Federal Reserve Bank of Minneapolis
Research Department

Forthcoming, *Journal of Econometrics*

Bayesian Reduced Rank Regression in Econometrics

John Geweke*

Working Paper 540

Revised January 1995

ABSTRACT

The reduced rank regression model arises repeatedly in theoretical and applied econometrics. To date the only general treatments of this model have been frequentist. This paper develops general methods for Bayesian inference with noninformative reference priors in this model, based on a Markov chain sampling algorithm, and procedures for obtaining predictive odds ratios for regression models with different ranks. These methods are used to obtain evidence on the number of factors in a capital asset pricing model.

*Geweke, Federal Reserve Bank of Minneapolis and University of Minnesota. The original version of this paper was prepared for the First Riverboat Conference on Bayesian Econometrics and Statistics, Basel-Amsterdam, April 29–May 4, 1992. It has been substantially improved by comments from Luc Bauwens, John Burkett, Guofu Zhou, and two anonymous referees, but the author is responsible for any errors. This work was supported in part by National Science Foundation Grant SES–9210070. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

This paper introduces procedures for Bayesian inference in the reduced rank regression model,

$$Y = X\Theta + ZA + E. \tag{1}$$

This is a multivariate regression, organized with observations by row, and equations by column: Y is the matrix of dependent variables, X and Z are matrices of explanatory variables, and E is the matrix of disturbances. The matrix of coefficients, Θ, is assumed to have rank q, where q is an integer less than the number of columns in X or Y. The matrix of coefficients, A, is not assumed to have reduced rank. Development of asymptotic sampling theory for the reduced rank regression model began with Anderson (1951) and has continued with Joreskog and Goldberger (1975), Izenman (1975), Davies and Tso (1982), and Velu (1991). Only Zhou (1993) appears to have addressed finite sample questions from a frequentist perspective.

Bayesian inference in the model (1) *per se* appears not to have been taken up in the literature, although two special cases have received attention. The first is the incomplete simultaneous equation model (Dreze 1976; Dreze and Richard 1983; Zellner, Min, and Dallaire 1993). The second is the cointegration model for vector autoregressions (Bauwens and Lubrano 1993, Kleibergen and van Dijk 1993). Both cases are detailed subsequently. The contribution of this paper is to provide a Bayesian treatment of the general model (1) using recently developed computational procedures, and to suggest a method for inferring q.

The paper begins by deriving posterior distributions and conditional posterior distributions for two alternative normalizations and a family of reference priors. Section 3 shows how the conditional posterior distributions may be used in a Gibbs sampling algorithm for the computation of posterior moments, and discusses questions of convergence and identification. Development to

this point assumes that q = rank (Θ) is known. This assumption is removed in Section 4, which shows how to compute predictive odds ratios for alternative values of q. In Section 5 an example with artificial data illustrates both the convergence properties of the computational algorithm and the ability of the predictive odds ratio to identify a true model. The paper concludes with an application of these procedures to inference for the number of factors in an asset pricing model.

2. Prior and Posterior Distributions

In standard notation the multivariate regression model is

$$\underset{n\times L}{Y} = \underset{n\times p}{X}\ \underset{p\times L}{\Theta} + \underset{n\times k}{Z}\ \underset{k\times L}{A} + \underset{n\times L}{E} \tag{2}$$

where $Y = [y_{ij}]$ is a matrix of dependent variables, with typical entry y_{ij} providing observation i on variable j; $X = [x_{ij}]$ and $Z = [z_{ij}]$ are matrices of explanatory variables; and Θ and A are matrices of unknown coefficients. Disturbances are independent and identically distributed across observations; that is, taking $E = [\epsilon_1,\ldots,\epsilon_n]'$, then $\epsilon_i \sim \text{IIDN}(0,\Sigma)$. Sometimes it is more convenient to express (2) in "row-stacked form" by taking $y = \text{vec}(Y)$, $\theta = \text{vec}(\Theta)$, $\alpha = \text{vec}(A)$, and $\epsilon = \text{vec}(E)$. Then,

$$\underset{nL\times 1}{y} = (I_L \otimes X)\underset{pL\times 1}{\theta} + (I_n \otimes Z)\underset{kL\times 1}{\alpha} + \epsilon, \quad \epsilon \sim N(0, \Sigma \otimes I_n).$$

Bayesian inference in this model is straightforward for certain reference priors. For example, given the improper prior density $p(\Sigma,\Theta,A) \propto |\Sigma|^{-(L+1)/2}$, the log posterior density kernel is

$$-\left[\frac{n+L+1}{2}\right]\log|\Sigma| - \frac{1}{2}\,\text{tr}[(Y-X\Theta-ZA)'(Y-X\Theta-ZA)\Sigma^{-1}], \tag{3}$$

the marginal distribution of Θ is matricvariate "t", and the marginal distribution of Σ is inverted Wishart [Zellner (1971, Chapter 8)].

2.1 Prior and Unconditional Posterior Distributions

The reduced rank regression model adds to (2) the specification rank $(\Theta) = q < \max(p,L)$. This is equivalent to the parametric specification

$$Y = X\Theta + ZA + E, \quad \Theta = \underset{p\times q}{\Psi} \cdot \underset{q\times L}{\Phi} \tag{4}$$

with no further rank restrictions on either Ψ or Φ. Given an improper, flat prior distribution for Ψ and Φ, these matrices are unidentified in the strong sense that a posterior distribution for their elements does not exist. This paper employs two alternative normalizations that identify Ψ and Φ. For each normalization proper reference priors for all parameters are then introduced.

In Normalization 1, $\Phi = [I_q | \Phi^*]$, and the reference prior is

$$|\Sigma|^{-(L+\underline{\nu}+1)/2} \exp\left[-\frac{1}{2}\,\mathrm{tr}\underline{S}\Sigma^{-1}\right] \exp\left[-\frac{\tau^2}{2}\,(\mathrm{tr}\Phi^{*\prime}\Phi^* + \mathrm{tr}\Psi'\Psi + \mathrm{tr}A'A)\right].$$

Thus, the reference prior is the product of an independent inverted Wishart distribution for Σ with $\underline{\nu}$ degrees of freedom and matrix parameter $\underline{S}$, and independent $N(0,\tau^{-2})$ shrinkage priors for each element of Φ^* and Ψ. The log of the posterior density kernel is then

$$\begin{aligned} &-\left[\frac{L+\underline{\nu}+n+1}{2}\right]\log|\Sigma| - \frac{1}{2}\,\mathrm{tr}(Y_1-X\Psi-ZA_1)'(Y_1-X\Psi-ZA_1)\Sigma^{11} \\ &- \mathrm{tr}(Y_1-X\Psi-ZA_1)'(Y_2-X\Psi\Phi^*-ZA_2)\Sigma^{21} \\ &- \frac{1}{2}\,\mathrm{tr}(Y_2-X\Psi\Phi^*-ZA_2)'(Y_2-X\Psi\Phi^*-ZA_2)\Sigma^{22} - \frac{1}{2}\,\mathrm{tr}(\underline{S}\Sigma^{-1}) \\ &- \frac{\tau^2}{2}\,(\mathrm{tr}\Phi^{*\prime}\Phi^* + \mathrm{tr}\Psi'\Psi + \mathrm{tr}A'A) \end{aligned} \tag{6}$$

where $Y = [\underset{n\times q}{Y_1} | \underset{n\times s}{Y_2}]$, $A = [\underset{k\times q}{A_1} | \underset{k\times s}{A_2}]$, and

$$\Sigma^{-1} = \begin{bmatrix} \underset{q\times q}{\Sigma^{11}} & \underset{q\times s}{\Sigma^{12}} \\ \\ \underset{s\times q}{\Sigma^{21}} & \underset{s\times s}{\Sigma^{22}} \end{bmatrix}$$

($s = L-q$). Of course there are $\binom{L}{q}$ variants of this normalization. A reordering of the equations leaves the model unchanged, but if the reordering changes the variables included in Y_1 (equivalently, Y_2) then the normalization is affected and so is the posterior distribution of Θ.

Rearrangement of the equations for Y_1 and Y_2 in normalization 1 leads to

$$Y_1 = X\Psi + ZA_1 + E_1$$

$$Y_2 = X\Psi\Phi^* + ZA_2 + E_2 = Y_1\Phi^* + Z[A_2 - A_1\Phi^*] + [E_2 - E_1\Phi^*],$$

where

$$E = [\underset{s\times q}{E_1} | \underset{s\times s}{E_2}].$$

The second block of equations is the incomplete simultaneous equation model, for which Bayesian inference was taken up in Dreze (1976) and Dreze and Richard (1983). For earlier treatments by maximum likelihood see Anderson and Rubin (1949, 1950) and Koopmans (1950).

In Normalization 2, $\Psi' = [I_q | \Psi^{*\prime}]$, and the reference prior is $|\Sigma|^{-(L+\underline{\nu}+1)/2}$ $\exp[-(1/2)\mathrm{tr}\underline{S}\Sigma^{-1}]\exp[-(\tau^2/2)(\mathrm{tr}\Phi'\Phi + \mathrm{tr}\Psi^{*\prime}\Psi^* + \mathrm{tr}A'A)]$. The log of the posterior density kernel is

$$-\left[\frac{L+\underline{\nu}+n+1}{2}\right]\log|\Sigma| \tag{7}$$

$$-\frac{1}{2}\,\mathrm{tr}\{[\underline{S} + (Y - X_1\Phi - X_2\Psi^*\Phi - ZA)'(Y - X_1\Phi - X_2\Psi^*\Phi - ZA)]\Sigma^{-1}\}$$

$$-\frac{\tau^2}{2}\,(\mathrm{tr}\Phi'\Phi + \mathrm{tr}\Psi^{*\prime}\Psi^* + \mathrm{tr}A'A)$$

where $X = [X_{1_{n\times q}} | X_{2_{n\times r}}] (r = p-q)$. Here, reorganizing the columns of X can affect the normalization and the implied posterior distribution of Θ.

Normalization 2 is convenient for inference in a cointegrated vector autoregression:

$$Y = X \begin{bmatrix} I_q \\ \Psi^* \end{bmatrix} \Phi + ZA + E.$$

Given an L-component vector time series $\{y_t\}$, row t of Y is $\Delta y_t'$, row t of X is y_{t-1}', and row t of Z is $[\Delta y_{t-1}', \ldots, \Delta_{t-m}']$. The cointegrating vectors are the columns of Ψ, and Φ provides the loading factors (Engle and Granger 1987). Bayesian inference in this context has been taken up by Bauwens and Lubrano (1993) and Kleibergen and van Dijk (1993).

The posterior distributions indicated by (6) and (7) are intractable: the multiplicative interaction of Ψ and Φ precludes analytical expressions for posterior moments or marginal densities of the parameters. The approach here is to provide a straightforward numerical method for the computation of posterior moments that may be used in a variety of applications. The key to this approach is the relative simplicity of some conditional posterior distributions.

2.2 Conditional Posterior Distributions

The respective conditional posterior distributions $\Sigma|(\Phi,\Psi,A,X,Y)$, $A|(\Phi,\Psi,\Sigma,X,Y)$, $\Phi|(\Psi,A,\Sigma,X,Y)$, and $\Psi(\Phi,A,\Sigma,X,Y)$ are all members of the multivariate normal family. This remains true when Φ is replaced by Φ^* (Normalization 1) or Ψ is replaced by Ψ^* (Normalization 2). Here we discuss these distributions, beginning with the simplest. Mathematical details are deferred to Appendix A. Throughout, let $\Theta = \Psi\Phi$ and $Y^* = Y - ZA$.

Conditional distribution of Σ. Given Φ, Ψ, and A, the log conditional posterior density kernel for Σ is

$$-\left[\frac{L+\underline{\nu}+n+1}{2}\right] \log|\Sigma| - \frac{1}{2}\, \mathrm{tr}\{[\underline{S} + (Y^*-X\Theta)'(Y^*-X\Theta)]\Sigma^{-1}\}.$$

The conditional distribution of Σ is therefore inverted Wishart,

$$\Sigma \,|\, (\Phi,\Psi,A,X,Y) \sim IW[n + \underline{\nu},\ \underline{S} + (Y^*-X\Theta)'(Y^*-X\Theta)] \tag{8}$$

[Zellner (1971, pp. 395-99)].

Conditional distribution of A. Given Φ, Ψ, and Σ, the log conditional posterior density kernel for A is

$$-\left[\frac{1}{2}\right] \mathrm{tr}[(Y-X\Theta-ZA)'(Y-X\Theta-ZA)\Sigma^{-1}] \ - \ \left[\frac{\tau^2}{2}\right] \mathrm{tr}A'A$$

the same as that for a multivariate regression model with dependent variable matrix $Y - X\Theta$, explanatory variable matrix Z, and the shrinkage prior. Hence

$$\begin{aligned} &\mathrm{vec}(A) \,|\, (\Phi,\Psi,\Sigma,X,Y) \\ &\quad \sim N\{[\Sigma^{-1} \otimes Z'Z+\tau^2 I_{Lk}]^{-1}[\Sigma^{-1} \otimes Z'Z]\mathrm{vec}(\hat{A}),\ [\Sigma^{-1} \otimes Z'Z+\tau^2 I_{Lk}]^{-1}\} \end{aligned} \tag{9}$$

where $\hat{A} = (Z'Z)^{-1}Z'(Y-X\Theta)$.

Conditional distribution of Φ. In Normalization 2 the log conditional posterior density kernel for Φ is

$$-\frac{1}{2}\ \mathrm{tr}[(Y^*-X\Psi\Phi)'(Y^*-X\Psi\Phi)\Sigma^{-1}] \ - \ \frac{\tau^2}{2}\ \mathrm{tr}\Phi'\Phi$$

of the same form as (3) for the multivariate regression model with X replaced by $X\Psi$ and Θ by Φ, and ZA omitted, plus the shrinkage prior. Hence

$$\begin{aligned} &\mathrm{vec}(\Phi) \,|\, (\Psi,\Sigma,Y^*,X) \\ &\quad \sim N\{[\Sigma^{-1} \otimes (\Psi'X'X\Psi) + \tau^2 I_{Lq}]^{-1}[\Sigma^{-1} \otimes (\Psi'X'X\Psi)]\mathrm{vec}(\hat{\Phi}), \\ &\qquad [\Sigma^{-1} \otimes (\Psi'X'X\Psi) + \tau^2 I_{Lq}]^{-1}\} \end{aligned} \tag{10}$$

where $\hat{\Phi} = (\Psi'X'X\Psi)^{-1}\Psi'X'Y^*$.

Conditional distribution of Ψ. Given Φ, construct the $L \times L$ nonsingular matrix

$$C = \left[\underset{L\times q}{\Phi^+} \,\middle|\, \underset{L\times s}{\Phi^0}\right],$$

where Φ^+ is the Moore-Penrose generalized inverse of Φ (so that $\Phi\Phi^+ = I_q$) and the columns of Φ^+ and Φ^0 are orthogonal. Postmultiplication of (4) by C leaves

$$[Y^*\Phi^+ | Y^*\Phi^0] = [X\Psi | 0] + [E\Phi^+ | E\Phi^0]. \tag{10}$$

The effect of this transformation is to "strip" Φ from the right-hand side of (4) leaving Ψ in the position of a coefficient matrix. However, the new system is not a multivariate regression, since there are restrictions on the coefficients of the last $s = L - q$ equations in the system. It is instead a set of seemingly unrelated regressions, with Σ known. It is well known that the conditional posterior distribution of the coefficients is multivariate normal in this situation, and general expressions [for example, Zellner (1971, Section 8.5)] apply. Appendix A2 derives the more compact result specific to (10), for the case $\tau^2 = 0$,

$$\mathrm{vec}(\Psi) | (\Phi,\Sigma,X,Y^*) \sim N[\mathrm{vec}(\hat{\Psi}), (\tilde{\Sigma}^{11})^{-1} \otimes (X'X)^{-1}],$$

$$\hat{\Psi} = \hat{\Theta}[\Phi^+ + \Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}]$$

where $\tilde{\Sigma}^{ij}$ denotes the partitioning of $\tilde{\Sigma}^{-1} = (C'\Sigma C)^{-1}$ into its first q and last $s = L - q$ rows and columns, and $\hat{\Theta} = (X'X)^{-1}X'Y^*$. Note that in the special case $\Sigma_{12} = 0$, $\Sigma^{12} = \tilde{\Sigma}_{12} = \tilde{\Sigma}^{12} = 0$ as well, and (10) follows directly from the subsystem

$$Y^*\Phi^+ = X\Psi + E\Phi^+$$

which is a multivariate regression. Reintroducing the shrinkage prior for $\Psi(\tau^2 > 0)$

$$\begin{aligned} &\mathrm{vec}(\Psi) | (\Phi,\Sigma,Y^*,X) \\ &\quad \sim N\{[(\tilde{\Sigma}^{11}) \otimes (X'X) + \tau^2 I_{pq}]^{-1}\mathrm{vec}(\hat{\Psi}), [(\tilde{\Sigma}^{11}) \otimes (X'X) + \tau^2 I_{pq}]^{-1}\}. \end{aligned} \tag{11}$$

Conditional distribution of Φ^*. The log posterior density kernel is immediate from (6). Appendix A3 shows that for $\tau^2 = 0$,

$$\mathrm{vec}(\Phi^*)\,|\,(\Psi,\Sigma,X,Y^*) \sim N[\mathrm{vec}(\hat{\Phi}^*),\ (\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1}] \tag{12}$$

$$\hat{\Phi}^* = (\Psi'X'X\Psi)^{-1}\Psi'X'Y_1^*\Sigma^{12}(\Sigma^{22})^{-1} - \Sigma^{12}(\Sigma^{22})^{-1} + (\Psi'X'X\Psi)^{-1}\Psi'X'Y_2^*.$$

(Consideration of the special case $\Sigma_{12} = 0$ again provides some insight into the result.) For $\tau^2 > 0$,

$$\mathrm{vec}(\Phi^*)\,|\,\Psi,\Sigma,X,Y^*) \sim N\{[(\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1} + \tau^2 I_{q(L-q)}]^{-1}\mathrm{vec}(\hat{\Phi}^*),$$
$$[(\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1} + \tau^2 I_{q(L-q)})]^{-1}\}.$$

Conditional distribution of Ψ^*. Beginning from the log posterior density kernel indicated by (7) and assuming $\tau^2 = 0$, Appendix A4 derives

$$\mathrm{vec}(\Psi^*)\,|\,(\Phi,\Sigma,X,Y^*) \sim N[\mathrm{vec}(\hat{\Psi}^*),\ (\tilde{\Sigma}^{11})^{-1} \otimes (X_2'X_2)^{-1}] \tag{15}$$

$$\hat{\Psi}^* = \hat{\Theta}_2[\Phi^+ + \Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}] - \hat{\Theta}_{21}$$

where $\hat{\Theta}_2 = (X_2'X_2)^{-1}X_2'Y^*$ and $\hat{\Theta}_{21} = (X_2'X_2)^{-1}X_2'X_1$. For $\tau^2 > 0$,

$$\mathrm{vec}(\Psi^*)\,|\,(\Phi,\Sigma,X,\Psi) \sim N\{[(\tilde{\Sigma}^{11})^{-1} \otimes (X_2'X_2)^{-1} + \tau^2 I_{q(p-q)}]^{-1}\mathrm{vec}(\hat{\Psi}^*),$$
$$[(\tilde{\Sigma}^{11})^{-1} \otimes (X_2'X_2)^{-1} + \tau^2 I_{q(p-q)}]^{-1}\}.$$

3. Computation of Posterior Moments and Densities

In this work a Gibbs sampling algorithm [Gelfand and Smith (1990)] is used to produce sequences of drawings from the parameter space that are neither independently nor identically distributed, but converge in distribution to the posterior distribution whose log kernel density is either (6) or (7). Consistent with the discussion of Section 2.2 adopt the following notation and groupings of parameters:

$\lambda_1 = \text{vec}(\Phi)$ or $\lambda_1 = \text{vec}(\Phi^*)$

$\lambda_2 = \text{vec}(\Psi)$ or $\lambda_2 = \text{vec}(\Psi^*)$

$\lambda_3 = \text{vec}(A)$

$\lambda_4 = \text{veclt}(\Sigma)$

$\lambda' = (\lambda_1', \lambda_2', \lambda_3', \lambda_4')$.

The algorithm is easy to construct. Begin with arbitrary initial values $\Psi^{(0)}$ for Ψ, $A^{(0)}$ for A, and $\Sigma^{(0)}$ for Σ in the support of the prior distribution. Then for j = 1, 2, ...

(i) Given $\Psi^{(j-1)}$, $A^{(j-1)}$, and $\Sigma^{(j-1)}$, draw $\Phi^{(j)}$ from the distribution (10) or $\Phi^{*(j)}$ from (12);

(ii) Given $\Phi^{(j)}$, $A^{(j-1)}$, and $\Sigma^{(j-1)}$, draw $\Psi^{(j)}$ from the distribution (11) or $\Psi^{*(j)}$ from (15).

(iii) Given $\Phi^{(j)}$, $\Psi^{(j)}$, and $\Sigma^{(j-1)}$, draw $A^{(j)}$ from the distribution (9);

(iv) Given $\Phi^{(j)}$, $\Psi^{(j)}$, and $A^{(j)}$ draw $\Sigma^{(j)}$ from the distribution (8);

These four steps constitute a single pass of the Gibbs sampler. After each pass a function of interest $g(\lambda^{(j)})$ can be computed, and after m passes $m^{-1}\sum_{j=1}^{m} g(\lambda^{(j)})$ provides a numerical approximation to $E[g(\lambda)]$. This section takes up the justification for this procedure.

3.1 Convergence and Numerical Approximation

Weak sufficient conditions for convergence of the Gibbs sampler are provided by Tierney (1991, 1994). Roberts and Smith (1992) provide conditions that are stronger but often much easier to verify in econometric models. That is the case here.

Let $p(\lambda)$ denote the posterior density of the parameter vector λ. Roberts and Smith (1992) show that if

(1) $p(\lambda)$ is lower semicontinuous at 0;

(2) $\int p(\lambda) d\lambda_j$ is locally bounded (j = 1,2,3,4);

(3) the support of $p(\lambda)$ is connected;

(4) $E[g(\lambda)] = \int g(\lambda)p(\lambda)d\lambda$ is convergent;

then,

$$m^{-1}\sum_{j=1}^{m} g(\lambda^{(j)}) \xrightarrow{a.s.} E[g(\lambda)].$$

Condition (1) requires that for all λ with $p(\lambda) > 0$, there exists an open neighborhood $N_\lambda \supset \lambda$ and $\epsilon > 0$ such that for all $\tilde{\lambda} \in N_\lambda$, $p(\tilde{\lambda}) \geq \epsilon > 0$. This condition is easily verified by inspection of the log posterior density kernel (6) or (7).

The posterior density kernel is the product of the bounded likelihood function for Ψ, Φ, A, and Σ, and independent prior distributions for each group of parameters. Hence condition (2) is satisfied. Proper prior distributions for Ψ and Φ are important here. For example, if the prior distribution for Φ^* were flat, then conditional on rank $(\Psi) < q$ condition (2) would be violated. This has been documented in some detail by Kleibergen and van Dijk (1991) in simultaneous equation models. While condition (2) is satisfied for $\tau^2 > 0$, as a practical matter it is natural to suspect that convergence of the Gibbs sampler could be slow if τ^2 is small.

Condition (3) is obviously satisfied. If the prior distribution is proper and the product of $|g(\lambda)|$ and the likelihood function is uniformly bounded, then condition (4) is satisfied. This will be the case for the functions of interest introduced in Section 4 and used in the examples in Sections 5 and 6.

3.2 Evaluation of numerical accuracy

A compelling advantage of Monte Carlo integration methods in general is that accuracy may be assessed through a central limit theorem [for example, Geweke (1989, Theorem 2)]. In the case of the Gibbs sampler this strategy is complicated by the fact that the process $\{\lambda^{(j)}\}$ is neither

independently nor identically distributed. The limiting distribution of $m^{1/2}(\bar{g}_m - E[g(\lambda)])$ is known to be normal under several sets of assumptions. Some require that the support of λ be bounded and therefore do not apply to our problem. Others [for example, Nummelin (1984, Corollary 7.3)] pertain to bounded $g(\lambda)$, and consequently apply to the computation of posterior probabilities but not posterior expectations of parameters. Even in these cases there are no subsidiary results supporting estimation of the variance of the limiting distribution. The strategy adopted here is to employ an estimated variance that would be appropriate if $\{\lambda^{(j)}\}$ were a serially correlated but identically distributed process, and then make certain checks for internal consistency.

Under the assumption that $\{\lambda^{(j)}\}$ is identically distributed and serially correlated the approximation of $E[g(\lambda)]$ is equivalent to the classical problem of mean estimation in time series analysis. A full development is given in Geweke (1991) and is only outlined here. Given that $g(\lambda^{(j)})$ has finite mean and variance, $\bar{g}_m = m^{-1}\sum_{j=1}^{m} g(\lambda^{(j)}) \rightarrow \bar{g} \equiv E[g(\lambda)]$. Under weak conditions [Hannan 1970, Section 2.2)] the spectral density $S(\omega)$ of $g(\lambda^{(j)})$ exists; and $m^{1/2}(\bar{g}_m - \bar{g}) \Rightarrow N[0,S(0)]$ [Hannan 1970, Theorem 4.11)]. If $\hat{S}_m(\omega)$ is a consistent (in m) estimator of $S(\omega)$, then the accuracy of $\bar{g}_m$ as an approximation of $\bar{g}$ may be assessed by the numerical standard error (NSE) $[m^{-1}\hat{S}_m(0)]^{1/2}$. Many consistent estimators of $S(0)$ are available.

Especially in the absence of a central limit theorem that pertains to all $g(\lambda)$, and of a demonstrated consistent estimator of the variance term in the limiting normal distribution, it is important to assess the adequacy of the computed NSE's. In the work reported here that is done by repeating the computations with different initial conditions and a different seed for the random number generator. It was always the case that differences in computed posterior moments were consistent with computed NSE's and the assumption of normality.

4. Predictive Odds Ratios for Alternative Ranks

To this point we have proceeded as if q were known. In most applications this will not be true and so analysis to this point is conditional. For example, in a cointegrated vector autoregression $q = \text{rank}(\Theta)$ is the number of cointegrating vectors, whose value is often the object of the research. Another example in which the value of q is central is presented in Section 6.

When q is unknown, the analysis may be carried out for several alternative values of q, and one may then compare the performance of the different models. In general, the computation of Bayes factors is difficult in a numerical approach, because standard Monte Carlo methods do not yield the necessary integrals as a byproduct. Here we outline a general method for obtaining predictive factors for model comparison, presented fully in Geweke (1994). Chib (1994) has described a procedure specific to the Gibbs sampler that should be well suited to the approach taken in Sections 2 and 3, as well.

Denote by y_s row s of Y, by x_s row s of X, by Y_s the first s rows of Y, and by X_s the first s rows of X. Given normalization $j (j = 1,2)$ and $\text{rank}(\Theta) = q$ ("model jq") let λ_{jq} denote the vector of parameters for the model. Then we may denote the conditional density $y_s | (x_s, \lambda_{jq})$ under normalization j and $\text{rank}(\Theta) = q$ by $f_{jq}(y_s | x_s, \lambda_{jq})$. Let the prior probability of model j be p_{jq} and denote the prior density $f_{jq0}(\lambda_{jq})$. Let

$$L_{jqt}(\lambda_{jq} | Y_t, X_t) = \prod_{s=1}^{t} f_{jq}(y_s | x_s, \lambda_{jq})$$

denote the partial likelihood through observation t. Then conditional on model jq, Y_t, and X_t, the posterior density for λ_{jq} is

$$p_{jqt}(\lambda_{jq} | Y_t, X_t) = f_{jq0}(\lambda_{jq}) L_{jqt}(\lambda_{jq} | Y_t, X_t) / \int f_{jq0}(\lambda_{jq}) L_{jqt}(\lambda_{jq} | Y_t, X_t) d\lambda_{jq}.$$

The predictive likelihood for observations u + 1 through t, given model jq and observations 1 through u, is

$$\hat{p}^t_{jqu} = \int p_{jqu}(\lambda_{jq} | Y_u, X_u) \prod_{s=u+1}^{t} f_{jq}(y_s | x_s, \lambda_{jq}) d\lambda_{jq}. \tag{16}$$

The predictive Bayes factor in favor of model jq versus model j′q′ for observations u + 1 through t is $\hat{B}^t_{jq|j'q',u} = \hat{p}^t_{jqu}/\hat{p}^t_{j'q'u}$ and the predictive odds ratio is $PRED^t_{jq|j'q'} = p_{jq}\hat{p}^t_{jqu}/p_{j'q'}\hat{p}^t_{j'q'u}$. These decompositions are widely known, but direct numerical approximation of the integral in (16) is difficult for t/u large, because the function of interest $\Pi^t_{s=u+1} f_{jqs}(y_s | x_s, \lambda_{jq})$ is then highly concentrated relative to the more disperse posterior density $p_{jqu}(\lambda_{jq} | Y_u, X_u)$.

A further decomposition facilitates this computation and provides useful diagnostic tools. Suppose $u = s_0 < s_1 < \ldots < s_m = t$. Expanding (16), recalling that $\int p_{jqs_0}(\lambda_{jq} | Y_{s_0}, X_{s_0}) d\lambda_{jq} = 1$, and adopting the notational convention $\Pi^{s_0}_{s=s_0+1} f_{jqs}(y_s | x_s, \lambda_{jq}) = 1$, we have

$$\hat{p}^t_{jqu} = \prod_{\ell=1}^{m} \hat{p}^{s_\ell}_{jqs_{\ell-1}}. \tag{17}$$

(The result may be verified by expansion of the product followed by cancellations.) By choosing a sequence of s_ℓ such that $s_\ell - s_{\ell-1}$ is not too large, the function of interest $\Pi^{s_\ell}_{s=s_{\ell-1}+1} f_{jqs}(y_s | x_s, \lambda_{jq})$ may be made sufficiently diffuse relative to the posterior density $p_{jqs_{\ell-1}}(\lambda_{jq} | Y_{s_{\ell-1}}, X_{s_{\ell-1}})$ that

$$E_{jqs_{\ell-1}}\left[\prod_{s=s_{\ell-1}+1}^{s_\ell} f_{jqs}(y_s | x_s, \lambda_{jq})\right] = \int p_{jqs_{\ell-1}}(\lambda_{jq} | Y_{s_{\ell-1}}, X_{s_{\ell-1}}) \left[\prod_{s=s_{\ell-1}+1}^{s_\ell} f_{jqs}(y_s | x_s, \lambda_{jq})\right] d\lambda_{jq}$$

can be well approximated by Monte Carlo methods. (Geweke 1994 provides additional details on the increase in numerical efficiency provided by this decomposition.)

The particular choice $s_\ell - s_{\ell-1} = 1$ provides a complete decomposition of the predictive likelihood. Then (17) is of interest as a model diagnostic: an unusually low value of $\hat{p}^{s_\ell}_{s_{\ell-1}}$ indicates

that observation s_ℓ is improbable conditional on model jq and the previous observations. There is a corresponding decomposition of the Bayes factor

$$\hat{B}^t_{jq|j'q',u} = \hat{p}^t_{jqu}/\hat{p}^t_{j'q'u} = \prod_{\ell=1}^{m} (\hat{p}^{s_\ell}_{jqs_{\ell-1}}/\hat{p}^{ts_\ell}_{j'q's_{\ell-1}}) = \prod_{\ell=1}^{m} \hat{B}^{s_\ell}_{jq|j'q',s_{\ell-1}}.$$

5. Some Numerical Evidence

To illustrate the use of predictive odds ratios to provide evidence about alternative ranks, we present results using artificial data before turning to an application in the next section. In this illustration, $n = 332$ observations from a system with $L = 12$ dependent variables and $p = 7$ covariates, were generated. The values of n, L, and p match those in the application in the next section. The coefficient matrix Θ is indicated in Table 1; it has rank $q = 3$. The disturbances are normally distributed with mean 0, and variance matrix Σ also indicated in Table 1. The first covariate is an intercept, and the remaining covariates are normally distributed with mean 0, and variance matrix indicated in Table 1.

Two experiments illustrate the properties of the predictive odds ratio. In both experiments the prior probability of each of the 14 models (jq; $j = 1,2$; $q = 1,\ldots,7$) is the same, and the value of the prior parameter $\alpha = 1$. Similar results were obtained for $\alpha = 10^{-3}$. In both experiments all numerical approximations to predictive densities were computed using 1,000 iterations of the Gibbs sampler preceded by a burn-in of 100 iterations. More iterations (up to 10,000) and a longer Gurn-in (up to 1,000 iterations) had no effects on posterior moments, up to numerical standard error computed as described in Geweke (1991).

In the first experiment, five independent subsamples of the data consisting of 48 observations each were used to examine the behavior of the predictive odds ratio in small samples. The first $u = 25$ observations of each sample were used to form the posterior distribution, and observations

$u + 1 = 26$ through $t = 48$ were used to form the predictive density. This predictive density was computed using the complete decomposition (17): $s_1 = 26$, $s_2 = 27$, ..., $s_m = s_{23} = 48$. Thus, as a byproduct of this experiment we have available predictive odds ratios using the posterior density formed from the first 25 observations, and predictive densities from observations 26 through $t(t = 26,...48)$. The predictive odds ratios may be transformed into predictive probabilities for the 14 models (normalizations $j = 1,2$ and ranks $q = 1,...,7$) in the obvious way.

The results of the first experiment are presented in Table 2. For each of the five samples, posterior probabilities for predictive densities with 2, 5, 10, and 23 observations, respectively, are presented. For the 14 models (two normalizations and seven ranks) marginal predictive probabilities for the two normalizations and for the seven ranks, are given. (Predictive probabilities for ranks conditional on normalization, not shown, are similar to the presented marginal predictive probabilities for ranks, in every case.)

The predictive probabilities for the normalizations vary considerably with the dates entering into the predictive density, but overall there is no tendency to favor one normalization over another. This outcome cannot be deduced, but it is plausible. The alternative normalizations, together with their respective shrinkage priors, constituted different prior distributions for the coefficient matrix Θ. It is reasonable to anticipate that the differences in the prior distributions are minor and should be dominated by the 25 observations in the posterior distribution. This appears to be the case.

The rank $q = 1$ model receives negligible predictive probability, even with only two observations in the predictive density. The predictive probability of the rank $q = 2$ model drops quickly as observations are added to the predictive density. The predictive probability of the population model, $q = 3$, rises from less than 0.5 with two observations in the predictive density, to more than 0.95 with 23 observations, and is modal in four of the five cases with five observations in the predictive density. This increased predictive probability comes at the expense of larger values

of q. As the number of observations in the predictive density increases, predictive probability for $q \geq 4$ becomes monotone decreasing in q, as well as small. This documents the well-known consistency of Bayesian model selection (Schwarz 1978).

The second experiment explores the effects of choosing alternative sample sizes, u, for the predictive density. Sample sizes of $u = 24$, 48, 96, 192, and 288 for the posterior density were considered, together with predictive densities incorporating $t - u = 24$, 48, 120, and 240 observations as permitted by the sample, as well as $t = 332$ in every case. Table 3 displays the results of this experiment, using the same format as Table 2. Once again, neither normalization is dominated by the other in predictive probability. For posterior distributions with $u = 24$ and $u = 48$ observations the limiting behavior of the predictive probabilities for all alternative ranks q is evident as $t - u$ increases. For $u = 96$, 192, or 288, the models $q = 1$ and $q = 2$ receive negligible predictive probability. But while $q = 3$ is modal when $u = 96$, the limiting concentration of predictive probabilities on $q = 3$ is not evident in any of these three cases.

These results are consistent with the asymptotic behavior of posterior odds ratios. Suppose u is fixed, while $t - u \rightarrow \infty$. Then the results of Schwarz (1978) apply: predictive probability must concentrate on $q = 3$. Log predictive odds ratios will be proportional asymptotically to $t - u$ for $q \leq 3$, and proportional asymptotically to $\log(t-u)$ for $q \geq 3$, so models with $q < 3$ are ruled out well before those with $q > 3$. On the other hand, suppose $t - u$ is fixed while $u \rightarrow \infty$. Then the posterior distributions for Θ and Σ collapse to points, which are the same for $q \geq 3$. The limiting distribution of the predictive probability for ranks $q = 1$, 2, and 3 is nondegenerate, while the predictive probabilities for ranks $q = 4$, 5, 6, and 7 become identical to the predictive probability for rank $q = 3$.

6. Application to a Capital Asset Pricing Model

This paper concludes with an application to an asset pricing model. The objective in this application is to find exact predictive probabilities for the rank of the coefficient matrix, which is the number of factors in the asset pricing model.

6.1 The Model

Let R_{it} denote the realized return on asset i in period t, for some collection of assets i = 1, ..., n and periods t = 1, ..., T. Let z_t denote a p-dimensional vector of information available in period t, including the risk-free rate of return. Assume that the expectation of the returns conditional on the information z_{t-1} is linear in z_{t-1}:

$$R_{it} = \delta_i' z_{t-1} + u_{it}, \quad E(u_{it} | z_{t-1}) = 0 (i = 1,\ldots,n;\ t = 1,\ldots,T).$$

Assume further, as is typically done in the asset pricing literature, that $(u_{it},\ldots,u_{nt}) \sim IIDN(0,\Sigma)$. In matrix notation,

$$R = Z\Delta + U \tag{18}$$

where R has element R_{it} in row t and column i, $Z' = [z_0,\ldots,z_{T-1}]$, $\Delta = [\delta_1,\ldots\delta_n]$, and U has element u_{it} in row t and column i. With the obvious change in notation, this is (1) without the additional matrix of explanatory variables "Z" of that equation.

A k-factor asset pricing model may be expressed

$$E(R_{it} | z_{t-1}) = E(\tilde{R}_{0t} | z_{t-1}) + \sum_{h=1}^{q} \beta_{ih} E(\tilde{R}_{ht} - \tilde{R}_{0t} | z_{t-1}) \ (i = 1,\ldots,n;\ t = 1,\ldots,T) \tag{19}$$

where the $\tilde{R}_{ht}$ are returns on k unobserved hedge portfolios, and $\tilde{R}_{0t}$ is the return on a portfolio uncorrelated with the hedge. Gibbons and Ferson (1985) show, using an iterated expectations argument, that (19) implies rank $(\Delta) = q$ in (18). So long as $q < \min(n,p)$ this places a restriction

on Δ of exactly the same form as (4) for Θ (where once again the matrix of variables "Z" in the earlier equation does not appear).

Gibbons and Ferson (1985) express the rank restriction

$$\Delta = [\Delta^* | \Delta^* C]. \tag{20}$$

This makes the further assumption that the first k columns of Δ are linearly independent, which is also implied by normalization 1. They test the restriction (20) using asymptotic sampling theory for maximum likelihood estimators and find that a one-factor model cannot be rejected. Ferson and Harvey (1991) develop a similar model which leads to the same restrictions in a multivariate regression model for returns. They use monthly data on returns. Applying a multi-step procedure, they also conclude in favor of a one-factor model.

6.2 Data and Priors

The data consist of 332 monthly observations on returns and information, from May 1959 through December 1986. The returns pertain to 12 portfolios of securities of New York Stock Exchange listed firms, grouped by two-digit standard industrial classifications. The 12 portfolios are described in Table 4. The vector of information z_t consists of six proxies for the economic risks that influence security returns, plus an intercept. The six variables are also described in Table 4. The data are the same used in Ferson and Harvey (1991) and were graciously supplied by the authors. Further description and details of construction may be found in Ferson and Harvey (1991).

All results presented here pertain to a prior distribution with $\tau^2 = \underline{\nu} = 1$ and $\underline{S} = I$. Variants on this prior distribution ($\tau^2 = \underline{\nu} = 0.01$, $\underline{S} = 0.01I$, and $\tau^2 = \underline{\nu} = 10$, $\underline{S} = 10I$) produce no substantial changes in the results.

6.3 Results

Two sets of results are presented, in Tables 5 and 6 respectively. The first set uses a rolling sequence of data sets. Each set is four years (48 observations) long, with the first two years used to provide the posterior density and the last two years the predictive density. The second set uses the entire sample, with varying divisions of data between the posterior and predictive densities.

The results strongly support a small number of factors: often $q = 1$, but for some subperiods there is evidence of $q > 1$. A striking feature of the result is that normalization 1 is strongly preferred to normalization 2.

In Table 5, the evidence is consistent with $q = 1$ except for three subperiods: 5/69–4/70, 5/71–4/73, and 5/79–4/81. In the first two of these periods the modal predictive probability is $q = 2$, and only in the latter subperiod is there any substantial evidence for $q > 2$. In interpreting Table 6, it is important to keep in mind the point noted at the end of Section 5: when the sample period is long and the predictive period is short, predictive probabilities do not discriminate well among models that have been overfit. Given this consideration the results in Table 6 are consistent with those in Table 5: the evidence favors $q = 1$ for most subperiods, and $q > 1$ during 5/71–4/73 and 5/79–4/81 accounts for substantial departures from $q = 1$. The support for $q = 1$ is consistent with many investigations of the number of factors in asset pricing models, for example, Gibbons and Ferson (1985).

When $q = 1$ each column of the coefficient matrix is proportional to the first column in normalization 1, and each row of the coefficient matrix is proportional to the first row in normalization 2. The first column of the coefficient matrix corresponds to the coefficients for the first portfolio, in this application, whereas the first row corresponds to the intercept. Theory suggests that the intercept term should be negligible, and empirical results (Ferson and Harvey 1991, Table 6) support this prediction. Thus, normalization 2 imposes a structure inconsistent with the theory. In

Table 5, except for 5/71–4/73, every 24–month predictive period favors normalization 1 by a predictive odds ratio of at least 70:1, and every 12–month predictive period favors it by at least 4:1. These very high predictive odds ratios support the theory. In Table 6 the results are more mixed, but with the exception of the bottom panel predictive densities not involving the 5/71–4/73 period favor normalization 1.

More detailed examination of the modal predictive probabilities for subperiods shows that using 5/69–4/71 as the sample period, the modal predictive probability shifts gradually from $q = 1$ to $q = 2$ over the period 1/71–3/73. Using 5/77–4/79 as the sample period, the modal predictive probability moves sharply from $q = 1$ to $q = 2$ in 1/80. Events during these periods could provide a focus for determining the causes of the shifts apparent in Table 5. More generally, a model with regime shifts or time-varying parameters might be appropriate.

Appendix A:

Conditional Posterior Distributions in the Reduced Rank Regression Model

The reduced rank regression model is

$$\underset{n\times L}{Y} = \underset{n\times p}{X}\ \underset{p\times L}{\Theta} + \underset{n\times k}{Z}\ \underset{k\times L}{A}\ \underset{n\times L}{E} = X\ \underset{p\times q}{\Psi}\ \underset{q\times L}{\Phi} + E; \quad \mathrm{vec}(E) \sim N(0, \Sigma \otimes I_n).$$

This appendix derives distributions for Σ, Φ, and Ψ, respectively, conditional on the other parameter groups, at the limit points $\underline{\nu} = \tau^2 = 0$, $\underline{S} = 0$, of the prior distributions described in Section 2. This reduces notational clutter, and reintroduction of the proper priors ($\underline{\nu} > 0$, $\tau^2 > 0$, $\underline{S}$ p.d.) is straightforward. All derivations begin from the expression

$$-\left[\frac{n+L+1}{2}\right]\log|\Sigma| - \frac{1}{2}\,\mathrm{tr}[(Y^*-X\Psi\Phi)'(Y^*-X\Psi\Phi)\Sigma^{-1}] \tag{A1}$$

where $Y^* = Y - ZA$. This is the log of each proper conditional posterior density function kernel, subject to any normalizations.

A1. Posterior Distribution of Φ and Σ Conditional on Ψ

Let $\tilde{X} = X\Psi$, and write (A1) in the form

$$-\left[\frac{n+L+1}{2}\right]\log|\Sigma| - \frac{1}{2}\,\mathrm{tr}[(Y^*-\tilde{X}\Phi)'(Y^*-\tilde{X}\Phi)].$$

This is the log posterior density kernel for the parameters of the multivariate regression model, with prior density $p(\Sigma,\Phi) \propto |\Sigma|^{-(L+1)/2}$ [Zellner (1971, pp. 225–27)]. Consequently,

$$\Sigma\,|\,(\Psi,\Phi,X,Y^*) \sim IW(S,T-q) \tag{A2}$$

$$\mathrm{vec}(\Phi)\,|\,(\Psi,\Sigma,X,Y^*) \sim N(\mathrm{vec}(\hat{\Phi}),\ \Sigma \otimes (\tilde{X}'\tilde{X})^{-1}),$$

where $\hat{\Phi} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y^* = (\Psi'X'X\Psi)^{-1}\Psi'X'XY^*$ and $S = (Y^*-\tilde{X}\hat{\Phi})'(Y^*-\tilde{X}\hat{\Phi}) = Y^{*\prime}Y^* - Y^{*\prime}\tilde{X}\hat{\Phi}$.

A2. Posterior Distribution of Ψ Conditional on Φ and Σ

From (A1) the conditional log posterior density kernel is

$$-\frac{1}{2}\,\mathrm{tr}[(Y^*-X\Psi\Phi)'(Y^*-X\Psi\Phi)\Sigma^{-1}]. \tag{A3}$$

Construct the $L \times L$, nonsingular matrix $C = [\Phi^+_{L\times q}|\Phi^0_{L\times s}](s \equiv L-q)$ as follows. Denote the singular value decomposition [Press *et al.*, (1986, pp. 52–64)] of Φ,

$$\underset{q\times L}{\Phi} = \underset{q\times q}{U'}\left[\underset{q\times q}{D} \mid \underset{q\times s}{0}\right]\underset{L\times L}{V} = U'[D|0]\begin{bmatrix} \underset{q\times L}{V_1} \\ \underset{s\times L}{V_2} \end{bmatrix}$$

where U and V are orthonormal and D is a diagonal matrix whose diagonal elements are strictly positive with probability one. Then the Moore-Penrose generalized inverse [Theil (1971, pp. 268–70)] of Φ is

$$\underset{L\times q}{\Phi^+} = \underset{L\times q}{V_1'}\,\underset{q\times q}{D^{-1}}\,\underset{q\times q}{U}, \quad \text{and let } \underset{L\times s}{\Phi^0} = V_2'; \quad \Phi\Phi^+ = I_q, \quad \Phi\Phi^0 = \Phi^{+\prime}\Phi^0 = \underset{q\times s}{0}.$$

Now re-express the log conditional posterior density kernel (A3) as

$$-\frac{1}{2}\,\mathrm{tr}[C'^{-1}C'(Y-X\Psi\Phi)'(Y-X\Psi\Phi)]CC^{-1}\Sigma^{-1}]$$

$$= -\frac{1}{2}\,\mathrm{tr}[(Y^*C-X\Psi\Phi C)'(Y^*C-X\Psi\Phi C)(C'\Sigma C)^{-1}]$$

which is that for the model

$$[\tilde{Y}_1^*|\tilde{Y}_2^*] \equiv Y^*[\Phi^+|\Phi^0] = [X\Psi|0] + [E\Phi^+|E\Phi^0] = [X\Psi|0] + [\tilde{E}_1|\tilde{E}_2]. \tag{A4}$$

If we define $\tilde{y}^*_{(1)} = \mathrm{vec}(\tilde{Y}_1^*)$, $\tilde{y}^*_{(2)} = \mathrm{vec}(\tilde{Y}_2^*)$, $\tilde{\epsilon}_{(1)} = \mathrm{vec}(\tilde{E}_1)$, $\tilde{\epsilon}_{(2)} = \mathrm{vec}(\tilde{E}_2)$, $\psi = \mathrm{vec}(\Psi)$, then

$$\begin{bmatrix} \tilde{y}^*_{(1)} \\ \tilde{y}^*_{(2)} \end{bmatrix} = \begin{bmatrix} I_q \otimes X \\ 0 \end{bmatrix}\psi + \begin{bmatrix} \tilde{\epsilon}_{(1)} \\ \tilde{\epsilon}_{(2)} \end{bmatrix} = Z\psi + \tilde{\epsilon}, \quad \mathrm{var}(\tilde{\epsilon}) = \tilde{\Sigma} \otimes I_n$$

where $\tilde{\Sigma} = C'\Sigma C$. The conditional posterior distribution of ψ is normal with variance

$$[Z'(\tilde{\Sigma}^{-1} \otimes I_n)Z]^{-1} = (\tilde{\Sigma}^{11})^{-1} \otimes (X'X)^{-1}$$

where $\tilde{\Sigma}^{11}$ denotes the first q rows and columns of $\tilde{\Sigma}^{-1}$. The mean of this normal distribution is

$$\begin{aligned}\hat{\psi} &= [(\tilde{\Sigma}^{11})^{-1} \otimes (X'X)^{-1}]Z'(\tilde{\Sigma}^{-1} \otimes I_n)\tilde{y}^* \\ &= [(\tilde{\Sigma}^{11})^{-1} \otimes (X'X)^{-1}][(\tilde{\Sigma}^{11} \otimes X')\tilde{y}^*_{(1)} + (\tilde{\Sigma}^{12} \otimes X')]\tilde{y}^*_{(2)} \\ &= [I_q \otimes (X'X)^{-1}X']\tilde{y}^*_{(1)} + \{[(\tilde{\Sigma}^{11})^{-1}\tilde{\Sigma}^{12}] \otimes (X'X)^{-1}X'\}\tilde{y}^*_{(2)}.\end{aligned} \tag{A5}$$

If we take $T \equiv (\tilde{\Sigma}^{11})^{-1}\tilde{\Sigma}^{12}$ then the second summand in (A5) is

$$\begin{bmatrix} (X'X)^{-1}X' \sum_{j=1}^{s} \tilde{y}^*_{q+j}\tau_{1j} \\ \vdots \\ (X'X)^{-1}X' \sum_{j=1}^{s} \tilde{y}^*_{q+j}\tau_{qj} \end{bmatrix}.$$

Hence $\hat{\psi} = \mathrm{vec}(\hat{\Psi})$, with

$$\begin{aligned}\hat{\Psi} &= (X'X)^{-1}X'[\tilde{Y}^*_1 + \tilde{Y}^*_2 T'] = (X'X)^{-1}X'[Y^*\Phi^+ + Y^*\Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}] \\ &= \hat{\Theta}[\Phi^+ + \Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}]\end{aligned}$$

where $\hat{\Theta} \equiv (X'X)^{-1}X'Y^*$.

A3. Posterior Distribution of Φ^* Conditional on Ψ and Σ

From (A1), the log of the conditional posterior density kernel is

$$-\frac{1}{2}\,\mathrm{tr}[(Y^* - \tilde{X}\Phi)'(Y^* - \tilde{X}\Phi)\Sigma^{-1}] \tag{A6}$$

where $\tilde{X} = X\Psi$ and $\Phi = [I_q | \Phi^*]$, with Φ^*: $q \times s (s = L - q)$. Rewrite the term $Y^* - \tilde{X}\Phi$ in (A6),

$$Y^* - \tilde{X}\Phi = \begin{bmatrix} \underset{n\times q}{Y^*_1} \,|\, \underset{n\times s}{Y^*_2} \end{bmatrix} - [\tilde{X} | \tilde{X}\Phi^*] = [Y^*_1 - \tilde{X} | Y^*_2] - [0 | X\Phi^*] = [\tilde{Y}^*_1 | Y^*_2] - [0 | X\Phi^*].$$

Denote $\tilde{y}^*_{(1)} \equiv \mathrm{vec}(\tilde{Y}^*_1)$, $y^*_{(2)} \equiv \mathrm{vec}(Y^*_2)$, $\phi^* = \mathrm{vec}(\Phi^*)$. Then the log kernel (A6) may be expressed

$$-\frac{1}{2}\,\mathrm{tr}\left\{\left[\begin{bmatrix}\tilde{y}^*_{(1)}\\ y^*_{(2)}\end{bmatrix} - \begin{bmatrix}0\\ I_s \otimes \tilde{X}\end{bmatrix}\phi^*\right]'(\Sigma^{-1} \otimes I_n)\left[\begin{bmatrix}\tilde{y}^*_{(1)}\\ y^*_{(2)}\end{bmatrix} - \begin{bmatrix}0\\ I_s \otimes \tilde{X}\end{bmatrix}\phi^*\right]\right\}. \tag{A7}$$

Let

$$Z \equiv \begin{bmatrix}0\\ I_s \otimes \tilde{X}\end{bmatrix}.$$

Then (A7) implies that, conditional on Σ and Ψ, $\phi^* \sim N(\hat{\phi},V)$ with

$$\begin{aligned} V &= [Z'(\Sigma^{-1} \otimes I_n)Z]^{-1} = [(I_s \otimes \tilde{X}')(\Sigma^{22} \otimes I_n)(I_s \otimes \tilde{X})]^{-1} = [\Sigma^{22} \otimes \tilde{X}'\tilde{X}]^{-1} \\ &= (\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1} \end{aligned}$$

where Σ^{22} denotes the last s rows and columns of Σ^{-1}. The conditional mean is

$$\begin{aligned} \hat{\phi}^* &= [(\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1}]Z'(\Sigma^{-1} \otimes I_n)\begin{bmatrix}\tilde{y}^*_{(1)}\\ y^*_{(2)}\end{bmatrix} \\ &= [(\Sigma^{22})^{-1} \otimes (\Psi'X'X\Psi)^{-1}][(\Sigma^{21} \otimes \tilde{X}')\tilde{y}^*_{(1)} + (\Sigma^{22} \otimes \tilde{X}')y^*_{(2)}] \\ &= [(\Sigma^{22})^{-1}\Sigma^{21} \otimes (\tilde{X}'\tilde{X})^{-1}\tilde{X}']\tilde{y}^*_{(1)} + [I_s \otimes (\tilde{X}'\tilde{X})^{-1}\tilde{X}']y^*_{(2)}. \end{aligned}$$

Reorganizing this expression as $\hat{\Phi}^*(\hat{\phi}^* = \mathrm{vec}(\hat{\Phi}^*))$,

$$\begin{aligned} \hat{\Phi}^* &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'[\tilde{Y}^*_1\Sigma^{12}(\Sigma^{22})^{-1} + Y^*_2] \\ &= (\Psi'X'X\Psi)^{-1}\Psi'X(Y^*_1 - X\Psi)\Sigma^{12}(\Sigma^{22})^{-1} + (\Psi'X'X\Psi)^{-1}\Psi'X'Y^*_2 \\ &= (\Psi'X'X\Psi)^{-1}\Psi'X'Y^*_1\Sigma^{12}(\Sigma^{22})^{-1} - \Sigma^{12}(\Sigma^{22})^{-1} + (\Psi'X'X\Psi)^{-1}\Psi'X'Y^*_2. \end{aligned}$$

When Φ is normalized in this way (A2) no longer applies; instead,

$$\Sigma\,|\,(\Psi,\Phi,X,Y^*) \sim IW(S,n)$$

with $S = (Y^* - X\Psi\Phi)'(Y^* - X\Psi\Phi)$.

A4. Posterior Distribution of Ψ^* Conditional on Φ and Σ

When $\Psi' = [I_q | \Psi^{*\prime}]$, let $X = [X_{1_{n\times q}} | X_{2_{n\times r}}](r = p-q)$ and write (A4) as

$$[\tilde{Y}_1^* | \tilde{Y}_2^*] \equiv [Y^*\Phi^+ - X_1 | Y^*\Phi^0] = [X_2\Psi^* | 0] + [E\Phi^+ | E\Phi^0] = [X_2\Psi^* | 0] + [\tilde{E}_1 | \tilde{E}_2].$$

(Note $\tilde{Y}_2^*$ is defined differently here than in Section A2.) If we define $\tilde{y}_{(1)}^* = \mathrm{vec}(\tilde{Y}_1^*)$, $\tilde{y}_{(2)} = \mathrm{vec}(\tilde{Y}_2^*)$, $\tilde{\epsilon}_{(1)} = \mathrm{vec}(\tilde{E}_1)$, $\tilde{\epsilon}_{(2)} = \mathrm{vec}(\tilde{E}_2)$, $\psi^* = \mathrm{vec}(\Psi^*)$, then

$$\begin{bmatrix} \tilde{y}_{(1)}^* \\ \tilde{y}_{(2)}^* \end{bmatrix} = \begin{bmatrix} I_q \otimes X_2 \\ 0 \end{bmatrix} \psi^* + \begin{bmatrix} \tilde{\epsilon}_{(1)} \\ \tilde{\epsilon}_{(2)} \end{bmatrix} = Z\psi^* + \tilde{\epsilon}, \quad \mathrm{var}(\tilde{\epsilon}) = \tilde{\Sigma} \otimes I_n$$

where $\tilde{\Sigma} = A'\Sigma A$. The conditional posterior distribution for ψ^* is normal with variance

$$[Z'(\tilde{\Sigma}^{-1} \otimes I_n)Z]^{-1} = (\tilde{\Sigma}^{11})^{-1} \otimes (X_2'X_2)^{-1}.$$

The mean of this normal distribution is

$$\hat{\psi} = [(\tilde{\Sigma}^{11})^{-1} \otimes (X_2'X_2)^{-1}]Z'(\tilde{\Sigma}^{-1} \otimes I_n)\tilde{y}^*.$$

Following the same manipulations as with (A5), obtain

$$\hat{\Psi}^* = (X_2'X_2)^{-1}X_2'[Y^*\Phi^+ - X_1 + Y^*\Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}] = \hat{\Theta}_2[\Phi^+ + \Phi^0\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}] - \hat{\Theta}_{21}$$

where $\hat{\Theta}_{22} = (X_2'X_2)^{-1}X_2'Y^*$ and $\hat{\Theta}_{21} = (X_2'X_2)^{-1}X_2'X_1$.

Appendix B:

Applying Shrinkage Priors in the Reduced Rank Regression Model

Given a flat prior, conditional posterior distributions for ψ and ϕ are of the form

$$N\left(\underset{mn\times 1}{r}, \underset{m\times m}{A} \otimes \underset{n\times n}{B}\right).$$

Given a normal prior distribution $N(0,\tau^{-2}I_{mn})$, the conditional posterior distribution is normal with precision $A^{-1} \otimes B^{-1} + \tau^2 I_{mn}$, and mean $[A^{-1} \otimes B^{-1}+\tau^2 I_{mn}]^{-1}(A^{-1} \otimes B^{-1})r$. Direct inversion of the posterior precision matrix requires on the order of $(mn)^3$ floating point operations.

Let A and B have respective diagonalizations $A = C\Lambda C'$ and $B = EME'$. Computation of these diagonalizations requires on the order of $m^3 + n^3$ floating point operations. Posterior precision is

$$(C \otimes E)(\Lambda^{-1} \otimes M^{-1}+\tau^2 I_{mn})(C' \otimes E')$$

and its inverse,

$$(C \otimes E)(\Lambda^{-1} \otimes M^{-1}+\tau^2 I_{mn})^{-1}(C' \otimes E')$$

involves trivial computations.

References

Anderson, T. W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics* 29, 813–28.

Anderson, T. W. and H. Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 20: 46–63.

_____, 1950, The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 21: 570–82.

Bauwens, L. and M. Lubrano, 1993, Identification restrictions and posterior densities in cointegrated Gaussian VAR systems, CORE Discussion Paper 9418. Forthcoming in *Advances in econometrics,* vol. 11B, ed., T. M. Fomby and R. Carter Hill. JAI Press.

Chib, S., 1994, Marginal likelihood from the Gibbs output, John M. Olin School of Business Working Paper, Washington University.

Davies, P. T. and M. K. S. Tso, 1982, Procedures for reduced rank regression, *Applied Statistics* 31, 244–55.

Dreze, J. H., 1976, Bayesian limited information analysis of the simultaneous equation model, *Econometrica* 44: 1045–75.

Dreze, J. H. and J. F. Richard, 1983, Bayesian analysis of simultaneous equations systems, in: Z. Griliches and M. D. Intriligater, eds., *Handbook of Econometrics,* Vol. 1, Amsterdam: North-Holland Publishing Co.

Engle, R. F. and C. W. J. Granger, 1987, Co-integration and error correction: Representation, estimation and testing, *Econometrica* 55: 251–76.

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

Gelfand, A. and A. F. M. Smith, 1990, Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85, 389–409.

Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* 57: 1317–40.

_____, 1991, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in: J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian statistics,* Vol. 4, Oxford: Oxford University Press.

_____, 1993, Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics* 8: S91–40.

_____, 1994, Bayesian comparison of econometric models, Working Paper 532, Federal Reserve Bank of Minneapolis.

Gibbons, M. R., and W. Ferson, 1985, Testing asset pricing models with changing expectations and an unobservable market portfolio, *Journal of Financial Economics* 14: 217–36.

Hannan, E. J., 1970, *Multiple time series,* New York: Wiley.

Izenman, A. J., 1975, Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis* 5, 248–64.

Joreskog, K. G. and A. S. Goldberger, 1975, Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of the American Statistical Association* 70, 631–39.

Keilbergen, F. and H. K. van Dijk, 1991, Bayesian simultaneous equations model analysis: On the existence of structural posterior moments, Unpublished manuscript.

_____, 1993, On the shape of the likelihood-posterior in cointegration models, Unpublished manuscript.

Koopmans, T. C. (ed.), 1950, *Statistical inference in dynamic economic models,* Cowles Commission Monograph No. 10. New York: John Wiley.

Nummelin, E., 1984, *General irreducible Markov chains and non-negative operators,* Cambridge: Cambridge University Press.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1986, *Numerical recipes: The art of scientific computing,* Cambridge: Cambridge University Press.

Roberts, G. O. and A. F. M. Smith, 1992, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, Research report no. 92–30, University of Cambridge Statistical Laboratory.

Schwarz, G., 1978, Estimating the dimension of a model, *The Annals of Statistics* 6: 461–64.

Theil, H., 1971, *Principles of econometrics,* New York: Wiley.

Tierney, L., 1991, Exploring posterior distributions using Markov chains, in: *Computer science and statistics: 23rd symposium on the interface,* 563–70. Alexandria, VA: American Statistical Association.

_____, 1994, Markov chains for exploring posterior distributions, *Annals of Statistics,* forthcoming. (Also, Technical Report 560, University of Minnesota School of Statistics.)

Velu, R. P., 1991, Reduced rank models with two sets of regressors, *Applied Statistics* 40, 159–70.

Zellner, A., 1971, *An introduction to Bayesian inference in econometrics,* New York: Wiley.

Zellner, A., C. Min, and D. Dallaire, 1993, Bayesian analysis of simultaneous equation and related models using the Gibbs sampler and convergence checks, H. G. B. Alexander Research Foundation Working Paper, University of Chicago.

Zhou, G., 1993, Small sample rank tests with applications to asset pricing, Mimeo, John M. Olin School of Business, Washington University.

Table 1

Parameters for Generation of Artificial Data*

Coefficient Matrix Θ

1.00	.00	.00	2.00	−1.00	.00	.00	.00	.00	.00	1.00	−1.00
.00	1.00	.00	.00	.00	−3.00	2.00	.00	.00	.00	−1.00	3.00
.00	.00	1.00	.00	.00	.00	.00	3.00	−3.00	4.00	2.00	2.00
1.00	.00	.00	2.00	−1.00	.00	.00	.00	.00	.00	1.00	−1.00
.00	1.00	.00	.00	.00	−3.00	2.00	.00	.00	.00	−1.00	3.00
.00	1.00	.00	.00	.00	−3.00	2.00	.00	.00	.00	−1.00	3.00
.00	.00	1.00	.00	.00	.00	.00	3.00	−3.00	4.00	2.00	2.00

Disturbance Variance Matrix Σ

1.473											
−.346	.650										
.530	.173	1.197									
−.165	−.029	.036	1.027								
.363	.146	.564	−.004	1.730							
−.147	−.094	.185	.271	.471	1.114						
.011	.017	.088	−.094	−.331	−.409	.546					
−.673	.096	−.470	.137	−.007	−.131	.274	1.407				
−.015	.129	.414	−.036	−.059	−.535	.553	.332	1.027			
−.212	−.018	.110	.336	−.138	.127	.234	.118	.248	1.462		
−.159	−.087	−.174	.071	.659	.076	−.107	.300	.032	.010	.524	
−.077	−.001	.088	−.134	.233	.274	−.108	−.132	−.118	.043	.081	.130

Covariate Variance Matrix

.630					
.396	.839				
−.300	.281	1.535			
−.125	−.033	.264	.269		
.256	.606	−.161	−.456	1.572	
.217	.509	.666	.393	−.192	1.178

*The coefficient matrix was chosen directly. Note that the product of the first three columns and the first three rows determines the matrix. The disturbance variance matrix was generated by forming a 12 × 12 matrix A of independent $N(0,0.3^2)$ variates, and then taking $\Sigma = A'A$: thus $E(\sigma_{ii}) = 1.080$, and $E(\sigma_{ij}) = 0$, $[E(\sigma_{ij}^2)]^{1/2} = 0.312$. Similarly the covariate variance matrix V was generated by forming a 7 × 7 matrix B of independent N(0,0.42) variates, and taking V to be the last six rows and columns of B′B: thus $E(v_{ii}) = 1.120$, $E(v_{ij}) = 0$, and $[E(v_{ij}^2)]^{1/2} = 0.423$.

Table 2

Predictive Probabilities: Artifical Data

Observations		Normalization		Rank(q)						
Post.	Pred.	1	2	1	2	3	4	5	6	7
1–25	26–27	.508	.492	.000	.007	.083	.099	.239	.378	.194
1–25	26–30	.566	.434	.000	.000	.186	.300	.178	.285	.051
1–25	26–35	.520	.480	.000	.000	.386	.285	.095	.207	.027
1–25	26–48	.293	.707	.000	.000	.990	.005	.004	.000	.000
49–73	74–75	.519	.481	.000	.013	.392	.140	.226	.127	.102
49–73	74–78	.534	.466	.000	.000	.421	.187	.219	.085	.087
49–73	74–83	.379	.621	.000	.000	.712	.181	.092	.009	.005
49–73	74–96	.731	.269	.000	.000	.988	.011	.001	.000	.000
97–121	122–123	.592	.408	.000	.023	.302	.257	.156	.159	.103
97–121	122–126	.638	.362	.000	.000	.302	.201	.094	.281	.121
97–121	122–131	.602	.398	.000	.000	.580	.136	.069	.137	.078
97–121	122–144	.105	.895	.000	.000	.955	.037	.005	.003	.000
145–169	170–171	.527	.473	.000	.069	.481	.161	.086	.108	.094
145–169	170–174	.706	.294	.000	.001	.733	.065	.045	.080	.076
145–169	170–179	.578	.422	.000	.000	.765	.075	.020	.085	.055
145–169	170–192	.996	.004	.000	.000	.996	.002	.001	.000	.000
193–217	218–219	.403	.597	.000	.027	.386	.347	.110	.076	.054
193–217	218–222	.331	.669	.000	.001	.905	.070	.019	.004	.001
193–217	218–227	.499	.501	.000	.000	.715	.213	.064	.007	.001
193–217	218–240	.176	.824	.000	.000	.972	.027	.001	.000	.000

Table 3

Predictive Probabilities: Artificial Data

Observations		Normalization		Rank(q)						
Post.	Pred.	1	2	1	2	3	4	5	6	7
1–24	25–48	.285	.715	.000	.000	.991	.004	.004	.001	.000
1–24	25–96	.256	.744	.000	.000	.999	.001	.000	.000	.000
1–24	25–144	.403	.597	.000	.000	1.000	.000	.000	.000	.000
1–24	25–240	.260	.740	.000	.000	1.000	.000	.000	.000	.000
1–24	25–332	.183	.817	.000	.000	1.000	.000	.000	.000	.000
1–48	49–72	.783	.217	.000	.000	.966	.032	.000	.000	.000
1–48	49–120	.552	.448	.000	.000	.973	.023	.004	.000	.000
1–48	49–168	.510	.490	.000	.000	.985	.015	.001	.000	.000
1–48	49–268	.505	.495	.000	.000	.944	.039	.016	.000	.000
1–48	49–332	.426	.574	.000	.000	1.000	.000	.000	.000	.000
1–96	97–120	.568	.432	.000	.000	.739	.143	.100	.014	.004
1–96	97–168	.545	.455	.000	.000	.924	.061	.014	.001	.000
1–96	97–216	.617	.383	.000	.000	.634	.155	.208	.001	.001
1–96	97–332	.675	.325	.000	.000	.882	.118	.000	.000	.000
1–192	193–216	.511	.489	.000	.000	.080	.133	.170	.182	.434
1–192	193–264	.795	.205	.000	.000	.341	.285	.157	.054	.164
1–192	193–312	.719	.281	.000	.000	.128	.251	.449	.101	.071
1–192	193–332	.497	.503	.000	.000	.091	.182	.433	.120	.174
1–288	289–312	.472	.573	.000	.000	.229	.253	.217	.187	.114
1–288	289–332	.165	.835	.000	.000	.120	.138	.303	.282	.155

Table 4

Data Used in Example

Portfolio Number	Two-Digit SIC Codes	Industry Name
1	13, 29	Petroleum
2	60–69	Finance/Real Estate
3	25, 30, 36–37, 50, 55, 57	Consumer Durables
4	10, 12, 14, 24, 26, 28, 33	Basic Industries
5	1, 20, 21, 54	Food/Tobacco
6	15–17, 32, 52	Construction
7	34–35, 38	Capital Goods
8	40–42, 44, 45, 47	Transportation
9	46, 48, 49	Utilities
10	22–23, 31, 51, 53, 56, 59	Textiles/Trade
11	72–73, 75, 80, 82, 89	Services
12	27, 58, 70, 78–79	Leisure

Covariates

1. Value-weighted NYSE index return less 1–month Treasury bill return.
2. Monthly real per capita growth of personal consumption expenditures for nondurable goods, seasonally adjusted.
3. Monthly return of corporate bonds rated Baa by Moody's Investor Services less the long-term U.S. government bond return (CRSP).
4. Change in the difference between the average monthly yield of a 10–year Treasury bond and a 3–month Treasury bill.
5. Unexpected inflation rate: the difference between the actual and the forecasted inflation rate, formed from a time-series model for percentage changes in the CPI for all urban consumers, not seasonally adjusted.
6. 1–month Treasury bill return less the monthly rate of inflation, as measured by the CPI.

Table 5

Predictive Probabilities: Asset Pricing Model

Observations		Normalization		Rank(q)						
Post.	Pred.	1	2	1	2	3	4	5	6	7
5/59–4/61	5/61–4/62	.968	.032	.953	.047	.000	.000	.000	.000	.000
5/59–4/61	5/61–4/63	.994	.006	.983	.017	.000	.000	.000	.000	.000
5/61–4/63	5/63–4/64	1.000	.000	.929	.070	.001	.000	.000	.000	.000
5/61–4/63	5/63–4/65	1.000	.000	1.000	.000	.000	.000	.000	.000	.000
5/63–4/65	5/65–4/66	.832	.168	.968	.032	.000	.000	.000	.000	.000
5/63–4/65	5/65–4/67	.986	.014	.999	.001	.000	.000	.000	.000	.000
5/65–4/67	5/67–4/68	.656	.344	.985	.015	.000	.000	.000	.000	.000
5/65–4/67	5/67–4/69	.995	.005	1.000	.000	.000	.000	.000	.000	.000
5/67–4/69	5/69–4/70	1.000	.000	.094	.903	.002	.000	.000	.000	.000
5/67–4/69	5/69–4/71	1.000	.000	.998	.002	.000	.000	.000	.000	.000
5/69–4/71	5/71–4/72	.583	.417	.771	.226	.003	.000	.000	.000	.000
5/69–4/71	5/71–4/73	.760	.240	.331	.669	.000	.000	.000	.000	.000
5/71–4/73	5/73–4/74	1.000	.000	1.000	.000	.000	.000	.000	.000	.000
5/71–4/73	5/73–4/75	1.000	.000	1.000	.000	.000	.000	.000	.000	.000
5/73–4/75	5/75–4/76	.997	.003	.997	.003	.000	.000	.000	.000	.000
5/73–4/75	5/75–4/77	1.000	.000	.999	.000	.000	.000	.000	.000	.000
5/75–4/77	5/77–4/78	.958	.042	1.000	.000	.000	.000	.000	.000	.000
5/75–4/77	5/77–4/79	.993	.007	1.000	.000	.000	.000	.000	.000	.000
5/77–4/79	5/79–4/80	1.000	.000	.087	.909	.005	.000	.000	.000	.000
5/77–4/79	5/79–4/81	1.000	.000	.012	.576	.413	.000	.000	.000	.000
5/79–4/81	5/81–4/82	.988	.012	.991	.009	.000	.000	.000	.000	.000
5/79–4/81	5/81–4/83	1.000	.000	1.000	.000	.000	.000	.000	.000	.000
5/81–4/83	5/83–4/84	1.000	.000	1.000	.000	.000	.000	.000	.000	.000
5/81–4/83	5/83–4/85	1.000	.000	1.000	.000	.000	.000	.000	.000	.000

Table 6

Predictive Probabilities: Asset Pricing Model

Observations		Normalization		Rank(q)						
Post.	Pred.	1	2	1	2	3	4	5	6	7
5/59–4/61	5/61–4/65	.997	.003	.997	.003	.000	.000	.000	.000	.000
5/59–4/61	5/61–4/69	.996	.004	.998	.002	.000	.000	.000	.000	.000
5/59–4/61	5/61–4/73	.983	.017	1.000	.000	.000	.000	.000	.000	.000
5/59–4/61	5/61–4/81	.967	.033	.988	.012	.000	.000	.000	.000	.000
5/59–4/61	5/61–12/86	.921	.079	.929	.071	.000	.000	.000	.000	.000
5/59–4/63	5/63–4/67	.996	.004	.681	.277	.037	.000	.000	.000	.000
5/59–4/63	5/63–4/71	.589	.411	.990	.010	.000	.000	.000	.000	.000
5/59–4/63	5/63–4/75	.021	.979	.988	.012	.000	.000	.000	.000	.000
5/59–4/63	5/63–4/85	.966	.033	.444	.555	.000	.000	.000	.000	.000
5/59–4/63	5/63–12/86	.284	.716	.766	.234	.000	.000	.000	.000	.000
5/59–4/67	5/67–4/71	.007	.993	.999	.000	.000	.000	.000	.000	.000
5/59–4/67	5/67–4/75	.000	1.000	1.000	.000	.000	.000	.000	.000	.000
5/59–4/67	5/67–4/79	.000	1.000	1.000	.000	.000	.000	.000	.000	.000
5/59–4/67	5/67–12/86	.004	.996	.996	.004	.000	.000	.000	.000	.000
5/59–4/71	5/71–4/75	.366	.634	.638	.316	.029	.016	.000	.001	.000
5/59–4/71	5/71–4/79	.395	.605	.622	.368	.007	.002	.000	.000	.000
5/59–4/71	5/71–4/83	.980	.020	.054	.938	.004	.004	.000	.000	.000
5/59–4/71	5/71–12/86	.931	.069	.073	.909	.009	.009	.000	.000	.000
5/59–4/75	5/75–4/79	.833	.167	.716	.195	.039	.022	.024	.004	.001
5/59–4/75	5/75–4/83	.899	.101	.614	.311	.015	.022	.031	.001	.005
5/59–4/75	5/75–12/86	.746	.253	.105	.481	.081	.078	.209	.009	.038
5/59–4/79	5/79–4/83	.233	.767	.028	.132	.067	.154	.378	.038	.204
5/59–4/79	5/79–12/86	.198	.802	.001	.035	.083	.074	.589	.031	.186