature on stock-return predictability, see Kaul (1995).

[12] Reporting the p-value as a flexible measure of the evidence, as opposed to rejecting or accepting a null versus an alternative, is generally associated with the views of R. A. Fisher, in contrast to the views of Neyman and Pearson generally associated with the accept/reject decision. See Fisher (1973).

[13] For a Bayesian approach to testing the hypothesis of return predictability, see Kothari and Shanken (1995).

[14] A separate issue regarding the interpretation of p-values arises in a contrast between Bayesian and frequentist approaches to hypothesis testing. For example, Shanken (1987) presents a scenario in which any p-value less than 0.46 would, based on a Bayesian posterior-odds ratio, constitute evidence against a null hypothesis of portfolio efficiency. See Lindley (1957) for an early discussion of this issue.

[15] See Kendall and Stuart (1979).

sharply with the economic significance of stock-return predictability as reflected in the investor's asset allocation decisions.[16]

Another example of a contrast between a p-value and the economic significance of sample evidence can be constructed using results reported by Brown (1979), who examines asset-allocation in an i.i.d. setting. In a stocks-versus-cash allocation decision, Brown compares ω_B, the optimal stock allocation chosen by a Bayesian investor with noninformative prior beliefs, to ω_C, the allocation that would be optimal if sample estimates were simply treated as true parameters. For example, if the Sharpe ratio of stocks computed using a sample of 16 monthly (simple) returns is 0.2, Brown reports that $\omega_B/\omega_C = 0.82$.[17] Brown does not examine measures of statistical significance, but it is easily seen that, with a sample Sharpe ratio of 0.2 and 16 observations, the t-statistic for the hypothesis of a zero expected excess stock return is equal to $\sqrt{16} \cdot (0.2) = 0.8$, and the one-sided p-value is 0.22. This p-value contrasts sharply with the economic significance of the unconditional equity premium as reflected in the investor's asset allocation decision. That is, even though the sample evidence for a non-zero unconditional equity premium seems weak, when judged by the p-value, the investor, rather than allocating his entire portfolio to cash, chooses a stock allocation equal to 82 percent of the allocation that would be chosen if the true Sharpe ratio were known to be 0.2. Although, to our knowledge, the potential contrast between p-values and asset allocations in an i.i.d. setting has not been previously noted, such a comparison is distinct from the question of whether the most recent values of a set of predictive variables should impact the investor's asset allocation, which is the question we address in assessing the economic significance of stock-return predictability.

In computing optimal asset allocations and comparing certainty equivalent returns, we compute expected utility with respect to the Bayesian investor's predictive pdf. Thus, expected utility is as perceived by the investor, conditional on the data set Φ_T, and the relative desirability and optimality of an allocation is judged based on that conditional expected utility. Given the investor's prior beliefs, the optimal allocation ω^* is determined by the data, and we can denote such a dependence as $\omega^*(\Phi_T)$. Intentionally omitted from our investor's asset-allocation decision, however, is a consideration of the "typical performance" of $\omega^*(\Phi_T)$ when applied repeatedly to data sets generated randomly from a given distribution. The performance in repeated samples of the decision rule $\omega^*(\Phi_T)$, where the data set Φ_T is viewed as random, invokes the frequentist concept of "risk." The *risk function* of a decision rule, defined on

[16] The contrast between the p-value and economic significance in that example is not restricted to cases where the number of observations is small. If, instead of the 16-observation sample, there are $T = 200$ observations with $T_i = 100$ observations per state, $M_1 = 56$, and $M_2 = 50$, then the one-tailed p-value is 0.24. The stock-allocation is 0 percent in state 2, whereas the stock allocation in state 1 ranges from 29 percent when $c = 1$ to 21 percent when $c = 21$.

[17] This result obtains with both the quadratic and negative exponential preferences considered by Brown. The *Sharpe ratio* is defined as the ratio of expected excess return to the standard deviation of the return.

the parameter space Θ, is equal to the expected utility (or loss) with respect to the joint probability distribution of Φ_T and r_{T+1}, as determined by a given value of θ.

The frequentist risk function is often used to compare the performance of decision rules, or to compare models that give rise to different rules. For some values of θ, such as when θ_1 and θ_2 in the previous example are sufficiently close to each other, an asset-allocation decision rule based on an i.i.d. model, wherein the investor ignores the predictive state variable and simply pools the returns data, might have lower frequentist risk than the decision rule in (16) based on the two-state model. For other values of θ, where θ_1 and θ_2 are sufficiently far apart, the two-state decision rule might have lower frequentist risk. In general, selecting the decision rule with the lowest frequentist risk requires knowledge about the parameter vector θ, and neither we nor the investor know the true value of θ. The frequentist repeated-sampling criterion can be related to a conditional Bayesian decision in the following sense. "Averaging" the frequentist risk function over different values of θ, or more formally, integrating the risk function over the prior pdf for θ, gives the *Bayes risk* of the decision rule. If this integral is finite, then the conditional Bayesian decision rule will also minimize Bayes risk.[18]

II. Asset Allocation Based on Regression Evidence

A. The Conditional Distribution of the Stock Return

The continuously compounded excess stock return r_t is the dependent variable in the regression

$$r_t = x'_{t-1}b + \epsilon_t, \tag{19}$$

where $x'_{t-1} = (1\ y'_{t-1})$, and the $N \times 1$ vector y_{t-1} contains N "predictive" variables that are observed at the end of month $t - 1$. The disturbances ϵ_t, $t = 1, 2, \ldots, T$, are assumed to be independent mean-zero draws from a normal distribution with variance σ^2_ϵ. Although we assume $\mathrm{E}\{\epsilon_t | x_{t-1}\} = 0$, the vector y_{t-1} is in general stochastic, and some elements of y_{t-1} can be correlated with past disturbances.[19] Such correlation obviously arises when y_{t-1} contains lagged values of r_t, but it is likely to arise more generally for many variables commonly used to predict stock returns. For example, numerous previous studies specify y_{t-1} to include the dividend yield at the end of month $t - 1$, which is likely to be negatively correlated with the unexpected return in that month, ϵ_{t-1}.[20] Thus, the regression in (19) departs somewhat from the stan-

[18] See Berger (1985) for extensive discussions and comparisons of frequentist and Bayesian decision principles.

[19] See Stambaugh (1986) and Nelson and Kim (1993) for treatments of this problem in a frequentist setting.

[20] The first study to investigate the ability of dividend yields to predict stock returns is, to our knowledge, Rozeff (1984). Later studies include Fama and French (1988) and Goetzmann and Jorion (1995).

dard Bayesian regression framework, in which y_{t-1} is assumed to be either nonstochastic or stochastic but distributed independently of the disturbances with a distribution involving neither b nor σ_ϵ.[21]

Our approach to allowing for the stochastic properties of the regressors is to assume that r_t is the first element of y_t and then to model y_t as a first-order vector autoregression (VAR). That is, we assume

$$y_t' = x_{t-1}'B + u_t', \tag{20}$$

where B is an $(N + 1) \times N$ matrix of regression coefficients whose first column is b, and u_t is an N-vector of disturbances whose first element is ϵ_t. Such a VAR representation was proposed previously by Kandel and Stambaugh (1987) to model the predictability of stock returns.[22] We assume that the vectors u_t, $t = 1, \ldots, T$, are independent mean-zero draws from a multivariate normal distribution with a covariance matrix equal to Σ. The T observations for this VAR are represented in the matrix notation,

$$Y = XB + U, \tag{21}$$

where

$$Y = \begin{bmatrix} y_1' \\ y_2' \\ \cdot \\ \cdot \\ y_T' \end{bmatrix}, \quad X = \begin{bmatrix} 1 & y_0' \\ 1 & y_1' \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & y_{T-1}' \end{bmatrix}, \quad \text{and } U = \begin{bmatrix} u_1' \\ u_2' \\ \cdot \\ \cdot \\ u_T' \end{bmatrix}. \tag{22}$$

The joint probability density function for the elements of U is given by

$$p(U|\Sigma) \propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2} \text{tr } U'U\Sigma^{-1} \right], \tag{23}$$

where "tr" denotes the trace operator. Following an approach common to many Bayesian time-series models, our analysis takes the initial observation y_0 as effectively nonstochastic.[23] In other words, we essentially assume that the investor's prior beliefs about the model's parameters do not depend on y_0, even though the investor's information set Φ_T includes both the "pre-sample" observation y_0 as well as the "sample" observations $y_1, \ldots, y_T$. (Note that, other than y_0 and a vector of ones, X simply contains the first $T - 1$ rows of Y.) A change of variables from U to Y gives the likelihood function,

$$p(Y|B, \Sigma, x_0) \propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2} \text{tr } (Y - XB)'(Y - XB)\Sigma^{-1} \right], \tag{24}$$

[21] See, for example, Zellner (1971, page 59) for the assumptions in the standard Bayesian regression model.

[22] See also Campbell (1991) and Hodrick (1992).

[23] See, for example, Hamilton (1994, p. 358).

since the Jacobian of the transformation from U to Y is equal to unity.[24]

The following sample statistics for the above model are useful in the subsequent analysis:

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{25}$$

$$\hat{\sigma}_r^2 = \frac{1}{T}\sum_{t=1}^{T} (r_t - \bar{r})^2, \tag{26}$$

$$\bar{y} = \frac{1}{T}\sum_{t=0}^{T-1} y_t, \tag{27}$$

$$\hat{V}_y = \frac{1}{T}\sum_{t=0}^{T-1} (y_t - \bar{y})(y_t - \bar{y})', \tag{28}$$

$$\hat{b} = (X'X)^{-1}X'y = \begin{bmatrix} \bar{r} - \hat{\beta}'\bar{y} \\ \hat{\beta} \end{bmatrix} \tag{29}$$

$$R^2 = 1 - \frac{(r - X\hat{b})'(r - X\hat{b})}{T\hat{\sigma}_r^2}, \tag{30}$$

and r is the first column of Y.

The remaining assumption necessary in obtaining the conditional distribution of the stock return is the specification of the investor's prior beliefs about the model's parameter values. We consider two alternative specifications. In the first, we assume that the investor's beliefs are given by the "diffuse" prior,

$$p(B, \Sigma) \propto |\Sigma|^{-(N+2)/2}, \tag{31}$$

which is intended to represent vague or noninformative prior beliefs about the parameters.[25] With this prior distribution and the likelihood function in (24), the predictive pdf is easily obtained from known results (see Appendix, part A).

[24] Define the $TN \times 1$ vectors $\tilde{u} = \text{vec}(U)$ and $\tilde{y} = \text{vec}(Y)$, where vec( ) creates a column vector by stacking the (transposed) rows of the matrix. It is then easily verified that the Jacobian of the transformation from U to Y equals unity, since the $TN \times TN$ matrix $\partial\tilde{u}'/\partial\tilde{y}$ is lower triangular with all diagonal elements equal to unity.

[25] This prior specification for B and Σ can be found, for example, in Zellner (1971, chapter 8), who discusses its foundations in the invariance theory due to Jeffreys (1961). As Zellner also discusses, one result often obtained using such priors is that confidence regions for parameter values obtained from the posterior distribution correspond closely to frequentist confidence regions for parameter estimators.

When $T > 2N + 1$, the predictive pdf for r_{T+1} is given by a Student t distribution and can be written in terms of y_T and the above sample statistics:

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)}\left(\frac{1}{(\nu-2)\sigma_T^2}\right)^{1/2}\left[1+\left(\frac{1}{\nu-2}\right)\frac{(r_{T+1}-\mu_T)^2}{\sigma_T^2}\right]^{-(\nu+1)/2}, \tag{32}$$

where

$$\mu_T = \mathrm{E}\{r_{T+1}|\Phi_T\} = \bar{r} + \hat{\beta}'(y_T - \bar{y}), \tag{33}$$

$$\sigma_T^2 = \mathrm{var}\{r_{T+1}|\Phi_T\} = \frac{T}{T-2(N+1)}(1-R^2)\left[1+\frac{1}{T}(1+q)\right]\hat{\sigma}_r^2, \tag{34}$$

$$q = (y_T - \bar{y})'\hat{V}_y^{-1}(y_T - \bar{y}), \text{ and} \tag{35}$$

$$\nu = T - 2N. \tag{36}$$

In the second specification of the investor's prior beliefs, we construct the prior distribution as a posterior distribution for the parameter values that would result from combining the diffuse prior in (31) with a hypothetical "prior" sample of size T_0 in which the sample R-squared is *exactly* equal to zero. Except for this no-predictability feature, the hypothetical prior sample is assumed to be otherwise similar to the actual sample, in that the prior sample is assumed to produce the same values as the actual sample for the statistics corresponding to $\bar{r}$, $\hat{\sigma}_r^2$, $\bar{y}$, and $\hat{V}_y$.[26] With this "no-predictability" informative prior, the predictive pdf for r_{T+1} is, by basic principles of Bayesian analysis, the same as that obtained using the diffuse prior and a sample of size $T^* = T + T_0$ that combines the actual sample with the hypothetical prior sample. Thus, the predictive pdf is in precisely the same form as (32), where μ_T, σ_T^2 and ν, although redefined, are still written in terms of y_T and the same sample statistics in (25) through (28):

$$\mu_T = \mathrm{E}\{r_{T+1}|\Phi_T\} = \bar{r} + \left(\frac{T}{T^*}\right)\hat{\beta}'(y_T - \bar{y}), \tag{37}$$

$$\sigma_T^2 = \mathrm{var}\{r_{T+1}|\Phi_T\} = \frac{T^*}{T^*-2(N+1)}\left[1-\left(\frac{T}{T^*}\right)^2 R^2\right]\left[1+\frac{1}{T^*}(1+q)\right]\hat{\sigma}_r^2, \tag{38}$$

$$\nu = T^* - 2N. \tag{39}$$

Note that, since $\bar{y}$ and $\hat{V}_y$ are assumed to be the same in the actual and hypothetical samples, q is still defined as in (35).

[26] Using statistics from the actual sample to supply some of the parameters for the prior distribution can be termed "empirical Bayes," at least under the broad interpretation of that classification. See, for example, Maritz and Lwin (1989, p. 14).

The strength of the "no-predictability" prior is determined by T_0, the size of the hypothetical prior sample. In Part B of the Appendix, we report an investigation of the marginal prior distribution of $\tilde{R}^2$, the true population R-squared for the excess-return regression in (19). When T_0 is held fixed across different values of N, we find that, as N increases, the prior assigns greater mass to higher values of $\tilde{R}^2$. In the analysis below, we investigate the asset-allocation decision across a range of values for N. It seems desirable that the results based on the informative prior not be driven by differences in prior beliefs that depend on N. Thus, our objective is to specify T_0 such that the implied priors for $\tilde{R}^2$ are similar across various specifications of N. We find that this objective is satisfied better by specifying T_0 as an increasing function of N, calculated as follows. The number of distinct parameters in B and Σ is equal to $\frac{1}{2}N(N+1)$, and the number of data entries in the hypothetical prior sample of T_0 months is $N \cdot T_0$. In all of the calculations reported below, we set $T_0 = 75 \cdot (N+1)$, which is equivalent to 50 data entries per parameter. The number of variables that we consider ranges from 2 to 50, so the size of the hypothetical prior sample ranges from 225 months (18.75 years) to 3825 months (318.75 years).

It is easily verified that the quantities $\hat{\beta}'(y_T - \bar{y})$ and $q = (y_T - \bar{y})'\hat{V}_y^{-1}(y_T - \bar{y})$ are invariant with respect to nonsingular linear transformations of y_t. Therefore, under either specification of the prior distribution, μ_T, σ_T, and hence the predictive pdf for r_{T+1}, are invariant with respect to nonsingular linear transformations of y_t in which the first element of the transformed vector remains r_t. In other words, the predictive pdf for r_{T+1} is unaffected by the degree of contemporaneous correlation among the N predictive variables.

The predictive variance σ_T^2, in both (34) and (38), incorporates uncertainty about parameter values, i.e., estimation risk.[27] Suppose that we hold constant the values of the sample statistics in (25) through (30). Given the number of predictive variables, N, and their most recent values, y_T, the predictive variance σ_T^2 approaches $(1 - R^2)\hat{\sigma}_r^2$ as T becomes large. In a finite sample, the positive difference between σ_T^2 and that limiting value reflects the presence of estimation risk. Uncertainty about the parameters also results in a positive relation between σ_T^2 and the distance of y_T from the sample mean $\bar{y}$, as measured by q. In this sense, the predictive pdf for r_{T+1} exhibits conditional heteroskedasticity. This source of heteroskedasticity is distinct, however, from conditional heteroskedasticity incorporated in the likelihood function (regression model). The latter extension would be an interesting direction for future research. Finally, note that, given T and y_T, the predictive variance is increasing in N, due to the presence of the term $-2(N+1)$ in the denominator of both (34) and (38). This effect is analogous to that in the standard frequentist setting for a multiple regression with $N+1$ regressors (including the inter-

[27] A number of previous studies investigate the economic significance of stock-return predictability by examining the performance of asset-allocation strategies constructed by essentially treating sample parameter estimates as true values. See, for example, Breen, Glosten, and Jagannathan (1989), Solnik (1993), and Lo and MacKinlay (1995).

cept), where the unbiased estimate of the residual variance is obtained by dividing the sum of the squared fitted residuals by $T - N - 1$.

B. Computing Optimal Asset Allocations

The predictive pdf in (32) is the conditional distribution used in computing expected utility, $\mathrm{E}\{v(W_{T+1})|\Phi_T\}$. To our knowledge, this study is the first to report calculations of optimal portfolio allocations in a non-i.i.d. setting using a regression-based predictive pdf. Such a calculation was proposed much earlier, however. Zellner and Chetty (1965) suggest using a regression-based predictive pdf to compute investment portfolio weights that maximize expected utility, although they do not specify a utility function or perform such calculations.[28] We discuss in this subsection a number of issues related to computing the optimal asset allocations.

Given the Student t form for the conditional distribution $p(r_{T+1}|\Phi_T)$, the integration in (4) extends from $-\infty$ to ∞. As noted earlier, we restrict attention in this study to asset allocations that do not involve short selling either the risky or riskless asset, i.e., the stock allocation ω obeys $0 \leq \omega \leq 1$. When $A > 1$, expected utility is equal to $-\infty$ when $\omega = 1$, although the optimal ω can be very close 1.[29] We simply restrict $\omega \leq 0.99$ throughout. The maximization problem is solved numerically, since we are unaware of an exact analytic solution.[30] In virtually all cases, however, the optimal allocation to stocks is well approximated by

$$\hat{\omega}^* = \frac{\mu_T}{A\sigma_T^2} + \frac{1}{2A}, \tag{40}$$

[28] In fact, Zellner and Chetty (1965) specify a multivariate regression model with a diffuse prior identical to (31). In their framework, however, all regressors are assumed to be nonstochastic, and the N equations in the multivariate regression result from the consideration of N assets. In our single-asset framework, the N equations reflect the use of N stochastic regressors.

[29] When $\omega < 1$, wealth is bounded above zero, so that utility, and thus expected utility, are bounded from below. When $\omega = 1$, wealth can be arbitrarily close to zero, so that utility is unbounded from below. In that case, the lower tail of the predictive pdf does not shrink rapidly enough as utility approaches $-\infty$, and this property essentially reflects the leptokurtosis of the Student t distribution. The integral exists in the limit as $T \to \infty$, in which case the predictive Student t pdf converges to its limiting normal distribution. In a similar vein, although the moments of the simple rate of return $R_{T+1} = \exp(r_{T+1})$ do not enter our analysis, the conditional mean and variance of R_{T+1} do not exist when T is finite, which follows from a similar observation about the moment-generating function of the Student t distribution, as noted in Kendall and Stuart (1977, p. 63). The nonexistence of certain integrals under the Student t distribution enters more directly in earlier studies that use simple returns. Brown (1979), for example, encounters the nonexistence of expected utility for *any* nonzero allocation to stocks when utility is specified as negative exponential, so he defines expected utility in that case using a normal approximation to the Student t predictive pdf.

[30] 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).

subject to the [0, 0.99] bounds. Equation (40) gives the exact solution when the investor rebalances continuously, the instantaneous interest rate is a constant i_{T+1}, and the continuously compounded stock return over the discrete period from T to $T + 1$ has mean $\mu_T + i_{T+1}$, variance σ_T^2, and is an infinitely divisible normal random variable.[31] If (40) is rewritten slightly as

$$\hat{\omega}^* = \frac{\alpha_T - i_{T+1}}{A\sigma_T^2}, \tag{41}$$

where $\alpha_T = \mu_T + \frac{1}{2}\sigma_T^2 + i_{T+1}$, the expected instantaneous rate of return on the stock, then $\hat{\omega}^*$ is seen as a familiar mean-variance result.

As will be discussed in a later subsection, we analyze optimal asset allocations for a variety of samples that differ with respect to sample size (T), number of predictive variables (N), and the regression R^2. The remaining sample quantities required to construct the predictive pdf are $\bar{r}$, $\hat{\sigma}_r$, i_{T+1} (the riskless interest rate for month $T + 1$), $\hat{\beta}'(y_T - \bar{y})$, and $q = (y_T - \bar{y})' \hat{V}_y^{-1}(y_T - \bar{y})$. The first three quantities, $\bar{r}$, $\hat{\sigma}_r$, and i_{T+1}, are held constant across all samples. We set $\bar{r} = 0.49$ percent and $\hat{\sigma}_r = 5.60$ percent, the sample estimates for the 804-month period from January 1927 through December 1993 using the continuously compounded monthly return on the value-weighted portfolio of the New York Stock Exchange in excess of the continuously compounded one-month T-bill rate, and we set $i_{T+1} = 0.235$ percent, the continuously compounded monthly yield on the Treasury bill with 27 days to maturity as of 12/31/93.[32]

For each specification of T, N, and R^2, we wish to investigate the behavior of the optimal stock allocation ω^* over a range of values for the vector of predictive variables, y_T. The value of y_T enters q, as will be discussed below, but the key role for y_T is in determining the one-step-ahead fitted value from the regression, $\bar{r} + \hat{\beta}'(y_T - \bar{y})$. Thus, rather than specifying completely the vector y_T, we simply specify a value for $\hat{\beta}'(y_T - \bar{y})$. This difference between the fitted regression value and the sample average return can be stated in units of the fitted values' sample standard deviation. The series of in-sample fitted regression values, $\bar{r} + \hat{\beta}'(y_t - \bar{y})$, $t = 0, \ldots, T - 1$, has sample standard deviation equal to $\sqrt{R^2}\hat{\sigma}_r$. We specify δ such that

$$\hat{\beta}'(y_T - \bar{y}) = \delta\sqrt{R^2}\hat{\sigma}_r. \tag{42}$$

That is, the fitted value of r_{T+1} based on y_T is specified as being δ sample standard deviations of the fitted values away from the overall sample mean, $\bar{r}$. We compute the optimal allocation ω^* for five specifications of δ: -1.0, -0.5,

[31] A random variable is "infinitely divisible" if, for all n, the random variable can be expressed as a sum of n independent and identically distributed random variables. See Ingersoll (1987, chapter 12). The derivation of (40) is a straightforward application of results contained in Merton (1969), and an expression equivalent to (40) also arises as a solution to a special case of the dynamic consumption-investment problem analyzed in that study.

[32] The stock returns and T-bill rates are obtained from the CRSP files.

0, 0.5, and 1.0. An alternative approach would be to consider a fixed range of values for the fitted excess return, say $\bar{r}$ plus or minus 100 basis points, but this approach gives a range of fitted values that is too modest when R^2 is high and too extreme when R^2 is low. With $R^2 = 0.15$, for example, the in-sample standard deviation of the fitted values is 217 basis points, so a fitted value for r_{T+1} might easily lie well outside the 100-basis-point range. On the other hand, when $R^2 = 0.002$, the standard deviation of the fitted values is only 25 basis points, so it would be quite unlikely that a fitted value would differ from $\bar{r}$ by as much as 100 basis points. The samples we consider in the subsequent analysis include a broad range of R^2 values, so calibrating a range for the current values of the predictive variables using δ produces a plausible set of one-step-ahead fitted predictions that might arise from a given regression.

The quantity $q = (y_T - \bar{y})'\hat{V}_y^{-1}(y_T - \bar{y})$ summarizes the "standardized" differences between the current values of the predictive variables and their sample means. We consider samples where the number of observations (T) and the number of predictive variables (N) cover a wide range. Since our analyses of these various cases do not employ actual data, our aim is to specify a reasonable value of q for a given combination of T and N. If $N = 1$, then q is determined uniquely by the deviation of the fitted one-step-ahead regression prediction from the overall mean. Specifically, if $N = 1$ then $q = \delta^2$, where δ is defined as in (42). For $N \geq 2$, however, this simple correspondence between q and the fitted regression prediction no longer exists. In general, for a given value of δ, q has a lower bound of δ^2 but no upper bound. We consider only samples in which $N \geq 2$, so we simply specify, for a given T and N, a value of q that is constant across different realizations of y_T. If q is held constant across realizations of y_T, then a larger value of q produces smaller differences between optimal stock allocations at different one-step-ahead fitted regression predictions. This effect can be seen most easily by noting that σ_T^2 appears in the denominator of $\hat{\omega}^*$ in (40), and σ_T^2 is increasing in q (equation (34) or (38)). Given this study's orientation, we wish to be conservative in representing the differences in optimal allocations arising from different realizations of the fitted regression prediction, so we wish to select a value for q from the high end of its plausible range. The sampling distribution for q depends on the degree of serial dependence in the elements of y_t, with positive serial dependence leading to larger values for q.[33] Here again we follow a conservative approach. For each T and N, we take q as the 99th percentile of a Monte Carlo distribution for 5000 samples generated with each element of y_t following an AR(1) process with normal disturbances and autocorrelation coefficient equal to 0.99.[34]

[33] This statement is based on our Monte Carlo evidence. We are unaware of an analytic finite-sample result in the presence of serial dependence.

[34] We generate y_t with a zero mean vector and a scalar variance-covariance matrix, but this is without loss of generality, since, as noted previously, q is invariant under nonsingular linear transformations of y_t. The effect of autocorrelation on the sampling distribution of q can be substantial. If $y_0, y_1, \ldots, y_T$ are serially independent draws from a multivariate normal distribution, then $(T - N)/(N(T + 1))\ q$ is distributed as $F_{N,T-N}$. See, for example, Anderson (1984, chapters 5 and 7). In that i.i.d. case, the 99th percentiles of q when $T = 804$ are equal to 9.2 with

Table I

Optimal Stock Allocations and Comparisons of Certainty Equivalents with 804 Observations and 25 Variables

Optimal stock allocations and certainty equivalents are computed with respect to a Bayesian investor's predictive pdf based on regression evidence in which the (unadjusted) sample R-squared is equal to R^2. The fitted one-month-ahead regression prediction of the excess stock return, r_{T+1}, is δ sample standard deviations of the fitted values away from the sample average excess return $\bar{r}$. The investor's relative risk aversion is equal to A. The "Comparison of Certainty Equivalents" gives the difference in certainty equivalent monthly returns between the optimal allocation and the allocation that would have been chosen for $\delta = 0$. The p-value is computed for the hypothesis that the regression's slope coefficients are jointly equal to zero. The no-predictability informative prior is equivalent to a posterior that combines diffuse prior beliefs with a hypothetical T_0-month sample in which all estimated slopes are exactly zero. T_0 is specified such that the hypothetical sample contains 50 data entries per parameter, which gives $T_0 = 1950$ when $N = 25$.

			Optimal Stock Allocation (percent) for δ Equal to					Comparison of Certainty Equivalents (basis pts.) for δ Equal to			
R^2	p-value	A	−1	−0.5	0	0.5	1	−1	−0.5	0.5	1.0
Panel A: Diffuse prior											
0.025	0.75	1	0	57	99	99	99	38.6	6.0	0.0	0.0
0.025	0.75	2	0	28	61	93	99	28.5	7.2	7.2	24.0
0.025	0.75	5	0	11	24	37	50	11.4	2.9	2.9	11.5
0.055	0.01	1	0	25	99	99	99	80.8	18.1	0.2	0.3
0.055	0.01	2	0	12	62	99	99	55.8	16.3	15.3	39.9
0.055	0.01	5	0	5	25	45	65	22.2	6.5	6.5	26.1
Panel B: No-predictability informative prior											
0.025	0.75	1	99	99	99	99	99	0.0	0.0	0.0	0.0
0.025	0.75	2	53	68	83	99	99	4.0	1.0	1.0	3.0
0.025	0.75	5	21	27	33	39	46	1.6	0.4	0.4	1.6
0.055	0.01	1	76	99	99	99	99	1.1	0.0	0.0	0.0
0.055	0.01	2	38	61	84	99	99	8.9	2.2	1.9	4.9
0.055	0.01	5	15	24	33	43	52	3.5	0.9	0.9	3.5

C. Results with $T = 804$ and $N = 25$

Table I reports optimal asset allocations in samples with $T = 804$ observations and $N = 25$ predictive variables. As noted in the introduction, although this number of predictive variables is large by many standards, we analyze this case in order to provide some perspective on the impact of potential data-mining concerns. Results are presented for two different outcomes for the unadjusted sample R^2, 0.025 and 0.055. Under standard assumptions for a regression model, these R^2 values produce p-values of 0.75 and 0.01. Each p-value is the probability of observing a sample with an R^2 greater than the

$N = 2$ and 46.2 with $N = 25$. The 99th percentiles of the simulated distributions in these two cases are 12.2 and 72.9.

given value, computed using the result that, under the standard regression-model assumptions,

$$F = \left(\frac{T - N - 1}{N}\right)\left(\frac{R^2}{1 - R^2}\right) \tag{43}$$

is distributed (central) F with N and $T - N - 1$ degrees of freedom under the null hypothesis that the true population R-squared is zero.[35]

It is useful to note that the p-values reported in Table I, merely transformations of the R^2 values, need not represent the correct tail probabilities under the null. In particular, the standard regression-model assumptions need not hold in the presence of lagged stochastic regressors, as entertained in our framework.[36] (Both the sign and the magnitude of the error in the standard p-value depend on the elements in the true B and Σ.) Therefore, we present the standard p-value simply as a widely employed summary measure.

We begin with $R^2 = 0.025$ because, as noted in the introduction, studies using data starting in 1927 typically report an R^2 of about that magnitude when regressing a monthly stock return on the lagged return and only a few other predictive variables. If the small set of predictive variables is obtained by "mining" a larger set of 25 variables, then the sample R^2 produced by the larger set must, by construction, be at least 0.025 and would most likely be considerably larger. (Recall that R^2, defined in (30), is the *unadjusted* R-squared.) Thus, we suggest that the asset allocations reported for $N = 25$ when $R^2 = 0.025$ provide a conservative "worst-case" characterization of the importance of the reported regression evidence to a Bayesian investor who includes all 25 variables in the regression model.

When $R^2 = 0.025$, the fitted regression values' sample standard deviation, $\sqrt{R^2}\hat{\sigma}_r$, is 89 basis points per month, so the one-step-ahead fitted regression prediction for r_{T+1} ranges from -40 basis points to 138 basis points as δ ranges from -1 to 1 (recall equation (42)). As reported in Table I, an investor with relative risk aversion (A) equal to 2 and diffuse prior beliefs would choose a stock allocation at the lower bound of 0 percent at $\delta = -1$, whereas that same investor would allocate 61 percent to stocks when $\delta = 0$. If that investor's prior beliefs are given instead by the no-predictability informative prior, those allocations are instead equal to 53 percent at $\delta = -1$ and 83 percent at $\delta = 0$. Under both sets of prior beliefs, the investor would allocate the upper bound of 99 percent to stocks when $\delta = 1$. With the informative prior, the percentage allocations change slightly less in absolute magnitude across values of δ, but the stock allocations are higher in general. The latter effect is due to the lower variance of the predictive distribution, σ_T^2, that occurs with the informative prior. In general, however, we see that even when the investor relies on a model containing 25 possible predictive variables and the sample R^2 of that

[35] See, for example, Judge et al. (1985).

[36] See Stambaugh (1986) and Nelson and Kim (1993).

regression is only 0.025, the optimal asset allocation depends strongly on the current values of the predictive variables.

The sensitivity of the optimal asset allocation to the current values of the predictive variables provides a metric by which to characterize the economic significance of the sample regression evidence. That is, the sample evidence is translated into implications about actions. Additional insight into the investor's perceived importance of these actions is provided by a related metric, introduced in Section I.*B*, that compares expected utilities associated with optimal and suboptimal allocations. Let ω^a denote the optimal allocation chosen by the investor when $\delta = 0$. When $\delta \neq 0$, we compare the investor's expected utility for the portfolio formed using the allocation ω^a, which is then suboptimal, to the investor's expected utility for the portfolio formed using the optimal allocation ω^*. The expected utilities for both portfolios are computed under the same predictive pdf, i.e., based on the same given value for δ, and each expected utility is converted to a certainty equivalent return (CER).

For each of the four nonzero values of δ, Table I reports the "comparison of certainty equivalents," which is the CER for the portfolio formed using the optimal ω^* minus the CER for the portfolio formed using ω^a. In the case where $R^2 = 0.025$, $A = 2$, and the prior is diffuse, for example, we see from Table I that, when $\delta = 1$, the optimal allocation of 99 percent to stocks produces a CER that is 24.1 basis points (per month) higher than an allocation of 61 percent to stocks, the optimal allocation for $\delta = 0$. With the informative no-predictability prior, again when $\delta = 1$, the CER for the optimal 99 percent stock allocation is 3.0 basis points higher than the CER for an 83 percent allocation, the optimal allocation for $\delta = 0$.

In addition to the cases discussed above, in which $A = 2$ and $R^2 = 0.025$, Table I also reports results for investors with other coefficients of relative risk aversion ($A = 1$ and $A = 5$) and a regression with an R^2 value equal to 0.055. Higher risk aversion is associated with lower stock allocations, less sensitivity of the optimal allocations to the current value of the predictive variables, and smaller values for the comparisons of certainty equivalents. The larger R^2 value produces a p-value of about 0.01, which, in terms of standard statistical characterizations, provides a contrast to the p-value of 0.75 when $R^2 = 0.025$. When judged in terms of economic significance, however, the contrast between the two regression outcomes is not as sharp. In both cases, the optimal asset allocation is sensitive to the fitted regression values. A more extensive comparison of the economic and statistical significance of the regression evidence is presented in the next subsection.

As discussed earlier, data mining is probably the principal motivation for considering cases with large numbers of predictive variables. To consider only such cases, however, would almost surely assign the data-mining issue undue weight in our overall investigation of the role of sample regression evidence in asset allocation. Far from clear is the extent to which published regression results reflect the outcome of data mining, as either an intended perpetration or, as more commonly suggested, an unintended outcome of conducting research with some knowledge of the past successes and failures of other studies.

Indeed, some of the early studies offer theoretical motivations, such as the argument by Keim and Stambaugh (1986) that expected future returns should be positively related to current values of variables that move inversely with levels of asset prices. In order to analyze more thoroughly the potential role of regression evidence in the asset-allocation decision, the next subsection considers wide ranges for both the sample size (T) and the number of predictive variables (N).

D. Economic Significance and Regression Statistics

We select a set of (T, N) combinations designed to include both small and large values for each quantity. Specifically, we let $T = 7$, 60, and 804, $N = 2$, 10, 25, and 50, and we consider all seven of the (T, N) combinations in which $T > 2N$.[37] In the VAR framework, the lagged return always appears as one of the predictive variables, so when $N = 2$ the predictive variables include the lagged return and one additional variable. We include a sample size of $T = 7$ when $N = 2$ because it is the smallest sample for which the predictive variance σ_T^2 in (34) exists when the prior is diffuse. A five-year sample of size $T = 60$ would no doubt be considered quite small for an empirical study of stock-return predictability, but we find that a sample of that size can still provide some interesting contrasts between economic significance and various statistical measures. As noted earlier, $T = 804$ is the number of months in the period from January 1927 through December 1993.

In the analysis here, rather than first specifying the statistical measures summarizing a regression and then examining the implications of that regression evidence for asset allocation, we proceed in the opposite direction. Since our ultimate interest centers on the economic significance of the sample evidence, as characterized by the implications of that evidence for the asset-allocation decision, we take a given degree of economic significance as a starting point and then, for each (T, N) combination, we derive the statistical measures for a sample that would produce that degree of economic significance.

The degree of economic significance is specified in terms of the sensitivity of the optimal asset allocation ω^* to the current values of the predictive variables. Specifically, we return to the approximation in (40), which can be rewritten, using (37), (38), and (42), as

$$\hat{\omega}^* = \left[\frac{\bar{r} + \frac{1}{2}\sigma_T^2}{A\sigma_T^2}\right] + \gamma\delta, \tag{44}$$

[37] This condition is imposed so that, when the prior is diffuse, the number of degrees of freedom in the Student t predictive pdf is greater than zero (equation (36)).

where

$$\gamma = \left(\frac{\left(T/T^*\right)\sqrt{R^2}}{1 - \left(T/T^*\right)^2 R^2} \right) \left(1 - \frac{2(N+1)}{T^*} \right) \left(A\hat{\sigma}_r \left[1 + \frac{1}{T^*}(1+q) \right] \right)^{-1}. \quad (45)$$

As defined earlier, $T^* = T + T_0$, where $T_0 = 75 \cdot (N + 1)$ with the informative no-predictability prior and $T_0 = 0$ with the diffuse prior. The first term on the right-hand side of (44) does not depend on the current values of the predictive variables, y_T.[38] Those values enter the second term, where γ, which we interpret as the degree of economic significance, measures the sensitivity of $\hat{\omega}^*$ to the current values of the predictive variables, summarized by δ as in (42). Specifically, γ gives the (approximate) difference in optimal allocations when the one-step-ahead fitted regression predictions differ by one sample standard deviation of the fitted predictions (except when ω is constrained by the [0, 0.99] bounds).

Once T and N are specified, then R^2 is the only unknown quantity on the right-hand side of (45). We specify a value for γ and then solve (45) for R^2. This R^2 value can be used, along with T and N, to compute an array of additional statistics that might be used either in hypothesis testing and model selection or in more general descriptions of the strength of the sample regression evidence. From this array we simply choose two that, in some sense, illustrate the diversity of such statistics. The first is the p-value for the F statistic in (43), the same statistic reported in Table I. The second statistic is based on the Schwarz criterion. Schwarz (1978) develops this model-selection criterion in a large-sample setting, but, as shown by Klein and Brown (1984), the Schwarz criterion can also be used for model-selection in a finite-sample Bayesian setting. Brown and Klein derive a posterior odds ratio for the comparison of two models when the prior for the parameters in each model is intended to be noninformative.[39] We use their result to construct the odds ratio that compares the given regression model to a model in which returns are assumed to be i.i.d. Specifically,

$$\ln(O^*) = -\frac{1}{2}\left[T \ln(1 - R^2) - N \ln(T)\right], \quad (46)$$

where O^* gives the odds in favor of the regression model. That is, the odds ratio favors the regression model if $\ln(O^*) > 0$. The asset-allocation decision we analyze does not involve model selection, as explained earlier, but we report

[38] Even though y_T appears in q in equation (35), recall from the previous discussion that we specify the same (high) value for q across different values of y_T.

[39] The model assumptions and noninformative prior specifications differ from those in the Bayesian framework used here.

$\ln(O^*)$ here simply to broaden our choices of statistics used in standard analyses of regression evidence.[40]

Table II reports results for γ specified alternately as 20, 30, and 40 percent, and where the investor has diffuse prior beliefs and relative risk aversion $A = 2$. As explained above, the value of γ, combined with T and N (columns 1 and 2), gives the R^2 (column 3) as well as the p-value and $\ln(O^*)$ (columns 4 and 5). The R^2 is then used to compute the reported allocations and certainty-equivalent comparisons by the numerical methods described earlier, using the Student t predictive pdf. The approximation in (40) is sufficiently accurate so that, in most cases, the differences in allocations corresponding to unit differences in δ are quite close to γ (except, of course, when the [0, 0.99] range for ω is binding).

Given the above discussion, the three values T, N, and R^2 jointly determine the statistical measures as well as the asset-allocation results reported in a given row of Table II. This functional dependence essentially implies that either the p-value or $\ln(O^*)$ could, in principle, be used along with T and N to compute the asset-allocation results. As we see, though, this mapping between the statistical measures and the asset-allocation results produces no simple pattern in Table II. Most of the p-values are large—generally greater than 0.5 and often nearly 1.0. Small p-values occur only in the case where $T = 60$ and $N = 25$. All of the values of $\ln(O^*)$ are negative, indicating that the odds ratio favors the i.i.d. model over the regression model, and those values tend to become more extreme as N increases relative to T.

It is interesting to compare the results when $T = 7$ to those of the simple two-state, two-outcome example in the previous section, where the number of time-series observations is also small (sixteen). The specifications of the models are quite different, but in both cases a sample that produces a large p-value (or odds favoring the i.i.d. model) contains sufficient evidence of predictability to exert a substantial influence on the asset-allocation decision.

Also reported in Table II are results for $T = \infty$. Recall from the discussion following equation (34) that, for any finite N, the predictive variance is then equal to the limiting value $(1 - R^2)\hat{\sigma}^2$. In this case, where there exists no estimation risk, values of γ between 20 and 40 percent are obtained with R^2 values that are small, ranging from 0.001 to 0.002. The results in this infinite-sample case are similar to those obtained when $T = 804$ and $N = 2$. In other words, when the number of predictive variables is small, estimation risk becomes relatively unimportant for a sample containing 804 observations. Note that estimation risk still plays a nontrivial role when $N = 2$ and $T = 60$. The importance of estimation risk in this case is attributable primarily to our conservative approach in specifying a large value for q in (34). If q is instead set to δ^2, its minimum value, then the results for $T = 60$ and $N = 2$ are also close

[40] Statistics that can also be computed from T, N, and R^2 include the adjusted R-squared, $\bar{R}^2 = R^2 - [N/(T - N - 1)]/(T - N - 1)](1 - R^2)$, and statistics that compare the regression model to the i.i.d. model using various other model-selection criteria. See, for example, Sawa (1978) and Amemiya (1980) for discussions of such criteria.

Table II

Samples Providing Given Degrees of Economic Significance for an Investor with Diffuse Prior Beliefs and Relative Risk Aversion Equal to 2

Optimal stock allocations and certainty equivalents are computed with respect to a Bayesian investor's predictive pdf based on regression evidence with an (unadjusted) sample R-squared equal to R^2, T observations, and N predictive variables. The fitted one-month-ahead regression prediction of the excess stock return, r_{T+1}, is δ sample standard deviations of the fitted values away from the sample average excess return $\bar{r}$. The value γ, defined in equations (44) and (45) and interpreted as the degree of economic significance, is the (approximate) difference in optimal allocations corresponding to a unit difference in δ. For each T and N, we specify γ, which then implies the remaining values in the table. The "Comparison of Certainty Equivalents" gives the difference in certainty equivalent monthly returns between the optimal allocation and the allocation that would have been chosen for $\delta = 0$. The p-value is computed for the hypothesis that the regression's slope coefficients are jointly equal to zero, and O^* is an odds ratio, defined in equation (46) in the text, that compares the regression model to a model with no predictability.

					Optimal Stock Allocation (percent) for δ Equal to					Comparison of Certainty Equivalents (basis pts.) for δ Equal to			
T	N	R^2	p-value	ln O^*	-1	-0.5	0	0.5	1	-1	-0.5	0.5	1.0
					$\gamma = 20$ percent								
7	2	0.063	0.877	−1.72	10	21	33	44	56	16.3	4.2	4.1	16.6
60	2	0.001	0.968	−4.06	57	67	77	87	97	1.9	0.5	0.5	1.9
60	10	0.033	0.998	−19.45	14	24	35	45	55	10.2	2.6	2.6	10.3
60	25	0.618	0.016	−22.29	7	17	27	37	47	42.3	10.7	11.2	45.2
804	2	0.001	0.804	−6.47	80	90	99	99	99	1.1	0.2	0.0	0.0
804	10	0.001	1.000	−33.02	58	68	78	88	98	1.8	0.5	0.5	1.8
804	25	0.002	1.000	−82.62	40	50	60	70	80	2.8	0.7	0.7	2.8
804	50	0.007	1.000	−164.39	26	36	46	56	66	4.7	1.2	1.2	4.7
∞	—	0.001	—	—	83	93	99	99	99	0.8	0.1	0.0	0.0
					$\gamma = 30$ percent								
7	2	0.125	0.766	−1.48	2	16	33	51	67	32.9	8.6	8.7	34.3
60	2	0.003	0.930	−4.02	47	62	77	92	99	4.2	1.1	1.1	3.9
60	10	0.070	0.955	−18.31	5	20	35	50	65	22.0	5.5	5.5	22.3
60	25	0.725	0.000	−12.50	0	13	28	43	59	68.4	17.4	18.3	73.6
804	2	0.001	0.613	−6.20	70	85	99	99	99	2.7	0.6	0.0	0.0
804	10	0.002	0.997	−32.48	49	64	79	94	99	4.1	1.0	1.0	3.7
804	25	0.006	1.000	−81.38	30	45	60	75	90	6.3	1.6	1.6	6.3
804	50	0.016	1.000	−160.90	16	31	46	61	76	10.5	2.6	2.6	10.6
∞	—	0.001	—	—	73	88	99	99	99	2.1	0.4	0.0	0.0
					$\gamma = 40$ percent								
7	2	0.190	0.657	−1.21	0	11	34	57	76	50.4	14.0	14.2	55.0
60	2	0.004	0.880	−3.96	37	57	77	98	99	7.5	1.9	1.9	5.9
60	10	0.113	0.787	−16.89	0	15	35	55	76	36.8	9.4	9.4	37.9
60	25	0.785	0.000	−5.07	0	9	29	49	70	88.9	24.6	25.0	101.3
804	2	0.002	0.420	−5.82	60	80	99	99	99	4.9	1.1	0.0	0.0
804	10	0.004	0.970	−31.73	39	59	79	99	99	7.3	1.8	1.8	5.6
804	25	0.010	1.000	−79.66	20	40	60	80	99	11.1	2.8	2.8	11.1
804	50	0.027	1.000	−156.17	6	26	46	66	86	18.4	4.6	4.6	18.6
∞	—	0.002	—	—	63	83	99	99	99	4.0	0.8	0.0	0.0

to the infinite-sample results. The possibility that estimation risk can become unimportant with such a modest sample size also appears in earlier results in an i.i.d. setting. Although Brown (1979) characterizes his analysis for the i.i.d. model as demonstrating a "measurable difference" between the optimal asset allocations in finite and infinite samples, it can also be seen from his reported results that, in samples of only 50 observations, the optimal stock allocation is 94 percent or more of the allocation chosen in an infinite sample.[41]

Table III reports the same analysis as in Table II, except that the informative no-predictability prior replaces the diffuse prior. The table omits some of the cases with $T = 7$ and all of the cases with $T = 60$ when $N = 25$. In the omitted cases, the informative prior precludes those effectively small samples from producing the given degree of economic significance (γ), even with an outcome of $R^2 = 1$. With the larger values of T, however, it is still the case that the given degree of economic significance is often accompanied by large p-values and large negative values of $\ln(O^*)$. Note also that, as γ increases, the p-value can decrease considerably. For example, in the case of ($T = 804$, $N = 50$), the p-value is 0.997 for $\gamma = 20$ percent but 0.000 for $\gamma = 40$ percent. Similarly, when $T = 60$ and $N = 10$, the p-value drops from 0.337 to 0.001 as γ increases from 20 to 30 percent. As in Table II, however, there appears to be no simple correspondence between the statistical measures and the asset-allocation results.

III. Conclusions

We view this study as an initial attempt to assess the economic significance of empirical evidence about stock-return predictability by examining an investor's conditional Bayesian portfolio decision. The specific choices we make here in implementing this general approach, such as the forms of the prior distribution and the likelihood function, are dictated in large part by tractability. Extending the analysis to include richer specifications would be worthwhile, although probably not without computational challenges.

The research conducted here can also be extended along a number of other dimensions. We confine our attention to a single stock portfolio, but additional risky assets could be introduced into the allocation decision as well.[42] A recent study by Lo and MacKinlay (1995) uses a cross-section of assets to construct a portfolio that is "maximally predictable" by a given set of predictive variables, and it would be interesting to investigate the desirability of such a portfolio from the perspective of a Bayesian investor.

The regression framework we employ assumes that the regression disturbances are homoskedastic. A number of studies conclude, however, that stock returns exhibit conditional heteroskedasticity (e.g., French, Schwert, and

[41] See Brown (1976, p. 114 and table 8.1).

[42] 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.

Table III

Samples Providing Given Degrees of Economic Significance for an Investor with Informative No-Predictability Prior Beliefs and Relative Risk Aversion Equal to 2

Optimal stock allocations and certainty equivalents are computed with respect to a Bayesian investor's predictive pdf based on regression evidence with an (unadjusted) sample R-squared equal to R^2, T observations, and N predictive variables. The fitted one-month-ahead regression prediction of the excess stock return, r_{T+1}, is δ sample standard deviations of the fitted values away from the sample average excess return $\bar{r}$. The value γ, defined in equations (44) and (45) and interpreted as the degree of economic significance, is the (approximate) difference in optimal allocations corresponding to a unit difference in δ. For each T and N, we specify γ, which then implies the remaining values in the table. The "Comparison of Certainty Equivalents" gives the difference in certainty equivalent monthly returns between the optimal allocation and the allocation that would have been chosen for $\delta = 0$. The p-value is computed for the hypothesis that the regression's slope coefficients are jointly equal to zero, and O^* is an odds ratio, defined in equation (46) in the text, that compares the regression model to a model with no predictability. The no-predictability informative prior is equivalent to a posterior that combines diffuse prior beliefs with a hypothetical T_0-month sample in which all estimated slopes are exactly zero. T_0 is specified such that the hypothetical sample contains 50 data entries per parameter, which, for N = 2, 10, 25, and 50, gives values for T_0 of 225, 825, 1950, and 3825.

					Optimal Stock Allocation (percent) for δ Equal to					Comparison of Certainty Equivalents (basis pts.) for δ Equal to			
T	N	R^2	p-value	ln O^*	−1	−0.5	0	0.5	1	−1	−0.5	0.5	1.0
					γ = 20 percent								
7	2	0.605	0.156	1.31	80	90	99	99	99	1.2	0.3	0.0	0.0
60	2	0.014	0.677	−3.68	76	86	96	99	99	1.3	0.3	0.2	0.4
60	10	0.192	0.337	−14.07	64	74	84	94	99	1.7	0.4	0.4	1.5
804	2	0.001	0.705	−6.34	81	91	99	99	99	1.1	0.2	0.0	0.0
804	10	0.003	0.991	−32.20	69	79	89	99	99	1.5	0.4	0.4	1.2
804	25	0.011	0.999	−79.34	63	73	83	93	99	1.7	0.4	0.4	1.6
804	50	0.034	0.997	−153.34	60	70	80	90	99	1.8	0.4	0.4	1.8
∞	—	0.001	—	—	83	93	99	99	99	0.8	0.1	0.0	0.0
					γ = 30 percent								
60	2	0.031	0.413	−3.16	66	81	96	99	99	3.0	0.7	0.3	0.6
60	10	0.431	0.001	−3.55	54	69	84	99	99	3.8	0.9	0.9	2.8
804	2	0.002	0.455	−5.90	71	86	99	99	99	2.5	0.5	0.0	0.0
804	10	0.007	0.852	−30.65	59	74	89	99	99	3.5	0.9	0.8	1.9
804	25	0.024	0.797	−73.94	53	68	83	98	99	3.8	0.9	1.0	2.9
804	50	0.076	0.125	−135.35	50	65	80	95	99	4.0	1.0	1.0	3.5
∞	—	0.001	—	—	73	88	99	99	99	2.1	0.4	0.0	0.0
					γ = 40 percent								
60	2	0.054	0.204	−2.42	56	76	96	99	99	5.3	1.3	0.4	0.8
60	10	0.764	0.000	22.86	44	64	84	99	99	6.7	1.7	1.6	4.0
804	2	0.003	0.247	−5.29	61	81	99	99	99	4.6	1.0	0.0	0.0
804	10	0.012	0.453	−28.47	49	69	89	99	99	6.1	1.5	1.2	2.7
804	25	0.042	0.108	−66.30	43	63	83	99	99	6.7	1.7	1.6	4.3
804	50	0.135	0.000	−108.89	40	60	80	99	99	7.2	1.8	1.8	5.2
∞	—	0.002	—	—	63	83	99	99	99	4.0	0.8	0.0	0.0

Stambaugh (1987)). An interesting extension of our framework would be to analyze the asset-allocation problem when conditional heteroskedasticity is included in the regression model. Given the importance of the variance in the asset-allocation decision, incorporating conditional heteroskedasticity into the model could change the manner by which the current values of the predictive variables influence the optimal allocation. For example, if both the conditional mean and the conditional variance of the excess stock return are increasing in the same predictive variable, then the influence of that predictive variable on the optimal allocation may be less than when homoskedasticity is assumed. Of course, incorporating conditional heteroskedasticity into the model could also change the manner by which the current values of the predictive variables influence the conditional mean and other properties of the predictive pdf.

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. We know, however, that in regressions of stock returns on dividend yields and other predictive variables, 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 the role of the investment horizon in a buy-and-hold asset-allocation problem. Such an investigation might reveal whether the differences in R-squared values between short and long horizons are economically meaningful.

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. With logarithmic utility, one of the preference specifications we consider, the solution to the one-month problem is still correct in such a setting. With other specifications of preferences, however, the problem becomes more complicated. 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.[43] 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, transac-

[43] 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). While completing the current revision, we received a recent working paper by Barberis (1995), who extends the diffuse-prior VAR model in Section II to investigate optimal asset allocations for multiperiod investment horizons.

tion costs would present an additional challenge to any modeling effort with dynamic rebalancing.

Appendix

A. The Predictive Pdf

Although the VAR model in (21) employs assumptions different from those in the traditional multivariate regression model (MVRM), the likelihood function in (24) is identical to that obtained in the MVRM. Hence, we can simply follow the analysis of the MVRM provided by Zellner (1971, pp. 233–236), who develops the predictive pdf using the same diffuse prior as in (31). As Zellner shows, the predictive pdf for y_{T+1} is in the multivariate Student t form,

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

where

$$x'_T = (1 \; y'_T), \tag{A.2}$$

$$\hat{B} = (X'X)^{-1}X'Y, \tag{A.3}$$

$$S = (Y - X\hat{B})'(Y - X\hat{B}), \tag{A.4}$$

and

$$g = 1 - x'_T(X'X + x_T x'_T)^{-1}x_T. \tag{A.5}$$

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

$$p(y_{T+1}|\Phi_T) = \frac{\nu^{(\nu/2)}\Gamma[(\nu + N)/2]|G|^{(1/2)}}{(\Gamma(1/2))^N\Gamma(\nu/2)} \cdot [\nu + (y'_{T+1} - x'_T\hat{B})G(y'_{T+1} - x'_T\hat{B})']^{-(\nu+N)/2}, \tag{A.6}$$

where $G = g\nu S^{-1}$ and $\nu = T - 2N$. The predictive distribution of r_{T+1} (the first element of y_{T+1}) is a univariate Student t (US t) pdf:[44]

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma(1/2)\Gamma(\nu/2)}\left(\frac{g}{S_{11}}\right)^{1/2}\left[1 + \frac{g}{S_{11}}(r_{T+1} - x'_T\hat{b})^2\right]^{-(\nu+1)/2}, \tag{A.7}$$

where S_{11}, the (1, 1) element of S, is given by

$$S_{11} = (y - X\hat{b})'(y - X\hat{b}). \tag{A.8}$$

[44] See Zellner (1971, p. 387).

Let

$$h = \frac{g\nu}{S_{11}}, \tag{A.9}$$

and substitute (A.9) into (A.7) to get the form of the US t pdf as in Zellner (1971, page 366),

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)}\left(\frac{h}{\nu}\right)^{1/2}\left[1 + \frac{h}{\nu}(r_{T+1} - \mu_T)^2\right]^{-(\nu+1)/2}. \tag{A.10}$$

where

$$\mu_T = E\{r_{T+1}|\Phi_T\} = x'_T\hat{b}. \tag{A.11}$$

Substituting (A.2) and (29) into (A.11) yields (33). The second moment about the mean of the US t pdf is:

$$\sigma_T^2 = \text{var}\{r_{T+1}|\Phi_T\} = \frac{\nu}{(\nu-2)h}. \tag{A.12}$$

Using (A.12), we obtain

$$\frac{h}{\nu} = \frac{1}{(\nu-2)\sigma_T^2}. \tag{A.13}$$

Substituting (A.11) and (A.13) into (A.10) yields (32). To obtain (34) we first rewrite (A.12) using (A.9):

$$\sigma_T^2 = \frac{S_{11}}{(\nu-2)g}. \tag{A.14}$$

Next, note that, since $\nu = T - 2N$,

$$\nu - 2 = T - 2(N+1). \tag{A.15}$$

Using (A.8) and the definition of R^2 in (30), we get

$$S_{11} = T(1-R^2)\hat{\sigma}_r^2. \tag{A.16}$$

To simplify the expression for g in (A.5), observe that[45]

$$g = 1 - x'_T(X'X + x_T x'_T)^{-1}x_T = \frac{1}{1 + x'_T(X'X)^{-1}x_T}. \tag{A.17}$$

[45] See, for example, Zellner (1971, p. 73).

Using (21), we get

$$\left(\frac{1}{T}\right)X'X = \begin{bmatrix} 1 & \bar{y}' \\ \bar{y} & (\bar{y}\bar{y} + \hat{V}_y) \end{bmatrix}. \tag{A.18}$$

Inverting (A.18) yields

$$(X'X)^{-1} = \frac{1}{T}\begin{bmatrix} 1 + (\bar{y}'\hat{V}_y^{-1}\bar{y}) & -\bar{y}'\hat{V}_y^{-1} \\ -\hat{V}_y^{-1}\bar{y} & \hat{V}_y^{-1} \end{bmatrix}. \tag{A.19}$$

Using (A.17), (A.19), and (35), we obtain

$$\frac{1}{g} = 1 + x_T'(X'X)^{-1}x_T = 1 + \frac{1}{T}(1 + (y_T - \bar{y})'\hat{V}_y^{-1}(y_T - \bar{y})) = 1 + \frac{1}{T}(1 + q). \tag{A.20}$$

Substituting (A.15), (A.16), and (A.20) into (A.14) yields (34).

B. Informative Priors

The informative no-predictability prior is equivalent to the posterior distribution obtained by combining the diffuse prior in (31) with a sample of size T_0 in which a regression of the excess stock return on the N predictive variables produces an R-squared of exactly zero. We wish to specify T_0 such that prior beliefs about stock-return predictability are similar across various specifications of N. In summarizing those prior beliefs, we use a measure that offers simple intuitive appeal. From the joint prior distribution for the parameter matrices B and Σ, we compute a prior distribution for the "true" or population R-squared value,

$$\tilde{R}^2 = 1 - \frac{\sigma_\epsilon^2}{\sigma_r^2}, \tag{B.1}$$

where σ_ϵ^2, defined previously, is the (1,1) element of Σ, and σ_r^2 is the unconditional variance of the excess stock return r_t. (Recall from equation (30) that R^2 denotes the regression outcome in the actual sample.) For some values of B and Σ, the unconditional variance σ_r^2 does not exist, so $\tilde{R}^2$ is then undefined. Specifically, if we partition B as

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \tag{B.2}$$

where B_1 is $1 \times N$ and B_2 is $N \times N$, then a necessary and sufficient condition for the existence of the unconditional variance-covariance matrix

$$V_y \equiv \text{cov}\{y_t, y_t'\} \tag{B.3}$$

is that the eigenvalues of B_2 lie inside the unit circle. In that case, V_y is the solution to

$$V_y = B_2 V_y B_2' + \Sigma, \tag{B.4}$$

and σ_r^2 is the (1, 1) element of V_y.[46]

We examine the prior distribution of $\tilde{R}^2$ conditional on the existence of V_y. The general approach is as follows. The joint prior distribution for B_2 and Σ can be obtained analytically, as discussed below, and we draw values of B_2 and Σ repeatedly from that distribution. For each draw satisfying the condition that the eigenvalues of B_2 lie inside the unit circle, we compute $\tilde{R}^2$ using (B.4) and then (B.1). The frequency distribution of these $\tilde{R}^2$ values provides an approximation to the prior distribution. The prior probability that V_y does not exist can also be estimated in this process, and we report below the results of that calculation as well.

The informative prior for B_2 and Σ is determined by the following statistics for the hypothetical prior sample of size T_0: $\hat{B}_{2,0}$, the $N \times N$ matrix of OLS regression slopes for the VAR in (21), $\hat{\Sigma}_0$, the variance-covariance matrix of the residuals of that VAR, and $\hat{V}_{y,0}$, the variance-covariance matrix of y_t based on observations 0 through $T_0 - 1$. Given these statistics, the prior for B_2, conditional on Σ, is given by a multivariate Normal distribution,

$$p(\text{vec}(B_2)|\Sigma) \sim N\left(\text{vec}(\hat{B}_{2,0}), \frac{1}{T_0} \hat{V}_{y,0}^{-1} \otimes \Sigma\right), \tag{B.5}$$

where vec( ) denotes the vector formed by stacking the successive transposed rows of the matrix, and the marginal prior for Σ is an inverted Wishart,

$$p(\Sigma) \sim IW(T_0 \hat{\Sigma}_0, T_0 - N - 1). \tag{B.6}$$

The informative no-predictability prior specifies that, in the hypothetical prior sample, a regression of the excess stock return on the predictive variables produces an R-squared of exactly zero. That is, all N elements in the first row of $\hat{B}_{2,0}$ are set to zero. The predictive pdf for r_{T+1} does not depend on the other $N - 1$ rows of $\hat{B}_{2,0}$, but the latter values do enter the prior distribution for B_2 and Σ. As explained in Section II, the (1, 1) element of $\hat{\Sigma}_0$ is set equal to $\hat{\sigma}_r^2$, the actual sample variance of excess stock returns. The other elements of $\hat{\Sigma}_0$ do not enter the predictive pdf for r_{T+1}, but they do enter the prior for B_2 and Σ. We assume, for simplicity, that

$$\hat{V}_{y,0} = \hat{B}_{2,0} \hat{V}_{y,0} \hat{B}_{2,0}' + \hat{\Sigma}_0, \tag{B.7}$$

which will obtain, for example, if observation 0 is identical to observation T. Recall that the predictive pdf for r_{T+1} is invariant under linear transformations of y_t that preserve r_t as the first element. It can also be verified, given the

[46] See Wei (1990, pp. 339–341).

joint prior distribution for B_2 and Σ in (B.5) and (B.6), that the marginal prior for $\tilde{R}^2$ is also invariant under such transformations. Therefore, without loss of generality, we set

$$\hat{V}_{y,0} = \hat{\sigma}_r^2 I_N, \tag{B.8}$$

where I_N is the $N \times N$ identity matrix. For this set of orthogonalized variables, we specify

$$\hat{B}_{2,0} = \begin{bmatrix} 0 & 0 \\ 0 & \rho I_{N-1} \end{bmatrix}. \tag{B.9}$$

In other words, returns are totally unpredictable in the hypothetical sample, each of the other N–1 variables has sample autocorrelation equal to ρ, and all sample cross autocorrelations are equal to zero. Combining (B.7), (B.8), and (B.9) then gives

$$\hat{\Sigma}_0 = \hat{\sigma}_r^2 \begin{bmatrix} 1 & 0 \\ 0 & (1 - \rho^2) I_{N-1} \end{bmatrix}. \tag{B.10}$$

Although the predictive pdf for r_{T+1} depends on $\hat{\sigma}_r^2$, the prior for $\tilde{R}^2$ does not. Therefore, the simplifications in (B.7) through (B.10) allow the marginal prior distribution for $\tilde{R}^2$, given T_0 and N, to be specified completely by the scalar value ρ. We wish to specify T_0 such that this prior distribution is similar across different values of N. It appears that such an objective is not accomplished satisfactorily by holding T_0 fixed. For example, suppose that, for all N, we set $T_0 = 1200$. That is, the prior is then equivalent to 100 years of hypothetical data in which a regression of returns on the predictive variables produces a sample R-squared of exactly zero. With $T_0 = 1200$, we draw repeatedly from the priors for B_2 and Σ in (B.5) and (B.6) and tabulate the frequency distribution of the resulting values of $\tilde{R}^2$. This procedure is performed with four different values of N (2, 10, 25, and 50) and with four different values of ρ (0, 0.5, 0.8, and 0.95). In each case, the draws of (B_2, Σ) continue until there are 2000 draws in which the eigenvalues of B_2 lie inside the unit circle, so that the estimated prior for $\tilde{R}^2$, conditional on that value being defined, is based on a frequency distribution of 2000 values. Figure 2 displays those marginal priors for $\tilde{R}^2$. The case in which $N = 50$ and $\rho = 0.95$ is omitted; this case proved computationally infeasible, due to the apparently high frequency of draws in which V_y does not exist as well as the difficulty in solving (B.4) for many draws in which V_y does exist. For each value of ρ, observe that the prior assigns greater mass to higher values of $\tilde{R}^2$ as N increases. For example, with $\rho = 0.8$, virtually all of the prior mass lies *below* $\tilde{R}^2 = 0.01$ with $N = 2$, whereas with $N = 25$ or $N = 50$, virtually all of the prior mass lies *above* $\tilde{R}^2 = 0.01$.

The results in Figure 2 suggest that, in order to obtain similar priors across different values of N, T_0 must increase with N. With T_0 observations on each of N variables, the hypothetical sample contains $N \cdot T_0$ data entries. We follow the simple approach of specifying T_0 so as to hold constant the number of data

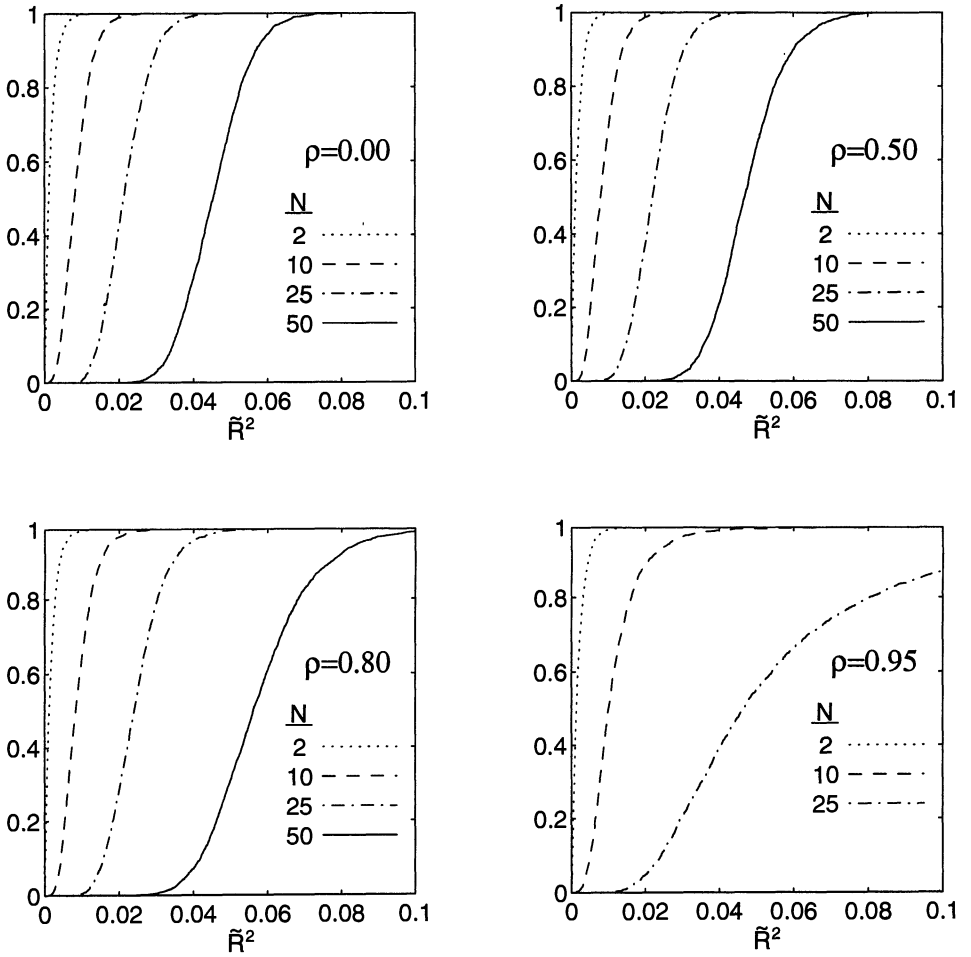

Figure 2. Prior distributions for $\tilde{R}^2$ with T_0 constant across N. Each graph plots, for a given value of ρ, the cumulative prior distributions of $\tilde{R}^2$ obtained for different values of N, the number of predictive variables. The size of the hypothetical prior sample, T_0, is equal to 1200. The parameter $\tilde{R}^2$ is the true R-squared in a regression of the excess stock return on the predictive variables, and ρ is the sample autocorrelation of each of the $N - 1$ predictive variables (excluding the return) as specified in the hypothetical prior sample.

entries per parameter. There are $\frac{1}{2}N(N + 1)$ distinct parameters in B and Σ, so a sample of size T_0 provides $\frac{2}{3}T_0/(N + 1)$ data entries per parameter. We set $T_0 = 75(N + 1)$, which gives 50 data entries per parameter. This choice produces a prior for $\tilde{R}^2$ that lies between those for $N = 10$ and $N = 25$ in Figure 2. (Specifically, with $T_0 = 75(N + 1)$, a value of $T_0 = 1200$ implies $N = 15$.) Figure 3 displays the marginal priors for $\tilde{R}^2$ for the same combinations of N and ρ used in Figure 2, and the case in which $N = 50$ and $\rho = 0.95$ is now included as well. In Figure 3, as well as in Figure 2, it can be seen that the priors assign more mass to higher values of $\tilde{R}^2$ as ρ increases, although this

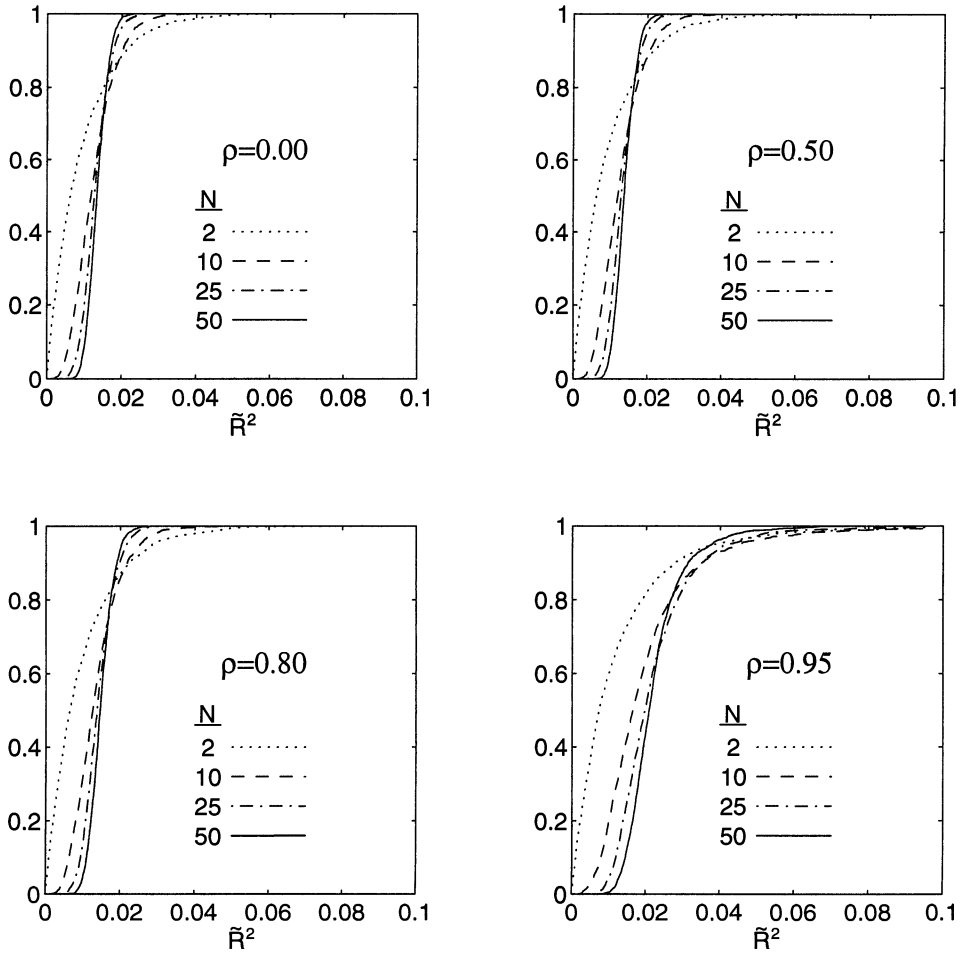

Figure 3. Prior distributions for $\tilde{R}^2$ with T_0 increasing in N. Each graph plots, for a given value of ρ, the cumulative prior distributions of $\tilde{R}^2$ obtained for different values of N, the number of predictive variables. The size of the hypothetical prior sample, T_0, is equal to $75(N + 1)$, which is equivalent to 50 data entries per parameter. The parameter $\tilde{R}^2$ is the true R-squared in a regression of the excess stock return on the predictive variables, and ρ is the sample autocorrelation of each of the $N - 1$ predictive variables (excluding the return) as specified in the hypothetical prior sample.

effect is most evident in moving from $\rho = 0.8$ to $\rho = 0.95$. The priors for $\tilde{R}^2$ in Figure 3, however, exhibit much less variation across N than do the priors in Figure 2. In fact, it appears that the objective of increasing T_0 with N so as to obtain similar priors across different values of N is accomplished reasonably well by holding constant the number of data entries per parameter.

As explained earlier, in constructing each frequency distribution of $\tilde{R}^2$, the total number of draws of (B_2, Σ) is $2000 + m$, where m is the number of draws in which the eigenvalues of B_2 do not lie inside the unit circle. Thus, the prior

Table IV

N	$T_0 = 1200$	$T_0 = 75 \cdot (N + 1)$
2	0	0.0074
10	0	0.0109
25	0.1127	0.0005
50	—	0

probability that V_y does not exist can be estimated as $m/(2000 + m)$. This estimated probability equals zero ($m = 0$) for all cases with ρ equal to 0, 0.5, and 0.8. For $\rho = 0.95$, this estimated probability is as in Table IV.

REFERENCES

Amemiya, Takeshi, 1980, Selection of regressors, *International Economic Review* 21, 331–353.

Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis* (John Wiley and Sons, New York).

Barberis, Nicholas, 1995, How big are hedging demands? Evidence from long-horizon asset allocation, Working paper, Harvard University.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).

Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* 44, 1177–1189.

Brennan, Michael J., Eduardo S. Schwartz, and Ron Lagnado, 1993, Strategic asset allocation, Working paper no. 11-93, UCLA.

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, in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, Eds: *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).

Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

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.

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.

Fisher, Ronald Aylmer, Sir, 1973, *Statistical Methods and Scientific Inference* (Hafner Press, New York).

Forsythe, George E., Michael A. Malcolm, and Cleve 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 and Duke University.

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, 1995, A longer look at dividend yields, *Journal of Business* 68, 483–508.

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 and University of California, Berkeley.

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

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making* (Rowman and Littlefield, Savage, Md.).

Jeffreys, Harold, Sir, 1961, *Theory of Probability* (Oxford, Clarendon).

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.

Judge, George E., W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee, 1985, *The Theory and Practice of Econometrics* (John Wiley and Sons, New York).

Kandel, Shmuel, and Robert F. Stambaugh, 1987, Long-horizon returns and short-horizon models, CRSP Working paper 222, University of Chicago.

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.

Kaul, Gautam, 1995, Predictable components in stock returns, Working paper, University of Michigan.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kendall, Maurice, Sir, and Alan Stuart, 1977, *The Advanced Theory of Statistics*, Vol. 1 (Macmillan, New York).

Kendall, Maurice, Sir, and Alan Stuart, 1979, *The Advanced Theory of Statistics*, Vol. 2 (Macmillan, New York).

Klein, Roger W., and Vijay S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Klein, Roger W., and Stephen J. Brown, 1984, Model selection when there is minimal prior information, *Econometrica* 52, 1291–1312.

Kothari, S. P., and Jay Shanken, 1995, Book-to-market, dividend yield, and expected market returns: A time-series analysis, Working paper, University of Rochester.

Lamoureux, Christopher G., and Goufu Zhou, 1995, Temporary components of stock returns: What do the data tell us? Working paper, University of Arizona and Washington University.

Lindley, D. V., 1957, A statistical paradox, *Biometrika* 44, 187–192.

Lo, Andrew W., and A. Craig MacKinlay, 1995, Maximizing predictability in the stock and bond markets, Working paper, MIT and University of Pennsylvania.

Maritz, J. S., and T. Lwin, 1989, *Empirical Bayes Methods* (Chapman and Hall, London).

McCulloch, Robert, and Peter E. Rossi, 1990, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, *Journal of Financial Economics* 28, 7–38.

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).

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Rozeff, Michael, 1984, Dividend yields are equity risk premiums, *Journal of Portfolio Management* 11, Fall, 68–75.

Sawa, Takamitsu, 1978, Information criteria for discriminating among alternative regression models, *Econometrica* 46, 1273–1291.

Schwarz, Gideon, 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

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

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 156, University of Chicago.

Wei, William S., 1990, *Time Series Analysis: Univariate and Multivariate Methods* (Addison-Wesley, Redwood City, Calif.).

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

Zellner, Arnold, and V. Karuppan Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* 60, 608–616.