EFFICIENT INFERENCE IN RANDOM
COEFFICIENT MODELS WITH MULTICOLLINEARITY
IN THE TIME SERIES REGRESSIONS

Working Paper No. 228

Robert K. Rayner
Roger L. Wright*

The University of Michigan

FOR DISCUSSION PURPOSES ONLY

Introduction

In applications involving pooled cross-sectional and time series data, one often has several variables describing individuals (units, firms) which exhibit cross-sectional variability, but little or no time series variability for each individual. This type of data presents a problem for random coefficient regression (RCR) pooling. The usual method of estimation for this model (Swamy, 1971) is a feasible Aitken procedure in which the first step is the estimation of a time series regression for each individual. If some explanatory variables are fixed or nearly fixed over time for one or more individuals, there are problems in implementing this first step.

This paper provides an extension of Swamy's RCR model to the case where there may be multicollinearity in some or all of the individual time series regressions. The first section of this paper describes an application in finance which motivates our interest in this problem. In Section Two, we describe a generalization of Swamy's two-stage estimator of the mean ($\overline{\beta}$) of the vector of coefficients conditional on known variance-covariance parameters. Section Three deals with the problem of predicting the coefficients of a particular individual, again assuming known variance-covariance parameters. In Section Four, estimation and prediction methods are developed assuming unknown variance-covariance parameters. Then we discuss some empirical results on the finance problem introduced in Section One. A brief summary of the paper is given in Section Six.

1. Estimating Cost of Equity Capital

Rate regulation for electric utilities depends heavily on estimates of

the cost of the firm's equity capital. The capital asset pricing model is

one of the many techniques that is used to provide these estimates.

According to capital asset pricing theory, the cost of equity capital

(in excess of the risk free rate) to firm i in period t is given by $E(y_{it})$

in the model

$$E(y_{it}) = \gamma_i E(R_t) \tag{1.1}$$

Here R_t is the excess return in period t of the market portfolio of all

stocks. Equation (1.1) states that systematic risk, γ_i, determines differ-

ential expected returns among securities and that there is a linear relation-

ship between γ_i and expected return, i.e., cost of capital.

In a rate case application, an estimate of γ_i may be obtained from a

time series regression of the form

$$y_i = R\gamma_i + \epsilon_i \tag{1.2}$$

where y_i and R are T x 1 vectors of the observed excess rates of return on

the ith stock and the market, respectively, γ_i is an unknown scalar

parameter, and ϵ_i is a T x 1 vector of disturbances. The estimate g_i of

γ_i is then used to help estimate cost of capital to firm i. However, the

question arises as to how well g_i predicts the firm's true γ_i, which is

what is needed in estimating cost of capital.

We might expect to get better estimates of cost of capital by "comparing"

relevant characteristics of the firm to those of other firms like it, i.e.,

other firms in the industry. As Myers (1978) has noted:

> The distinction between industry and firm systematic risk is important
> in rate cases. It is hard to estimate a regulated firm's cost of capital
> if data on only that firm are available. This is true regardless of the
> approach taken. It is necessary to broaden the sample.

Rosenberg and McKibbon (1973) have devised a pooling approach which is potentially useful for predicting γ_i for an electric utility. They estimate the systematic risk for firm i in period t using a model in which systematic risk is assumed to be linearly related to certain accounting variables,

$$\gamma_{it} = W_{it}\beta_i$$

In this paper we use the recent theoretical results of Bowman (1979) to help specify the model for systematic risk, and then use a modified Rosenberg-McKibbon (R-M) pooling approach to predict γ_i for electric utilities.

Bowman has shown that there is a theoretical relationship between a firm's systematic risk γ_i and the firm's debt ratio. We note that debt ratio is a variable which exhibits variability across firms, but little or no variability over time, for each firm. Thus, we investigate the model

$$\gamma_i = W_i\beta_i \tag{1.3}$$

in which γ_i is an unknown scalar parameter, W_i is a 1 x 2 vector (including the intercept) of observations assumed to be constant over time, and β_i is a 2 x 1 vector of unknown parameters.

Then, substitution of (1.3) into (1.2) results in

$$y_i = RW_i\beta_i + \epsilon_i \tag{1.4}$$

$$= X_i\beta_i + \epsilon_i$$

where y_i and R are T x 1 vectors of observations on each of the N firms and the market, respectively, and X_i is the T x 2 matrix defined by $X_i = RW_i$.

Note that for each i, (1.4) is a singular model, because X_i is of rank 1. Now, if we are willing to assume that $\beta_i = \beta$ (i = 1,2,...,N) multicollinearity may not be a problem in the pooling model. But a more reasonable assumption is that there are firm differences, and so we consider random coefficient regression pooling.

RCR is a parsimonious specification that allows for cross-sectional variation in model parameters. In RCR pooling, each coefficient vector is considered to be a random draw from the same multivariate normal distribution. This seems reasonable for a homogeneous industry like the electric utilities industry.

Random coefficient regression is a promising modeling approach for other reasons. It is well suited for drawing inferences from a sample to a population (see, for example, Dielman, Nantell and Wright, 1980). Also, the procedure is computationally efficient.

In RCR pooling it is assumed that the coefficient vector β_i is fixed over time. Although not always so, this may be a reasonable assumption in this application, at least over the relatively short time period of the last four years. For a portfolio of electric utilities, Warga (1980) has found that there are dramatic fluctuations in model parameters around the period of the 1973 oil embargo, but that the parameters are relatively stable after 1975. Using monthly data from 1976-1979, we have a reasonable number of observations for the individual time series regressions.

Overall, random coefficient regression seems to be a good choice of methodology for the cost-of-capital problem, and so it is the approach that is used here. Coefficients in (1.4) are assumed to be fixed over time for each firm, but they are allowed to vary randomly across firms. Specifically, we model cost of capital by

$$y_i = (RW_i)\beta_i + \varepsilon_i \quad (i = 1,2,\ldots,N) \tag{1.5}$$

$$= X_i\beta_i + \varepsilon_i$$

where $\beta_i \sim N_2(\bar{\beta}, \Delta)$, $X_i = RW_i$, and $\varepsilon_i \sim N_T(0, \sigma_{ii}I)$.

Because X_i is not of full rank, Swamy's (1971) feasible Aitken procedure

for estimating $\bar{\beta}$ cannot be used directly. The first step in that method is

the estimation of individual time series regressions. In the next section

we show how to estimate $\bar{\beta}$, conditional on known variance-covariance param-

eters, when there is multicollinearity in the time series regressions.

2. Estimation With Known Variance-Covariance Parameters

This section introduces the random coefficient regression (RCR) pooling model which is useful in many applications including our financial example. Our version of the RCR model is similar to the pooling models studied extensively by Swamy and others, but features singularity (i.e., multicollinearity) in the time series datasets describing some or all of the individuals (e.g., securities). In this section we will describe an extension of Swamy's two-stage procedure for computing Aitken's generalized least squares estimators in the presence of this multicollinearity. Throughout this section, all variance-covariance parameters are assumed to be known.

Consider the following model:

$$y_i = X_i \beta_i + \varepsilon_i \quad (i = 1, 2, \ldots, N) \tag{2.1}$$

where y_i and ε_i are $T \times 1$ vectors, X_i is a $T \times k$ matrix, and β_i is a $k \times 1$ vector. The vector y_i and the matrix X_i contain observed variables characterizing individual i, while β_i and ε_i are unobserved random vectors.

We assume that

(a) Rank $(X_i) = r_i \leq k$, but rank $(\underline{X}) = k$ where $\underline{X}' = [X_1' \ X_2' \ \ldots \ X_N']$ is the $k \times NT$ matrix comprised of the X_i,

(b) β_i is $N_k(\overline{\beta}, \Delta)$ where Δ is positive semidefinite,

(c) ε_i is $N_T(0, \sigma_{ii}I)$ with $\sigma_{ii} > 0$, and

(d) $\beta_1, \ldots, \beta_N, \varepsilon_1, \ldots, \varepsilon_N$ are mutually independent.

In addition, throughout this section and the next, it is assumed that σ_{ii} and Δ are known.

Equation (2.1) can be rewritten as

$$y_i = X_i \bar{\beta} + u_i \quad \text{where} \tag{2.2}$$

$$u_i = X_i(\beta_i - \bar{\beta}) + \varepsilon_i.$$

Here u_i is $N_T(0, \Omega_i)$ with known variance-covariance matrix

$$\Omega_i = X_i \Delta X_i' + \sigma_{ii} I,$$

and $u_1 \ldots u_N$ are mutually independent. Since $\underline{X}$ is of full rank, the best

linear unbiased estimator of $\bar{\beta}$ is uniquely determined by Aitken's generalized

least squares estimator:

$$\bar{b} = (\sum_{i=1}^{N} X_i' \Omega_i^{-1} X_i)^{-1} \sum_{i=1}^{N} X_i' \Omega_i^{-1} y_i. \tag{2.3}$$

Moreover, the variance-covariance matrix of the estimator $\bar{b}$, denoted here as

$\bar{C}$, is

$$\bar{C} = (\sum_{i=1}^{N} X_i' \Omega_i^{-1} X_i)^{-1}.$$

We now turn to the case that each X_i is of full rank, i.e., $r_i = k$.

This situation has been extensively studied by Swamy (1971). Swamy shows that

$\bar{b}$ can be efficiently computed following a two-stage procedure. In the first

stage of this procedure, ordinary regression statistics are calculated which

summarize the time series observations of each individual. In the second

stage, these N sets of time series regression statistics are pooled to estimate

the population parameter $\bar{\beta}$.

Specifically, in the first-stage time series analysis of each individual

i, the following statistics are calculated:

$$b_i = (X_i'X_i)^{-1}X_i'y_i \quad \text{and}$$

$$C_i = \Delta + \sigma_{ii}(X_i'X_i)^{-1}.$$

Here b_i is the vector of ordinary regression coefficients summarizing the

time series data describing individual i, and C_i is the variance-covariance

matrix of b_i taking into account both the variance of β_i around $\bar{\beta}$ and

also the conditional variance of b_i as an estimator of β_i:

$$C_i = E(b_i - \bar{\beta})(b_i - \bar{\beta})'$$

$$= E(\beta_i - \bar{\beta})(\beta_i - \bar{\beta})' + E[(b_i - \beta_i)(b_i - \beta_i)' \mid \beta_i].$$

Swamy's second-stage, cross-sectional analysis follows from the GLS represen-

tation of $\bar{\beta}$ (2.3) together with the identities

$$X_i'\Omega_i^{-1}X_i = C_i^{-1} \quad \text{and}$$

$$X_i'\Omega_i^{-1}y_i = C_i^{-1}b_i.$$

In fact,

$$\bar{b} = (\sum_{i=1}^{N} C_i^{-1})^{-1} \sum_{i=1}^{N} C_i^{-1}b_i \quad \text{and} \qquad\qquad (2.4)$$

$$\bar{C} = (\sum_{i=1}^{N} C_i^{-1})^{-1}.$$

So in the second stage, $\bar{b}$ is calculated as a weighted average of the b_i.

The computational advantage of (2.4) over (2.3) is substantial since (2.3)

involves the inverse of each Ω_i which is a T x T matrix, while (2.4) in-

volves the inverse of matrices of size k x k.

In many applications, including the finance example described in Section

One, Swamy's two-stage procedure cannot be directly applied since the X_i are

of less than full rank (i.e., multicollinear) so that the b_i are undefined.

The remainder of this section describes a generalization of Swamy's

two-stage procedure which preserves its intuitive appeal and computational

advantages and is applicable to the multicollinear case in which some of the

r_i are less than k.

This generalization utilizes the rank factorization (Rao, 1973, p. 19)

of each X_i with rank r_i less than k. Using the rank factorization, we

compute matrices R_i and W_i such that $X_i = R_iW_i$, where R_i is of size

$T \times r_i$ and is of full rank r_i, and W_i is of size $r_i \times k$ and is also of

rank r_i. (If X_i is of full rank k, we take $R_i = X_i$ and $W_i = I$.)

Then we make the following definitions:

$$\gamma_i = W_i \beta_i, \tag{2.5}$$

$$g_i = (R_i'R_i)^{-1}R_i'y_i, \quad \text{and}$$

$$H_i = W_i \Delta W_i' + \sigma_{ii}(R_i'R_i)^{-1}.$$

To motivate these definitions, we note that the model (2.1) can be rewritten as

$$y_i = R_i \gamma_i + \varepsilon_i$$

and g_i is the ordinary least squares, time series estimator of γ_i with

variance-covariance matrix H_i if the variability of γ_i is taken into

account.

By employing these definitions and a matrix identity of Rao (1973, exer-

cise 2.9, p. 33), we obtain the identities

$$X_i'\Omega_i^{-1}X_i = W_i'H_i^{-1}W_i \quad \text{and} \tag{2.6}$$

$$X_i'\Omega_i^{-1}y_i = W_i'H_i^{-1}g_i.$$

These identities together with (2.3) give

$$\bar{b} = (\sum_{i=1}^{N} W_i'H_i^{-1}W_i)^{-1} \sum_{i=1}^{N} W_i'H_i^{-1}g_i, \tag{2.7}$$

$$\bar{C} = (\sum_{i=1}^{N} W_i'H_i^{-1}W_i)^{-1}.$$

So Swamy's second-stage weighted average (2.4), which is only applicable

if each X_i is of full rank, generalizes to (2.7) which represents the best

linear unbiased estimator $\bar{b}$ as a cross-sectional, generalized MANOVA estimator

in which the g_i, computed from the time series regressions, play the role of

a vector of observed dependent variables, the W_i are matrices of independent

variables, and the H_i represent the variance-covariance matrices of the

observations g_i. The form of (2.7) is consistent with the fact that each g_i is $N_{r_i}(W_i\bar{\beta}, H_i)$ and $g_1, \ldots, g_N$ are mutually independent.

The rank factorization of each X_i plays an essential role in this representation of $\bar{b}$. The first factor R_i, is used only in the ith first-stage time series regression, i.e., in the computation of g_i and H_i. These statistics from the N time series regressions are used in the second-stage pooling in conjunction with the set of second factors W_i which determine the relationships between the observed g_i and the underlying β_i. Hence the two factors of X_i are related respectively to the two stages of analysis; the R_i comprise the time series observations on each individual, and the set of W_i, together with the time series regression statistics g_i and H_i, comprise the cross-sectional information characterizing the sample of individuals.

In concluding this section we note that in many applications involving cross-sectional time series datasets, including our finance example, it is natural to formulate a model which has two components corresponding to the time series and cross-sectional stages of analysis (see, for example, Amemiya (1978) and Hanushek (1974)). In these cases, the rank factorization may arise naturally. In the situations that we have in mind, a time series model, say

$$y_i = R_i \gamma_i + \varepsilon_i,$$

is specified to represent the process generating the time series observations of each individual i (i = 1,2,...,N). Here each R_i is T x r_i and of full rank r_i, and γ_i is a r_i x 1 vector of time-invariant, unobserved characteristics of individual i. A second cross-sectional model,

$$\gamma_i = W_i \beta_i,$$

is specified which expresses the unobserved γ_i as an observed linear

transformation W_i of a vector β_i (k x 1) of unobserved, time-invariant characteristics of the individual. These vector β_i are assumed to comprise a random sample from a multivariate normal population with unknown mean $\bar{\beta}$ and known variance-covariance matrix Δ.

If the cross-sectional model is imbedded into the time series model, we obtain the RCR model (2.1) where each X_i is singular, having rank decomposition $X_i = R_i W_i$.

3. Predicting With Known Variance-Covariance Parameters

In many applications, there is a need to predict the value of β_i (a
k x 1 vector) for a specific individual i. For example, in our finance
application, interest is in predicting the β_i which represents the ith
security's characteristics that determine its systematic risk in conjunction
with its observed characteristics W_i.

In this section we consider the problem of efficient prediction of β_i.
We assume that the RCR model (2.1) is applicable to an available pooled cross-
sectional time series database describing a sample of N individuals, and that
all variance-covariance parameters are known. The case that each X_i is of
full rank is discussed first, followed by the case where some or all of the
X_i may be less than full rank.

A natural predictor of β_i when X_i is of full rank is the OLS pre-
dictor, $b_i = (X_i'X_i)^{-1}X_i'y_i$, based on the ith individual. b_i is
unbiased, and of all predictors that are linear in y_i, it is the most effi-
cient. However, as Kadiyala and Oberhelman (1979) note, b_i uses information
only on the ith individual and ignores information about all of the other
individuals. Therefore, they consider predictors of β_i which are linear
in $\underline{y}$, i.e., of the form $A_i\underline{y}$ where A_i is a (k x NT) matrix of constants.
They show that the unbiased, linear predictor of β_i with the lowest mean
squared error is given by

$$b^c_i = \bar{b} + \Delta X_i'\Omega_i^{-1}y_i - \Delta X_i'\Omega_i^{-1}X_i\bar{b} \tag{3.1}$$

which simplifies to

$$b^c_i = \Delta(\Delta + \sigma_{ii}(X_i'X_i)^{-1})^{-1}b_i + (\sigma_{ii}(X_i'X_i)^{-1})(\Delta + \sigma_{ii}(X_i'X_i)^{-1})^{-1}\bar{b} \tag{3.2}$$

As Kadiyala and Oberhelman have observed, equation (3.2) has an intuitive
interpretation. The predictor is simply a weighted average of the estimator

b_i on the ith individual, and the estimator $\bar{b}$. In this formula,

$\sigma_{ii}(X_i'X_i)^{-1}$ is the variance-covariance matrix of b_i, and Δ is the

variance-covariance matrix of the β_i's. The larger the variance of the OLS

estimator b_i, the greater the weight given to $\bar{b}$, the estimator of the mean

vector; and the greater the dispersion of the β_i, the greater the weight

given to the estimator b_i based on observations on the ith individual.

The extension of equation (3.1) to the case that rank $(X_i) \leq k$ is

straightforward. In the Kadiyala-Oberhelman derivation of the "best" pre-

dictor of β_i, it is sufficient that $\underline{X}$ be of full rank, so (3.1) remains

valid with our assumptions. Using (2.6), we rewrite (3.1) to obtain the

more computationally efficient formula

$$b^c_i = \bar{b} + \Delta W_i'(W_i \Delta W_i' + \sigma_{ii}(R_i'R_i)^{-1})^{-1}(g_i - W_i\bar{b}) \qquad (3.3)$$

Equation (3.3) provides the predictor of β_i that has the smallest mean

squared error (in the class of linear, unbiased predictors), when rank

$(X_i) \leq k$.

Equation (3.3) appears at first not to have an easy interpretation. How-

ever, we note that it is, in fact, closely related to an intuitively appealing

estimator which we derive now.

Although it has been observed previously that β_i may not be estimable

under generalized rank conditions on X_i, it is possible to get a g-inverse

solution for β_i in the equation

$$W_i \beta_i = g_i \qquad (3.4)$$

Given an estimate $\bar{b}$ of $\bar{\beta}$, an intuitively appealing estimator of β_i for the

i^{th} individual is the minimum norm g-inverse solution of (3.4). (For a

suitably chosen positive definite matrix Q, the norm of a vector α is defined

by $||\alpha|| = (\alpha'Q\alpha)^{1/2}$). Rao (1973, p. 48) shows that the minimum norm

g-inverse, with respect to the 0 vector, is given by

$$Q^{-1}W_i'(W_iQ^{-1}W_i')^-,$$

where "−" denotes the g-inverse. Thus, after transforming (3.4) to an origin at $\bar{b}$,

$$W_i(\beta_i - \bar{b}) = g_i - W_i\bar{b},$$

the minimum norm solution to (3.4) is given by

$$\bar{b} + Q^{-1}W_i'(W_iQ^{-1}W_i')^-(g_i - W_i\bar{b})$$

and choosing $Q = \Delta^{-1}$, we get

$$\bar{b} + \Delta W_i'(W_i\Delta W_i')^-(g_i - W_i\bar{b}) \tag{3.5}$$

The expression in (3.5) is very similar to the predictor in (3.3). However, (3.5) does not allow for variability in b_i, conditional on known β_i.

In some applications, it is of interest to predict $\gamma_i = W_i\beta$. For example, in our finance problem, we are interested in systematic risk γ_i. It is easy to show that the best predictor of $W_i\beta_i$ is $W_ib^c_i$, and from (3.3)

$$W_ib^c_i = W_i\bar{b} + W_i\Delta W_i'\ (W_i\Delta W_i + \sigma_{ii}(R_i'R_i)^{-1})^{-1}(g_i - W_i\bar{b})$$

which can be written as

$$W_ib^c_i = (W_i\Delta W_i')(W_i\Delta W_i' + \sigma_{ii}(R_i'R_i)^{-1})^{-1}g_i$$

$$+ \sigma_{ii}(R_i'R_i)^{-1}(W_i\Delta W_i' + \sigma_{ii}(R_i'R_i)^{-1})^{-1}W_i\bar{b} \tag{3.6}$$

Equation (3.6) admits to the same kind of interpretation as did (3.2).

In this section we have derived efficient predictors for β_i and γ_i $= W_i\beta_i$ for the case that rank $(X_i) \leq k$. As in the previous section, results were obtained assuming known variance-covariance parameters in model (2.1). In practice, the variance-covariance parameters must be estimated, and so it is relevant to ask how prediction is affected. Kadiyala and Oberhelman cite some Monte Carlo evidence that b^c_i, computed from (3.3) with estimated parameters, is particularly warranted when N is large and T is small.

4. Estimation with Unknown Variance-Covariance Parameters

In this section we consider estimation in the RCR model with unknown variance-covariance parameters and multicollinearity in some or all of the individual time series regressions. A computationally efficient form for the likelihood function of the parameters in the case rank $(X_i) \leq k$ is derived first. Then, because it is not clear whether the parameters are identified, we look at identification in the RCR model in the presence of collinearity. The information matrix for the parameters is derived, and two conditions are obtained which, in conjunction with assumptions (a) thru (d) in Section One, are sufficient for the information matrix to be of full rank, thus establishing local identifiability (see Rothenberg (1971)). Then Fisher's method of scoring (Rao, 1973, pp. 366-374), a procedure for maximum likelihood estimation, is proposed for estimating the Δ matrix. We conclude this section with an iterative estimation procedure based on equation (3.3).

Introducing some additional notation we write the system of equations (2.2) as

$$\underline{y} = \overline{\underline{X}\beta} + \underline{u}$$

with $\underline{y}' = (y_1' \ \dots \ y_N')$ and $\underline{u}' = (u_1' \ \dots \ u_N')$ and $E(\underline{u}\,\underline{u}') = \Omega$. The vectors $u_1, \dots, u_N$ are mutually independent, so that $\Omega = \text{diag}(\Omega_i)$. We assume that Δ depends on a finite number of unknown parameters $\delta' = (\delta_1 \ \dots \ \delta_M)$. Let $\Theta' = (\delta'\sigma')$, where $\sigma' = (\sigma_{ii} \ \dots \ \sigma_{NN})$.

Then by our assumptions, the likelihood function for the parameters $\overline{\beta}$ and Θ is given by

$$L(\overline{\beta}, \Theta | \underline{y}, \underline{X}) = (2\pi)^{-NT/2} \ \prod_{i=1}^{N} |X_i \Delta X_i' + \sigma_{ii} I_T|^{-1/2}$$

$$\cdot \exp\left\{-1/2 \sum_{i=1}^{N} (y_i - X_i\overline{\beta})'(X_i \Delta X' + \sigma_{ii} I_T)^{-1}(Y_i - \overline{X}_i \beta)\right\} \qquad (4.1)$$

Using the rank factorization of the X_i, the definitions (2.5) and again the

matrix identity in Rao, equation (4.1) can be written as

$$L(\bar{\beta}, \Theta | \underline{y}, \underline{X}) = (2\pi)^{-NT/2} \prod_{i=1}^{N} \{\sigma_{ii}^{-(T-r_i)/2} |R_i'R_i|^{-1/2} |H_i|^{-1/2}\}$$

$$\cdot \exp\{-1/2 \sum_{i=1}^{N} \frac{(T-r_i)}{\sigma_{ii}} s_{ii} + (g_i - W_i\bar{\beta})' H_i^{-1}(g_i - W_i\bar{\beta})\} \quad (4.2)$$

where s_{ii} is the OLS estimator of σ_{ii} from the ith time series equation.

Equation (4.2) provides substantial computational advantages over (4.1),

because the matrices in (4.2) are on the order of r_i x r_i, whereas those in

(4.1) are T x T. (cf. Swamy, 1971, pp. 111-112).

We now look at identification in the RCR model with generalized rank

assumptions. It is clear that if rank $(X_i) < k$, then β_i from the ith time

series regression is unidentified, but what of the parameters $\bar{\beta}$ and Θ? It

is not clear from (4.2) whether they are identified or not.

We use the information matrix to examine parameter identifiability. Ac-

cording to Rothenberg, a necessary and sufficient condition for the $\bar{\beta}$ and Θ

parameters to be locally identified is that the information matrix be nonsingu-

lar. Magnus (1978) has shown that the information matrix is given by

$$\begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & 1/2(\frac{\partial vec\Omega^{-1}}{\partial\Theta})' (\Omega \otimes \Omega)(\frac{\partial vec\Omega^{-1}}{\partial\Theta}) \end{bmatrix} \quad (4.3)$$

In (4.3), the notation vec Ω^{-1} means the $(NT)^2$ x 1 vector obtained

from Ω^{-1} by taking the first column of Ω^{-1}, then the 2nd and so forth;

and $\otimes$ is the Knonecker product operator. Also, we use the definition of matrix

derivative as Dhrymes (1978), so that $\frac{\partial vec\Omega^{-1}}{\partial\Theta}$ is an $(NT)^2$ by (M + N) matrix

of first partial derivatives.

By assumption, $\underline{X}'\Omega^{-1}\underline{X}$ is of full rank, and hence $\bar{\beta}$ is locally identified.

A necessary and sufficient condition for Θ to be locally identified is that the matrix $1/2(\frac{\partial \text{vec}\Omega^{-1}}{\partial\Theta})'(\Omega \otimes \Omega)(\frac{\partial \text{vec}\Omega^{-1}}{\partial\Theta})$ be nonsingular. Now Ω is nonsingular, and so $\Omega \otimes \Omega$ is nonsingular. Thus, we need to establish conditions for which $\frac{\partial \text{vec}\Omega^{-1}}{\partial\Theta}$ (or equivalently $\frac{\partial \text{vec}\Omega}{\partial\Theta}$) is of full rank $(M + N)$.

We approach the problem by deriving a convenient expression for $\frac{\partial \text{vec}\Omega}{\partial\Theta}$, using the matrix formulas in Dhrymes and again the rank factorization idea and Rao's matrix identity. The Ω matrix is block diagonal, so that

$$\text{rank } (\frac{\partial \text{vec}\Omega}{\partial\Theta}) = \text{rank } [(\frac{\partial \text{vec}\Omega_1}{\partial\Theta})' \cdots (\frac{\partial \text{vec}\Omega_N}{\partial\Theta})']'. \tag{4.4}$$

Also, it can be shown that

$$\frac{\partial \text{vec}\Omega_1}{\partial\Theta} = [(R_i \otimes R_i)(W_i \otimes W_i) \frac{\partial \text{vec}\Delta}{\partial\delta} \vdots \frac{\partial \text{vec}\sigma_{ii}I}{\partial\sigma}]. \tag{4.5}$$

Then, after substituting (4.5) into the matrix

$$V = [(\frac{\partial \text{vec}\Omega_1}{\partial\Theta})' \cdots (\frac{\partial \text{vec}\Omega_N}{\partial\Theta})']'$$

of (4.4), it follows that

$$\text{rank } (\frac{\partial \text{vec}\Omega}{\partial\delta}) = \text{rank } (V) = \text{rank } (V'V). \tag{4.6}$$

and so we need conditions for which the matrix $V'V$ is of full rank. It can be shown that sufficient for $V'V$ to be of full rank, and hence Θ identified, are:

(1) The matrix $W(\frac{\partial \text{vec}\Delta}{\partial\delta})$ is of rank M, where

   $W = (W_i \otimes W_i' \cdots W_N' \otimes W_N')'$, and

(2) $T > r_i$, for $i = 1, 2, \ldots, N$.

Assumptions (a) thru (d) with conditions (1) and (2) above are sufficient for identifiability of all of the RCR parameters.

Now consider estimation of $(\bar{\beta}, \delta, \sigma)$. It does not seem to be computationally feasible to compute the maximum likelihood estimator of σ since matrices of order N^2 are encountered. However, using rank factorization, an ordinary regression estimator s_{ii} is easily calculated and is an unbiased, consistent $(T \to \infty)$ estimator of σ_{ii} (Goldberger, 1964, pp. 269-272). Conditional on these estimates of σ_{ii}, maximum likelihood estimation of $\bar{\beta}$ and δ is computationally feasible using Fisher's method of scoring (Rao, 1973, pp. 366-374). We use scoring on δ combined with (2.7) for estimation of $\bar{\beta}$.

To implement this procedure, it is necesary to derive for δ the first-order conditions and information matrix. Starting from equation (13) in Magnus, we can show that the first-order condition on δ is given by

$$-1/2 \sum_{i=1}^{N} (\frac{\partial \text{vec}\,\Delta}{\partial \delta})'(W_i \otimes W_i)'\text{vec } H_i^{-1}$$

$$+1/2 \sum_{i=1}^{N} (\frac{\partial \text{vec}\,\Delta}{\partial \delta})'(W_i \otimes W_i)'\text{vec } (H_i^{-1}(g_i - W_i\bar{\beta})(g_i - W_i\bar{\beta})'H_i^{-1}). \quad (4.7)$$

Similarly, the information matrix is given by

$$1/2 \sum_{i=1}^{N} (\frac{\partial \text{vec}\,\Delta}{\partial \delta})'(W_i \otimes W_i)'(H_i^{-1} \otimes H_i^{-1})(W_i \otimes W_i)(\frac{\partial \text{vec}\,\Delta}{\partial \delta}). \quad (4.8)$$

Unfortunately, there is no assurance that a scoring procedure using (4.7)-(4.8) will give positive semidefinite estimates of Δ. This problem of "negative variance components" is one that has long been troublesome for random coefficient regression and other variance components models (see, for example, Swamy, 1971, pp. 107-111). A potential remedy for this problem is suggested

by the papers of Box (1966) and Dent-Hildreth (1977). Box shows that in certain specific cases, a problem of constrained maximization of the likelihood function can sometimes be converted to one involving unconstrained optimization without introducing additional local optima. We are investigating the transformation $\Delta = TT'$, where T is lower triangular, as a method of obtaining a positive semidefinite estimate of the Δ matrix. Unconstrained optimization is used, with the search being conducted over the elements of the T matrix rather than over the elements of Δ.

Equations (4.7) and (4.8) need to be modified to take advantage of the $\Delta = TT'$ transformation. Letting t be the M x 1 vector of elements of T where $\Delta = TT'$, it can be shown that

$$\frac{\partial \mathrm{vec}\,\Delta}{\partial \delta} = \frac{\partial \mathrm{vec}TT'}{\partial t} = (T \otimes I) \frac{\partial \mathrm{vec}T}{\partial t} + (I \otimes T) \frac{\partial \mathrm{vec}T'}{\partial t} , \text{ so that}$$

$$(W_i \otimes W_i) \frac{\partial \mathrm{vec}\,\Delta}{\partial \delta} = (W_i T \otimes W_i) \frac{\partial \mathrm{vec}T}{\partial t} + (W_i \otimes W_i T) \frac{\partial \mathrm{vec}\ T'}{\delta t} \qquad (4.9)$$

While equations (4.7) and (4.8), in conjunction with (4.9), may appear to be formidable at first glance, our initial investigations indicate that the formulas are actually computationally efficient and relatively easy to program. The matrix $\frac{\partial \mathrm{vec}T}{\partial t}$, for example, is k^2 x M, but each column of this matrix contains only one 1, all other elements in the column being 0. Computational advantage can be had using this fact. And although we went to some trouble to obtain formulas (4.7)-(4.9), computational advantage may be obtained because the first-order conditions and the information matrix are given explicitly, rather than their having to be computed numerically. Also, in practice it may not be necessary to compute and invert the information matrix at each iteration (Rao, 1973, p. 370).

By way of summarizing this part of the paper, we note that even though the transformation $\Delta = TT'$ is not one to one, our limited investigations (at this time) show that this lack of uniqueness appears not to cause convergence problems, and that this transformation, with the method of scoring, may provide computationally efficient, positive semidefinite estimates of the Δ matrix of the RCR model. In Section Five we present some empirical results which were obtained using formulas (4.1) thru (4.9).

At this point we briefly discuss another potential estimation method for $(\bar{\beta}, \Delta, \sigma)$. The equation (3.3) suggests an iterative procedure of estimating Δ by

$$\sum_{i=1}^{N} (b_i{}^c - \bar{b})(b_i{}^c - \bar{b})'/N.$$

The procedure may at least be useful in providing a starting value for a maximum likelihood search.

5. An Application: Predicting Systematic Risk for Electric Utilities

In this section we return to the financial application of Section One. The model we use here is oversimplified and is not intended as a meaningful test of financial hypotheses. Its purpose is to illustrate the procedures developed in the previous sections. We are pursuing a more comprehensive analysis of systematic risk in financial markets. However this simple illustration suggests that pooling is a promising approach.

A sample was taken of 64 electric utilities that have data available on both the CRSP and COMPUSTAT data tapes over the period January 1976 thru December 1979. Estimation was done using the 24 monthly observations from 1976-77; forecasting was done over the period 1978-79.

Debt ratio is a variable that is available from annual data. Because it typically changes very little over time periods as short as several years, we simply calculated W_i as the average of the two annual values 1976-77, for each electric utility in the sample.

The procedures developed in the previous sections are applicable because all of the X_i matrices are singular. The σ_{ii} were estimated from the individual time series regressions. The identity matrix was used as an initial estimate of the Δ matrix, and $\bar{b}$ was calculated from the GLS formula (2.7). Scoring formulas (4.7)-(4.9) were then used to provide a new estimate of $\bar{\beta}$, and an iterative procedure was continued until stable values to four decimal places were obtained for the estimates of $\bar{\beta}$ and Δ. The results are reported below:

$\bar{b}$ = (0.6455, -0.0928)', with t-statistics of 1.64 and -0.10 respectively; and $(\delta_{11}, \delta_{21}, \delta_{22})$ = (0.3569, -0.0640, 0.1417).

Neither $\overline{b}$ coefficient is significant at the .05 level of significance. Thus, we are not able to conclude here that debt ratio is a determinant of systematic risk γ_i, as the theoretical finance literature suggests. A relationship may still exist, however, because there is evidence of multicollinearity in our model. The condition number, the ratio of the largest singular value of the variance-covariance matrix of $\overline{b}$ to the smallest, is 40.9, and multicollinearty may cause problems in parameter estimation when the condition number is greater than 10-30, see Belsley, Kuh and Welsch (1980). It appears that there may not be enough cross-sectional variability of debt ratio in our sample.

The main concern of this application is prediction of systematic risk γ_i. We did a mean squared error study, using the following forecasting methods:

1. A firm-specific method in which the estimates g_i of γ_i over the period 1976-77 were used to predict g_i calculated over the forecast period 1978-79.

2. A pooling method in which the forecasting formula (3.4), evaluated over the period 1976-77, was used to predict g_i calculated from the period 1978-79.

The results are reported below:

Method 1, Root MSE: 0.2636,

Method 2, Root MSE: 0.1857.

These results are encouraging. In this application, the pooling approach has a root MSE that is 30% less than the firm-specific method which is the approach usually taken.

6. Summary

This paper is concerned with efficient inference in Swamy's random coefficient regression (RCR) model when there is singularity, i.e., multicollinearity in some or all of the individual time series regressions. In particular, efficient estimation of the parameters of the RCR model is considered, as is efficient prediction of the response β_i of the ith individual. Methods are developed first under the assumption of known variance-covariance parameters, and then with this assumption relaxed.

We derive sufficient conditions for identifiability of all of the parameters of the RCR model. Then Fisher's method of scoring is developed for estimation of the Δ matrix parameters. A transformation, suggested by the papers of Box and Dent-Hildreth, is incorporated into the scoring procedure. Our initial investigations suggest that scoring with this transformation produces computationally efficient, positive semidefinite estimates of the Δ matrix.

The problem of multicollinearity in the time series regressions is motivated by an application in finance, namely, estimating cost of capital to utilities. The model is introduced in Section One and some preliminary empirical results are given in Section Five. These results are encouraging; in this particular application, using pooling methods there was a 30% reduction in root mean squared forecast error compared to forecasts from the usual forecasting technique.

References

1. Amemiya, T. (1978), "A Note on a Random Coefficients Model," International Economic Review. 19, 793-796.

2. Belsley, D. A., Kuh, E. and Welsch, R. E. (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York: John Wiley & Sons.

3. Bowman, R. G. (1979), "The Theoretical Relationship Between Systematic Risk and Financial (Accounting) Variables," Journal of Finance, 34, 617-630.

4. Box, M. J. (1966), "A Comparison of Several Current Optimization Methods, and the Use of Transformations in Constrained Problems," Computer Journal, 9, 67-77.

5. Dent, W. T. and Hildreth, C. (1977), "Maximum Likelihood Estimation in Random Coefficient Models," Journal of the American Statistical Association, 72, 69-72.

6. Dielman, T., Nantell, T. J. and Wright, R. L. (1980), "Price Effects of Stock Repurchasing: A Random Coefficient Regression Approach," Journal of Financial and Quantitative Analysis, 15, 171-189.

7. Dhrymes, P. J. (1978), Mathematics for Econometrics, New York: Springer-Verlag.

8. Goldberger, A. S. (1964), Econometric Theory, New York: John Wiley & Sons.

9. Hanushek, E. A. (1974), "Efficient Estimators for Regression Regression Coefficients," The American Statistician, 28, 66-67.

10. Kadiyala, K. R. and Oberhelman, D. (1979), "Response Predictions in Regressions on Panel Data," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 92-97.

11. Magnus, J. R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," Journal of Econometrics, 7, 281-312.

12. Myers, S. C. (1978), "On the Use of Modern Portfolio Theory in Public Utility Cases: Comment," Financial Management Association, 7, Number 3, 66-76.

13. Rao, C. R. (1973), Linear Statistical Inference and Its Applications, Second Edition, New York, John Wiley & Sons.

14. Rosenberg, B. and McKibbon, W. (1973), "The Prediction of Systematic and Specific Risk in Common Stocks." Journal of Financial and Quantitative Analysis, 8, 317-334.

15. Rothenberg, T. J. (1971), "Identification in Parametric Models." Econometrica, 39, 577-591.

16. Swamy, P. A. V. B. (1971), Statistical Inference in Random Coefficient Regression Models, Berlin: Springer-Verlag.

17. Swamy, P. A. V. B. (1973), "Criteria, Constraints and Multicollinearity in Random Coefficient Regression Models," Annals of Economics and Social Measurement, 24, 429-450.

18. Tracy, D. S. and Dwyer, P. S. (1969), "Multivariate Maxima and Minima With Matrix Derivatives," Journal of the American Statistical Association, 64, 1576-1594.

19. Warga, A. D. (1980), "Risk Instability in the Electric Utility Industry," Unpublished Doctoral Dissertation, University of Michigan.