Volume Title: Annals of Economic and Social Measurement, Volume 2, number 2

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm73-2

Publication Date: April 1973

Chapter Title: On the Use of Survey Sample Weights in the Linear Model

Chapter Author: Richard Porter

Chapter URL: http://www.nber.org/chapters/c9887

Chapter pages in book: (p. 141 - 158)

Annals of Economic and Social Measurement, 2/2, 1973

ON THE USE OF SURVEY SAMPLE WEIGHTS IN THE LINEAR MODEL*

BY RICHARD D. PORTER

If individuals have different coefficients in a linear model, then the choice of regression technique for estimating population averages depends on the sample design. We examine various estimators of the random coefficient model for panel data, where the random component arises from the random selection of individuals out of a finite population.

1. INTRODUCTION

1.1. *Problem*

Sample surveys such as the Current Population Survey are a rich source of economic data. If the sample is drawn according to the principles of sample survey theory, each member will have an attached weight. For example, suppose there are two strata A and B and that a sample is drawn in which members in A are sampled at a rate 6:1,000 (six per thousand population individuals in A) whereas members in B are sampled at a rate of 3:1,000. Then to compute a population total, say the total wage bill for the population as a whole, it is sensible to give twice as much weight to an earnings measurement in B as to an earnings measurement in A, that is, the weights will be proportional to the inverse of the probability of being selected. But when different classes or strata are sampled at different rates, should the associated weights be used in estimating a behavioral econometric model? And how should they be used? In practice we usually have more information about the method by which the sample was drawn than just sampling weights for each observation. We also know the type of sampling procedure (such as simple random sampling with replacement, simple random sampling without replacement, stratified random sampling, single-stage Cluster sampling, multi-stage sampling) as well as detailed probability descriptions of the procedure. We often know the probability that any unit will be drawn as well as the joint probability that any pair of units will be drawn. As before, this information about the sampling design can be incorporated into estimates of population totals, standard error estimates for the estimated population totals, and so forth. But what use should we make of this information in estimating a behavioral econometric model?

In the econometric literature, opinions divide. Some authors advocate that the sample weights be used in linear econometric models in a way which is similar to the use of weights in computing finite population totals: they recommend using weighted least-squares.[1] Other writers argue that such sample survey

* I wish to thank my colleagues, John Paulus, Joe Sedransk, P.A.V.B. Swamy and my discussant, Professor Arnold Zellner, for useful criticisms and comments. Thanks also go to my summer assistant, Ken Wise of Northfield Park and M.I.T., for valuable advice and invaluable Fig Newtons. An expanded version of this paper is available from the author.

[1] See Klein and Morgan (1951), Klein (1953, pp. 305–313), Hu and Stormsdorfer (1970), and Cohen, Rea, and Lerman (1970, pp. 193–194).

information is irrelevant for econometric models.[2] Most econometric textbook authors do not discuss this issue.[3]

1.2. *Homogeneous Coefficients: The Choice of the Regression Technique Does Not Depend on the Sample Design*

If the coefficients in the behavioral model are homogeneous throughout the population, then the sample design does not affect the validity of the usual (least-squares) estimates. To pursue this point consider the following example.

Suppose there are q possible samples of size n that can be drawn from a population of size N according to the sampling design chosen and that the probability of selecting each sample is known. To represent this probability model for sampling we construct a random variable S taking on q distinct values $s_1, s_2, \ldots, s_q$ with associated probabilities $p_1, p_2, \ldots, p_q$. Let the regression model for any sample, say the sth, be given by

(1) $$\mathbf{y}_s = X_s \boldsymbol{\beta} + \mathbf{u}_s,$$

where X_s is a $n \times k$ matrix of regressors, $\mathbf{y}_s$ is a $n \times 1$ vector of regressands, $\boldsymbol{\beta}$ is a fixed $k \times 1$ vector of unknown coefficients and $\mathbf{u}_s$ is a $n \times 1$ vector of unobserved disturbances. We treat X_s as fixed so that the only source of variation in $\mathbf{y}_s$ is due to the variation in the disturbance vector $\mathbf{u}_s$. We postulate that $\mathbf{u}_s$ is generated by a classical probability mechanism which is independent of the sampling design and exhibits the usual properties

(2) $$E_c(\mathbf{u}_s | X_s) = \mathbf{0} \qquad \text{for all } s,$$

(3) $$E_c(\mathbf{u}_s \mathbf{u}_s' | X_s) = \sigma^2 I \qquad \text{for all } s,$$

where E_c denotes the expectator operator. We distinguish the operator by the subscript c, where c stands for the classical probability mechanism generating the disturbances. Assume X_s has full column rank for all s so that the least-squares estimator of $\boldsymbol{\beta}$, namely

(4) $$\mathbf{b}(s) = (X_s' X_s)^{-1} X_s \mathbf{y}_s$$

exists.

To evaluate properties of $\mathbf{b}$ remember that we must take into account two sources of random variation: that caused by the random selection of individuals and that caused by the random variation in the disturbance vector. Since the unconditional expectation $E\mathbf{b}(s)$ is the sum of the conditional expectations, we have

(5) $$E[\mathbf{b}(s)] = \sum_{i=1}^{q} E_c[\mathbf{b}(s) | S = s_i] p_i,$$

[2] See Cramer (1971, p. 143), Fleischer and Porter (1970, pp. 99–111), and Roth (1971). I became aware of several of these references by reading Roth's memorandum, Roth (1971).

[3] See e.g., Dhrymes (1970), Goldberger (1964), Goldberger (1968), Johnston (1963), Kmenta (1971), Malinvaud (1966), Theil (1971), Zellner (1971). A notable exception is Klein's pioneering textbook, Klein (1953); Champernowne—Champernowne (1969)—takes up survey sampling theory but does not relate it to the regression model.

where $E_c[\mathbf{b}(s)|S = s_i]$ represents the conditional expected value of $\mathbf{b}(s)$ given the event $S = s_i$. Given our specification for $\mathbf{u}$ we can show that $\mathbf{b}(s)$ is an unbiased estimator of $\boldsymbol{\beta}$. Inserting (4) and (1) into (5) and simplifying gives

$$(6) \qquad E[\mathbf{b}(s)] = \sum_{i=1}^{q} E_c[\boldsymbol{\beta} + (X_s'X_s)^{-1}X_s'\mathbf{u}_s|S = s_i]p_i = \boldsymbol{\beta} \sum_{i=1}^{q} p_i = \boldsymbol{\beta}.$$

The crucial relations used to derive (6) are (a) X_s is fixed for a given sample and (b) $E_c[\mathbf{u}_s|S = s_i] = \mathbf{0}$. The assumption that $\mathbf{u}$ does not depend on the sampling procedure is critical for establishing (b).

If we restrict our analysis to be conditioned upon the particular X matrix which is drawn, then the Gauss–Markov theorem holds and the least-squares estimator will be a best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$.[4] Indeed, it would appear that $\mathbf{b}(s)$ will have these optimal properties when we also allow for sampling variations.[5]

The implication of the foregoing analysis is that for homogeneous populations we are not obliged to incorporate the structure of the sampling plan into our regression analysis. Of course, the sample design is important regardless of whether coefficients are homogeneous or heterogeneous.

1.3. *Outline of the Paper*

In the rest of the paper we adopt the assumption that the coefficients differ across individuals. Then it appears that the choice of the regression technique depends on the sample design so we explore some procedures for combining the information on the sample design with the specification of the behavioral model to obtain estimates of certain population parameters. In Section 2 we review some results from sample survey theory. We employ these results in Section 3 to form estimators for the *random coefficient* regression model based on panel data. Here the "random" component in the coefficient arises solely from the random selection of individuals. Although this problem has been intentively studied recently,[6] the analysis has implicitly proceeded under the assumption of random sampling from an infinite population. We consider the more usual sampling design in which sampling is done without replacement from a finite population with unequal probabilities. See Konijn (1962) for a related contribution when the data source is a single cross section.[7]

[4] See, e.g. Theil (1971, p. 119).

[5] The proof follows the standard proof of the Gauss–Markov theorem, Theil (1971, pp. 119–120). The proof consists of showing that the covariance matrix of the least squares estimator, say V, is

$$V = \sum_{i=1}^{q} p_i\sigma^2(X_{si}'X_{si})^{-1},$$

while any other linear unbiased estimator, say $A_s y_s$—where A_s may be functionally dependent on s—has a covariance matrix equal to

$$V + \sum_{i=1}^{q} p_i(A_{si}'A_{si}).$$

[6] See Rao (1965), Zellner (1966), Swamy (1968), (1970), (1971), (1972), Theil (1971), Lindley and Smith (1972) and Schmalensee (1972).

[7] I am grateful to Professor Zellner for bringing Konijn's valuable study to my attention.

2. Sampling from Finite Populations

In this section we review some elements of sampling theory from finite populations.[8] The object of this theory is descriptive: to estimate finite population totals or averages.

2.1. *Simple Random Sampling Without Replacement*

We start with the concept of an ordered *random sample*. Let the finite population being sampled consist of N items, numbered $1, 2, \ldots, N$. An ordered sample from this population is an arrangement of the items in a particular order. For example, if the population consists of three elements $\{1, 2, 3\}$, there are six possible ordered samples of size two: $(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)$.[9] When each of these ordered samples appears with equal frequency in repetitive sampling, the sample is called an *ordered random sample*. Define the product $N(N-1)\ldots(N-n+1) = \pi(N, n)$. Probabilities are herein computed in accord with the equivalence law of ordered random sampling:

Theorem 1 (The Equivalence Law of Ordered Random Sampling)

If an ordered random sample of size n is drawn from a population of size N, then on any particular one of the n draws, each of the N items has the same probability $1/N$ of appearing.

Proof. See Hodges and Lehmann (1970, pp. 55–59).

The theorem generalizes to more than one item in a general way but we need consider only:

Theorem 2

Any pair of items, say I and J, has the same probability $1/\binom{N}{2}$ of appearing on any 2 specified draws. (Note that we do not indicate the *order* in which I and J appear on the two specified draws.)

Proof. Without loss of generality suppose that the two draws are the first and the second. If I appears on the first and J appears on the second, the remaining items can be drawn in $\pi(N-2, n-2) = (N-2), (N-3)\ldots(N-n+1)$ ways; alternatively, J may appear on the first and I on the second in $\pi(N-2, n-2)$ ways. Thus, the probability of $\{I, J\}$ on draws 1 and 2 is $2\pi(N-2, n-2)/\pi(N, n) = 1/\binom{N}{2} = 2/(N)(N-1)$.

Suppose we are not interested in an *ordered* random sample but in an *unordered* random sample. We can obtain an unordered random sample by first drawing an ordered random sample and then disregarding the order.[10]

Let y designate the variable which we are measuring in the population; y may be a scalar or a vector. For the present we will let y be a scalar. The value of y for the first item in the population is y_1, for the second y_2, and so forth. If we consider

[8] Hodges and Lehmann (1970. Sections 2.3, 4.3, 7.2, 9.1 and 10.3). Kendall and Stuart (1966. Chapters 39–40), and Cochran (1963) are useful introductions to the sampling theory. We draw on them in this section.

[9] Note that we use braces "{ }" when the order is irrelevant and parentheses "( )" when the order becomes important.

[10] See Hodges and Lehmann (1970, p. 54).

a random drawing of one item, say $\tilde{y}$, from this population, its expected value and variance are

$$E[\tilde{y}] = \sum_{i=1}^{N} y_i \Pr(\tilde{y} = y_i) = \sum_{i=1}^{N} y_i(1/N) \equiv \mu \tag{7}$$

$$\operatorname{Var}[\tilde{y}] = \sum_{i=1}^{N} (y_i - \mu)^2(1/N) \equiv \sigma^2. \tag{8}$$

Note that the population mean and variance, μ and σ^2, are generated by a very simple probability mechanism: the random drawing of one item from this population.

It will simplify matters if we adopt the following notational conventions. Let $p_i(r)$ be the probability that the ith person is selected on the rth draw. Let $p_{ij}(r, s)$ be the probability that the i and jth individuals are selected on the r and sth draws respectively. Let $\bar{N} = \{1, 2, \ldots, N\}$ and $\bar{n} = \{1, 2, \ldots, n\}$. As a shorthand we will write

$$\dot{\sum} w_i = \sum_{i=1}^{N} w_i, \qquad \dot{\sum} w_{ij} = \sum_{j=1}^{N} \sum_{\substack{i=1 \\ i \neq j}}^{N} w_{ij}, \qquad \sum w_i = \sum_{i=1}^{n} w_i,$$

and

$$\sum w_{ij} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} w_{ij}.$$

We next draw an ordered random sample, say $(\tilde{y}, \tilde{y}, \ldots, \tilde{y}_n)$, from this population. By Theorem 1, each $\tilde{y}$ has the same probability distribution:

$$p_i(r) = 1/N \qquad \text{for all } i \in \bar{N} \text{ and } r \in \bar{n}. \tag{9}$$

Consequently for each $r \in \bar{n}$

$$E(\tilde{y}_r) = \dot{\sum} y_i N^{-1} = \mu \tag{10}$$

$$\operatorname{Var}(\tilde{y}_r) = \dot{\sum}(y_i - \mu)^2 N^{-1} = \sigma^2. \tag{11}$$

In view of the proof of Theorem 2 we have

$$p_{ij}(r, s) = 1/N(N-1) \qquad \text{for all } r, s \in \bar{n}, r \neq s \text{ and } i, j \in \bar{N}, i \neq j. \tag{12}$$

Thus the covariance between $\tilde{y}_r$ and $\tilde{y}_s$ is equal for all r and s. If C is this common covariance, C satisfies

$$\operatorname{Var}\left(\sum \tilde{y}_i\right) = n\sigma^2 + (n^2 - n)C.$$

When $n = N, \sum \tilde{y}_i$ is a constant with zero variance so $N\sigma^2 + N(N-1)C = 0$ and

$$C = -\sigma^2/(N-1). \tag{13}$$

We now consider the problem of estimating μ. It is convenient to cast this problem in the format of a linear model. Let ε_i be a variable defined by $\varepsilon_i \equiv y_i - \mu$, for i in $\bar{N}$. If we observe the entire population, μ is known exactly; this implies that

ε_i, $i = 1, 2, \ldots, N$ are known quantities. However the sample values $\tilde{\varepsilon}_r \equiv \tilde{y}_r - \mu$, r in $\bar{n}$, are random variables with the following properties:

(14) $\quad p_i(r) = N^{-1}$ $\quad$ for all $i \in \bar{N}$ and all $r \in \bar{n}$

$\quad\quad p_{ij}(r, s) = 1/N(N-1)$ $\quad$ for all $i, j \in \bar{N}$, $i \neq j$ and $r, s \in \bar{n}$, $r \neq s$.

Our sample $(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)$ thus belongs in the following setup:

(15)
$$\tilde{\mathbf{y}} = \mathbf{1}\mu + \tilde{\boldsymbol{\varepsilon}}$$

(16)
$$E(\tilde{\boldsymbol{\varepsilon}}) = \mathbf{0}$$

(17)
$$E\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}' = \sigma^2\Omega, \Omega = (1-\rho)I + \rho\mathbf{1}\mathbf{1}'$$

where Ω is a $n \times n$ matrix, $\rho \equiv -(N-1)^{-1}$ and $\tilde{\mathbf{y}}' = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)$, $\tilde{\boldsymbol{\varepsilon}}' = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_n)$, $\mathbf{1}' = (1, 1, \ldots, 1)$ are $1 \times n$ vectors. For the model of (15), (16), and (17) the best linear unbiased estimator (BLUE) of μ is, of course, the Aitken generalized least-squares estimator

(18)
$$\hat{\mu} = (\mathbf{1}'\Omega^{-1}\mathbf{1})^{-1}\mathbf{1}'\Omega^{-1}\tilde{\mathbf{y}}.$$

Let

(19)
$$r = \{1 + (n-1)\rho\}.$$

One can easily verify that[11]

(20)
$$\Omega^{-1} = \frac{1}{(1-\rho)r}[rI - \rho\mathbf{1}\mathbf{1}']$$

so that

$$\mathbf{1}'\Omega^{-1} = r^{-1}\mathbf{1}'$$
$$\mathbf{1}'\Omega^{-1}\mathbf{1} = r^{-1}n.$$

Thus

(21)
$$\hat{\mu} = rn^{-1}r^{-1}\mathbf{1}'\tilde{\mathbf{y}} = \frac{\sum y_i}{n} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\tilde{\mathbf{y}}.$$

That is, the Aitken estimator and the ordinary least-squares estimator are identical in this case.

2.2. *Simple Random Sampling Without Replacement With Unequal Probabilities*

We now relax the assumption that all individuals have an equal chance of being selected on each draw and permit probabilities of being drawn to differ between individuals and from drawing to drawing. Most sample designs are special cases of this scheme.[12] As before let $p_{ij}(r, s)$ be the probability that the i and jth individuals are selected on the r and sth draws respectively in a sample of size n from a population of size N: i and j range from 1 to N and r and s from

[11] This result is well known. See, for example, Kendall and Stuart (1966, p. 167).
[12] See Kendall and Stuart (1966, p. 177 ff).

1 to n where $i \neq j$ and $r \neq s$. The probability that the ith person is selected on the rth draw, $p_i(r)$, is

(22) $$p_i(r) = \sum_{\substack{s=1 \\ r \neq s}}^{n} \sum_{\substack{j=i \\ i \neq j}}^{N} p_{ij}(r, s)/(n-1).$$

Since someone is always selected at the rth drawing,

(23) $$\sum_{i=1}^{N} p_i(r) = 1.$$

Let π_i be the probability that the ith person is selected in the sample,

(24) $$\pi_i = \sum_{r=1}^{n} p_i(r).$$

Finally, let π_{ij} be the joint probability that the ith and jth persons are selected in the sample,

(25) $$\pi_{ij} = \sum_{\substack{r=1 \\ r \neq s}}^{n} \sum_{s=1}^{n} p_{ij}(r, s).$$

Since $p_{ij}(r, s) = p_{ji}(s, r)$, we have, of course, that $\pi_{ij} = \pi_{ji}$.

For our purposes, it will suffice to characterize the sampling design in terms of π_i and π_{ij}. From (23) and (24) we find

(26) $$\dot{\sum} \pi_i = n.$$

From (22), (24), and (25) we get

(27) $$\sum_{j=1}^{N} \pi_{ij} = (n-1)\pi_i$$

(28) $$\sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \pi_{ij} = n(n-1).$$

Before, we were careful to distinguish between the labelling of observations in the sample and that in the population. The second person in our sample will not usually be the second person in the population. However, now we will label the sample observations in the order in which they are drawn and *not* distinguish between the order in the sample and the order in the population. As long as we are considering symmetric functions of sample observations this notational convention will not lead us astray.

A result we shall often call upon is the following:

Theorem 3

Suppose a sample of size n, $y_1, y_2, \ldots, y_n$ is drawn from a population of size N. Then for any function g

(29) $$E[\sum g(y_i)] = \dot{\sum} \pi_i g(y_i)$$

(30) $$E[\sum g(y_i, y_j)] = \dot{\sum} \pi_{ij} g(y_i, y_j).$$

Proof. So that there is no ambiguity let us first write out (29) and (30) fully:

$$E\left[\sum_{i=1}^{n} g(y_i)\right] = \sum_{i=1}^{N} \pi_i g(y_i) \tag{29'}$$

$$E\left[\sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} g(y_i, y_j)\right] = \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \pi_{ij} g(y_i, y_j). \tag{30'}$$

To prove (29) note that

$$E[\sum g(y_i)] = \sum_i Eg(y_i). \tag{31}$$

But by definition $Eg(y_i) = \sum_j g(y_j)p_j(i)$. Thus

$$E[\sum g(y_i)] = \sum_i \sum_j g(y_j)p_j(i) = \sum_j g(y_j) \sum_i p_j(i) = \sum_j g(y_j)\pi_j.$$

This proves (29): equation (30) follows by a similar argument.

We can use Theorem 3 to obtain a linear unbiased estimator of the population mean, μ.

$$\mu = \frac{1}{N} \sum_{i=1}^{N} y_i. \tag{32}$$

Suppose the same weight v_i is to be assigned to an individual whenever he is selected. A linear estimator will have the form

$$\hat{\mu} = \sum_{i=1}^{n} v_i y_i, \tag{33}$$

with the weights to be determined by the unbiasedness condition. Using (29) with $g(y_i) = v_i y_i$ we find

$$E[\hat{\mu}] = \sum_{i=1}^{N} (v_i y_i)\pi_i. \tag{34}$$

Then equating coefficients in (32) and (34) we must have

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i}. \tag{35}$$

3. Survey Sampling and the Random Coefficient Regression Model for Panel Data

3.1. *Introduction*

Recently, there has been renewed interest in the random coefficient regression model.[13] A specification leading to a random coefficient regression model occurs in the survey sampling framework. Suppose the population consists of N individuals

[13] See the references in footnote 6. Also see Hildreth and Houck (1968). Swamy (1971). (1972) provides an extensive bibliography on this literature.

and let the economic relationship for the ith unit be given by

$$\mathbf{y}_i = X_i\boldsymbol{\beta}_i + \mathbf{u}_i,\ i \in \bar{N}, \tag{36}$$

where $\mathbf{y}_i$ is a $T \times 1$ vector of observations on the dependent variable, X_i is a $T \times K$ matrix of observations with rank K on K independent variables, $\boldsymbol{\beta}_i$ is a $K \times 1$ vector of *non-random* coefficients and $\mathbf{u}_i$ is a $T \times 1$ vector of disturbance terms with mean zero for each i.

It is convenient to think of T as the number of time periods so that, for example, the tth element of $\mathbf{y}_i$ and $\mathbf{u}_i$ refer to the tth period. We allow for heterogeneity across individuals: each unit has its own coefficient vector.

The random coefficient model arises when a sample is drawn from a population. At the beginning of the first sampling period n individuals are randomly selected out of the population. In T successive periods the *same* n individuals are sampled. Assembling the observations on the n individuals for T periods we have[14]

$$\begin{aligned} \mathbf{y}_1 &= X_1\boldsymbol{\beta}_1 + \mathbf{u}_1 \\ \mathbf{y}_2 &= X_2\boldsymbol{\beta}_2 + \mathbf{u}_2 \\ &\vdots \\ \mathbf{y}_n &= X_n\boldsymbol{\beta}_n + \mathbf{u}_n. \end{aligned} \tag{37}$$

The random selection of individuals determines the random coefficient model for the system in (37). Let the population coefficient vector of interest be given by[15]

$$\boldsymbol{\beta} = \frac{1}{N}\sum \boldsymbol{\beta}_i. \tag{38}$$

We will develop various estimators for $\boldsymbol{\beta}$ under two sampling schemes: simple random sampling without replacement and random sampling without replacement with unequal probabilities.

3.2. *Simple Random Sampling Without Replacement*

In simple random sampling the units are drawn without replacement with equal probabilities. We shall make the following specification initially for the system of observations in (37) which came from the population in (36).

Assumption 3.1:

1. The number of units sampled (n) and the number of time periods (T) are such that $n > K$ and $T > K$.
2. For each unit i in the population, the independent variables are fixed in repeated samples on $\mathbf{y}_i$. The rank of $X \equiv [X_1', X_2', \ldots, X_n']'$ is K for every possible sample drawn.
3. The disturbance vectors $\mathbf{u}_i$ ($i \in \bar{N}$) are independently distributed each having mean zero. The variance–covariance matrix of $\mathbf{u}_i = \sigma_{ii}I_T$.
4. The n units are drawn by simple random sampling without replacement from the population of N units.

[14] As in Section 2 we *do not* distinguish between the labeling order in the sample and the population.

[15] We could carry out the analysis for other population concepts such as $\boldsymbol{\beta}^* = \sum w_i\boldsymbol{\beta}_i$, where w_i are known weights.

As was stressed in the introduction, there are two different sources of random variation in this model, one being the behavioral random error, the $\mathbf{u}$ vectors, and the other being the variation in β vectors caused by the random selection of individuals. In evaluating expectations of random variables it will often be convenient to distinguish these two sources of variation. We shall use the shorthand S to denote the summation over individual units, i.e., the variation caused by sampling. And we shall let ε denote the integration over the behavioral random errors, the $\mathbf{u}$'s.

Since the method of sampling is simple random sampling, the results reviewed in Section 2 apply directly to the $\boldsymbol{\beta}$'s. In particular, from (10) we have

$$E_S(\boldsymbol{\beta}_i) = \boldsymbol{\beta}, \qquad i \in \bar{n}. \tag{39}$$

We shall define the variance-covariance matrix for the population by

$$\Delta = \sum \frac{(\boldsymbol{\beta}_i - \boldsymbol{\beta})(\boldsymbol{\beta}_i - \boldsymbol{\beta})'}{N}. \tag{40}$$

We assume that Δ is positive definite. The sampling errors

$$\boldsymbol{\delta}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta} \qquad i \in \bar{n} \text{ have zero mean values.} \tag{41}$$

Using (11) we have

$$E(\boldsymbol{\delta}_i\boldsymbol{\delta}_i') = \Delta, \qquad i \in \bar{n}. \tag{42}$$

Finally, the matrix version of (13) is (43):

$$E(\boldsymbol{\delta}_i\boldsymbol{\delta}_j') = -\frac{\Delta}{N-1} \qquad i, j \in \bar{n}, \qquad i \neq j. \tag{43}$$

For the model of (3.1) we shall consider two estimates. The first will be the simple average of the least-squares estimators of each unit in the sample. The second estimator is an approximate Aitken estimator.

Average Least-Squares Estimator

Let $\mathbf{b}$ be the first estimator,

$$\mathbf{b} = \frac{1}{n}\sum \mathbf{b}_i, \tag{44}$$

where

$$\mathbf{b}_i = (X_i'X_i)^{-1}X_i'\mathbf{y}_i. \tag{45}$$

Considering the variation in $\mathbf{u}_i$ above we have the usual result that

$$E_\varepsilon(\mathbf{b}_i|S) = \boldsymbol{\beta}_i \tag{46}$$

where $E_\varepsilon(\mathbf{b}_i|S)$ denotes the conditional expected value of $\mathbf{b}_i$ given the ith unit is drawn. From (39) and (46) we obtain

$$E(\mathbf{b}) = \frac{1}{n}\sum E_S E_\varepsilon[\mathbf{b}_i|S] = \frac{1}{n}\sum E_S(\boldsymbol{\beta}_i) = \boldsymbol{\beta}. \tag{47}$$

That is, $\mathbf{b}$ is an unbiased estimator of $\boldsymbol{\beta}$.

Next we determine the variance–covariance matrix for $\mathbf{b}$ and an estimator of it. The error between $\mathbf{b}$ and $\boldsymbol{\beta}$ is

$$\mathbf{b} - \boldsymbol{\beta} = n^{-1}[\sum(\boldsymbol{\delta}_i + (X_i'X_i)^{-1}X_i'\mathbf{u}_i)]. \tag{48}$$

It will simplify notation to introduce P_i by

$$P_i = \sigma_{ii}(X_i'X_i)^{-1}. \tag{49}$$

The variance–covariance matrix of $\mathbf{b}$, say S_{bb}, is

$$S_{bb} = E(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'.$$

Evaluating S_{bb} we find

$$S_{bb} = \frac{\Delta(N - n)}{n(N - 1)} + \frac{1}{Nn}\sum^{\cdot} P_i. \tag{50}$$

To obtain an estimate of S_{bb} we shall first evaluate the matrix S_b,

$$S_b = \sum \mathbf{b}_i\mathbf{b}_i' - \frac{1}{n}(\sum \mathbf{b}_i)(\sum \mathbf{b}_i'). \tag{51}$$

Substituting

$$\mathbf{b}_i = \boldsymbol{\beta}_i + (X_i'X_i)^{-1}X_i'\mathbf{u}_i$$

into (51) and taking expectations gives

$$E[S_b] = (n - 1)\Delta + (n - 1)\frac{\Delta}{N - 1} + \frac{(n - 1)}{N}\sum^{\cdot} P_i. \tag{52}$$

Let

$$M_i = I - X_i(X_i'X_i)^{-1}X_i' \tag{53}$$

$$\mathbf{e}_i = M_i\mathbf{y}_i. \tag{54}$$

As is well known

$$s_{ii} = \frac{\mathbf{e}_i'\mathbf{e}_i}{T - K} \tag{55}$$

is an unbiased estimator of σ_{ii} so that

$$\frac{1}{n}\sum s_{ii}(X_i'X_i)^{-1} \tag{56}$$

is an unbiased estimator of $(1/N)\sum^{\cdot} P_i$. In view of (52) and (56), an unbiased estimator of Δ is

$$\hat{\Delta} = \left[\frac{S_b}{n - 1} - \frac{1}{n}\sum \hat{P}_i\right]\frac{N - 1}{N}, \tag{57}$$

where

$$\hat{P}_i = s_{ii}(X_i'X_i)^{-1}. \tag{58}$$

Thus an unbiased estimate of S_{bb} will be

$$\hat{S}_{bb} = \hat{\Delta}\frac{(N-n)}{n(N-1)} + \frac{1}{n^2}\sum \hat{P}_i. \tag{59}$$

A possible operational difficulty with the estimator for Δ, $\hat{\Delta}$, is that it may not be positive definite or even positive semi-definite. A *necessary condition* for $\hat{\Delta}$ to be positive semi-definite is $n > K$.[16] However, this difficulty does not extend to the estimator for S_{bb}.

An Approximate Aitken Estimator

Assuming that the estimate of Δ is positive definite, we can create an estimator for **β** which uses more of the model specification than the average least square estimator, **b**. This Aitken estimator has the property that it will be dependent on the *particular* X matrix which is drawn. To form this estimator of **β** we follow Swamy (1971, Chapter 4), and write the sample system of nT observations (37) together as

$$\mathbf{y} = X\boldsymbol{\beta} + D(X)\boldsymbol{\delta} + \mathbf{u}, \tag{60}$$

where

$$\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_n')'$$

$$X = [X_1', X_2', \ldots, X_n']'$$

$$D(X) = \begin{bmatrix} X_1 & 0 & 0 & \ldots & 0 \\ 0 & X_2 & 0 & \ldots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \ldots & & X_n \end{bmatrix}^{17}$$

$$\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \ldots, \boldsymbol{\delta}_n')',$$

$$\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_n')'.$$

Conditional on X the $nT \times 1$ disturbance vector for (60), $D(X)\boldsymbol{\delta} + \mathbf{u}$, has the following variance covariance matrix

$$E[\{D(X)\boldsymbol{\delta} + \mathbf{u}\}\{D(X)\boldsymbol{\delta} + \mathbf{u}\}'|X] \equiv H(\boldsymbol{\theta}) =$$

$$\begin{bmatrix} X_1\Delta X_1' + \sigma I & -zX_1\Delta X_2' & \ldots & -zX_1\Delta X_n' \\ -zX_2\Delta X_1' & X_2\Delta X_2' + \sigma I & \ldots & -zX_2\Delta X_n' \\ \vdots & & & \vdots \\ -zX_n\Delta X_1' & & \ldots & X_n\Delta X_n' + \sigma I \end{bmatrix}, \tag{61}$$

where $z = 1/(N-1)$ and $\sigma = (1/N)\sum \sigma_{ii}$. The matrix $H(\boldsymbol{\theta})$ is a symmetric $nT \times nT$ matrix. It is functionally dependent on X, z and an unknown $\frac{1}{2}[K(K+1)+2]$

[16] See Schmalensee (1972, p. 6) for a proof of this result for Swamy's specification of the random coefficient model, Swamy (1971, Chapter 4). That proof carries over to our specification.
[17] The zeroes in D are $T \times K$ null matrices.

vector of parameters, $\boldsymbol{\theta}$, containing the distinct elements of Δ and σ arranged in a particular order. It can readily be shown that $H(\boldsymbol{\theta})$ has an inverse.[18] Conditional on X, the BLUE of $\boldsymbol{\beta}$ is the Aitken estimator,

(62) $$\mathbf{b}(\boldsymbol{\theta}) = (X'H(\boldsymbol{\theta})^{-1}X)^{-1}X'H(\boldsymbol{\theta})^{-1}\mathbf{y}.$$

Since Δ and σ are unknown, $\mathbf{b}(\boldsymbol{\theta})$ is not operational. We can, however, form an approximate Aitken estimator by substituting unbiased estimates for Δ and σ. Thus let $H(\hat{\boldsymbol{\theta}})$ be the $nT \times nT$ matrix formed by substituting $\hat{\Delta}$ for Δ and $s = (1/n)\sum s_{ii}$ for σ into $H(\boldsymbol{\theta})$. The approximate Aitken estimator is

(63) $$\mathbf{b}(\hat{\boldsymbol{\theta}}) = (X'H(\hat{\boldsymbol{\theta}})^{-1}X)^{-1}X'H(\hat{\boldsymbol{\theta}})^{-1}\mathbf{y}.$$

We conjecture that under fairly general conditions $\mathbf{b}(\hat{\boldsymbol{\theta}})$ will have desirable asymptotic properties.[19]

3.3. *Random Sampling Without Replacement With Unequal Probabilities*

We now generalize from simple random sampling to random sampling without replacement with unequal probabilities. We again consider two estimators: a simple weighted average of the least-squares estimators and an approximate Aitken estimator.

We make the following assumption

Assumption 3.2:

(1)–(3) the same as Assumption 3.1 (1)–(3).

(4) Sampling is done without replacement with unequal probabilities. π_i will be the overall probability that the ith unit is drawn and π_{ij} the joint probability that the i and jth units are drawn.

Weighted Average of Least Squares

From (35) it follows that a natural estimator for $\boldsymbol{\beta}$ is a simple weighted average of the least-squares estimators, where the weights are inversely proportional to the probability of being selected in the sample. That is, consider the estimator $\mathbf{b}^*$,

(64) $$\mathbf{b}^* = \frac{1}{N}\sum \frac{\mathbf{b}_i}{\pi_i}.$$

Using (29) and (46) we find

$$NE(\mathbf{b}^*) = E_S\left[\sum \frac{E_e(\mathbf{b}_i|S)}{\pi_i}\right] = E_s\left[\sum \frac{\boldsymbol{\beta}_i}{\pi_i}\right] = \sum \boldsymbol{\beta}_i$$

[18] See appendix.

[19] See Swamy (1971), (1972) for a discussion of large sample properties when N is infinite. His analysis needs to be modified for our work. However, much of his analysis does carry over to the present problem. For T sufficiently large with n fixed, we can treat $\mathbf{b}_i$ $(i = 1, 2, \ldots, n)$, as if they were sample of size n from the population of $\boldsymbol{\beta}$'s, i.e., $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_N)$. Then we can combine the result with the central limit results of Hajek (1960) for finite populations, to get the full set of asymptotic properties of $\mathbf{b}(\hat{\boldsymbol{\theta}})$. Also, see Theil (1971, p. 399). If $\mathbf{u}$ and $\boldsymbol{\delta}$ are symmetrically distributed about the null vector, then we can use the type of argument developed by Kakwani (1967) to show that $\mathbf{b}(\hat{\boldsymbol{\theta}})$ is an unbiased estimator of $\boldsymbol{\beta}$.

so that $\mathbf{b}^*$ is an unbiased estimator of $\boldsymbol{\beta}$. Let $S_{b^*b^*}$ be the variance-covariance matrix for $\mathbf{b}^*$. Evaluating $S_{b^*b^*}$ we find

$$S_{b^*b^*} = \frac{1}{N^2}\left[\sum \frac{P_i}{\pi_i} + \sum \frac{\boldsymbol{\beta}_i\boldsymbol{\beta}_i'(1 - \pi_i)}{\pi_i} + \sum \frac{\boldsymbol{\beta}_i\boldsymbol{\beta}_j'(\pi_{ij} - \pi_i\pi_j)}{\pi_i\pi_j}\right]. \tag{65}$$

By inspection of (65) we recognize that an unbiased estimator of $S_{b^*b^*}$ is

$$\hat{S}_{b^*b^*} = \frac{1}{N^2}\left[\sum \frac{(1 - \pi_i)\mathbf{b}_i\mathbf{b}_i'}{\pi_i^2} + \sum \frac{\mathbf{b}_i\mathbf{b}_j'(\pi_{ij} - \pi_i\pi_j)}{\pi_{ij}\pi_j\pi_j} + \sum \frac{\hat{P}_i}{\pi_i}\right]. \tag{66}$$

An Approximate Aitken Estimator

We now develop an approximate Aitken estimator for this model. As before the analysis is conditioned on X.

To construct the Aitken procedure we would like to write an observation at, say, the rth draw as

$$\mathbf{y}_r = X_r\boldsymbol{\beta} + \mathbf{v}_r \tag{67}$$

where the disturbance $\mathbf{v}_r$ satisfies

$$E[\mathbf{v}_r|X] = \mathbf{0}. \tag{68}$$

However, for random sampling without replacement with unequal probabilities,

$$v_r = X_r(\boldsymbol{\beta}_r - \boldsymbol{\beta}) + \mathbf{u}_r$$

and

$$E[\mathbf{v}_r|X] = X_r\left(\sum_i \boldsymbol{\beta}_i p_i(r) - \beta\right). \tag{69}$$

Note that the expected value of $\mathbf{v}_r$ will not vanish unless $p_i(r) = 1/N$, i.e., we engage in simple random sampling. To avoid this problem we transform each draw in the following way. If the lth unit in the population is chosen on the rth draw write

$$e_r = N^{-1}p_l(r)^{-1}$$

and let

$$\tilde{\mathbf{y}}_r = \mathbf{y}_r e_r, \qquad \tilde{\boldsymbol{\beta}}_r = \boldsymbol{\beta}_r e_r, \qquad \tilde{\mathbf{u}}_r = \mathbf{u}_r e_r.$$

The transformed representation of the rth draw is then

$$\tilde{\mathbf{y}}_r = X_r\tilde{\boldsymbol{\beta}}_r + \tilde{\mathbf{u}}_r \tag{70}$$

and the expected value of $\tilde{\boldsymbol{\beta}}_r = \boldsymbol{\beta}$. The difficulty with this particular transformation is that the variance-covariance matrix for the transformed system of n draws depends on the draw-by-draw probabilities, the $p_i(r)$ and $p_{ij}(r, s)$ terms. To circumvent this complication we assume that the sample design satisfies the following equations,[20]

[20] If we interpret all quantities as referring to a particular stratum then whenever the number sampled (n) within a stratum is small relative to the number of units in the stratum (N), equations (71) and (72) are likely to be adequate approximations (within the stratum). See Cochran (1962, p. 260–262) for a description of a common method for selecting units with unequal probabilities but without replacement which will approximately satisfy these equations within a stratum. In this case the approximate Aitken estimator developed in the text will be defined for each stratum. An estimate of the overall population mean for all strata taken together can then be formed by suitably averaging the estimates from the different strata.

(71) $$p_i(r) = \frac{\pi_i}{n} \qquad \text{for all } r \in \bar{n} \text{ and } i \in \bar{N}$$

(72) $$p_{ij}(r, s) = \frac{\pi_{ij}}{n(n-1)} \qquad \text{for all } r, s \in \bar{n}, r \neq s \text{ and } i, j \in \bar{N}, i \neq j.$$

We now analyze the transformed system of equations having the form of equation (70) for all $r \in \bar{n}$, where $e_r = n/\pi_l N$ when the lth unit is chosen at the rth draw. The following results will be useful in this analysis. From (71) and (72) we can easily show that for draws r and s, $r \neq s$,

(73) $$E_S(z_r) = \frac{1}{n} \sum \pi_i z_i$$

(74) $$E_S(z_r z_s) = \frac{1}{n} \frac{1}{(n-1)} \sum \pi_{ij} z_i z_j.$$

From (73) we find

(75) $$E_S[\tilde{\boldsymbol{\beta}}_r] = \sum_i \tilde{\boldsymbol{\beta}}_i p_i(r) = \sum_i \frac{\boldsymbol{\beta}_i}{e_i} \frac{\pi_i}{n} = \frac{1}{N} \sum_i \boldsymbol{\beta}_i = \boldsymbol{\beta}.$$

Let $\boldsymbol{\delta}_r$ be the sampling error in the transformed random coefficient $\tilde{\boldsymbol{\beta}}_r$

(75) $$\boldsymbol{\delta}_r = \tilde{\boldsymbol{\beta}}_r - \boldsymbol{\beta}, \qquad r \in \bar{n}.$$

By construction

$$E_S[\boldsymbol{\delta}_r] = 0 \qquad r \in \bar{n}.$$

Each $\boldsymbol{\delta}_r$ will have the same variance-covariance matrix, say $\tilde{\Delta}$.

$$\tilde{\Delta} = E\boldsymbol{\delta}_r\boldsymbol{\delta}_r' = E\tilde{\boldsymbol{\beta}}_r\tilde{\boldsymbol{\beta}}_r' - \boldsymbol{\beta}\boldsymbol{\beta}'.$$

Evaluating $\tilde{\Delta}$ gives

(76) $$\tilde{\Delta} = \sum \frac{\boldsymbol{\beta}_i\boldsymbol{\beta}_i'(n - \pi_i)}{N^2\pi_i} - \frac{1}{N^2} \sum \boldsymbol{\beta}_i\boldsymbol{\beta}_j.$$

We assume $\tilde{\Delta}$ is positive definite. By inspection of (76) we infer that an unbiased estimator of $\tilde{\Delta}$ is

(77) $$\hat{\tilde{\Delta}} = \sum \frac{\mathbf{b}_i\mathbf{b}_i'(n - \pi_i)}{N^2\pi_i^2} - \sum \frac{\hat{P}_i(n - \pi_i)}{N^2\pi_i^2} - \frac{1}{N^2} \sum \frac{\mathbf{b}_i\mathbf{b}_j'}{\pi_{ij}}.$$

The covariance between $\boldsymbol{\delta}_r$ and $\boldsymbol{\delta}_s$, say $\tilde{\Delta}_c$, will be identical for all $r \neq s$ and satisfy

$$\tilde{\Delta}_c = -\frac{\tilde{\Delta}}{N-1} \qquad \text{for all } r \text{ and } s \in \bar{n}, r \neq s.$$

Using the foregoing results, the system of nT observations may be written as

(78) $$\tilde{\mathbf{y}} = X\boldsymbol{\beta} + D(X)\boldsymbol{\delta} + \tilde{\mathbf{u}}$$

where

$$\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \ldots, \tilde{\mathbf{y}}'_n)'$$

$$\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}'_1, \tilde{\mathbf{u}}'_2, \ldots, \tilde{\mathbf{u}}'_n)'$$

$$\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, \ldots, \boldsymbol{\delta}'_n)'$$

and X and $D(X)$ are given beneath (60). Given X the disturbance in (78) has a variance-covariance matrix $G(\boldsymbol{\varphi})$

(79)
$$G(\boldsymbol{\varphi}) = \begin{bmatrix} X_1\tilde{\Delta}X'_1 + \tilde{\sigma}I & X_1\tilde{\Delta}_c X'_2 & \ldots & X_1\tilde{\Delta}_c X'_n \\ X_2\tilde{\Delta}_c X'_1 & X_2\tilde{\Delta}X'_2 + \tilde{\sigma}I & \ldots & X_2\tilde{\Delta}_c X'_n \\ \vdots & & & \vdots \\ X_n\tilde{\Delta}_c X' & & \ldots & X_n\tilde{\Delta}X'_n + \tilde{\sigma}I \end{bmatrix}$$

where $\boldsymbol{\varphi}$ contains the distinct unknown parameters elements of $\tilde{\Delta}$ and $\tilde{\sigma}$, with $\tilde{\sigma} = (n/N^2)\sum(\sigma_{ii}/\pi_i)$.

If Δ and $\tilde{\sigma}$ were known,

$$\mathbf{b}(\boldsymbol{\varphi}) = (X'G(\boldsymbol{\varphi})^{-1}X)^{-1}X'G(\boldsymbol{\varphi})^{-1}\mathbf{y}^*$$

would be the BLUE of $\boldsymbol{\beta}$. An approximate Aitken estimator may be formed by substituting $\hat{\tilde{\Delta}}, \hat{\tilde{\Delta}}_c = \hat{\tilde{\Delta}}/(N-1)$, and $\hat{\tilde{\sigma}} = (n/N^2)\sum(s_{ii}/\pi_i^2)$ for $\tilde{\Delta}$, $\tilde{\Delta}_c$ and $\tilde{\sigma}$ into $G(\boldsymbol{\varphi})$ to obtain $G(\hat{\boldsymbol{\varphi}})$; the estimator is

(80)
$$b(\hat{\boldsymbol{\varphi}}) = (X'G(\hat{\boldsymbol{\varphi}})^{-1}X)^{-1}X'G(\hat{\boldsymbol{\varphi}})^{-1}\tilde{\mathbf{y}}.$$

If $\hat{\tilde{\Delta}}$ is not positive definite (or at least positive semidefinite) we face a negative variance problem.[21] There does not appear to be an easy solution to the negative variance problem. One can never be sure whether or not the result arises because of a model misspecification or is just an anomaly of a given sample.

An Extension

It is not difficult to see how these results may be generalized to permit contemporaneous correlation between **u**'s in the population. That is, consider *Assumption 3.3.*

(1), (2), (4) same as corresponding conditions in Assumption 3.2.

(3) The disturbance vectors $\mathbf{u}_i$, $(i \in \bar{N})$ each have mean zero and $E\mathbf{u}_i\mathbf{u}'_j = \sigma_{ij}I$ for all i and j.

The correct unbiased estimator of $\tilde{\Delta}$ becomes

(81)
$$\hat{\tilde{\Delta}} = \sum \frac{\mathbf{b}_i\mathbf{b}'_i(n-\pi_i)}{N^2\pi_i^2} - \sum\frac{\hat{P}_i(n-\pi_i)}{N^2\pi_i^2} - \frac{1}{N^2}\sum\frac{\mathbf{b}_i\mathbf{b}'_j}{\pi_{ij}} + \frac{1}{N^2}\sum\frac{s_{ij}\hat{P}'_iX'_iX_j\hat{P}_j}{s_{ii}s_{jj}\pi_{ij}}.$$

where

$$s_{ij} = \mathbf{y}'_iM_iM_j\mathbf{y}_j/\text{trace}(M_iM_j)$$

[21] See Swamy (1971) and Schmalensee (1972) for discussions of this problem and additional references.

The matrix $G(\varphi)$ and therefore $G(\hat{\varphi})$ changes also for Assumption 3.3. The iith block diagonal matrix is still

$$X_i \tilde{\Delta} X_i' + \tilde{\sigma} I$$

but the ijth off-diagonal matrix becomes

$$X_i \tilde{\Delta}_c X_j' + \tilde{\sigma}_c I, \qquad \text{where } \tilde{\sigma}_c = \frac{1}{n(n-1)} \sum \pi_{ij} \sigma_{ij} e_i e_j.$$

4. Future Extensions

In this paper we explore the consequences of using information on the design of a sample survey to estimate population averages in a linear model. An analysis of the sampling properties of the alternative estimators considered awaits further study.

Finally, we treat the sample design as being given exogenously. It may prove illuminating to relax this assumption and rank alternative sample designs on the basis of their precision in estimating population averages in a linear model.

Economist, Board of Governors
Federal Reserve System

Appendix: Inverse of $H(\theta)$

$H(\theta)$ may be written as

(1)
$$H(\theta) = D[Z \otimes \Delta]D' + \sum \otimes I$$
$$= R + DBD'$$

where $\otimes$ is the Kronecker product symbol.

$$R = \sum \otimes I_T, \qquad \sum = \sigma I_n, \qquad B = Z \otimes \Delta,$$

and $Z = (z_{ij})$ is an equicorrelated matrix with $z_{ii} = 1$,

$$z_{ij} = -z, \; i \neq j.$$

Since Δ is positive definite (by assumption) Δ^{-1} exists. The inverse of Z is readily found, see Rao (1965, p. 53, problem 2(ii)).

Now

(2)
$$R^{-1} = \sum^{-1} \otimes I_T$$

(3)
$$B^{-1} = Z^{-1} \otimes \Delta^{-1}.$$

Finally, using a result given in Rao (1965, p. 29, problem 29), we find

(4)
$$(R + DBD')^{-1} = R^{-1} - R^{-1}D(D'R^{-1}D)^{-1}D'R^{-1}$$
$$+ R^{-1}D(D'B^{-1}D)^{-1}((D'R^{-1}D)^{-1} + B)^{-1}(D'R^{-1}D)^{-1}DR'^{-1}.$$

Inspecting the r.h.s. of (4) we note that in view of (2) and (3), the largest matrix to be inverted is nK by nK. If Δ is positive semidefinite, $H(\theta)$ is also nonsingular.

Champernowne, D. G. (1969). *Uncertainty and Estimation in Economics*. Volume Two. San Francisco: Holden-Day.

Cochran, W. G. (1963). *Sampling Techniques*. Second Edition. New York: John Wiley and Sons.

Cohen, M. S., Rea, A. and Lerman, R. I. (1970). *A Micro Model of Labor Supply*. BLS Staff Paper 4. Government Printing Office.

Cramer, J. S. (1971). *Empirical Econometrics*. Amsterdam: North-Holland.

Dhrymes, P. J. (1970). *Econometrics*. New York: Harper and Row.

Fleisher, B. M. and Porter, R. D. (1970). *The Labor Supply of Males 45–59: A Preliminary Report*. Columbus: Center for Human Resource Research, The Ohio State University.

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

——— (1968). *Topics in Regression Analysis*. New York: The MacMillan Company.

Hajek, J. (1960). Limiting Distributions in Simple Random Sampling From a Finite Population. *Pub. Math. Inst. Hungarian Acad. Sci.*, 5, 361–374.

Hildreth, C. and Houck, J. P. (1968). Some Estimators for a Linear Model with Random Coefficients. *Journal of the American Statistical Association* 63, 584–595.

Hodges, J. L. and Lehmann, E. L. (1970). *Basic Concepts of Probability and Statistics*. Second Edition. San Francisco: Holden-Day.

Hu, T. W. and Stromsdorfer, E. W. (1970). A Problem of Weighting Bias in Estimating the Weighted Regression Model. *Proceedings of the Business and Economics Section of the American Statistical Association*, 513–516.

Johnston, J. (1963). *Econometric Methods*. New York: McGraw-Hill.

Kakwani, N. C. (1967). The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators. *Journal of the American Statistical Association*, 62, 141–142.

Kendall, M. G. and Stuart, A. (1966). *The Advanced Theory of Statistics*. Volume Three. New York: Hafner Publishing Company.

Klein, L. R. and Morgan, J. (1951). Results of Alternative Statistical Treatments of Sample Survey Data. *Journal of the American Statistical Association*, 46, 442–460.

Klein, L. R. (1953). *A Textbook of Econometrics*. Evanston: Row, Peterson and Company.

Kmenta, J. (1971). *Elements of Econometrics*. New York: The MacMillan Company.

Konijn, H. S. (1962). Regression Analysis in Sample Surveys. *Journal of the American Statistical Association*, 57, 590–606.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society*, B, 34, 1–41.

Malinvaud, E. (1966). *Statistical Methods of Econometrics*. Chicago: Rand McNally and Company.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons.

Roth, J. (1971). Application of Regression Analysis to Sample Survey Data. Unpublished memorandum, Social Security Administration.

Schmalensee, R. (1972). Variance Estimation in a Random Coefficient Regression Model. Report #72-10. Department of Economics, University of California, San Diego.

Swamy, P. A. V. B. (1968). Statistical Inference in a Random Coefficient Regression Model. Ph.D. thesis. University of Wisconsin, Madison.

——— (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, 38, 311–323.

——— (1971). *Statistical Inference in Random Coefficient Regression Models*. New York: Springer-Verlag.

——— (1972). Linear Models with Random Coefficients. To be published in Zarembka, P. (editor), *Frontiers in Econometrics*. New York: Academic Press.

Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley and Sons.

Zellner, A. (1966). On the Aggregation Problem: A New Approach to a Troublesome Problem. Report #6628. Center for Mathematical Studies in Business and Economics, University of Chicago. Published in Fox, K. A., et al. (editors), *Economic Models, Estimation and Risk Programming*: Essays in Honor of Gerhard Tintner. New York: Springer-Verlag, 1969.

——— (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons.