Portfolio Selection and Asset Pricing Models

ĽUBOŠ PÁSTOR*

ABSTRACT

Finance theory can be used to form informative prior beliefs in financial decision making. This paper approaches portfolio selection in a Bayesian framework that incorporates a prior degree of belief in an asset pricing model. Sample evidence on home bias and value and size effects is evaluated from an asset-allocation perspective. U.S. investors' belief in the domestic CAPM must be very strong to justify the home bias observed in their equity holdings. The same strong prior belief results in large and stable optimal positions in the Fama–French book-to-market portfolio in combination with the market since the 1940s.

FINANCE THEORY HAS PRODUCED A VARIETY of models that attempt to provide some insight into the environment in which financial decisions are made. How should these models be used by financial decision makers? Empirical finance typically approaches a theoretical model by testing whether its implications are supported by the data. Based on the result of a hypothesis test, the model is either rejected or not rejected. It is not clear what such an outcome implies about the usefulness of the model for decision making. If the model is not rejected, should it be used as the truth? And if it is rejected, should it be discarded as worthless? Such a simplistic approach, based solely on the result of a hypothesis test, fails to capture many aspects of both the model and the data that could potentially be useful to a decision maker. Instead, it might be reasonable to assume that financial models are neither perfect nor useless. By definition, every model is a simplification of reality. Hence, even if the data fail to reject the model, the decision maker may not necessarily want to use the model as a dogma. At the same time, the notion that models implied by finance theory could be entirely worthless seems rather extreme. Hence, even if the data reject the model, the decision maker may want to use the model at least to some degree.

* Graduate School of Business, University of Chicago. The paper is based on my dissertation at the Wharton School, University of Pennsylvania. I am grateful to the members of my dissertation committee, Don Keim, Karen Lewis, Craig MacKinlay, Frank Schorfheide, and especially to the committee chair Rob Stambaugh, for their numerous helpful comments. The comments of Greg Bauer, Michael Brandt, Mark Britten-Jones, Chris Géczy, Chris Jones, Ron Kaniel, Krishna Ramaswamy, Jay Shanken (the referee), René Stulz (the editor), and seminar participants at Columbia University, Harvard University, MIT, New York University, the University of California at Los Angeles, the University of Chicago, the University of Pennsylvania, the University of Rochester, Yale University, and the 1999 meetings of the American Finance Association are also appreciated. Of course, I am fully responsible for all the weaknesses of the paper.

A natural approach to using financial models in decision making can be developed in a Bayesian framework. A model can be used as a point of reference around which the decision maker can center his prior beliefs. These prior beliefs are combined with the data, which may violate the implications of the model, and the revised beliefs are used to make decisions. The relative importance of the sample evidence versus the model depends on the strength of the violations of the model in the data relative to the strength of the prior belief in the model.[1] Following such an approach, this paper explores how asset pricing models can be used in portfolio selection.

The goal of portfolio selection is to find an optimal allocation of wealth across a number of assets. At least two approaches to portfolio selection are commonly used in finance. A "data-based" approach assumes a functional form for the distribution of asset returns and estimates its parameters from the time series of returns. For example, sample estimates of the mean and covariance matrix of asset returns can be used to compute the optimal weights in a mean-variance framework. This approach ignores the potential usefulness of asset pricing models.[2] Asset pricing models imply an alternative approach to portfolio selection. In this "model-based" approach, the optimal portfolio of every investor is a combination of benchmark portfolios that expose the investor only to priced sources of risk.[3] For example, under the Capital Asset Pricing Model, the market portfolio is the single benchmark, and is therefore the optimal portfolio of every investor. This approach makes no use of the time series of returns on the nonbenchmark assets.

The two typical approaches to portfolio selection essentially reflect two extreme views about the validity of asset pricing models. The first approach regards asset pricing models as useless, and the second approach considers one of these models to be a perfect description of reality. Such polar views could be adopted as a result of a hypothesis test, as described at the outset.[4] However, the portfolio literature is silent about what happens in between. For example, what if an investor thinks highly of a certain pricing model, but is concerned that the model may not hold exactly due to mild violations of its assumptions in the real world?[5]

[1] The idea of using financial models to form prior beliefs in decision making is also mentioned in Stambaugh (1998).

[2] Examples of asset pricing models include the Capital Asset Pricing Models (CAPM) of Sharpe (1964) and Lintner (1965), the intertemporal CAPM of Merton (1973), and models based on the arbitrage pricing theory of Ross (1976). The precise meaning of the term *data-based* approach is clarified in Section I.A.

[3] Examples of benchmark portfolios are factor-mimicking portfolios, whose returns mimic the realizations of the factors in a factor-based asset pricing model.

[4] Frequentist tests of asset pricing models are too numerous to list. Bayesian tests include Shanken (1987), Harvey and Zhou (1990), McCulloch and Rossi (1990, 1991), Kandel, McCulloch, and Stambaugh (1995), and Geweke and Zhou (1996).

[5] For instance, Jagannathan and Wang (1996) argue that "We have to keep in mind that the CAPM, like any other model, is only an approximation of reality. Hence, it would be rather surprising if it turns out to be "100 percent accurate"."

This paper approaches the portfolio selection problem in a Bayesian framework that incorporates the investor's prior degree of confidence in an asset pricing model. The degree of confidence can range from a dogmatic belief in the model to a belief that the model is useless. As the degree of skepticism about the model grows, the resulting optimal allocation moves away from a combination of benchmark portfolios toward the allocation obtained in the data-based approach. We explore how fast the optimal allocation moves from one extreme to the other in response to sample evidence and what determines the strength of the influence of sample evidence on the optimal allocation. The approach developed in the paper uses both an asset pricing model and the time series of asset returns to find the optimal portfolio.

The investor specifies an informative prior distribution on the assets' mispricing α within an asset pricing model. As is typical in Bayesian analysis, the sample mispricing $\hat{\alpha}$ is "shrunk" toward the prior mean of α to obtain the posterior mean of α, which is used in portfolio analysis. The most natural choice for the prior mean of α is zero, the value implied by the model. Prior confidence in the model's implication that $\alpha = 0$ is expressed through σ_α, the prior standard deviation of α. Due to the shrinkage in α, the sample mean is shrunk toward the expected return implied by the pricing model. Shrinking the sample mean reduces the sensitivity of the optimal weights to the sampling error in $\hat{\alpha}$. The weights have less extreme values and are more stable over time than in the data-based approach.

The idea of specifying an informative prior on α is proposed in Pástor and Stambaugh (1999), where the authors suggest a Bayesian approach to estimating costs of equity and analyze the sources of uncertainty in the cost of equity estimates for individual firms. The part of our methodology that leads to the posterior distribution can be viewed as a multivariate extension of a part of the methodology in Pástor and Stambaugh. However, the focus of this paper is quite different. Instead of examining an estimation problem, we concentrate on the decision problem of forming the optimal portfolio of multiple risky assets.

In this paper, the investor's prior beliefs are centered around an asset pricing model. In a related study, Black and Litterman (1992) suggest using the CAPM as a benchmark toward which the investor can shrink his subjective views about expected returns. The extent of the deviations from the CAPM depends on the investor's degree of confidence in his subjective views. That study makes no direct use of sample information about expected returns. In contrast, our approach shrinks the sample means toward their values implied by the model. The extent of the deviations from the model depends on the strength of the violations of the model in the data as well as on the investor's degree of confidence in the model.

The focus of the empirical analysis is to investigate the extent to which optimal holdings depart from the market portfolio, which plays a central role in finance theory. In our mean-variance examples, wealth is allocated between the market portfolio and an asset (or assets) with nonzero sample mispricing $\hat{\alpha}$ within the CAPM. It is well known that one should invest

(disinvest) in any asset whose α is positive (negative), since combining the asset with the market portfolio increases the portfolio's Sharpe ratio.[6] However, the true value of α is unknown. How much attention should the investor pay to a nonzero value of the sample estimate of α? The impact of $\hat{\alpha}$ on the optimal allocation is investigated for different assets and different prior beliefs about α.

The empirical analysis investigates the home bias in the equity holdings of U.S. investors and the issues of investing based on value and size. The home bias puzzle is associated with the observation that investors' equity holdings typically include a substantially larger proportion of domestic equities than is suggested by standard portfolio theory. U.S. investors hold only about eight percent of their equity holdings in foreign equities, although their optimal allocation in foreign equities based on the sample moments of asset returns is more than 40 percent.[7] However, several recent studies cannot reject the hypothesis that the global mean-variance efficient portfolio puts zero weights on non-U.S. stocks. These two different ways of looking at the data, one based on point estimates and the other on the result of a hypothesis test, lead to different conclusions about the home bias and about the benefits of international diversification for U.S. investors.

This paper assesses the evidence in the data from an asset allocation perspective. A U.S. investor confronted with the data decides how much to invest in foreign stocks. In our framework, the bias toward domestic equities can simply reflect a certain degree of prior confidence in the domestic CAPM. However, we find that U.S. investors' belief in the global mean-variance efficiency of the U.S. market portfolio must be very strong to justify the home bias observed in their equity holdings. Their actual holdings are consistent with the prior belief that the annual mispricing of a foreign stock portfolio within the domestic CAPM is in the tight interval between -2 percent and 2 percent.

Surprisingly, even the same strong prior belief in the CAPM is significantly revised by the sample evidence about the Fama and French (1993) book-to-market portfolio (HML, or "high minus low"). Consider an investor who allocates his wealth between HML and the market portfolio, and who believes that the annual mispricing of HML within the CAPM is between -2 percent and 2 percent. As of January 1997, this investor should optimally invest 40 percent of his wealth in HML, despite his strong belief in the

[6] An asset with a nonzero α that is combined with a passive portfolio is sometimes referred to as an "active portfolio," following Treynor and Black (1973). The portfolio's Sharpe ratio is the ratio of its expected excess return and the standard deviation of its return. Adding an asset with mispricing α to the market portfolio increases the portfolio's squared Sharpe ratio by $(\alpha/\sigma)^2$, where σ^2 is the residual variance from the market model regression. See Gibbons, Ross, and Shanken (1989).

[7] See Lewis (1999). The foreign portfolio in her example is Morgan Stanley's EAFE index, and her sample period is January 1970 through December 1996. The home bias puzzle is also present from the perspective of the international CAPM. The weight of non-U.S. equity in the value-weighted world market portfolio is about 60 percent.

mean-variance efficiency of the market portfolio. Moreover, this investor's optimal position in HML is large and rather stable, mostly between 20 percent and 40 percent, in every month since the early 1940s. The optimal positions in HML for investors with weaker beliefs in the CAPM are even larger and still fairly stable. For example, an investor who is completely skeptical about the CAPM should have held approximately 50 to 80 percent of his wealth in HML in the last five decades. The robust optimal weights in HML are primarily due to the robust value premium over the last 60 years.

The rest of the paper is organized as follows. Section I first describes the data-based approach to portfolio selection and then develops our "model-and-data-based" Bayesian methodology. Section II describes the data used in the empirical analysis. In Sections III through V, optimal combinations of the market portfolio with a number of different assets are explored. Section III examines the home bias puzzle. Section IV looks into investing based on value and size. Section V explores the effect of imposing prior beliefs that depart from the model. Section VI concludes.

I. Methodology

The methodology section is divided into four subsections. The first subsection lays out the portfolio selection problem and discusses the data-based approach to this problem. The remaining subsections develop a methodology that can be used to compute optimal weights in the presence of a nontrivial prior degree of belief in an asset pricing model. The second subsection specifies the assumptions on the stochastic behavior of returns and the resulting likelihood function. The third subsection describes the prior distribution on the model parameters. The final subsection describes how the predictive distribution of returns is obtained.

A. Portfolio Selection

Consider a risk-averse investor with a one-period investment horizon who must allocate funds between a riskless asset and a portfolio of $(N + K)$ risky assets, K of which are benchmark portfolios. The returns on the benchmark portfolios replicate the realizations of K priced sources of risk in a certain asset pricing model. The $(N + K)$ risky assets are referred to as "investable assets," and the N risky assets are referred to as "nonbenchmark assets" or simply "assets." The investor is assumed to consider the past to be informative about the future. The allocation decision is made based on the information set Φ containing a finite history of returns on the investable assets and prior information. The investor believes that his portfolio decision has no effect on the probability distribution of asset returns. The markets are assumed to be frictionless, with no transaction costs or taxes.

Let W denote the investor's current wealth, and δ the proportion of the wealth invested in the riskless asset. The optimal value of δ depends on the degree of the investor's risk aversion and is not investigated in this paper.

Let w denote the $(N + K) \times 1$ vector of the weights in the portfolio of the investable assets, where $w'\iota_{N+K} = 1$ and ι_{N+K} is an $(N + K)$-vector of ones. The investor's wealth one period later is

$$W_{+1} = W(1 + r_f + (1 - \delta)w'r_{+1}), \tag{1}$$

where r_f stands for the rate of return on the riskless asset and r_{+1} is the $(N + K) \times 1$ vector of the next-period returns on the investable assets in excess of r_f. The investor chooses w to maximize the expected utility of the next-period wealth:

$$\max_w \int u(W_{+1})p(r_{+1}|\Phi)\,dr_{+1}, \tag{2}$$

where u is the investor's utility function and $p(r_{+1}|\Phi)$ is the probability density of r_{+1} conditional on Φ, often referred to as the predictive density.[8] Although the predictive density is in general unknown, the density $p(r_{+1}|\theta,\Phi)$ is usually assumed to be known, where θ denotes the parameters of the statistical model that describes the stochastic behavior of returns. However, θ is unknown. The simplest way to deal with this challenge is to treat the sample estimates $\hat{\theta}$ as their true values. However, such an approach ignores the estimation risk in the estimates and hence understates the true level of uncertainty faced by the investor. As shown by Zellner and Chetty (1965), Brown (1976), Klein and Bawa (1976), and others, estimation risk can be accounted for in a Bayesian framework. Instead of using $p(r_{+1}|\hat{\theta},\Phi)$, the predictive density can be obtained as

$$p(r_{+1}|\Phi) = \int p(r_{+1}|\theta,\Phi)p(\theta|\Phi)\,d\theta. \tag{3}$$

The posterior distribution of θ, $p(\theta|\Phi)$, is proportional to the product of the prior distribution and the likelihood function,

$$p(\theta|\Phi) \propto p(\theta)L(\theta;\Phi). \tag{4}$$

One approach to specifying the likelihood function is to assume that, in each period, the joint distribution of the excess returns on the investable assets is multivariate normal with parameters E and V. If the prior distribution of $\theta \equiv (E,V)$ is noninformative and all investable assets have return histories of the same length, the resulting predictive density is a multivariate Student t and the tangency portfolio weights are the same as the weights

[8] Throughout the paper, p is a generic notation for any probability density function.

obtained when E and V are simply replaced by their maximum-likelihood estimates. Such an approach corresponds to our approach when $N = 0$ and the investor seeks an optimal mix of the benchmark portfolios.

Stambaugh (1997) assumes a standard noninformative prior on θ and derives closed-form expressions for the first two moments of the predictive density when the assets have return histories that differ in length. In our framework, the K benchmarks may have longer histories than the N assets. When our investor is completely skeptical about the pricing model implied by the K benchmarks, the results essentially coincide with the results obtained in Stambaugh's unequal-history framework. The expressions for the first two moments of the predictive density $p(r_{+1}|\Phi)$, $\tilde{E}$ and $\tilde{V}$, are given in Appendix A. As in the equal-history framework, estimation risk is included in $\tilde{V}$, which exceeds the maximum-likelihood estimate of the covariance matrix by a positive definite matrix. Unlike in the equal-history framework, the weights in the tangency portfolio are affected by estimation risk and in general differ from the weights produced by the maximum-likelihood estimates of E and V. Although more complicated approaches based solely on the data can be constructed, our designation "data-based approach" refers to the approach that obtains the weights using the moments given in Appendix A.

The remainder of Section I proposes an alternative methodology for obtaining $\tilde{E}$ and $\tilde{V}$. Unlike the data-based approach described above, the new methodology allows the investor to use an asset pricing model and incorporate his prior degree of confidence in the model's pricing abilities. Regardless of the approach used to obtain $\tilde{E}$ and $\tilde{V}$, the $(N + K) \times 1$ vector of the weights in the portfolio with the maximum Sharpe ratio is

$$w^* = \frac{\tilde{V}^{-1}\tilde{E}}{\iota'_{N+K}\tilde{V}^{-1}\tilde{E}}. \tag{5}$$

The above expression is a standard mean-variance result, which requires the return on the riskless asset to be smaller than the return on the global minimum variance portfolio of risky assets. A risk-averse mean-variance investor optimally chooses a portfolio with the maximum Sharpe ratio. For simplicity, the examples presented in this paper focus on the familiar mean-variance case and calculate the optimal portfolio weights using equation (5). However, our procedure can be used to construct the entire predictive distribution of returns on the investable assets, not only its first two moments. As a result, the portfolio choice problem in equation (2) can also be solved for utility functions that involve higher order moments such as skewness and kurtosis.

Note that our investor should not be viewed as a representative investor. The equilibrium in asset markets cannot be supported if all investors have the same perception of the likelihood function and identical nondogmatic prior beliefs. For example, with identical imperfect beliefs in the CAPM, all investors would deviate from the market portfolio in the direction pointed by

the data. If all investors perceive the same likelihood, the existence of an equilibrium requires heterogeneity of investors' prior beliefs, with some prior beliefs about mispricing centered at nonzero values.

B. Likelihood

Suppose that L returns on K benchmark portfolios are available and denote the $L \times K$ matrix of those returns in excess of r_f by F^L. Let F_t denote the tth row of F^L, $t = 1,\ldots,L$. Also, suppose that $T \leq L$ returns on N risky assets are available in the most recent periods $t = L - T + 1,\ldots,L$. Let R denote the $T \times N$ matrix of those returns in excess of r_f. Let R_t denote the $(t - L + T)$th row of R, $t = L - T + 1,\ldots,L$, and let F^T denote the $T \times K$ submatrix of F^L corresponding to the same period as R. The multivariate regression of the asset returns on the benchmark returns can be written as

$$R = XB + U, \qquad \text{vec}(U) \sim N(0, \Sigma \otimes I_T), \tag{6}$$

where $X = [\iota_T \; F^T]$, "vec" denotes an operator that stacks the columns of a matrix into a vector, "$\otimes$" denotes the Kronecker product, and

$$B = \begin{bmatrix} \alpha' \\ B_2 \end{bmatrix} \tag{7}$$

is a $(K + 1) \times N$ matrix containing the $N \times 1$ vector of the intercepts α and the $K \times N$ matrix of the slopes (benchmark loadings) B_2 from the regression. The rows of the disturbance matrix U are assumed to be serially uncorrelated and homoskedastic with an $N \times N$ covariance matrix Σ. Also denote the $N(K + 1) \times 1$ vector

$$b = \text{vec}(B') = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \tag{8}$$

where $\beta = \text{vec}(B_2')$ is an $NK \times 1$ vector.

The benchmark returns are assumed to be distributed as i.i.d. normal,

$$F_t \sim N(E_F, V_F), \tag{9}$$

where E_F is $1 \times K$ and V_F is $K \times K$. The benchmark returns are also assumed to be independent over time and independent of U.

The assumptions about the stochastic behavior of returns imply that the likelihood function for the parameters (B, Σ, E_F, V_F) can be factored into a product of two normal likelihood functions,

$$p(R, F^L | B, \Sigma, E_F, V_F) = p(R | F^T, B, \Sigma) p(F^L | E_F, V_F). \tag{10}$$

The likelihood function for the regression parameters is

$$p(R|F^T,B,\Sigma) \propto |\Sigma|^{-T/2}\exp\{-\tfrac{1}{2}\mathrm{tr}(R-XB)'(R-XB)\Sigma^{-1}\}$$
$$\propto |\Sigma|^{-T/2}\exp\{-\tfrac{1}{2}\mathrm{tr}\,S\Sigma^{-1} - \tfrac{1}{2}\,\mathrm{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\}, \qquad (11)$$

where $S = (R - X\hat{B})'(R - X\hat{B})$, $\hat{B} = \begin{bmatrix}\hat{\alpha}'\\ \hat{B}_2\end{bmatrix} = (X'X)^{-1}X'R$, tr is the trace operator, $|\Sigma|$ denotes the determinant of Σ, and $\propto$ means "proportional to." The likelihood function for the benchmark moments is

$$p(F^L|E_F,V_F) \propto |V_F|^{-L/2}\exp\{-\tfrac{1}{2}\mathrm{tr}(F^L - \iota_L E_F)'(F^L - \iota_L E_F)V_F^{-1}\}$$
$$\propto |V_F|^{-L/2}\exp\left\{-\tfrac{1}{2}\mathrm{tr}\,QV_F^{-1} - \frac{L}{2}\,\mathrm{tr}(E_F - \hat{E}_F)'(E_F - \hat{E}_F)V_F^{-1}\right\}, \qquad (12)$$

where $Q = (F^L - \iota_L\hat{E}_F)'(F^L - \iota_L\hat{E}_F)$ and $\hat{E}_F = (1/L)\sum_{t=1}^{L}F_t$.

C. Prior

The prior distribution of the parameters (B,Σ,E_F,V_F) is specified to be noninformative about all of the parameters except for α, the first row of the matrix B. The decision maker is assumed to have no prior information about the moments of the benchmark returns, the benchmark loadings, and the residual covariance matrix. The only prior information is about the assets' mispricing α, which corresponds to a certain degree of belief in the validity of the asset pricing model. For example, if the CAPM is considered, although there is no prior information about the location of the market portfolio in a mean-variance space, there may be some prior belief about the market portfolio's degree of inefficiency. The regression parameters and the benchmark moments are assumed to be independent in the prior:

$$p(B,\Sigma,E_F,V_F) = p(B,\Sigma)p(E_F,V_F). \qquad (13)$$

The prior on the regression parameters B and Σ is a normal-inverted-Wishart prior:

$$b|\Sigma \sim N(\bar{b},\Psi(\Sigma)) \qquad (14)$$

$$\Sigma^{-1} \sim W(H^{-1},\nu), \qquad (15)$$

where

$$\bar{b} = \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix} \tag{16}$$

$$\Psi(\Sigma) = \begin{bmatrix} \sigma_\alpha^2 \left(\frac{1}{s^2} \Sigma \right) & 0 \\ 0 & \Omega \end{bmatrix} \tag{17}$$

$$E(\Sigma) = s^2 I_N. \tag{18}$$

The notation $W(H^{-1},\nu)$ stands for a Wishart distribution with parameter matrix H^{-1} and ν degrees of freedom. The prior is made essentially noninformative about Σ by setting ν equal to 15, so that the prior contains only about as much information as a sample of 15 observations. $E(\Sigma)$ denotes the prior expectation of Σ. From the properties of the inverted Wishart distribution (see, e.g., Anderson (1984)), $E(\Sigma) = H/(\nu - N - 1)$, where $H = s^2(\nu - N - 1)I_N$ is consistent with equation (18). For simplicity, equation (18) assumes that, in the prior, the regression residuals are uncorrelated and the residual variances are equal across assets. In the posterior, sample evidence overwhelms the noninformative prior on Σ, so the residuals are in general correlated and the variances are unequal across assets.

In equation (17), $\Psi(\Sigma)$ is an $N(K+1) \times N(K+1)$ matrix, whose (1,1) block is the $N \times N$ prior covariance matrix of α conditional on Σ. The (1,2) and (2,1) blocks of $\Psi(\Sigma)$ are zero matrices corresponding to the prior covariances between the intercepts and the slopes from the regression in equation (6). These prior covariances are assumed to be zero for tractability reasons, unlike in Pástor and Stambaugh (1999). The (2,2) block of $\Psi(\Sigma)$ is the $NK \times NK$ prior covariance matrix Ω of the regression slopes β. The prior is made noninformative about β by setting Ω equal to a diagonal matrix with extremely large diagonal elements. Since $\bar{b}$ is independent of Σ, the unconditional prior covariance matrix of b equals

$$V_b \equiv \text{cov}(b,b') = E\{\Psi(\Sigma)\} = \begin{bmatrix} \sigma_\alpha^2 \left(\frac{1}{s^2} E(\Sigma) \right) & 0 \\ 0 & \Omega \end{bmatrix} = \begin{bmatrix} \sigma_\alpha^2 I_N & 0 \\ 0 & \Omega \end{bmatrix}, \tag{19}$$

where I_N denotes an identity matrix of rank N. As a result, the marginal prior distribution for the $N \times 1$ vector α of asset mispricings has a mean of $\bar{\alpha}$ and a covariance matrix of $\sigma_\alpha^2 I_N$.[9]

[9] The elements of α are uncorrelated in the prior, but they are in general correlated in the posterior. The marginal prior for α is easily shown to be a Student t distribution with $\nu - N + 1$ degrees of freedom. In the empirical examples with $N = 1$, the prior for α therefore has $\nu = 15$ degrees of freedom and its 95 percent critical values are $\pm 2.1\sigma_\alpha$.

If the benchmark returns are constructed as excess returns or returns on zero-investment positions, as is the case throughout the paper, then the pricing model implies that α is an N-vector of zeros (see Huberman, Kandel, and Stambaugh (1987)). If the elements of α are a priori centered at the model-predicted value of zero (i.e., if $\bar{\alpha}$ is a zero vector), the value of σ_α represents a prior degree of belief that the pricing model holds. An investor's choice of σ_α may be influenced by many factors, including the results of the studies that test the implications of the model. For example, if most previous studies are unable to reject the model, the investor might choose to specify a low value for σ_α. Hypothesis tests could therefore play a role in decision making through the specification of the prior degree of belief in the model. Note that if the prior degree of belief in the model is formed based on the result of a hypothesis test, the data used to update the prior should not overlap with the data used to conduct the hypothesis test.

The elements of α can also be centered at nonzero values, such as the forecasts of α provided by a security analyst (see Treynor and Black (1973)). In that case, σ_α no longer represents a degree of belief in the model, but rather a degree of belief in the accuracy of the forecast. This case is illustrated in Section V.

Note that, in equation (17), the conditional prior covariance matrix of α is proportional to the residual covariance matrix Σ. The motivation for the prior link between the intercepts and the residual covariance matrix is the same as in Pástor and Stambaugh (1999) and comes from MacKinlay (1995). It is well known that $\alpha'\Sigma^{-1}\alpha$ is the difference between two maximum squared Sharpe ratios—one obtainable by combining the N assets with the K benchmark portfolios, and the other by combining only the benchmark portfolios (see Gibbons et al. (1989)). If α is not linked to Σ, then, as argued by MacKinlay, very large Sharpe ratios could potentially be obtained by combining the assets with the benchmarks. If $\bar{\alpha}$ is a zero vector, our link between α and Σ reduces the probability of very large Sharpe ratios by adjusting the prior variances of the elements of α downward if the diagonal elements of Σ are small. If the conditional covariance matrix of β were also proportional to Σ, our prior on the regression parameters would be the natural conjugate prior. Nevertheless, this covariance matrix is made independent of Σ in the prior, since there is no theoretical reason why it should be dependent.[10]

The prior on the benchmark moments E_F and V_F is a standard diffuse prior for the parameters of a multivariate normal distribution:

$$p(E_F, V_F) \propto |V_F|^{-(K+1)/2}. \tag{20}$$

This prior reflects noninformative beliefs about E_F and V_F and is discussed in detail in Box and Tiao (1973).

[10] MacKinlay and Pástor (1998) impose a stronger form of the $\alpha - \Sigma$ link in the likelihood function for a larger number of assets. They find that imposing the link can lead to more precise expected return estimates and improved portfolio selection.

D. Predictive Return Density

Recall that Bayesian portfolio selection is based on the predictive density of the (excess) returns on the investable assets. Let F_{L+1} denote the $1 \times K$ vector of the next-period benchmark returns. The predictive density of the benchmark returns equals

$$p(F_{L+1}|F^L) = \int p(F_{L+1}|E_F,V_F,F^L)p(E_F,V_F|F^L)\,dE_F dV_F. \tag{21}$$

This predictive density corresponds to a multivariate Student t distribution with $L - K$ degrees of freedom, as shown in Appendix B.

Let R_{L+1} denote the $1 \times N$ vector of the next-period returns on the nonbenchmark assets. The predictive density of the asset returns is

$$\begin{aligned} p(R_{L+1}|\Phi) &= \int p(R_{L+1}|F_{L+1},\Phi)p(F_{L+1}|\Phi)\,dF_{L+1} \\ &= \int p(R_{L+1}|B,\Sigma,F_{L+1},\Phi)p(B,\Sigma|\Phi)p(F_{L+1}|F^L)\,dF_{L+1}dBd\Sigma. \end{aligned} \tag{22}$$

The density $p(R_{L+1}|B,\Sigma,F_{L+1},\Phi)$ is normal since

$$R_{L+1} = X_{L+1}B + U_{L+1}, \qquad U_{L+1} \sim N(0,\Sigma), \tag{23}$$

where $X_{L+1} = [1\ F_{L+1}]$. The predictive benchmark return density $p(F_{L+1}|F^L)$ is discussed in Appendix B and the posterior density $p(B,\Sigma|\Phi)$ is discussed in Appendix C. It is possible to make draws of R_{L+1} from the predictive density (equation (22)) as follows. First, draw the benchmark returns F_{L+1} as described in Appendix B. Second, draw the regression parameters B and Σ using Gibbs sampling as described in Appendix C. Finally, draw the asset returns R_{L+1} conditional on B, Σ, and F_{L+1} from the normal density implied by equation (23). A large number of such draws of R_{L+1} yield a simulated predictive distribution of the asset returns. The first two moments of the joint predictive density of the returns on the $(N + K)$ investable assets, $\tilde{E}$ and $\tilde{V}$, are estimated (to an arbitrary degree of precision) across the predictive draws of $[R_{L+1}\ F_{L+1}]$, except that the predictive benchmark moments are taken directly from equations (B5) and (B6) in Appendix B. The optimal weights in the investable assets are computed as shown in equation (5).

In order to present some intuition about the mean of the predictive return density, it is convenient to consider the simplest case with one benchmark and one nonbenchmark asset ($K = 1$, $N = 1$). In this case, the regression in equation (6) can be written as

$$R_t = \alpha + \beta F_t + u_t, \qquad u_t \sim N(0,\sigma^2), \qquad t = L - T + 1,\ldots,L. \tag{24}$$

The posterior means of α and β can be shown to be equal to

$$\tilde{\alpha} = (1 - w_{\hat{\alpha}})\bar{\alpha} + w_{\hat{\alpha}}\hat{\alpha} \tag{25}$$

$$\tilde{\beta} = \hat{\beta} + \xi. \tag{26}$$

In the above,

$$w_{\hat{\alpha}} = \frac{\hat{\text{var}}(F_t)}{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t^2 + \hat{\text{var}}(F_t)} \tag{27}$$

$$\xi = \frac{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t}{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t^2 + \hat{\text{var}}(F_t)}(\hat{\alpha} - \bar{\alpha}), \tag{28}$$

where $\bar{F}_t = (1/T)\sum_{t=L-T+1}^{L} F_t$, $\bar{F}_t^2 = (1/T)\sum_{t=L-T+1}^{L} F_t^2$, $\hat{\text{var}}(F_t) = \bar{F}_t^2 - (\bar{F}_t)^2$, and $E(\sigma^2)$ stands for the prior expected value of the residual variance σ^2. The weight $w_{\hat{\alpha}}$ indicates how much attention the investor pays to the sample violation of the model, $\hat{\alpha}$. In line with intuition, $w_{\hat{\alpha}}$ increases as the belief in the model becomes weaker (as σ_α increases), as the amount of sample evidence increases (as T increases), and as the expected precision of $\hat{\alpha}$ increases (as $E(\sigma^2)$ decreases).

As the value of ξ in equation (28) is typically very small relative to $\hat{\beta}$, the posterior mean of β in equation (26) is close to $\hat{\beta}$. Intuitively, since there is no prior information about β, the revised beliefs about β center on the sample estimate. Also note that the absolute value of ξ increases as σ_α decreases. For $\sigma_\alpha = 0$, the market model regression is essentially run without the intercept. Since $\tilde{\beta}$ is (in reasonable samples) very close to $\hat{\beta}$ even for $\sigma_\alpha = 0$, whether the regression is run with or without the intercept has little effect on the resulting estimate of β.

The mean $E(R_{L+1}|\Phi)$ of the predictive density of asset returns equals $\tilde{\alpha} + \tilde{\beta}\hat{E}_F$, where $\hat{E}_F$ is a long-series average of benchmark returns taken from equation (12). The posterior mean of α in equation (25) is simply a weighted average of the prior mean of α and its sample estimate. It follows for $\bar{\alpha} = 0$ that the mean of the predictive density is close (up to a small value of $\xi\hat{E}_F$) to a weighted average of $\hat{\alpha} + \hat{\beta}\hat{E}_F$ and $\hat{\beta}\hat{E}_F$. The value of $\hat{\alpha} + \hat{\beta}\hat{E}_F$ is close to $\hat{\alpha} + \hat{\beta}\bar{F}_t$, the sample mean of the nonbenchmark asset returns over the most recent T periods, unless there is a substantial difference between the sample means of the benchmark returns over the most recent L and T periods. That is, the sample mean of the asset returns is essentially shrunk toward a simple model estimate of the expected return, $\hat{\beta}\hat{E}_F$. Hence the expected return estimate used in portfolio selection is close to a shrinkage estimator of

expected return introduced to portfolio analysis by Jobson, Korkie, and Ratti (1979) and later applied by Jorion (1985, 1986, 1991) and Frost and Savarino (1986). In those studies, however, the prior distribution ignores the potential usefulness of asset pricing models. As a result, the sample mean is shrunk toward a value that is not related to an asset pricing model, such as a grand mean of asset returns. In contrast, the sample means of the nonbenchmark asset returns in this study are shrunk toward values implied by finance theory. The sample means of the benchmark returns are not shrunk. If it is believed that those sample means contain substantial estimation error, the noninformative prior in equation (20) can be replaced by an informative prior.

The studies cited in the previous paragraph find that portfolio performance can be improved by specifying an informative prior that reduces the estimation error in the expected return estimate. This reduction is helpful since the weights in equation (5) are known to be quite sensitive to changes in the expected return estimate. A substantial part of the estimation error in a sample mean is due to the error in $\hat{\alpha}$. When the prior belief in the model is strong (i.e., when σ_α is small), $w_{\hat{\alpha}}$ in equation (27) is close to zero and the resulting expected return estimate contains only a small fraction of the estimation error in $\hat{\alpha}$. Therefore, instead of resorting to approaches that may not be easy to justify economically (such as imposing artificial constraints on asset holdings or shrinking sample means to an ad hoc value), it is possible to reduce the sampling variability of the optimal portfolio weights by specifying at least a mild degree of belief in the model in the form of a relatively small σ_α.

II. Data

The focus of the empirical analysis is to investigate the extent to which optimal holdings depart from the market portfolio. Although the methodology presented in the previous section is applicable when K benchmarks are mixed with N nonbenchmark assets, the empirical examples presented in the following sections are restricted to $K = 1$, where the single benchmark is the market portfolio.

In the empirical examples, different assets are optimally combined with the U.S. market portfolio. In Section III, which investigates the home bias puzzle, the nonmarket asset is a portfolio of foreign stocks. Section IV looks into investing based on value and size by combining the Fama–French book-to-market and size portfolios with the market. In Section V, which explores the effect of including prior beliefs that depart from the model, a portfolio of small stocks is combined with the market. A brief description of the data follows.

The proxy for the market portfolio used throughout the study is the value-weighted portfolio of all stocks listed on the New York Stock Exchange (VW NYSE). The monthly returns on VW NYSE are obtained from the Center for Research in Security Prices (CRSP). The average return on VW NYSE in

excess of the return on a one-month Treasury bill in January 1926 through December 1996 is 0.66 percent per month, and the standard deviation is 5.49 percent per month.

The portfolio of foreign stocks is the "World-Except-U.S." portfolio (WXUS) provided by Morgan Stanley Capital International. This portfolio is a value-weighted portfolio of the 22 most developed equity markets outside the United States. On average, 60 percent of the market capitalization in each country is included in the index. The U.S. dollar returns on WXUS including reinvested dividends are obtained from Datastream. Between January 1970 and December 1996, the mean monthly return in excess of the return on a one-month U.S. Treasury bill is 0.57 percent, and the standard deviation 4.95 percent.

The Fama–French book-to-market portfolio (HML) buys stocks with a high ratio of book value to market value and shorts the same dollar amount of stocks with a low ratio. The size portfolio ("small minus big," or SMB) buys small-cap stocks and shorts the same dollar amount of large-cap stocks. For more details on the construction of HML and SMB, see Fama and French (1993). This study uses a series of returns on HML and SMB that begins in July 1927.[11] Over the 70-year period up to December 1996, the average return on HML, sometimes referred to as the *value premium*, is 0.45 percent per month and the average return on SMB, sometimes referred to as the *size premium*, is 0.17 percent per month. The standard deviations of the HML and SMB returns are 3.12 percent and 3.26 percent, respectively.

The small stock portfolio analyzed in this study is the "9-10 Fund" administered by Dimensional Fund Advisors (DFA). The DFA 9-10 Fund, launched in 1982, is a passive fund designed to closely follow the returns on small-capitalization stocks. This set of small-cap stocks roughly corresponds to the stocks in the bottom two size-based deciles provided by the CRSP.[12] The CRSP's ninth and tenth deciles contain those NYSE, AMEX, and NASDAQ-traded stocks whose market capitalization falls into the bottom two deciles based on NYSE breakpoints. The CRSP 9-10 index therefore includes approximately the smallest 20 percent of all stocks, and DFA loosely tracks the return on these small stocks. DFA's mean excess return during January 1982 through December 1996 is 0.72 percent per month, and the standard deviation 4.88 percent.

III. Home Bias

Investors tend to hold a substantially larger proportion of their equity in domestic equities than is suggested either by the weight of their country in the value-weighted world equity portfolio or by the sample return moments used in the standard mean-variance framework. This phenomenon is docu-

[11] The data were generously provided by Ken French.

[12] For a detailed description of the DFA 9-10 Fund and its differences from the CRSP 9-10 index, see Keim (1999).

mented in a number of studies and is often referred to as the "home bias puzzle".[13] For example, U.S. investors' foreign equity holdings account for only about eight percent of their total equity holdings.[14] In contrast, a simple mean-variance illustration based on the sample moments of returns, such as in Lewis (1999) and Britten-Jones (1999), implies that U.S. investors' optimal weight in foreign equities is about 40 percent. Hence, the point estimates of the mean and the covariance matrix of returns suggest that U.S. investors would benefit by increasing the extent of their international diversification.

On the other hand, evidence presented in the recent studies by Bekaert and Urias (1996), Gorman and Jorgensen (1997), and Britten-Jones (1999) indicates that the U.S. investors' gains from international diversification may not be statistically significant. Britten-Jones and Gorman and Jorgensen cannot reject the hypothesis that the global mean-variance efficient portfolio has no exposure to non-U.S. stocks. In other words, global mean-variance efficiency of the U.S. market portfolio is not rejected. Hence, the conclusion about the benefits of international diversification based on the hypothesis test is different from the conclusion based on the point estimates. The disagreement in the conclusions from these two different perspectives begs for further investigation.

This paper assesses the sample evidence on home bias and the benefits of international diversification from an asset allocation perspective. "Explaining" the home bias puzzle is not the purpose of this analysis. Sensible explanations of home bias would certainly include some aspects omitted from our analysis, such as information asymmetries, transaction costs, the possibility of investing in foreign index futures, etc. Instead, our analysis is intended to provide a different way of looking at the data, based on a simple decision problem of a U.S. investor. A U.S. investor is allowed to invest in two assets, a foreign stock portfolio, and the U.S. market portfolio, the latter of which plays the role of a benchmark asset. The choice of the U.S. market as a benchmark is motivated by the previous studies, which assess the benefits of international diversification by testing whether the U.S. market is globally mean-variance efficient. The investor has a certain degree of belief in the global efficiency of the U.S. market portfolio, sometimes referred to as a belief in the domestic CAPM. This section investigates how strongly a typical U.S. investor must believe in the domestic CAPM in order to justify the level of his domestic equity holdings.

The investor would benefit from investing in foreign equities if and only if the intercept α from a regression of foreign equity excess returns on U.S. equity excess returns is positive. Recall that the foreign stock portfolio used

[13] One of the first studies that documents the home bias puzzle is Levy and Sarnat (1970). Recent studies of home bias and the benefits of international diversification include Kang and Stulz (1997), Errunza, Hogan, and Hung (1999), and Lewis (1999).

[14] This recent estimate is based on Bohn and Tesar (1996). Bohn and Tesar also report that the home bias has eroded with time; 15 years earlier, the fraction of foreign shares in U.S. investors' equity portfolios was only 2.5 percent. French and Poterba (1991) report a six percent estimate as of December 1989.

in this study is Morgan Stanley's World-Except-U.S. portfolio (WXUS), as described earlier. Based on the period from January 1973 through December 1996, the results from the regression of the excess returns of WXUS on the excess returns of VW NYSE are $\hat{\alpha} = 0.24$ percent per month (2.89 percent annualized), $\hat{\beta} = 0.56$, $\hat{\sigma}^2 = 0.0018$, and $R^2 = 25.27$ percent. The standard error of WXUS's $\hat{\alpha}$ is 3.05 percent per year, resulting in a t-statistic for $\hat{\alpha}$ of 0.95. In view of the positive estimate of α and in the absence of a prior belief that α is negative, portfolio theory indeed suggests that a U.S. investor with some skepticism about the domestic CAPM should invest in foreign stocks.

It is shown below that the statistical insignificance of $\hat{\alpha}$ does not prevent a Bayesian investor from investing in foreign stocks. In contrast, if the investment decision is based solely on testing the hypothesis that $\alpha = 0$, an investor facing an insignificant $\hat{\alpha}$ approaches portfolio selection as if $\alpha = 0$. In other words, since the hypothesis is not rejected, the investor maintains his implicit prior belief that $\alpha = 0$. If the test revealed $\hat{\alpha}$ to be significantly different from zero, such an investor would approach portfolio selection as if $\alpha = \hat{\alpha}$, since no other information about α is available. That is, the investor's implicit prior would be revised all the way to $\alpha = \hat{\alpha}$. Such "binary" updating of the prior seems less appealing than the "smooth" updating provided in our Bayesian approach.

Before looking at the optimal allocations, let us make a short detour to explain how the prior parameters are specified. The values of ν and Ω are specified such that the priors of Σ and β are noninformative. The value of s^2 is estimated from a 36-month period prior to the sample period. The return history on WXUS is available back to January 1970. The regression of the excess returns of WXUS on the excess returns of VW NYSE from January 1970 through December 1972 produces a sample estimate of the residual variance equal to 0.0013, so WXUS's $s^2 (\equiv E(\sigma^2))$ is set equal to 0.0013. The values of $\bar{\alpha}$ and σ_α are varied in order to express different beliefs about the expected mispricing and about the validity of the domestic CAPM. The CAPM predicts that α is equal to zero. For $\bar{\alpha} = 0$, as σ_α grows, so does the degree of skepticism that the CAPM holds.

Part A of Table I reports the optimal percentage weight in WXUS when it is combined with VW NYSE to form the portfolio with the maximum Sharpe ratio. The resulting portfolio is optimal from the viewpoint of an investor whose prior beliefs about expected mispricing and about the validity of the CAPM are varied across a number of values. WXUS's $\bar{\alpha}$ takes on the model-predicted value of zero and the values of ± 5 percent per year. Part B of the table reports the corresponding perceived maximum Sharpe ratios, which indicate how valuable (ex ante) the optimal portfolios are to the investor. Parts C and D report the posterior means and standard deviations of WXUS's α.

With perfect confidence in the domestic CAPM, represented by $\bar{\alpha} = \sigma_\alpha = 0$, nonzero values of $\hat{\alpha}$ are believed to be due solely to sampling error. Therefore, the optimal weight in WXUS is zero and all the wealth is invested in the U.S. market portfolio. As σ_α grows while $\bar{\alpha} = 0$, the belief in the model becomes weaker and the investor pays some attention to WXUS's positive

Table I

Optimal Weight in Foreign Stocks in a Two-Asset Portfolio with the U.S. Market

The foreign stock portfolio is the "M.S.C.I. World-Except-U.S." (WXUS) portfolio provided by Morgan Stanley Capital International. The U.S. market portfolio is proxied by the value-weighted portfolio of NYSE stocks (VW NYSE). The optimal weight in the foreign stock portfolio is given by the first element of $\tilde{V}^{-1}\tilde{E}/\iota_2'\tilde{V}^{-1}\tilde{E}$ and the maximum Sharpe ratio by $\sqrt{\tilde{E}'\tilde{V}^{-1}\tilde{E}}$, where $\tilde{E}$ and $\tilde{V}$ are the first two moments of the predictive density of the returns on the investable assets, obtained using our "model-and-data-based" methodology proposed in Section I. The maximum Sharpe ratio is the ex ante Sharpe ratio perceived by an investor who forms an optimal portfolio of the two assets. The intercept from the regression of the excess returns on WXUS on the excess returns on VW NYSE is denoted by α. The prior distribution of α has a mean of $\bar{\alpha}$ and a standard deviation of σ_α. Over the sample period of January 1973 through December 1996, the OLS estimates of the market model regression coefficients are $\hat{\alpha} = 2.89$ percent per year (with a standard error of 3.05 percent per year) and $\hat{\beta} = 0.56$. The values of $\bar{\alpha}$, σ_α, and the posterior mean and standard deviation of α are annualized percentage values.

Expected Prior Mispricing ($\bar{\alpha}$)	Prior Standard Deviation of α (σ_α)						
	0	1%	2%	3%	5%	10%	∞
A. Optimal percentage weight in foreign stocks							
0	0.00	7.65	20.57	29.92	38.99	44.70	47.00
+5%	70.54	67.77	62.34	57.69	52.42	48.64	47.00
−5%	−174.10	−119.69	−48.55	−8.67	23.22	40.55	47.00
B. Maximum Sharpe ratio							
0	0.1208	0.1210	0.1227	0.1251	0.1287	0.1317	0.1331
+5%	0.1548	0.1515	0.1456	0.1412	0.1369	0.1342	0.1331
−5%	0.1542	0.1426	0.1267	0.1210	0.1232	0.1295	0.1331
C. Posterior mean of α ($\tilde{\alpha}$)							
0	0.00	0.39	1.10	1.68	2.30	2.72	2.89
+5%	5.00	4.72	4.20	3.77	3.33	3.02	2.89
−5%	−5.00	−3.95	−1.99	−0.41	1.27	2.41	2.89
D. Posterior standard deviation of α							
0	0.00	1.12	1.89	2.33	2.72	2.95	3.05
+5%	0.00	1.12	1.88	2.33	2.72	2.95	3.05
−5%	0.00	1.13	1.90	2.34	2.72	2.96	3.05

sample estimate of α. As a result, the optimal weight in WXUS increases with σ_α, up to the level of 47.00 percent for $\sigma_\alpha = \infty$. This value is close to 46.48 percent, the value predicted by the data-based approach introduced in Section I.A.[15] The table also shows what happens between the two extremes. If there is some doubt about the CAPM, represented by $\sigma_\alpha > 0$, it is optimal for the investor to account for the positive $\hat{\alpha}$ and invest in WXUS. For example, an investor whose prior belief in the CAPM is represented by $\sigma_\alpha = 1$ percent per year should optimally invest about eight percent of wealth in foreign

[15] The small difference between the two values is due to the fact that the functional forms of the noninformative priors in the two cases are different, which results in small differences in "degree-of-freedom-type" adjustments.

stocks, and an investor with $\sigma_\alpha = 3$ percent per year should invest 30 percent. The actual allocation in foreign stocks observed in the equity holdings of U.S. investors is consistent with σ_α of about one percent, which represents a strong belief in the mean-variance efficiency of the U.S. market portfolio.

The prior with $\sigma_\alpha = 1$ percent seems to be very strong. This prior essentially rules out mispricing of the foreign stock portfolio larger than two percent in absolute value. Moreover, this prior places a large weight on the prior view relative to sample evidence. Part C of Table I shows that, for $\sigma_\alpha = 1$ percent, the posterior mean of α is only 0.39 percent, which is much closer to the prior mean $\bar{\alpha} = 0$ than to the sample value $\hat{\alpha} = 2.89$ percent. The prior on α is strong relative to the sample evidence since it shrinks $\hat{\alpha}$ close to $\bar{\alpha}$. Finally, imagine that the prior with $\sigma_\alpha = 1$ percent were to be formed as a posterior after an investor with no prior information observes a hypothetical sample with the same sample statistics as the observed sample. With no prior information, the posterior standard deviation of alpha from the hypothetical sample is essentially equal to the standard error of $\hat{\alpha}$ in the hypothetical sample. Note that the standard error of WXUS's observed $\hat{\alpha}$ is greater than three percent per year and that the standard error is inversely proportional to the square root of the sample size. In order to make the standard error of $\hat{\alpha}$ from the hypothetical sample as low as one percent per year, the hypothetical sample must be more than nine times as long as the observed sample—that is, more than 216 years long.[16]

The results in Table I provide a "snapshot" view of the home bias issue. Optimal allocations are computed in January 1997 based on the sample period from January 1973 through December 1996. More information on home bias is provided by analyzing the optimal allocations computed in previous years. For every portfolio formation month between January 1978 and January 1997, optimal weights in the foreign stock portfolio are computed based on a sample of returns that ends one month prior to the portfolio formation month.[17] Two different approaches to specifying the starting dates of the sample periods are pursued. In the "cumulative sample period" approach, all sample periods begin in January 1973. In the "10-year rolling sample period" approach, the starting dates roll over time together with the ending dates such that all sample periods are 10 years long. Using rolling sample periods could be useful when there is suspicion that the regression coefficients change over time in a manner that is difficult to model explicitly.

Figure 1 plots the optimal allocations in foreign stocks for every portfolio formation month from January 1978 to January 1997. For each month, four different allocations are reported, corresponding to $\sigma_\alpha = 1$ percent, 2 percent, 3 percent, and ∞. Throughout, $\bar{\alpha} = 0$. The first plot in Figure 1 is based on cumulative sample periods. Note that the allocations in January 1997 cor-

[16] Thanks to Bill Schwert for suggesting this interpretation. In the same spirit, Kandel and Stambaugh (1996) used a hypothetical prior sample to construct prior beliefs in their study of stock return predictability.

[17] When one asset is combined with one benchmark at many different dates, the moments of the predictive density are not computed using the computationally intensive Gibbs sampling procedure described in Appendix C, but are instead approximated based on equations (D4) and (D5) in Appendix D. The approximation is very precise, as explained in Appendix D.

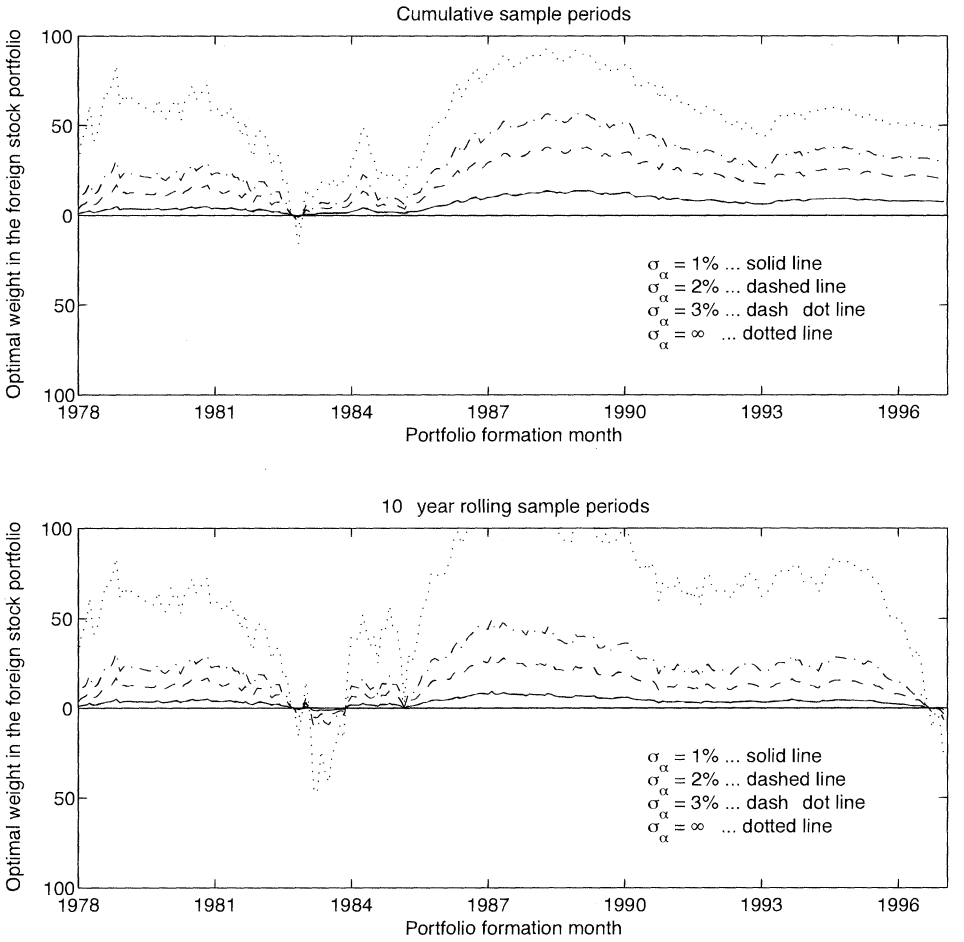

Figure 1. Optimal percentage weights in foreign stocks in a two-asset portfolio with the U.S. market. The foreign stock portfolio is the Morgan Stanley "World-Except-U.S." (WXUS) portfolio. The prior distribution of α, the intercept from the regression of the excess returns on WXUS on the excess returns on the value-weighted NYSE index, has a mean of zero and an annualized standard deviation of σ_α. Each portfolio formation month follows the sample period over which the optimal weights are estimated. In the first plot, the sample periods begin in January 1973 and end in each month between December 1977 and December 1996. In the second plot, the sample periods are moving 10-year windows, except for the first five years, in which the periods begin in January 1973.

respond to those reported in the first row of Table I. In this plot, the optimal weight in foreign stocks is consistently positive, with a minor exception in the last three months of 1982. For $\sigma_\alpha = \infty$, the optimal weights in WXUS tend to be large and unstable, although they never exceed 100 percent. Since $\sigma_\alpha = \infty$ yields essentially the same results as standard mean-variance optimization with sample means and covariances (except for the difference in the lengths of the return histories for VW NYSE and WXUS), it is often the

large weights corresponding to $\sigma_\alpha = \infty$ that are shown as evidence of home bias in U.S. investors' holdings. In the presence of some prior belief in the mean-variance efficiency of the U.S. market portfolio ($\sigma_\alpha < \infty$), the optimal weights in foreign stocks are lower and more stable over time. However, it is interesting that these weights are sometimes substantial even for fairly low values of σ_α. For example, the weights can be as high as 57 percent for $\sigma_\alpha =$ 3 percent and as high as 14 percent for $\sigma_\alpha = 1$ percent.

The second part of Figure 1 plots the optimal weights in WXUS based on 10-year rolling sample periods. The weights in the first five years are identical to those from cumulative regressions, since fewer than 10 years of data are available by then. Although the results from rolling regressions are fairly similar to those from cumulative regressions, three differences emerge. First, the weights for $\sigma_\alpha = \infty$ tend to have more extreme values than in cumulative regressions. With shorter sample periods, $\hat{\alpha}$'s tend to have more extreme values, which results in more extreme weights. Second, the weights obtained for finite values of σ_α are now somewhat smaller relative to those obtained for $\sigma_\alpha = \infty$. Since the sample periods are now shorter, less attention is paid to sample evidence than in the case of cumulative regressions, for a given value of σ_α. Finally, the optimal weights in foreign stocks based on 10-year rolling regressions are decreasing since 1995, and even turn negative in the second half of 1996. In other words, if the first two moments of returns are estimated based on the returns in the preceding decade, home bias disappears. Of course, this result could simply reflect the recent long-lasting bull market in the United States and the fact that 10-year periods are too short to estimate means. Britten-Jones (1999) also obtains an optimal weight in the U.S. equity that exceeds 100 percent based on the past 10 years of data, and shows that the standard error in the weight estimate is huge (about 70 percent).

An alternative approach to investigating the home bias in the equity holdings of U.S. investors is to take the world market portfolio (WRLD) as the benchmark. The optimal weight in the U.S. market in a two-asset portfolio with WRLD depends on the sample mispricing of the U.S. market relative to WRLD as well as on the prior degree of belief in the international CAPM. The returns on WRLD used in the following experiment are the U.S. dollar returns on the world market portfolio provided by Morgan Stanley Capital International, obtained from Datastream. A regression of excess VW NYSE returns on excess WRLD returns in January 1973 through December 1996 yields $\hat{\alpha} = 0.10$ percent per month (1.22 percent annualized), $\hat{\beta} = 0.87$, $\hat{\sigma}^2 =$ 0.0007, and $R^2 = 69.72$ percent. The t-statistic of VW NYSE's $\hat{\alpha}$ is 0.67. With no belief in the efficiency of the world market portfolio ($\sigma_\alpha = \infty$), the positive $\hat{\alpha}$ implies the allocations of 53 percent in VW NYSE and 47 percent in WRLD, which amounts to about 30 percent in non-U.S. stocks.[18] As the belief in the

[18] The 30 percent foreign stock allocation for $\sigma_\alpha = \infty$ differs from the 47 percent allocation obtained previously with the U.S. market as a benchmark. This difference reflects the difference in the lengths of the return histories in the two cases. Recall that the methodology requires the benchmark return history to be no shorter than the nonbenchmark history. With VW NYSE as a benchmark, its return history begins in January 1926. With WRLD as a benchmark, VW NYSE's history begins in January 1973, the beginning of WRLD's history in our sample.

international CAPM becomes stronger (i.e., as σ_α decreases), the optimal allocation moves toward 100 percent in WRLD. As a result, no value of σ_α can produce a weight in foreign stocks smaller than the 30 percent obtained for $\sigma_\alpha = \infty$. Hence this alternative setup, in which the benchmark model is the international CAPM, cannot reconcile the observed allocations with the optimal allocations. Our conclusions related to the home bias puzzle reflect the approach based on the domestic CAPM. As argued earlier, choosing the U.S. market as the benchmark is consistent with the existing literature. The goal is to analyze to what extent a U.S. investor who is fully invested in the U.S. market should diversify internationally after observing the return data.

In summary, we find that a typical U.S. investor's confidence in the domestic CAPM must be very strong to justify his low holdings of foreign stocks. In particular, he must believe that the mispricing of the foreign stock portfolio within the domestic CAPM is no larger than two percent per year. Unless the prior belief of a typical U.S. investor is so strong, home bias remains a puzzle.

IV. Investing Based on Value and Size

A number of studies document that value stocks (stocks with high book-to-market ratios) tend to outperform growth stocks (stocks with low book-to-market ratios). Fama and French (1993) provide such evidence for the United States in the period following July 1963.[19] Davis (1994) shows that value stocks outperform growth stocks also in the period from July 1940 through June 1963. Fama and French (1998) find the evidence of a positive value premium in 12 out of 13 major stock markets around the world. In a recent study, Davis, Fama, and French (1998) document the robustness of the value premium in the United States from July 1929 through June 1997. The authors find that the average return on a value-minus-growth portfolio during the 34 years prior to July 1963 is positive and similar in magnitude to the average return after July 1963.

It might also be informative to examine the estimates of the value premium on a month-by-month basis. The top row in Figure 2 plots the rolling estimates of the value premium in each month between July 1937 and December 1996. The estimates are computed as 10-year moving averages of the returns on HML, the Fama–French book-to-market portfolio. Surprisingly, the value premium estimates are positive in every month. For comparison, the middle row of Figure 2 plots the rolling estimates of the market premium, computed as 10-year moving averages of the excess market returns. The market premium estimates are negative in 49 out of 715 months. Curiously, the ex post value premium is more stable than the market premium over the last 60 years. In contrast, the ex post size premium (computed as

[19] Kothari, Shanken, and Sloan (1995) conjecture that the outperformance of growth stocks by value stocks may be due to a selection bias present in the COMPUSTAT database. Chan, Jegadeesh, and Lakonishok (1995) and Fama and French (1996) argue against this conjecture.

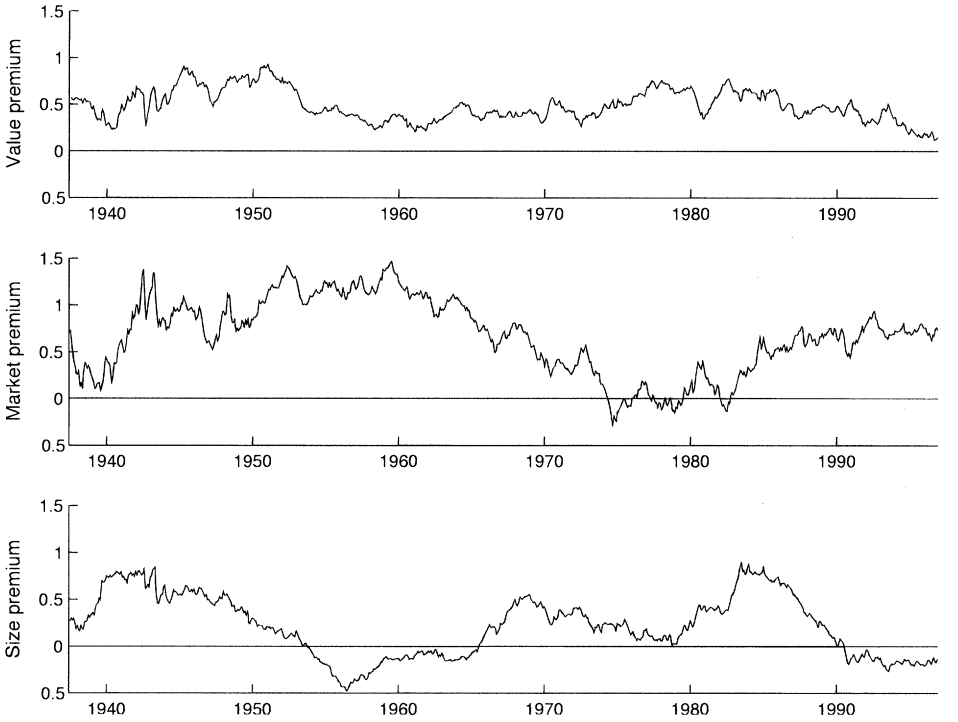

Figure 2. The value premium, the market premium, and the size premium. The premiums are estimated as 10-year moving averages of the returns on the Fama–French HML portfolio, the excess returns on the value-weighted NYSE index, and the returns on the Fama–French SMB portfolio, respectively. The estimates are percentage monthly values.

10-year moving averages of the returns on SMB, the Fama–French size portfolio) plotted at the bottom of Figure 2 is much less stable over time. For example, the premium is negative between 1953 and 1965 as well as throughout the 1990s. Our evidence is consistent with the evidence in Davis et al. (1998) who find that the size premium in the last seven decades is low and not very reliable. It is also interesting that both the value premium and the size premium exhibit a downward trend since 1982. Their current rolling estimates are 0.14 percent and −0.13 percent per month, respectively, compared to 0.45 percent and 0.17 percent per month based on the entire sample. The ex post value premium is statistically significant based on the whole sample ($t = 4.18$), whereas the size premium is not ($t = 1.54$).

One way of assessing the economic significance of the empirical evidence on value and size is to consider the Fama–French portfolios HML and SMB as investable assets. Investing one dollar in HML or SMB is interpreted as investing one dollar in cash and taking a zero-investment position of one dollar long and one dollar short in HML or SMB. Consider an investor with a certain prior degree of belief in the CAPM who is fully invested in the

(U.S.) market portfolio. We investigate what proportion of wealth the investor reallocates in the Fama–French portfolios after he observes their past returns and updates his prior beliefs.

The returns on HML and SMB are available from July 1927 through December 1996. The first five years from this period are used to estimate the prior parameters. In equation (18), s^2 is set equal to 0.0011, the average of the diagonal elements of the sample estimate of the residual covariance matrix from the prior period. This value also happens to be close to the estimates of the residual variances from the sample period, which are equal to 0.0009 for both HML and SMB. The multivariate regression of the returns on HML and SMB on the excess market returns is run across the remaining 774 months. The regression sample estimates for HML and SMB, respectively, are $\hat{\alpha} = [0.37\ 0.04]'$ percent per month ($[4.46\ 0.52]'$ percent annualized), and $\hat{\beta} = [0.14\ 0.24]'$. The annualized standard errors on the $\hat{\alpha}$'s are 1.29 percent and 1.32 percent, resulting in the t-statistics on the $\hat{\alpha}$'s of 3.45 for HML and 0.39 for SMB.

Part A of Table II shows the optimal percentage weights in the book-to-market and size portfolios when these are combined with the market in January 1997. Throughout the table, $\bar{\alpha} = [0\ 0]'$. As σ_α grows, the optimal allocation approaches the allocation computed by the data-based approach, 71.19 percent in HML and 2.23 percent in SMB. The weight in HML moves surprisingly fast toward the large value obtained in the data-based approach. For example, for $\sigma_\alpha = 1$ percent, the optimal weight in book-to-market is already 40 percent, and for $\sigma_\alpha = 2$ percent, the book-to-market position is already 60 percent. In order for the optimal position in HML to be close to zero, the prior belief in the CAPM must be extremely strong, much stronger than $\sigma_\alpha = 1$ percent. The passive strategy of investing only in the market portfolio is clearly dominated (ex ante) by an active strategy that takes a substantial position in the Fama–French book-to-market portfolio. Sample evidence about book-to-market is powerful enough to overwhelm even strong prior beliefs in the CAPM.[20]

The weight in SMB is close to zero for any σ_α primarily because its $\hat{\alpha}$ is small, but also due to some interaction with HML. When HML is excluded and the wealth is allocated only between the market and SMB, the SMB weight for $\sigma_\alpha = \infty$ is 19 percent. Apparently, including HML drives SMB out of the optimal portfolio. When SMB is excluded and the wealth is allocated only between the market and HML, the HML weights are very close to those reported in Table II.[21] Including SMB seems to have no effect on the optimal portfolio of the market and HML.

[20] The finding that sample evidence can overcome even strong priors in portfolio selection is similar in spirit to one of the conclusions of Kandel and Stambaugh (1996). In that study, a Bayesian investor allocating funds between the market portfolio and cash is given sample evidence about the predictability of stock returns. The authors find that this sample evidence can exert a substantial influence on the investor's portfolio decision, even when the investor's prior beliefs are weighted against predictability.

[21] The detailed results are not reported in separate tables to save space, but they are available upon request.

Table II

Optimal Weights in the Fama–French Book-to-Market and Size Portfolios in a Three-Asset Portfolio with the U.S. Market

HML is the Fama–French book-to-market portfolio, which finances a long position in high-book-to-market stocks by a short position in low-book-to-market stocks. SMB is the Fama–French size portfolio, which finances a long position in small-capitalization stocks by a short position in big-capitalization stocks. Investing one dollar in HML or SMB is interpreted as investing one dollar in cash and taking a zero-investment position of one dollar long and one dollar short in HML or SMB. The U.S. market portfolio is proxied by the value-weighted portfolio of NYSE stocks (VW NYSE). The optimal weights in HML and SMB are given by the first two elements of $\tilde{V}^{-1}\tilde{E}/\iota_3'\tilde{V}^{-1}\tilde{E}$ and the maximum Sharpe ratio by $\sqrt{\tilde{E}'\tilde{V}^{-1}\tilde{E}}$, where $\tilde{E}$ and $\tilde{V}$ are the first two moments of the predictive density of the returns on the investable assets, obtained using our "model-and-data-based" methodology proposed in Section I. The maximum Sharpe ratio is the ex ante Sharpe ratio perceived by an investor who forms an optimal portfolio of the three assets. The vector of the intercepts from the regression of the HML and SMB returns on the excess returns on VW NYSE is denoted by α. The prior distribution of both elements of α has a mean of zero and a standard deviation of σ_α. Over the sample period of July 1932 through December 1996, the OLS estimates of the market model regression coefficients are $\hat{\alpha} = [4.46\ 0.52]'$ percent per year (with standard errors of $[1.29\ 1.32]'$ percent per year) and $\hat{\beta} = [0.14\ 0.24]'$. The values of σ_α and the posterior mean and standard deviation of α are annualized percentage values.

	Prior Standard Deviation of α (σ_α)						
Asset	0	1%	2%	3%	5%	10%	∞
A. Optimal percentage weights in book-to-market and size							
HML	0.00	39.94	59.74	65.71	69.23	70.83	71.37
SMB	0.00	1.20	1.90	2.11	2.23	2.28	2.30
B. Maximum Sharpe ratio							
	0.1208	0.1273	0.1460	0.1578	0.1673	0.1724	0.1742
C. Posterior mean of α ($\tilde{\alpha}$)							
HML	0.00	1.44	2.92	3.62	4.11	4.37	4.46
SMB	0.00	0.17	0.34	0.42	0.48	0.51	0.52
D. Posterior standard deviation of α							
HML	0.00	0.74	1.05	1.17	1.24	1.28	1.29
SMB	0.00	0.75	1.07	1.19	1.27	1.31	1.32

The HML weights in optimal portfolios with the market are large for three reasons. First, HML's $\hat{\alpha} = 4.46$ percent per year is large. Second, HML's prior residual variance is small and T is large, implying a large weight on $\hat{\alpha}$ in equation (25). The two observations imply that the posterior mean of α is large. Finally, HML's posterior mean of the residual variance is small because the prior of σ^2 is noninformative and the sample estimate of the residual variance is small. The small posterior mean of the residual variance in combination with the large posterior mean of α results in large weights on HML, as shown in Appendix E.

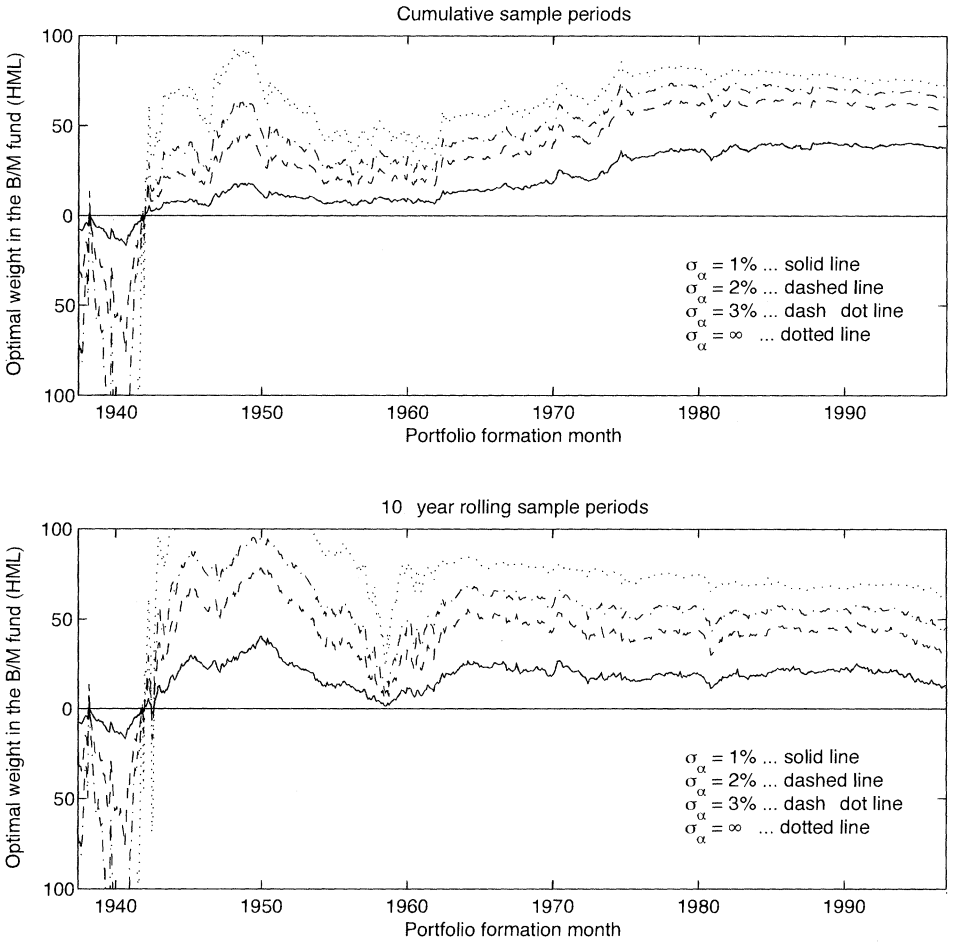

Figure 3. Optimal percentage weights in the Fama–French book-to-market portfolio in a two-asset portfolio with the U.S. market. The prior distribution of α, the intercept from the regression of the returns on the Fama–French book-to-market portfolio (HML) on the excess returns on the value-weighted NYSE index, has a mean of zero and an annualized standard deviation of σ_α. Each portfolio formation month follows the sample period over which the optimal weights are estimated. In the first plot, the sample periods begin in July 1932 and end in each month between June 1937 and December 1996. In the second plot, the sample periods are moving 10-year windows, except for the first five years, in which the periods begin in July 1932.

Figure 3 plots the optimal allocations in HML for every month between July 1937 and January 1997. This figure reveals three striking findings. First, the optimal weights in HML are consistently positive since the early 1940s, based on both cumulative and rolling regressions. In other words, for more than half a century, investors holding the market should also include

book-to-market in their optimal portfolios. Second, the optimal weights in HML are large, even for small values of σ_α and relatively short sample periods used in rolling regressions. For example, the weights for $\sigma_\alpha = 1$ percent average 23 percent for cumulative regressions and 20 percent for 10-year rolling regressions since 1943. That is, even investors with strong beliefs in the CAPM should take substantial positions in HML. This finding indicates that the HML results in Table II are not driven solely by the 70-year length of the sample period. Finally, the optimal weights are quite stable over time, even for large values of σ_α. For example, based on rolling regressions, the optimal weights range from 62 percent to 85 percent for $\sigma_\alpha = \infty$ since 1963. The weights range from 43 percent to 69 percent for $\sigma_\alpha = 3$ percent and from 12 percent to 28 percent for $\sigma_\alpha = 1$ percent during that period. Based on cumulative regressions, the optimal weights range from 31 percent to 41 percent for $\sigma_\alpha = 1$ percent and from 55 percent to 84 percent for $2 \leq \sigma_\alpha \leq \infty$ in the last two decades. Overall, the results in Figure 3 underline the robustness of the finding that optimal portfolios in the last 50 years involve sizable and fairly stable positive positions in the book-to-market portfolio. Since HML's market beta is fairly small, the robust value premium observed in Figure 2 is reflected in robust optimal weights in book-to-market, even if the confidence in the CAPM is very strong.

Figure 4 plots the optimal allocations in SMB for every month between July 1937 and January 1997. There is no consistent pattern in the optimal weights in SMB. For both cumulative and rolling regressions, the weights are rather unstable and change sign several times. As of January 1997, the cumulative regression recommends taking a position of up to one-fifth of the wealth in SMB, whereas the regression based on the last 10 years of data recommends taking a large short position. Between 1966 and 1990, the weights based on rolling regressions are always positive, consistent with the evidence from some earlier studies. However, the weights are mostly negative outside that period, suggesting that the size effect could be specific to certain time periods.[22]

V. Departures of Prior Beliefs from the Model

The examples presented so far consider scenarios in which informative prior beliefs center on the CAPM. This section discusses an example in which prior beliefs depart from the model. If fundamental research by a financial analyst concludes that a certain asset is mispriced by the CAPM, prior beliefs about α can center on a nonzero forecast $\bar{\alpha}$ produced by the analyst. In such a scenario, σ_α no longer represents a degree of confidence in the CAPM,

[22] The size effect is identified in Reinganum (1981) based on the 1963 to 1977 data and in Keim (1983) based on the 1963 to 1979 data. Banz (1981) uses the data since 1936, and notes that the size effect is not very stable through time. Brown, Kleidon, and Marsh (1983) also note the instability of the size effect.

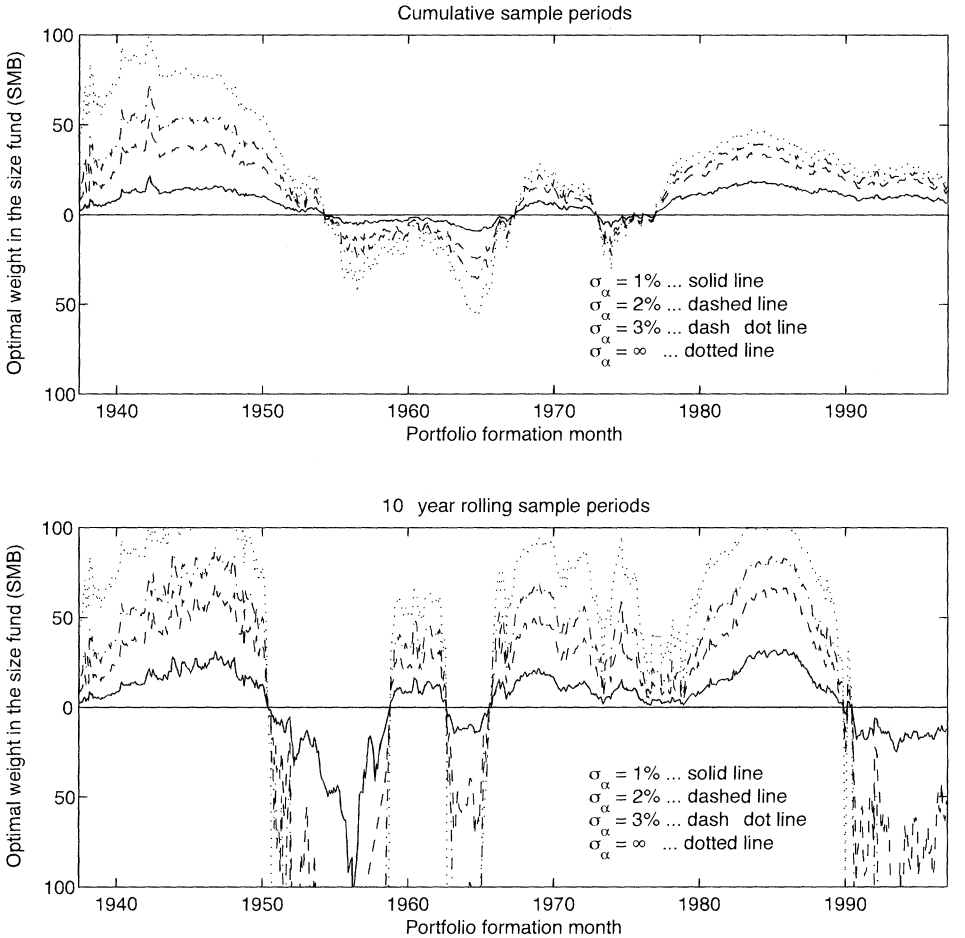

Figure 4. Optimal percentage weights in the Fama–French size portfolio in a two-asset portfolio with the U.S. market. The prior distribution of α, the intercept from the regression of the returns on the Fama–French size portfolio (SMB) on the excess returns on the value-weighted NYSE index, has a mean of zero and an annualized standard deviation of σ_α. Each portfolio formation month follows the sample period over which the optimal weights are estimated. In the first plot, the sample periods begin in July 1932 and end in each month between June 1937 and December 1996. In the second plot, the sample periods are moving 10-year windows, except for the first five years, in which the periods begin in July 1932.

but rather a degree of confidence in the analyst's forecast. The sample mean is no longer shrunk toward the expected return implied by the CAPM, but rather toward the analyst's forecast of the expected return.

This section computes optimal allocations between the small-stock DFA 9-10 Fund (DFA) and the market portfolio. Using all available monthly return data (January 1982 to December 1996) on DFA, the sample estimates from the market model regression are $\hat{\alpha} = -0.07$ percent per month (-0.83 percent annualized), $\hat{\beta} = 0.98$, $\hat{\sigma}^2 = 0.0008$, and $R^2 = 67.77$ percent. The standard

error on DFA's $\hat{\alpha}$ is 2.52 percent per year, resulting in the t-statistic on $\hat{\alpha}$ of -0.33. Over the past 15 years, small stocks fail to outperform the market, both before and after the risk adjustment via the CAPM. In light of this evidence, an investor whose nondogmatic prior beliefs center on the CAPM should optimally short the small stock portfolio.[23] The focus in this section is instead on prior beliefs that depart from the CAPM. For example, the investor may form his prior belief about the mispricing of small stocks based on the pre-1982 data. Based on such earlier data, the studies by Banz (1981), Reinganum (1981), and Keim (1983) document the so-called size effect, according to which small-capitalization stocks outperform large-capitalization stocks.

One way of obtaining a forecast $\bar{\alpha}$ around which to center the prior beliefs about α is to use the pre-1982 series of returns on the CRSP 9-10 index. The description of DFA in Section II suggests that the CRSP 9-10 index can serve as a proxy for DFA in the period before DFA was created. The January 1926 to December 1981 market model regression of monthly excess returns on the CRSP 9-10 index on excess market returns produces an estimated intercept of 2.82 percent per year, with an annualized standard error of 2.10 percent. Therefore, prior beliefs with $\bar{\alpha} = 2.82$ percent are given special attention in our analysis. The sample residual variance from the 1926 to 1982 regression is 0.0020, so DFA's $E(\sigma^2)$ is set equal to this value.

Part A of Table III reports the optimal percentage weights in DFA in combination with VW NYSE, as of January 1997. DFA's $\bar{\alpha}$ takes on the model-predicted value of zero as well as the values of ± 5 percent and 2.82 percent. When $\bar{\alpha} = 0$, the investor shorts DFA in response to its negative $\hat{\alpha}$. If the investor a priori believes that DFA is systematically mispriced by the CAPM ($\bar{\alpha} \neq 0$), such a prior belief is substantially reflected in the posterior. Consider the prior belief that $\bar{\alpha} = 2.82$ percent, formed based on the pre-1982 data. The optimal weight in DFA moves from 131.16 percent for $\sigma_\alpha = 0$ to -40.56 percent for $\sigma_\alpha = \infty$. If the investor's prior belief in $\bar{\alpha} = 2.82$ percent is formed on the basis of the standard error of $\bar{\alpha}$ from the prior period (2.10 percent per year), all the wealth should be invested in DFA and nothing in the market. That is, the prior belief that $\bar{\alpha} = 2.82$ percent is strong enough to overcome the effect of DFA's negative $\hat{\alpha} = -0.83$ percent per year. It appears that security analysis capable of producing nonzero $\bar{\alpha}$ and $\sigma_\alpha < \infty$ could play an important role in asset allocation.

VI. Conclusion

Finance theory can be used to form informative prior beliefs in financial decision-making. This paper develops a portfolio selection methodology that allows a Bayesian investor to include a certain degree of belief in an asset pricing model. In the extreme cases of complete confidence and complete skepticism about the model, the resulting optimal allocations correspond to

[23] Short optimal position in small stocks is broadly consistent with the evidence presented in the second panel of Figure 4. Throughout the 1990s, the optimal weight in the "small-minus-big" size portfolio (SMB) computed from the last 10 years of data is negative.

Table III

Optimal Weight in Small Stocks in a Two-Asset Portfolio with the U.S. Market

The small stock portfolio is the DFA 9-10 Fund, a small stock fund run by Dimensional Fund Advisors. The U.S. market portfolio is proxied by the value-weighted portfolio of NYSE stocks (VW NYSE). The optimal weight in the small stock portfolio is given by the first element of $\tilde{V}^{-1}\tilde{E}/\iota_2'\tilde{V}^{-1}\tilde{E}$ and the maximum Sharpe ratio by $\sqrt{\tilde{E}'\tilde{V}^{-1}\tilde{E}}$, where $\tilde{E}$ and $\tilde{V}$ are the first two moments of the predictive density of the returns on the investable assets, obtained using our "model-and-data-based" methodology proposed in Section I. The maximum Sharpe ratio is the ex ante Sharpe ratio perceived by an investor who forms an optimal portfolio of the two assets. The intercept from the regression of the excess returns on the small stock portfolio on the excess returns on VW NYSE is denoted by α. The prior distribution of α has a mean of $\bar{\alpha}$ and a standard deviation of σ_α. Over the sample period of January 1982 through December 1996, the OLS estimates of the market model regression coefficients are $\hat{\alpha} = -0.83$ percent per year (with a standard error of 2.52 percent per year) and $\hat{\beta} = 0.98$. The values of $\bar{\alpha}$, σ_α, and the posterior mean and standard deviation of α are annualized percentage values.

Expected Prior Mispricing ($\bar{\alpha}$)	Prior Standard Deviation of α (σ_α)						
	0	1%	2%	3%	5%	10%	∞
A. Optimal percentage weight in small stocks							
0	0.00	−2.26	−7.74	−14.06	−24.17	−34.68	−40.56
+5%	218.04	205.79	174.79	136.98	72.41	0.86	−40.56
−5%	−241.48	−230.85	−204.64	−173.80	−123.48	−70.36	−40.56
+2.82%	131.16	122.21	100.11	74.03	31.18	−14.58	−40.56
B. Maximum Sharpe ratio							
0	0.1208	0.1208	0.1209	0.1211	0.1217	0.1226	0.1233
+5%	0.1908	0.1835	0.1668	0.1496	0.1290	0.1208	0.1233
−5%	0.1916	0.1863	0.1739	0.1606	0.1421	0.1281	0.1233
+2.82%	0.1471	0.1437	0.1363	0.1294	0.1223	0.1211	0.1233
C. Posterior mean of α ($\tilde{\alpha}$)							
0	0.00	−0.05	−0.16	−0.29	−0.49	−0.71	−0.83
+5%	5.00	4.68	3.89	2.98	1.53	0.02	−0.83
−5%	−5.00	−4.77	−4.20	−3.55	−2.51	−1.43	−0.83
+2.82%	2.82	2.61	2.12	1.55	0.65	−0.30	−0.83
D. Posterior standard deviation of α							
0	0.00	0.60	1.10	1.49	1.95	2.33	2.52
+5%	0.00	0.60	1.12	1.50	1.96	2.34	2.52
−5%	0.00	0.60	1.11	1.49	1.95	2.34	2.52
+2.82%	0.00	0.60	1.11	1.49	1.95	2.34	2.52

allocations previously studied in the literature. This paper investigates what happens in between. Compared to their counterparts in the data-based approach, the optimal portfolio weights are less sensitive to sampling error and tend to have less extreme values. The optimal portfolio reflects the implications of the model as well as the time series of asset returns.

In the empirical illustrations, sample evidence on home bias and value and size effects is evaluated from an asset-allocation perspective. The empirical evidence from earlier studies provides different conclusions about the benefits of international diversification and the related home bias puzzle. Whereas the sample estimates of return moments indicate that U.S. investors should invest a substantial part of their wealth abroad, the hypothesis that the U.S. market portfolio is globally mean-variance efficient is not rejected. Our framework provides a different perspective on the benefits of international diversification. U.S. investors' home equity bias can in principle be rationalized by a certain prior degree of belief in the global tangency of the U.S. market. However, we find that such prior beliefs must be very strong. U.S. investors' actual holdings of domestic versus foreign securities are consistent with a prior belief that the mispricing of a portfolio of foreign stocks within the domestic CAPM is between -2 percent and 2 percent per year.

Surprisingly, the same strong prior belief is significantly revised by the sample evidence about the Fama–French book-to-market portfolio. The current optimal allocation of an investor with such a strong belief in the efficiency of the market portfolio involves a 40 percent weight in book-to-market. Moreover, for more than half a century, the optimal allocations in book-to-market are large and fairly stable. Evidently, sample evidence can substantially revise even strong prior beliefs about the optimal allocation. The robust optimal positions in book-to-market are primarily due to the fact that the value premium in the U.S. data is very robust, in contrast to the size premium.

This study uses theoretically motivated prior information about expected returns in portfolio selection. In contrast, essentially no prior information about the covariance matrix is used. From an estimation perspective, the focus on expected returns could be helpful since it is well known that means are in general estimated with much less precision than covariances and that the tangency portfolio weights are very sensitive to small changes in expected returns. Nevertheless, using prior information to impose some structure on the covariance matrix could potentially also be beneficial, especially for a large number of assets. For example, Ledoit (1994) argues that shrinking the sample covariance matrix toward an identity matrix can be useful when constructing optimal portfolios. MacKinlay and Pástor (1998) provide some theoretical justification for using a plain identity matrix as a covariance matrix in portfolio selection. Our methodology allows the investor to include prior information about the residual covariance matrix of asset returns. By simply increasing the degrees of freedom in the prior distribution of the residual covariance matrix, the sample matrix can be shrunk arbitrarily close to a matrix specified a priori.

Potential extensions of the methodology developed here include relaxing the assumption that the investment opportunity set does not change over time. For example, conditional expected benchmark returns could be modeled as linear functions of state variables, as in Ferson and Harvey (1991). It could also be interesting to allow the investor to have a multiperiod investment horizon and hedge against changes in the investment opportunity set.[24] Another direction for future research is to allow the investor to specify prior beliefs about the validity of several asset pricing models. Empirical illustrations with different asset pricing models and different nonbenchmark assets might be informative, too. Finally, the out-of-sample performance of investment strategies with different degrees of belief in a model could be investigated. Preliminary results reveal substantial payoffs to incorporating some prior belief in the CAPM in portfolio selection.

Appendix A. Moments of the Predictive Density for Data-Based Portfolio Selection with Unequal History Lengths

This appendix provides the expressions for $\tilde{E}$ and $\tilde{V}$, the first two moments of the predictive density of returns on the investable assets, in the case of *data-based* portfolio selection with unequal history lengths. Recall that L vectors F_t (of dimension $1 \times K$), $t = 1, \ldots L$, of returns on K benchmark portfolios are available, together with $T \leq L$ vectors R_t (of dimension $1 \times N$), $t = L - T + 1, \ldots, L$, of returns on N nonbenchmark assets. Let $\hat{B}_2$ and S denote the statistics from the regression of the asset returns on the benchmark returns, as defined in equation (11). Define

$$\hat{E}_{F,T} = \frac{1}{T} \sum_{t=L-T+1}^{L} F_t'$$

$$\hat{E}_{A,T} = \frac{1}{T} \sum_{t=L-T+1}^{L} R_t'$$

$$\hat{E}_F = \frac{1}{L} \sum_{t=1}^{L} F_t'$$

$$\hat{E}_A = \hat{E}_{A,T} + \hat{B}_2'(\hat{E}_F - \hat{E}_{F,T})$$

[24] Even if the true investment opportunities do not change over time, the possibility of future learning about the parameter values could induce the investor to hedge against changes in the perceived investment opportunity set, as shown by Brennan (1997) in a continuous-time framework. The continuous-time literature that addresses the role of parameter uncertainty in portfolio selection also includes Gennotte (1986), Feldman (1992), and Xia (1998), among others. The discrete-time multiperiod problem with return predictability and parameter uncertainty is addressed in Barberis (2000). These papers focus on multiperiod decision making and do not consider the role of asset pricing models in portfolio selection.

$$\hat{V}_F = \frac{1}{L}\sum_{t=1}^{L}(F_t - \hat{E}_F)'(F_t - \hat{E}_F)$$

$$\hat{V}_{F,T} = \frac{1}{T}\sum_{t=L-T+1}^{L}(F_t - \hat{E}_F)'(F_t - \hat{E}_F)$$

$$\hat{V}_{AF} = \hat{B}_2\hat{V}_F$$

$$\hat{V}_{FA} = \hat{V}_{AF}'$$

$$\hat{V}_A = S/T + \hat{B}_2\hat{V}_F\hat{B}_2'.$$

The following expressions for $\tilde{E}$ and $\tilde{V}$ are presented in Propositions 1 and 2 of Stambaugh (1997) and are proved in the appendix to that paper.

$$\tilde{E} = \begin{bmatrix} \hat{E}_F \\ \hat{E}_A \end{bmatrix} \tag{A1}$$

$$\tilde{V} = \begin{bmatrix} \tilde{V}_F & \tilde{V}_{FA} \\ \tilde{V}_{AF} & \tilde{V}_A \end{bmatrix}, \tag{A2}$$

where

$$\tilde{V}_F = \left(\frac{L+1}{L-N-K-2}\right)\hat{V}_F$$

$$\tilde{V}_{FA} = \left(\frac{L+1}{L-N-K-2}\right)\hat{V}_{FA}$$

$$\tilde{V}_{AF} = \tilde{V}_{FA}'$$

$$\tilde{V}_A = \kappa S/T + \left(\frac{L+1}{L-N-K-2}\right)\hat{B}_2'\hat{V}_F\hat{B}_2$$

$$\kappa = \left(\frac{T}{T-N-2}\right)\left(1 + \frac{1}{T}\left[1 + \left(\frac{L+1}{L-N-K-2}\right)\right.\right.$$
$$\left.\left. \times \operatorname{tr}(\hat{V}_{F,T}^{-1}\hat{V}_F) + (\hat{E}_F - \hat{E}_{F,T})'\hat{V}_{F,T}^{-1}(\hat{E}_F - \hat{E}_{F,T})\right]\right).$$

Appendix B. Predictive Distribution of Benchmark Returns

The predictive density of the benchmark returns equals

$$p(F_{L+1}|F^L) = \int p(F_{L+1}|E_F,V_F,F^L)p(E_F,V_F|F^L)\,dE_F dV_F. \tag{B1}$$

The first term after the integral sign, $p(F_{L+1}|E_F,V_F,F^L)$, is simply a normal density with mean E_F and covariance matrix V_F. The second term, $p(E_F,V_F|F^L)$, is the posterior density of the benchmark moments and follows standard results. Define the statistics

$$\hat{E}_F = \frac{1}{L}\sum_{t=1}^{L} F_t' \tag{B2}$$

$$\hat{V}_F = \frac{1}{L}\sum_{t=1}^{L} (F_t - \hat{E}_F)'(F_t - \hat{E}_F). \tag{B3}$$

Then the posterior of V_F^{-1} is a Wishart distribution with parameter matrix $(L\hat{V}_F)^{-1}$ and $(L - 1)$ degrees of freedom. The posterior of E_F given V_F is normal with mean $\hat{E}_F$ and covariance matrix V_F/L. Therefore, draws of F_{L+1} from its predictive density can be obtained in three steps. First, draw V_F from its inverted Wishart posterior, then draw E_F from its normal posterior given V_F, and, finally, draw F_{L+1} from its normal density given E_F and V_F.

The moments of the predictive density in equation (B1) are obtainable analytically. Following Zellner (1971),

$$p(F_{L+1}|F^L) \propto \left[(L-K) + (F_{L+1} - \hat{E}_F')\left(\frac{L+1}{L-K}\hat{V}_F\right)^{-1}(F_{L+1} - \hat{E}_F')'\right]^{-L/2}. \tag{B4}$$

Hence, the predictive density of the benchmark returns is a multivariate Student t with $L - K$ degrees of freedom, and its first two moments are

$$E(F_{L+1}|F^L) = \hat{E}_F \tag{B5}$$

$$\operatorname{cov}(F_{L+1}',F_{L+1}|F^L) = \frac{L+1}{L-K-2}\hat{V}_F. \tag{B6}$$

Appendix C. Posterior Distribution of Regression Parameters

We derive the joint posterior density $p(B,\Sigma|\Phi)$ and describe an algorithm that can be used to obtain a large number of draws of B and Σ from their

joint posterior. The joint prior density of the regression parameters can be written as[25]

$$\begin{aligned} p(B,\Sigma) &= p(B|\Sigma)p(\Sigma) \\ &\propto |\Psi(\Sigma)|^{-1/2}\exp\{-\tfrac{1}{2}(b-\bar{b})'\Psi(\Sigma)^{-1}(b-\bar{b})\} \\ &\quad \times |\Sigma|^{-(\nu+N+1)/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}H\} \\ &\propto |\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}(b-\bar{b})'\Psi(\Sigma)^{-1}(b-\bar{b})\} \\ &\quad \times |\Sigma|^{-(\nu+N+1)/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}H\} \\ &\propto |\Sigma|^{-(\nu+N+2)/2}\exp\{-\tfrac{1}{2}(b-\bar{b})'\Psi(\Sigma)^{-1}(b-\bar{b}) - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}H\}. \end{aligned} \tag{C1}$$

Combining this prior density with the likelihood function in equation (11) gives the joint posterior density for the regression parameters

$$\begin{aligned} p(B,\Sigma|R,F^T) &\propto p(B,\Sigma)p(R|F^T,B,\Sigma) \\ &\propto |\Sigma|^{-(T+\nu+N+2)/2}\exp\{-\tfrac{1}{2}(b-\bar{b})'\Psi(\Sigma)^{-1}(b-\bar{b}) - \tfrac{1}{2}\operatorname{tr}H\Sigma^{-1} \\ &\qquad - \tfrac{1}{2}\operatorname{tr}S\Sigma^{-1} - \tfrac{1}{2}\operatorname{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\}. \end{aligned} \tag{C2}$$

Since

$$\begin{aligned} (b-\bar{b})'\Psi(\Sigma)^{-1}(b-\bar{b}) &= (\alpha-\bar{\alpha})'s^2\Sigma^{-1}(\alpha-\bar{\alpha})/\sigma_\alpha^2 + (\beta-\bar{\beta})'\Omega^{-1}(\beta-\bar{\beta}) \\ &= \operatorname{tr}((\alpha-\bar{\alpha})'s^2\Sigma^{-1}(\alpha-\bar{\alpha})/\sigma_\alpha^2) + (\beta-\bar{\beta})'\Omega^{-1}(\beta-\bar{\beta}) \\ &= \operatorname{tr}(\Sigma^{-1}(\alpha-\bar{\alpha})(\alpha-\bar{\alpha})'s^2/\sigma_\alpha^2) + (\beta-\bar{\beta})'\Omega^{-1}(\beta-\bar{\beta}), \end{aligned} \tag{C3}$$

the joint posterior of B and Σ takes the form

$$\begin{aligned} p(B,\Sigma|\Phi) \propto |\Sigma|^{-(T+\nu+N+2)/2}&\exp\{-\tfrac{1}{2}(\beta-\bar{\beta})'\Omega^{-1}(\beta-\bar{\beta})\} \\ &\times \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}C\}, \end{aligned} \tag{C4}$$

[25] Note that $|\Psi(\Sigma)| \propto |\Sigma|$. The densities of the multivariate normal and inverted Wishart distributions can be found in Zellner (1971, Appendix B).

where

$$C = S + H + (\alpha - \bar{\alpha})(\alpha - \bar{\alpha})'s^2/\sigma_\alpha^2 + (B - \hat{B})'X'X(B - \hat{B}). \tag{C5}$$

It follows from equation (C4) that the conditional posterior of Σ given B is inverted Wishart with parameter matrix C and $T + \nu + 1$ degrees of freedom:

$$p(\Sigma|B,\Phi) \propto |\Sigma|^{-(T+\nu+N+2)/2}\exp\{-\tfrac{1}{2}\text{tr}\,\Sigma^{-1}C\}. \tag{C6}$$

The tractability of this conditional distribution is due to zero correlations between α and β in the prior. With a nonzero prior correlation between α and β, the simplification in equation (C3) does not obtain.

In the case of $N = 1$, developed in Pástor and Stambaugh (1999), it is possible to complete the square on B in the conditional posterior of B given Σ. This conditional posterior is multivariate normal and thus is easy to sample from. However, for $N > 1$, the conditional posterior does not resemble any known distribution. The marginal posterior distribution of B is

$$p(B|\Phi) = \int p(B,\Sigma|\Phi)\,d\Sigma. \tag{C7}$$

The properties of the inverted Wishart distribution can be used to integrate out of Σ in equation (C4), and the following marginal posterior of B is obtained:

$$p(B|\Phi) \propto |C|^{-(T+\nu+1)/2}\exp\{-\tfrac{1}{2}(\beta - \bar{\beta})'\Omega^{-1}(\beta - \bar{\beta})\}. \tag{C8}$$

The assumption that no prior information about the assets' betas is available is reflected in the prior covariance matrix of β, Ω. Since this matrix is diagonal and contains very large values, the value of $(\beta - \bar{\beta})'\Omega^{-1}(\beta - \bar{\beta})$ is close to zero and the marginal posterior density of B in equation (C8) simplifies into

$$p(B|\Phi) \propto |S + H + (\alpha - \bar{\alpha})(\alpha - \bar{\alpha})'s^2/\sigma_\alpha^2 + (B - \hat{B})'X'X(B - \hat{B})|^{-(T+\nu+1)/2}. \tag{C9}$$

The components of $B = [\alpha\ B_2']'$ can be drawn from the above density using Gibbs sampling, a method proposed by Geman and Geman (1984) and more recently described in Casella and George (1992). As shown below, $\alpha|B_2,\Phi$ and $B_2|\alpha,\Phi$ turn out to be distributed as multivariate and generalized ("matric-

variate") Student t, respectively. For every posterior draw of B from the distribution in equation (C9), Σ can be drawn from its posterior in equation (C6). Some matrix manipulation on equation (C9) leads to

$$p(\alpha|B_2,\Phi) \propto |1 + c^*(\alpha - \alpha^*)'D^{-1}(\alpha - \alpha^*)|^{-(T+\nu+1)/2}, \tag{C10}$$

where

$$c^* = T + \frac{s^2}{\sigma_\alpha^2}$$

$$\alpha^* = \frac{T\hat{\alpha} + s^2\bar{\alpha}/\sigma_\alpha^2 - (B_2 - B_2)'(F^T)'\iota_T}{T + s^2/\sigma_\alpha^2} \tag{C11}$$

$$\begin{aligned} D = {} & S + H + T\hat{\alpha}\hat{\alpha}' + \bar{\alpha}\bar{\alpha}'s^2/\sigma_\alpha^2 - \hat{\alpha}\iota_T'F^T(B_2 - \hat{B}_2) - (B_2 - \hat{B}_2)'(F^T)'\iota_T\hat{\alpha}' \\ & + (B_2 - \hat{B}_2)'(F^T)'(F^T)(B_2 - \hat{B}_2) - c^*\alpha^*(\alpha^*)'. \end{aligned}$$

The above density is a multivariate Student t distribution with mean α^*, covariance matrix $D/(c^*(T + \nu - N - 1))$, and $(T + \nu - N + 1)$ degrees of freedom. As a result, draws of $(\alpha|B_2,\Phi)$ can be obtained in a straightforward manner. Similarly,

$$p(B_2|\alpha,\Phi) \propto |A + (B_2 - B_2^*)'M^*(B_2 - B_2^*)|^{-(T+\nu+1)/2}, \tag{C12}$$

where

$$M^* = (F^T)'(F^T)$$

$$B_2^* = \hat{B}_2 + ((F^T)'(F^T))^{-1}(F^T)'\iota_T(\hat{\alpha} - \alpha)' \tag{C13}$$

$$\begin{aligned} A = {} & S + H + (\alpha - \bar{\alpha})(\alpha - \bar{\alpha})'s^2/\sigma_\alpha^2 + T(\alpha - \hat{\alpha})(\alpha - \hat{\alpha})' - \hat{B}_2'(F^T)'\iota_T(\alpha - \hat{\alpha})' \\ & - (\alpha - \hat{\alpha})\iota_T'F^T\hat{B}_2 + \hat{B}_2'((F^T)'(F^T))\hat{B}_2 - (B_2^*)'M^*B_2^*. \end{aligned}$$

The above density is in the form of a generalized or matric-variate Student t distribution, which is described for example in Box and Tiao (1973). The N columns of B_2 can be drawn from multivariate Student t distributions. The Gibbs chain is initialized at $B_2 = \hat{B}_2$ and repeated draws are made from the distributions in equations (C10) and (C12). After an initial burn-in stage, the draws of α and B_2 are made from their joint posterior in equation (C9). For every such draw of B, Σ is drawn from an inverted Wishart distribution in equation (C6). The results in our multi-asset examples are based on Gibbs chains of 300,000 draws, after the first 1,000 draws are discarded. The Gibbs

chain appears to converge to the target distribution immediately, and produces virtually identical results when rerun with a different seed in the random number generator.

Special Case: One Nonbenchmark Asset (N = 1)

When $N = 1$, the posteriors of the regression parameters are obtainable using the univariate methodology proposed in Pástor and Stambaugh (1999). The posterior draws of the $(K + 1) \times 1$ vector $b \equiv B$ and the scalar $\sigma^2 \equiv \Sigma$ can be obtained by Gibbs sampling. The conditional posterior for b given σ is normal with mean $\tilde{b}_\sigma$ and covariance matrix M^{-1},

$$b|\sigma^2,\Phi \sim N(\tilde{b}_\sigma, M^{-1}), \tag{C14}$$

where

$$M = \Psi(\sigma)^{-1} + \frac{1}{\sigma^2} X'X \tag{C15}$$

and

$$\tilde{b}_\sigma = \begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix} = M^{-1}\left[\Psi(\sigma)^{-1}\bar{b} + \frac{1}{\sigma^2} X'X\hat{b}\right]. \tag{C16}$$

Note that, since $\tilde{b}_\sigma$ is a (matrix) weighted average of the prior mean $\bar{b}$ and the sample estimate $\hat{b}$, the sample estimate $\hat{b}$ is essentially "shrunk" toward its prior mean $\bar{b}$. The degree of shrinkage is determined in accordance with intuition by the precisions of $\bar{b}$ and $\hat{b}$ conditional on σ.

The conditional posterior of σ^2 given b from equation (C6) simplifies into

$$\sigma^2|b,\Phi \sim \frac{C}{\chi^2_{T+\nu+1}}, \tag{C17}$$

an inverted gamma distribution with $(T + \nu + 1)$ degrees of freedom.[26] Recall that C is defined in equation (C5). The Gibbs chain is initiated at the value of σ equal to its sample estimate and repeated draws are made from the conditional densities in equations (C14) and (C17). After a number of draws, the effect of the parameter values used to initiate the chain disappears and the draws are then made from the joint posterior $p(b,\sigma|\Phi)$.

[26] Since Pástor and Stambaugh (1999) do not assume a zero prior correlation between α and β, the conditional posterior of σ^2 in their study is not inverted gamma. Pástor and Stambaugh draw σ^2 using a Metropolis–Hastings algorithm with an inverted gamma proposal density. For a description of the Metropolis–Hastings algorithm, see Chib and Greenberg (1995).

Special Case: One Nonbenchmark Asset and One Benchmark (N = 1, K = 1)

For $K = 1$, the matrices M, $\Psi(\sigma)$, and $X'X$ in equation (C16) are all 2×2 matrices, so equation (C16) can be made more explicit using straightforward algebra. Further simplification of $\tilde{b}_\sigma$ is possible due to the assumption that the investor has no prior knowledge about the asset's β, since all the terms involving the reciprocal of the prior standard deviation of β (σ_β) are very close to zero and can be neglected. This simplification shows that $\tilde{b}_\sigma$ does not depend on σ, so $\tilde{b}_\sigma$ also equals the unconditional posterior mean of b. The elements of $\tilde{b}_\sigma$ can be written as

$$\tilde{\alpha} = (1 - w_{\hat{\alpha}})\bar{\alpha} + w_{\hat{\alpha}}\hat{\alpha} \tag{C18}$$

$$\tilde{\beta} = \hat{\beta} + \xi. \tag{C19}$$

In the above,

$$w_{\hat{\alpha}} = \frac{\hat{\text{var}}(F_t)}{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t^2 + \hat{\text{var}}(F_t)} \tag{C20}$$

$$\xi = \frac{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t}{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t^2 + \hat{\text{var}}(F_t)}(\hat{\alpha} - \bar{\alpha}), \tag{C21}$$

where $\bar{F}_t = (1/T)\sum_{t=L-T+1}^{L} F_t$, $\bar{F}_t^2 = (1/T)\sum_{t=L-T+1}^{L} F_t^2$, $\hat{\text{var}}(F_t) = \bar{F}_t^2 - (\bar{F}_t)^2$, and $E(\sigma^2)$ stands for the prior expected value of the residual variance σ^2.

Note that the weighting on $\hat{\alpha}$ and $\bar{\alpha}$ in the multivariate case in equation (C11) is very close to that from the univariate case. The main determinants of the weighting, σ_α, s^2, and T, enter in the same way in both cases.

In the presence of prior information about β (i.e., when $\sigma_\beta < \infty$), $\tilde{b}_\sigma$ depends on σ as well as σ_β, so the simplification of $\tilde{b}_\sigma$ into the expressions in equations (C18) and (C19) does not obtain. However, with an assumption that $\bar{F}_t = 0$, $\tilde{b}_\sigma$ can be simplified even for a finite σ_β. In particular, $\tilde{\alpha}$ can now be approximated by the expression in equation (C18), with $w_{\hat{\alpha}}$ taking an even simpler form of

$$w_{\hat{\alpha}} = \frac{1}{\dfrac{E(\sigma^2)}{T\sigma_\alpha^2} + 1}. \tag{C22}$$

The conditional posterior mean of β given σ is a weighted average of the prior mean of β and its sample estimate, but the unconditional posterior mean of β, $\tilde{\beta}$, is not tractable. Of course, $\tilde{\beta}$ can still be obtained using Gibbs sampling.

Appendix D. Moments of the Predictive Density for $N = 1$, $K = 1$

We provide the expressions for the approximate unconditional posterior moments of b and σ^2 as well as for the resulting $\tilde{E}$ and $\tilde{V}$, the first two moments of the predictive density of returns on the investable assets. The unconditional posterior mean of b is given by equation (C16) and later by equations (C18) and (C19). The unconditional posterior covariance matrix of b, obtainable by the variance decomposition rule, can be shown to equal

$$\operatorname{cov}(b,b'|\Phi) = \frac{E(\sigma^2|\Phi)}{T\left[\dfrac{E(\sigma^2)}{T\sigma_\alpha^2}\bar{F}_t^2 + \hat{\operatorname{var}}(F_t)\right]}\begin{pmatrix} \bar{F}_t^2 & -\bar{F}_t \\ -\bar{F}_t & \dfrac{E(\sigma^2)}{T\sigma_\alpha^2}+1 \end{pmatrix}, \tag{D1}$$

where $E(\sigma^2|\Phi)$ is the posterior mean of σ^2, shown below.

The unconditional posterior moments of σ^2 can be well approximated based on equation (C17) by the posterior moments of σ^2 conditional on $\tilde{b}_\sigma$, the posterior mean of b shown in equation (C16). Note from equations (C5) and (C17) that the posterior of σ^2 depends on b through the last two terms in equation (C5). Those terms are in general quite small relative to $(S + H)$, and the results are almost unaffected when these terms are ignored. Instead of ignoring them, b in those terms is replaced by its posterior mean, and the results are virtually indistinguishable from those obtained by Gibbs sampling. Based on Zellner (1971, Appendix A), the first two approximate posterior moments of σ^2 are

$$E(\sigma^2|\Phi) = \frac{\bar{C}}{T+\nu-1} \tag{D2}$$

$$\operatorname{var}(\sigma^2|\Phi) = \frac{2\bar{C}^2}{(T+\nu-1)^2(T+\nu-3)}, \tag{D3}$$

where $\bar{C} = S + H + (\tilde{\alpha} - \bar{\alpha})^2 s^2/\sigma_\alpha^2 + (\tilde{b}_\sigma - \hat{b})'X'X(\tilde{b}_\sigma - \hat{b})$.

The predictive moments for the investable assets are defined as

$$\tilde{E} = [E(R_{L+1}|\Phi) \quad E(F_{L+1}|\Phi)]' \tag{D4}$$

$$\tilde{V} = \begin{bmatrix} \operatorname{var}(R_{L+1}|\Phi) & \operatorname{cov}(R_{L+1},F_{L+1}|\Phi) \\ \operatorname{cov}(R_{L+1},F_{L+1}|\Phi) & \operatorname{var}(F_{L+1}|\Phi) \end{bmatrix}. \tag{D5}$$

The expressions for $E(F_{L+1}|\Phi)$ and $\text{var}(F_{L+1}|\Phi)$ are given in equations (B5) and (B6). It follows by the law of iterative expectations that

$$E(R_{L+1}|\Phi) = E(X_{L+1}|\Phi)\tilde{b}_\sigma = \tilde{\alpha} + \tilde{\beta}\hat{E}_F. \tag{D6}$$

Recall that the posterior means of α and β, $\tilde{\alpha}$ and $\tilde{\beta}$, are defined in equation (C16), and $\hat{E}_F$ is defined in equation (B2). The result in equation (D6) does not rely on any approximation. Using the variance decomposition rule, it can be shown that

$$\begin{aligned}\text{var}(R_{L+1}|\Phi) = {} & E(\sigma^2|\Phi) + \text{tr}\{\text{cov}(X'_{L+1},X_{L+1}|\Phi)[\text{cov}(b,b'|\Phi) + \tilde{b}_\sigma\tilde{b}'_\sigma]\} \\ & + E(X_{L+1}|\Phi)\text{cov}(b,b'|\Phi)E(X_{L+1}|\Phi)'\end{aligned} \tag{D7}$$

$$\text{cov}(R_{L+1},F_{L+1}|\Phi) = \tilde{\beta}\,\text{var}(F_{L+1}|\Phi). \tag{D8}$$

These two elements of the predictive covariance matrix can be evaluated using equations (D1) and (D2).

Appendix E. Optimal Weight in the Nonbenchmark Asset for $N = 1$, $K = 1$

The optimal weight in the nonbenchmark asset can be obtained from equation (B5) using equations (D4) and (D5). An alternative representation of the weight is obtained below. This equivalent representation serves to provide some insight into the components of the weight that are not immediately recognizable from equations (D4) and (D5).

In order to simplify notation, denote the predictive moments in equations (D4) and (D5) by

$$\tilde{E} = \begin{bmatrix}\tilde{E}_A \\ \tilde{E}_F\end{bmatrix} \quad \text{and} \quad \tilde{V} = \begin{bmatrix}\tilde{V}_A & \tilde{V}_{AF} \\ \tilde{V}_{AF} & \tilde{V}_F\end{bmatrix}. \tag{E1}$$

The vector of the optimal weights in the nonbenchmark asset and in the benchmark, respectively, is proportional to

$$\tilde{V}^{-1}\tilde{E} = \frac{\tilde{V}_F}{\tilde{V}_A\tilde{V}_F - \tilde{V}_{AF}^2} \times \begin{bmatrix}\tilde{E}_A - (\tilde{V}_{AF}/\tilde{V}_F)\tilde{E}_F \\ \tilde{E}_F(\tilde{V}_A/\tilde{V}_F) - \tilde{E}_A(\tilde{V}_{AF}/\tilde{V}_F)\end{bmatrix}. \tag{E2}$$

Since equation (D8) implies that $(\tilde{V}_{AF}/\tilde{V}_F)$ is equal to the posterior mean of β, $\tilde{\beta}$, the optimal weight in the nonbenchmark asset is proportional to

$$\begin{aligned} w_A &\propto \frac{\tilde{V}_F}{\tilde{V}_A\tilde{V}_F - \tilde{V}_{AF}^2} \times (\tilde{E}_A - \tilde{\beta}\tilde{E}_F) \\ &= \frac{\tilde{V}_F}{\tilde{V}_A\tilde{V}_F - \tilde{V}_{AF}^2} \times \tilde{\alpha} \\ &= \frac{\tilde{\alpha}}{\tilde{V}_A - \tilde{V}_{AF}^2/\tilde{V}_F} \\ &= \frac{\tilde{\alpha}}{\tilde{V}_A - \tilde{\beta}^2\tilde{V}_F} \\ &= \frac{\tilde{\alpha}}{\tilde{V}_u}, \end{aligned} \tag{E3}$$

where $\tilde{\alpha}$ is the posterior mean of α and $\tilde{V}_u$ denotes the predictive residual variance, the variance of the next-period disturbance u_{L+1} from the market model regression. Using the variance decomposition rule,

$$\tilde{V}_u = \text{var}(u_{L+1}|\Phi) = E(\text{var}(u_{L+1}|\sigma^2,\Phi)) = E(\sigma^2|\Phi) \equiv \tilde{\sigma}^2, \tag{E4}$$

where σ^2 denotes the residual variance and $\tilde{\sigma}^2$ its posterior mean. The optimal weight in the nonbenchmark asset is therefore proportional to the ratio of the posterior means of α and σ^2:

$$w_A \propto \frac{\tilde{\alpha}}{\tilde{\sigma}^2}. \tag{E5}$$

This simple result is a special case of the result derived by Stevens (1998), except that the unknown values of α and σ^2 are replaced by their posterior means.

REFERENCES

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

Banz, Rolf, 1981, The relationship between returns and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

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

Bekaert, Geert, and Michael S. Urias, 1996, Diversification, integation, and emerging market closed-end funds, *Journal of Finance* 51, 835–869.

Black, Fisher, and Robert Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* 48, 28–43.

Bohn, Henning, and Linda L. Tesar, 1996, U.S. equity investment in foreign markets: Portfolio rebalancing or return chasing?, *American Economic Review* 86, 77–81.

Box, George E.P., and George C. Tiao, 1973, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, Reading, Mass.).

Brennan, Michael J., 1997, The role of learning in dynamic portfolio decisions, Working paper 3-97, UCLA.

Britten-Jones, Mark, 1999, The sampling error in estimates of mean-variance efficient portfolio weights, *Journal of Finance* 54, 655–671.

Brown, Philip, Allan W. Kleidon, and Terry A. Marsh, 1983, New evidence on the nature of size-related anomalies in stock prices, *Journal of Financial Economics* 12, 33–56.

Brown, Stephen J., 1976, Optimal portfolio choice under uncertainty: A Bayesian approach, Ph.D. dissertation, University of Chicago.

Casella, George, and Edward I. George, 1992, Explaining the Gibbs Sampler, *The American Statistician* 46, 167–174.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok, 1995, Evaluating the performance of value versus glamour stocks: The impact of selection bias, *Journal of Financial Economics* 38, 269–296.

Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis–Hastings algorithm, *The American Statistician* 49, 327–335.

Davis, James, 1994, The cross-section of realized stock returns: The pre-COMPUSTAT evidence, *Journal of Finance* 49, 1579–1593.

Davis, James, Eugene F. Fama, and Kenneth R. French, 1998, Characteristics, covariances, and average returns: 1929–1997, Working paper, Kansas State University.

Errunza, Vihang, Ked Hogan, and Mao-Wei Hung, 1999, Can the gains from international diversification be achieved without trading abroad?, *Journal of Finance* 54, 2075–2107.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1996, The CAPM is wanted, dead or alive, *Journal of Finance* 51, 1947–1958.

Fama, Eugene F., and Kenneth R. French, 1998, Value versus growth: The international evidence, *Journal of Finance* 53, 1975–1999.

Feldman, David, 1992, Logarithmic preferences, myopic decisions, and incomplete information, *Journal of Financial and Quantitative Analysis* 27, 619–629.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

French, Kenneth R., and James M. Poterba, 1991, International diversification and international equity markets, *American Economic Review* 81, 222–226.

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.

Geman, S., and D. Geman, 1984, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 609–628.

Gennotte, Gerard, 1986, Optimal portfolio choice under incomplete information (with discussion), *Journal of Finance* 41, 733–749.

Geweke, John, and Guofu Zhou, 1996, Measuring the pricing error of the arbitrage pricing theory, *Review of Financial Studies* 9, 557–587.

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

Gorman, Larry R., and Bjorn N. Jorgensen, 1997, Domestic versus international portfolio selection: A statistical examination of the home bias, Working paper, Northwestern University.

Harvey, Campbell R., and Guofu Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics* 26, 221–254.

Huberman, Gur, Shmuel Kandel, and Robert F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *Journal of Finance* 42, 1–9.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jobson, J. D., B. Korkie, and V. Ratti, 1979, Improved estimation for Markowitz portfolios using James–Stein type estimators, in *Proceedings of the American Statistical Association, Business and Economic Statistics Section* (American Statistical Association, Washington).

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, 717–727.

Kandel, Shmuel, Robert McCulloch, and Robert F. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies* 8, 1–53.

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.

Kang, Jun-Koo, and René M. Stulz, 1997, Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan, *Journal of Financial Economics* 46, 3–28.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13–32.

Keim, Donald B., 1999, An analysis of mutual fund design: The case of investing in small-cap stocks, *Journal of Financial Economics* 51, 173–194.

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

Kothari, S. P., Jay Shanken, and Richard G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185–224.

Ledoit, Olivier, 1994, Portfolio selection: Improved covariance matrix estimation, manuscript, MIT.

Levy, Haim, and Marshall Sarnat, 1970, International diversification of investment portfolios, *American Economic Review* 60, 668–675.

Lewis, Karen K., 1999, Trying to explain home bias in equities and consumption, *Journal of Economic Literature*, forthcoming.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

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

MacKinlay, A. Craig, and Ľuboš Pástor, 1998, Asset pricing models: Implications for expected returns and portfolio selection, Working paper, University of Pennsylvania.

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.

McCulloch, Robert, and Peter E. Rossi, 1991, A Bayesian approach to testing the arbitrage pricing theory, *Journal of Econometrics* 49, 141–168.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Pástor, Ľuboš, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance*, 54, 67–121.

Reinganum, Marc R., 1981, Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values, *Journal of Financial Economics* 9, 19–46.

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

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

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Stambaugh, Robert F., 1998, The role of financial models in empirical research, Keynote address at the 1998 Annual Meetings of the Northern Finance Association.

Stevens, Guy V. G., 1998, On the inverse of the covariance matrix in portfolio analysis, *Journal of Finance* 53, 1821–1827.

Treynor, Jack, and Fisher Black, 1973, How to use security analysis to improve portfolio selection, *Journal of Business* 46, 65–86.

Xia, Yihong, 1998, Learning about predictability: The effects of parameter uncertainty in asset allocation in a dynamic setting, Working paper, UCLA.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley & 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.