Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence
Author(s): James J. Heckman
Source: *Annales de l'inséé*, No. 30/31, The Econometrics of Panel Data (Apr. - Sep., 1978), pp. 227-269
Published by: GENES on behalf of ADRES
Stable URL: http://www.jstor.org/stable/20075292
Accessed: 25-07-2017 02:42 UTC

ANNALES DE L'INSEE – N° 30-31/1978

Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence

James J. HECKMAN *

This paper develops statistical models for the analysis of discrete longitudinal data. Many interesting stochastic processes are generated by a model in which discrete events arise from a dichotomization of latent continuous variables. The latent variables are given various dynamic specifications which give rise to alternative stochastic processes. Probabilities of runs patterns for the temporal sequences of discrete events generated by these models are investigated. The models are used to investigate whether conditional relationships between the probability of experiencing an event and previous experience of the event arise from spurious correlation or a real effect of previous experience. Simple runs tests and generalized linear probability estimators are developed and applied to test the hypothesis of no real effect.

* University of Chicago. This research was partially supported by an ASPER-DOL grant to the National Bureau of Economic Research and by NSF grant SOC 77-27136. Valuable comments have been received from Gary Chamberlain, Zvi Griliches, Jan Hoem, Tom Marcurdy, Marc Nerlove and Jose Scheinkman. Chris Flinn, Guilherme Sedlacek and Ralph Shnelvar provided highly competent research assistance. I retain responsibility for all remaining errors.

This paper considers simple answers to the following question. From an observed series of discrete events from individual histories, such as the spells of unemployment experienced by workers or the labor force participation histories of married women, is it possible to explain the frequently noted empirical regularity that individuals who experience an event in the past are more likely to experience the event in the future? There are two distinct explanations for this regularity. One explanation is that individuals who experience the event are altered by their experience. A second explanation is that individuals may differ in their propensity to experience the event. If individual differences are stable over time, individuals who experience the event in the past are likely to experience the event in the future, even though the actual experience of the event has not modified individual behavior. The question considered in this paper is whether or not it is possible to determine which explanation is more appropriate.

This question is of considerable substantive interest. Two examples are offered. The first is drawn from recent work in the theory of unemployment. Phelps (1972) argues that short term economic policies that alleviate unemployment tend to lower aggregate unemployment rates in the long run by preventing the loss of work enhancing market experience. This view assumes that unemployment has a real effect on the unemployed. Cripps and Tarling (1974) maintain a diametrically opposite view in their analysis of the determinants of the incidence and duration of unemployment. They argue that individuals differ in their propensity to experience the event of unemployment, and that the actual experience of being unemployed does not affect the probability of becoming unemployed in the future.

To take another example, recent work on the dynamics of female labor supply assumes that entry and exit from the labor force for a given woman follows a Bernoulli statistical model (Heckman and Willis, 1977). This view of female labor force dynamics ignores considerable evidence that work experience per se raises wage rates and hence may increase the probability that a woman works in the future, even if initial entry into the work force is random. In interpreting cross section labor supply functions, it is important to know whether or not the Bernoulli view of labor force turnover is correct.

The model presented in this paper is essentially a discrete data analogue of the familiar continuous variate distributed lag model with serial correlation in the disturbances that has received much attention in the econometric literature (Balestra and Nerlove, 1966; Malinvaud, 1970). The methodology developed here is based on previous work by the author (Heckman, 1978a). The purpose of this paper is to extend that work to develop a model applicable to panel data and to develop simple estimators, based on runs patterns in the data, that enable analysts to use regression methods to answer the question posed. This work is very much in the spirit of previous work by David (1947), Goodman (1958) and Denny and Yakowitz (1978) who develop runs tests to distinguish between Markov and Bernoulli models. The methods developed here permit the analyst to distinguish among a greater variety of interesting stochastic models for discrete data than is considered in those papers.

A second topic of attention in this paper is the analysis of the usefulness and limitations of cross section data for the estimation of microdynamic models. The main questions considered are (a) when can cross section data be used to fully characterize a dynamic process? (b) If it cannot be so used, what information about the dynamic process can be retrieved from the cross section? (c) How many panel observations does one need to estimate the parameters of particular models?

This paper is in five parts. Part one is a technical introduction to the remainder of the paper that sets out the framework for the ensuing analysis. Part two outlines the basic statistical model used to discriminate between "true" and "spurious" state dependence. There, it is noted that under general conditions, unconditional probability models can be used to distinguish between the two types of dependence using only two periods of data drawn from the life histories of individuals. Part three presents a general stochastic model, and various specializations of it that appear in the literature. Special attention is devoted to the implications of these models for "runs" patterns, and consideration is given to the question of whether or not cross section data can be used to estimate some, or all, of the parameters of these models. Part four collects the results on runs patterns developed in part three, and presents a general regression model that can be used to perform "runs tests" when simple methods--based on the assumption that individuals are observationally identical in "control" variables

--are inapplicable. Part five presents a brief empirical example of the methods developed in part four. The paper concludes with a summary and references to related work.

I. INTRODUCTION

All of the statistical models considered here are based on the following ideas. The analyst has at his disposal a random sample of I individuals on each of whom there are individual histories recording the presence or absence of an event in each of T equally spaced discrete time periods. An event occurs at time t for individual i only if a continuous latent random variable $X_i(t)$ crosses a threshold, assumed to be zero for convenience. The event occurs and dummy variable $d_i(t) = 1$ if and only if $X_i(t) > 0$, otherwise the event does not occur, and $d_i(t) = 0$.

Two remarks are in order at the outset of the discussion. First, there may be more than two possible outcomes at any time rather than the dichotomy of outcomes considered here. The polytomous case will be considered in another paper that is a fairly direct generalization of this paper. Second, the introduction of a latent continuous random variable into the analysis may appear to be an unnecessary complicating idea. On the contrary, it simplifies the analysis greatly, links the current work with previous work in econometrics, and leads to a natural framework for formulating choice theoretic econometric models.

One example of the value of this framework comes in the analysis of the time series of the labor force participation of married women. In that analysis, $X_i(t)$ may be interpreted as the difference between the remaining lifetime utility of woman i evaluated at time t if she is in the labor force and her remaining lifetime utility at time t if she is not. In this example, it is natural to assume that the difference in utilities $(X_i(t))$ is a continuous random variable (see, e.g., McFadden, 1976). If $X_i(t)$ is positive, the woman participates, otherwise she does not.

In an analysis of the incidence of unemployment, it is natural to formulate a model in terms of the difference between reservation wages and offered market wages (see Kiefer and Neumann, 1977). If this difference is positive, the individual becomes (or remains) unemployed.

Sometimes it is possible to observe a continuous random variable $X_i(t)$ that generates the discrete random variable $d_i(t)$ so that $X_i(t)$ is more than a theoretical construct (e.g., hours of work in the labor force participation example).

By making alternative stochastic assumptions about the nature of random variable $X_i(t)$, it is possible to generate a rich class of statistical models for discrete data that can be interpreted within a choice theoretic context.

Random variable $X_i(t)$ may always be decomposed into two components: a purely stochastic component, $\varepsilon_i(t)$, and a function of measured variables that affect current choices, $V_i(t)$. $V_i(t)$ may or may not be correlated with $\varepsilon_i(t)$. Utilizing this notation, we may write

(1.1) $X_i(t) = V_i(t) + \varepsilon_i(t)$.

The distribution of the $d_i(t)$, $t = 1,\dots,T$, $i = 1,\dots,I$ critically depends on the assumptions made about the distribution of $\varepsilon_i(t)$ and $V_i(t)$. To simplify the argument in this paper, it is assumed that the $\varepsilon_i(t)$ are jointly normally distributed when a distributional assumption is required so that the analysis presented here is similar to the multivariate probit structure of Ashford and Sowden (1970) as extended by the author (Heckman, 1978a). The bulk of the analysis in the paper consists of examining the implications of alternative stochastic specifications of $V_i(t)$ and $\varepsilon_i(t)$, and the problem of inferring the underlying stochastic structure from observations drawn from panel data.

Before a more formal analysis is begun, it is useful to sketch the basic idea underlying the methodology developed here, and to demonstrate the relevance of the methodology for the question at hand, i.e., whether or not it is possible to use panel data to determine whether or not there is a "real" effect of occupancy of a state on the future probability of occupying the state. Having presented the main idea, a more formal analysis is then pursued.

II. THE BASIC MODEL

Suppose that there is a sample of I randomly sampled individuals who are observationally identical at time $t = 1$. There are two observations available on each person in the sample, one at time t = 1 and a second at t = 2. Events that occur prior to the sample are assumed to have no effect on decisions made during the sample period. The issue at hand is to ascertain whether or not such data can be used to determine whether current occupancy of a state affects the future probability of occupying the state.

Utilizing the notation established in the Introduction, let $V_i(1) = \bar{V}$, a constant, so that at time one everyone is observationally identical in the measured traits that affect the probability of experiencing the event. As before, person i experiences the event ($d_i(1) = 1$) if and only if $X_i(1) > 0$, i.e., if

$$\varepsilon_i(1) > -\bar{V}$$

which occurs with probability

$$\Pr(\varepsilon_i(1) > -\bar{V}).$$

If $\varepsilon_i(1)$ is a normally distributed random variable, this probability is a familiar "probit" probability

$$\Pr(\varepsilon_i(1) > -\bar{V}) = \Phi\left(\frac{\bar{V}}{\sigma_\varepsilon}\right) = 1 - \Phi\left(-\frac{\bar{V}}{\sigma_\varepsilon}\right) \tag{2.1}$$

where Φ is the cumulative distribution of the standard normal, and σ_ε is the standard error of $\varepsilon_i(1)$. At time t = 1, if the sample is sufficiently large, proportion $\Phi(\bar{V}/\sigma_\varepsilon)$ of the sample experience the event while $1 - \Phi(\bar{V}/\sigma_\varepsilon)$ do not.

The hypothesis that there is a real effect of past occupancy of a state on future behavior requires that people who experience an event in time period one have their V_i changed so that subsequent choice probabilities are altered. One simple way to capture this notion is to define random variable $X_i(2)$ in the following way

$$X_i(2) = \bar{V} + \gamma d_i(1) + \varepsilon_i(2). \tag{2.2}$$

The expectation of $\varepsilon_i(2)$ (conditioned on $\bar{V}$) is zero, and its variance is σ_ε^2, the same as that of random variable $\varepsilon_i(1)$. The correlation between $\varepsilon_i(2)$ and $\varepsilon_i(1)$ is ρ. The event occurs in time period two ($d_i(2) = 1$) if

$$X_i(2) > 0 \iff \varepsilon_i(2) > -[\bar{V} + \gamma d_i(1)]. \tag{2.3}$$

If γ is a positive number, individuals who experience the event in the first time period are more likely to experience the event in the second time period. In the example of the labor force participation of women, γ may be interpreted as the effect of current work experience on future earning power, a relationship that is assumed to be positive. Defining $d_i(1) = 1$ if woman i participates in the market in time period one, it is plausible (but not logically necessary) that γ is positive since higher wage women are more likely to participate in the market.

For two conceptually distinct reasons, persons who experience the event in time period one are more likely to experience the event in time period two. The first reason has just been discussed, and has a sound economic motivation. The second reason arises even if $\gamma = 0$ and is a consequence of the correlation between $\varepsilon_i(1)$ and $\varepsilon_i(2)$. This correlation gives rise to the "mover-stayer" problem.

To understand this point, set $\gamma = 0$. The conditional probability that a person who experiences the event in the first period experiences the event in the second period is

$$(2.4) \quad \Pr(d_i(2) = 1 | d_i(1) = 1) = \frac{\int_{-\bar{V}}^{\infty} \int_{-\bar{V}}^{\infty} f(t_1, t_2) dt_1 dt_2}{\int_{-\infty}^{\infty} \int_{-\bar{V}}^{\infty} f(t_1, t_2) dt_1 dt_2}$$

where $f(t_1, t_2)$ is a standardized bivariate normal density, and σ_ε is arbitrarily fixed at one. If $\rho = 0$, so that $\varepsilon_i(1)$ and $\varepsilon_2(2)$ are independent random variables, the bivariate normal distribution function is the product of two univariate normal distribution functions and

$$(2.5) \quad \Pr(d_i(2) = 1 | d_i(1) = 1) = \Pr(d_i(2) = 1)$$
$$= \Phi(-\bar{V}).$$

The conditional probability is a monotonically increasing function of ρ. If $\rho > 0$, individuals who experience the event at $t = 1$ are more likely to experience the event at $t = 2$, and the dependence grows with the value of ρ. If the correlation between $\varepsilon_i(1)$ and $\varepsilon_i(2)$ were perfect ($\rho = 1$), individuals who experience the event in the first time period are certain to experience the event in the second time period. Even if the correlation is not perfect, the information that an individual has experienced the event at time $t = 1$ conveys information about his likelihood of experiencing the event at time $t = 2$ even if $\gamma = 0$. Thus in equality (2.3), if $\rho > 0$, $d_i(1)$ and $d_i(2)$ are positively correlated so that a simple probit model applied to the second period data would lead to upward biased estimates of γ. It is this bias that constitutes essence of the "mover-stayer" problem in the current context.

The question posed in the Introduction to this paper can now be reformulated in more precise terms. Is $\gamma \neq 0$, so that there is a genuine effect of occupancy of a state on the future probability of occupying the state (i.e., there is true state dependence at the micro level), or is $\rho \neq 0$, so that there is "spurious dependence," or are both parameters nonzero? If so, what is the relative importance of each effect on explaining the measured sequence of discrete events? The distinction between spurious and true state dependence is that spurious dependence does not arise as a consequence of previous choices while true state dependence does arise from the actual experience of past events.

A method for estimating γ, $\bar{V}$ and ρ (and other parameters in a more general context to be elaborated below) has been developed by the author in other papers (Heckman, 1978a, b)). The essential idea underlying the method will be exposited here but the reader is referred to the other papers for a more rigorous development of the properties of the estimators. (Amemiya (1978) has proposed alternative estimators for this model.)

In the hypothetical sample of our example there are four distinct outcomes illustrated in the following contingency table with hypothetical numbers in each cell.

	$d_i(2) = 1$	$d_i(2) = 0$
$d_i(1) = 1$	N_{11}	N_{10}
$d_i(1) = 0$	N_{01}	N_{00}

Clearly $\sum_{j=0}^{1} \sum_{i=0}^{1} N_{ij} = I$.

The probability of the four events in the general case of $\gamma \neq 0$ and $\rho \neq 0$ (assuming $\sigma_\varepsilon = 1$) is, letting "Pr" stand for probability,

$$(2.6a)\quad P_{11} = \Pr(d_i(1) = 1 \wedge d_i(2) = 1) = \int_{-\bar{V}}^{\infty} \int_{-(\bar{V}+\gamma)}^{\infty} f(t_1, t_2)\,dt_1 dt_2$$

$$(2.6b)\quad P_{10} = \Pr(d_i(1) = 1 \wedge d_i(2) = 0) = \int_{-\bar{V}}^{\infty} \int_{-\infty}^{-(\bar{V}+\gamma)} f(t_1, t_2)\,dt_1 dt_2$$

$$(2.6c) \quad P_{01} = \Pr(d_i(1) = 0 \wedge d_i(2) = 1) = \int_{-\infty}^{-\bar{V}} \int_{-\bar{V}}^{\infty} f(t_1, t_2)dt_1dt_2$$

$$(2.6d) \quad P_{00} = \Pr(d_i(1) = 0 \wedge d_i(2) = 0) = \int_{-\infty}^{-\bar{V}} \int_{-\infty}^{-\bar{V}} f(t_1, t_2)dt_1dt_2.$$

Utilizing the four cells of data in the contingency table (only three of which yield independent information), one can estimate the parameters $\bar{V}$, γ and ρ by the method of maximum likelihood (abstracting from the important practical problem of empty cells). These estimators have desirable large sample properties.

The basic source of identification of the parameter γ comes from the following insight. If there is no true state dependence at the individual level ($\gamma = 0$), the probability that $d_i(2) = 1$ and $d_i(1) = 0$ (P_{01}) is the same as the probability that $d_i(2) = 0$ and $d_i(1) = 1$ (P_{10}). But if γ is positive, the first probability, P_{01}, exceeds the second probability, P_{10}. The ordering of the unconditional probabilities is used to infer the presence or absence of true state dependence at the micro level. As previously established, an examination of conditional probabilities is not informative on the matter.

Another way to state the same point is to consider the regions of integration for the density $f(t_1, t_2)$ used to define the probabilities P_{01} and P_{10}. Figure 1 corresponds to the case of $\gamma = 0$. The area under the density in region DBC yields P_{01}. The area under the density in region D'BC' yields P_{10}. Under the assumption that the variance of $\varepsilon_i(1)$ is the same as that of $\varepsilon_i(2)$, an assumption consistent with the assumption of underlying stationarity in the distribution of the latent variables, B lies on a 45° line from the origin, and $P_{01} = P_{10}$.

Next, consider the case in which $\gamma \neq 0$, and suppose that it is positive in keeping with our previous example. The appropriate regions of integration are DBC (for P_{01}) which is the same as in the previous diagram and C'B'D" (for P_{10}) which has a smaller area than D'BC' in Figure 1. The reduction in area is given by the strip D"B'BD'. Accordingly, $P_0 > P_{10}$. (If $\gamma < 0$, $P_{01} < P_{10}$.)

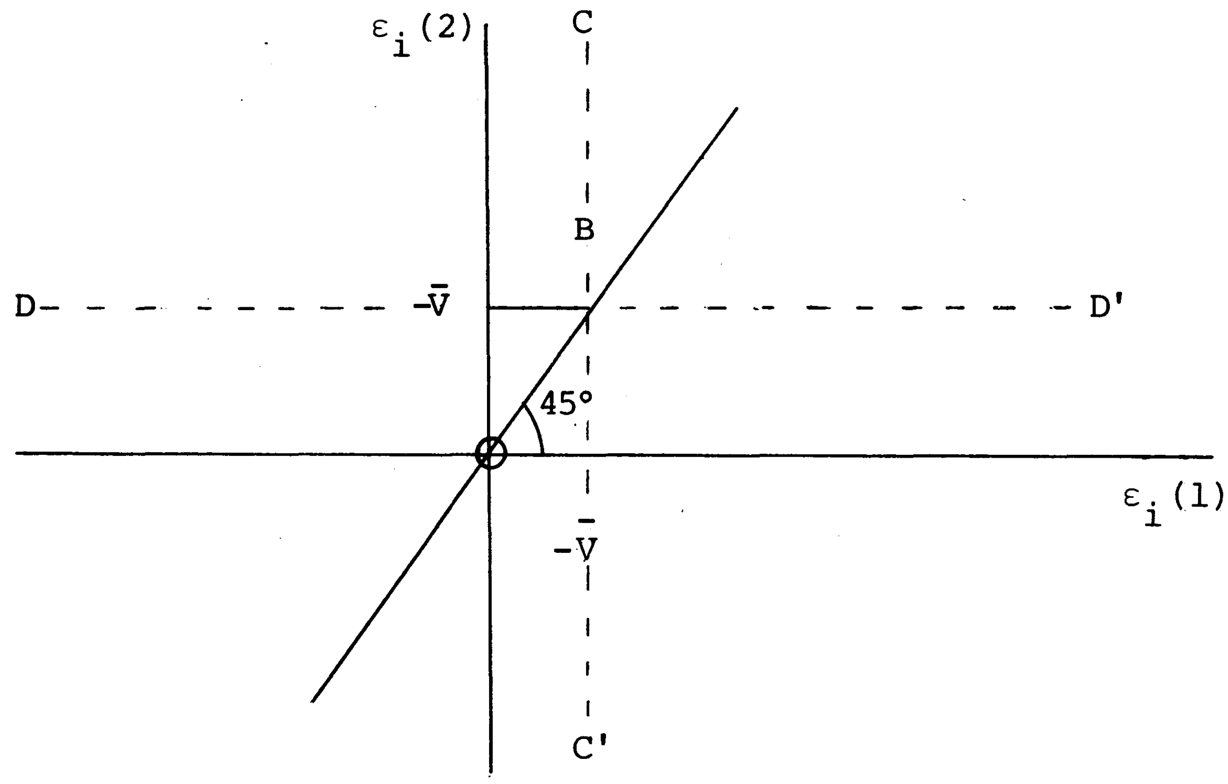

Figure 1: $\gamma = 0$

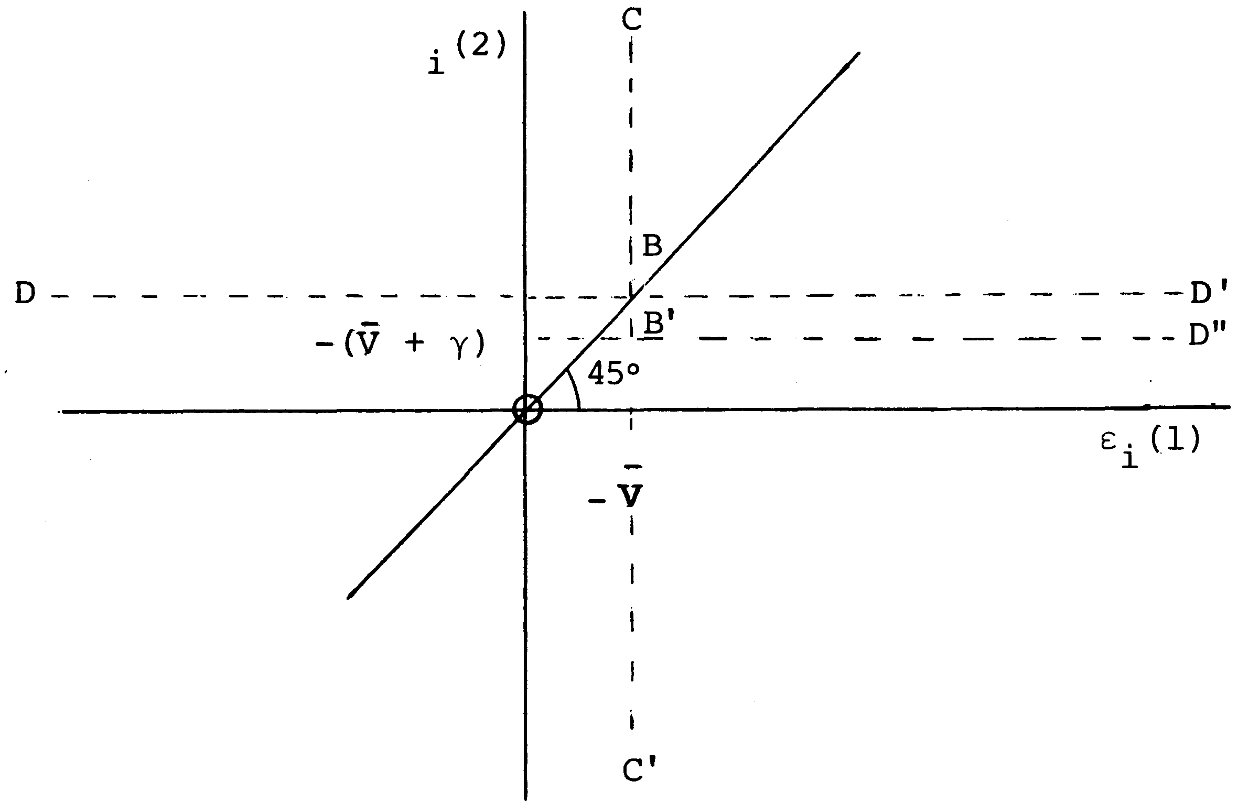

Figure 2: $\gamma > 0$

The main point to be drawn from this example is that sequences of unconditional probabilities provide the key to identification of the parameters γ and ρ. As long as the off-diagonal cells of the contingency table are not empty, one can estimate ρ, γ and $\bar{V}$. Thus, it is required that there be some individuals who change state. If $\rho = 1$, the off-diagonal cells of the contingency table would be empty, and one could not estimate γ (but one could estimate $\bar{V}$ from the first period marginal frequencies of the table). Recall that the case of $\rho = 1$

implies a perfect correlation in status over time, so that no individual would be observed to change state.

Having deliberately simplified the analysis in order to focus on main points, it is important to separate inessential from essential assumptions. Recall that we assumed (1) that before the first period, everyone was observationally identical (i.e., $V_i(1) = \bar{V}$), (2) the variance in $\varepsilon_i(2)$ is the same as the variance in $\varepsilon_i(1)$--an implicit assumption of stationarity in the underlying latent variables, (3) only two periods of data were available, and (4) since $\bar{V}$ is a constant, there is no correlation between $\bar{V}$ and $\varepsilon_i(1)$ and $\varepsilon_i(2)$. The consequences of relaxing each of these assumptions will now be briefly discussed in turn. The complete development of these points is the subject of this paper and a companion paper (1978b).

Rarely do we possess samples of observationally identical people. If we can measure differences, such heterogeneity creates no difficulty in applying the procedure previously sketched. Let the observed variables at time t for individual i be recorded in a 1xK vector $Z_i(t)$. The measured traits may change over time and may include both age and time related variables. If $V_i(t)$ is assumed to be a linear function of $Z_i(t)$, we may write

$$(2.7a) \quad X_i(1) = Z_i(1)\beta + \varepsilon_i(1)$$

and

$$(2.7b) \quad X_i(2) = Z_i(2)\beta + \gamma d_i(1) + \varepsilon_i(2)$$

where β is a Kx1 vector of coefficients. As before,

$$(2.8) \quad \begin{aligned} d_i(t) &= 1 \quad \text{iff} \quad X_i(t) > 0 \\ d_i(t) &= 0 \quad \text{otherwise for } t = 1,2,\ i = 1,\ldots,I. \end{aligned}$$

This is the multivariate probit model with structural shift in the equations (Heckman, 1978a). If the $Z_i(t)$ are independent of the $\varepsilon_i(t)$, the coefficients β, γ and ρ may be estimated under general conditions outlined in that paper. Permitting the $Z_i(t)$ to change in different time periods generates a time inhomogeneous stochastic process.

The second item on the list is less easily disposed of. Without some variation in sample measurable characteristics (the $Z_i(t)$), the restriction that the variance in $\varepsilon_i(1)$ is the same as that of $\varepsilon_i(2)$ is essential. Without this assumption, or without knowledge of how the variances are related, the parameters of the model are not identified (in particular, one cannot separately estimate γ and ρ).[1] However, if there is sample variation in relevant measurable characteristics ($Z_i(t)$), one can remove this restriction and test the hypothesis of a general nonstationary distribution for $\varepsilon_i(1)$ and $\varepsilon_i(2)$. (This follows from Heckman, 1978a, part III.) Conditions for identifiability in this model are analogous to identification conditions in an ordinary time series model with first order serial correlation and a lagged value of the dependent variable. In that model, identification of correlation and lag coefficient parameters is secured only through sample variation in exogenous variables.

The restriction to two periods of data in the previous example was made solely for expository convenience. Having more periods of data available serves to enlarge the potential for determining the true nature of intertemporal correlation in the latent variables $\varepsilon_i(t)$, and the time pattern in the structural shift parameter γ. These points will be developed below.

Finally, the assumption that $\bar{V}$ is independent of $\varepsilon_i(t)$ is not innocuous especially when it is recognized that in most practical problems individuals differ in their initial conditions. For example, most panel studies only record in detail a short stretch of an individual's life cycle. At best, an analyst has only a brief summary of the life history of the individual before the panel study is conducted. Controlling for individual differences in relevant presample experience by including regressor variables in the analysis gives rise to a problem if the regressors are correlated with the disturbances. If some of the regressors are related to past choices, and disturbances are serially correlated, it is plausible that there is correlation between the regressors and the current disturbances.

There are two cases here; one easy and one hard. Take the easy case first. Suppose that $V_i(t) = Z_i(t)\beta + E_i(t)\phi$ where $E_i(t)$ is correlated with $\varepsilon_i(t)$ while the variables in $Z_i(t)$ are not. If it is the case that we may write

(2.9) $\quad E_i(t) = \bar{E}_i(t) + \mu_i(t)$

where $\bar{E}_i(t)$ is a deterministic function and $\mu_i(t)$ is a normal random variable, the composite disturbance term for the model of equation (2.2) may be written as

(2.10) $\mu_i(t)\phi + \varepsilon_i(t)$

and one may proceed as in the analysis surrounding equations (2.7a) and (2.7b) with the composite term replacing the error term $\varepsilon_i(t)$, and the exogenous determinants of $E_i(t)$ included as regressors in an augmented $Z_i(t)$ vector. Note that under the null hypothesis of $\phi = 0$, this procedure is exact even if $\mu_i(t)$ is non-normal.

The hard case comes when $\mu_i(t)$ is non-normal and $\phi \neq 0$. In this case, the latent variables $(\mu_i(t)\phi + \varepsilon_i(t))$ are no longer normally distributed and the multivariate probit structure is inappropriate. While this more general model is theoretically well defined, it may be computationally intractable.

This section has discussed the basic idea underlying this paper, and a set of related issues that arise in applying the idea to data. The remainder of the paper is devoted to a more detailed examination of the set of issues discussed in this section, and to more general issues that arise in using panel data to estimate discrete dynamic models.

III. ALTERNATIVE MODELS FOR DISCRETE PANEL DATA AND THEIR IMPLICATIONS FOR RUNS PATTERNS

This section considers a variety of models for discrete panel data generated from alternative hypotheses about $V_i(t)$, $\varepsilon_i(t)$, and the nature of structural shift in the equations (γ in Section II). In investigating this menu of models, two questions are posed. These are: (1) When can cross section data be used to fully characterize a dynamic process, and if it cannot be so used, what information about the dynamic process can be retrieved from the cross section, and how many panel observations does one need to estimate the parameters of the model? (2) From observed sequences of discrete events, is it possible to infer the underlying stochastic structure of the unobservable random variables?

Both questions are important. The first because we are often restricted to using cross section data, and we frequently draw inferences from cross sections about individual dynamic processes. When do we err in this procedure? Also, in the rare case when we can design panel samples, it is helpful to know the length of the panel that is required in order to estimate the structure of a model. The second question is important because the temporal sequence of discrete events can be used to shed light on the underlying dynamic process (as shown in Section II). The answer to the second question is useful in enabling us to design simple regression estimators to make inferences about true dynamic processes.

All of the models considered in this section are special cases of the following general model

$$(3.1) \qquad X_i(t) = V_i(t) + \varepsilon_i(t), \quad t = 1,\ldots,T, \quad i = 1,\ldots,I.$$

$$d_i(t) = 1 \quad \text{iff} \quad X_i(t) > 0, \; d_i(t) = 0 \text{ iff } X_i(t) \leq 0.$$

$V_i(t)$ may be written as

$$(3.2) \qquad V_i(t) = Z_i(t)\beta + \sum_{\ell=1}^{t-1} \gamma(\ell) d_i(t-\ell) + \phi \sum_{j=1}^{t-1} \prod_{\ell=1}^{j} d_i(t-\ell)$$

where $Z_i(t)$ is independent of $\varepsilon_i(t)$ (so that all contemporaneous and predetermined variables with $\varepsilon_i(t)$ have been substituted out in reduced form), and $\gamma(\ell)$ is a coefficient that measures the effect of experience of the event ℓ periods ago on current values of $X_i(t)$. The final term on the right hand side of equation (3.2) measures the effect of the cumulated recent spell of experience in the state for those still in the state on the current value of $X_i(t)$. The second from the last term measures the effect of all previous experience of the event on the current value of $X_i(t)$.

A complete specification of the model requires a statement of the initial conditions assumed for $d_i(t)$. If the stochastic process described by equations (3.1) and (3.2) begins at time $t = 1$, the appropriate initial conditions are $d_i(j) = 0$, $j = 0, -1, \ldots$. However, if the stochastic process has been operating prior to time $t = 1$, one requires information on the relevant presample values of $d_i(j)$. In this paper, a reduced form approach is pursued. It is assumed that when presample values of $d_i(j)$ are required to estimate a model, they can be well approximated by linear stochastic functions of exogenous variables. The conditional expectations of these functions are allocated to the $Z_i(t)$ vector and the error term is allocated to the disturbance, $\varepsilon_i(t)$. An alternative and somewhat more satisfactory structural approach to this problem is pursued in a companion paper (Heckman, 1978b). However, as noted in the conclusion of Section II, the reduced form procedure utilized in this paper introduces no error in testing the null hypothesis of no state dependence.

When a distributional assumption is required for the disturbances, it is assumed that $\varepsilon_i(t)$ is normally distributed. Thus letting $\varepsilon_i = (\varepsilon_i(1),\ldots,\varepsilon_i(T))$,

$$\varepsilon_i \sim N(0, \Sigma)$$

where Σ is a TxT variance covariance matrix. Random sampling is assumed so that the ε_i are independent of all $\varepsilon_{i'}$, $i \neq i'$, $i = 1,\ldots,I$ where I is the number of individuals in the sample.

For notational convenience, array the $d_i(t)$, $t = 1,\ldots,T$ into a 1xT vector $d_i = (d_i(1),\ldots,d_i(T))$. Define $V_i(t)$ by

$$(3.3)\quad V_i(t) \equiv Z_i(t)\beta + \sum_{\ell=1}^{t-1} \gamma(\ell) d_i(t-\ell) + \phi \sum_{j=1}^{t-1} \prod_{\ell=1}^{j} d_i(t-\ell).$$

Array the $V_i(t)$ into a 1xT vector of the $V_i(t)$, i.e.,

$$(3.4)\quad V_i = (V_i(1),\ldots,V_i(T)).$$

Define D as the diagonal matrix formed from the diagonal of Σ. Define V_i^* by

$$(3.5)\quad V_i^* \equiv D^{-1/2} V_i.$$

Define correlation matrix Σ^* by

$$(3.6)\quad \Sigma^* \equiv D^{-1/2} \Sigma\, D^{-1/2}.$$

Finally, denote a 1xT vector of ones by ι. In this notation, the probability of d_i is

$$(3.7)\quad \Pr(d_i|Z_i) = F(V_i^* * (2d_i - \iota);\ \Sigma^* * [2d_i - \iota][2d_i - \iota]')$$

where the operation "*" denotes a Hadamard product,[2] and F is the cumulative distribution function of a T-variate multivariate normal.

Note that the operation in equation (3.5) scales each element of V_i by the appropriate standard deviation. For example, in period t,

$$V_i^*(t) = Z_i(t)\beta/\sigma_{tt}^{1/2} + \sum_{\ell=1}^{t-1} \frac{\gamma(\ell)}{\sigma_{tt}^{1/2}} d_i(t-\ell) + \frac{\phi}{\sigma_{tt}^{1/2}} \sum_{j=1}^{t-1} \prod_{\ell=1}^{j} d_i(t-\ell).$$

It is possible to estimate proportionalities among the variances since the elements of β are fixed in all time periods. Thus, if the normalization $\sigma_{11} = 1$ is adopted, and β is non-zero, one can estimate β, ϕ, and σ_{tt}, $t = 2,\ldots,T$. Also, it is clearly the case that if $\beta \neq 0$, one can estimate each element of $\gamma(\ell)$, $\ell = 1,\ldots,T-1$. (Recall that $\beta \neq 0$ is a necessary condition for identification of the model in the general nonstationary case.)

The sample likelihood function $\mathcal{L}$ is

$$(3.8)\quad \mathcal{L} = \prod_{i=1}^{I} F(V_i^* * (2d_i - \iota);\ \Sigma^* * (2d_i - \iota)(2d_i - \iota)').$$

This is to be maximized with respect to the parameters Σ^*, β, $\gamma(\ell)$, $\ell = 1,\ldots,T-1$, ϕ, and σ_{tt}, $t = 2,\ldots,T$ (assuming $\sigma_{11} = 1$). Under the conditions just specified, these parameters are identified. (For a complete analysis, see Heckman, 1978a, b. The second paper develops random and fixed factor analytic schemes that greatly simplify computation with the model.)

This model is sufficiently flexible to accomodate a wide variety of stochastic models that appear in the literature. For example, if $Z_i(t) = 1$, $\gamma(\ell) = 0$, $\ell = 2,\ldots,T-1$, $\phi = 0$, and $\varepsilon_i(t)$ is independently identically distributed, equations (3.1) and (3.2) generate a time homogeneous first order Markov process. If $\varepsilon_i(t)$ has a components of variance structure ($\varepsilon_i(t) = \eta_i + U_i(t)$, where $U_i(t)$ is iid), and all the other previous assumptions are maintained, a compound first order Markov process is generated by equations (3.1) and (3.2). If $\phi \neq 0$, $\gamma(\ell) = 0$, $\ell = 1,\ldots,T-1$, $Z_i(t) = 1$, and $\varepsilon_i(t)$ is iid, a renewal process is generated by equation (3.2). If $\phi = 0$, $\gamma(\ell) = 0$, $\ell = 1,\ldots,T-1$, $Z_i(t) = 1$, and $\varepsilon_i(t)$ follows an ARMA serial correlation scheme, Coleman's latent Markov model (1964) emerges. If $\phi = 0$, $\gamma(\ell) = 0$, $\ell = 1,\ldots,T$, $Z_i(t) = 1$, and $\varepsilon_i(t)$ is iid, a simple Bernoulli model results. In all these models, heterogeneity in measured characteristics can be introduced by admitting nontrivial regressor variables into the $Z_i(t)$ vector. In particular, age and year effects can be incorporated in $Z_i(t)$.

In order to focus the discussion, we consider a sequence of models starting with the simplest and most familiar: a homogeneous Bernoulli model.

III.1 Homogeneous Bernoulli Model

Let $V_i(t) = \bar{V}$, and assume that $\varepsilon_i(t)$ is iid. Each person has an identical probability of experiencing an event ($d_i(t) = 1$) in each of T sample periods. That probability is

$$(3.9) \quad \Pr(\varepsilon_i(t) > -\bar{V}) = \Phi(\bar{V}/\sigma_\varepsilon) = \bar{P}.$$

The expected duration in the state for someone just entering the state is

$$(3.10) \quad \frac{1}{1-\bar{P}}.$$

The expected number of periods in the state is $\bar{P}T$ with variance $(\bar{P})(1-\bar{P})T$. Cross section estimators of $\bar{P}$ taken from a sample of observations at a point in time are consistent estimators of $\bar{P}$. A long time series on one person or a large cross section at a point in time yield equally efficient estimators if sample sizes are equal. Panel data are not required to estimate $\bar{P}$.

In T trials, the probability of j "successes" ($d_i(t) = 1$) and T - j "failures" is

$$\binom{T}{j}\bar{P}^j(1-\bar{P})^{T-j}.$$

The random variables $d_i(t)$, $t = 1,\ldots,T$ are "exchangeable" (see Feller, 1971), p. 228) as are the $\varepsilon_i(t)$. The probability of any sequence of "runs" with j successes is the same as any other sequence with j successes.

This model can be modified to take account of known differences in personal characteristics. It is not necessary to assume that $V_i(t) = \bar{V}$. If $V_i(t)$ is assumed to be a linear function of known exogenous variables that are independent of $\varepsilon_i(t)$, one may write

$$V_i(t) = Z_i(t)\beta.$$

Depending on the content of the $Z_i(t)$ "regressor" vector, one may generate a nonstationary time inhomogeneous stochastic process at the micro level. For example, the $Z_i(t)$ vector may include "age." As before, one can estimate the parameters of the model from a cross section or a time series on a single person using standard probit analysis provided that there is sufficient sample variation in the regressor matrix $Z_i(t)$. For example, if education is included as a regressor in $Z_i(t)$, and education does not change over the sample period, a time series for one person would not afford estimates of the coefficient of education. If there are genuine "year" effects--such as the effect of unemployment on participation, cross section data would not permit estimation of an effect of aggregate unemployment on participation. If there are K parameters in the β vector, one requires at least K distinct configurations of data (i.e., K linearly independent $Z_i(t)$ vectors).

Of course, if the variables included in $Z_i(t)$ change over the sample period, the previous discussion of "runs" patterns is invalid. Depending on the content of the $Z_i(t)$ vectors, it is possible to generate any ordering of the probabilities for the various runs patterns.

III.2 A Random Effects Bernoulli Model $V_i(t) = \bar{V}$

$\varepsilon_i(t)$ has a components of variance structure.

$$(3.11) \quad \varepsilon_i(t) = \phi_i + U_i(t)$$

where $U_i(t)$ is iid with mean zero and variance σ_u^2. ϕ_i is a random variable with mean zero and variance σ_ϕ^2 drawn from distribution $f(\phi)$. $\varepsilon_i(t)$ is a convolution of two random variables.

Over sample period T, a woman carries a fixed component ϕ_i. Given ϕ_i, the probability that person i experiences the event at time t ($d_i(t) = 1$) is

$$(3.12) \quad \Pr(\varepsilon_i(t) > -\bar{V}|\phi) = \Pr(U_i(t) > -(\bar{V} + \phi)).$$

The average probability in the population is

$$(3.13) \quad \bar{P} = \Pr(\varepsilon_i(t) > -\bar{V}) = \int \Pr(U_i(t) > -(\bar{V}_i + \phi))f(\phi)d\phi.$$

The average probability can be determined from a cross section as can the expected number of periods in the state $\bar{P}T$, but the variance of $\bar{P}T$ cannot be estimated from a cross section. Cross section estimators of $\bar{P}$ taken from a sample of observations at a point in time are consistent estimators of $\bar{P}$. A long time series on person i or a large cross section at a point in time lead to estimates of different parameters. If both samples become large, the first sample estimates $\frac{\bar{V} + \phi_i}{(\sigma_u^2)^{1/2}}$ while the second estimates $\frac{\bar{V}}{(\sigma_u^2 + \sigma_\phi^2)^{1/2}}$. Panel data on a sample of individuals are required to estimate $\frac{\sigma_\phi^2}{\sigma_u^2 + \sigma_\phi^2}$ the intraclass correlation coefficient (see Heckman and Willis, 1975). Only two years of panel data are required.

The average duration in the state cannot be estimated from cross section data because

$$(3.14) \quad E_\phi((1 - P) \sum_{j=0}^{\infty} jP^j) = E_\phi(\sum_{j=1}^{\infty} P^j) \geq \sum_{j=1}^{\infty} (E_\phi(P))^j, \quad 0 < P < 1,$$

by Jensen's inequality. (E_ϕ denotes the expectation with respect to the density of ϕ, $f(\phi)$.) Use of an estimated cross section probability to estimate the duration over-estimates the amount of sample turnover between states. Use of an estimated probability for each person (taken from a long series) leads to correct estimates of average duration in the state for that person.

In T trials, the probability of j "successes" and T-j "failures" is the same for any sequence with j successes in any order. This is a consequence of intertemporal exchangeability in the underlying random variable $\varepsilon_i(t)$.

As in the previous case, it is straightforward to relax the assumption that $V_i(t) = \bar{V}$. Making the same assumptions as were made in modifying model (3.1) to accommodate observed heterogeneity, one may write

$$(3.15) \quad V_i(t) = Z_i(t)\beta$$

and under the conditions previously given, β is estimable (see Heckman and Willis, 1975). As before, the results on the relative probabilities of various runs patterns are vitiated for an arbitrary sequence of $Z_i(t)$ regressors.

This model is based on a strictly stationary process that is not ergodic. Unlike the situation in stationary time series analysis, it is in principle possible to test for ergodicity in panel data precisely because the analyst has access to a multiplicity of time series on individuals.

III.3 A Model with General Correlation: $V_i(t) = \bar{V}$, $\varepsilon_i(t)$ Follows an Autoregressive-moving Average (ARMA) Process

This model is essentially the latent Markov model due to Coleman (1964). Define polynomials in the lag operator B(L) and C(L). Let

$$B(L)\varepsilon_i(t) = C(L)U_i(t) \tag{3.16}$$

where $U_i(t)$ is white noise, and C(L) is invertible and the roots of $|C(L)| = 0$ and $|B(L)| = 0$ lie outside the unit circle so that the process is stationary.

The probability that person i experiences an event at time t ($d_i(t) = 1$) is

$$\bar{P} = \Pr(\varepsilon_i(t) > -\bar{V}) = \Phi(\frac{\bar{V}}{g}) \tag{3.17}$$

where g^2 is the variance of $B^{-1}(L)C(L)U_i(t)$. This is the average probability in the population. Over sample period T, the expected number of periods in the state is $\bar{P}T$, a number that can be estimated from a cross section. Panel data are required to estimate the variance normalized elements of B(L) and C(L).

The average duration in the state cannot be estimated from cross section data because

$$\sum_{j=0}^{\infty} jP(j) \geq \frac{1}{1 - \bar{P}} \tag{3.18}$$

(where P(j) is the probability of being in the state j successive time periods) if the intertemporal correlations among all disturbances are positive.[3] Thus, as in the previous case, actual turnover is less than the turnover predicted from cross section probability $\bar{P}$.

In T trials, the probability of j "successes" and T-j "failures" is not the same for any sequence with j successes in any order. This is a consequence of the lack of exchangeability in the underlying random variable $\varepsilon_i(t)$. As a consequence of assuming stationarity in the time series of latent variables, "mirror image" sequences have identical probabilities. For an example, a sequence of trials recorded as (1, 0, 1, 1) has the same probability as a sequence (1,1,0,1).

This is a consequence of the intrinsic time reversibility of a stationary series. The term "mirror image" is derived from the following thought experiment: imagine putting the first sequence up against a mirror and noting its reflection.

This result on runs for any stationary process for the latent variables is of sufficient interest to warrant an independent discussion. While the result is true for any stationary process, it is most easily proved for the case of Gaussian variables.

Theorem. For any stationary Gaussian process, mirror image sequences of discrete events are equiprobable.

Proof. Let ε' be a 1xT vector with elements $\varepsilon(1),\ldots,\varepsilon(T)$. From the assumed normality we may write

$$\varepsilon \sim N(0, \Sigma)$$

where Σ is a TxT covariance matrix. From stationarity, $\sigma_{ij} = \sigma(|i - j|)$. The "mirror image" of ε is $\tilde{\varepsilon}$ defined by

$$\tilde{\varepsilon} = P\varepsilon$$

where P is a skew diagonal permutation matrix ($P_{ij} = 1$ for $j = T-1 + 1$, $P_{ij} = 0$ otherwise). Then

$$\tilde{\varepsilon} \sim N(0, P\Sigma P')$$

but from stationarity, $P\Sigma P' = \Sigma$ since $\sigma_{T-i,\, T-j} = \sigma(|i - j|) = \sigma_{i,j}$. The theorem follows immediately. QED.[4]

As in the other cases, this result on "runs" is invalid if the exogenous variables $Z_i(t)$ change in an arbitrary way over time. If the highest order term in the lag operator B(L) is p and the highest order term in the lag operation C(L) is q, in general p + q periods of panel data are required in order to identify the elements of the operators B(L) and C(L). If there are factors common to both operators, these cannot be identified and less than p + q periods of panel data are required (two periods for each commonality in factors).

Several special cases of the ARMA model are of interest in practical work. The simplest model is a first order Markov process $B(L) = 1 - \rho L$, $C(L) = 1$, or a moving average model, $B(L) = 1$, $C(L) = 1 - bL$. A model that combines the "permanenttranitory" model with the serial correlation model has $B(L) = 1 - \rho L$, $C(L) = 1 - L$. Note that this error process is not invertible, and is not ergodic. Using panel data, one can test for this model against alternative ergodic models.

III.4 Models with True State Dependence

Let $V_i(t) = \bar{V}$, $\varepsilon_i(t)$ is iid with mean zero.

$$(3.19)\quad X_i(t) = \bar{V} + \gamma \sum_{j=1}^{t-1} d_i(j) + \varepsilon_i(t).$$

$$(3.20)\quad d_i(j) = 1 \quad \text{iff} \quad X_i(j) > 0$$
$$= 0 \text{ otherwise, } j = 1,\ldots,T.$$

It is assumed that the process begins at $t = 1$.

The probability that person i experiences the event at time t ($d_i(t) = 1$) is

$$(3.21)\quad \bar{P}_i(t) = \Pr\left(\varepsilon_i(t) > -\bar{V} - \gamma \sum_{j=1}^{t-1} d_i(j)\right)$$
$$= \Phi\left(\frac{\bar{V} + \gamma \sum_{j=1}^{t-1} d_i(j)}{\sigma_\varepsilon}\right).$$

The average probability in the population at time t is

$$(3.22)\quad \bar{P} = \sum_{j=0}^{t-1} \Phi\left(\frac{\bar{V} + \gamma j}{\sigma_\varepsilon}\right) \tilde{P}(j,\ t-1),$$

where $\tilde{P}(j, t-1)$ is the probability of j spells of the event in the t-1 previous time periods. For $\bar{V} > 0$ and $\gamma > 0$, $\Phi(\ )$ is a concave function of j so that from Jensen's inequality

$$(3.23)\quad \bar{P} < \Phi\left(\frac{\bar{V} + \gamma \sum_{j=1}^{t-1} E(d_i(j))}{\sigma_\varepsilon}\right).$$

In this model, a cross section probit analysis can be used to estimate $\bar{V}/\sigma_\varepsilon$ and $\gamma/\sigma_\varepsilon$ if one knows the number of past spells of the event for each person in the sample. However, from the nonlinearity of Φ, use of the representative individual's previous experience does not lead to a reliable prediction of the average probability in the sample.

From a single cross section, one can estimate the structural parameters of the model, and hence one can estimate the expected number of times the individual will experience the event over horizon T and the expected duration in the state for each individual. Panel data are not required to estimate any parameter of the model. Also, subject to standard conditions for identification in a probit model, a time series of observations of length greater than two for a single person will be enough to estimate the parameters of the model.

In T trials, the probability of j "successes" and T-j "failures" is not the same for any sequence of j successes in any order. Because of the time irreversibility inherent in the nonstationary process induced by the random variable $\sum_{j=0}^{t-1} d_i(j)$, "mirror image" sequences will not have identical probabilities. Assuming $\gamma > 0$, a sequence (1,0,1,1) is more probable than the sequence (1,1,0,1). Intuitively, if occupancy of a state raises the probability of future occupancy, a later "drop out" is less likely than an earlier one.

Thus,

$$\text{Probability }(1,0,1,1) = \Phi(\bar{V}/\sigma_\varepsilon)\Phi(-\frac{(\bar{V}+\gamma)}{\sigma_\varepsilon})\Phi(\frac{\bar{V}+\gamma}{\sigma_\varepsilon})\Phi(\frac{\bar{V}+2\gamma}{\sigma_\varepsilon})$$

$$\text{Probability }(1,1,0,1) = \Phi(\bar{V}/\sigma_\varepsilon)\Phi(\frac{\bar{V}+\gamma}{\sigma_\varepsilon})\Phi(-\frac{(\bar{V}+2\gamma)}{\sigma_\varepsilon})\Phi(\frac{\bar{V}+2\gamma}{\sigma_\varepsilon}).$$

Since $\gamma > 0$, the first sequence is more probable. (Compare the second term in the first sequence with the third term in the second sequence. The product of the other terms in each sequence is the same for both sequences.)

The conclusion of this analysis is robust with respect to the introduction of heterogeneity (of the type discussed in subsection III.2 above) into the model. To see this, simply add permanent component ϕ to the model (following the notation in subsection III.2) and write

Probability (1,0,1,1) =

$$= \int\Phi\left(\frac{\bar{V}+\phi}{(\sigma_u^2+\sigma_\phi^2)^{1/2}}\right)\Phi\left(-\frac{(\bar{V}+\gamma+\phi)}{(\sigma_u^2+\sigma_\phi^2)^{1/2}}\right)\Phi\left(\frac{\bar{V}+\gamma+\phi}{(\sigma_u^2+\sigma_\phi^2)^{1/2}}\right)\Phi\left(\frac{\bar{V}+2\gamma+\phi}{(\sigma_u^2+\sigma_\phi^2)^{1/2}}\right)f(\phi)d\phi$$

Probability (1,1,0,1) =

$$= \int \Phi\left(\frac{\bar{V} + \phi}{(\sigma_u^2 + \sigma_\phi^2)^{1/2}}\right)\Phi\left(\frac{\bar{V} + \gamma + \phi}{(\sigma_u^2 + \sigma_\phi^2)^{1/2}}\right)\Phi\left(-\frac{(\bar{V} + 2\gamma + \phi)}{(\sigma_u^2 + \sigma_\phi^2)^{1/2}}\right)\Phi\left(\frac{\bar{V} + 2\gamma + \phi}{(\sigma_u^2 + \sigma_\phi^2)^{1/2}}\right) f(\phi)d\phi$$

A term by term comparison of the kernels of the integrals reveals that the second sequence is less probable than the first.

The same inequality does not hold if the unobserved component of the latent variable follows a general ARMA process. In this case it is not possible to order the sequences of discrete events, but the key result that different sequences of runs are of different probability remains intact.

The assumption that past experience in a state has the same effect on the probability of current occupancy of the state no matter when the experience occurred (a constant γ in equation (3.19)) is a strong and intuitively unacceptable description of behavior in many situations. A more general model would allow for depreciation of the effect of previous state occupancy. The simplest such model would be a specialization of equations (3.1) and (3.2)

$$(3.24) \quad X_i(t) = Z_i(t)\beta + \sum_{\ell=1}^{t-1} \gamma(\ell) d_i(t-\ell) + \varepsilon_i(t).$$

If $Z_i(t) = 1$, $\gamma(\ell) = 0$, $\ell = 2,\ldots,T$, and $\varepsilon_i(t)$ is iid, this equation generates a first order Markov process. The only past experience relevant to current choices is the experience of the event in the preceding period. A second order process constrains $\gamma(\ell) = 0$, $\ell = 3,\ldots,T$. One can admit unobserved heterogeneity into the model by adopting a components of variance scheme for $\varepsilon_i(t)$ following the analysis of subsection III.2 or by introducing a more general correlation along the lines of the ARMA scheme of subsection III.3. One can admit time and period effects and measured differences in personal characteristics by admitting nontrivial regressors to enter the $Z_i(t)$ vector.

As an alternative to the classical Markov model, one can introduce a "geometric Markov" model in which $\gamma(\ell)$ is parameterized so that

$$\gamma(\ell) = n(m)^\ell, \; n > 0, \; 0 < m < 1,$$

so that depreciation of past experience is geometric.

In all of these Markov-type models, if the assumption that $\varepsilon_i(t)$ is iid is maintained, cross section data can be used to estimate the parameters of the model provided that both the extent and timing of the relevant previous experience are known. In order to separately identify n and m, at least two years of past experience are required, and all individuals cannot have the same experience.

If the iid assumption for $\varepsilon_i(t)$ is relaxed, for two reasons a cross section cannot be used to estimate the parameters of the model. First, cumulated experience is no longer independent of the error term in the probit function. Second, panel data are required to estimate the correlation structure of the latent variables. Methods for estimating this more general model are presented in a companion paper (Heckman, 1978b). Given the requisite retrospective data, more years of panel data are required to estimate the parameters of the model than were required in the model of subsection III.3. In fact, in certain cases one requires the entire history of a person to estimate the model. This problem is discussed at length in the companion paper.

In contrast with the Markov-type dependence just discussed, there is a conceptually distinct model more closely related to renewal theory (see Karlin and Taylor, 1975). Once the individual is in a state, an accumulation process begins. In the language of human capital theory, "specific capital" accumulation may arise. Accumulation continues until the individual leaves the state, at which time the "specific capital" is lost. A simple model that captures this notion is

$$(3.25) \quad X_i(t) = Z_i(t)\beta + \phi \sum_{j=1}^{t-1} \prod_{\ell=1}^{j} d_i(t-\ell) + \varepsilon_i(t).$$

"ϕ" measures the effect of specific capital on current $X_i(t)$. Once the individual leaves the state, the slate is wiped clean, and there is no memory (Jovanovic, 1978).

Combining the models of equation (3.24) with those of equation (3.25) leads to the general model of equations (3.1) and (3.2). For a more complete description of the formulation and estimation of such models see Heckman (1978b).

III.5 True State Dependence versus Spurious State Dependence

The principle distinction between the models developed in subsections III.1 - III.3 and those presented in subsection III.4 is that in the first group of models, previous choices (or events) do not directly determine current choices while in the second group of models they do. This is not to say that there is no correlation between current and past choices in the first group of models. With the exception of the Bernoulli model of subsection III.1, there is such correlation. In the first group of models, a correlation between events arises because of serial correlation in unobservables while in the second group of models the correlation arises, at least in part, because of a direct effect of past choices on current behavior. It is this latter, structural, dependence that is termed "true" state dependence in this paper while the former "statistical" or "correlational" dependence is termed spurious state dependence. For both reasons, one would estimate that the conditional probability of experiencing an event in the current period is a function of previous experience of the event.

IV. USE OF RUNS TO DISCRIMINATE AMONG MODELS

This section collects the results on "runs" from models outlined in subsections III.1 - III.4 to demonstrate how empirical knowledge of these runs can be used to discriminate among certain alternative models. The models presented in subsections III.1 and III.2 are based on exchangeable random variables. In T trials, all runs with j successes in any order have equal probability of occurence. The model outlined in subsection III.3 provides an example of a discrete model based on stationary non-exchangeable random variables. "Mirror image" sequences are equally likely but all sequences with j successes are not. The model presented in subsection III.4 with a common value of γ for all lagged values of $d_i(t-j)$ is an example of a nonstationary sequence. In this case, even mirror image sequences are not equally likely. Given data on runs in samples of sufficient size that sample proportions yield reliable estimates of population probabilities, it is possible to use the runs data to discriminate among these models.[6]

A simple example will serve to fix ideas. Let the sample time period (T) be three. The probabilities of different runs for the different models are tabulated in Table 1. As before, the first element in each row refers to the first event. Thus (1,1,0) corresponds to "success, success, failure."

TABLE 1

PROBABILITIES OF RUNS FOR ALTERNATIVE MODELS

Model III.1 Homogeneous Bernoulli Model	P(110) = P(011) = P(101) P(100) = P(001) = P(010)
Model III.2 Heterogeneous Bernoulli Model	P(110) = P(011) = P(101) P(100) = P(001) = P(101)
Model III.3 Stationary Latent Variable Model "Latent Markov model"	P(110) = P(011) $\neq$ P(101) P(001) = P(100) $\neq$ P(010)
Model III.4 Model with True State Dependence, Latent Variables Uncorrelated, $\gamma > 0$, γ the same for all periods.	P(011) > P(101) > P(110) P(001) > P(010) > P(100)
Model III.4' Model with True State Dependence, Latent Variables Correlated, $\gamma > 0$, γ the same for all periods.	All probabilities distinct but unordered.

A nested test among these competing models can be based on the following simple idea. Each run pattern has an associated multinomial probability. Estimated multinomial proportions have a well defined asymptotic distribution (see, e.g., Rao, 1973). Utilizing a set of nested hierarchical tests, one can first test whether or not probabilities are distinct but unordered (model III.4'). If that hypothesis is rejected, one can test if model III.3 is concordant with the data, and so on until one reaches the most restrictive models of Table 1 (i.e., models III.1 and III.2). An empirical example of this procedure is presented in the next section. This test procedure can readily be generalized to panel data in which more than three observations per person appear.

In order to test between the homogeneous Bernoulli model and the heterogeneous Bernoulli model, note that in the homogeneous model the probability of occurence of the event, $\bar{P}$, can be estimated from a pooled cross section. The ratio of the recorded number of success to the total number of trials is an unbiased estimator of $\bar{P}$. The estimated probability can be used to predict the proportion of sample observations in each runs

runs pattern, including the (111) and (000) sequences. A standard Pearson χ^2 test can be used to test the hypothesis that a homogeneous Bernoulli model generates the data. As a consequence of Jensen's inequality, if a heterogeneous Bernoulli model fits the data better, use of estimated values of $\bar{P}$ to predict runs patterns will result in an under-prediction of the proportion of the sample that does not change state and an over-prediction of the proportion that does. An example of this phenomenon is offered in the next section.

A practical difficulty with these simple testing procedures is that people are rarely observationally identical. This problem can be surmounted if competing models are estimated by the method of maximum likelihood. As discussed in Section III, a wide variety of alternative models are nested within the general model of equations (3.1) and (3.2). The full development of such models is deferred to another paper (Heckman, 1978b).

The question considered in this paper is whether or not it is possible to use simple estimators to explore data before computationally more cumbersome maximum likelihood methods are employed.

Fortunately, the answer to this question is in the affirmative. The probability of a given sequence of events can be written as

$$\text{(4.1)} \quad \Pr(d_i(1),\ldots,d_i(T) \mid Z_i(1),\ldots,Z_i(T)) = F(Z_i(1),\ldots,Z_i(T) \mid d_i(1),\ldots,d_i(T))$$

where $F(\cdot)$ is the cumulative multivariate distribution function of the latent variables $\varepsilon_i(1),\ldots,\varepsilon_i(T)$ corresponding to the dichotomization of these variables (given the exogenous variables) that generates the given sequence of discrete events. The value of this multivariate integral depends on the correlation among unobservables, the values of the structural parameters of the model, and the values of the exogenous variables. For each of the 2^T possible sequences of outcomes, one may define a dummy variable $D(d_i(1),\ldots,d_i(T))$ that equals one when a given sequence occurs and is zero otherwise. Clearly

$$\text{(4.2)} \quad \sum_{d_i(1)=0}^{1} \cdots \sum_{d_i(T)=0}^{1} D(d_i(1),\ldots,d_i(T)) = 1.$$

Each of the $F(\cdot)$ functions corresponding to a unique sequence of events can be approximated by a power series expansion in the exogenous variables. If only the linear terms of the expansion are retained, this procedure produces a linear probability model. To simplify the exposition, suppose that linear terms of the expansion provide a good approximation in the sense that the remainder term of each expansion for each sequence is negligible. This assumption can always be relaxed by adding higher order terms to the model.

Collect the variables $Z_i(1),\ldots,Z_i(T)$ into a "super vector" Z_i, and write the regression model

$$(4.3)\quad D(d_i(1),\ldots,d_i(T)) = Z_i\beta(d_i(1),\ldots,d_i(T)) + \varepsilon_i(d_i(1),\ldots,d_i(T))$$

where $\beta(d_i(1),\ldots,d_i(T))$ is a suitably dimensioned vector of coefficients of distinct exogenous variables corresponding to the first partials of the cumulative distribution function written in equation (4.1). $\varepsilon_i(d_i(1),\ldots,d_i(T))$ is a random variable with mean zero and variance $Z_i\beta(d_i(1),\ldots,d_i(T))\cdot(1 - Z_i\beta(d_i(1),\ldots,d_i(T))$ (see Goldberger, 1964). The disturbance in each equation is defined so that

$$E(\varepsilon_i(d_i(1),\ldots,d_i(T))\cdot Z_i|Z_i) = 0$$

where "$E(a|b)$" denotes the conditional expectation of a given b.

Restriction (4.2) coupled with equation (4.3) implies that for arbitrary values of Z_i,

$$(4.4)\quad \sum_{d_i(1)=0}^{1} \cdots \sum_{d_i(T)=0}^{1} Z_i\beta(d_i(1),\ldots,d_i(T)) = 1.$$

Thus, summing across all the equations for the probability of each participation sequence, the sum of the intercepts must be one and the sum of slope coefficients must be zero. This condition is familiar in budget studies (e.g., Kmenta, 1971 or Theil, 1971) and is explicitly formulated for the model discussed here in Lee, Judge and Zellner (1970, pp. 183-190). This constraint can be imposed by dropping any one of the 2^T equations.

One can fit the generalized linear probability model and estimate the $\beta(d(1),...,d_i(T))$ vectors by least squares. Generalized least squares estimators may be calculated following the procedures outlined in Lee, et al. They also develop restricted estimators that bound the value of estimated probabilities to lie within the unit interval.

The hypotheses of equal probability for certain runs patterns given the Z_i can now be translated into hypotheses about the $\beta(\cdot)$ vectors associated with each runs patterns. Nested tests for equality among runs can be based on nested tests for equality for the associated regression coefficients (see, e.g., Theil, 1971, pp. 312-317). This point is best made by way of an example. Table 2 demonstrates how the hypotheses of Table 1 may be tested within the general regression model of Lee, et al.

TABLE 2

THE HYPOTHESES OF TABLE 1 TRANSLATED INTO THE GENERALIZED LINEAR PROBABILITY MODEL

Model	Test
Model III.1	$\beta(110) = \beta(011) = \beta(101)$
Model III.2	$\beta(100) = \beta(001) = \beta(010)$
Model III.3	$\beta(110) = \beta(011)$
	$\beta(100) = \beta(001)$
Model III.4 Model III.4'	All regression coefficients across equations are distinct

Using the appropriate estimated variance-covariance matrix for the GLS estimators, one can test the nested hypotheses using χ^2 statistics utilizing the asymptotic normality of the estimators. An example of this procedure and its value in determining the appropriate model is provided in the next section.[7]

It is important to note that the runs tests proposed in this section have been used to distinguish among only a limited subset of the models discussed in Section III. In particular, no runs tests have been offered here that permit analysts to distinguish between models with nonstationary disturbances without true state dependence and models with stationary errors with true state dependence. All of the models considered in this section assume stationarity in the latent variables. Methods are developed elsewhere (Heckman, 1978b) that permit analysts to distinguish between these two models. These methods rely critically on the assumption that individuals differ in measured exogenous variables.

Note further that only one type of "true" state dependence has explicitly been considered. This is the form of dependence that assumes that previous experience has the same effect on current choices no matter when it occurs (i.e., in the notation of equation (3.2), $\gamma(\ell) = \gamma$, $\ell = 1,\ldots,T-1$). This assumption is relatively innocuous. Assuming stationarity in the disturbances, no model with true state dependence generates the complete set of runs patterns that are associated with the Bernoulli models or the stationary latent variable model. Accordingly, runs tests that reject these models indicate the presence of true state dependence although they do not reveal the precise form of the true state dependence.[8]

Finally, it is important to note that in order to utilize the generalized linear probability model to estimate models with true state dependence, a careful treatment of relevant presample experience of the event is required. The procedure pursued in this paper is to replace the relevant presample experience with a set of exogenous variables that determine presample experience. As noted at the end of Section II, this procedure is always correct under the null hypothesis of no true state dependence and requires no special assumptions.

V. TWO APPLICATIONS OF THE RUNS METHODOLOGY TO DATA

This section presents two examples of the use of the runs procesures discussed in the preceding section. The examples are drawn from an extensive empirical analysis of the dynamics of female labor force participation. The empirical work presented here is offered solely for the purpose of illustrating techniques. However, the empirical results reported in this paper are in accord with the empirical results reported in a more comprehensive analysis presented in a companion paper (Heckman, 1978c).

The first example is a simple application of runs tests that ignore measured heterogeneity. The second example is an application of runs tests that control for measured heterogeneity utilizing the regression methods discussed in the preceding section. The two empirical examples pose an interesting contrast. Without controlling for measured heterogeneity, one cannot reject the null hypothesis of no true state dependence; controlling for measured characteristics, one can.

Both examples are based on a sample of 198 women age 45-59 in 1968 who were married to the same spouse in the seven years of panel data drawn from the probability sample of the Michigan Panel Survey of Income Dynamics. For a complete description of these data, the reader is referred to Morgan, et al (1974). Three years of labor force participation histories are utilized here (for the years 1968, 1969 and 1970). Participation is defined to occur if a woman works for money sometime during the survey year.

The basic runs data for the analysis are presented in the second column of Table 3. "0" corresponds to no work in the year and "1" corresponds to work. It is clear from the table that roughly 40 percent of the women work all three years and roughly 40 percent never work.

In the first example, women are assumed to be observationally identical in the relevant control variables. This assumption enables us to use simple runs tests proposed in the preceding section.

First, consider the test for a homogeneous Bernoulli model (model III.1). The average participation probability in the three year sample is .48. Under the maintained hypothesis, this is the estimated probability of participation for each woman for each year. Inspection of the third column of Table 3 reveals that the homogeneous Bernoulli model underpredicts the "no turnover" cells (000 and 111) and overpredicts the turnover cells. This is the usual manifestation of the mover-stayer problem. The χ^2 statistic clearly rejects the hypothesis of a homogeneous Bernoulli model.

Next, consider the random effects Bernoulli model (model III.2). This model predicts equal numbers of observations in turnover cells with the same number of successes (in any order).

TABLE 3

χ^2 TESTS FOR ALTERNATIVE MODELS

Runs Patterns (1968, 1969, 1970)	Number of Observations in Each Cell	Number Predicted under Homogeneous Bernoulli Model	Number Predicted under Random Effects Bernoulli Model	Number Predicted for General Stationary Latent Variable Model
0 0 0	87	27.8	. . .	. . .
0 0 1	5	25.7	4-2/3	4.5
0 1 0	5	25.7	4-2/3	. . .
1 0 0	4	25.7	4-2/3	4.5
1 1 0	8	23.7	6-1/3	9
1 0 1	10	23.7	6-1/3	. . .
0 1 1	1	23.7	6-1/3	9
1 1 1	78	21.8	. . .	. . .
χ^2	. . .	362.6[a]	7.20[c]	.33[e]
Theoretical χ^2 at 5%	. . .	12.60	9.49	5.99
Significance level	. . .	(6 d. f.)[b]	(4 d. f.)[d]	(2 d. f.)[f]

[a] χ^2 is computed by comparing the actual number in each cell to the predicted number, squaring the difference, dividing by the predicted number in each cell, and summing across cells. The predicted probability of work in each year is .48.

[b] Six degrees of freedom result from the fact that one degree of freedom is used to estimate the sample proportion, and there are only seven independent cells.

[c] Computed by comparing the predicted number, obtained by forming the appropriate "1 out of 3" or "2 out of 3" average and comparing with the actual number.

[d] Two degrees of freedom are lost because predicted proportions are estimated for each of the two patterns.

[e] Predicted from "mirror image" cells (001 and 100, 110 and 011). For each mirror image configuration, a mean number is computed which is then compared with the actual number in the usual fashion.

[f] A degree of freedom is lost in each configuration because the predicted number in each sequence is estimated from the sample.

The empirical results are displayed in column 4 of Table 3. A Pearson χ^2 statistic does not lead to rejection of this hypothesis.

Finally, the results for the general stationary latent variable model (model III.3) are displayed in the fourth column of Table 3. This model fits the data better in the sense of having a lower Pearson χ^2 than any other model that appears in the table.

A clear conclusion from the simple analysis is that a heterogeneous Bernoulli model describes the data quite well, but a general stationary latent variable model does even better although the differences in fit are not statistically significant. There is certainly little evidence of "true" state dependence in these data.

A crucial, and highly questionable assumption in the preceding analysis is that women are observationally identical in relevant control variables. When this assumption is relaxed, one rejects the null hypothesis of no state dependence.

A simple regression procedure for testing the hypothesis of no true state dependence when observations differ in relevant control variables is outlined in Section IV. This procedure is based on the generalized linear probability model.

Although this model is thoroughly described in Lee, et al (1970), it may be useful to summarize the main details of their procedure as it is applied here. It is not necessary to restate their results on consistency and asymptotic normality of the estimators.

Utilizing the notation of Section IV, write the following eight equation system for each person in the sample

$$(5.1) \quad D(j, k, \ell)_i = Z_i\beta(j, k, \ell) + \varepsilon_i(j, k, \ell) \quad i = 1,\ldots,I \quad j, k, \ell = 0, 1.$$

The procedure to be used is as follows. First estimate each equation by least squares. Utilize the estimated first round estimates of $\beta(j, k, \ell)$ to estimate the interequation variance-covariance matrix for the disturbances for each observation. The variance-covariance matrix is

$\Omega_i =$

$$
= \begin{bmatrix}
Z_i\beta(111)(1-Z_i\beta(111)) & -Z_i\beta(111)Z_i\beta(011) & \cdots & -Z_i\beta(111)Z_i\beta(000) \\
-Z_i\beta(111)Z_i\beta(011) & Z_i\beta(011)(1-Z_i\beta(011)) & \cdots & \cdot \\
\vdots & -Z_i\beta(101)Z_i\beta(011) & \cdots & \ddots \\
-Z_i\beta(111)Z_i\beta(000) & -Z_i\beta(011)Z_i\beta(000) & \cdots & Z_i\beta(000)(1-Z_i\beta(000))
\end{bmatrix}
$$

For expositional convenience, the equations are ordered in a way corresponding to the ordering of cells in the first column of Table 3, except that the order of the equations is exactly inverse to the order of the cells presented in the table. Because of the accounting identity (4.2), Ω_i is singular. Deleting the final row and columns, corresponding to the final equation in the model, results in a reduced Ω_i, $\tilde{\Omega}_i$. The inverse of $\tilde{\Omega}_i$ is

$$(5.2) \quad B_i = \tilde{\Omega}_i^{-1} =$$

$$
= \begin{bmatrix}
\frac{1}{Z_i\beta(111)} + \frac{1}{Z_i\beta(000)} & \frac{1}{Z_i\beta(000)} & \cdots & \frac{1}{Z_i\beta(000)} \\
\frac{1}{Z_i\beta(000)} & \frac{1}{Z_i\beta(011)} + \frac{1}{Z_i\beta(000)} & & \frac{1}{Z_i\beta(000)} \\
\vdots & & \ddots & \vdots \\
\frac{1}{Z_i\beta(000)} & \frac{1}{Z_i\beta(000)} & \cdots & \frac{1}{Z_i\beta(001)} + \frac{1}{Z_i\beta(000)}
\end{bmatrix}.
$$

Array the $\beta(j, k, \ell)$ vectors into a grand vector, β, deleting the coefficients for the omitted equation, and preserving the ordering of equations previous established. The GLS estimator of β is

$$(5.3) \quad \hat{\beta} = [\sum_{i=1}^{I} (B_i \otimes Z_i'Z_i)]^{-1} [\sum_{i=1}^{I} B_i Y_i \otimes Z_i']$$

where Y_i is a 7x1 vector corresponding to the left hand side of equation (5.1) with the final equation, corresponding to D(000) deleted. In practice, B_i is estimated from the least squares estimators of equations (5.1). "$\otimes$" denotes a Kronecker product. "'" denotes transpose. A fairly direct extension of a theorem due to Amemiya (1977) proves that the one round GLS estimators for the generalized linear probability model converge in asymptotic distribution to the maximum likelihood estimator.

The variance-covariance matrix for $\hat{\beta}$ is

$$(5.4)\quad \left(\sum_{i=1}^{I} (B_i \otimes Z_i'Z_i)\right)^{-1} .$$

First round least squares estimates of $\beta(j, k, \ell)$ can be used to form estimates of B_i.

In order to test linear restrictions on β of the generalized form

$$(5.5)\quad 0 = R\beta,$$

the Wald statistic is

$$(5.6)\quad (R\hat{\beta})'[R(\sum_{i=1}^{I} (B_i \otimes Z_i'Z_i))^{-1}R']^{-1}R\hat{\beta}$$

which is asymptotically distributed as a χ^2 statistic with the number of degrees of freedom given by the rank of the R matrix.[9] An estimated value of B_i can be used in place of the true (unknown) value of B_i.

The hypotheses of Table 2 can be translated into restrictions on the R matrix. For example, the joint restriction that $\beta(011) = \beta(101) = \beta(110)$ and $\beta(100) = \beta(010) = \beta(001)$ requires that R be written as

$$R = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \otimes I_K$$

where K is the number of regressor variables in the Z_i vector. It is assumed that the vectors in β are ordered so as to conform with the R matrix. The Wald statistic associated with this restriction is a χ^2 statistic with 4K degrees of freedom.

Note that in order to estimate the model and test the restrictions it is not required that the predicted values of the dependent variables of equation (5.1) lie inside the unit interval in order to compute the inverse of the appropriate variance-covariance matrix (B_i in equation (5.2)) needed to compute the GLS estimator (equation (5.3)). However, in any small sample, it is required that

$$(5.7) \quad \sum_{i=1}^{I} (B_i \otimes Z_i'Z_i)$$

be positive definite in order to ensure the existence of GLS estimators. Thus in practice it might be necessary to delete certain observations to achieve estimates although in large samples this becomes unnecessary. Any rule that deletes observations to achieve a finite sample GLS estimator does not affect the asymptotic properties of the procedure. Since the argument that justifies the use of an estimated value of B_i in place of the actual value of B_i is only valid in large samples, the choice of a rule to ensure a well defined small sample GLS estimator has no consequences for the optimality of the procedure.

The empirical results obtained from this procedure are reported in Table 4. The regressor variables used to control for observed heterogeneity (apart from an intercept term) are (a) the number of children less than six, (b) the number of children living at home, (c) the unemployment rate for the county in which the woman resides, (d) the wife's education, (e) family income exclusive of wife's earnings, (f) previous labor force

TABLE 4

TESTS AMONG ALTERNATIVE MODELS USING THE GENERALIZED LINEAR PROBABILITY MODEL

	Wald Statistic (intercept only)	Wald Statistic (with regressors)
Random Effects Bernoulli Model	7.6	165
Theoretical χ^2 (5%)	9.49 (4 d.f.)	41.3 (28 d.f.)[a]
General Stationary Latent Variable Model	4.6	90.2
Theoretical χ^2 (5%)	5.99 (2 d.f.)	23.7 (14 d.f.)[b]

[a]Since there are seven regressors (including intercept) in each equation, the constraints implied by the random effects Bernoulli model lead to 28 restrictions across equations.

[b]Since there are seven regressors (including intercept), the constraints implied by the general stationary latent variable model lead to 14 restrictions across equations.

- - -

experience predicted by a set of family background and regional variables to avoid a potential simultaneity bias that might arise from using what is essentially the lagged value of the endogenous variables as a control variable. In order to avoid a high degree of collinearity that would arise if all variables for each year were used, three year averages of the variables were employed in the empirical analysis. Finally, to ensure the existence of a small sample GLS estimator, roughly ten percent of the observations were dropped to ensure a positive definite value for expression (5.7).[10]

The first column of Table 4 reports results based on the generalized linear probability model regressors (but with intercept). Not surprisingly, the results in this column are in accord with the χ^2 statistics reported in Table 3. There is no evidence of state dependence in the data. The most surprising aspect of Table 4 is that <u>when regressors are added</u>, both models are decisively rejected by the data. From tests not reported here, the regressors clearly belong in the model. When they are included, and the hypothesis of no state dependence is tested, it is decisively rejected by the data.

SUMMARY AND DISCUSSION OF RELATED WORK

This paper develops statistical models for discrete panel data. The group of models presented here is based on the notion that discrete events are generated by latent variables crossing thresholds. These latent variable models are used to generate a rather wide class of stochastic processes for the analysis of discrete longitudinal data.

The models considered here are put to use to answer the following substantive question. Does a conditional relationship between the <u>current</u> probability of experiencing an event and previous experience of the event arise because past experience altered individual behavior, or because past experience is a proxy for an individual's propensity to experience the event? Not surprisingly, unconditional, <u>structural</u> probability models can be used to answer the question posed, and such models are presented in this paper. The procedures developed in this paper are the discrete data analogue to ordinary simultaneous equation procedures widely used in the econometric literature to analyze continuous variate structural equation models.

The focus of the analysis in this paper has been on the development of <u>simple</u> estimators designed to answer the question posed. These simple estimators exploit runs patterns in the data to test among competing hypotheses. Two sorts of models are considered : simple runs tests along the lines of those proposed by David (1947) and Goodman (1958) for a more restrictive class of models in which control variables are ignored, and more complicated models that allow for individual variation in measured attributes. Empirical examples of both procedures are presented.

A second focus of attention has been the analysis of the usefulness and limitations of panel data. A variety of discrete microdynamic models were investigated, and data requirements needed to estimate these models have been considered.

A companion paper (Heckman, 1978b) develops more fully the link between the discrete stochastic models discussed here, and more familiar models for stochastic processes that appear in the literature. Using maximum likelihood, it is possible to estimate a general stochastic model that contains many models previously advanced in the literature as special cases of a more general model. This generality permits analysts to perform nested statistical tests to discriminate among competing specifications.

The companion paper also presents simple methods for estimating competing models by maximum likelihood. Factor analytic schemes are developed, and a multivariate probit model with fixed effects is explored. Both approaches greatly simplify the computational burden of the maximum likelihood procedure but, as noted in that paper, they are not without their costs.

The companion paper also presents alternative solutions to the problem that "control variables" may be correlated with the error term. This paper has "solved" the problem by adopting a reduced form approach, i.e., substituting out potentially troublesome correlated variables by using reduced form equations. This approach is always correct for testing the null hypothesis of no state dependence. The companion paper presents a fixed effects model that permits a more structural approach to the formulation and estimation of discrete dynamic models.

FOOTNOTES

1. More precisely, if $V_i(1) = \bar{V}_1$ and $V_i(2) = \bar{V}_2$, $\bar{V}_1 \neq \bar{V}_2$, a situation algebraically equivalent to a case of $\sigma_{\varepsilon_1} \neq \sigma_{\varepsilon_2}$, $\bar{V}_1$ is estimable, but $\bar{V}_2$, ρ and γ are not uniquely estimable without further restrictions. The proof is trivial and hence is deleted. Intuitively, one cannot hope to estimate four parameters from the three cells of independent data that are available.

2. A Hadamard product of two matrices is defined for matrices with identical numbers of rows and columns by the following operation: $A * B = C$ where $c_{ij} = a_{ij}b_{ij}$, all i and j. See C. R. Rao (1973).

3. This condition is sufficient but not necessary.

4. Moreover, one can prove in all cases except the "permanent-transitory" case that there is no other admissible permutation except P in the text. I am indebted to J. A. Scheinkman for this point.

5. In particular, it is required that a person change state at least once.

6. For a discussion of runs tests used to discriminate between Bernoulli and Markov models, see David (1947) or Goodman (1958).

7. In order to test between the homogeneous Bernoulli model and the heterogeneous Bernoulli model in the presence of measured heterogeneity within the linear probability model framework, the following procedure can be used. Fit the ordinary single equation linear probability to a pooled cross section. The dependent variable in such an analysis is a dummy variable with the value of one or zero depending on whether or not the individual experiences the event in a single period. The model is fitted on a pooled sample in which observations for all individuals and all time periods are treated alike. Predicted values of the proportions in each runs pattern can be compared with the predictions based on the probabilities estimated from the generalized linear probability model corresponding to model III.2, including the predicted probabilities for the no turnover cells ((111) and (000)). Comparing the two models in terms of goodness of fit in predicting the eight cells permits a test between the two specifications.

8. For runs tests that distinguish between homogeneous Markov and homogeneous Bernoulli models, see David (1947) or Goodman (1958).

9. Since the GLS estimator converges in distribution to the maximum likelihood estimator, use of the Wald statistic is justified by a theorem of Rao (1973, pp. 418-419). An alternative, and in this case, equivalent justification for test statistic (5.6) is based on the limiting form of the classical F statistic used to test hypotheses in multiple equation systems (see, e.g., Theil, p. 314). In large samples, the classical F statistic and Wald statistic are identical except for a degrees of freedom adjustment.

Finally, use of the Wald statistic can be justified by a direct appeal to the limiting distribution of GLS estimators.

10. The rule employed to ensure positive definiteness was to drop all observations for which the estimated value of B_i is not positive definite.

REFERENCES

Amemiya, Takeshi. "The Estimation of a Simultaneous Equation Generalized Probit Model." Stanford University, May 1977. Forthcoming, Econometrica, 1978.

__________. "Some Theorems on the Linear Probability Model." International Economic Review 18, no. 3 (October 1977).

Ashford, J. R. and R. R. Sowden. "Multivariate Probit Analysis." Biometrics 26 (1970): 535-546.

Balestra, P. and M. Nerlove. "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model." Econometrica 34 (July 1966): 585-612.

Coleman, J. S. Models of Change and Response Uncertainty. Englewood Cliffs, N.J.: Prentice Hall, 1964.

Cripps, T. F. and R. J. Tarling. "An Analysis of the Duration of Male Unemployment in Great Britain 1932-1973." The Economic Journal 84, no. 334 (June 1974): 289-316.

David, F. N. "A Power Function for Tests of Randomness in a Sequence of Alternatives." Biometrika 34 (1947): 335-339.

Denny, J. and S. Yakowitz. "Admissible Run-Contingency Type Tests for Independence and Markov Dependence." Journal of the American Statistical Association 73 (March 1978): 177-181.

Feller, W. An Introduction to Probability Theory and Its Applications, Vol. II. New York: John Wiley & Sons, 1971.

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

Goodman, Leo. "Simplified Runs Tests and Likelihood Ratio Tests for Markov Chains." Biometrika 45 (1958): 181-197.

Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equation Systems." April 1973 (Revised April 1977). Forthcoming in Econometrica, 1978a.

__________. "Statistical Models for Discrete Longitudinal Data." University of Chicago, 1978b.

__________. "Heterogeneity and State Dependence in Dynamic Models of Labor Supply." University of Chicago, 1978c.

Heckman, James J. and R. Willis. "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women." Journal of Political Economy 85, no. 1 (February 1977): 27-58.

__________. "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In Household Production and Consumption, N. Terleckyj (ed.), National Bureau of Economic Research, 1975. Pp. 99-138.

Jovanovic, B. "State Dependence in a Continuous Time Stochastic Model of Worker's Behavior." Discussion Paper 77-7808, Columbia University, January 1978.

Karlin, .Samuel and Howard M. Taylor. A First Course in Stochastic Processes, 2nd. ed. New York: Academic Press, 1975.

Kiefer, N. and G. Neumann. "Estimating Wage Offer Distributions and Reservation Wages." University of Chicago, June 1977.

Kmenta, J. Elements of Econometrics. New York: The Macmillan Co., 1971.

Lee, T. C.; G. Judge and A. Zellner. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. Amsterdam: North-Holland Publishing Co., 1970.

Malinvaud, E. Statistical Methods of Econometrics. Amsterdam: North-Holland Publishing Co., 1970.

McFadden, Daniel. "Quantal Choice Analysis: A Survey." Annals of Economic and Social Measurement 5, no. 4 (December 1976): 363-390.

Morgan, J., et al. Five Thousand American Families--Patterns of Economic Progress, Vols. 1 and 2. Ann Arbor: University of Michigan, Institute for Social Research, 1974.

Rao, C. R. Linear Statistical Inference and Its Applications, 2nd. ed. New York: John Wiley & Sons, 1973.

Theil, H. Principles of Econometrics. New York: John Wiley & Sons, 1971.