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

by

Shmuel Kandel
Robert F. Stambaugh

27-94

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
Philadelphia, PA 19104-6367

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

by*

Shmuel Kandel
Recanati Graduate School of Business Administration
Tel-Aviv University
and
The Wharton School
University of Pennsylvania

and

Robert F. Stambaugh
The Wharton School
University of Pennsylvania
and
National Bureau of Economic Research

October 7, 1994
revised July 19, 1995

Abstract

Sample evidence about the predictability of monthly stock returns is considered from the perspective of an investor allocating funds between stocks and cash. A regression of stock returns on a set of predictive variables might seem weak when described by usual statistical measures, but such measures can fail to convey the economic significance of the sample evidence when it is used by a risk-averse Bayesian investor to update prior beliefs about the regression relation and to compute an optimal asset allocation. Even when those prior beliefs are weighted substantially against predictability, the current values of the predictive variables can exert a strong influence on the portfolio decision.

---

*The authors are grateful for comments by René Stulz, three anonymous referees, and workshop participants at Baruch College (CUNY), Northwestern University, Ohio State University, Rutgers University, 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.

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

Abstract

Sample evidence about the predictability of monthly stock returns is considered from the perspective of an investor allocating funds between stocks and cash. A regression of stock returns on a set of predictive variables might seem weak when described by usual statistical measures, but such measures can fail to convey the economic significance of the sample evidence when it is used by a risk-averse Bayesian investor to update prior beliefs about the regression relation and to compute an optimal asset allocation. Even when those prior beliefs are weighted substantially against predictability, the current values of the predictive variables can exert a strong influence on the portfolio decision.

1. Introduction

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. Suppose that this investor is given the results of running the following regression using monthly data from January 1927 through December 1993,

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

where r_t is the continuously compounded NYSE return in month t, in excess of the continuously compounded T-bill rate for that month, and x_{t-1} is a vector of "predictive" variables that are observed at the end of month $t-1$. The investor is provided with the OLS estimate $\hat{b}$, the regression's R^2, and any other desired statistics, including those often published in the academic literature. Now suppose that the investor is also given the most recent vector of the predictive variables, x_T, where December 1993 is denoted as month T. To what extent will the investor's asset allocation decision depend on x_T? The average excess return 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 -30 bp. If the investor would allocate about 75% to stocks if $x'_T\hat{b}$ were equal to the average excess return of 49 bp, how much less will the investor allocate to stocks when $x'_T\hat{b}$ is actually 79 bp lower than that long-run average? How much does the investor value the ability to allocate less than 75% 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. Our analytical framework, although simplified in a number of respects, proves tractable in addressing the questions posed above and, we suggest, illustrates the potential insights offered by this type of approach.

We find that characterizations of sample evidence based on standard statistical measures need not convey the economic significance of that evidence. Suppose, for example, that the unadjusted sample R-squared for the regression in (1) is equal to $R^2 = 0.02$, which is fairly typical of values reported in studies using monthly data beginning in 1927.[1] Based on the standard F statistic, an $R^2 = 0.02$ implies a small p-value for the hypothesis that the true slope coefficients are jointly zero, provided that the number of regressors is small, say three or four. Of course, if the small set of reported regressors is selected from a much larger number of regressors, say 100, then $R^2 = 0.02$ implies a larger p-value computed with appropriate recognition of the size of the larger set.[2] In the worst case, albeit unlikely, the R^2 in the regression on all 100 variables might still be only 0.02, in which case the p-value based on the standard F statistic with 100 regressors would be nearly 1. Even in that extreme scenario, however, the sample evidence can exert an influence on the investor's asset-allocation decision. In fact, returning to the questions posed in the numerical example above, we find that a log-utility investor with vague prior beliefs about the parameters in that 100-variable regression would, after updating those beliefs using the regression evidence, allocate only 35% to stocks with $x_T'\hat{b}$ equal to -30 bp, whereas the same investor would indeed allocate about 75% to stocks if $x_T'\hat{b}$ were instead equal to the long-run average of 49 bp. The ability to allocate 35% instead of 75% is worth about 15 bp to that investor, valued in terms of differences in a certainty-equivalent monthly return. If that investor's prior beliefs are, instead of being vague, weighted against predictability to a degree equivalent to having observed a century of prior data in which the sample R-squared is *exactly* zero, then that investor would still allocate about 30% less to stocks when the fitted value is -30 bp instead of the long-run average 49 bp, and the ability to do so would still be worth about 4 bp to the investor.

The paper 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 2, is organized around an example of the asset-allocation decision within a simple two-state, two-outcome setting. Section 3 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

[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]See Foster and Smith (1994).

features of the data, such as heteroskadasticity, 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 4 concludes the paper and suggests directions for future research.

2. Analyzing the Asset-Allocation Decision: General Approach

2.1. 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 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

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

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 solution for 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.[3] 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[4]

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

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 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.[5] 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.[6] 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.).[7] This study analyzes a portfolio decision where the investor instead uses a model in which returns can possess predictability.

[3]This conditional Bayesian decision approach is discussed further in subsection 2.3.. See also Berger (1985).

[4]See Zellner, 1971, p. 14.

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

[6]See Zellner (1971, pp. 320–327).

[7]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.

2.2. Economic Significance of Stock-Return Predictability: 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% or −40%. 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 $\theta = (\theta_1, \theta_2)$, 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%. 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

$$p(\theta_i) = \frac{[\theta_i(1-\theta_i)]^{a-1}}{B(a,a)}, \; i = 1, \; 2, \tag{8}$$

where $a \geq 1$ 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 a for the prior distribution in (8): $a = 1$, $a = 6$, and $a = 21$. When $a = 1$, the prior distribution for each of the θ_i's is the Bayes-Laplace uniform prior on $(0, 1)$. As a increases, the prior distribution becomes more concentrated around 0.5. The implied prior distributions of $(\theta_1 - \theta_2)$, for the three values of a, are numerically evaluated and displayed as dashed curves in Figure 1. The larger is a, 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:

$$\Phi_T = \{\Psi_T, s_T\}. \tag{9}$$

In our example, the sample Ψ_T includes $T = 16$ pairs with $T_i = 8$ months for each state, and this 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% stock return will be realized in M_i of those months is given by the binomial likelihood function,

$$p(M_i|\theta_i, T_i) = \begin{pmatrix} T_i \\ M_i \end{pmatrix} \theta_i^{M_i}(1-\theta_i)^{T_i - M_i}, \quad i = 1,\ 2. \tag{10}$$

Combining the prior distribution in (8) with the likelihood function in (10) yields the marginal posterior distribution for θ_i, a Beta distribution:[8]

$$p(\theta_i|\Phi_T) = p(\theta_i|M_i, T_i) = \frac{\theta_i^{M_i+a-1}(1-\theta_i)^{T_i-M_i+a-1}}{B((M_i+a),(T_i-M_i+a))} \quad ,i = 1,\ 2. \tag{11}$$

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 a, by the solid curves in Figure 1. All of these posterior distributions center at positive values, but, the larger is a, 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 $\hat{\theta}_j$ denotes the predictive probability that the return will be 40%. To obtain $\hat{\theta}_j$, first recall that

$$p(R_{T+1} = 40\%|\Phi_T = \{\Psi_T, s_T = j\}, \theta_j) = \theta_j, \tag{12}$$

and then substitute (12) into (7) to get

$$\begin{aligned} \hat{\theta}_j = p(R_{T+1} = 40\%|\Phi_T = \{\Psi_T, s_T = j\}) &= \int p(R_{T+1} = 40\%|\Phi_T, \theta_j) \cdot p(\theta_j|\Phi_T) d\theta_j \\ &= \int \theta_j \cdot p(\theta_j|\Phi_T) d\theta_j \\ &= \frac{(M_j + a)}{(T_j + 2a)}. \end{aligned} \tag{13}$$

[8]See Zellner, 1971, p. 39

With the predictive distribution in (13), the investor's optimization problem in (4) becomes

$$\max_{0 \leq \omega_j \leq 1} [\hat{\theta}_j \ln(1 + 0.4\omega_j) + (1 - \hat{\theta}_j) \ln(1 - 0.4\omega_j)], \tag{14}$$

and its solution is

$$\omega_j = \begin{cases} 0 & \text{if } (2\hat{\theta}_j - 1) \leq 0 \\ \left(\frac{2\hat{\theta}_j - 1}{0.4}\right) & \text{if } 0 < (2\hat{\theta}_j - 1) < 0.4 \\ 1 & \text{if } (2\hat{\theta}_j - 1) \geq 0.4 \, . \end{cases} \tag{15}$$

The number of sample months where the 40% stock return follows state 2 is equal to four, exactly half of the sample size for that state. Hence, the posterior distribution of θ_2 is centered around 0.5, and $\hat{\theta}_2 = 0.5$ for all values of a in the prior. Since the predictive distribution of the stock return in this case is symmetric around zero, any risk-averse investor will refrain from investing any money in stock when $s_T = 2$. It can be verified easily from (15) that $\omega_2 = 0$ at $\hat{\theta}_2 = 0.5$. If $s_T = 1$, however, then the predictive probability $\hat{\theta}_1$ is greater than 0.5, and the optimal stock allocation ω_1 is positive. These values are given below for each value of a, in addition to the expected stock return

$$\hat{R}_{T+1} = E\{R_{T+1} | \Phi_T\} = \hat{\theta}_1 \cdot 40\% + (1 - \hat{\theta}_1) \cdot (-40\%), \tag{16}$$

and the corresponding expected monthly return on the optimal portfolio,

$$\hat{R}_p = \omega_1 \cdot \hat{R}_{T+1}. \tag{17}$$

a	$\hat{\theta}_1$	$\hat{R}_{T+1}$	ω_1	$\hat{R}_p$	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 $a = 1$, the sample evidence leads the investor to choose a stock allocation of 100% if state 1 is observed but zero if state 2 is observed. The stock allocation for state 1 is decreasing in a, but that allocation is still 20% for $a = 21$.

Additional insight into the economic significance of the sample evidence on stock-return predictability can be obtained by comparing the levels of the investor's expected utility associated with optimal and suboptimal asset allocations. For a given state $s_T = j$, we compare the investor's certainty equivalent return (CER) for the optimal allocation to the investor's CER for a suboptimal allocation, where the latter allocation would have been optimal had a different state $i \neq j$ occurred.[9] The CER for both allocations is computed under the same pdf, the predictive pdf for $s_T = j$. Suppose we make this comparison, for example, when state 1 is observed. We then compare the optimal allocation ω_1 to a suboptimal allocation of zero in stock (the optimal allocation if state 2 is observed). The difference between the CER of the optimal allocation and the CER of the suboptimal allocation is given above as ΔCER for each of the three values of a. This measure ranges from 8.58% when $a = 1$ to 0.32% when $a = 21$. 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.

2.3. 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.[10] 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.[11] 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. Rather, our objective, as stated earlier, is simply to use the asset-

[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). In the current example with the logarithmic utility function, the CER is given by

$$\text{CER} = e^{\bar{v}} - 1.$$

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

[11] 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).

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.

Although interpretations of p-values no doubt differ across readers, with some readers attaching more importance than others to reported p-values of, say 1%, we also suggest that p-values of 30% 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, $\theta_1 = \theta_2$.[12] This test is based on the conditional distribution of M_1 (the number of periods with a 40% return following state 1) given the two-way table's row and column sums. 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 sharply with the economic significance of stock-return predictability as reflected in the investor's asset allocation decisions.

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 $\omega_B/\omega_c = 0.82$.[13] Although Brown does not examine measures of statistical significance, 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% of the allocation that would be chosen if the true Sharpe ratio were known to be 0.2.

In computing optimal asset allocations and comparing certainty equivalent returns, we

[12]See Kendall and Stuart (1979).

[13]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.

compute expected utility with respect to the Bayesian investor's predictive pdf. Thus, expected utility is as perceived by the investor, conditional on the observed data 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 to data sets other than the one observed. The performance in repeated samples of the decision rule $\omega(\Phi_T)$, where Φ_T is viewed as random, invokes the frequentist concept of "risk." The *risk function* of a decision rule, defined on 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 (15) 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. Of course, neither we nor the investor know the true value of θ.[14] More importantly, though, we suggest that the typical investor making the asset-allocation decision observes a given sample of returns and will probably not observe another non-overlapping sample of a similar size in his or her lifetime. In other words, although performance in repeated samples might be relevant for decision rules in other applications, perhaps even some in finance, it is difficult to motivate such a consideration for the stocks-versus-cash asset-allocation decision. Thus, we suggest that the conditional Bayesian approach is more relevant to our investigation of the economic significance of the empirical evidence in a given sample. The next section applies this approach using a linear regression model.

A previous application of the conditional Bayesian decision approach to investigating the economic significance of empirical evidence is provided by McCulloch and Rossi (1990), who use such a framework to assess the empirical evidence regarding departures from the Arbitrage Pricing Theory (APT) of Ross (1976). They compare an investor's certainty equivalent for a portfolio that is optimal under beliefs that the APT holds exactly to the

[14]Integrating the risk function over a prior pdf for θ gives the *Bayes risk* of the decision rule, which can often provide a link between conditional Bayesian decisions and frequentist-based decision rules (subject to existence and regularity conditions). See Berger (1985) for extensive discussions and comparisons of frequentist and Bayesian decision principles.

certainty equivalent for a portfolio that is optimal under beliefs that allow for departures from the exact pricing relation. In each case, the optimal portfolio and certainty equivalent are computed with respect to the investor's predictive pdf, conditional on the given sample.[15] As in this study, the performance of a portfolio decision rule in repeated samples (frequentist risk) is not addressed.

3. Asset Allocation Based on Regression Evidence

3.1. The Conditional Distribution of the Stock Return

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

$$r_t = x'_{t-1}b + \epsilon_t \ , \tag{18}$$

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_ϵ^2. 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.[16] 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}.[17] Thus, the regression in (18) departs somewhat from the standard 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.[18]

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

$$y'_t = x'_{t-1}B + u'_t, \tag{19}$$

[15] Our approach differs from that of McCulloch and Rossi in one key respect. We compute certainty equivalents for two alternative portfolio allocations using one common predictive pdf, rather than two predictive pdf's arising from different models (prior beliefs).

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

[17] 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 (1994).

[18] See, for example, page 59 of Zellner (1971).

where B is a $(N+1) \times N$ 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.[19] 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,

$$Y = XB + U, \tag{20}$$

where

$$Y = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix}, \; X = \begin{bmatrix} 1 & y_0' \\ 1 & y_1' \\ \vdots & \vdots \\ 1 & y_{T-1}' \end{bmatrix}, \text{ and } U = \begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_T' \end{bmatrix}. \tag{21}$$

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

$$p(U|\Sigma) \propto |\Sigma|^{-T/2} \exp[-\frac{1}{2}\mathrm{tr}U'U\Sigma^{-1}], \tag{22}$$

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.[20] 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,

$$p(Y|B, \Sigma, x_0) \propto |\Sigma|^{-T/2} \exp[-\frac{1}{2}\mathrm{tr}(Y - XB)'(Y - XB)\Sigma^{-1}], \tag{23}$$

since the Jacobian of the transformation from U to Y is equal to unity.[21]

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

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{24}$$

$$\hat{\sigma}_r^2 = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2, \tag{25}$$

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

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

[21]Define the $TN \times 1$ vectors $\tilde{u} = \mathrm{vec}(U)$ and $\tilde{y} = \mathrm{vec}(Y)$, 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.

$$\bar{y} = \frac{1}{T}\sum_{t=0}^{T-1} y_t, \tag{26}$$

$$\hat{\Sigma}_y = \frac{1}{T}\sum_{t=0}^{T-1}(y_t - \bar{y})(y_t - \bar{y})', \tag{27}$$

$$\hat{b} = (X'X)^{-1}X'y = \begin{bmatrix} \bar{r} - \hat{\beta}'\bar{y} \\ \hat{\beta} \end{bmatrix} \tag{28}$$

$$R^2 = 1 - \frac{(y - X\hat{b})'(y - X\hat{b})}{T\hat{\sigma}_r^2}, \tag{29}$$

and y 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,

$$p(B, \Sigma) \propto |\Sigma|^{-(N+2)/2}, \tag{30}$$

which is intended to represent vague or noninformative prior beliefs about the parameters.[22] With this prior distribution and the likelihood function in (23), the predictive pdf is easily obtained from known results (see Appendix). The predictive pdf for r_{T+1}, which is given by a Student t distribution, can then be written in terms of y_T and the above sample statistics:

$$p(r_{T+1}|\Phi_T) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)}\left(\frac{1}{(\nu-2)\sigma_T^2}\right)^{1/2}\left[1 + \left(\frac{1}{\nu-2}\right)\frac{(r_{T+1} - \mu_T)^2}{\sigma_T^2}\right]^{-(\nu+1)/2}, \tag{31}$$

where

$$\mu_T = \mathrm{E}\{r_{T+1}|\Phi_T\} = \bar{r} + \hat{\beta}'(y_T - \bar{y}), \tag{32}$$

$$\sigma_T^2 = \mathrm{var}\{r_{T+1}|\Phi_T\} = \frac{T}{T - 2(N+1)}(1 - R^2)\left[1 + \frac{1}{T}(1+q)\right]\hat{\sigma}_r^2, \tag{33}$$

$$q = (y_T - \bar{y})'\hat{\Sigma}_y^{-1}(y_T - \bar{y}), \text{ and} \tag{34}$$

$$\nu = T - 2N. \tag{35}$$

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

[22]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.

the diffuse prior in (30) with a hypothetical "prior" sample of size T_0 in which the sample R-squared is *exactly* equal to zero. Except for this extreme 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 $\hat{\Sigma}_y$.[23] 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 (31), where μ_T, σ_T^2 and ν, although redefined, are still written in terms of y_T and the same sample statistics in (24) through (27):

$$\mu_T = \mathrm{E}\{r_{T+1}|\Phi_T\} = \bar{r} + \left(\frac{T}{T^*}\right)\hat{\beta}'(y_T - \bar{y}), \tag{36}$$

$$\sigma_T^2 = \mathrm{var}\{r_{T+1}|\Phi_T\} = \frac{T^*}{T^* - 2(N+1)}\left[1 - \left(\frac{T}{T^*}\right)^2 R^2\right]\left[1 + \frac{1}{T^*}(1+q)\right]\hat{\sigma}_r^2, \tag{37}$$

$$\nu = T^* - 2N. \tag{38}$$

Note that, since $\bar{y}$ and $\hat{\Sigma}_y$ are assumed to be the same in the actual and hypothetical samples, q is still defined as in (34). In all of the calculations reported below, we set $T_0 = 1200$ months. In other words, the informative prior distribution is equivalent to having observed 100 years of data in which the estimated regression slopes on all predictive variables are equal to zero.

The predictive variance σ_T^2, in both (33) and (37), incorporates uncertainty about parameter values, i.e., estimation risk.[24] Suppose that we hold constant the values of the sample statistics in (24) through (29). 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. 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 (33) and (37). This effect is analogous to that in the standard frequentist setting for a multiple regression with $N+1$ regressors (including

[23]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).

[24]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).

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

3.2. Computing Optimal Asset Allocations

The predictive pdf in (31) is the conditional distribution used in computing expected utility, $\mathrm{E}\{v(W_{T+1})|\Phi_T\}$. To our knowledge, the present 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.[25] 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 to $-\infty$ 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.[26] We simply restrict $\omega \leq .99$ throughout. The maximization problem is solved numerically, since we are unaware of an exact analytic solution. [27] In virtually all cases, however, the optimal allocation to stocks is well approximated by

$$\hat{\omega} = \frac{\mu_T}{A\sigma_T^2} + \frac{1}{2A}, \tag{39}$$

[25]In fact, Zellner and Chetty specify a multivariate regression model whose parameters are given a diffuse prior identical to (30). 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.

[26]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.

[27]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, 1] bounds. Equation (39) 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.[28] If (39) is rewritten slightly as

$$\hat{\omega} = \frac{\alpha_T - i_{T+1}}{A\sigma_T^2}, \tag{40}$$

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 $q = (y_T - \bar{y})'\hat{\Sigma}_y^{-1}(y_T - \bar{y})$. The first three quantities, $\bar{r}$, $\hat{\sigma}_r$, and i_{T+1}, are held constant across all samples. We set $\bar{r} = 0.49\%$ and $\hat{\sigma}_r = 5.60\%$, 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 $i_{T+1} = 0.235\%$, the continuously compounded monthly yield on the Treasury bill with 27 days to maturity as of 12/31/93.[29]

In each sample considered, 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, holding constant the other characteristics of the sample. 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, $\bar{r} + \hat{\beta}'(y_T - \bar{y})$. The difference between this 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 $\sqrt{R^2}\hat{\sigma}_r$. Thus, if the fitted value of r_{T+1} based on y_T is δ sample standard deviations of the fitted values away from the overall sample mean, $\bar{r}$, then

$$\hat{\beta}'(y_T - \bar{y}) = \delta\sqrt{R^2}\hat{\sigma}_r. \tag{41}$$

As explained above, we specify $\hat{\sigma}_r^2$ and R^2 for each sample, so a fitted value based on y_T

[28] 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 (39) is a straightforward application of results contained in Merton (1969), and an expression equivalent to (39) also arises as a solution to a special case of the dynamic consumption-investment problem analyzed in that study.

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

is determined simply by specifying, in addition, a value for δ. For each sample considered, we compute the optimal allocation ω for five realizations of δ: -1.0, -0.5, 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 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 $q = (y_T - \bar{y})'\hat{\Sigma}_y^{-1}(y_T - \bar{y})$ 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 (41). 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 (39), and σ_T^2 is increasing in q (equation (33) or (37)). 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.[30] Here again we follow a conservative approach. For each T and N, we take q as the 99th percentile of a Monte Carlo

[30]This statement is based on our Monte Carlo evidence. 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, chapters 5 and 7 of Anderson (1984). We are unaware of an analytic finite-sample result in the presence of serial dependence.

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.[31]

3.3. Results with T = 804 and N =100

Table 1 reports optimal asset allocations in samples with $T = 804$ observations and $N = 100$ predictive variables. As noted in the introduction, although this number of predictive variables is implausibly 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 under four different outcomes for the unadjusted sample R^2 (0.02, 0.06, 0.12, and 0.18), and these R^2 values produce a wide range of p-values, shown in the second column. This p-value is the probability of observing a sample with an R^2 greater than the value in the first column when the true population R-squared is zero, computed using the result that, under that hypothesis,

$$F = \left(\frac{T-N-1}{N}\right)\left(\frac{R^2}{1-R^2}\right) \tag{42}$$

is distributed (central) F with N and $T-N-1$ degrees of freedom.[32]

With 100 predictive variables, obtaining an R^2 of only 0.02 is very unlikely, even if returns are truly unpredictable, so the p-value in that case is nearly 1.0. (Of course, if the F-statistic is computed using instead a small value for N, then an R^2 of 0.02 produces a very small p-value.) We begin with $R^2 = 0.02$ because, as noted in the introduction, a number of studies using data beginning in 1927 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 much larger set of 100 variables, then the sample R^2 produced by the larger set must, by construction, be at least 0.02 and would most likely be considerably larger. (Recall that R^2, defined in (29), is the *unadjusted* R-squared.) Thus, we suggest that the asset allocations reported for $N = 100$ when $R^2 = 0.02$ provide a conservative "worst-case" characterization of the importance of the reported regression evidence to a Bayesian investor who includes all 100 variables in the regression model.

When $R^2 = 0.02$, the fitted regression values' sample standard deviation, $\sqrt{R^2}\hat{\sigma}_r$, is 79

[31] We generate y_t with a zero mean vector and a scalar variance-covariance matrix, but this is without loss of generality, since q is invariant up to nonsingular linear transformations of y_t.

[32] See, for example, Judge et al. (1985). This distributional result relies on the standard regression-model assumptions and generally does not hold, for example, in the presence of lagged stochastic regressors entertained by the regression framework used here.

basis points per month, so the one-step-ahead fitted regression prediction for r_{T+1} ranges from -30 basis points to 128 basis points as δ ranges from -1 to 1 (recall equation (41)). As reported in table 1, an investor with relative risk aversion (A) equal to 2 and diffuse prior beliefs would choose a stock allocation of 18% at $\delta = -1$, whereas that same investor would allocate 56% to stocks when $\delta = 1$, a threefold increase. If that investor's prior beliefs are given instead by the no-predictability informative prior, those allocations range from 35% to 70%. 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 100 possible predictive variables and the sample R^2 of that regression is only 0.02, 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 the previous section, that compares expected utilities associated with optimal and suboptimal allocations. Let ω_0 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 optimal allocation ω_0, which is then suboptimal, to the investor's expected utility for the portfolio formed using the 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 1 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_0. In the case where $R^2 = 0.02$, $A = 2$, and the prior is diffuse, for example, we see from table 1 that, when $\delta = 1$, the optimal allocation of 56% to stocks produces a CER that is 7.7 basis points (per month) higher than an allocation of 37% to stocks, the optimal allocation for $\delta = 0$. With the informative no-predictability prior, again when $\delta = 1$, the CER for the optimal 70% stock allocation is 2.8 basis points higher than the CER for a 52% allocation, the optimal allocation for $\delta = 0$.

In addition to the cases discussed above, in which $A = 2$ and $R^2 = 0.02$, Table 1 also reports results for investors with other coefficients of relative risk aversion ($A = 1$ and $A = 5$)

and for regressions with larger R^2 values. In general, as might be anticipated, 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 higher R^2 values provide some insight into the contrast between a standard statistical characterization of the regression evidence and the economic significance of the evidence, as characterized by the asset-allocation results. A further investigation of this contrast is pursued in the next subsection.

As noted earlier, $R^2 = 0.02$ is an unlikely sample outcome with 100 predictive variables. Even if returns are truly unpredictable, a more typical outcome in that case is $R^2 = 0.12$, as indicated by the p-value of 0.595. With that higher but still "insignificant" value for R^2, the asset-allocation results are quite strong even under the informative no-predictability prior. The $A = 2$ investor then invests only 9% in stocks at $\delta = -1$ but invests 96% in stocks at $\delta = 1$, and the certainty-equivalent comparisons for both allocations are approximately 17 basis points. Recall that, following (41), with a given range for δ, a higher R^2 leads to a wider range for the one-step-ahead fitted regression prediction for r_{T+1}. With $R^2 = 0.12$, that fitted monthly excess return ranges from -145 basis points to 243 basis points as δ ranges from -1 to 1.

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 a conscious 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 a wide ranges for both the sample size (T) and the number of predictive variables (N).

3.4. Economic Significance and Regression Statistics

The preceding analysis contains several illustrations of the potentially sharp contrast between the economic significance of sample evidence and standard interpretations of statistical measures commonly used to summarize that evidence. That contrast is investigated further in this subsection.

We begin by selecting 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, 100, and 200, and we consider all nine of the (T, N) combinations in which $N < T$. 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 (33) 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 (39), which can be rewritten, using (36), (37), and (41), as

$$\hat{\omega} = \left[\frac{\bar{r} + \frac{1}{2}\sigma_T^2}{A\sigma_T^2}\right] + \gamma\,\delta, \tag{43}$$

where

$$\gamma = \left(\frac{\left(\frac{T}{T^*}\right)\sqrt{R^2}}{1 - \left(\frac{T}{T^*}\right)^2 R^2}\right)\left(1 - \frac{2(N+1)}{T^*}\right)\left(A\hat{\sigma}_r\left[1 + \frac{1}{T^*}(1+q)\right]\right)^{-1}. \tag{44}$$

As defined earlier, $T^* = T + T_0$, where $T_0 = 1200$ with the informative no-predictability prior and $T_0 = 0$ with the diffuse prior. The first term on the right-hand side of (43) does not depend on the current values of the predictive variables, y_T.[33] Those values enter the second term, where γ, which we interpret as the degree of economic significance, measures

[33]Even though y_T appears in q in equation (34), recall from the previous discussion that we specify the same (high) value for q across different values of y_T.

the sensitivity of $\hat{\omega}$ to the current values of the predictive variables, summarized by δ as in (41). 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, 1] bounds).

Once T and N are specified, then R^2 is the only unknown quantity on the right-hand side of (44). We specify a value for γ and then solve (44) 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 (42), the same statistic reported in table 1. 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.[34] 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,

$$\ln(O^*) = -\frac{1}{2}\left[T\ln(1-R^2) - N\ln(T)\right], \tag{45}$$

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 $\ln(O^*)$ here simply to broaden our choices of statistics used in standard analyses of regression evidence.[35]

Table 2 reports results for γ specified alternately as 20%, 30%, and 40%, 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 (39) 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 $[0,1]$ range for ω is binding).

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

[35] Statistics that can also be computed from T, N, and R^2 include the adjusted R-squared, $\bar{R}^2 = R^2 - [(T-1)/(T-N-1)](1-R^2)$, 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.

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 2. 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 2. Most of the p-values are large—generally greater than 0.5 and often nearly 1.0. Small p-values occur in cases where N is large relative T, especially for the higher values of γ, such as the cases ($T = 60$, $N = 25$) and ($T = 804$, $N = 200$). 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. For a given T and γ, the p-value is not monotonic in N, as demonstrated both when $T = 60$ and $T = 804$.

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 contains sufficient evidence of predictability to exert a substantial influence on the asset-allocation decision.

Table 3 reports the same analysis as in table 2, except that the informative no-predictability prior replaces the diffuse prior. The table omits all of the cases with $T = 7$ and some of the cases with $T = 60$ when $N = 10$ or 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 from 20% to 40%, the p-value can decrease considerably. For example, in the case of ($T = 804$, $N = 2$), the p-value is 0.275 for $\gamma = 20\%$ but 0.006 for $\gamma = 40\%$. In the case of ($T = 804$, $N = 200$), in table 2 and table 3, the p-value goes from nearly 1.0 to 0.001 as γ increases from 20% to 40%. As in table 2, however, there appears to be no simple correspondence between the statistical measures and the asset-allocation results.

4. 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.[36] 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 have concluded, however, that stock returns exhibit conditional heteroskedasticity [e.g., French, Schwert, and Stambaugh (1987)]. An interesting extension of our framework would be to analyze the asset-allocation problem when conditional heteroskedasticity is included in the regression model.

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-

[36]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.

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.[37] Addressing the latter would require that each month the investor not only incorporate a new observation of the predictive variables into the conditional mean but also update his beliefs about all parameters of the predictive distribution. Of course, transaction costs would present an additional challenge to any modeling effort with dynamic rebalancing.

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

Appendix

Although the VAR model in (20) employs assumptions different from those in the traditional multivariate regression model (MVRM), the likelihood function in (23) 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 (30). 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

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

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

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

and

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

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

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

where $V = g\nu S^{-1}$ 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:[38]

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

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

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

Let

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

[38]See Zellner (1971, page 387).

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

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

where

$$\mu_T = E\{r_{T+1}|\Phi_T\} = x_T\hat{b}. \qquad \text{(A.11)}$$

Substituting (A.2) and (28) into (A.11) yields (32). The second moment about the mean of the US t pdf is:

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

Using (A.12), we obtain

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

Substituting (A.11) and (A.13) into (A.10) yields (31). To obtain (33) we first rewrite (A.12) using (A.9):

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

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

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

Using (A.8) and the definition of R^2 in (29), we get

$$S_{11} = T(1-R^2)\sigma_r^2. \qquad \text{(A.16)}$$

To simplify the expression for g in (A.5), observe that[39]

$$g = 1 - x_T'(X'X + x_T x_T')^{-1}x_T = \frac{1}{1+x_T'(X'X)^{-1}x_T}. \qquad \text{(A.17)}$$

Using (20), we get

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

Inverting (A.18) yields

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

Using (A.17),(A.19), and (34), we obtain

$$\frac{1}{g} = 1 + x_T'(X'X)^{-1}x_T = 1 + \frac{1}{T}\left(1 + (y_T-\bar{y})'\hat{\Sigma}_y^{-1}(y_T-\bar{y})\right) = 1 + \frac{1}{T}(1+q). \qquad \text{(A.20)}$$

Substituting (A.15), (A.16), and (A.20) into (A.14) yields (33).

[39]See, for example, Zellner (1971, page 73).

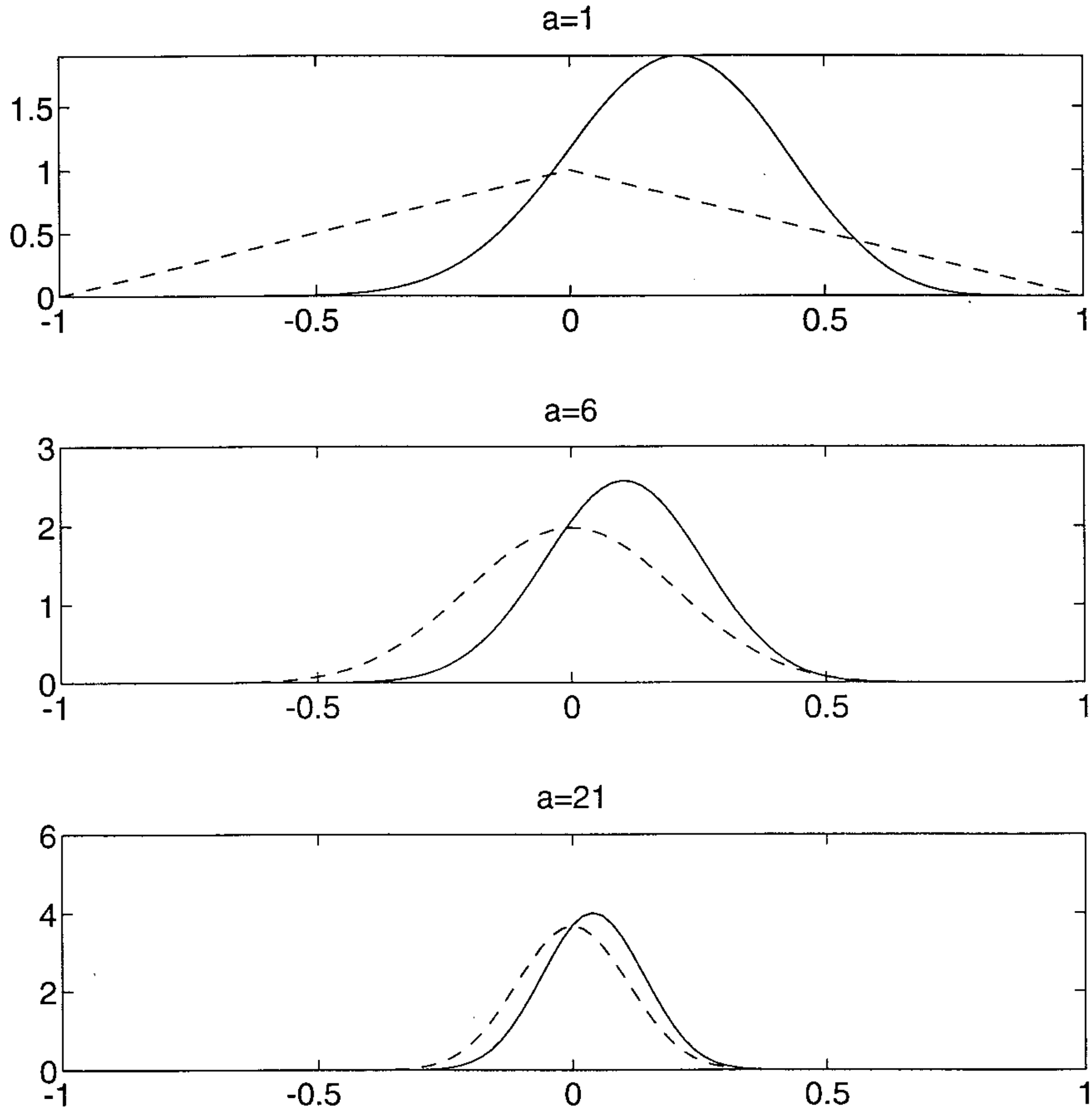

Figure 1. Prior and posterior distributions for $(\theta_1 - \theta_2)$. The plot displays, for $a = 1$, $a = 6$, and $a = 21$, the prior distribution (dashed curve) and the posterior distribution (solid curve) for the quantity $(\theta_1 - \theta_2)$.

Table 1

Optimal Stock Allocations and Comparisons of Certainty Equivalents with 804 Observations and 100 Variables

			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
A. Diffuse prior											
0.02	1.000	1	35	55	74	93	99	15.4	3.9	3.9	13.5
0.06	1.000	1	5	40	75	99	99	47.9	12.0	10.9	27.5
0.12	0.595	1	0	24	77	99	99	94.4	25.7	17.2	39.2
0.18	0.001	1	0	9	79	99	99	133.1	41.3	20.8	45.4
0.02	1.000	2	18	27	37	47	56	7.7	1.9	1.9	7.7
0.06	1.000	2	2	20	37	55	73	23.8	6.0	6.0	24.1
0.12	0.595	2	0	12	38	65	91	47.0	12.8	12.9	51.7
0.18	0.001	2	0	5	39	74	99	66.2	20.5	20.7	81.8
0.02	1.000	5	7	11	15	19	22	3.0	0.8	0.8	3.0
0.06	1.000	5	1	8	15	22	29	9.5	2.4	2.4	9.5
0.12	0.595	5	0	5	15	26	36	18.7	5.1	5.1	20.3
0.18	0.001	5	0	2	16	30	43	26.4	8.2	8.2	32.9
B. No-predictability informative prior (100 years)											
0.02	1.000	1	69	87	99	99	99	4.0	0.7	0.0	0.1
0.06	1.000	1	43	74	99	99	99	13.8	2.8	0.1	0.1
0.12	0.595	1	18	61	99	99	99	29.3	6.2	0.1	0.2
0.18	0.001	1	0	52	99	99	99	45.2	9.8	0.1	0.2
0.02	1.000	2	35	43	52	61	70	2.8	0.7	0.7	2.8
0.06	1.000	2	22	37	52	68	83	8.4	2.1	2.1	8.5
0.12	0.595	2	9	31	53	75	96	17.0	4.3	4.3	17.2
0.18	0.001	2	0	26	53	80	99	25.8	6.5	6.5	25.4
0.02	1.000	5	14	17	21	24	28	1.1	0.3	0.3	1.1
0.06	1.000	5	9	15	21	27	33	3.4	0.8	0.8	3.4
0.12	0.595	5	4	12	21	30	39	6.8	1.7	1.7	6.8
0.18	0.001	5	0	10	21	32	43	10.3	2.6	2.6	10.3

Optimal stock allocations and certainty equivalents are computed with respect to a Baysian 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.

Table 2

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

					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
					$\gamma = 20\%$								
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	100	0.021	1.000	-325.84	17	27	37	47	57	8.1	2.0	2.0	8.2
804	200	0.129	1.000	-613.41	10	20	30	40	50	19.9	5.0	5.0	20.1
					$\gamma = 30\%$								
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	100	0.045	1.000	-315.76	7	22	37	52	67	17.8	4.5	4.5	18.0
804	200	0.228	0.834	-564.89	0	15	30	45	61	39.5	10.0	10.0	40.3
					$\gamma = 40\%$								
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	0.1
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	100	0.076	1.000	-302.78	0	18	38	58	78	30.5	7.7	7.7	31.0
804	200	0.317	0.001	-515.55	0	11	31	51	71	58.9	15.6	15.7	63.6

Optimal stock allocations and certainty equivalents are computed with respect to a Baysian 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 (43) and (44) 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 (45) in the text, that compares the regression model to a model with no predictability.

Table 3

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 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
					$\gamma = 20\%$								
60	2	0.231	0.001	3.77	82	92	99	99	99	1.0	0.2	0.0	0.0
60	10	0.333	0.018	-8.31	69	79	89	99	99	1.5	0.4	0.4	1.2
60	25	0.566	0.059	-26.10	54	64	74	84	94	2.0	0.5	0.5	2.0
804	2	0.003	0.275	-5.39	82	92	99	99	99	0.9	0.1	0.0	0.0
804	10	0.004	0.968	-31.69	71	81	91	99	99	1.5	0.4	0.3	0.9
804	25	0.007	1.000	-80.89	58	68	78	88	98	1.9	0.5	0.5	1.9
804	100	0.026	1.000	-323.98	32	42	52	62	72	3.6	0.9	0.9	3.6
804	200	0.087	1.000	-632.21	20	30	40	50	60	6.6	1.7	1.7	6.7
					$\gamma = 30\%$								
60	2	0.518	0.000	17.80	72	87	99	99	99	2.4	0.5	0.0	0.0
60	10	0.748	0.000	20.93	59	74	89	99	99	3.5	0.9	0.8	2.0
804	2	0.007	0.055	-3.77	72	87	99	99	99	2.3	0.4	0.0	0.0
804	10	0.010	0.645	-29.50	61	76	91	99	99	3.4	0.9	0.6	1.5
804	25	0.015	0.986	-77.47	48	63	78	93	99	4.2	1.0	1.0	3.8
804	100	0.057	1.000	-310.72	22	37	52	67	82	8.1	2.0	2.0	8.1
804	200	0.190	0.998	-584.23	10	25	40	55	70	14.6	3.7	3.7	14.8
					$\gamma = 40\%$								
60	2	0.919	0.000	71.36	62	82	99	99	99	4.4	0.9	0.0	0.1
804	2	0.013	0.006	-1.50	62	82	99	99	99	4.3	0.9	0.0	0.1
804	10	0.017	0.175	-26.41	51	71	91	99	99	5.9	1.5	0.9	2.1
804	25	0.027	0.663	-72.66	38	58	78	98	99	7.4	1.8	1.9	5.7
804	100	0.101	0.934	-291.85	12	32	53	73	93	14.2	3.6	3.6	14.3
804	200	0.323	0.001	-512.09	0	20	40	60	80	25.4	6.4	6.4	25.7

Optimal stock allocations and certainty equivalents are computed with respect to a Baysian 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 (43) and (44) 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 (45) 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 100-year sample in which all estimated slopes are exactly zero.

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.

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, Anderson Graduate School of Management, UCLA, Los Angeles, CA.

Brent, Richard P., 1973, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, N.J.

Brown, Stephen J., 1979, "Optimal Portfolio Choice Under Uncertainty: A Bayesian Approach," chapter 8 in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.

Campbell, John Y., 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, Iowa City, IA and Duke University, Durham, NC.

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19, 3–29.

Frost, Peter A. and James E. Savarino, 1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis* 21, 293–305.

Gennotte, Gerard, 1986, "Optimal Portfolio Choice Under Incomplete Information," *Journal of Finance*, 41, 733–746.

Goetzmann, William N. and Philippe Jorion, 1994, "A Longer Look at Dividend Yields," working paper, Yale University, New Haven, CT and University of California at Irvine, Irvine, CA.

Grauer, Robert R. and Nils H. Hakansson, 1992, "Stein and CAPM Estimators of the Means in Asset Allocation: A Case of Mixed Success," working paper, Simon Fraser University, Burnaby, BC, and University of California, Berkeley, CA.

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

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 no. 222, University of Chicago, Chicago, IL.

Kandel, Shmuel and Robert F. Stambaugh, 1991, "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics* 27, 39–71.

Karatzas, Ionnis and Xing-Xiong Xue, 1991, "A Note on Utility Maximization Under Partial Observations," *Mathematical Finance*, 1, 57–70.

Kaul, Gautam, 1995, "Predictable Components in Stock Returns," working paper, University of Michigan, Ann Arbor, MI.

Keim, Donald B. and Robert F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics* 17, 357–390.

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

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

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 Brown, Stephen J., 1984, "Model Selection When There is Minimal Prior Information," *Econometrica* 52, 1291–1312.

Lo, Andrew W. and A. Craig MacKinlay, 1995, "Maximizing Predictability in the Stock and Bond Markets," working paper, MIT, Cambridge, MA, and University of Pennsylvania, Philadelphia, PA.

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*, 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.

Solnik, Bruno, 1993, "The Performance of International Asset Allocation Strategies Using Conditioning Information," *Journal of Empirical Finance* 1, 33–55.

Stambaugh, Robert F., 1986, "Bias in Regressions with Lagged Stochastic Regressors," CRSP working paper no 156, University of Chicago, Chicago, IL.

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

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.