Research Support
University of Michigan Business School

August 1997

Hierarchical Bayes Methods for Market Model Estimation and Portfolio Selection

Working Paper #9712-16

by

Martin R. Young
University of Michigan Business School
Peter J. Lenk
University of Michigan Business School

Hierarchical Bayes Methods for Market Model Estimation and Portfolio Selection

Martin R. Young   Peter J. Lenk

myoung@umich.edu   plenk@umich.edu

University of Michigan School of Business

Department of Statistics and Management Science

Ann Arbor, MI 48109–1234

August, 1997

Abstract

The market model is an important construct for both portfolio managers and researchers in modern finance. For practitioners, market model coefficients are used to guide the construction of optimal portfolios. For academicians, the market model parameters play a fundamental role in explaining equilibrium asset prices and other market phenomena. This paper presents a hierarchical modeling procedure which can substantially improve the accuracy of market model parameter estimates, through incorporation of cross-sectional information. It is shown that this improvement in parameter estimation accuracy translates into substantial improvement in portfolio performance. Expressions are derived which characterize the sensitivity of portfolio performance to parameter estimation error. Evidence with NYSE data suggests that the hierarchical estimation technique leads to superior out-of-sample portfolio performance, when compared to alternative estimation approaches.

KEY WORDS: *Beta, Estimation Risk, Markov Chain Monte Carlo, Sensitivity, Shrinkage*

1 Introduction

Stochastic linear models for asset returns play a central role in modern financial theory and practice. Such models describe the return on each particular asset in a market as varying linearly with some exogenous factors: $y_{jt} = \alpha_j + \beta_j'\mathbf{f}_t + \epsilon_{jt}, \quad j = 1, \ldots, p, \quad t = 1, \ldots, n,$ where y_{jt} is the return of stock j at time t, $\mathbf{f}_t = (f_{t1}, \ldots, f_{tr})'$ are the values of the underlying factors at time t, α_j is the intercept and $\beta_j = (\beta_{j1}, \ldots, \beta_{jr})'$ are the factor coefficients for stock j, and ϵ_{jt} is a random deviation, independent of $\mathbf{f}_t$, with mean zero and variance v_j. For example, the traditional single factor market model (e.g., Elton and Gruber 1995) describes the covariation of asset returns via the model:

$$y_{jt} = \alpha_j + \beta_j m_t + \epsilon_{jt}, \tag{1}$$

where the single factor m_t represents a market index such as the S&P 500 index which measures the general performance of the stock market. More recently, Fama and French (1995) modeled the covariation of stock returns via the three factor model

$$y_{jt} - r_{ft} = \alpha_j + \beta_{1j}(m_t - r_{ft}) + \beta_{2j}(\text{SML}_t - r_{ft}) + \beta_{3j}(\text{HML}_t - r_{ft}) + \epsilon_{jt}, \tag{2}$$

where r_{ft} is the "risk-free" rate of return, SML_t ("Small Minus Large") is the return on a portfolio composed of the firms with smallest market value minus the return on a portfolio composed of firms with large market value, and HML_t ("High Minus Low") is the return on a portfolio composed of the firms with a high ratio of book value to market value minus the return on a portfolio composed of stocks with low values of this ratio.

The parameter values (α_j, β_j, v_j) for the different assets can be used to select portfolios that achieve optimal tradeoffs between risk and return (Elton and Gruber 1995). In addition, these parameters appear as explanatory variables in financial models for asset pricing (e.g., Fama and MacBeth 1973; Fama and French 1992; Fama and French 1995). The parameter values (α_j, β_j, v_j) of course cannot be directly observed, but can only be imperfectly estimated from finite samples of data. This unavoidable estimation error presents a problem to both theorists and practitioners.

The sensitivity of financial models to estimation error suggests two remedies: reducing estimation error, and correctly accounting for it. Vasicek (1973) noted in the context of a single factor model such as (1) that the estimate of a beta coefficient for a particular security could be improved by use of cross-sectional information. For example, if the betas of the stocks tend to range between 0.5 and 1.5, then a beta estimate of 2.0 is more likely to be an over-estimate than an under estimate. Thus, the estimate can possibly be improved by shrinking it towards the cross-sectional mean of 1.0. Barry (1973) developed a Bayesian approach to portfolio analysis, which includes the use of predictive distributions to account for parameter uncertainty. Rosenberg and James (1976) suggested that the incorporation of fundamental financial quantities such as firm size or liquidity - could lead to improved estimates of betas. More recently, Jorion (1986), Frost and Savarino (1986), and Board and Sutcliffe (1994) used Stein type shrinkage estimators to improve upon the usual least-squares procedure for obtaining parameter estimates for the portfolio selection problem, and Karolyi (1992) used multiple shrinkage to help estimate betas.

In this paper, we introduce a hierarchical market model and Bayesian estimation procedure which incorporate the above improvements to the simple least squares estimation procedure in a unified model. This hierarchical modeling procedure, which jointly models both the cross-sectional and the time-series variation in stock returns, has the desirable feature that the degree of adjustment of the usual least-squares estimates is determined automatically for each security, based on relevant attributes of the data. For securities whose parameter estimates are uncertain, either because the security has only been observed for a short period of time, or because the security's returns do not closely fit the market model; the parameter estimates will be substantially modified toward a cross-sectional mean, which may be determined by fundamental characteristics of the firm, such as firm size or industry sector. For securities whose least-squares parameter estimates are reliable, in the sense of having small standard errors, the estimates will be relatively unmodified.

In addition, the paper explores the impact of estimation error on portfolio selection. We demonstrate that the improvement in estimation accuracy achieved by hierarchical Bayes estimation leads to improved portfolio selection. We develop analytic measures which characterize the sensitivity of optimal portfolio allocations with respect to parameter estimation error. One result obtained is that *small* idiosyncratic variance for a given asset indicates potentially *large* error in the estimated portfolio allocation for that asset.

The paper has the following outline. Section 2 defines the hierarchical market model. Section 3 presents a simulation study of the hierarchical Bayesian (HB) parameter estimation and portfolio selection procedure, and in Section 4, historical data from the NYSE are analyzed by HB and alternative methods; with both the simulated and the real data, it is shown that the HB method out-performs the competitors, in terms of out-of sample forecast accuracy and portfolio performance. Section 5 presents a sensitivity analysis. Appendix A describes the numerical algorithm used to implement the HB method.

2 The Hierarchical Market Model

The linear factor model (3) relates the return y_{jt} for asset or firm j at time t to the returns $\mathbf{f}_t$ on a vector of economic factors by a simple linear regression:

$$y_{jt} = \alpha_j + \boldsymbol{\beta}_j \mathbf{f}_t + \epsilon_{jt}, \quad j = 1, \ldots, p, \quad t = 1, \ldots, n, \tag{3}$$

where the ϵ_{jt} are independent residuals, which will be assumed to have a normal distribution with mean 0 and variance v_j: $\mathrm{N}(0,\ v_j)$.

The hierarchical market model describes the cross sectional variation in the parameters α_j, $\boldsymbol{\beta}_j$, and v_j across the firms in the population. Here, the parameters $(\alpha_j, \boldsymbol{\beta}_j)$ are assumed to be related to covariates $\mathbf{z}_j^{\alpha}$ and $\mathbf{z}_j^{\beta}$ according to the linear regression equations:

$$\alpha_j = \theta_0' \mathbf{z}_j^{\alpha} + u_{j0}, \quad u_{j0} \sim \mathrm{N}(0,\ \Lambda_0), \tag{4}$$

$$\beta_{jk} = \theta_k' \mathbf{z}_j^{\beta} + u_{jk}, \quad u_{jk} \sim \mathrm{N}(0,\ \Lambda_k), \quad j = 1, \ldots, p, \quad k = 1, \ldots, r, \tag{5}$$

with u_{jk} mutually independent for $k = 0, 1, \ldots, r$. The covariates $\mathbf{z}_j^{\alpha}$ and $\mathbf{z}_j^{\beta}$ may contain such variables as firm size, leverage, and other accounting numbers, as well as indicator variables representing the industry segment for the jth firm.

The firm specific variances v_j may also be related to fundamental quantities of the firm. Let $\mathbf{z}_j^{\tau}$ be a vector of fundamental variables, and let $\tau_j = \log(v_j)$. τ_j can be modeled as:

$$\tau_j = \psi' \mathbf{z}_j^{\tau} + w_j, \quad w_j \sim \mathrm{N}(0,\ \delta). \tag{6}$$

Equation (6) establishes the prior distribution on the unknown firm specific variance $v_j = \exp(\tau_j)$.

Equations (3)-(6) describe a hierarchical regression model (Lindley and Smith 1972), sometimes referred to as a population model (Wakefield, Smith, Racine-Poon, and Gelfand 1994): the multiple shrinkage estimator of George (1986) is a related approach. Wakefield, Smith, Racine-Poon, and Gelfand (1994) and Müeller and Rosner (1994) describe alternative models that could be used for the Gaussian hyper-prior distributions; these alternatives include a Student-t distribution, and a mixture of Gaussian distributions.

As is discussed in section 3.2, the selection of optimal portfolios is based upon an estimate of the joint distribution of future returns $\{y_{jt}\}$. To obtain this distribution, one must model both the conditional distribution of the returns y_{jt} given the factor returns $\mathbf{f}_t$, and also the marginal distribution of the factor returns. Here, this latter will be treated as a multivariate normal:

$$\mathbf{f}_t \sim \mathrm{N}(\boldsymbol{\mu}_f,\ \boldsymbol{\Omega}_f). \tag{7}$$

Equations (3) and (7) together imply that the joint moments for the returns y_{jt} are given by:

$$\mathrm{E}[y_{jt}] = \alpha + \boldsymbol{\beta}_j'\boldsymbol{\mu}_f, \quad \mathrm{Var}[y_{jt}] = \boldsymbol{\beta}_j'\boldsymbol{\Omega}_f\boldsymbol{\beta}_j + v_j, \quad \mathrm{Cov}[y_{jt}, y_{kt}] = \boldsymbol{\beta}_j'\boldsymbol{\Omega}_f\boldsymbol{\beta}_k. \tag{8}$$

2.1 Prior Distributions on Model Hyperparameters

The parameters $\{\boldsymbol{\theta}_k, \boldsymbol{\Lambda}_k\}$, $k = 0, \ldots, r$, ψ, and δ in the prior distributions (4)-(6) are typically referred to as *hyperparameters*; to complete the specification of the Bayesian model, one will require prior distributions on these hyperparameters, as well as on the parameters $\boldsymbol{\mu}_f$ and $\boldsymbol{\Omega}_f$ which characterize the distribution of the independent variables in the market model (3). In this paper, we use non-informative priors for all of these parameters, in order to allow the data, rather than the priors, to determine posterior conclusions. The priors on the location parameters θ, ψ, and μ_f are (improper) uniform distributions over all possible parameter values. The prior on ${\Omega_f}^{-1}$ is also taken to be uniform. The prior on the Λ_k and on δ are taken to be $\mathrm{IG}(a_0,\ a_1)$, where $v \sim \mathrm{IG}(a_0,\ a_1)$ denotes that v^{-1} has the gamma distribution with mean a_0/a_1 and variance a_0/a_1^2. In the applications described in this paper. we use $a_0 = 1$, $a_1 = 0.1$, to provide proper, but very diffuse priors.[1]

The joint posterior distribution for the unknown model parameters in equations (3) - (7) cannot be evaluated analytically. Appendix A. though, describes how the model can be analyzed numerically, using a Markov chain Monte Carlo algorithm (Roberts and Smith 1993). Section 3 describes a simulation study designed to evaluate the performance of the hierarchical Bayes estimator.

3 Estimation Accuracy and Portfolio Performance

This section evaluates the hierarchical Bayes (HB) and least squares (LS) estimators in terms of estimation accuracy and portfolio performance. We show in section 3.1 that the HB estimators have significantly smaller estimation error than the LS estimators, and in section 3.2 we show that this improvement in estimation accuracy leads to improved portfolio performance. In section 5, we develop analytic expressions characterizing the sensitivity of portfolio performance to estimation accuracy.

[1]Berger (1985, page 187) discusses the fact that the usual Jeffreys' non informative prior cannot be used for the scale parameters in a hierarchical model, since such priors can lead to improper posterior distributions.

3.1 Parameter Estimation Accuracy

For the simulation study used in this section and the next, random datasets were generated from the hierarchical model defined by equations (3)-(7), with $r = 1$, and with $z_j^\alpha = z_j^\beta = z_j^\tau = 1$, $j = 1, \ldots, p$. For each of p stocks, values of α, β, and v were assigned by random generation from the laws given by (4)-(6). Then, for each time period t, a value of the (scalar) factor index was generated from a $N(\mu_f, \Omega_f)$ distribution, and for each stock j a value of the return y_{jt} was generated by equation (3). This model corresponds to the standard single factor model, as in equation (1).

Four different settings, or blocks, for model parameters were used in the simulation. The settings were selected to mimic typical monthly stock returns and to demonstrate the performance of HB and LS estimates with different sample sizes and amounts of parameter heterogeneity. Block 1 mimics two years (24 months) of monthly data; the cross-sectional means and standard deviations for α, β, and τ in this block were set equal to values estimated from a random sample of 500 New York Stock Exchange companies during 1988-1991; $E[\alpha] = 0.00$, $SD[\alpha] = 0.70$, $E[\beta] = 1.00$, $SD[\beta] = 0.25$, $E[v] = 105$, $SD[v] = 175$. Blocks 2 and 3 change the sample size, to 72 months and 12 months. In Block 4, the sample size is 24 months, but the cross-sectional parameter heterogeneity is increased by a factor of 4, relative to Block 1: $SD[\alpha] = 2.80$, $SD[\beta] = 1.00$. For each simulation block, the number of firms generated equalled 30, the mean and standard deviation of the market index f_t were 1.0 and 4.0 respectively, and the cross-sectional correlations between α, β, and τ were equal to 0.0. 100 replications were simulated for each block. In the following, Blocks 1, 2, and 3 will be referred to as "NYSE(24)", "NYSE(72)", and "NYSE(12)", and Block 4 will be referred to as "HIGHHET(24)" (for "high heterogeneity").

Table 1 compares the mean absolute error (MAE) of the hierarchical Bayes (HB) and least squares (LS) estimates for the simulations. The table lists the mean and standard deviation of the MAE taken over all 100 simulation replications within a block. In the NYSE(24) block, the MAE's are substantially lower for the HB estimates: the differences are significant at $p < .001$ for all comparisons of HB and LS. using a non parametric sign test. The comparison for the alpha coefficient estimates is noteworthy over the 100 replications, the average HB error is less than 22% of the average LS error.

The results from the NYSE(72) block show that even for a large sample size, HB offers substantial advantages relative to LS: average HB error is less than 31% of the average LS error. Also, the results from the HIGHHET(24) block show that, even with cross-sectional parameter heterogeneity much higher than is seen in actual returns data, the HB estimator is considerably more efficient than the LS estimator.

Researchers analyzing stock market time series data face the problem of choosing an appropriate length of data for analysis. If the dataset is too long, then the parameters may not be constant over the entire period of observation, while if the dataset is too short, the data may be insufficient to accurately estimate model parameters. Here, it is shown that an HB estimation approach may help with this conflict. Comparing the estimation error of betas in the NYSE(72) and NYSE(12) blocks. it is seen that the average LS error with 72 time series observations, 0.21, is greater than the average HB error with only 12 time series observations, 0.19. Thus, HB may permit the efficient analysis of time series short enough that parameters may reasonably be assumed to be constant over the observation period.[2]

The hierarchical Bayes method makes some additional assumptions about the underlying data generating process which are not required with the least squares method. In this simulation experiment, the assumptions are not violated, but in practice, if the

[2]Alternatively, the HB framework can be used to explicitly model time-varying parameters, in the spirit of West and Harrison (1989).

assumptions are grossly incorrect, then the least squares estimator could out perform the shrinkage estimator in terms of estimation accuracy. Section 4 describes an experiment with real stock returns data, and it is seen that, in this one setting with non simulated data, the HB method again substantially outperforms the least squares estimator.

3.2 Portfolio Performance

Portfolio allocations w were chosen to maximize the expected utility, where the utility function was taken to be the negative exponential $U(y) = 1 - \exp(-\lambda y)$, with λ being a parameter expressing aversion to risk (e.g., Frost and Savarino 1986). In general, maximizing the expected utility will require numerical optimization, but in the case in which the distribution of future returns y_{jt} is well approximated by a normal distribution, with cross sectional mean and variance μ and Σ, the distribution of any portfolio $\mathbf{w}'\mathbf{y} = \sum_{j=1}^{p} w_j y_{jt}$ will also be normal, and the allocation vector which maximizes the expected utility will be the solution to the Markowitz (1952) optimization problem

$$\max_{\mathbf{w}} \mathbf{w}'\hat{\mu} - \frac{\lambda}{2}\mathbf{w}'\hat{\Sigma}\mathbf{w}, \quad \text{subject to} \quad \mathbf{w}'\mathbf{1} = 1, \tag{9}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ represent estimates of the joint moments, and $\mathbf{1}$ is a vector of 1's. The maximizer of (9) is given by $\mathbf{w}^* = \lambda^{-1}\hat{\Sigma}^{-1}\hat{\mu} + \frac{1-\lambda^{-1}\mathbf{1}'\hat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}\hat{\Sigma}^{-1}\mathbf{1}$. The least squares estimates of the moments μ and Σ can be determined from the respective estimates of parameters (α_j, β_j, v_j) and (μ_f, Ω_f), using the formulas (8). For Bayes methods, the appropriate variance matrix to use is the so-called *predictive* variance (Barry 1973); the predictive variance incorporates uncertainty about model parameters in addition to the usual sampling uncertainty about future observed returns. The method for obtaining the hierarchical Bayes predictive moments is described in Appendix A.3. Press (1982), and Quintana (1992) provide further discussion of the use of Bayesian techniques in portfolio selection.

The quality of an implied portfolio allocation vector $\mathbf{w}$ can be evaluated by the expected utility under the true model for the future returns. Let this true model be denoted by $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0)$. The true expected utility can be converted to a certainty equivalent (Elton and Gruber 1995). which is the return, in dollars, such that the investor will be indifferent between holding the risky portfolio $\mathbf{w}'\mathbf{y}$, and holding the certainty equivalent. In our case. the certainty equivalent C is found by solving $1 - \exp(-\lambda C) = \mathrm{E}[1 - \exp(-\lambda \mathbf{w}'\mathbf{y})]$. and the solution is $C = \mathbf{w}'\boldsymbol{\mu}_0 - \frac{\lambda}{2}\mathbf{w}'\Sigma_0\mathbf{w}$. The benchmark for the certainty equivalent is the value obtained when $\mathbf{w}$ is computed by solving problem (9) using the true moments $\boldsymbol{\mu}_0, \Sigma_0$.

The bottom rows of Table 1 display the quality of portfolios formed on the basis of the true mean and covariance, as well as the hierarchical Bayes and least squares estimates of these moments. For each simulation replication. the portfolios are obtained by solving the maximization problem (9). In all blocks, the performance of the HB portfolios, as measured by the mean certainty equivalent, is significantly better than that of the LS portfolios ($p < .001$). In the NYSE(24) block, the mean optimal certainty equivalent if the true model parameters were known was 1.29 (1.29% per month), while the the mean certainty equivalent for the portfolio based on HB estimates was 0.46. However, the mean certainty equivalent for the LS portfolios was 12.85. Thus, the deterioration in portfolio performance due to estimation error is much more severe for the LS than for the HB method. A similar pattern is observed in the other simulation blocks.

[Insert Table 1 Here]

4 Example: N.Y.S.E. Stock Returns

The previous section compared the performance of the hierarchical Bayesian method to the performance of a non–pooled estimator. ordinary least squares, using simulated data. This section reports the results of an experiment in which monthly returns data from the New York Stock Exchange were analyzed via hierarchical Bayesian methods, and an alternative shrinkage method, the multiple shrinkage estimator of Karolyi (1992). In this experiment. monthly returns data on 500 randomly selected securities in the Center for Research in Stock Prices (C.R.S.P.) database were obtained for 19 four year intervals: 1955 1959. 1957 1961, ..., 1991-1994. For each four year interval, the data from the first two years were used to obtain parameter estimates for a linear factor model, along with the implied optimal portfolio allocations, using hierarchical Bayes and multiple shrinkage techniques. The parameter estimates were then compared to the ordinary least squares estimates obtained over the third and fourth years of the interval, to evaluate out-of-sample parameter estimation accuracy of the two methods. Also, the estimated portfolio weights were applied to the returns in the third and fourth years to assess the quality of the portfolio selections.

The particular factor model used in the analyses was

$$y_{jt} = \alpha_j + \beta_{1j} m_t + \beta_{2j} \mathrm{SML}_t + \epsilon_{jt}, \quad \epsilon_{jt} \sim \mathrm{N}(0,\ v_j), \tag{10}$$

where m_t, the overall market factor, denotes the C.R.S.P. value weighted market index, and SML_t, the "size sensitivity" factor, represents the return on a portfolio composed of all the firms with smallest market value (lowest size decile in the C.R.S.P. dataset) minus the return on a portfolio composed of firms with large market value (highest size decile in the C.R.S.P. dataset). See Fama and French (1995) for further discussion of this factor model.

The multiple shrinkage method of Karolyi (1992) was originally introduced in order to obtain superior estimates of β coefficients in the single index model (1). The method produces an estimate $\hat{\beta}_j$ for stock j which is a weighted average of, say, $L+1$ different estimates: $\hat{\beta}_j = \frac{\sum_{l=0}^{L} \nu_l \hat{\beta}_{jl}}{\sum_{l=0}^{L} \nu_l}$. The first estimate, $\hat{\beta}_{j0}$, is just the ordinary least squares estimate. The other L estimates $\hat{\beta}_{jl}$, $l = 1, \ldots, L$, are cross sectional group means for the OLS coefficients for example, $\hat{\beta}_{j1}$ might be the average OLS β estimate for all companies in the same size class as firm j, and $\hat{\beta}_{j2}$ the average OLS β estimate for all companies in the same industry as firm j. The weights ν_l are set equal to the relative precision of each of the estimates $\hat{\beta}_{jl}$: the weight ν_0 on the OLS estimate is just the inverse of the sampling variance for the OLS estimator, and the weights ν_l, $l = 1, \ldots, L$ are the inverses of the cross sectional variances of the β estimates within the L classes.

Chan and Chen (1988) suggests using firm size as an instrument for predicting market beta. For the data analysis in this section, the classes used for the multiple shrinkage estimator were the firm size (market value) decile, and the industry segment, where the different industry classifications used were "manufacturing" (SIC code $\in [2000, 3999]$), "utilities" (SIC $\in [4900, 4999]$), "finance/insurance" (SIC $\in [6000, 6999]$), "services" (SIC $\in [7000, 8999]$), and "other": the industries represented in the last class included agriculture, mining, construction, and retail trade. The multiple shrinkage method of Karolyi was applied, using these classes, for obtaining pooled estimates for the β coefficients, the α coefficients, and the idiosyncratic variances v, in model (10).

For the hierarchical Bayes method, the same information set - the market values and the industry classifications for the firms was used to construct the predictor covariates for the hierarchical prior distributions for α_j, β_{jk}, and v_j, as was used to form classes for the multiple shrinkage method. In this application, $\mathbf{z}^{\alpha}$, $\mathbf{z}^{\beta}$, and $\mathbf{z}^{\tau}$ were specified as $\mathbf{z}_j^{\alpha} = \mathbf{z}_j^{\beta} = \mathbf{z}_j^{\tau} = \mathbf{z}_j$, where $\mathbf{z}_j = (1, z_{j1}, z_{j2}, z_{j3}, z_{j4}, z_{j5})$, with z_{j1} through z_{j4} defined as indicator variables for manufacturing, utilities, finance/insurance, and services firms, respectively, and z_{j5} defined as the size decile for firm j. The Markov chain used in the hierarchical Bayes method was run for 2000 iterations, with the first 1000 samples discarded.

For each estimation method, portfolio weights were selected as the solution to optimization problem (9), with risk aversion parameter λ fixed at 10.0. The MS estimates of the moments $\boldsymbol{\mu}$ and Σ were determined from the estimates of parameters $(\alpha_j, \boldsymbol{\beta}_j, v_j)$ and $(\boldsymbol{\mu}_f, \Omega_f)$, using equation (8). The calculation of the HB estimates of $\boldsymbol{\mu}$ and Σ is described in Appendix A.3.

4.1 Parameter Estimates

Figure 1 plots the posterior means for the coefficients of $\mathbf{z}$ in the prior equations (4) (6), for the 19 two-year estimation periods between 1955-1956 and 1991–1992. The signs for several of the coefficients are consistent over time. For example, the coefficient for z_2, the indicator for utilities, is consistently negative, both for predicting idiosyncratic variance (Figure 2(b)), and market beta (Figure 2(c)). The small market beta for utilities suggests that these firms have relatively little undiversifiable risk; the sensitivity analysis to be presented in section 5, however, demonstrates that the low idiosyncratic variance for utilities may render the firms highly sensitive to estimation risk. As expected, z_5, firm market value, is highly related to the size sensitivity measure β_2 (Figure 2(d)); it is also the case that larger firms have consistently lower market beta β_1, and, as is suggested in Malkiel and Xu (1997), lower idiosyncratic risk v.

4.2 Prediction Accuracy and Portfolio Performance

Table 2 describes the accuracy with which the hierarchical Bayes (HB) and multiple shrinkage (MS) estimates from the leading two-year periods predict least squares estimates from the trailing, two year hold out periods. The table lists, for each of the methods, the mean and standard deviation, over the 19 analyses, of the mean absolute error (MAE), across all 500 firms, in predicting future least squares parameter values. The row in the table labeled "# HB" refers to the number of periods in which the HB method outperformed the alternative estimator: the p-value reported is from the associated binomial sign test. For example, the MAE for the HB estimator for the β coefficients was smaller than that for the MS estimator in 18 of the 19 analyses; this difference is statistically significant ($p < .001$). For the other model parameters, α and v, the HB estimator was more accurate as well.

Table 2 also displays the out of sample performance of the estimated portfolios, in terms of certainty equivalents, the risk adjusted measure of portfolio performance. It is seen that the HB method significantly outperforms the MS method in terms of portfolio performance: in 14 out of 19 periods, the HB portfolio had a higher risk adjusted return than the MS portfolio.

The HB and the MS methods both employ shrinkage, and both make use of fundamental information on the firm to estimate model parameters. One possible reason that the HB method could outperform the MS method in some settings is that the MS formula does not correspond exactly to a Bayes estimate, in that the formula takes into account the precisions of the different estimates used to form the weighted average, but does not take into account the correlations between the estimates. While in the present setting the HB method out performed the MS method, one can expect that in many cases the two methods could have similar performance.

[Insert Table 2 Here]

5 Sensitivity Analysis

The previous sections demonstrated the considerable extent to which estimation error degrades portfolio performance. In this section, the relationship between mis estimated market model parameters and mis-estimated portfolio weights is explored in greater detail.

Sensitivity analysis for portfolio selection in the general mean-variance framework has been discussed in Best and Grauer (1991) and Chopra and Ziemba (1993). In these studies, it was seen that portfolio selection is sensitive to errors in parameter estimates, and particularly to errors in estimates of the asset means. In this section, the sensitivity analysis is specialized to the case of the factor model for stock returns, in which setting the intercept parameters α_j are critical in determining asset means. It is shown that, for data with similar characteristics to N.Y.S.E. stock returns: (a) the error in estimating w_j^*, the optimal weight for stock j, is due largely to the error in estimating α_j, and less to estimating α_k, $k \neq j$; and (b) the sampling variance of the estimated weight for stock j, $\hat{w}_j^*$, is approximately *inversely* proportional to $\hat{v}_j$. The findings suggest that the value $1/\hat{v}_j$ can serve as a useful diagnostic in portfolio optimization: a security with a high value of this quantity is "suspect", in the sense that the security's estimated portfolio weight is likely to be far from the true optimal weight.

The sensitivity analysis derived in this section will be conducted with respect to the least squares estimator, though is appropriate asymptotically for the hierarchical Bayes estimator as well. Attention will be directed to the common special case in which there is just a single independent factor in model (3):

$$y_{jt} = \alpha_j + \beta_j f_t + \epsilon_{jt}, \quad \epsilon_{jt} \sim \mathrm{N}(0, v_j);$$

this restriction simplifies the necessary notation. In this setting, if the factor f_t in equation (5) has been scaled to have zero mean and variance v_M, and if the ϵ_{jt} are independent of ϵ_{kt} for $k \neq j$, then the sampling distributions of the model parameters $\{\alpha_j, \beta_j, v_j, j = 1, \ldots, p\}$ are all independent of each other, with sampling variances given by $\text{Var}[\hat{\alpha}_j] = v_j/n$, $\text{Var}[\hat{\beta}_j] = v_j/(nv_M)$, $\text{Var}[\hat{v}_j] = 2v_j^2/n$. It follows from the independence of the parameter estimates that the first order approximation to the sampling variances for the optimal portfolio weights $w_j^*, j = 1, \ldots, p$, is given by

$$\text{Var}[\hat{w}_j^*] \approx \sum_{k=1}^{p}(\frac{\partial w_j}{\partial \alpha_k})^2 \cdot \text{Var}[\hat{\alpha}_k] + \sum_{k=1}^{p}(\frac{\partial w_j}{\partial \hat{\beta}_k})^2 \cdot \text{Var}[\hat{\beta}_k] + \sum_{k=1}^{p}(\frac{\partial w_j}{\partial \hat{v}_k})^2 \cdot \text{Var}[\hat{v}_k]. \quad (11)$$

Because of the special importance of estimating the asset means in portfolio analysis (Chopra and Ziemba (1993) find that "errors in means are about eleven times as important as errors in variances") attention will be focused on estimation of the α coefficients; the variance of $\hat{w}_j^*$ will thus be approximated by $\text{Var}[\hat{w}_j^*] \approx \sum_{k=1}^{p}(\frac{\partial w_j}{\partial \alpha_k})^2 \cdot \text{Var}[\hat{\alpha}_k]$. It is shown in Appendix B that $\frac{\partial w_j}{\partial \alpha_j}$ will tend to be more significant in magnitude than $\frac{\partial w_j}{\partial \alpha_k}$, $j \neq k$; further, the analysis in Appendix B and the computations below show that $\frac{\partial w_j}{\partial \alpha_j}$ can be very accurately approximated by $1/\lambda\hat{v}_j$. Altogether, this leads to the simple approximation: $\text{Var}[\hat{w}_j^*] \approx (\frac{\partial w_j}{\partial \alpha_j})^2 \cdot \text{Var}[\hat{\alpha}_j] \approx (\lambda\hat{v}_j)^{-2} \cdot v_j/n \approx 1/(\lambda^2 n\hat{v}_j) \propto 1/\hat{v}_j$; i.e., the uncertainty about w_j is inversely related to the estimated idiosyncratic variance $\hat{v}_j$.

The simulation described in Section 3 helps to clarify the relationship between parameter estimation error and portfolio weight estimation error. Table 3 describes the correlations between the squared errors in estimating w_j^*, and: (a) the squared errors in estimating α_j (ERR(alpha)); (b) the squared errors in estimating β_j (ERR(beta)); (c) the squared errors in estimating v_j (ERR(var)); (d) the squared errors in estimating α_j, multiplied by the sensitivity $1/\hat{v}_j^2$ (ERR-SENS(alpha)); and (e) the estimated variance of w_j^*, $1/\hat{v}_j$ (EST-VAR(w)). The correlations are computed across the 30 securities for each simulation replication; Table 3 lists the average and standard deviation of these correlations, over the 100 simulation replications. The rows corresponding to ERR(alpha), ERR(beta), and ERR(var) confirm the central importance of estimating the α parameters in determining portfolio weights: the average correlation between mis estimation of α and mis estimation of w^* is generally very high. The rows corresponding to ERR-SENS(alpha) demonstrate the additional importance of the sensitivity factors $1/\hat{v}_j^2$, as the mean correlations reported are even greater than those for ERR(alpha). In practice, ERR(alpha) cannot be observed, since its calculation depends on the true, but unknown, value of α_j. However, Table 3 shows that the observable quantity $1/\hat{v}_j$ (EST-VAR(w)) tends to be positively correlated with the error in estimating w_j^*: the measure $1/\hat{v}_j$ can thus serve as a diagnostic in assessing possible errors in portfolio allocation. In each of the blocks, the correlation of $1/\hat{v}_j$ with the exact value of $\frac{\partial w_j^*}{\partial \alpha_j}$, presented in Appendix B, is over .95.

Classical portfolio theory (e.g., Elton and Gruber 1995) has held that the idiosyncratic variance of a security, v_j, is economically inconsequential, since idiosyncratic risks can, in principle, be diversified away. Malkiel and Xu (1997), though, show that idiosyncratic variance appears to be cross sectionally correlated with ex-post expected returns. This section has demonstrated a second respect in which idiosyncratic variance may be significant: the uncertainty of an estimate of optimal portfolio weight w_j^* can in practice be quantified by the simple expression $\hat{v}_j^{-1}$.

[Insert Table 3 Here]

6 Conclusion

This paper advocates the use of hierarchical Bayes methods for estimating market model parameters and for selecting portfolios. These methods automatically and optimally use cross-sectional data to improve upon parameter estimates for each individual firm.

The hierarchical Bayes procedure incorporates parameter uncertainty into the estimate of predictive variance. thus allowing for rational management of estimation risk.

The improvement in estimation accuracy leads to improved portfolio performance. This paper shows through a sensitivity analysis that the optimal portfolio weights are most strongly affected by estimation error when the idiosyncratic variances are small.

A Markov Chain Monte Carlo Estimation

A.1 Gibbs Sampling

Exact finite sample inferences on the parameters of the hierarchical regression model can be obtained using Gibbs sampling (Gelfand and Smith 1990; Müeller 1991). Gibbs sampling is a particular variant of the class of procedures known as "Markov chain Monte Carlo methods" (Roberts and Smith 1993), in which parameter vectors are randomly generated from a Markov chain whose stationary distribution is equal to the joint posterior distribution of the model parameters. A Gibbs sampler involves iterative resampling from the full conditional posterior distributions of all of the model parameters (Gelfand and Smith 1990). For the hierarchical market model of Section 2. all of the associated parameters can be sampled directly, except for the variances v_j, and for these parameters a Metropolis Hastings step can be imbedded (Tierney 1994). The derivations of the conditional distributions follow from standard results in Bayesian analysis (Zellner 1971). Details are provided in the following sections.

A.2 The Full Conditional Posterior Distributions

A.2.1 Generating μ_f and Ω_f

Given the observed factor returns f_t, $t = 1, \ldots, n$, the conditional posterior distribution of μ_f is $N(\sum_{t=1}^{n} f_t/n, \ \Omega_f/n)$, and the conditional posterior for ${\Omega_f}^{-1}$ is $W(\frac{1}{2}\sum_{t=1}^{n}(f_t - \mu_f)(f_t - \mu_f)', \ \frac{n}{2})$, where $V \sim W(T, \ d)$ denotes that V has the Wishart probability density proportional to $|T|^{d/2}|V|^{(d-r-1)/2}\exp(-\frac{1}{2}\operatorname{Tr}(TV))$, and expected value dT^{-1}.

A.2.2 Generating α_j

Let $y^*_{jt} = y_{jt} - \beta'_j f_t$. Then, conditional on β_j, and f_t, the quantity y^*_{jt} is normally distributed with mean α_j and variance v_j. Since the prior distribution for α_j is $N(\theta'_0 z^{\alpha}_j, \ \Lambda_0)$, the conditional posterior for α_j is $N((\Lambda_0^{-1}\theta'_0 z^{\alpha}_j + nv_j^{-1}\bar{y}^*_j)(\Lambda_0^{-1} + nv_j^{-1})^{-1}, \ (\Lambda_0^{-1} + nv_j^{-1})^{-1})$, where $\bar{y}^*_j = n^{-1}\sum_{t=1}^{n} y^*_{jt}$.

A.2.3 Generating β_j

Now let $X = ((f_{tk}))^{1\le k\le r}_{1\le t\le n}$, and $y^*_j = (y_{jt} - \alpha)_{1\le t\le n}$ denote the regression data for the market model for stock j. From (5), the prior mean for β_j is $\theta' z^{\beta}_j$, and the prior variance is $\Lambda = \operatorname{diag}(\Lambda_k)$. The full conditional posterior distribution for β_j is then $N(\eta'_j, \ \Lambda'_j)$, where the posterior parameters are $\eta'_j = (\Lambda^{-1} + v_j^{-1}X'X)^{-1}(\Lambda^{-1}\theta' z^{\beta}_j + v_j^{-1}X'y^*_j)$, $\Lambda'_j = (\Lambda^{-1} + v_j^{-1}(X'X))^{-1}$.

A.2.4 Generating v_j and τ_j

The conditional posterior of v_j is not of a standard form. However, $\tau_j = \log(v_j)$ can be generated from its correct distribution through use of an imbedded Metropolis chain (Hastings 1970: Tierney 1994). Let $\pi(\tau_j)$ denote the exact conditional posterior density for τ_j. and let $\tau_j^{(g)}$ be the value of τ_j after g iterations of the Markov chain algorithm. In a Metropolis sampler. a new candidate value of τ_j, τ_j^* is generated from some density $f(\tau)$ which approximates the desired conditional posterior $\pi(\tau)$. With probability $\max\{1, \frac{f(\tau_j^*)/\pi(\tau_j^*)}{f(\tau_j^{(g)})/\pi(\tau_j^{(g)})}\}$. this value is accepted as the $g+1$st sample of τ_j: otherwise. the old value is retained, and the Markov chain proceeds. The τ_j so generated will. asymptotically. have the correct distribution $\pi(\tau_j)$.

By Bayes theorem, the exact conditional posterior distribution $\pi(\tau_j)$ is proportional to the product of the likelihood and the prior for τ_j; the likelihood of the data is proportional to $\exp(-\frac{n}{2}\tau_j - \frac{S_j}{2}e^{-\tau_j})$, where $S_j = \sum_{t=1}^{n}(y_{jt} - \alpha_j - \beta_j'\mathbf{f}_t)^2$, and the prior for τ_j is proportional to $\exp(-\frac{1}{2\delta}(\tau_j - \psi'\mathbf{z}_j^\tau)^2)$. The likelihood, as a function of τ_j, can be approximated closely by a normal distribution by matching the modes, and the second derivatives at the modes. The resulting normal approximation has mean $\log(S_j/n)$ and variance $2/n$. Given a normal prior and approximately normal likelihood, the conditional posterior of τ_j is also approximately normal, with mean $(\frac{n}{2}\log(S_j/n)+\delta^{-1}\psi'\mathbf{z}_j^\tau)/(\frac{n}{2}+\delta^{-1})$, and variance $(\frac{n}{2}+\delta^{-1})^{-1}$; this normal distribution forms a suitable generating density f for the Metropolis chain. In simulation studies, the normal approximation is seen to be very accurate. with the acceptance probability usually close to 1.0, and seldom less than 0.9. Given a sample of τ_j, the corresponding value of v_j is given by $v_j = \exp(\tau_j)$.

A.2.5 Generating Θ, Λ

Let $\boldsymbol{\alpha}$ denote the α coefficients for all securities, arrayed as a vector, and let $\mathbf{B}_k$ similarly denote the β coefficients associated with the kth factor. Let $\mathbf{Z}^\alpha$ be the set of fundamental variables related to α, and $\mathbf{Z}^\beta$ the corresponding set of variables related to $\boldsymbol{\beta}$. Then the full conditional distributions for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_k$. $k = 1, \ldots, r$ are $N((\mathbf{Z}^{\alpha\prime}\mathbf{Z}^\alpha)^{-1}\mathbf{Z}^{\alpha\prime}\boldsymbol{\alpha},\ \Lambda_0(\mathbf{Z}^{\alpha\prime}\mathbf{Z}^\alpha)^{-1})$ and $N((\mathbf{Z}^{\beta\prime}\mathbf{Z}^\beta)^{-1}\mathbf{Z}^{\beta\prime}\mathbf{B}_k,\ \Lambda_k(\mathbf{Z}^{\beta\prime}\mathbf{Z}^\beta)^{-1})$. respectively. The full conditional distributions for

Λ_0 and Λ_k, $k = 1, \ldots, r$, are $\text{IG}(a_0 + p/2,\ a_1 + \sum_{j=1}^{p} (\alpha_j - \theta_0' \mathbf{z}_j^{\alpha})^2/2)$ and $\text{IG}(a_0 + p/2,\ a_1 + \sum_{j=1}^{p} (\beta_{jk} - \theta_k' \mathbf{z}_j^{\beta})^2/2)$.

A.2.6 Generating ψ, δ

Let τ denote the log variances τ for all securities, arrayed as a vector, and let $\mathbf{Z}^{\tau}$ denote the set of fundamental variables related to τ. The conditional posterior for ψ is $\text{N}((\mathbf{Z}^{\tau\prime}\mathbf{Z}^{\tau})^{-1}\mathbf{Z}^{\tau\prime}\tau,\ \delta(\mathbf{Z}^{\tau\prime}\mathbf{Z}^{\tau})^{-1})$, and the conditional posterior for δ is $\text{IG}(a_0 + p/2,\ a_1 + \sum_{j=1}^{p} (\tau_j - \psi' \mathbf{z}_j^{\tau})^2/2)$.

A.3 Predictive Moments

Let $\tilde{\mathbf{Y}}$ represent the future returns for the vector of assets under analysis, and let $\mathbf{S}$ denote the set of all model parameters. Then the predictive moments are defined as $\hat{\mu} = \text{E}[\tilde{\mathbf{Y}}|\mathcal{D}] = \text{E}[\text{E}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]]$, $\hat{\Sigma} = \text{Var}[\tilde{\mathbf{Y}}|\mathcal{D}] = \text{E}[\text{Var}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]] + \text{Var}[\text{E}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]]$. The inner conditional expectations and variances $\text{E}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]$ and $\text{Var}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]$ are determined by equation (8); thus, one can generate realizations of these conditional moments at each step of the Markov chain. The estimate of the predictive mean $\text{E}[\tilde{\mathbf{Y}}|\mathcal{D}]$ will then be the sample average of the generated values of $\text{E}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]$, and the estimate of the predictive variance will be the sample average of the generated values of $\text{Var}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]$ plus the sample variance of the generated values of $\text{E}[\tilde{\mathbf{Y}}|\mathbf{S}, \mathcal{D}]$.

A.4 Initial Conditions

The algorithm requires initial values for starting the Markov chain. Initial values for the parameters α_j, β_j, v_j can be obtained by using ordinary least squares regression – i.e., the usual estimators for these parameters. Initial estimates for Θ, Λ, ψ and δ can be obtained by multivariate regression of these estimates of α_j's β_j's and τ_j's versus the z_j^{α}'s. z_j^{β}'s and z_j^{τ}'s.

A.5 Markov Chain Monte Carlo Algorithm: A Summary

The Gibbs sampling algorithm for generating samples from the posterior distribution of the model parameters can be summarized as follows:

1. Obtain preliminary estimates of the parameters α_j, β_j, v_j, via ordinary LS regression.

2. Based on the initial estimates of α_j, β_j, v_j, obtain preliminary estimates of the parameters Θ, Λ, ψ, and δ, via LS regression.

3. Repeat for G Gibbs iterations:

   (a) Generate samples of factor moments μ_f and Ω_f.

   (b) For each security $j = 1, \ldots, p$, generate samples of α_j, β_j, τ_j and v_j from their respective posterior conditional densities.

   (c) Given the new samples of α_j, β_j, and τ_j generate samples of hyperparameters Θ, Λ, ψ and δ.

   (d) Given the samples of μ_f, Ω_f, and α_j, β_j, v_j, $j = 1, \ldots, p$, compute values for $\mathrm{E}[\mathrm{Y}|\mathrm{s}, \mathcal{D}]$ and $\mathrm{Var}[\mathrm{Y}|\mathrm{s}, \mathcal{D}]$, using equation (8).

The estimate of a particular parameter, say β_j, is obtained by its posterior mean: $\mathrm{E}[\beta_j|\mathcal{D}] = \frac{1}{G-B-1} \sum_{g=B+1}^{G} \beta_j^{(g)}$, where B is the number of initial samples discarded, and $\beta_j^{(g)}$ is the value of β_j generated in step 3b above during the gth iteration of the Markov chain. The predictive moments $\mathrm{E}[\mathrm{Y}|\mathcal{D}]$ and $\mathrm{Var}[\mathrm{Y}|\mathcal{D}]$, are determined by the formulas: $E[Y|\mathcal{D}] = \frac{1}{G-B-1}\sum_{g=B+1}^{G} E[Y|s,\mathcal{D}]^{(g)}$. $Var[Y|\mathcal{D}] = \frac{1}{G-B-1}\sum_{g=B+1}^{G} Var[Y|s,\mathcal{D}]^{(g)} + \frac{1}{G-B-1}\sum_{g=B+1}^{G} (E[Y|s,\mathcal{D}]^{(g)} - E[Y|\mathcal{D}])^2$, where $E[Y|s,\mathcal{D}]^{(g)}$ and $Var[Y|s,\mathcal{D}]^{(g)}$ are the values of the conditional moments generated in step 3d during the gth iteration of the Markov chain.

B Sensitivities of Portfolio Weights

Section 5 derived an approximation to the variance of the estimated portfolio weights, in terms of the derivatives of these weights with respect to model parameters (α, β, v). In this section, the relevant derivatives are derived.

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, $\boldsymbol{v} = (v_1, \ldots, v_p)'$, and $\mathbf{V} = \text{diag}(\boldsymbol{v})$. The vector of mean returns of the securities is: $\boldsymbol{\mu} = \boldsymbol{\alpha} + \mu_M \boldsymbol{\beta}$, and the covariance matrix is: $\Sigma = \mathbf{V} + v_M \boldsymbol{\beta}\boldsymbol{\beta}'$. The optimal portfolio weights from the previous section are:

$$w^* = \lambda^{-1}(\Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}})\boldsymbol{\mu} + \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}, \tag{12}$$

where

$$\Sigma^{-1} = \mathbf{V}^{-1} - (\frac{v_M}{1 + v_M \boldsymbol{\beta}'\mathbf{V}^{-1}\boldsymbol{\beta}})\mathbf{V}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}^{-1}. \tag{13}$$

The sensitivity of the optimal weights with respect to $\boldsymbol{\alpha}$ is:

$$\frac{\partial w^*}{\partial \boldsymbol{\alpha}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}}\frac{\partial w^*}{\partial \boldsymbol{\mu}} = \lambda^{-1}(\Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}), \tag{14}$$

where $\frac{\partial w^*}{\partial \boldsymbol{\alpha}}$ is a $p \times p$ matrix with (i,j)th element equal to $\frac{\partial w_i^*}{\partial \alpha_j}$

It can be shown that:

$$\frac{\partial w_i^*}{\partial \alpha_i} = (\lambda v_i)^{-1}\left[1 - \left(\frac{1}{v_i \mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)\left(\frac{1 + v_M u[S_\beta^2 + (\bar{\beta} - \beta_i)^2]}{1 + v_M \sum_{k=1}^{p} \beta_k^2 / v_k}\right)\right] \tag{15}$$

$$\frac{\partial w_i^*}{\partial \alpha_j} = -\left(\lambda v_i v_j \mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1}\left(\frac{1 + v_M u[S_\beta^2 + (\bar{\beta} - \beta_i)(\bar{\beta} - \beta_j)]}{1 + v_M \sum_{k=1}^{p} \beta_k^2 / v_k}\right) \text{ for } i \neq j \tag{16}$$

where $u = \sum_{k=1}^{p} v_k^{-1}$; $\bar{\beta} = u^{-1}\sum_{k=1}^{p} \beta_k / v_k$, and $S_\beta^2 = u^{-1}\sum_{k=1}^{p}(\beta_k - \bar{\beta})^2 / v_k$. The analysis of these results is facilitated by recognizing that the matrix Σ^{-1} is equivalent to a conditional covariance matrix. Let $\mathbf{y}$ be a multivariate normal random variable with mean $\mathbf{0}$ and covariance matrix $\lambda^{-1}\Sigma^{-1}$. Then the conditional variance of $\mathbf{y}$ given its sum, $\mathbf{1}'\mathbf{y}$, is the same as $\frac{\partial \boldsymbol{w}^*}{\partial \boldsymbol{\alpha}}$. Consequently, $\frac{\partial w_i^*}{\partial \alpha_i}$, as given in Equation (15), must be positive, as it is the conditional variance of Y_i given $\mathbf{1}'\mathbf{y}$. Although $\frac{\partial w_i^*}{\partial \alpha_j}$ in Equation (16) can be positive if either β_i or β_j is much greater than $\bar{\beta}$ while the other is much less than $\bar{\beta}$, for the typical values of the market risk it is usually negative. Moreover, $\frac{\partial \boldsymbol{w}^{*\prime}}{\partial \boldsymbol{\alpha}}\mathbf{1} = \mathbf{0}$, so that $\frac{\partial w_i^*}{\partial \alpha_i} = -\sum_{j=1;j\neq i}^{p} \frac{\partial w_i^*}{\partial \alpha_j}$. Thus, when Equation (16) is negative for all $j \neq i$, w_i^* tends to be more sensitive to estimation error in α_i than to estimation error in α_j for $j \neq i$.

Matrix expressions for the partial derivatives of the optimal weights with respect to $\boldsymbol{\beta}$ and $\boldsymbol{v}$ can be obtained as well, but they do not have simple interpretations. Formulas for these derivatives can be obtained from the authors.

In practice, the exact values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{v}$ in equations (12)-(16) will be unknown, and so will be replaced by their respective parameter estimates.

References

Barry, C. B. (1973). Portfolio analysis under uncertain means, variances, and covariances. *Journal of Finance 29*, 515-522.

Berger. J. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York, NY: Springer Verlag.

Best, M. J. and R. R. Grauer (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means. *Review of Financial Studies 4*, 315-342.

Board, J. L. G. and C. M. S. Sutcliffe (1994). Estimation methods in portfolio selection and the effectiveness of short sales restrictions: U. K. evidence. *Management Science 40*, 516–534.

Chan, K. C. and N.-F. Chen (1988). An unconditional asset-pricing test, and the role of firm size as an instrumental variable for risk. *Journal of Finance 43*, 309-325.

Chopra, V. K. and W. T. Ziemba (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management Winter*, 559-582.

Elton, E. J. and M. J. Gruber (1995). *Modern Portfolio Theory and Investment Management* (5th ed.). New York, NY: John Wiley and Sons.

Fama. E. and K. R. French (1992). The cross section of expected stock returns. *Journal of Finance 47*, 427-465.

Fama, E. and K. R. French (1995). Explaining the cross section of expected stock returns. *Journal of Finance 50*, 427-465.

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 71*, 607-636.

Frost, P. A. and J. E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis 21*, 293 305.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

George, E. I. (1986). A formal Bayes multiple shrinkage estimator. *Communications in Statistics, A 15*, 2099–2114.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97 109.

Jorion, P. (1986). Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis 21*, 279 292.

Karolyi, G. A. (1992). Predicting risk: Some new generalizations. *Management Science 38*, 57–74.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B 34*, 1-41.

Malkiel, B. G. and Y. Xu (1997). Risk and return revisited. *Journal of Portfolio Management 23*, 9-14.

Markowitz, H. M. (1952). Portfolio Selection. *Journal of Finance 7*, 77–91.

Müeller, P. (1991). A generic approach to posterior integration and Gibbs sampling. Technical Report 91-09, Department of Statistics, Purdue University.

Müeller, P. and G. Rosner (1994). A semiparametric Bayesian population model with hierarchical mixture priors. Technical Report 94-17, Duke University Institute of Statistics & Decision Sciences.

Press, S. J. (1982). *Applied Multivariate Analysis*. Melbourne, FL: Krieger Publishing Company, Inc.

Quintana, J. M. (1992). Optimal portfolios of forward currency contracts. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, pp. 753–762. New York: Oxford University Press.

Roberts, G. O. and A. F. M. Smith (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B 55*, 3–23.

Rosenberg, B. and G. James (1976). Prediction of beta from investment fundamentals. *Financial Analysts Journal 32*, 60–72.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics 4*, 1701–1762.

Vasicek, O. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance 28*, 1233–1239.

Wakefield, J. C., A. F. M. Smith, A. Racine-Poon, and A. E. Gelfand (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics 43*, 201–221.

West, M. and J. Harrison (1989). *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

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

Table 1: Estimation accuracy (mean absolute error. MAE) and portfolio performance (certainty equivalent) of Hierarchical Bayes (HB) and Least Squares (LS) estimators on simulated data. 100 simulation replications per block. MEAN and SD denote mean and standard deviation over the 100 replications.

		NYSE(24)		NYSE(72)		NYSE(12)		HIGHHET(24)	
		MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
Alpha	HB	0.33	0.09	0.27	0.04	0.42	0.20	0.90	0.16
MAE	LS	1.54	0.36	0.88	0.18	2.17	0.47	1.56	0.37
Beta	HB	0.17	0.03	0.13	0.02	0.19	0.03	0.30	0.07
MAE	LS	0.38	0.10	0.21	0.04	0.55	0.17	0.38	0.11
Variance	HB	22.45	7.46	14.49	6.24	32.31	12.50	23.86	8.40
MAE	LS	24.25	8.29	14.75	6.32	35.23	13.06	25.35	9.08
Certainty	TRUE	1.29	0.35	1.36	0.36	1.38	0.41	12.45	5.70
Equivalent	HB	0.46	0.34	0.78	0.36	0.38	0.53	5.58	5.11
	LS	-12.85	4.70	-2.04	1.18	-46.98	24.02	-10.32	17.63

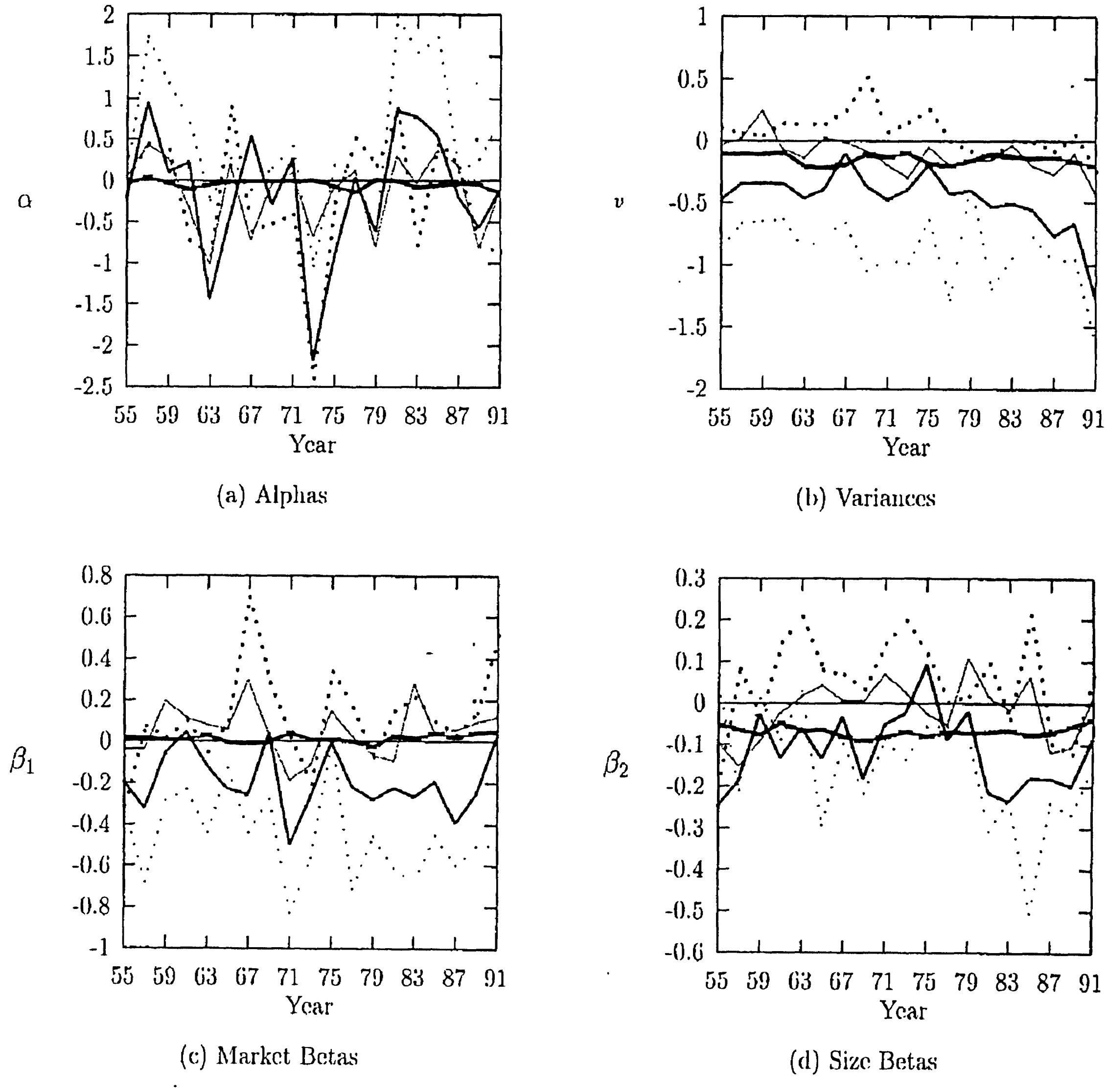
Table 2: Estimation accuracy (mean absolute error) and portfolio performance (certainty equivalent) of Hierarchical Bayes (HB) and Multiple Shrinkage (MS) estimators in N.Y.S.E. cross validation study. 1955 1994. # HB denotes the number of periods. out of 19, in which the HB estimator outperformed the MS estimator: P value is the associated significance level (binomial sign test).

		MEAN	SD	# HB	P value
Alpha	HB	1.57	0.30	18	$< .0001$
	MS	1.64	0.33		
Beta	HB	0.41	0.07	18	$< .0001$
	MS	0.43	0.07		
Variance	HB	2.33	0.54	11	0.18
	MS	2.34	0.53		
Certainty	HB	-49.58	30.56	14	0.01
Equivalent	MS	-78.56	51.76		

Table 3: Correlations between squared error of estimated portfolio weights, and the potential sources of this error. for simulated data. MEAN and SD denote mean and standard deviation of the correlations over the 100 simulation replications.

		CORRELATIONS			
		NYSE(24)	NYSE(72)	NYSE(12)	HIGHHET(24)
ERR(alpha)	MEAN	0.656	0.754	0.480	-0.018
	SD	0.161	0.128	0.216	0.130
ERR(beta)	MEAN	-0.035	0.011	-0.007	-0.130
	SD	0.154	0.183	0.180	0.093
ERR(var)	MEAN	0.112	0.008	0.171	-0.131
	SD	0.199	0.203	0.223	0.071
ERR-SENS(alpha)	MEAN	0.892	0.922	0.856	0.596
	SD	0.095	0.075	0.167	0.279
EST-VAR(w)	MEAN	0.377	0.247	0.512	0.514
	SD	0.199	0.180	0.217	0.241

Figure 1: Posterior means of hierarchical model coefficients.
Manufacturing: ___ Utility: .... Finance: ___ Service: .... Firm Size: ___



(a) Alphas

(b) Variances

(c) Market Betas

(d) Size Betas