EFFICIENT INFERENCE IN A RANDOM COEFFICIENT REGRESSION MODEL

BY P. A. V. B. SWAMY

In this paper an attempt is made to estimate a regression equation using a time series of cross sections. It is assumed that the coefficient vector is distributed across units with the same mean and the same variance-covariance matrix. The distribution of the coefficient vector is assumed to be invariant to translations along the time axis. A consistent and an asymptotically efficient estimator for the mean vector and an unbiased estimator for the variance-covariance matrix of the coefficient vector have been suggested. Some asymptotic procedures for testing linear hypotheses on the means and the variances of coefficients have been described. Finally, the estimation procedure is applied in the analysis of annual investment data, 1935–54, for eleven firms.

1. INTRODUCTION

As observed by Klein [8, pp. 216–18], it is unlikely that interindividual differences observed in a cross section sample can be explained by a simple regression equation with a few independent variables. In such situations the coefficient vector of a regression model can be treated as random to account for interindividual heterogeneity. In deriving the production function, the supply function for factors, and the demand function for a product, Nerlove [10, pp. 34–35 and Ch. 4] found it appropriate to treat the elasticities of output with respect to inputs and of factor supplies and product demand as random variables differing from firm to firm. Kuh [9] and Nerlove [20, pp. 157–187] treated the intercept as random and the slopes as fixed parameters in estimating a relationship from a time series of cross sections. In a recent note, Zellner [17] applied a regression model with random coefficients to the aggregation problem and showed that for such models and for a certain range of specifying assumptions, there would be no aggregation bias in the least squares estimates of coefficients in a macro equation obtained by aggregating a micro equation over all micro units. Thus the value of specifying a random coefficient regression (RCR) model may be substantial in econometric work. Under certain assumptions, we will study RCR models which treat both intercept and slope coefficients as random variables and will develop appropriate statistical inference procedures.

For empirical implementation of an RCR model specified in this paper, panel data provide the base. By panel data we mean temporal cross section data obtained by assembling cross sections of several years, with the same cross section units appearing in all years. The cross section units could be households or firms or any economic units.

The plan of the paper is as follows. In Section 2 we specify the model that we are analyzing and describe the proposed estimation procedure. For conditions

1 This paper forms a part of the author's doctoral dissertation submitted to the University of Wisconsin. This research was supported by the National Science Foundation under Grant GS-1350. The author would like to express his indebtedness to Professor A. S. Goldberger, under whose guidance this research was carried out, and to Professor A. Zellner for helpful comments and correction.
generally encountered, we propose an estimation procedure that yields consistent and asymptotically efficient estimators for the parameters of the model. In Section 3 we indicate the procedure of constructing forecast intervals for the predictions of coefficients. Section 4 is devoted to establishing the properties of estimators constructed in Section 2. In Section 5 we turn to consideration of two asymptotic tests of linear hypotheses on the parameters of RCR models. In Section 6 we describe an asymptotic test of equality between coefficient vectors in $N$ relations with heteroskedastic disturbances. Then the estimation and testing procedures presented in Sections 3–5 are applied in Section 7. Lastly, we present some concluding remarks in Section 8.

2. EFFICIENT ESTIMATORS OF RCR EQUATIONS USING PANEL DATA

Consider the following model:

\[(2.1) \quad \hat{y}_i = X_i \beta_i + u_i = \begin{pmatrix} X_1 & \ldots & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}, \quad (i = 1, 2, \ldots, N). \]

There are $T$ observations on each of the $N$ individual units. Observed are $\hat{y}_i$ and $X_i$ for $i = 1, 2, \ldots, N$. The matrix $X_i$ $(i = 1, 2, \ldots, N)$ is of rank $A$, containing observations on $A$ nonstochastic regressors, $x_{itA}(t = 1, 2, \ldots, T; \lambda = 1, 2, \ldots, A)$; $\beta_i$ and $u_i$ are unobserved random vectors.

We assume that for $i, j = 1, 2, \ldots, N$:

\[(2.2a) \quad E(u_i) = 0, \quad E(u_i u'_j) = \begin{cases} \sigma_i I & \text{if } i = j, \\
0 & \text{if } i \neq j; \end{cases} \]

\[(2.2b) \quad E(\hat{\beta}_i) = \bar{\beta}; \]

\[(2.2c) \quad E(\hat{\beta}_i - \bar{\beta})(\hat{\beta}_j - \bar{\beta}) = \begin{cases} A & \text{if } i = j, \\
0 & \text{if } i \neq j; \end{cases} \]

\[(2.2d) \quad \beta_i \text{ and } u_j \text{ are independent;} \]

\[(2.2e) \quad \beta_i \text{ and } \beta_j \text{ for } i \neq j \text{ are independent.} \]

Let $\hat{\beta}_i = \bar{\beta} + \delta_i$ $(i = 1, 2, \ldots, N)$ where $\delta_i$ is a $A \times 1$ vector of random elements. Assumptions (2.2b) and (2.2c) are equivalent to saying that $E\delta_i = 0$ and

\[E\delta_i \delta'_j = \begin{cases} A & \text{if } i = j, \\
0 & \text{if } i \neq j; \end{cases} \]

respectively. Assumption (2.2e) implies that $\delta_i$ and $\delta_j$ for $i \neq j$ are independent.

We can write the equation in (2.1) as

\[(2.4) \quad \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} = \begin{bmatrix} X_1 & \ldots & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} X_1 & \ldots & X_N \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}. \]
or more compactly as

\[(2.4a) \quad y = X\beta + D\delta + u\]

where \(y \equiv [y_1', y_2', \ldots, y_N']\), \(X \equiv [X_1', X_2', \ldots, X_N']\), \(\delta \equiv [\delta_1', \delta_2', \ldots, \delta_N']\), \(u \equiv [u_1', u_2', \ldots, u_N']\), the 0's are \(T \times A\) null matrices, and \(D\) denotes the block-diagonal matrix on the r.h.s. of (2.4). Under the assumptions (2.2a)–(2.2e), the \(NT \times 1\) disturbance vector \(D\delta + u\) has the covariance matrix

\[
V(\theta) = \begin{bmatrix}
X_1\Delta X_1' + \sigma_{11}I & 0 & \ldots & 0 \\
0 & X_2\Delta X_2' + \sigma_{22}I & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & X_N\Delta X_N' + \sigma_{NN}I
\end{bmatrix}
\]

where the 0's are \(T \times T\) null matrices. The matrix \(V(\theta)\) is symmetric of dimension \(NT \times NT\) and is a function of the \(X_i\)'s and an unknown \(\frac{1}{2}[A(A + 1) + 2N]^{-1}\)-element parametric vector \(\theta\) containing all the distinct elements of \(\Delta\) and \(\sigma_{ii}\) \((i = 1, 2, \ldots, N)\) arranged in an order. We assume that each element of \(V(\theta)\) has continuous first order derivatives. In some \(\frac{1}{2}[A(A + 1) + 2N]^{-1}\)-dimensional interval \(A\) that contains \(\theta_0\), the true value of \(\theta\), as an interior point, \(V(\theta)\) is positive definite for all \(\theta \in A\).

The model (2.1) represents a temporal cross section situation with the subscript \(t\) representing time and the subscript \(i\) representing a micro unit. Assumption (a) implies constant variances and zero covariances from period to period as well as the absence of any auto or serial correlation of the disturbance terms. The \(\sigma_{ii}\) is the variance of the \(ith\) micro unit’s disturbance term for any time period. In other words the disturbances are heteroskedastic and they have different variances for different micro units. Furthermore, they are structurally and temporally uncorrelated. Assumptions (b) and (c) imply that the regression coefficient vectors \(\beta_i\) are random and uncorrelated across micro units, but follow the same distribution with mean \(\bar{\beta}\) and variance-covariance matrix \(\Delta\). This distribution is assumed to be stable over time. Specification of a RCR model is appealing in the analysis of cross section data since it permits corresponding coefficients to be different for different individual units. A regression model with random intercept is a particular case of the specification (2.1).

Wald [16] and Hildreth and Houck [7] considered the problem of estimating parameters in a random coefficient model and suggested some estimators that require only a single sample of \(T\) observations on each of the variables in the model. Their specifications are different from the specification in (2.2) and in Swamy [13]. Hildreth and Houck have attempted to estimate \(2A\) parameters with \(T\) observations whereas in this paper we attempt to estimate \(\frac{1}{2}[A(A + 1)] + A + N\) parameters with \(NT\) observations. If we treat the coefficient vector \(\beta_i\) as fixed but different for different individual units, then the number of coefficient vectors to be estimated grows with the number of units in the sample. Data on each individual unit should be used to estimate its own coefficient vector. This difficulty will not be present if we can make assumptions (b) and (c) in (2.2).
We now regard (2.4a) as a single equation regression model and try to obtain estimators with desirable properties for the parameters of the model. The parameters to be estimated are the elements of $\beta$ and $\theta$.

To estimate $\beta$ we apply Aitken's generalized least squares to (2.4a). Thus a minimum variance linear unbiased estimator of $\beta$ is

\[
\begin{align*}
\bar{\beta}(\theta) &= (X'V(\theta)^{-1}X)^{-1}X'V(\theta)^{-1}Y \\
&= \left[ \sum_{j=1}^{N} X'_j \{ X_j \Delta X'_j + \sigma_{jj} I \}^{-1} X_j \right]^{-1} \sum_{i=1}^{N} X'_i \{ X_i \Delta X'_i + \sigma_{ii} I \}^{-1} Y_i.
\end{align*}
\]

Applying the matrix result in Rao [12, p. 29, (2.9)] to (2.6) we have

\[
\begin{align*}
\bar{\beta}(\theta) &= \sum_{i=1}^{N} W_i^* b_i \\
\text{where} \\
W_i^* &= \left[ \sum_{j=1}^{N} \{ \Delta + \sigma_{jj}(X'_jX_j)^{-1} \}^{-1} \right]^{-1} \{ \Delta + \sigma_{ii}(X'_iX_i)^{-1} \}^{-1}
\end{align*}
\]

and $b_i = (X'_iX_i)^{-1}X'_iY_i$.

The variance-covariance matrix of the estimator $\bar{\beta}(\theta)$ is easily shown to be $(X'V(\theta)^{-1}X)^{-1}$ or

\[
\begin{align*}
C(\theta) &= \left[ \sum_{j=1}^{N} X'_j \{ X_j \Delta X'_j + \sigma_{jj} I \}^{-1} X_j \right]^{-1} \\
&= \left[ \sum_{j=1}^{N} \{ \Delta + \sigma_{jj}(X'_jX_j)^{-1} \}^{-1} \right]^{-1}.
\end{align*}
\]

The estimator (2.7) will be recognized as the weighted average of the estimators $b_i (i = 1, 2, \ldots, N)$ with weights inversely proportional to their covariance matrices. It is to be noted that (2.7) is equal to the estimator suggested by Rao [11] if $X_1 = X_2 = \ldots = X_N$.

We shall now assume specific distributions for the variables involved.

\[
\begin{align*}
(2.10a) \quad &\quad u_i \sim N_T(0, \sigma_{ii} I), \\
(2.10b) \quad &\quad \beta_i \sim N_A(\bar{\beta}, \Delta).
\end{align*}
\]

Under these additional assumptions $\bar{\beta}(\theta)$ is a maximum likelihood estimator of $\bar{\beta}$. In addition to this property $\bar{\beta}(\theta)$ is also a minimum variance estimator not only within the class of linear unbiased estimators but also within the class of all unbiased estimators because the variance of $\bar{\beta}(\theta)$ can easily be shown to be equal to the Cramer-Rao lower bound (cf. [12, p. 265]).

Despite these optimal properties it is impossible to use (2.7) in practice since $\Delta$ and $\sigma_{ii} (i = 1, 2, \ldots, N)$ are usually unknown. What we propose to do is to employ estimators of $\Delta$ and the $\sigma_{ii}$'s in constructing the Aitken estimator.
An unbiased estimator of $\sigma_{ii}$ is given by

\begin{equation}
(2.11) \quad s_{ii} = \frac{y_i'M_iy_i}{T - A}
\end{equation}

where $M_i = I - X_i(X'_iX_i)^{-1}X'_i$. An unbiased estimator of $A$ is given by

\begin{equation}
(2.12) \quad \hat{A} = \frac{S_b}{N - 1} - \frac{1}{N} \sum_{i=1}^{N} s_{ii}(X'_iX_i)^{-1}
\end{equation}

where

\begin{equation}
S_b = \sum_{i=1}^{N} b_i b_i' - \frac{1}{N} \sum_{i=1}^{N} b_i \sum_{i=1}^{N} b_i'.
\end{equation}

Given that we have the estimates $s_{ii} (i = 1, 2, \ldots, N)$ and $\hat{A}$ we can form the estimator

\begin{equation}
(2.13) \quad \bar{b}(\hat{\theta}) = C(\hat{\theta}) \sum_{i=1}^{N} \{ \hat{A} + s_{ii}(X'_iX_i)^{-1}\}^{-1} b_i
\end{equation}

where

\begin{equation}
C(\hat{\theta}) = \left[ \sum_{j=1}^{N} \{ \hat{A} + s_{jj}(X'_jX_j)^{-1}\}^{-1} \right]^{-1}
\end{equation}

3. CONSTRUCTION OF FORECAST INTERVALS

Under the set of assumptions in (2.2) and (2.10),

\begin{equation}
(3.1) \quad \frac{(T - A)s_{ii}}{\sigma_{ii}} \sim \chi^2_{(T - A)}
\end{equation}

According to Rao's Lemma 2 [11], the function $b_i$ in (2.7) is the best linear unbiased predictor of $\beta_i$. The prediction error $(b_i - \beta_i) = (X'_iX_i)^{-1}X'_i\delta_i$, which is a linear function of $\delta_i$, is independent of the quadratic form $y'_iM_iy_i = u'_iM_iu_i$ because $(X'_iX_i)^{-1}X'_iM_i = 0$. It follows from the above argument that

\begin{equation}
(3.2) \quad \frac{\ell'(b_i - \beta_i)}{[s_{ii}\ell'(X'_iX_i)^{-1}\ell]\frac{1}{2}} \sim t_{(T - A)} \quad (i = 1, 2, \ldots N)
\end{equation}

where $\ell$ is a $(A \times 1)$ vector of arbitrary elements.

A forecast interval for $\ell'\beta_i$ with $(1 - \alpha)$ probability is given by

\begin{equation}
(3.3) \quad \ell'\beta_i \pm [s_{ii}\ell'(X'_iX_i)^{-1}\ell]\frac{1}{2} t_{\frac{1}{2}\alpha} \quad (i = 1, 2, \ldots N)
\end{equation}

where $t_{\frac{1}{2}\alpha}$ is the upper $\frac{1}{2}\alpha$ point of the $t$ distribution with $(T - A)$ d.f.

\footnote{The estimator $\hat{A}$ is not a maximum likelihood estimator of $A$ and in some numerical problems it will yield negative estimates for the variances of coefficients. An instructive numerical example on this problem is discussed in Swamy [14].}
If we set up the forecast interval (3.3) for each element in \( \beta_i \), or for different linear combinations, \( \ell' \beta_i \), we will get many intervals covering their respective elements of \( \beta_i \) or linear combinations, \( \ell' \beta_i \). But these say little about the probability of the forecast intervals simultaneously covering their respective elements of \( \beta_i \) or linear combinations \( \ell' \beta_i \). For this we have to construct simultaneous forecast regions. First recall the familiar formula for the probability that the elements of an observational vector on \( \beta_i \) are contained in an ellipsoid in the \( A \)-dimensional \( \beta_i \) space. That is,

\[
\text{Pr} \left\{ \frac{(b_i - \beta_i)'[X'X]_{ii}(b_i - \beta_i)}{A_{ii}} \leq F_a \right\} = 1 - \alpha
\]

where \( F_a \) is the upper \( \alpha \) point of the \( F \) distribution with \( A, T - A \) d.f.

Using the Cauchy-Schwartz inequality we can derive from (3.4) the simultaneous forecast intervals which cover all linear combinations of \( \beta_i \) with probability \( 1 - \alpha \):

\[
\text{Pr} (\ell' \beta_i \in \ell' b_i \pm s_{ii} A_{xx} \ell' (X'X)^{-1} \ell^\top)^{1/2} \text{ for all } \ell \) = 1 - \alpha.
\]

For any particular linear combination, \( \ell' \beta_i \),

\[
\text{Pr} (\ell' \beta_i \in \ell' b_i \pm s_{ii} A_{xx} \ell' (X'X)^{-1} \ell^\top)^{1/2} \geq 1 - \alpha
\]

so that the confidence coefficient is greater than \( 1 - \alpha \) (cf. [12, p. 198]).

4. ASYMPTOTIC PROPERTIES OF \( \hat{\theta} \)

We assume that

\[
\lim_{T \to \infty} T^{-1} X'X = \Sigma
\]

exists and is nonsingular.

4.1. Consistency of \( \hat{\theta} \)

From equations (4.1) and (2.9),

\[
\lim_{T \to \infty} \text{var} [\hat{\theta}] = 0.
\]

Hence \( \text{plim} \hat{\theta} = \beta \).

Following the same argument as in Goldberger [3, pp. 269–72], we can easily show that

\[
\text{plim}_{T \to \infty} \left| \hat{\theta} - \frac{S_{\beta}}{N - 1} \right| = 0
\]

where

\[
S_{\beta} = \sum_{i=1}^{N} \beta_i \beta_i' - \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i',
\]
and that
\[ \text{(4.4)} \quad \text{plim}_{T \to \infty} s_{it} = \sigma_{it}, \]
\[ \text{(4.5)} \quad \text{plim}_{T \to \infty} |\hat{\beta}_t - \beta^*_t| = 0, \]
and
\[ \text{(4.6)} \quad \text{plim}_{T \to \infty} |\tilde{\beta}(\hat{\theta}) - \boldsymbol{m}_{\beta}| = 0 \]
where \( \boldsymbol{m}_{\beta} = (1/N)\sum_{i=1}^{N} \beta_i \).

Hence, by the invariance property of consistency [3, p. 130]
\[ \text{(4.7)} \quad \text{plim}_{N \to \infty} \tilde{\beta}(\hat{\theta}) = \bar{\beta}. \]

Thus, \( \tilde{\beta}(\hat{\theta}) \) is a consistent estimator of \( \bar{\beta} \). From (4.6) it follows that the asymptotic distribution of \( \tilde{\beta}(\hat{\theta}) \) is the same as the distribution of \( \boldsymbol{m}_{\beta} \) (cf. [12, p. 101]). Since the variance of the latter is equal to \( \Delta/N \), we can treat \( \Delta/N \) as the estimated asymptotic variance-covariance matrix of \( \tilde{\beta}(\hat{\theta}) \).

4.2. Asymptotic Efficiency of \( \tilde{\beta}(\hat{\theta}) \)

The consistent estimator \( \tilde{\beta}(\hat{\theta}) \) is said to be efficient if
\[ \text{(4.9)} \quad \sqrt{N} \left| \tilde{\beta}(\hat{\theta}) - \bar{\beta} - \frac{1}{N} \frac{\partial \log \text{P}(b_1, b_2, \ldots, b_N|\beta, \Delta, \sigma_{ii})}{\partial \beta} N[I(\tilde{\beta})]^{-1} \right| \to 0 \]
in probability, where \( \text{P}(b_1, b_2, \ldots, b_N|\beta, \Delta, \sigma_{ii}) \) is the joint probability density of the random vectors \( b_1 \ldots b_N \), and \( I(\tilde{\beta}) \) is Fisher’s information measure on the parameter \( \tilde{\beta} \) (cf. [12, p. 285]).

That \( \tilde{\beta}(\hat{\theta}) \) does indeed satisfy the condition (4.9) can be easily shown as follows. Under the additional assumptions (2.10a) and (2.10b) we have
\[ \text{(4.10)} \quad \frac{1}{N} \frac{\partial \log \text{P}(b_1, b_2, \ldots, b_N|\beta, \Delta, \sigma_{ii}(i = 1, 2, \ldots, N))}{\partial \beta} N[I(\tilde{\beta})]^{-1} = \tilde{\beta}(\hat{\theta}) - \bar{\beta} \]
and
\[ \text{(4.11)} \quad \text{plim}_{T \to \infty} \sqrt{N} \left| \tilde{\beta}(\hat{\theta}) - \bar{\beta} - (\tilde{\beta}(\hat{\theta}) - \bar{\beta}) \right| = 0. \]

Notice here that the definition of asymptotic efficiency given in (4.9) can be adopted only if we know the form of the distribution of the \( b_i \)'s. Without making the assumptions (2.10a) and (2.10b) we can still show, using (4.2), (4.3), and (4.4), that
\[ \text{(4.12)} \quad \text{plim}_{T \to \infty} |\tilde{\beta}(\hat{\theta}) - \bar{\beta}(\hat{\theta})| = 0. \]

According to a convergence theorem in [12, p. 101] the asymptotic distribution of \( \tilde{\beta}(\hat{\theta}) \) is the same as that of \( \tilde{\beta}(\hat{\theta}) \), because the difference of these two quantities has zero probability limit. Since it is known that under general conditions the asymptotic distribution of \( \sqrt{N}(\tilde{\beta}(\hat{\theta}) - \bar{\beta}) \) is normal, the asymptotic distribution of \( \sqrt{N}(\tilde{\beta}(\hat{\theta}) - \bar{\beta}) \) is also normal. Since the variance of the limiting distribution of
$b(\hat{\theta})$ is equal to $A/N$, which is the reciprocal of Fisher's information measure on $\hat{\theta}$ in a sample of size $N$ from the population of $\beta_i$, $b(\hat{\theta})$ is the best asymptotically normal estimator (cf. [12, p. 284]).

5. ASYMPOTIC TESTS OF LINEAR HYPOTHESES ON THE PARAMETERS OF RCR MODELS

We now require assumptions (2.10a) and (2.10b) to derive the following asymptotic tests.

5.1. Asymptotic Tests of Hypotheses on the Mean of Coefficient Vectors

In this section we suggest a criterion of testing the hypothesis, $\hat{\beta} = \beta_0$ (a pre-assigned value), under the assumptions (a) to (e) in (2.2), and the assumptions (a) and (b) in (2.10).

Using the results in (4.3), (4.4), and (4.5), we can show that

$$\left[ b(\hat{\theta}) - \beta \right]'C(\hat{\theta})^{-1}[b(\hat{\theta}) - \beta] \to 0$$

in probability as $T \to \infty$ and $N$ is fixed. According to the limit theorem in [12, p. 101] the asymptotic distribution of $\left[ b(\hat{\theta}) - \beta \right]'C(\hat{\theta})^{-1}[b(\hat{\theta}) - \beta]$ is the same as the distribution of $N(N-1)(m_\beta - \beta)'S_\beta^{-1}(m_\beta - \beta)$. The distribution of the latter can be derived following the derivation of Hotelling’s distribution in Rao [12, p. 458].

Under the assumptions in (2.2) and (2.10) the asymptotic distribution of

$$\frac{N - A}{A(N - 1)}[b(\hat{\theta}) - \beta]'C(\hat{\theta})^{-1}[b(\hat{\theta}) - \beta]$$

is $F$ with $A, N - A$ d.f. Since the exact finite sample distribution of (5.2) is not known, the power of a test based on (5.2) in small samples cannot be discussed here. However, asymptotically the test of $H_0 : \beta = \beta_0$ based on (5.2) is equivalent to the test of the same hypothesis based on Hotelling’s $T^2$ statistic. The optimal properties of the latter are well known (cf. Anderson [1, pp. 115–118]).

5.2. Asymptotic Tests of Linear Hypotheses on the Variances of Coefficients

It follows from (4.3) that

$$\left| \ell' \bar{\Delta} \ell - \ell' \frac{S_\beta}{N - 1} \ell \right| \to 0$$

in probability as $T \to \infty$ and $N$ is fixed, where $\ell$ is a $(A \times 1)$ vector of fixed elements. Since $S_\beta$ is distributed as Wishart with parameters $N - 1$ and $A$, $\ell' S_\beta \ell$ is $\chi^2 \sigma^2_1$ distributed with $N - 1$ d.f., where $\sigma^2_1$ is equal to $\ell' \bar{\Delta} \ell$ (cf. [12, p. 452]). Therefore, according to the limit theorem in [12, p. 101], the asymptotic distribution of
\[(N - 1)\ell' \hat{\Delta} \ell \text{ is } \chi^2 \sigma_1^2 \text{ with } N - 1 \text{ d.f. The statistic}
\]
\[
\begin{align*}
(5.4) \quad & (N - 1) \frac{\ell' \hat{\Delta} \ell}{\ell' \Delta_0 \ell} \\
& \text{can be used to test the hypothesis that } \Delta = \Delta_0, \text{ not all elements of which are zero.}
\end{align*}
\]

6. ASYMPTOTIC TESTS OF EQUALITY BETWEEN FIXED COEFFICIENT VECTORS IN N RELATIONS WITH HETEROSEDASTIC DISTURBANCES

Before attempting to estimate any model under the set of assumptions in (2.2), it is better to test whether the coefficient vectors \( \beta_i \) \((i = 1, 2, \ldots N)\) are fixed and are all equal. Depending on the outcome of this preliminary test we can decide whether assumptions (2.2b) to (2.2e) are plausible for the situation under study. Consider the following hypothesis:

\[
(6.1) \quad H_0: \beta_1 = \beta_2 = \ldots = \beta_N = \beta.
\]

The above hypothesis states that the coefficient vectors, \( \beta_i \)'s, are fixed and the sample units under study are homogeneous in so far as the \( \beta_i \)'s are concerned. Further if (6.1) is valid, there will be no aggregation bias involved in simple aggregation of linear relations of (2.1) type (cf. Theil [15]). If hypothesis (6.1) is not true, then the data on all units cannot be pooled to estimate a single relationship between variables. One way of getting around this difficulty is to make assumptions (b) to (e) in (2.2), if they are reasonable, and pool the data to estimate a relationship of the type (2.1).

If the hypothesis (6.1) is true, then the \( b_i \) \((i = 1, 2, \ldots N)\) in (2.7) are \( N \) unbiased and independent estimators of the same parametric vector \( \beta \). The distribution of \( (b_i - \beta) \) is assumed to be \( N_A(0, \sigma_{ii}(X_i'X_i)^{-1}) \). Consider the homogeneity statistic

\[
(6.2) \quad H_\beta = \sum_{i=1}^{N} \frac{(b_i - \hat{\beta})'X_i'X_i(b_i - \hat{\beta})}{s_{ii}}
\]

where

\[
\hat{\beta} = \left[ \sum_{i=1}^{N} \frac{X_i'X_i}{s_{ii}} \right]^{-1} \sum_{i=1}^{N} \frac{X_i'X_i}{s_{ii}} b_i.
\]

If the \( b_i \)'s do not estimate the same parametric vector, then the differences are reflected in the statistic \( H_\beta \) and therefore it may be used to test \( H_0 \) in (6.1). The asymptotic distribution of \( H_\beta \) is \( \chi^2 \) with \( \Lambda(N - 1) \) d.f. as \( T \to \infty \) and \( N \) is fixed (cf. [12, pp. 323–24]). The asymptotic distribution of \( H_\beta/\Lambda(N - 1) \) can also be approximated by \( F \) with \( \Lambda(N - 1), N(T - \Lambda) \) d.f. (cf. Zellner [18]).

7. APPLICATION OF METHODS IN THE ANALYSIS OF DATA

To illustrate application of the methods discussed above, we utilize the investment equation developed by Grunfeld [6] and described by Boot and de Wit [2], and Griliches and Wallace [5]. Grunfeld's investment function involves a firm's
current gross investment, \( y_{it} \), being dependent on the value of firm’s outstanding shares at the beginning of the year, \( x_{1t-1,i} \), and on its beginning-of-year capital stock, \( x_{2t-1,i} \). That is, the micro investment function for a firm is

\[
y_{it} = \beta_{0i} + \beta_{1i}x_{1t-1,i} + \beta_{2i}x_{2t-1,i} + u_{it}
\]

with the subscripts \( i \) and \( t \) denoting an observation for the \( i \)th firm in the \( t \)th year.

Herein we present estimates of (7.1) for eleven corporations\(^3\) in the U.S. by the method described above. The annual data, 1935–54, for each corporation are taken from [2].

On the assumption that the same equation of the type (7.1) is applicable to all corporations in our cross section sample of size eleven, we have a system of eleven equations, each of which relates to a single corporation. We write the set of eleven equations in matrix notation as

\[
y_i = X_i\beta + u_i \quad (i = 1, 2, \ldots, 11)
\]

where \( y_i \) is a \( 20 \times 1 \) vector of observations on \( y_{it} \), \( X_i \) is a \( 20 \times 2 \) matrix of observations on \( x_{1t-1,i} \) and \( x_{2t-1,i} \), \( \beta_i \) is a \( 2 \times 1 \) vector of random coefficients and \( u_i \) is a \( 20 \times 1 \) vector of additive disturbances. The intercept term \( \beta_{0i} \) is eliminated by expressing the observations on each variable as deviations from their respective firm means.

Before adopting the random coefficient approach to estimate the model (7.2), we conduct a preliminary test of the following hypothesis using (6.2):

\[
H_0: \beta_1 = \beta_2 = \ldots = \beta_N = \beta,
\]

where \( \beta_i \), for every \( i \), is a \( 2 \times 1 \) vector of fixed coefficients. The value of the statistic in (6.2) on the basis of our sample data is 14.4521. This value is well above the five per cent value of \( F \) with 20 and 187 d.f. so that the data do not support the hypothesis that the coefficient vectors in (7.2) are the same for all corporations.

Under these conditions it may be reasonable to assume that the coefficient vector in (7.2) is random across units. We now make assumptions (a) to (e) in (2.2) for \( i,j = 1, 2, \ldots, 11 \). Assumption (b) in (2.2) implies that the data on all the 11 corporations contain some information on \( \bar{\beta} \), the means of coefficients, and they can be utilized in estimating \( \bar{\beta} \). Adopting the formula (2.13) we have

\[
\bar{b}(\hat{\theta}) = [\bar{b}_1(\hat{\theta})\bar{b}_2(\hat{\theta})]' = \begin{bmatrix} 0.0843 & 0.1961 \\ 0.0104 & 0.0412 \end{bmatrix}.
\]

The figures given in the parentheses are the estimated asymptotic standard errors of \( \bar{b}_1(\hat{\theta}) \) and \( \bar{b}_2(\hat{\theta}) \). They are obtained by taking the square root of the diagonal elements of \( \hat{\Delta} \), where \( \hat{\Delta} \) is the estimate of the variance-covariance matrix of the \( \beta_i \)'s. \( \hat{\Delta} \) is obtained by using the formula (2.12):

\[
\hat{\Delta} = \begin{bmatrix} 0.0011 & -0.0002 \\ -0.0002 & 0.0187 \end{bmatrix}.
\]

On the basis of the asymptotic distribution of \( \ell' \hat{\Lambda} \ell \), which is derived in (5.4), we obtain the following interval estimates for the variances of the marginal distributions of \( \beta_{1i} \) and \( \beta_{2i} \):

\[
\begin{align*}
\begin{cases}
(0.0005 < \sigma_{\beta_{1i}}^2 < 0.0034) \\
(0.0091 < \sigma_{\beta_{2i}}^2 < 0.0575)
\end{cases}
\end{align*}
\]

where \( \sigma_{\beta_{1i}}^2 \) is the variance of the marginal distribution of \( \beta_{1i} \) and \( \sigma_{\beta_{2i}}^2 \) is the variance of the marginal distribution of \( \beta_{2i} \).

If the above interval estimates based on a single sample cover the true values of the variances of coefficients, then they indicate that the true value of \( \sigma_{\beta_{1i}}^2 \) is small, lying between 0.0005 and 0.0034.

We now turn to an application of (5.2) to test the hypothesis

\[
H_0: E_{\beta_{1i}} = 0.
\]

In the present application the value of the statistic (5.2) is 30.8720 and this falls well above the five per cent value of \( F \) with 2 and 9 d.f., so that the data cannot be regarded as consistent with the hypothesis (7.7).

Aggregating over \( i \) we can write the estimated aggregate relationship as

\[
\hat{\gamma}_t = 0.0843x_{1t-1} + 0.1961x_{2t-1}
\]

where

\[
y_t = \sum_{i=1}^{11} y_{it},
\]

\[
x_{1t-1} = \sum_{i=1}^{11} x_{1t-1,i},
\]

and

\[
x_{2t-1} = \sum_{i=1}^{11} x_{2t-1,i}.
\]

The above equation indicates that at the aggregate level the value of the firm's outstanding shares is an important variable in explaining the investment. By following different approaches Griliches and Wallace [5] and Gould [4] criticized the theory underlying the model (7.1).

8. CONCLUDING REMARKS

Assuming that the coefficients in a regression equation are random across units in a temporal cross section analysis we have presented an efficient method of estimating the mean and the variance-covariance matrix of a distribution which the coefficient vector follows. We have suggested a consistent and an asymptotically efficient estimator for the mean, and an unbiased and a consistent estimator for the variance-covariance matrix. We have developed two asymptotic procedures, one to test a linear hypothesis on the means of coefficients and another to test a
linear hypothesis on the variances of coefficients. We have also described an
asymptotic test of equality between fixed coefficient vectors in $N$ relations with
heteroskedastic disturbances.

There is a close connection between RCR models as treated in this paper and a
Bayesian approach to fixed coefficient models. For estimation purposes we treated
the coefficient vector $\hat{\beta}_i$ as a random variable without specifying the form of its
distribution and built an estimator for the mean $\bar{\beta}$ proceeding conditionally upon
a given sample estimate of the variance-covariance matrix $V(\hat{\beta})$, a fact that appears
to explain why our results in the main are large sample results. A Bayesian, while
analyzing a fixed coefficient model, would also assume that the actual value
of the coefficient vector in the sampled population was determined by a random
experiment. In such cases $\hat{\beta}_i$ is a random variable having a certain a priori distribu-

tion. He would put a prior pdf on the coefficient vector and combine it with the
likelihood of a sample via Bayes’ theorem. He would be able to integrate out the
nuisance parameters and obtain a posterior density for the coefficient vector.

In closing, we note that many, if not all, micro units are heterogeneous with
regard to the regression coefficient vector in a model. If we proceed blithely with
cross section analysis ignoring such heterogeneity, we may be led to erroneous
inferences. Application of the random coefficient approach to situations that satisfy
the specifying assumptions introduced above leads to correct results.

This last conclusion should not be generalized to all situations. Obviously, it
is not hard to imagine situations that do not satisfy the above specifying assump-
tions. As usual, each situation has to be analyzed carefully.

State University of New York, Buffalo

Manuscript received November, 1967; revision received August, 1968.

REFERENCES

Sons, 1958.
University of Chicago, 1958.
efficients,” Journal of the American Statistical Association, 63 (1968), 584–95.
to the Analysis of Growth Curves,” Biometrika, 52 (1965), 447–58.


