ON THE STATISTICAL THEORY OF UNOBSERVED COMPONENTS

Robert E. Hall

Number 117 August 20, 1973

The research reported here was carried out while the author was on a leave supported by the Social Science Research Council. Support was also provided by the National Science Foundation. Edward Leamer and Gary Chamberlain made useful comments on an earlier version. I am especially indebted to Jerry Hausman for lengthy discussions and for pointing out the key reference, Karlin and Studden (1966).

1. Introduction

Statistical models of unobserved components seem destined for an increasing role in econometric work. Especially in cross-sections, the differences in the values of the left-hand variables among observations with identical values of the right-hand variables are sufficiently large to justify careful analysis of the apparently random component of the behavior under study. The simple characterization of randomness implicit in the stochastic specification of the regression model seems inadequate when the right-hand variables in a problem account for only a small portion of the dispersion of the left-hand variable. Many recent authors have sought to attribute part of the randomness in their samples to variations within the population of characteristics that are not observed. For example, Griliches (1973) assigns part of the dispersion of earnings conditional on education to the unobserved differences in ability of individuals with equal amounts of education. Domencich and McFadden (1974) hypothesize a distribution of tastes within the population to explain choices of modes of transportation by individual commuters. The present paper takes up the following question: What can be discovered about the underlying distribution of characteristics from the observed body of data? Are the assumptions about the distributions of unobserved characteristics made by previous authors verifiable, or must they be accepted on pure faith?

A general statistical model suitable for this discussion is the following

$$y = h(x,\theta,u) \tag{1.1}$$

where y is the scalar left-hand variable, assumed to be qualitative (taking on only a finite number of integer values[1]), x is a vector of observed characteristics, θ is the unobserved characteristic, and u is a disturbance whose distribution may depend on x and θ. Apart from the presence of θ, this would be a regression model if the distribution of u did not depend on x and θ; in the qualitative case especially, however, this dependence is critical. Our discussion concerns the untangling of the separate effects of θ and u, where the role of x is subsidiary, so until Section 6 we consider the case of sampling from a population whose members are observationally identical, where it is appropriate to suppress x:

$$y = h(\theta,u) \tag{1.2}$$

All observations from the same individual are assumed to correspond to the same θ, but each one involves a new drawing from the distribution of u. Finally, we assume prior knowledge of $h(\theta,u)$ and of the distribution of u. The last assumption should become more plausible as the discussion progresses.

Models of unobserved components are particularly important in the study of the distribution of income. The major theme of the most influential recent work on income distribution, Christopher Jenck's book, Inequality (1972), is exactly that observed differences among individuals account for very little of the dispersion of income among them: "Neither

---

[1]If the left-hand variable is continuous, y can be defined by a set of intervals of values of the variable.

family background, cognitive skill, educational attainment, nor occupational status explains much of the variation in men's incomes. Indeed, when we compare men who are identical in all these respects, we find only 12 to 15 percent less inequality than among random individuals. How are we to explain these variations among men who seem to be similarly situated?" (p. 227). Jencks replies that unmeasured differences in motivation, ability, and especially luck account for the bulk of the dispersion in income. His discussion is limited by his failure to distinguish between unobserved differences among individuals, on the one hand, and differences in the experience of the same individual at different points in time, on the other. In the context of measuring income, this distinction is familiar to economists in Milton Friedman's notion of the permanent and transitory components of measured income. Jencks alludes briefly to the distribution of permanent income (footnote 1, p. 233) but the distinction has no role in his discussion.

The class of statistical models studied here provides a general framework for separating the two sources of the apparently random differences among individuals at a point in time. Systematic differences among individuals are indexed by the random variable θ, and differences in the experiences of a single individual by the random variable u. Friedman's model is a special case of the general model in which θ and u are simply added together

$$y = \theta + u \qquad (1.3)$$

Here θ is permanent income and u is transitory income. If y is observed for a few successive years, then it is tempting to estimate permanent income for an individual as the average income over the years:

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{1.4}$$

The difficulty is that the distribution of $\hat{\theta}$ among the members of the population has more dispersion than the distribution of θ. This problem arises most critically in Jencks' data, where T is 1, but even where T is 3 or 4 one does not know how much the distribution of $\hat{\theta}$ tells about the distribution of θ. What is needed, and what this paper supplies, is a method for extracting as much reliable information as possible about the distribution of θ.

One of the most carefully studied models of the kind treated in this paper is the mover-stayer model used by sociologists in analyzing industrial and other forms of mobility. Most mover-stayer models posit only two kinds of individuals, movers and stayers (two values of θ in the notation of this paper). In these models, the randomness in individual experience over time governed by u follows a Markov process. In Section 5, we indicate briefly how the methods of this paper can be applied to a rather general version of the mover-stayer model.

2. Mixtures of Probabilities

Suppose that for an individual of type θ, the distribution of y is the vector of probabilities $a(\theta)$:

$$\text{Prob}\ [y = i \mid \theta] = a_i(\theta) \tag{2.1}$$

We observe the average of this probability over all individuals:

$$\phi_i = \text{Prob}\ [y = i] = \int_0^1 a_i(\theta)dF(\theta) \quad , \qquad (2.2)$$

where $F(\theta)$ is the cumulative distribution of types of individuals in the population, that is, the fraction whose type is lower than θ. There is a substantial statistical literature dealing with problems of this form. In the vocabulary of that literature, equation 2.2 is a mixture. The distribution $a(\theta)$ is the kernel and $F(\theta)$ is the mixing distribution. A survey of the statistical theory of mixtures appears in Maritz (1970), Chapter 2. In addition, there is an important body of mathematical thought about problems of the sort considered here. In the mathematical literature, equation 2.2 is called a Tchebycheff system (see Karlin and Studden (1966), Chapters I through V). It appears that statistical and mathematical work in this area has proceeded almost completely independently. The mathematical theory is substantially more general and more fully developed, so it forms the basis for this paper.

Our problem is to obtain information about the distribution of the unobserved component, $F(\theta)$, given the observed probability ϕ and the known kernel $a(\theta)$. In this section we present theorems that give a fairly precise characterization of the limits of knowledge about $F(\theta)$. Most of these theorems are simply re-interpretations of results of Krein (1951) and other mathematical students of Tchebycheff systems.

We begin with the

> Assumption of Distinct Types: The matrix $[a(\theta_1),\ldots,a(\theta_M)]$ has rank M for any distinct set of types $\theta_1,\ldots,\theta_M$.

This assumption is the defining characteristic of a Tchebycheff system.

It rules out models where the probabilities associated with one particular type of individual can be expressed as a linear combination of the probabilities associated with M - 1 or fewer other types. This assumption does not seem unduly strong, and it is satisfied by the applications studied in this paper.

Next we define two useful constructions. First,

$$\Phi = \{\phi \text{ satisfying } \phi = \int_0^1 a(\theta)dF(\theta) \text{ for some } F(\theta)\}. \tag{2.3}$$

Here we consider all $F(\theta)$ that are non-decreasing, continuous from the left, and have a finite number of discontinuities. Φ is the set of all possible observed probabilities consistent with a given problem as defined by $a(\theta)$. Second,

$$V(\phi) = \{F(\theta) \text{ satisfying } \int_0^1 a(\theta)dF(\theta) = \phi\}. \tag{2.4}$$

$V(\phi)$ is the set of all distributions of unobserved types in the population that are consistent with a particular observed probability, ϕ. The essence of the problem is that $V(\phi)$ may contain a variety of distributions. Our characterization of the limits of knowledge about $F(\theta)$ deals, therefore, with the extremal members of $V(\phi)$.

The first theorem establishes that no observed probability proves that there are more than $(M + 2)/2$ different types in the population (proofs and references appear in the appendix to this section):

> Theorem 2.1: For any $\phi \in \Phi$, there exists a cdf, $F(\theta) \in V(\phi)$, with no more than $(M + 2)/2$ points of increase.

If we let f_j be the mass at one of the points of increase, θ_j, then Theorem 2.1 shows that it is always possible that ϕ is a discrete mixture:

$$\phi = \sum_{j=1}^{N} f_j \; a(\theta_j) \tag{2.5}$$

with $N \leq (M + 2)/2$. Here f_j is the fraction of the population having type θ_j. This result appears independently in the statistical literature on mixtures in the form of an identification theorem: Given ϕ, one can calculate unique f_j and θ_j satisfying

$$\sum_{j=1}^{N} f_j \; a(\theta_j) = \phi \tag{2.6}$$

only if $N \leq 1 + M/2$. See Teicher (1963), p. 1269.

The second theorem shows that for any observed ϕ (with one class of exceptions) we cannot rule out the possibility that a positive fraction of the population has an arbitrary type, θ^*:

> <u>Theorem 2.2</u>: Suppose ϕ is in the interior of Φ and suppose θ^* is an arbitrary type in $[0,1]$. Then there is a cdf, $F(\theta)$, in $V(\phi)$ with positive mass $\rho(\theta^*)$ at θ^*.

This result imposes a limitation on the form of knowledge about F that we can deduce from ϕ: Except in borderline cases, we will never be able to state that any particular type, or any range of types, is non-existent in the population. On the other hand, $\rho(\theta^*)$ may be close to zero; the theorem does not prevent us from finding useful bounds on the fraction of the population of a certain type or range of types.

The next theorem provides a bound on the fraction of the population of type θ^*:

<u>Theorem 2.3</u>: Consider the problem of finding probabilities $f_1,\ldots,f_N$ and types $\theta_J,\ldots,\theta_N$ obeying

$$\sum_{j=1}^{N} f_j\ a(\theta_j) = \phi \qquad (2.7)$$

where $\theta_1 = \theta^*$ and N, θ_2, θ_3 and J take on one of the following sets of values: If M is odd, either N = 1 + (M - 1)/2, J = 2, or N = 2 + (M - 1)/2, J = 4, $\theta_2 = 0$, $\theta_3 = 1$; if M is even, N = 1 + M/2, J = 3, and either $\theta_2 = 0$ or $\theta_2 = 1$. Then this system has a unique solution and f_1 is the maximal mass at θ^* for any $F \in V(\phi)$.

Thus the problem of finding the distribution of types that is most concentrated at θ^* is simply one of solving a system of M equations in M unknowns: N values of f_j and N - J + 1 values of θ_j. The solution is called the canonical representation of ϕ involving θ^*.

A related problem is to find bounds on the fraction of the population whose type is less than some value θ^*:

<u>Theorem 2.4</u> (Markov-Krein Theorem):

$$\sum_{j \ni \theta_j < \theta^*} f_j \le F(\theta^*) \le \sum_{j \ni \theta_j \le \theta^*} f_j$$

$$\text{for all } F \in V(\phi) \qquad (2.8)$$

where f_j and θ_j are the canonical representation involving θ^*.

The Markov-Krein theorem shows that the canonical representation is extremal not only with regard to the mass at θ^* but also with regard to the mass below θ^*. The upper and lower bounds on $F(\theta^*)$ differ by precisely the maximal mass, f_1.

Unfortunately, the mathematical theory of Tchebycheff systems does not provide bounds on the fraction of the population between two arbitrary types. We would like to be able to answer the following question: Suppose we have a pair of types θ_L and θ_H, and we let $P = F(\theta_H) - F(\theta_L)$, the fraction of the population between θ_L and θ_H. What are the largest and smallest values of P consistent with a particular ϕ? The Tchebycheff inequality answers this question for the particular case where ϕ gives the first two moments of $F(\theta)$. There is an extensive mathematical literature on generalizations of the Tchebycheff inequality (summarized in detail in Karlin and Studden (1966), chapters XII-XIV), but it does not contain any results of sufficient generality for our purposes. Mathematicians have been concerned exclusively with sharp bounds on P, that is, bounds that are attained by some $F \in V(\phi)$, or at least that are approached arbitrarily closely by members of $V(\phi)$.

Before going on to our approach to the problem of bounds on the probability in an interval on the θ axis, which involves non-sharp bounds, we need to deal with the fundamental problem of identifiability. What conditions are required for it to be possible to find out anything about the fraction of the population in an interval? There has been a good deal of work on the identifiability of mixtures (see Maritz (1970), pp. 20-35), all using a strict definition of identifiability: A mixing distribution is said to be identifiable if its exact form can be deduced from the value of ϕ. Strong assumptions about $F(\theta)$ are required for identifiability. A leading result in the statistical literature has already appeared here as Theorem 2.1.

A much weaker notion of identifiability seems appropriate in this paper:

> Definition: The probability P is identifiable if there is some ϕ in the interior of Φ such that $V(\phi)$ contains no distributions with $P = 0$.

We gain information about P if we can show that it is positive, that some fraction of the population has types between θ_L and θ_H. A problem, as defined by $a(\theta)$, has an identifiable P if there is some observed outcome ϕ for which P must be positive. It is a remarkable fact that no additional assumptions are needed to ensure identifiability in a Tchebycheff system:

> Theorem 2.5: Every P is identifiable.

Appendix to Section 2

General remark. Proofs of the results in this section are all taken from Karlin and Studden (1966) (hereafter K & S). They deal with a somewhat more general problem in which $F(\theta)$ is not required to obey $\int_0^1 dF(\theta) = 1$ and $a(\theta)$ is not required to satisfy $\Sigma a_i(\theta) = 1$. In their exposition, Φ is a convex cone, while under our assumptions it is a convex subset of the unit simplex. However, the results invoked here apply without modification, because our Φ is simply the intersection of their Φ and the unit simplex.

Proof of Theorem 2.1: If ϕ is on the boundary of Φ, apply Theorem II.2.1, K & S. Otherwise, apply their Corollary II.3.1. If M is odd, $N = (M + 1)/2$.

Proof of Theorem 2.2: The appropriate cdf, $F(\theta)$, can be taken as defined in Theorem 2.3. K & S, Theorem II.3.1, establish that the mass is positive.

Proof of Theorem 2.3: K & S, Theorem II.4.1 (attributed to Krein (1951)), show that the canonical representation involving θ^* assigns maximal mass to θ^*. Existence and uniqueness of the canonical representation follow from their Theorem II.3.1 and Corollary II.3.2, respectively.

Proof of Theorem 2.4: See K & S, Theorem III.2.1.

Proof of Theorem 2.5: We need to exhibit a ϕ such that all $F \in V(\phi)$ have positive mass in the interval $[\theta_L, \theta_H]$. Define $\theta_k = \frac{N-k+1}{N}\theta_L + \frac{k-1}{N}\theta_H$. If M is odd, let $N = (M + 1)/2$ and

$$\phi = \frac{1}{N}\sum_{k=1}^{N} a(\theta_k); \tag{A2.1}$$

otherwise, let $N = (M + 2)/2$ and

$$\phi = \frac{1}{N+1} \sum_{k=1}^{N} a(\theta_k) + \frac{1}{N+1} a(0) \qquad (A2.2)$$

For this ϕ, the values of θ_k and $f_k = \frac{1}{N}$ or $\frac{1}{N+1}$ are a canonical representation. By K & S's Lemma II.3.1, every $F(\theta) \in V(\phi)$ assigns positive mass to $[\theta_{k-1},\theta_k]$, so clearly P must be positive. Finally, K & S's Theorem II.2.1 establishes that ϕ is in the interior of Φ.

3. Bounds for intervals

The theory of the previous section gives bounds on the fraction of the population within a prescribed interval only when the interval starts at 0 or ends at 1. In this section we discuss a method for deriving bounds for an interval beginning at $\bar{\theta}_L$ and ending at $\bar{\theta}_H$, that is, bounds on $F(\bar{\theta}_H) - F(\bar{\theta}_L)$ over all $F(\theta) \in V(\phi)$. Our strategy is the following: We define a set $V_O(\phi,D_N)$ that encloses $V(\phi)$; V_O contains all distributions consistent with ϕ and some others as well. V_O is mathematically tractable and from it we can derive "outside bounds" as the maximum and minimum of $F(\theta_H) - F(\theta_L)$ within V_O. These are true bounds on the fraction of the population between θ_L and θ_H, but they understate the amount of information available because they are taken over a set that includes false distributions. We show that as the index of computational effort, N, rises, more and more of the false distributions are excluded from V_O, and the bounds derived from it become sharper and sharper. In fact, as N approaches infinity, $V_O(\phi,D_N)$ approaches $V(\phi)$, and the bounds approach the sharp bounds taken over $V(\phi)$.

We also define a set $V_I(\phi,D_N)$ that is enclosed by $V(\phi)$. It contains no false distributions but excludes some true distributions, so the "inside bounds" derived from it are uniformly too optimistic. The reason for computing them is that the difference between the outside and inside bounds is a measure of the pessimism of the outside bounds. Again, the inside bounds converge to the exact bounds as N increases.

In constructing V_O and V_I, we make use of a partition, D_N, of the θ-axis:

$$D_N = \{[\bar{\theta}_0,\bar{\theta}_1],[\bar{\theta}_1,\bar{\theta}_2],\ldots,[\bar{\theta}_{N-1},\bar{\theta}_N]\} \tag{3.1}$$

where $\bar{\theta}_0 = 0$ and $\bar{\theta}_N = 1$. Throughout, we will consider an arbitrary sequence of D_Ns with the properties that higher numbered partitions are refinements of lower numbered ones:

$$D_N \subset D_{N+1} \text{ , all } N \tag{3.2}$$

and that the partition becomes finer and finer:

$$\lim_{N\to\infty} [\max_{j=1,\ldots,N} |\bar{\theta}_j - \bar{\theta}_{j-1}|] = 0 \tag{3.3}$$

Further, we require that the $\bar{\theta}_j$ include $\bar{\theta}_L$ and $\bar{\theta}_H$:

$$D_3 = \{[0,\bar{\theta}_L],[\bar{\theta}_L,\bar{\theta}_H],[\bar{\theta}_H,1]\} \tag{3.4}$$

We begin our derivation of the outer bounds by defining

$$\underline{a}_{i,j} = \min_{\bar{\theta}_{j-1} \le \theta \le \bar{\theta}_j} a_i(\theta) \tag{3.5}$$

$$\bar{a}_{i,j} = \max_{\bar{\theta}_{j-1} \le \theta \le \bar{\theta}_j} a_i(\theta) \tag{3.6}$$

Now

$$\int_0^1 a_i(\theta)dF(\theta) \ge \sum_{j=1}^{N} \int_{\bar{\theta}_{j-1}}^{\theta_j} \underline{a}_{i,j}dF(\theta) = \sum_{j=1}^{N} \underline{a}_{i,j}p_j \tag{3.7}$$

and

$$\int_0^1 a_i(\theta)dF(\theta) \le \sum_{j=1}^{N} \int_{\bar{\theta}_{j-1}}^{\bar{\theta}_j} \bar{a}_{i,j}dF(\theta) = \sum_{j=1}^{N} \bar{a}_{i,j}p_j \tag{3.8}$$

These are the lower and upper Stieltjes sums of the integral with respect to D_N (see Apostol (1957), p. 203). Defining $\underline{A}$ and $\bar{A}$ in the obvious way, we have

$$\underline{A}p \leq \phi \leq \bar{A}p \tag{3.9}$$

In addition, we require $\Sigma p_i = 1$ and $p_i \geq 0$, all i. The set of solutions to this system of inequalities, $S_0(\phi, D_N)$, is a convex polyhedron and can be represented most compactly as the convex hull of its vertices. Calculation of the vertices is discussed in the appendix to this section. The set $S_0(\phi, D_N)$ contains all the probabilities consistent with the original problem and possibly some others as well. Our next step is to compare the information about $F(\theta)$ contained in the computable $S_0(\phi, D_N)$ with the information in the uncomputable $V(\phi)$. To put $S_0(\phi, D_N)$ in a comparable form, we define

$$V_0(\phi, D_N) = \{F(\theta) \,|\, F(\bar{\theta}_j) - F(\bar{\theta}_{j-1}) = p_j \ , \ j = 1, \ldots, N \text{ for some } p \in S_0(\phi, D_N)\} \tag{3.10}$$

V_0 contains all the distributions that have the appropriate mass in each interval. Then, from the construction of V_0,

$$V(\phi) \subset V_0(\phi, D_N) \ , \text{ any } D_N \tag{3.11}$$

Our procedure understates the information available about $F(\theta)$, in that it suggests that some distributions are compatible with the observed probabilities ϕ when in fact they are not. It never makes the opposite mistake.

What are the costs and benefits of using a finer set of endpoints? The only costs are computational; adding a refinement to the endpoints can never reduce the precision of our bounds: $V_0(\phi,D_{N+1}) \subset V_0(\phi,D_N)$. Breaking an interval into two intervals not only helps localize the probability within the original interval but refines the bounds on other probabilities as well, by reducing the imprecision introduced in formulas 3.7 and 3.8.

Finally, we show that the true set $V(\phi)$ can be approximated arbitrarily closely by using a sufficiently large set of intervals:

$$V(\phi) = \bigcap_{N=3}^{\infty} V_0(\phi,D_N) \tag{3.12}$$

That is, if $F(\theta)$ is not in $V(\phi)$, there is some set of endpoints, D_N, such that the fact is revealed: $F(\theta)$ is not in $V_0(\phi,D_N)$ either.

The results of this section show that the mathematically simple $S_0(\phi,D_N)$ provides information about $F(\theta)$ that has a rigorous interpretation, becomes more precise as the set of endpoints becomes more refined, and converges to the information in the mathematically intractable $V(\phi)$.

Since the cost of computing $S_0(\phi,D_N)$ rises rather sharply with N, it is useful to have information about the amount of imprecision introduced by a given partition, D_N, to evaluate the prospective benefits of using a finer partition. For this purpose we develop a set of bounds that are known to be attained (and are usually exceeded) in $V(\phi)$. These bounds set a lower limit on the looseness of the outside bounds already discussed.

Among the members of $V(\phi)$ are some distributions that assign probability only at the points $\bar{\theta}_0,\ldots,\bar{\theta}_N$. Such a distribution obeys

$$\phi = \sum_{j=1}^{N} a(\bar{\theta}_{j-1})p_j \qquad (3.13)$$

where $p_j = F(\bar{\theta}_j) - F(\bar{\theta}_{j-1})$. In matrix form

$$\phi = Ap \qquad (3.14)$$

This equation, together with the requirement that p is non-negative, defines a set of probabilities, $S_I(\phi,D_N)$. Again, this is a convex polyhedron and is fully characterized by its set of vertices. Every vector p in S_I corresponds to a distribution in $V(\phi)$ that assigns probability only at the points $\bar{\theta}_0,\ldots,\bar{\theta}_N$. We define $V_I(\phi,D_N)$ as the set of distributions corresponding to the set of probabilities, $S_I(\phi,D_N)$; each probability contributes only one distribution. Then $V(\phi)$ encloses $V_I(\phi,D_N)$, so the extremal members of V_I meet our purpose of indicating how closely the outside bounds can be attained. As the partition becomes finer, V_I becomes richer, and ultimately converges to V.

The methods and conclusions of this section are summarized in the following theorem:

Theorem 3.1: Let

$$V_I(\phi,D_N) = \{F(\theta) \mid F(\theta) = \sum_{j \ni \bar{\theta}_{j-1} < \theta} p_j \text{ for some } p \in S_I(\phi,D_N)\} \qquad (3.15)$$

and

$$V_0(\phi,D_N) = \{F(\theta) \mid F(\bar{\theta}_j) - F(\bar{\theta}_{j-1}) = p_j,\; j = 1,\ldots,N, \text{ for some } p \in S_0(\phi,D_N)\} \qquad (3.16)$$

Then the following three properties hold:

(i) Enclosure: $V_I(\phi,D_N) \subset V(\phi) \subset V_O(\phi,D_N)$ (3.17)

(ii) Monotonicity: $V_I(\phi,D_N) \subset V_I(\phi,D_{N+1})$ and

$$V_O(\phi,D_{N+1}) \subset V_O(\phi,D_N) \tag{3.18}$$

(iii) Convergence: $\bigcap_{N=3}^{\infty} V_I(\phi,D_N) \doteq V(\phi) =$

$$\bigcap_{N=3}^{\infty} V_O(\phi,D_N) \tag{3.19}$$

(The precise meaning of $\doteq$ is explained in the proof.)

The final task of this section is to show how to find bounds on $P = F(\theta_H) - F(\theta_L)$ once a suitably refined partition, D_N, has been selected and the vertices $p^{(1)},\ldots,p^{(k)}$ of $S_O(\phi,D_N)$ calculated. For any $F(\theta) \in V_O(\phi,D_N)$,

$$P = \sum_{j \in J} p_j \tag{3.20}$$

where $J = \{j \mid \theta_L \leq \bar{\theta}_j \leq \theta_H\}$. Since P is a linear function on a polyhedron, it attains its extreme values at the vertices. Thus we define

$$\underline{P}_O = \min_{k=1,\ldots,K} \sum_{j \in J} p_j^{(k)} \tag{3.21}$$

and

$$\bar{P}_O = \max_{k=1,\ldots,K} \sum_{j \in J} p_j^{(k)} \tag{3.22}$$

so

$$\underline{P}_O \leq P \leq \bar{P}_O \tag{3.23}$$

We can define inside bounds $\underline{P}_I$ and $\bar{P}_I$ by a similar computation on the vertices of $S_I(\phi, D_N)$. Then from Theorem 3.1,

$$\underline{P}_0 \leq \underline{P} \leq \underline{P}_I \leq \bar{P}_I \leq \bar{P} \leq \bar{P}_0 \tag{3.24}$$

where $\underline{P}$ and $\bar{P}$ are the exact bounds over $V(\phi)$. Note that when $\theta_L = 0$ or $\theta_H = 1$, $\underline{P}$ and $\bar{P}$ can be computed exactly by the methods of Section 2.

Appendix to Section 3

Computing the vertices of $S_0(\phi,D)$ Each vertex of $S_0(\phi,D)$ is a non-negative solution to

$$\begin{pmatrix} \underline{A} & \\ -\bar{A} & I \\ v & 0 \end{pmatrix} \begin{pmatrix} p \\ \psi \end{pmatrix} = \begin{pmatrix} \phi \\ -\phi \\ 1 \end{pmatrix}, \tag{A3.1}$$

where v is a vector of N ones and ψ is a vector of 2M slack variables. Further, to be a vertex, no more than 2M + 1 elements of p and ψ together may be nonzero. Suppose K of the elements of p are nonzero and 2M + 1 - K of the elements of ψ are nonzero. Then

$$\begin{pmatrix} \hat{A} & 0 \\ A^* & -I \end{pmatrix} \begin{pmatrix} \hat{p} \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \hat{\phi} \\ \phi^* \end{pmatrix}, \tag{A3.2}$$

where $\hat{A}$ contains the columns of $\underline{A}$, $-\bar{A}$ and v corresponding to the nonzero probabilities, $\hat{p}$, and contains the rows of $\underline{A}$ and $-\bar{A}$ for the zero values of the slack variables. The last row of $\hat{A}$ is v. A^* contains the remaining rows of $\underline{A}$ and $-\bar{A}$. $\hat{\psi}$ contains the nonzero slack variables. $\hat{\phi}$ consists of the elements of ϕ, with appropriate sign, for rows with zero slack variables and ϕ^* consists of the remaining elements of ϕ and $-\phi$. Since the system is block-triangular, it has the recursive solution

$$\hat{p} = \hat{A}^{-1}\hat{\phi} \tag{A3.3}$$

$$\hat{\psi} = A^*\hat{p}$$

$\hat{p}$ contains the non-zero elements of a vertex of $S_0(\phi,D_N)$ if $\hat{p} \geq 0$ and

$\hat{\psi} > 0$. The vertices of the set of all solutions can be calculated by generating systematically the solutions for all possible choices of the elements of $\hat{p}$ and $\hat{\psi}$.

Proof of Theorem 3.1:

(i) Enclosure

(a) Consider $F(\theta) \in V_I(\phi, D_N)$. Then since $Ap = \phi$, $\int_0^1 a(\theta)dF(\theta) = \phi$ and $F(\theta) \in V(\phi)$.

(b) Consider $F(\theta) \in V(\phi)$. Let $p_j = F(\bar{\theta}_j) - F(\bar{\theta}_{j-1})$. From formulas (3.7) and (3.8), $p \in S_O(\phi, D_N)$, so $F(\theta) \in V_O(\phi, D_N)$.

(ii) Monotonicity

(a) Consider $F(\theta) \in V_I(\phi, D_N)$ and let $p_1, \ldots, p_N$ be the associated vector of probabilities. Without loss of generality, assume that D_{N+1} differs from D_N by its $\bar{\theta}_{N-1}$. Then

$$\sum_{j=1}^{N-1} a(\bar{\theta}_{j-1})p_j + a(\bar{\theta}_N)p_N = \phi \tag{A3.5}$$

This shows that the vector

$$p = [p_1, \ldots, p_{N-1}, 0, p_N]' \tag{A3.6}$$

is in $S_I(\phi, D_{N+1})$, so $F(\theta) \in V_I(\phi, D_{N+1})$.

(b) Consider $F(\theta) \in V_O(\phi, D_{N+1})$. Let $\bar{\theta}_j^{\,N}$ and $\bar{\theta}_j^{\,N+1}$ be the points of the two partitions, let $\underline{A}^N$, $\bar{A}^N$, $\underline{A}^{N+1}$, and $\bar{A}^{N+1}$ be the corresponding matrices calculated from $a(\theta)$, let $p_j^{\,N} = F(\bar{\theta}_j^{\,N}) - F(\bar{\theta}_{j-1}^{\,N})$, $j = 1, \ldots, N$ and $p_j^{\,N+1} = F(\bar{\theta}_j^{\,N+1}) - F(\bar{\theta}_{j-1}^{\,N+1})$, $j = 1, \ldots, N+1$, and let $\underline{B}$ and $\bar{B}$ be matrices obtained by duplicating the column in $\underline{A}^N$ and $\bar{A}^N$

that corresponds to the interval containing the point in D_{N+1} but not in D_N. Now

$$\underline{A}^{N+1}p^{N+1} \leq \phi \leq \bar{A}^{N+1}p^{N+1}. \quad \text{(A3.7)}$$

Further, $\underline{B} \leq \underline{A}^{N+1}$ and $\bar{B} \geq \bar{A}^{N+1}$ by their constructions, so

$$\underline{B}p^{N+1} \leq \phi \leq \bar{B}p^{N+1} \quad \text{(A3.8)}$$

But $\underline{B}p^{N+1} = \underline{A}^N p^N$ and $\bar{B}p^{N+1} = \bar{A}^N p^N$, so

$$\underline{A}^N p^N \leq \phi \leq \bar{A}^N p^N \quad \text{(A3.9)}$$

and $p^N \in S_O(\phi, D_N)$. We conclude that $F(\theta) \in V_O(\phi, D_N)$.

(iii) Convergence

(a) By $\doteq$, we mean

$$\bigcup_{|\psi-\phi|<\varepsilon} \bigcup_{N=3}^{\infty} V_I(\psi, D_N) = \bigcup_{|\psi-\phi|<\varepsilon} V(\psi) \quad \text{(A3.10)}$$

for any $\varepsilon > 0$. For some ϕ on the boundary of Φ, $V_I(\phi, D_N)$ may be empty for all N, so it is impossible that $\bigcup_{N=3}^{\infty} V_I(\phi, D_N) = V(\phi)$ in all cases.

(1) Consider $F(\theta) \in \bigcup_{|\psi-\phi|<\varepsilon} \bigcup_{N=3}^{\infty} V_I(\psi, D_N)$. Then there is a ψ and an N^* such that $|\psi-\phi|<\varepsilon$ and $\psi \in V_I(\psi, D_N)$ for $N > N^*$. Thus

$$\psi = \sum_{j=1}^{N} a(\bar{\theta}_j)(F(\bar{\theta}_j) - F(\bar{\theta}_{j-1})) \text{ , all } N > N^* \quad \text{(A3.11)}$$

But the right-hand side of (A3.11) converges to the corresponding integral (see Apostol (1957), Exercise 9-4,

p. 243), so

$$\psi = \int_0^1 a(\theta)dF(\theta) \tag{A3.12}$$

and $F(\theta) \in V(\psi)$.

(2) Consider $F(\theta) \in \bigcup_{|\psi-\phi|<\varepsilon} V(\psi)$. Then

$$\left|\int_0^1 a(\theta)dF(\theta) - \phi\right| = \eta < \varepsilon \; . \tag{A3.13}$$

There exists a partition D_N such that

$$\left|\psi - \int_0^1 a(\theta)dF(\theta)\right| < \varepsilon - \eta \tag{A3.14}$$

where

$$\psi = \sum_{j=1}^{N} a(\bar{\theta}_{j-1})(F(\bar{\theta}_j) - F(\bar{\theta}_{j-1})) \tag{A3.15}$$

so $F(\theta) \in V_I(\psi,D_N)$. Now

$$|\psi - \phi| < \eta + \varepsilon - \eta = \varepsilon \tag{A3.16}$$

so $F(\theta) \in \bigcup_{|\psi-\phi|<\varepsilon} V_I(\psi,D_N)$ as required.

(b) In view of monotonicity, we need only show that $\bigcup_{N=3}^{\infty} V_O(\phi,D_N) \subset V(\phi)$. Consider $F(\theta) \in V^*(\phi,D_N)$, all N. Let

$$L_i(F,D_N) = \sum_{j=1}^{N} \underline{a}^N{}_{i,j} p_j{}^N \tag{A3.17}$$

and

$$U_i(F,D_N) = \sum_{j=1}^{N} \bar{a}_{i,j}^{\,N} p_j^{\,N} \tag{A3.18}$$

From the definition of the integral (2.2),

$$\lim_{N\to\infty} L_i(F,D_N) = \lim_{N\to\infty} U_i(F,D_N) =$$

$$\int_0^1 a(\theta)\,dF(\theta) \tag{A3.19}$$

But $L_i(F,D_N) \leq \phi$ and $U_i(F,D_N) \geq \phi$, for all N, so

$$\int_0^1 a(\theta)\,dF(\theta) = \phi \ , \tag{A3.20}$$

and $F(\theta) \in V(\phi)$.

4. The Sex Composition of Families

The following example illustrates the nature of the information about the distribution of an unobserved component in a simple case. Suppose that we observe a large number of apparently identical families with two children, and suppose further that a fraction ϕ_1 of the families have no girls, ϕ_2 have one girl, and ϕ_3 have two girls. Each family has a probability θ that a given child will be a girl. In terms of the general model given earlier, if y is the number of girls in a family,

$$y = h(\theta,u)$$

$$= u, \text{ where } u \text{ is binomial of order 2 with parameter } \theta. \qquad (4.1)$$

If all families have the same θ, then ϕ will be the binomial distribution:

$$\phi_1 = (1 - \theta)^2; \quad \phi_2 = 2\theta(1 - \theta); \quad \text{and} \quad \phi_3 = \theta^2 \qquad (4.2)$$

If θ varies among families, then ϕ will be the mixed binomial,[2]

$$\phi_1 = \int_0^1 (1 - \theta)^2 dF(\theta) \ ; \qquad (4.3)$$

$$\phi_2 = \int_0^1 2\theta(1 - \theta) dF(\theta) \ ; \qquad (4.4)$$

$$\phi_3 = \int_0^1 \theta^2 dF(\theta) \ . \qquad (4.5)$$

---

[2] This possibility has been discussed in the literature on mathematical demography (for example, Goodman (1961) and Weiler (1959)). This treatment of the sex composition of families is only an example and does not consider other aspects of the problem, especially the effects of the efforts of parents to influence the composition through stopping rules. On this, see Ben Porath and Welch (1972).

In our earlier notation, the kernel is:

$$a(\theta) = \begin{pmatrix} (1-\theta)^2 \\ 2\theta(1-\theta) \\ \theta^2 \end{pmatrix} . \tag{4.6}$$

Ben Porath and Welch (1972) report the following distribution for the sexes of the first two children of American families: $\phi_1 = 0.262$, $\phi_2 = 0.497$, and $\phi_3 = 0.241$. The mean of this distribution is 0.979, suggesting that if θ had a single value, it would be half the mean, 0.489. However, the binomial distribution with parameter $\theta = 0.489$ is [0.261, 0.500, 0.239], which has somewhat less dispersion than the observed ϕ. No single value of θ can explain the observed distribution of sexes, so we are forced to consider a distribution of the propensity to have girls, θ, within the population.

The theory of Tchebycheff systems discussed in Section 2 focuses attention on the canonical representations involving alternative values of θ^*, a preassigned type. Since M is 3 for this problem, either N = 2 and J = 2, in which case the canonical representation requires solving for f_1, f_2, and θ_2, or N = 3 and J = 4, in which case the canonical representation requires solving a linear system for f_1, f_2, and f_3. In both cases f_1 is the upper bound on the fraction of the population that has probability θ^* of having a girl, by Theorem 2.3. Further, from Theorem 2.4, when N = 2 and $\theta^* < \theta_2$, f_1 is the upper bound on the fraction of the population with θ less than θ^*, $F(\theta^*)$; when N = 2 and $\theta^* > \theta_2$, f_1 is the upper bound on the fraction of the population with θ at or above

θ^*, $1 - F(\theta^*)$, and when $N = 3$, $f_1 + f_2$ is the upper bound on $F(\theta^*)$ and $f_1 + f_3$ is the upper bound on $1 - F(\theta^*)$. Table 1 presents canonical representations for a variety of values of θ^*. For θ^* outside a short interval enclosing 0.489, the canonical representation has only one additional type, θ_2. The first part of Table 1 shows a variety of representations of this kind. When θ^* is extreme, the representation gives a low weight (f^*) to θ^* and a high weight to a θ_2 that is close to 0.489. As θ^* approaches 0.489, it receives higher weight and the second type, θ_2, becomes more extreme. At the critical points $\theta^* = 0.4868$ and $\theta^* = 0.4923$, θ_2 reaches 1, and we enter the region where the representation gives weight to three values of θ: the two extremes, $\theta = 0$ and $\theta = 1$,

Table 1

Canonical Representations for the Mixed Binomial Model of the Sex Composition of Families

θ^*	θ_2	f^*	f_2
0.0	.4923	.0058	.9942
0.30	.4968	.0372	.9628
0.40	.5050	.1479	.8521
0.48	.6358	.9390	.0610
0.50	.3571	.9265	.0735
0.60	.4769	.1022	.8978
0.70	.4829	.0304	.9696
1.00	.4868	.0053	.9947

θ^*	f_2 ($\theta = 0$)	f^*	f_3 ($\theta = 1$)
.49	.0034	.9944	.0022

and $\theta = \theta^*$. One such representation is shown in the second part of Table 1. According to this representation, the observed distribution of sex compositions could be generated by a population in which 99 44/100% of couples had a probability of 0.49 of having girls, 0.34% had nothing but boys, and 0.22% had nothing but girls.

From Table 1 we can derive the Markov-Krein bounds $\underline{P}$ and $\bar{P}$ for the fraction of the population with θ between 0 and θ_H. These bounds are presented in Table 2, along with the outside and inside bounds calculated by the methods of Section 3. All of the bounds agree that it is quite possible that no couple has a probability of having a girl below 0.48 and also possible that none has a probability above 0.50 (but

Table 2

Bounds on the Fraction of the Population with θ between 0 and θ_H

θ_H	$\underline{P}_O$	$\underline{P}$	$\underline{P}_I$	$\bar{P}_I$	$\bar{P}$	$\bar{P}_O$
0.30	0	0	0	.0236	.0372	.0375
0.40	0	0	0	.1018	.1479	.1500
0.48	0	0	0	.7545	.9390	.9543
0.49	.0033	.0034	.0038	.9853	.9966	.9979
0.50	.0530	.0735	.0742	1.0	1.0	1.0
0.60	.8720	.8978	.8990	1.0	1.0	1.0
0.70	.9657	.9696	.9697	1.0	1.0	1.0

Explanation:

$\underline{P}$, $\bar{P}$: Markov-Krein exact bounds, derived from Table 1.

$\underline{P}_O$, $\bar{P}_O$: Outside bounds, with endpoints 0, .15, .25, .30, .35, .38, .40, .42, .47, .48, .485, .49, .495, .50, .52, .58, .60, .62, .65, .70, .75, .85, 1.00. S_O has 2776 vertices.

$\underline{P}_I$, $\bar{P}_I$: Inside bounds; same points as above. S_I has 368 vertices.

there must be some couples with probabilities either below 0.48 or above 0.50). There may be as many as 3.72% with probabilities below 0.30, as many as 14.79% below 0.40, and as many as 93.90% below 0.48. At least 0.34% and possibly as much as 99.66% of the population have probabilities below 0.49. At least 7.35% have probabilities below 0.50, at least 89.78% below 0.60, and at least 96.96% below 0.70. The upper outside bound $\bar{P}_0$ and the lower inside bound $\underline{P}_I$ perform well as approximations to the sharp bounds. $\underline{P}_0$ does well except at 0.50, where it is quite pessimistic (although, of course, still a true bound). $\bar{P}_I$ is always much too small. Recalculation of the inside bounds with a finer partition would remedy this problem.

Table 3 presents bounds for various intervals that do not begin at zero. No Markov-Krein sharp bounds are available for these intervals,

Table 3

Bounds on the Fraction of the Population with θ between θ_L and θ_H

θ_L	θ_H	$\underline{P}_0$	$\underline{P}_I$	$\bar{P}_I$	$\bar{P}_0$
0.30	0.70	.9625	.9697	1.0000	1.0000
0.40	0.60	.8500	.8851	1.0000	1.0000
0.40	0.50	0	0	.9955	1.0000
0.40	0.48	0	0	.7545	.9543
0.48	0.50	0	0	.9944	.9954
0.48	0.49	0	0	.9853	.9953
0.49	0.50	0	0	.9943	.9944
0.50	0.70	0	0	.9258	.9470

For explanation, see Table 2.

so the only way to judge the sharpness of the outside bounds is through the inside bounds. At least 96.25% of all couples have probabilities between 0.30 and 0.70, and we know that there exists a distribution consistent with ϕ in which only 96.97% of the population lies between 0.30 and 0.70. On the other hand, it is possible that 99.54% of the population has θ between 0.48 and 0.50, and we know for sure that 99 44/100 % can be in this interval.

A fairly wide variety of distributions of the propensity to have girls is consistent with the observed data on the distribution of the number of girls among the first two children. Although little can be done to localize the distribution in the vicinity of 0.5, our methods give fairly specific information about the fraction of the population with extreme propensities. The data are not consistent with any distributions with large fractions of the population having extreme values of θ. An increase in the number of times each unit is observed, in this case the number of children, would refine our knowledge considerably. A study of sex composition that examined more than the first two children would need to deal explicitly with the problem of stopping rules, however.

5. Mixed Markov Processes and the Mover-Stayer Model

This section illustrates the application of the methods discussed earlier to a problem of considerable interest in the study of social mobility. Suppose there are two states that an individual may occupy in each period: poor or not poor, employed or not employed, lower class or middle class, or some other dichotomy. Suppose further that a Markov process governs transitions between the states; there is a probability θ

that an individual in the first state in one period will move to the second state in the next period, and a probability δ that an individual in the second will move to the first. The probabilities of remaining in the states are then $1 - \theta$ and $1 - \delta$ respectively. Models of this kind fitted to data on observed transitions of individuals under the assumption that θ and δ are the same for all of them have suffered from an important defect: They understate the probability that an individual will remain for many successive periods in the same state, even though they predict correctly the probability that an individual chosen at random from the inhabitants of one state will move to the other state in the next period (Blumen, Kogan, and McCarthy (1955)).

The mover-stayer model resolves this paradox by assuming that there are actually two kinds of people, movers, who have positive θ, and stayers, whose θs are zero. The probabilities of observed transitions are the mixture of two different Markov processes. Methods for estimating the parameters of the two processes and the single mixing probability have been developed by Goodman (1960). Recently Spilerman (1972) has proposed an extension of the model in which the observed probabilities are treated as the mixture of all of the powers of a particular transition matrix. None of the literature on the mover-stayer model takes advantage of the statistical theory of mixtures, however.

A natural generalization of the mover-stayer model is the mixture of all Markov processes. To keep within the confines of the theory developed in this paper, however, we will suppose that individuals differ only with respect to their probability of upward mobility, θ, and that δ is known and constant within the population. Then it is appropriate to study the distribution of the number of spells in the second state over a certain number of periods, T. Individuals with high values of θ will tend to have more spells than do those with low θ. We define the observed probability, ϕ, in the following way:

ϕ_1 = fraction of the population with no spells

$\vdots$

ϕ_i = fraction with i - 1 spells

$\vdots$

ϕ_M = fraction with M - 1 or more spells

Data on spells of unemployment during a year are reported by the U. S. Census Bureau in precisely this form, with M = 4.

We define $a_i(\theta)$ as the probability of i - 1 spells in T periods induced by a Markov process with parameters θ and δ. There is no simple closed form for $a_i(\theta)$, but it can be calculated from the following recursion: Let Q(t,i,j) be the probability of having i - 1 spells in t periods and of finishing in state j at time t. Then

$$Q(t+1,i,1) = (1 - \theta)Q(t,i,1) + \theta Q(t,i,2)$$

$$Q(t+1,i,2) = \delta Q(t,i-1,1) + (1 - \delta)Q(t,i,2) \qquad (5.1)$$

with

$$Q(0,i,j) = 0 \qquad \text{if } i \neq 1$$

$$Q(0,1,1) = p^*$$

$$Q(0,1,2) = 1 - p^*$$

$$Q(t,-1,1) = 0 \ , \ t = 1,\ldots,T \qquad (5.2)$$

Here p* is the probability of being in the first state at time 0 and might reasonably be taken as the steady-state probability of being in the first state:

$$p^* = \frac{\delta}{\theta + \delta} \tag{5.3}$$

Finally,

$$a_i(\theta) = Q(T,i,1) + Q(T,i,2)\ ,\ i = 1,\ldots,M - 1 \tag{5.4}$$

$$a_M(\theta) = 1 - a_1(\theta) - \ldots - a_{M-1}(\theta) \tag{5.5}$$

This puts the mixture into our standard form,

$$\phi = \int_0^1 a(\theta)dF(\theta)\ . \tag{5.6}$$

All of our earlier techniques can be applied to obtain information about the distribution of the probability of upward mobility among the population. The mover-stayer model is the special case where $F(\theta)$ concentrates all its probability at $\theta = 0$ and at one other value of θ. From Theorem 2.1, if our data distinguish only among no spells, one spell, and two or more ($M = 3$), then there is always a simple mover-stayer model that explains the observed ϕ, namely the canonical representation involving $\theta^* = 0$. Other distributions will also be consistent with ϕ, however, and if the data on the number of spells are richer, the simple mover-stayer model will not generally be able to explain ϕ. In any case, the assumption that there are exactly two types of people is a highly restrictive one; our methods provide a workable method for relaxing it.

6. Extensions

Many investigators are likely to be willing to make restrictive assumptions about the form of the distribution of the unobserved component in order to tighten the results by ruling out implausible distributions. This can be done through the conventional device of confining the distribution to a family indexed by a limited number of parameters. If the number of parameters is equal to the number of observed probabilities, then it is often straightforward to calculate $F(\theta)$ from ϕ. For example, if $a(\theta)$ is binomial and $F(\theta)$ is a beta distribution, then the parameters can be calculated directly from ϕ; see Maritz (1970), pp. 22-23. On the other hand, a weak parametrization that imposes nothing more than smoothness on $F(\theta)$ will usually have more than M parameters, so more than one member of the parametric family of distributions will be consistent with the observed ϕ. The problem then is essentially similar to the problem treated in this paper. In particular, if the family is linear in its parameters, the set of parameters consistent with ϕ is mathematically the same as the set S_I derived in Section 3. The family of distributions whose densities are step functions is an important example of such a family.

Second, in practice we do not observe the probabilities ϕ but only the corresponding frequencies, say $\hat{\phi}$. If we apply our methods to $\hat{\phi}$, then our bounds become random variables that estimate the bounds but are not truly bounds themselves. A confidence region enclosing $\hat{\phi}$ induces a confidence interval for each bound. The only serious problem in dealing with $\hat{\phi}$ arises when it does not lie in Φ. For example, in a small population it is possible that every family has one girl and one boy, but there is no mixture of binomial distributions that gives rise to the corresponding set of probabilies. Fortunately, if ϕ is in the interior of Φ, the probability that $\hat{\phi}$ lies outside Φ approaches zero as the sample size increases.

Third, in many applications the probabilities of alternative outcomes depend on the observed characteristics of the individual as well as on his unobserved type. The easy way to incorporate this dependence in our model is to let $F(\theta;x)$ be the distribution of θ within the subpopulation of individuals with characteristics x. Then the observed mixture also depends on x:

$$\phi(x) = \int_0^1 a(\theta)dF(\theta;x) \qquad (6.1)$$

Given $\phi(x)$ for a particular x, we can then apply our methods to derive information about $F(\theta;x)$. In practice, we specify $\phi(x)$ as a multinomial probability depending on x in a reasonably flexible way, using a multinomial logit or other convenient specification. Note that $\phi(x)$ does not have the same structure as $a(\theta)$--for example, the study of mixed Markov processes does not involve the estimation of the parameters of a Markov process. From $\phi(x)$, we calculate bounds on $F(\theta;x)$ for representative values of x.

7. Concluding Remarks

Unobserved differences among individuals are an important source of diversity in their observed behavior. For the case in which the probability distribution among the alternatives is a known function of the unobserved type, this paper has shown that exact but not complete knowledge of the distribution can be obtained. The assumptions of previous authors about these distributions can, in fact, be tested.

References

Tom Apostol, (1957) Mathematical Analysis, Reading, Massachusetts, Addison-Wesley.

Yoran Ben Porath and Finis Welch, (1972) "Unvertain Quality: Sex of Children and Family Size," Rand Publication R-1117-NIH/RF, Aug. 30, 1972, Revised October 1972.

Isadore Blumen, Marvin Kogan, and Philip J. McCarthy, (1955) The Industrial Mobility of Labor as a Probability Process, Cornell University.

Thomas Domencich and Daniel McFadden, (1974) Development of Behavioral Models of Urban Travel Demand, North-Holland, forthcoming.

Leo Goodman, (1961) "Statistical Methods for the Mover-Stayer Model," Journal of the American Statistical Association, Vol. 56, pp. 841-868.

Leo Goodman, (1961) "Some Possible Effects of Birth Congrol on the Human Sex Ratio," Annals of Human Genetics, Vol. 25, pp. 75-81.

Zvi Griliches,(1973) research in progress.

Christopher Jencks, et al, (1972) Inequality: A Reassessment of the Effect of Family and Schooling in America, New York, Basic Books.

Samuel Karlin and William J. Studden, (1966) Tchebycheff Systems: with Applications in Analysis and Statistics, New York, Wiley.

M. G. Krein, (1951) "The Ideas of P.L. Cebysev and A.A. Markov in the Theory of Limiting Values of Integrals and Their Further Developments," American Mathematical Society Translations, Series 2, Vol. 12, pp. 1-122.

J.S. Maritz, (1970) Empirical Bayes Methods, London, Methuen.

Seymour Spilerman, (1972) "Extensions of the Mover-Stayer Model," American Journal of Sociology, Vol. 78, Nov. 1972, pp. 599-626.

Henry Teicher, (1963) "Identifiability of Finite Mixtures," Annals of Mathematical Statistics, Vol. 34, pp. 1265-1269.

H. Weiler, (1959) "Sex Ratio and Birth Control," American Journal of Sociology, Vol. 65, pp. 289-299.