1. Bayesian Applications to the Investment Management Process

Biliana Bagasheva
Department of Statistics and Applied Probability
University of California, Santa Barbara
CA 93106 – 3110, USA
Email: biliana@pstat.ucsb.edu

Svetlozar (Zari) Rachev*
Chair-Professor, Chair of Econometrics, Statistics,
and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Postfach 6980, 76128 Karlsruhe, Germany
and
Department of Statistics and Applied Probability
University of California, Santa Barbara
CA 93106 – 3110, USA
Email: zari.rachev@statistik.uni-karlsruhe.de

John Hsu
Associate Professor
Department of Statistics and Applied Probability
University of California, Santa Barbara
CA 93106 – 3110, USA
Email: hsu@pstat.ucsb.edu

Frank Fabozzi
Frederick Frank Adjunct Professor of Finance
Yale School of Management
135 Prospect Street, Box 208200
New Haven, Connecticut, 06520-8200, USA
Email: fabozzi321@aol.com

*Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the Deutschen Forschungsgemeinschaft and the Deutscher Akademischer Austausch Dienst.

1. 1 Introduction

There are several tasks in the investment management process. These include setting the investment objectives, establishing an investment policy, selecting a portfolio strategy, asset allocation, and measuring and evaluating performance. Bayesian methods have been either used or proposed as a tool for improving the implementation of several of these tasks. There are principal reasons for using Bayesian methods in the investment management process. First, they allow the investor to account for the uncertainty about the parameters of the return-generating process and the distributions of returns for asset classes and to incorporate prior beliefs in the decision-making process. Second, they address a deficiency of the standard statistical measures in conveying the economic significance of the information contained in the observed sample of data. Finally, they provide an analytically and computationally manageable framework in models where a large number of variables and parameters makes classical formulations a formidable challenge.

The goal of this chapter is to survey selected Bayesian applications to investment management. In Section 1.2, we discuss the single-period portfolio problem, emphasizing how Bayesian methods improve the estimation of the moments of returns, primarily the mean. In Section 1.3, we describe the mechanism for incorporating asset-pricing models into the investment decision-making process. Tests of mean-variance efficiency are surveyed in Section 1.4. We explore the implications of predictability for investment management in Section 1.5 and then provide concluding remarks in Section 1.6.

1.2. The Single-Period Portfolio Problem

The portfolio choice problem represents a primary example of decision-making under uncertainty. Let r_{T+1} denote the vector ($N \times 1$) of next-period returns and W current wealth. We denote next-period wealth by $W_{T+1} = W(1 + \omega' r_{T+1})$ in the absence of a risk-free asset and $W_{T+1} = W(1 + r_f + \omega' r_{T+1})$ when a risk-free asset with return r_f is present. Let ω denote the vector of asset allocations (fractions of wealth allocated to the corresponding stocks). In a one-period setting, the optimal portfolio decision consists of choosing ω that maximizes the expected utility of next-period's wealth,

$$\max_{\omega} E\left(U\left(W_{T+1}\right)\right) = \max_{\omega} \int U\left(W_{T+1}\right) p(r \mid \theta) dr\,, \tag{1.1}$$

subject to feasibility constraints, where θ is the parameter vector of the return distribution and U is a utility function generally characterized by a quadratic or a negative exponential functional form. A key component of Eq. (1.1) is the distribution of returns $p(r \mid \theta)$, conditional on the unknown parameter vector θ. The traditional implementation of the mean-variance framework[1] proceeds with setting θ equal to its estimate $\hat{\theta}(r)$ based on some estimator of the data r (often the maximum likelihood estimator). Then, the investor's problem in Eq. (1.1) leads to the optimal allocation given by

$$\omega^{*} = \arg\max_{\omega} E\left(U\left(\omega' r\right) \mid \theta = \hat{\theta}(r)\right). \tag{1.2}$$

The solution in Eq. (1.2), known as the *certainty equivalent solution*, treats the estimated parameters as the true ones and completely ignores the effect of the estimation error on the optimal decision. The resulting portfolio displays high sensitivity to small changes in the estimated mean, variance, and covariance, and usually contains large long and short positions that are difficult to implement in practice.[2]

Starting with the work of (Zellner and Chetty 1965), several early studies investigate the effect parameter uncertainty plays on optimal portfolio choice by re-expressing Eq. (1.1) in terms of the predictive density func-

[1] The mean-variance selection rule of (Markowitz's 1952), given by $\min_{\omega} \omega'\Sigma\omega,\ s.t.\ \omega'\mu \geq \mu^{*}, \omega'\iota = 1$, where μ is the vector of expected returns , Σ is the covariance matrix of returns, and ι is a compatible vector of ones, provides the same set of admissible portfolios as the quadratic-type expected-utility maximization in Eq. (1.1). (Markowitz and Usmen 1996) point out that the conventional wisdom that the necessary conditions for application of mean-variance analysis are normal probability distribution and/or quadratic utility is a "misimpression" (Markowitz and Usmen 1996, p. 217). Almost optimal solutions are obtained using a variety of utility functions and distributions. For example, it is possible to weaken the distribution condition to members of the location-scale family. See (Ortobelli, Rachev, and Schwartz 2004).

[2] See, for example, (Best and Grauer 1991)

tion.[3] The predictive density function reflects estimation risk explicitly since it integrates over the posterior distribution, which summarizes the uncertainty about the model parameters, updated with the information contained in the observed data. The optimal Bayesian portfolio problem takes the form:

$$\max_{\omega} E_{\theta}\left\{E_{r|\theta}\left(U\left(W_{T+1}\right) \mid \theta\right)\right\} = \tag{1.3}$$
$$\max_{\omega} \iint U\left(W_{T+1}\right) p\left(r_{T+1} \mid \theta\right) p(\theta \mid r)$$
$$\max_{\omega} \int U\left(W_{T+1}\right) \left[\int p\left(r_{T+1} \mid \theta\right) p(\theta \mid r) d\theta\right] dr,$$

where by Bayes' rule, the posterior density $p(\theta \mid r)$ is proportional to the product of the sampling density (the likelihood function) and the prior density, $f(r \mid \theta) p(\theta)$.

The multivariate normal distribution is the simplest and most convenient choice of sampling distribution in the context of portfolio selection, even though empirical evidence does not fully support this model.[4] In the case where no particular information (intuition) about the model parameters is available prior to observing the data, the decision-maker has diffuse (non-informative) prior beliefs, usually expressed in the form of the Jeffrey's prior $p(\mu, \Sigma) \propto |\Sigma|^{-(N+1)/2}$, where μ and Σ are, respectively, the mean vector and the covariance vector of the multivariate normal return distribution, N is the number of assets in the investment universe, and $\propto$ denotes "proportional to". The joint predictive distribution of returns is then a multivariate Student-*t* distribution.

Informative prior beliefs are usually cast in a conjugate framework to ensure analytical tractability of the posterior and predictive distributions. The predictive distribution is multivariate normal only when the covariance Σ is assumed known and μ is asserted to have the conjugate prior $N\left(\mu_0 \iota, \tau^2 I\right)$, where μ_0 stands for the prior mean, ι is a vector of ones, and $\tau^2 I$ is the diagonal prior covariance matrix. When both parameters are unknown and conjugate priors are assumed (the conjugate prior for Σ in a multivariate setting is an inverse-Wishart with scale parameter S^{-1}, where

[3] See, for example, (Barry 1974; Winkler and Barry 1975; Klein and Bawa 1976; Brown 1976; Jobson, Korkie and Ratti 1979; Jobson and Korkie 1980; Chen and Brown 1983).

[4] For example, see (Fama 1965).

S is the sample covariance matrix), the predictive distribution is multivariate Student-*t*.[5]

(Klein and Bawa 1976) compare the Bayesian and certainty equivalent optimal solutions under the assumption of a diffuse prior for the parameters of the multivariate normal returns distribution ((Barry 1974) asserts informative priors) and show that in both cases the admissible sets are the same up to a constant. However, the optimal choice differs in the two scenarios since portfolio risk is perceived differently in each case. Both the optimal individual investor's portfolio and the market portfolio have lower expected returns in the Bayesian setting. (Brown 1976) shows that the failure to account for estimation risk leads to suboptimal solutions.

It is instructive to examine the posterior mean under the informative prior assumption. Assuming that $\Sigma = \sigma^2 I$, the i th element of μ's posterior mean has the form

$$\mu_i \mid r, \sigma^2 = \left(\frac{T}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \left(\frac{T}{\sigma^2} \bar{r}_i + \frac{1}{\tau^2} \mu_0 \right) \tag{1.4}$$

where $\bar{r}_i$ is the sample mean of asset *i*, and *T* is the sample size. The posterior mean is a weighted average of the prior and sample information; that is, the sample mean $\bar{r}_i$ of asset *i* is shrunk to the prior mean μ_0. The degree of shrinkage depends on the strength of the confidence in the prior distribution, as measured by the prior precision $1/\tau^2$. The higher the prior precision, the stronger the influence of the prior mean on the posterior mean. Shrinking the sample mean reduces the sensitivity of the optimal weights to the sampling error in it. As a result, weights take less extreme values and their stability over time is improved. The prior distribution of μ could be made uninformative by choosing a very large prior variance elements τ^2. In the extreme case of an infinite prior variance, the posterior mean coincides with the sample mean and the correction for estimation risk becomes insignificant (Brown 1979; Jorion 1985).

The approach of employing shrinkage estimators as a way of accounting for uncertainty is rooted in statistics and can be traced back to (James and Stein 1961), who recognized the inadmissibility of the sample mean in a multivariate setting under a squared loss function. The James-Stein estimator given by

[5] See, for example, (Brown 1976)

$$\hat{\mu}^{JS} = \delta \mu_0 \iota + (1-\delta)\bar{r}\,, \tag{1.5}$$

where $\bar{r} = (\bar{r}_{1,t}, \ldots, \bar{r}_{N,t})$ is the vector of sample means, has a uniformly lower risk than $\bar{r}$, regardless of the point μ_0 towards which the means are shrunk.[6] However, the gains are greater the closer μ_0 is to the true value. For the special case when the return covariance matrix has the form $\Sigma = \sigma^2 I$, σ^2 is known, and the number of assets *N* is greater than 2, the weight δ is given by

$$\delta = \min\left\{1, \frac{(N-2)T}{(\bar{r}-\mu_0\iota)'\Sigma^{-1}(\bar{r}-\mu_0\iota)}\right\}.$$

Within the portfolio selection context, the effort was initiated with the papers of (Jobson, Korkie, and Ratti 1979; Jobson and Korkie 1980, 1981) and developed by (Jorion 1985, 1986; Grauer and Hakansson 1990). (Dumas and Jacquillat 1990) discuss Bayes-Stein estimation in the context of currency portfolio selection.

While the choice of prior distributions is often guided by considerations of tractability, the parameters of the prior distributions (called hyperparameters) are determined in a rather subjective fashion. This has led some researchers to embrace the *empirical Bayes approach*, which uses sample information to determine the hyperparameter values and is at the heart of the Bayesian interpretation of shrinkage estimators. The shrinkage target is the grand mean of returns *M*:

$$P(\mu) \sim N(M, \tau\Sigma).^{7} \tag{1.6}$$

(Frost and Savarino 1986; and Jorion 1986) employ it in an examination of the portfolio choice problem, asserting the conjugate inverse-Wishart prior for Σ. They estimate the prior parameters via maximum likelihood, assuming equality of the means, variances, and covariances. Comparing certainty-equivalent rates of return, they find that the optimal portfolios

[6] (Berger 1980) points out that the inadmissibility of the sample mean in the frequentist case is translated into inadmissibility of the Bayesian rule under the assumption of diffuse (improper) prior.

[7] It is not unusual to assume that the degree of uncertainty about the mean vector is proportional to the volatilities of returns. A value of τ smaller than 1 reflects the intuition that uncertainty about the mean is lower than uncertainty about the individual returns.

obtained in the Bayesian setting with informative priors outperform the optimal choices under both the classical and diffuse Bayes frameworks.[8]

(Jorion 1986) assumes that Σ is known and is replaced by its sample estimator $\frac{T-1}{T-N-2}S$. Jorion derives the so-called Bayes-Stein estimator of expected returns – a weighted average of sample means and the mean of the global minimum variance portfolio $\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}\bar{r}$ (the solution to the variance minimization problem under the constraint that the weights sum to unity).[9] He finds that the Bayes-Stein shrinkage estimator outperforms significantly the sample mean, based on comparison of the empirical risk function.[10] (Grauer and Hakansson 1990) observe that the portfolio strategies based on the Bayes-Stein and the James-Stein estimators are only marginally better than the historic mean strategies.

(Frost and Savarino 1986) obtain a shrinkage estimator not only for the mean vector but also for the covariance matrix of the predictive returns distribution, thus contributing to a relatively neglected area. A reason why there are relatively more studies concerned only with uncertainty about the mean (see also the discussion of the Black and Litterman model below) may be that optimal portfolio choice is highly sensitive to estimation error in the expected means, while variances and covariances (although also unknown) are more stable over time ((Merton 1980)). However, given that the optimal investor decision is the result of the trade-off between risk and return, efficient variance estimation seems to be no less important than mean estimation.[11]

[8] A certainty-equivalent rate of return is the risk-free rate of return which provides the same utility as the return on a given combination of risky assets.

[9] (Dumas and Jacquillat 1990) argue that in the international context this result introduces country-specific bias. They advocate shrinkage towards a portfolio which assigns equal weights to all currencies.

[10] The empirical risk function is computed as the loss of utility due to the estimation risk $L(\varpi^*, \hat{\varpi}) = \frac{F_{\max} - F(\hat{q})}{|F_{\max}|}$ averaged over repeated samples, where ϖ^* is the solution to (1) when the true parameter vector $\boldsymbol{\theta}$ is known, $\hat{\varpi}$ is the portfolio choice on the basis of the sample estimate $\hat{\boldsymbol{\theta}}$, $F_{\max}$ and F are the corresponding values of the utility functions.

[11] See, for example, (Frankfurter, Phillips, and Seagle 1972).

1.3. Combining Prior Beliefs and Asset Pricing Models

(Ledoit and Wolf 2003) develop a shrinkage estimator for the covariance matrix of returns in a portfolio selection setting, choosing as a shrinkage target the covariance matrix estimated from Sharpe's (Sharpe 1963) single-factor model of stock returns. They join a growing trend in the shrinkage estimator literature of deriving the shrinkage target structure from a model of market equilibrium. Equivalently, the asset pricing model serves as the reference point around which the investor builds prior beliefs. There is a trade-off then between the degree of confidence in the validity of the model and the information content of the observed data sample. The influential work of Black and Litterman (Black and Litterman 1990, 1991, 1992) (BL) presumably constitute the first analysis employing this approach.[12] Their model allows for a smooth and flexible combination of an asset pricing model, the Capital Asset Pricing Model (CAPM), and investor's views. The CAPM is assumed to hold in general, and investors' beliefs about expected stock returns can be expressed in the form of deviations from the model predictions.[13] Interpretations of the BL methodology from the Bayesian point of view are scarce (Satchell and Scowcroft 2000; He and Litterman 1999; Lee 2000; Meucci 2005), although, undoubtedly, the BL decision-maker is Bayesian, and somewhat ambiguous.

The excess returns of the *N* assets in the investment universe are assumed to follow a multivariate normal distribution $r \sim N(\mu, \Sigma)$.[14] The implied equilibrium risk premiums Π are used as a proxy for the true equilibrium returns and the distribution of expected .equilibrium returns is centered on them, with a covariance matrix proportional to Σ:

$$\mu \sim N(\Pi, \tau\Sigma) \tag{1.7}$$

where the scalar τ indicates the degree of uncertainty in the CAPM.[15] The investor's views (linear combinations of expected asset returns) are ex-

[12] For example, (Jorion 1991) mentions the possibility of using the CAPM equilibrium forecasts to form prior beliefs but doesn't pursue the idea further.

[13] BL consider an equilibrium model, such as the CAPM, as the most appropriate neutral shrinkage target for expected returns, since equilibrium returns clear the market when all investors have homogeneous views.

[14] The covariance matrix Σ is estimated outside of the model (see (Litterman and Winkelmann 1998)) for the specific methodology) and considered as given.

[15] The equilibrium risk premiums Π are the expected stock returns in excess of the risk-free rate, estimated within the CAPM framework. In the setting of the BL model, the vector Π is determined by a procedure appropriately called "reverse optimization". The market-capitalization weights observed in the capi-

pressed as probability distributions of the expected returns on the so-called "view" portfolios:

$$P\mu \sim N(Q,\Omega), \tag{1.8}$$

where P is a (K x N) matrix whose rows correspond to the K view portfolio weights. The magnitudes of the elements ϖ_i of Ω represent the degree of confidence the investor has in each view.

There is no consensus as to which one of the distributions in Eqs. (1.7 and 1.8) defines the prior and which one the sampling density. (Satchell and Scowcroft 2000; Lee 2000; Meucci 2005) favor the position that the investor views constitute the prior information which serves to update the equilibrium distribution of expected returns (in the role of the sampling distribution). This interpretation is in line with the Bayesian tradition of using subjective beliefs to construct the prior distribution. On the other hand, He and Litterman's (He and Litterman 1999) reference to Eq. (1.8) as the prior also has grounds in the Bayesian theory. Suppose that we are able to take a sample from the population of future returns, in which our subjective belief about the expected stock returns is realized. Then, a view could be interpreted as the information contained in this hypothetical sample.[16] The sample size corresponds to the degree of confidence the investor has in his view.

The particular definition one adopts does not have a bearing on the results. Deriving the posterior distribution of expected returns is a straightforward application of conjugate analysis and yields the familiar result

$$\mu \mid \Pi, Q, \Sigma, \Omega, \tau \sim N\left(\tilde{\mu}, \tilde{V}\right) \tag{1.9}$$

where the posterior mean and covariance matrix are given by

$$\tilde{\mu} = \left((\tau\Sigma)^{-1} + P'\Omega^{-1}P\right)^{-1}\left((\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}Q\right) \tag{1.10}$$

and

tal market are considered the optimal weights ω^*. Using the estimate $\hat{\Sigma}$ of the covariance matrix, the risk premiums are backed out of the standard mean-variance result $\omega^* = (1/\lambda)\left(\hat{\Sigma}^{-1}\Pi / \iota'\hat{\Sigma}^{-1}\Pi\right)$, where λ is the coefficient of relative risk aversion.

[16] See (Black and Litterman 1992) for this interpretation. Interpreting prior belief in terms of a hypothetical sample is not uncommon in Bayesian analysis. See also Stambaugh (1999).

$$\tilde{V} = \left((\tau\Sigma)^{-1} + P'\Omega^{-1}P\right)^{-1}. \tag{1.11}$$

The estimator of expected returns in Eq. (1.10) clearly has the form of a shrinkage estimator (the weights of Π and Q sum up to 1). When the level of certainty about the equilibrium returns increases (τ approaches 0), their weight $\left((\tau\Sigma)^{-1} + P'\Omega^{-1}P\right)^{-1}(\tau\Sigma)^{-1}$ increases and the investor optimally holds the market portfolio. If, on the contrary, belief in the deviations from equilibrium returns is strong, more weight is put on the views. (Lee 2000) extends the BL model to the tactical allocation problem. The equilibrium risk premiums Π are replaced by the vector of expected excess returns corresponding to a neutral position with respect to tactical bets, i.e., to holding the benchmark portfolio.

Admittedly, the BL methodology does not make use of all of the available information in historical returns, particularly, the sample means. (Pastor 2000; Pastor and Stambaugh 1999) address this issue by developing a framework in which uncertainty in the validity of the asset pricing model is quantified in terms of the amount of model mispricing. The estimate of expected returns is a weighted average between the model prediction and the sample mean, thus incorporating the benefits of both the Bayes-Stein and the BL methodologies.[17]

Let the return generating process for the stock's excess return be

$$r_t = \alpha + \beta' f_t + \varepsilon_t \qquad t = 1,\ldots,T, \tag{1.12}$$

where f_t denotes a (*K* x *1*) vector of factor returns (returns to benchmark portfolios), and ε_t is a mean-zero disturbance term. Then, the slopes of the regression in Eq. (1.12) are stock's sensitivities (betas). The stock's expected excess return implied by the model is

$$E(r_t) = \beta' E(f_t) \tag{1.13}$$

That is, the model implies that $\alpha = 0$.[18] When the investor believes there is some degree of pricing inefficiency in the model, the expected excess return will reflect this through an unknown mispricing term:

---

[17] The investigation of model uncertainty is expanded and explicitly modeled in the context of return predictability using the Bayesian Model Averaging framework by (Avramov 2000; Cremers 2002), among others. See Section 5.

[18] α is commonly interpreted as a representation of the skill of an active portfolio manager. (Pastor and Stambaugh 2000) point out this interpretation is not infal-

$$E(r_t) = \alpha + \beta' E(f_t) \quad (1.14)$$

In a single factor model such as the CAPM, the benchmark portfolio is the market portfolio. In a multifactor model, the benchmarks could be zero-investment, non-investable portfolios whose behavior replicates the behavior of an underlying risk factor (sometimes called factor-mimicking portfolios)[19] or factors extracted from the cross-section of stock returns using principal components analysis.[20]. (Pastor 2000) investigates the implications for portfolio selection of varying prior beliefs about α. When beliefs about a pricing model are expressed, the prior mean of α, α_0, is set equal to zero. It could have a non-zero value, when, for example, the investor expresses uncertainty about an analyst's forecast. The prior variance σ_α of α reflects the investor's degree of confidence in the prior mean – a zero value of σ_α represents dogmatic belief in the validity of the model; $\sigma_\alpha = \infty$ suggests complete lack of confidence in its pricing power. (Pastor and Stambaugh 1999), investigating the cost of equity of individual firms, suggest that α_0 could be set equal to the average ordinary least squares estimate from a subset (cross-section) of firms sharing common characteristics.

(Pastor 2000) assumes normality of stock and factor returns, and conjugate uninformative priors for all parameters in Eq. (1.12) but α. In the special case of one stock and one benchmark, the optimal weight in the stock is shown to be proportional to the ratio of the posterior mean of α and the posterior mean of the residual variance, $\tilde{\alpha}/\tilde{\sigma}^2$. The posterior mean $\tilde{\alpha}$ has the form of a shrinkage estimator:

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix} = M^{-1}\left(\Psi^{-1}\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} + \left(\sigma^2 (X'X)^{-1}\right)^{-1}\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \right), \quad (1.15)$$

where

$$M = \Psi^{-1} + \left(\sigma^2 (X'X)^{-1}\right)^{-1},$$

lible. For example, the benchmarks used to define α might not price all passive investments.

[19] See, for example, (Fama and French 1993).

[20] See (Connor and Korajczyk 1986).

$\sigma^2 (X'X)^{-1}$ = (sample) covariance estimator of the least-squares estimators $\hat{\alpha}$ and $\hat{\beta}$,

$\left(\sigma^2 (X'X)^{-1}\right)^{-1}$ = sample precision matrix, and

Ψ^{-1} = prior precision matrix.

Pastor's results demonstrate greater stability of optimal portfolio weights, which take less extreme values. Examining the home bias that is observed in solutions to international asset allocation studies, Pastor finds that the holdings of foreign equity observed for U.S. investors is consistent with a prior standard deviation σ_α equal to 1% – evidence for strong belief in the efficiency of the U.S. market portfolio.[21]

Building upon the recognition of the fact that no model is completely accurate, (Pastor and Stambaugh 2000) undertake an empirical investigation comparing three asset pricing models from the perspective of optimal portfolio choice, while accounting for investment constraints. The models are: the CAPM, the Fama-French model, and the Daniel-Titman model[22] Pastor and Stambaugh explore the economic significance of different investors' perceptions of the degree of model accuracy by comparing the loss in certainty-equivalent return from holding portfolio *A* (the choice of an investor with complete faith in model *A*), when in fact the decision-maker has full confidence in model *B* or *C*. They observe that when the degree of certainty in a model is less than 100%, cross-model differences diminish (the certainty-equivalent losses are smaller). Investment constraints dramatically reduce the differences between models, which is in line with Wang's (Wang 1998) conclusion that imposing constraints acts to weaken the perception of inefficiency of the benchmark portfolio (see Section 4).

[21] Home bias is a term used to describe the observed tendency of investors to hold a larger proportion of their equity in domestic stocks than suggested by the weight of their country in the value-weighted world equity portfolio

[22] The (Fama and French 1993) model is a factor model in which expected stock returns are linear functions of the stock loadings on common pervasive factors. Book-to-market ratio and size-sorted portfolios are proxies for the factors. The Daniel and Titman (1997) model is a characteristic-based model. Expected returns are linear functions of firms' characteristics. Co-movements of stocks are explained with firms' possessing common characteristics, rather than being exposed to the same risk factors, as in the Fama-French model.

1.4. Testing Portfolio Efficiency

Empirical tests of mean-variance efficiency in the Bayesian context of both the CAPM and the Arbitrage Pricing Theory (APT) could be divided into two categories. The first one focuses on the intercepts of the multivariate regressions describing the CAPM

$$r_i = \alpha + \beta r_M + \varepsilon_i, \qquad i = 1, \ldots, N \tag{1.16}$$

and the APT

$$r_t = \alpha + \beta_1 f_{1,t} + \ldots + \beta_k f_{k,t} + u_t, \qquad t = 1, \ldots, T, \tag{1.17}$$

where returns are in risk-premium form (in excess of the risk-free rate), r_M in (1.16) is the market risk premium, $f_{j,t}$ is the risk premium (return) of factor *j* at time *t*, and β_j is return's exposure (sensitivity) to factor *j*. As in the previous section, the pricing implications of the CAPM and the APT yield the restriction that the elements of the parameter vector α are jointly equal to zero. Therefore, the null hypothesis of mean-variance efficiency is equivalent to the null hypothesis of no mispricing in the model.[23] The test relies on the computation of the posterior odds ratio.

At the heart of the tests in the second category lies the computation of the posterior distributions of certain measures of portfolio inefficiency. A strand of the pricing model testing literature focuses on the utility loss as a measure of the economic significance of deviations from the pricing restrictions, for example, by comparing the certainty-equivalent rate of return. (McCulloch and Rossi 1990) follow this approach.

1.4.1 Tests involving posterior odds ratios

(Shanken 1987; Harvey and Zhou 1990; McCulloch and Rossi 1991) employ posterior odds ratios to test the point hypotheses of the restrictions implied by the CAPM (the first two studies) and the APT (the third study).

The test of efficiency can be expressed in the usual way:

[23] When returns are expressed in risk-premium form, and expected returns are linear combinations of exposures to *K* sources of risk, the mean-variance efficient portfolio is a combination of the *K* benchmark (factor) portfolios and performing the test above in the context of the APT is equivalent to testing for mean-variance efficiency of this portfolio.

$$H_0 : \alpha = 0 \text{ vs. } H_1 : \alpha \neq 0 \tag{1.18}$$

The investor's belief that the null hypothesis is true is incorporated in the prior odds ratio, and then updated with the data to obtain the posterior odds ratio. The posterior odds ratio is the product of the ratio of predictive densities under the two hypotheses and the prior odds and is given by

$$G = \frac{p(\alpha = 0 \mid \mathbf{r})}{p(\alpha \neq 0 \mid \mathbf{r})} = \frac{p(\mathbf{r} \mid \alpha = 0)p(\alpha = 0)}{p(\mathbf{r} \mid \alpha \neq 0)p(\alpha \neq 0)}, \tag{1.19}$$

where $\mathbf{r}$ denotes the data.[24] It is often assumed that the prior odds is 1 when no particular prior intuition favoring the null or the alternative exists. Then, G becomes:

$$G = \frac{\int L(\beta, \Sigma \mid \alpha = 0)p_0(\beta, \Sigma)d\beta\, d\Sigma}{\int L(\beta, \Sigma, \alpha \mid \alpha \neq 0)p_1(\alpha, \beta, \Sigma)d\alpha\, d\beta\, d\Sigma}, \tag{1.20}$$

where $L(\beta, \Sigma \mid \alpha = 0)$ is the likelihood function $L(\alpha, \beta, \Sigma)$ evaluated at $\alpha = 0$. Since the posterior odds ratio is interpreted as the probability that the null is true divided by the probability that the alternative is true, a low value of the posterior odds provides evidence against the null hypothesis that the benchmark portfolio is mean-variance efficient.

Assume the disturbances in Eq. (1.16) are identically and independently distributed (*i.i.d.*) normal with a zero mean vector and a covariance matrix Σ. (Harvey and Zhou 1990) explore three distributional scenarios – a multivariate Cauchy distribution, a multivariate normal distribution, and a Savage density ratio approach. In the first two scenarios, the prior distribution under the null is taken to be a diffuse one:

$$p_0(\beta, \Sigma) \propto |\Sigma|^{-(N+1)/2}. \tag{1.21}$$

Under the alternative, the prior is

$$p_1(\alpha, \beta, \Sigma) \propto |\Sigma|^{-(N+1)/2} f(\alpha \mid \Sigma), \tag{1.22}$$

where $f(\alpha \mid \Sigma)$ is the prior density function of α (a multivariate Cauchy or a multivariate normal). Following (McCulloch and Rossi 1991), Harvey

[24] We assume that $p(\alpha = 0)$ and $p(\alpha \neq 0)$ are strictly greater than zero.

and Zhou investigate also the so-called Savage density ratio method,[25] asserting a conjugate prior under the alternative hypothesis, $p_1(\alpha,\beta,\Sigma) = N(\alpha,\beta \mid \Sigma)IW(\Sigma)$ (*N* denotes normal density, *IW* denotes inverted Wishart density)). The prior under the null is:

$$p_0(\beta,\Sigma) = p_1(\alpha,\beta,\Sigma \mid \alpha = 0) = \left.\frac{p_1(\alpha,\beta,\Sigma)}{\int p_1(\alpha,\beta,\Sigma)d\beta\, d\Sigma}\right|_{\alpha=0} \tag{1.23}$$

Large deviations of the intercepts from zero, under the multivariate normal prior, intuitively, provide greater evidence against the null hypothesis than large deviations from zero under the multivariate Cauchy prior. Therefore, the normal prior is expected to produce lower posterior odds ratio than the Cauchy prior.

The Savage density assumption leads to a simplification of the posterior odds. Assuming a prior odds ratio equal to 1,

$$G = \left.\frac{p(\alpha \mid \mathbf{r})}{p(\alpha)}\right|_{\alpha=0}, \tag{1.24}$$

where both the marginal posterior density of α in the numerator and the prior density in the denominator can be shown to be multivariate Student-*t* densitites.

In an examination of the efficiency of the market index, (Harvey and Zhou 1990) find that the posterior odds increase monotonically for increasing levels of dispersion in the prior distributions. Both the Cauchy and the normal priors provide evidence against the null. The posterior probability of mean-variance efficiency varies between 8.9% and 15.5% under the normal assumption, and between 26.2% and 27.2% under the Cauchy assumption. The Savage prior case is analyzed for three different prior assumptions of relative efficiency of the market portfolio, reflected in the choice of hyperparameters of β and Σ.[26] The Savage prior offers more

[25] The Savage density ratio method involves selecting a particular form of the prior density under the null, as in Eq. (1.23), which results in the simplification of the posterior odds ratio in Eq. (1.24).

[26] Relative efficiency is measured by the correlation ρ between the given benchmark index and the tangency portfolio; $\rho = 1$ implies efficiency of the benchmark. (Shanken 1987) shows that in the presence of a risk-free asset, ρ is equal to the ratio between the Sharpe measure (ratio) of the benchmark port-

evidence against the null, compared to the normal and Cauchy priors – the probability of efficiency is generally less than 1%.

(McCulloch and Rossi 1991) explore the pricing implications of the APT and observe great variability of the posterior odds ratio in response to changing levels of spread of the Savage prior.[27] The ratio in the high-spread specification exceeds the one in the low-spread case by more than 40 times when a five-factor model is considered. Overall, evidence against the null hypothesis is weak in the case of the one-factor model (except in the high-variance scenario) and mixed in the case of the five-factor model. McCulloch and Rossi caution, however, against drawing conclusions about the benefit of adding more factors to the one-factor model. The addition of factors needs to be analyzed in a different posterior-odds framework, in which the restriction of zero coefficients of the new factors is imposed.

1.4.2 Tests involving inefficiency measures

Investors are often less interested in an efficiency test offering a "binary" outcome (reject/do not reject) than in an investigation of the degree of inefficiency of a benchmark portfolio. (Kandel, McCulloch, and Stambaugh 1995) target this argument and develop a framework for testing the CAPM, in which the posterior distribution of an inefficiency measure is computed.[28] (Wang 1998) extends their analysis to incorporate investment constraints.

Denote by p the portfolio whose efficiency is being tested and by x the efficient portfolio with the same variance as p. Then, the observation that the expected return of p is less than or equal to the expected return of x immediately suggests an intuitive measure of portfolio p's inefficiency:

folio and Sharpe measure of the tangency portfolio (which is the maximum Sharpe measure).

[27] A parallel could be drawn between McCulloch and Rossi's (McCulloch 1990, 1991) investigation and the traditional two-pass regression procedure for testing the APT. The authors first extract the factors using the principal components approach of (Connor and Korajzcyk 1986) and then perform the Bayesian analysis. In contrast, (Geweke and Zhou 1996) adopt a single-stage procedure in which the posterior distribution of a measure of the APT pricing error is obtained numerically. Admittedly, the Geweke-Zhou approach could only be employed to a relatively small number of assets, in contrast to the McCulloch-Rossi approach.

[28] (Shanken 1987; Harvey and Zhou 1990) also discuss similar measures.

$$\Delta = \mu_x - \mu_p, \tag{1.25}$$

where μ_j denotes the expected return of portfolio *j*. The benchmark portfolio is efficient if and only if $\Delta = 0$. The non-negative value of Δ could also be interpreted as the loss of expected return from holding portfolio *p* instead of the efficient portfolio *x* (carrying the same risk as *p*). Another measure of inefficiency explored by Kandel, McCulloch, and Stambaugh is ρ, the correlation between *p* and any efficient portfolio. The posterior density of Δ and ρ does not have a closed-form solution under standard diffuse prior assumptions about the mean vector μ and the covariance matrix Σ of the risky asset returns. An application of the Monte Carlo methodology, however, makes its evaluation straightforward. Suppose the posterior density of the mean and covariance are given by $p(\mu \mid \Sigma, \mathbf{r})$ and $p(\Sigma \mid \mathbf{r})$, respectively. Then, a draw from the (approximate) posterior distribution of Δ and ρ is obtained by drawing repeatedly from the posterior distributions of μ and Σ and then computing the corresponding values of Δ and ρ.

Kandel, McCulloch, and Stambaugh observe an interesting divergence of results depending on whether or not a risk-free asset is available in the capital market. For example, in the absence of a risk-free asset, most of the mass of ρ's posterior distribution lies between -0.1 and 0.3, while when the risk-free asset is included, the posterior mass shifts to the interval 0.89 to 0.94 (suggesting a shift from a very weak to a very strong correlation between the benchmark and the efficient portfolio). Similarly, the posterior mass of Δ lies farther away from $\Delta = 0$ in the former than in the latter case. An investigation into the extent that the data influence the posterior of ρ reveals that informative, rather than diffuse, priors are necessary to extract the information of inefficiency contained in the data in the presence of a risk-free asset and, in general, the prior's influence on the posterior is strong. When the risk-free asset is excluded, the data update the prior better, and the results show that the benchmark portfolio (composed of NYSE and AMEX stocks) is highly correlated with the efficient portfolio.

The methodology of Kandel, McCulloch, and Stambaugh is easily adapted to account for investment constraints in testing for mean-variance efficiency of a portfolio. (Wang 1998) proposes to modify Δ in the following way to incorporate short-sale constraints:

$$\tilde{\Delta} = \max\left[x'\mu - x_p'\mu \mid x \geq 0,\ x'\Sigma x \leq x_p'\Sigma x_p\right], \qquad (1.26)$$

where x_p are the weights of the given benchmark portfolio under consideration, x are the weights of the efficient portfolio, and $x'\mu$ and $x_p'\mu$ are the expected portfolio returns, denoted by μ_x and μ_p, respectively, in (1.25). The constraint modification to reflect a 50% margin requirement[29] is $x_i \geq -0.5,\ i = 1,\ldots,N$. For each set of draws of the approximate posteriors of μ and Σ, the constrained optimization in (1.26) is performed and a draw of Δ is obtained.

Wang compares the posterior distributions of the inefficiency measures with and without investment constraints. When no constraints are imposed, the posterior mean of $\tilde{\Delta}$ is 20.9% (indicating that a portfolio outperforming the benchmark by 20% could be constructed). Imposing the 50% margin constraint brings the values of the posterior mean of $\tilde{\Delta}$ down to 8.37%, while when short sales are not allowed, the posterior mean decreases to 4.25%. Thus, the benchmark's inefficiency decreases as stricter investment constraints are included in the analysis. Additionally, (Wang 1998) observes that uncertainty about the degree of mispricing declines with the imposition of constraints, making the posterior distribution of $\tilde{\Delta}$ less dispersed.

1.5 Return Predictability

Predictability in returns impacts optimal portfolio choice in several ways. First, it brings in horizon effects. Second, it makes possible the implementation of market timing strategies. Third, it introduces different sources of hedging demand. In this section we will explore how these three consequences of predictability are examined in the Bayesian literature.

With the exception of (Kothari and Shanken 1997) who investigate a Bayesian test of the null hypothesis of no predictability, most of the predictability literature focuses on the implications of predictability for the optimal portfolio choice, rather than on accepting or rejecting the null hy-

[29] A 50% margin requirement is a restriction on the size of the total short sale position an investor could take. The short sale position can be no more than 50% of the invested capital.

pothesis, since portfolio performance and utility gains (losses) provide natural measures to assess predictability power.

1.5.1 The static portfolio problem

The vector autoregressive (VAR) framework is a convenient and compact tool to model the return-generating process and the dynamics of the endogenous predictive variables. For the simple case of one predictor, its form is:

$$\begin{aligned} r_t &= \alpha + \beta x_{t-1} + \varepsilon_t \\ x_t &= \theta + \rho x_{t-1} + u_t, \end{aligned} \tag{1.27}$$

where r_t is the excess stock return (return on a portfolio of stocks) in period t, x_{t-1} is a lagged predictor variable, whose dynamics is described by a first-order autoregressive model, and ε_t and u_t are correlated disturbances. The vector $(\varepsilon_t, u_t)'$ is assumed to have a bivariate normal distribution with a zero mean vector and a covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{pmatrix}.$$

The predictor is a variable such as the dividend yield, the book-to-market ratio, and interest rate variables, or lagged values of the continuously compounded excess return r_t.[30]

The dividend yield is considered a prime predictor candidate and all of the studies discussed below use it as the sole return predictor.

The investor maximizes the expected utility, weighted by the predictive distribution as in Eq. (1.1).

(Kandel and Stambaugh 1996) examine the problem in Eq. (1.1) in a static, single-period investment horizon setting, while (Barberis 2000) extends it to consider multi-period horizon stock allocations with optimal rebalancing. Kandel and Stambaugh investigate a no-predictability informative prior for B and Σ. They do so by constructing it as the posterior distribution that would result from combining the diffuse prior

---

[30] Numerous empirical studies of predictability have identified variables with predictive power. See, for example, (Fama 1991).

$p(B,\Sigma) \propto |\Sigma|^{-(N+2)/2}$ with a hypothetical sample identical to the real sample, save for a sample coefficient of determination R^2 equal to zero.[31] The behavior of the optimal stock allocations is analyzed over a range of values of the predictors, for a number of samples that differ by the number of predictors *N*, the sample size *T*, and the regression R^2. Kandel and Stambaugh's results confirm an intuitive relation between the optimal stock allocation and the current value of the predictor variable, x_T. Specifically, the greater the positive difference between the one-step ahead fitted value $\hat{b}\,x_T$ and the returns' long-term average $\bar{r} = \hat{b}\,\bar{x}$, the higher the stock allocation.

Kandel and Stambaugh put forward a related criterion for assessing the economic significance of predictability evidence. The optimal allocation ω^a in the case when $x_T = \bar{x}$ (where $\bar{x}$ is the long-term average of the predictor variable) is no longer optimal when $x_T \neq \bar{x}$. Then, a comparison of the certainty-equivalent returns associated with the expected utilities of the optimal allocations when $x_T = \bar{x}$ and when $x_T \neq \bar{x}$ allows one to examine the economic implications (if any).

(Kandel and Stambaugh 1996) emphasize the important departure of the evidence of economic significance from the evidence of statistical significance. For example, given an R^2 (unadjusted) from the predictive regression of only 0.025 (implying a p-value of 0.75 of the standard regression *F* statistic), the investor optimally allocates 0% of his wealth to stocks when predicted return $\hat{b}\,x_T$ is one standard deviation below its long-term average $\bar{r}$, but 61% when $\hat{b}\,x_T = \bar{r}$, under a diffuse prior and a coefficient of risk aversion equal to 2. Under the no-predictability informative prior, the allocations are, respectively, 53% and 83%. Therefore, statistical insignificance of the predictability evidence does not translate into economic insignificance.

The mechanism through which predictability affects portfolio choice is further enriched by the investigation of (Barberis 2000), who ties the Kandel and Stambaugh's framework to the issue of a varying investment horizon. Incorporating parameter uncertainty into the portfolio problem tends

[31] In a related paper, (Stambaugh 1999) characterizes the economic importance of the sample evidence of predictability by considering hypothetical samples carrying the same information content about B and Σ as the actual sample but differing in the value of y_T.

to reduce optimal stock holdings, and this horizon effect is, not surprisingly, stronger at a long-horizon than at a short-horizon. In contrast, when the possibility of predictable returns is taken into account, perceived risk of stocks by a buy-and-hold investor at long horizons diminishes because the variance of cumulative returns grows slower than linearly with the horizon. Thus, a higher proportion of wealth is allocated to stocks at long horizons compared with the case when returns are assumed to be *i.i.d.* and these differences increase with the horizon.[32] Analyzing the interaction of the two opposing tendencies, Barberis finds that introducing estimation risk, in a static setting, reduces the horizon effect for a risk-averse investor – the uncertainty about the process parameters adds to uncertainty about the forecasting power of the predictor(s) and increases risk at longer horizons. As a result, the 10-year buy-and-hold portfolio strategy of an investor with a risk aversion parameter of 10, who takes both predictability and uncertainty into account, results in up to a 50% lower allocation compared to the case of predictability only, with no estimation risk.

Both (Barberis 2000) and (Stambaugh 1999) explore the sensitivity of the optimal allocation to varying the initial predictor's value, x_0. Long-horizon allocations under uncertainty generally increase with the horizon for low starting values of the predictor and decrease for high starting values, leading to a lesser sensitivity to the predictor's starting value. Stambaugh demonstrates that treating x_0 as a stochastic realization of the same process that generated $x_1, x_2, ..., x_T$, compared to considering it fixed, brings in additional information about the regression parameters and changes their posterior means. He observes that, when estimation risk is incorporated, the long-horizon (in particular, 20-years) optimal allocation is often decreasing in the predictor, even though expected return is not. This pattern can be ascribed to the skewness of the predictive distribution. Incorporating uncertainty (particularly the uncertainty about the autoregressive coefficient of the predictor) induces positive skewness for low initial values of the predictor (leading to high allocations) and negative skewness for high initial values (leading to low allocations).

[32] Empirically observed mean-reversion in returns (negative serial correlation) helps explain the horizon effect. However, Barberis notes that predictability itself may be sufficient to induce this effect, if not mean-reversion. Specifically, the negative correlation between the unexpected returns and the dividend yield innovations is one condition for the horizon effect. See also (Avramov 2000), and Section 5 below.

1.5.2 The dynamic portfolio problem

As mentioned earlier, market-timing is one of the modifications to the portfolio allocation problem resulting from predictability. Suppose that an investor at time *T* with an investment horizon $T+\hat{T}$ has a dynamic strategy and rebalances at each of the dates $T+1,\ldots,T+\hat{T}-1$. The new intertemporal context of the problem allows us to consider a new aspect of parameter uncertainty[33] – not only does the investor not know the true parameters of the return generating process but the relationship between the returns and the predictors may also be time-varying. At time *T*, the Bayesian investor solves the portfolio problem taking into account that at each rebalancing date, the posterior distribution of the parameters is updated with the new information. It turns out that this "learning" (Bayesian updating) process plays an important role in the way the investment horizon affects optimal allocations.[34] The underlying factor driving changes in allocations across horizons is now a hedging demand – a risk-averse investor attempts to hedge against the perceived changes in the investment opportunity set (equivalently, in the state variables).[35]

(Barberis 2000) considers a discrete dynamic setting with *i.i.d.* stock returns to explore the effects of learning about the unconditional mean of returns and finds that uncertainty induces a very strong negative hedging demand at long horizons.[36] A long-horizon investor who admits the possibility of learning about the unconditional mean in the future allocates substantially less to stocks than an investor with a buy-and-hold strategy.

[33] An early discussion of the Bayesian dynamic portfolio problem in a discrete-time setting (without accounting for predictability) can be found in (Winkler and Barry 1975). (Grauer and Hakansson 1990) examine the performance of shrinkage and CAPM estimators in a dynamic, discrete-time setting.

[34] (Merton 1971; Williams 1977) show that incorporating learning in a dynamic problem leads to the creation of a new state variable representing the investor's current beliefs. Here, the new state variables are the posterior estimates of the unknown parameters, whose dynamics might be nonlinear. If learning is ignored, the current dividend yield is the only state variable, and it fully characterizes the predictive return distribution.

[35] Hedging demands are introduced by (Merton 1973). An investor who is more risk averse than the log-utility case (i.e., with a coefficient of risk aversion higher than 1) aims at hedging against reinvestment risk and increases his demand for stocks when their expected returns are low. Recall that expected stock returns are negatively correlated with realized stock returns.

[36] The intuition behind the negative hedging demand is that an unexpectedly large return leads to an upward revision of unconditional expected return

While the framework introduced by Barberis involves learning about the unconditional mean of returns only, (Brandt, Goyal, Santa-Clara, and Stroud 2004) address simultaneous learning about all model parameters. The utility loss from ignoring learning is substantial but is negatively related to the amount of past data available and to the investor's risk aversion parameter. Brandt et. al observe that the utility gains from accounting for uncertainty or for learning are of comparable size, and increasing with the horizon and the current predictor value. They break down the hedging demand and analyze its components – (1) the positive hedging component arising from the negative correlation between returns and changes in the dividend yield and (2) the negative hedging component due to the positive correlation between returns and changes in the model parameters. The aggregate effect can be positive at short horizons (up to five years) but turns negative for longer horizons.

Brandt et. al observe that learning about the mean of the dividend yield and about the correlation between returns and the dividend yield induce a positive hedging demand which could partially offset the negative hedging demand above.

A question of practical importance to investors is whether it is possible to take advantage of the evidence of predictability in practice. (Lewellen and Shanken 2002) offer an insightful answer which is unfortunately disappointing. They find that patterns in stock returns, like predictability, which a researcher observes, cannot be perceived by a rational investor.

1.5.3 Model Uncertainty

(Avramov 2000; Cremers 2002) address what could be viewed as a deficiency shared by the predictability investigations above – model uncertainty, introduced by selecting and treating a certain return-generating process as if it were the true process. At the heart of Bayesian Model Averaging (BMA) is computing a weighted Bayesian predictive distribution of the "grand" model, in which individual models are weighted by their posterior distributions.[37]

Suppose that each individual model has the form of a linear predictive regression:

$$r_t = x_{j,t-1}B_j + \varepsilon_{j,t}, \tag{1.28}$$

[37] If K variables are entertained as potential predictors, there are 2^K possible models.

where

r_t = (N x 1) vector of excess returns on N portfolios,

$x_{j,t-1} = (1, z_{j,t-1})$,

$z_{j,t-1}$ = (k_j x 1) vector of predictors, observed at the end of t - 1, that belong to model j,

B_j = (($k_j + 1$) x N) matrix of regression coefficients, and

$\varepsilon_{j,t}$ = disturbance of model j, assumed to be normally distributed with mean 0 and covariance matrix Σ_j (Avramov) or Σ (Cremers).[38] The framework requires that two groups of priors be specified – model priors (i.e., priors of inclusion of each variable in an individual model), and priors on the parameters B_j and Σ_j of each model. Each model could be viewed equally likely a priori, and assigned the diffuse prior $P(M_j) = 1/2^K$, where M_j, $j = 1, ..., K$ is the j th model. A different prior ties the model selection problem with the variable selection problem, as in (Cremers 2002):

$$P(M_j) = \rho^{k_j}(1-\rho)^{K-k_j}, \tag{1.29}$$

where ρ denotes the probability of inclusion of a variable in model j (assumed equal for all variables, but easily generalized to reflect different degrees of prior confidence in subsets of the predictors).[39]

No predictability (no confidence in any of the potential predictors) is equivalent to not including any of the explanatory variables in the regression in (1.28). Then, returns are *i.i.d.*, and, using (1.29), the model prior is $P(M_j) = (1-\rho)^K$.

The posterior probability of model M_j is given by

[38] Both Avramov and Cremers treat the regression parameters B_j as fixed. (Dangl, Halling and Randl 2005) consider a BMA framework with time-varying parameters.

[39] (Pastor and Stambaugh 1999) observe that when the set of models considered includes one with a strong theoretical motivation (e.g., the CAPM), assigning a higher prior model probability to it is reasonable.

$$P\left(M_j \mid \Phi_t\right)=\frac{P\left(\Phi_t \mid M_j\right)P\left(M_j\right)}{\sum_{j=1}^{2^K} P\left(\Phi_t \mid M_j\right)P\left(M_j\right)}, \tag{1.30}$$

where Φ_t denotes all sample information available up to time *t*. The marginal likelihood function $P\left(\Phi_t \mid M_j\right)$ is obtained by integrating out the parameters B_j and Σ_j:

$$P\left(\Phi_t \mid M_j\right)=\frac{L\left(B_j,\Sigma_j;\Phi_j,M_j\right)P\left(B_j,\Sigma_j \mid M_j\right)}{P\left(B_j,\Sigma_j \mid \Phi_j,M_j\right)}, \tag{1.31}$$

where $L\left(B_j,\Sigma_j;\Phi_j,M_j\right)$ is the likelihood function corresponding to model M_j, $P\left(B_j,\Sigma_j \mid M_j\right)$ is the joint prior and $P\left(B_j,\Sigma_j \mid \Phi_j,M_j\right)$ is the joint posterior of the model parameters.

The weighted predictive return distribution is given by:

$$P\left(R_{t+\hat{T}} \mid \Phi_t\right)=\sum_{j=1}^{2^K} P\left(M_j \mid \Phi_t\right)\int P\left(B_j,\Sigma_j \mid \Phi_t,M_j\right) \tag{1.32}$$

$$\text{x } P\left(R_{t+\hat{T}} \mid B_j,\Sigma_j,M_j,\Phi_t\right)dB_j$$

where $R_{t+\hat{T}}$ is the predicted cumulative return over the investment horizon $\hat{T}$.

To express prior views on predictability, Cremers considers three quantities directly related to it: the expected coefficient of determination, $E\left(R^2\right)$, the expected covariance of returns, $E(\Sigma)$, and the probability of variable inclusion, ρ. He asserts conjugate priors for the parameters and includes a hyperparameter which penalizes large models. (Avramov 2000) uses a prior specification for B_j and Σ_j based on the one of (Kandel and Stambaugh 1996). The size of the hypothetical prior sample, T_0, determines the strength of belief in lack of predictability (as T_0 increases, belief in predictability diminishes).

Both Cremers and Avramov find in-sample and out-of-sample evidence of predictability.[40] Avramov estimates a VAR model similar to Eq. (1.27). His variance decomposition of predicted stock returns into model risk, estimation risk, and uncertainty due to forecast error shows that model uncertainty plays a bigger role than parameter uncertainty. He finds that model uncertainty is proportional to the distance of the current predictor values from their sample means. To gauge the economic significance of accounting for model uncertainty, Avramov uses the difference in certainty equivalent metric and reaches an interesting result: the optimal allocation for a buy-and-hold investor is not sensitive to the investment horizon. This finding is contrary to the general findings of the Bayesian predictability literature. He ascribes the finding to the positive correlation between the unexpected returns and the innovations on the predictors with the highest posterior probability. The dividend yield, which is most often the only predictor in predictability investigations, has a lower posterior probability than the term premium and market premium predictors, and therefore, a smaller influence in the "grand" model (confirmed by Cremers' results).

1.6. Conclusion

The application of Bayesian methods to investment management is a vibrant and constantly evolving one. Space constraints did not allow us to review many worthy contributions.[41] Active research is being conducted in the areas of volatility modeling, time series models, and regime-switching models. Recent examples of stochastic volatility investigations include (Jacquier, Polson, and Rossi 1994; Mahieu and Schotman 1998; Uhlig 1997); time series models are explored by (Aguilar and West 2000; Kleibergen and Van Dijk 1993; Henneke, Rachev, and Fabozzi 2006); regime switching has been discussed by (Hayes and Upton 1986; So, Lam, and Li 1998), and employed by (Neely and Weller 2000).

Bayesian methods provide the necessary toolset when heavy-tailed characteristics of stock returns are analyzed. (Buckle 1995; Tsionas 1999) model returns with symmetric stable distribution, while (Fernandez and Steel 1998) develop and employ a skewed Student *t* parameterization.

---

[40] Other empirical studies of return predictability include (Lamoureux and Zhou 1996; Neely and Weller 2000; Shanken and Tamayo 2001; Avramov and Chordia 2005).

[41] For a more detailed discussion, see (Rachev, Hsu, Bagasheva, and Fabozzi 2007).

These investigations have been made possible thanks to great advances in computational methods, such as Markov Chain Monte Carlo (see Bauwens, Lubrano, and Richard 2000)).

The individual investment management areas mentioned above, several of which were surveyed in the previous sections, will continue to evolve in future works. We see the main challenge lying in their integration into coherent financial models. Without doubt, Bayesian methods are the indispensable framework for embracing and addressing the ensuing complexities.

References

Aguilar O, West M (2000) Bayesian dynamic factor models and portfolio allocation. J of Business and Economic Statistics 18 (3): 338-357

Avramov D (2000) Stock return predictability and model uncertainty. J of Financial Economics 64: 423-458

Avramov D, Chordia T (2005) Predicting stock returns, http://ssrn.com/abstract=352980

Barberis N (2000) Investing for the long run when returns are predictable. J of Finance 55 (1): 225-264

Barry C (1974) Portfolio analysis under uncertain means, variances, and covariances. J of Finance 29 (2): 515-522

Bauwens L, Lubrano M, Richard JF (2000) Bayesian inference in dynamic econometric models. Oxford University Press

Berger J (1980) Statistical decision theory. Springer, New York

Best MJ, Grauer RR (1991) Sensitivity analysis for mean-variance portfolio problems. Management Science 37 (8): 980-989

Black F, Litterman R (1990) Asset allocation: Combining investor views with market equilibrium. Goldman Sachs

Black F, Litterman R (1991) Global asset allocation with equities, bonds, and currencies. Fixed Income Research, Goldman Sachs

Black F, Litterman R (1992) Global portfolio optimization. Financial Analysts Journal Sept-Oct: 28-43

Brandt MW, Goyal A, Santa-Clara P, Stroud J (2004) A simulation approach to dynamic portfolio choice with an application to learning about return predictability. NBER Working Paper Series

Brown SJ (1976) Optimal portfolio choice under uncertainty: a Bayesian approach. Ph.D. thesis, University of Chicago

Brown SJ (1979) The effect of estimation risk on capital market equilibrium. J of Financial and Quantitative Analysis 14 (2): 215-220

Buckle DJ (1995) Bayesian inference for stable distributions. J of the American Statistical Association 90: 605-613

Chen S, Brown SJ (1983) Estimation risk and simple rules for optimal portfolio selection. 38(4): 1087-1093

Connor G, Korajczyk R (1986) Performance measurement with the arbitrage pricing theory: A new framework for analysis. J of Financial Economics 15:373-394

Cremers KJM (2002) Stock return predictability: A Bayesian model selection perspective. The Review of Financial Studies 15 (4): 1223-1249

Dangl T, Halling M, Randl O (2005) Equity return prediction: Are coefficients time-varying? 10th Symposium on Finance, Banking, and Insurance. University of Karlsruhe.

Daniel K, Titman S (1997) Evidence on the characteristics of cross sectional variation in stock returns. J of Finance 52 (1): 1-33

Dumas B, Jacquillat B (1990) Performance of currency portfolios chosen by a Bayesian technique: 1967-1985. J of Banking and Finance 14: 539-558

Fama E (1965) The behavior of stock market prices. J of Business 38:34-105

Fama E (1991) Efficient capital markets: II. J of Finance 46 (5): 1575-1617

Fama E, French K (1993) Common risk factors in the returns on stocks and bonds. J of Financial Economics 33: 3-56

Fernandez C, Steel M (1998) On Bayesian modeling of fat tails and skewness. J of the American Statistical Association 93: 359-371

Frankfurter GM, Phillips HE, Seagle JP (1971) Portfolio selection: the effects of uncertain means, variances, and covariances. J of Financial and Quantitative Analysis 6 (5): 1251-1262

Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. J of Financial and Quantitative Analysis 21 (3): 293-305

Geweke J, Zhou G (1996) Measuring the pricing error of the arbitrage pricing theory. The Review of Financial Studies 9 (2): 557-587

Grauer RR, Hakansson NH (1990) Stein and CAPM estimators of the means in asset allocation. International Review of Financial Analysis 4 (1): 35-66

Harvey C, Zhou G (1990) Bayesian inference in asset pricing tests. J of Financial Economics 26: 221-254

Hays PA, Upton DE (1986) A shifting regimes approach to the stationarity of the market model parameters of individual securities. J of Financial and Quantitative Analysis 21 (3): 307-321

He G, Litterman R (1999) The intuition behind Black-Litterman model portfolios. Investment Management Division, Goldman Sachs

Henneke J, Rachev S, Fabozzi F (2006) MCMC based estimation of MS-ARMA-GARCH models. Technical Reports, Department of Probability and Statistics, UCSB

Jacquier E, Polson NG, Rossi PE (1994) Bayesian analysis of stochastic volatility models. J of Business and Economic Statistics 12 (4): 371-389

James W, Stein C (1961) Estimation with quadratic loss. In: Proceedings of the 4th Berkeley Simposium on Probability and Statistics I. University of California Press, pp 361-379

Jobson JD, Korkie B (1980) Estimation of Markowitz efficient portfolios. J of the American Statistical Association 75 (371): 544-554

Jobson JD, Korkie B (1981) Putting Markowitz theory to work. Journal of portfolio management 7 (4): 70-74

Jobson JD, Korkie B, Ratti V(1979) Improved estimation for Markowitz portfolios using James-Stein type estimators. In: Proceedings of the American Statistical Association, Business and Economics Section. American Statistical Association, Washington

Jorion P (1985) International portfolio diversification with estimation risk. J of Business 58 (3): 259-278

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. J of Financial and Quantitative Analysis 21 (3): 279-292

Jorion P (1991) Bayesian and CAPM estimators of the means: Implications for portfolio selection. J of Banking and Finance 15: 717-727

Kandel S, McCulloch R, Stambaugh R (1995) Bayesian inference and portfolio efficiency. The Review of Financial Studies 8 (1): 1-53

Kandel S, Stambaugh R (1996) On the predictability of stock returns: An asset-allocation perspective. J of Finance 51 (2): 385-424

Kleibergen F, Van Dijk HK (1993) Non-stationarity in GARCH models: A Bayesian analysis. J of Applied Econometrics 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques: S41-S61

Klein R, Bawa V (1976) The effect of estimation risk on optimal portfolio choice. J of Financial Economics 3: 215-231

Kothari SP, Shanken J (1997) Book-to-market, dividend yield, and expected market returns: A time-series analysis. J of Financial Economics 44: 169-203

Lamoureux CG, Zhou G (1996) Temporary components of stock returns: What do the data tell us? The Review of Financial Studies 9 (4): 1033-1059

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J of Empirical Finance 10: 603-621

Lee W (2000) Advanced theory and methodology of tactical asset allocation. John Wiley & Sons, New York

Lewellen J, Shanken J (2002) Learning, asset-pricing tests, and market efficiency. J of Finance 57 (3): 1113-1145

Litterman R, Winkelmann K (1998) Estimating covariance matrices. Risk Management Series, Goldman Sachs

Mahieu RJ, Schotman PC (1998) An empirical application of stochastic volatility models. J of Applied Econometrics 13 (4): 333-359

Markowitz H (1952) Portfolio selection. J of Finance 7 (1): 77-91

Markowitz H, Usmen N (1996) The likelihood of various stock market return distributions, Part I: Principles of inference. J of Risk and Uncertainty 13: 207-219

McCulloch R, Rossi P (1990) Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory. J of Financial Economics 28: 7-38

McCulloch R, Rossi P (1991) A Bayesian approach to testing the arbitrage pricing theory. J of Econometrics 49: 141-168

Merton R (1971) Optimum consumption and portfolio rules in a continuous-time model. J of Economic Theory 3: 373-413

Merton R (1973) An intertemporal capital asset pricing model. Econometrica 41: 867-887

Merton R (1980) An analytic derivation of the efficient portfolio frontier. J of Financial and Quantitative Analysis 7 (4): 1851-1872

Meucci A (2005) Risk and asset allocation. Springer, Berlin Heidelberg New York

Neely CJ, Weller P (2000) Predictability in international asset returns: A reexamination. J of Financial and Quantitative Analysis 35 (4): 601-620

Ortobelli S, Rachev S, Schwartz E (2004) The problem of optimal asset allocation with stable distributed returns. In: Krinik AC, Swift RJ (eds) Stochastic processes and functional analysis, Lecture notes in pure and applied mathematics, Marsel Dekker: 295-347

Pastor L (2000) Portfolio selection and asset pricing models. J of Finance 55 (1): 179-223

Pastor L, Stambaugh R (1999) Costs of equity capital and model mispricing. J of Finance 54 (1): 67-121

Pastor L, Stambaugh R (2000) Comparing asset pricing models: An investment perspective. J of Financial Economics 56: 335-381

Rachev S, Hsu J, Bagasheva B, Fabozzi F (2007) Bayesian methods in Finance. John Wiley & Sons, forthcoming

Satchell S, Scowcroft A (2000) A demystification of the Black-Litterman model: managing quantitative and traditional portfolio construction. J of Asset Management 1-2: 138-150

Shanken J (1987) A Bayesian approach to testing portfolio efficiency. J of Financial Economics 19: 195-215

Shanken J, Tamayo A (2001) Risk, mispricing, and asset allocation: Conditioning on the dividend yield. NBER Working Paper Series

Sharpe WF (1963) A simplified model for portfolio analysis. Management Science 9: 277-293

So MKP, Lam K, Li WK (1998) A stochastic volatility model with Markov switching. J of Business and Economic Statistics 16 (2): 244-253

Stambaugh R (1999) Predictive regressions. J of Financial Economics 54: 375-421

Tsionas E (1999) Monte Carlo inference in econometric models with symmetric stable disturbances. J of Econometrics 88: 365-401

Uhlig H (1997) Bayesian vector autoregressions with stochastic volatility. Econometrica 65 (1): 59-73

Wang Z (1998) Efficiency loss and constraints on portfolio holdings. J of Financial Economics 48: 359-375

Williams JT (1977) Capital asset prices with heterogeneous beliefs. J of Financial Economics 5: 219-241

Winkler RL, Barry BB (1975) A Bayesian model for portfolio selection and revision. J of Finance 30 (1): 179-192

Zellner A, Chetty VK (1965) Prediction and decision problems in regression models from the Bayesian point of view. J of the American Statistical Association 60:608-616