Volume Title: Studies in Labor Markets

Volume Author/Editor: Sherwin Rosen, ed.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-72628-2

Volume URL: http://www.nber.org/books/rose81-1

Publication Date: 1981

Chapter Title: Heterogeneity and State Dependence

Chapter Author: James J. Heckman

Chapter URL: http://www.nber.org/chapters/c8909

Chapter pages in book: (p. 91 - 140)

3   Heterogeneity and State Dependence

James J. Heckman

In a variety of contexts, such as in the study of the incidence of accidents (Bates and Neyman 1951), labor force participation (Heckman and Willis 1977) and unemployment (Layton 1978), it is often noted that individuals who have experienced an event in the past are more likely to experience the event in the future than are individuals who have not experienced the event. The conditional probability that an individual will experience the event in the future is a function of past experience. There are two explanations for this empirical regularity.

One explanation is that as a consequence of experiencing an event, preferences, prices, or constraints relevant to future choices (or outcomes) are altered. In this case past experience has a genuine behavioral effect in the sense that an otherwise identical individual who did not experience the event would behave differently in the future than an individual who experienced the event. Structural relationships of this sort give rise to true state dependence as defined in this paper.

A second explanation is that individuals may differ in certain unmeasured variables that influence their probability of experiencing the event

James J. Heckman is Professor of Economics, University of Chicago, and Research Associate, National Bureau of Economic Research.

but that are not influenced by the experience of the event. If these variables are correlated over time, and are not properly controlled, previous experience may appear to be a determinant of future experience solely because it is a proxy for such temporally persistent unobservables. Improper treatment of unmeasured variables gives rise to a conditional relationship between future and past experience that is termed spurious state dependence.

The problem of distinguishing between structural and spurious dependence is of considerable substantive interest. To demonstrate this point, it is instructive to consider recent work in the theory of unemployment. Phelps (1972) has argued that short-term economic policies that alleviate unemployment tend to lower aggregate unemployment rates in the long run by preventing the loss of work-enchancing market experience. His argument rests on the assumption that current unemployment has a real and lasting effect on the probability of future unemployment. Cripps and Tarling (1974) maintain the opposite view in their analysis of the incidence and duration of unemployment. They assume that individuals differ in their propensity to experience unemployment and in their unemployment duration times and that differences cannot be fully accounted for by measured variables. They further assume that the actual experience of having been unemployed or the duration of past unemployment does not affect future incidence or duration. Hence in their model short-term economic policies have no effect on long-term unemployment. The model developed in this paper is sufficiently flexible to accommodate both views of unemployment and can be used to test the two competing theories.

As another example, recent work on the dynamics of female labor supply assumes that entry into and exit from the labor force can be described by a Bernoulli probability model (Heckman and Willis 1977). This view of the dynamics of female labor supply ignores considerable evidence that work experience raises wage rates and hence that such experience may raise the probability that a woman works in the future, even if initial entry into the work force is determined by a random process. The general model outlined in this paper extends the econometric model of Heckman and Willis by permitting (1) unobserved variables that determine labor force choices to be freely correlated, in contrast with the rigid permanent-transitory error scheme for the unobservables assumed in their model; (2) observed explanatory variables to change over time (in their model these variables are assumed to be time invariant); and (3) previous work experience to determine current participation decisions. Empirical work reported below reveals that these three extensions are important in correctly assessing the determinants of female labor supply and in developing models that can be used in policy simulation analysis.

The problem of distinguishing between the two explanations for the empirical regularity has a long history. The first systematic discussion of this problem is presented in the context of the analysis of accident proneness. The seminal work on this topic is by Feller (1943) and Bates and Neyman (1951). Bates and Neyman demonstrate that panel data on individual event histories are required in order to discriminate between the two explanations. Papers preceding the Bates and Neyman work unsuccessfully attempted to use cross-section distributions of accident counts to distinguish between true and spurious state dependence. (See Feller 1943; Heckman and Borjas 1980.)

The problem of distinguishing between spurious and true state dependence is very similar to the familiar econometric problem of estimating a distributed lag model in the presence of serial correlation in the errors (Griliches 1967, Malinvaud 1970, Nerlove 1978). It is also closely related to previous work on the mover-stayer model that appears in the literature on discrete stochastic processes (Goodman 1961, Singer and Spilerman 1976).

This paper presents a new approach to this problem. A dynamic model of discrete choice is developed and applied to analyze the employment decisions of married women. The dynamic model of discrete choice presented here extends previous work on atemporal models of discrete choice by McFadden (1973a) and Domencich and McFadden (1975) to an explicitly dynamic setting. Markov models, renewal models, and "latent Markov" models emerge as a special case of the general model considered here. The framework presented here extends previous work on the mover-stayer problem (Singer and Spilerman 1976) by broadening the definition of heterogeneity beyond the "mixing distribution" or "components of variance" models employed in virtually all current work on the analysis of discrete dynamic data.

The major empirical finding reported in this paper is that past employment experience is an important determinant of current employment decisions for women past the childbearing years, even after accounting for heterogeneity of a very general type. This relationship can be interpreted as arising *in part* from the impact of both general and specific human capital investment on current labor market choices, but it is consistent with other explanations as well. Empirical evidence on the importance of heterogeneity is presented. Estimates of structural state dependence based on procedures that improperly control heterogeneity dramatically overstate the impact of past employment on current choices.

The estimates for younger married women, most of whom are in their childbearing years, suggest a weak effect of past participation on current choices, but empirical evidence on the importance of heterogeneity is still strong.

This paper also presents evidence that the unobserved variables that

determine employment follow a stationary first-order Markov process. Initial differences in unmeasured variables tend to be eliminated with the passage of time. This homogenizing effect is offset in part by the impact of prior work experience that tends to accentuate initial differences in the propensity to work.

The empirical evidence on heterogeneity reported in this paper calls into question the implicit assumption maintained in previous work that addresses the problem of heterogeneity (Singer and Spilerman 1976; Heckman and Willis 1977). That work assumes that unmeasured variables follow a components of variance scheme; an individual has a "permanent" component to which a serially uncorrelated "transitory" component is added. The work reported here suggests that the heterogeneity process for married women cannot be modeled so simply. Unmeasured components are better described by the first-order Markov process. Omitted variables determining choices are increasingly less correlated as the time span between choices widens. Misspecification of the heterogeneity process gives rise to an erroneous estimate of the impact of the true effect of past employment on current employment probabilities.

This paper is in four parts. Part 3.1 provides an intuitive motivation to the problems and models considered in this paper. Part 3.2 presents the model used here, and discusses econometric issues that arise in implementing it. Part 3.3 presents estimates of the model and some qualifications. Part 3.4 presents an interpretation of the estimates. The paper concludes with a brief summary. An appendix presents a decomposition of estimated structural state dependence effects into wage and nonwage components.

3.1. Heterogeneity and State Dependence: An Intuitive Introduction

In order to motivate the analysis in this paper, it is helpful to consider four simple urn models that provide a useful framework within which to introduce intuitive notions about heterogeneity and state dependence. In the first scheme there are I individuals who possess urns with the same content of red and black balls. On T independent trials individual i draws a ball and then puts it back in his urn. If a red ball is drawn at trial t, person i experiences the event. If a black ball is drawn, person i does not experience the event. This model corresponds to a simple Bernoulli model and captures the essential idea underlying the choice process in McFadden's (1973a) work on discrete choice. From data generated by this urn scheme, one would not observe the empirical regularity described in the introduction. Irrespective of their event histories, all people have the same probability of experiencing the event.

A second urn scheme generates data that would give rise to the empirical regularity solely due to heterogeneity. In this model individuals

possess distinct urns which differ in their composition of red and black balls. As in the first model sampling is done with replacement. However, unlike the first model, information concerning an individual's past experience of the event provides information useful in locating the position of the individual in the population distribution of urn compositions.

The person's past record can be used to estimate the person-specific urn composition. The conditional probability that individual i experiences the event at time t is a function of his past experience of the event. The contents of each urn are unaffected by actual outcomes and in fact are constant. There is no true state dependence.

The third urn scheme generates data characterized by true state dependence. In this model individuals start out with identical urns. On each trial, the contents of the urn change *as a consequence of the outcome of the trial*. For example, if a person draws a red ball, and experiences the event, additional new red balls are added to his urn. If he draws a black ball, no new black balls are added to his urn. Subsequent outcomes are affected by previous outcomes because the choice set for subsequent trials is altered as a consequence of experiencing the event. This model is a generalized Polya urn scheme.[1]

A variant of the third urn scheme can be constructed that corresponds to a renewal model (Karlin and Taylor 1975). In this scheme new red balls are added to an individual's urn on successive drawings of red balls until a black ball is drawn, and then all of the red balls added to the most recent continuous run of drawings of red balls are removed from the urn. The composition of the urn is then the same as it was before the first red ball in the run was drawn. A model corresponding to fixed costs of labor force entry is a variant of the renewal scheme in which new red balls are added to an individual's urn only on the first draw of the red ball in any run of red draws.

The crucial feature that distinguishes the third scheme from the second is that the contents of the urn (the choice set) are altered as a consequence of previous experience. The key point is not that the choice set changes across trials but that it changes in a way that depends on previous outcomes of the choice process. To clarify this point, it is useful to consider a fourth urn scheme that corresponds to models with more general types of heterogeneity to be introduced more formally below.

In this model individuals start out with identical urns, exactly as in the first urn scheme. After each trial, but independent of the outcome of the trial, the contents of each person's urn are changed by discarding a randomly selected portion of balls and replacing the discarded balls with a randomly selected group of balls from a larger urn (say, with a very large number of balls of both colors). Assuming that the individual urns are not completely replenished on each trial, information about the outcomes of previous trials is useful in forecasting the outcomes of future trials, although the information from a previous trial declines with its

remoteness in time. As in the second and third urn models, previous outcomes give information about the contents of each urn. Unlike the second model, the fourth model is a scheme in which the information depreciates since the contents of the urn are changed in a random fashion. Unlike in the third model, the contents of the urn do not change as a consequence of any outcome of the choice process. This is the urn model analogue of Coleman's (1964) latent Markov model.

The general model presented below is sufficiently flexible that it can be specialized to generate data on the time series of individual choices that are consistent with samples drawn from each of the four urn schemes just described as well as more general schemes including combinations of the four. The principle advantage of the proposed model over previous models is that it accommodates very general sorts of heterogeneity and structural state dependence as special cases and permits the introduction of explanatory exogenous variables. The generality of the framework proposed here permits the analyst to combine models and test among competing specifications within a unified framework.

In the literature on female labor force participation, models of extreme homogeneity (corresponding to urn model one) and extreme heterogeneity (corresponding to urn model two with urns either all red or all black) are presented in a paper by Ben Porath (1973) which is a comment on Mincer's model (1962) of female labor supply. Ben Porath notes that cross-section data on female participation are consistent with either extreme model. Heckman and Willis (1977) pursue this point somewhat further and estimate a model of heterogeneity in female labor force participation probabilities that is the probit analogue of urn model two. They assume no state dependence. There is no previous work on female labor supply that estimates models corresponding to urn schemes three and four.

Urn model three is of special interest. It is consistent with human capital theory, and other models that stress the impact of prior work experience on current work choices. Human capital investment acquired through on the job training may generate structural state dependence. Fixed costs incurred by labor force entrants may also generate structural state dependence as a renewal process. So may spell-specific human capital. This urn model is also consistent with psychological choice models in which, as a consequence of receiving a stimulus of work, women's preferences are altered so that labor force activity is reinforced (Atkinson, Bower, and Crothers 1965).

Panel data can be used to discriminate among these models. For example, an implication of the second urn model is that the probability that a woman participates does not change with her labor force experience. An implication of the third model in the general case is that participation probabilities change with work experience. One method for

discriminating between these two models utilizes individual labor force histories of sufficient length to estimate the probability of participation in different subintervals of the life cycle. If the estimated probabilities for a given woman do not differ at different stages of the life cycle, there is no evidence of structural state dependence.[2]

A more general test among the first three urn models utilizes labor force history data of sufficient length for *each woman* in a sample to estimate a regression of current participation status on previous participation status (measured by dummy variables indicating whether or not a woman worked at previous stages in her life cycle). If previous labor force experience has no effect on the current probability of participation, the first and third urn models would describe the data. If past experience predicts current participation status, but not perfectly, the third model describes the data.

Considerable care must be taken in utilizing panel data to discriminate among the models. The second test must be performed on data drawn from the work history of one person. One could utilize data on the histories of a sample of people by permitting each person to have his own fixed effect or intercept in the regression just described. If one were to pool data on individuals to estimate the regression on the entire sample, and not allow each person to have his own intercept, one would risk the danger that individual differences in participation probabilities, which would be relegated to a disturbance term in a pooled regression across people that does not permit individual intercepts, will be correlated with past participation status. If some individuals have a higher probability of participation than others, and if these differences are relegated to the disturbance term of the regression of current participation status on past participation status, regression analysis would produce a spurious positive relationship between current and previous experience that would appear to demonstrate the presence of structural state dependence that did not, in fact, exist, since people with higher participation probabilities are more likely to be in the labor force in the current period as well as in the past.

This point can be stated somewhat more precisely. Let $d(i, t)$ be a dummy variable that assumes the value of one if woman i works in period t and is zero otherwise. Define $\varepsilon(i, t)$ as a disturbance with the following structure:

$$\varepsilon(i, t) = \phi(i) + U(i, t) , t = 1, \ldots , T$$

$$i = 1, \ldots , I$$

where $\phi(i)$ is an individual-specific effect and $U(i, t)$ is a mean zero random variable of innovations uncorrelated with other innovations $U(i, t')$, $t \neq t'$.[3] There are I individuals in the sample followed for T time periods.

For each individual, write the regression

(1) $$d(i, t) = \phi(i) + \delta \sum_{t' < t} d(i, t') + U(i, t) , t+1, \ldots , T$$

where $d(i, 0)$ is a fixed nonstochastic initial condition.[4] Note that $\phi(i)$ is the intercept in the regression. More general models allowing for depreciation in the effect of past participation on current participation could be written out, but for present purposes nothing is gained by increasing the level of generality of the model. If this regression were fit on data for a single individual, a statistically significant value for δ would indicate that the third urn scheme is more appropriate than the second, i.e., that there is evidence for true state dependence at the individual level. If δ were estimated to be zero, the second urn model would fit the data better.[5]

If regression 1 is computed across people and time, and no allowance is made for individual differences in intercepts, the regression model for the pooled sample could be written as

(2) $$d (i, t) = \bar{\phi}(i) + \delta \sum_{t' < t} d(i, t') + U(i, t) + \phi(i) - \bar{\phi}(i)$$

$$i = 1, \ldots , I$$

$$t = 1, \ldots , T$$

where $\bar{\phi}(i)$ is the average intercept in the population. The composite disturbance in the regression is $U(i, t) + \phi(i) - \bar{\phi}(i)$. Because of equation 1, the term $\sum_{t' < t} d(i, t')$ would be correlated with the composite disturbance. Regression estimates of δ would be upward biased because past work experience is positively correlated with the composite disturbance. This bias could be avoided by permitting each individual to have his own intercept.[6]

Note further that if there is some variable, such as the number of children, that belongs in equation 1, the effect of children estimated from equation 2 will be biased. If children depress participation, and $\delta > 0$, the estimated effect of children on the probability of participation will be upward biased. This follows from a standard simultaneous equation bias argument if current numbers of children are negatively correlated with previous participation, and cumulated previous experience is positively correlated with the error term. Thus, uncorrected heterogeneity not only leads to an overstatement of the state dependence effect but also leads to an understatement of the negative effect of children on participation.[7]

The empirical analysis in this paper could be based on more general versions of equations 1 and 2.[8] However, estimation in the generalized linear probability model gives rise to well-known econometric difficulties; the errors are heteroscedestic, and estimated values of probabilities may not lie inside the unit interval.[9] Moreover, the interpretation of the statistical model as an economic model is unclear.

Instead, the model used here is a dynamic extension of cross-section models of discrete choice developed by the author in other work. (Heckman 1981a b). The essential features of this model are described in the next section.

3.2   A Dynamic Model of Labor Supply

This section presents a dynamic model of discrete choice that can be used to analyze unemployment, labor force participation, and other dynamic events. For specificity we focus on a dynamic model of female employment. The model presented here is based on Heckman (1978b, 1981a b).

Women are assumed to make employment decisions in successive equispaced intervals of time. Each woman has two options in each period in her life cycle: to work or not to work.[10] Let $v(1, i, t)$ be the expected lifetime utility that arises if woman i works in period t. This utility is a function of all relevant decision variables including her expectations about demographic events, such as the birth of children and divorce, and state variables such as her stocks of human capital. "$v(1, i, t)$" is the highest level of lifetime utility that the woman can attain *given* that she works today. "$v(0, i, t)$" is the highest level of lifetime utility that the woman can attain *given* that she does not work today. Implicit in both value functions is the notion that subsequent employment decisions are optimally chosen *given* the current choice, and given any new information, unknown to the agent at t, that becomes known in future periods when future employment decisions are being made.

Employment occurs at age t for woman i if $v(1, i, t) > v(0, i, t)$, i.e., if the expected lifetime utility of employment at age t exceeds the expected lifetime utility that arises from nonemployment. This view of employment is consistent with a wide variety of economic models. In particular, as is demonstrated below, under special assumptions it is consistent with McFadden's (1973b) random utility model applied in an intertemporal context or models of lifetime decision making under perfect certainty developed by Ryder, Stafford, and Stephan (1976) and others. The model is also consistent with fixed costs of entry into and exit from the work force.

For the present analysis, the difference in utilities $V(i, t) = v(1, i, t) - v(0, i, t)$ is the relevant quantity. If $V(i, t)$ is positive, a woman works at time t; otherwise she does not.

The difference in utilities $V(i, t)$ may be decomposed into two components. One component $\bar{V}(i, t)$ is a function of variables that can be observed by the economist, while the other component $\varepsilon(i, t)$ is a function of variables that cannot be observed by the economist. The difference in utilities may thus be written as

$$V(i, t) = \bar{V}(i, t) + \varepsilon(i, t)$$

We record whether or not woman i works at time t by introducing a dummy variable $d(i, t)$ that assumes the value of one when a woman works and is zero otherwise. Thus, $d(i, t) = 1$ if $V(i, t) > 0$, while $d(i, t) = 0$ if $V(i, t) \leq 0$.

To make the model empirically tractable we assume that the difference in utilities $V(i, t)$ can be approximated by

$$(3) \qquad V(i, t) = Z(i, t)\beta + \sum_{t' < t} \delta(t, t') \, d(i, t')$$

$$+ \sum_{j} \lambda(t, t-j) \prod_{\ell = 1}^{j} d\,(i, t-\ell) + \varepsilon(i, t)$$

where $i = 1, \ldots, I$ and $t = 1, \ldots, T$. $E(\varepsilon(i, t)) = 0$, $E\,(\varepsilon(i, t)\varepsilon(i, t'))$

$$= \sigma_{t, \, t'}. \; E(\varepsilon(i, t)\varepsilon(i, t')) = 0, \, i \neq i'.$$

For the moment, we assume that the initial conditions of the process $d(i, 0), \ldots, d(i, -k), \ldots$, are fixed, nonstochastic constants.

"$Z(i, t)$" is a vector of exogenous variables that determine choices in period t. β is a suitably dimensioned vector of coefficients. Included among the components of $Z(i, t)$ are variables such as education, income of the husband, number of children, and the like, as well as expectations about future values of these variables.

The effects of prior work experience on choice in period t are captured by the second and third terms on the right hand side of equation 3. The second term indicates the effect of all prior work experience on choice in period t. The third term indicates the effect on choice in period t of work experience in the most recent continuous spell of work for those who have worked in period $t - 1$. The coefficients associated with these terms are written to allow for depreciation of the effects of previous work experience and to capture the idea that the effect of previous work experience depends on conditions prevailing in the period in which experience occurs as well as on conditions in period t.

Alternative specifications of δ and λ generate different models. For example, setting $\delta(t, t') = \delta$ and $\lambda(t, t-j) = 0$ generates a stochastic process for which the entire work history is relevant for determining choices in period t. Such a model is consistent with (but not necessarily limited to) models of general human capital. Setting $\delta(t, t') = \delta(t - t')$ for $t - t' \leq K$, $\delta(t, t') = 0$ otherwise, generates a Kth-order Markov process. A first-order Markov process is consistent with a model of fixed costs of labor force entry. Once an individual is working she need not pay further fixed costs to continue working. Setting $\delta(t, t') = 0$ for all t', and letting $\lambda(t, t-j)$ be free, generates a renewal process which describes spell-specific human capital accumulation. (See Jovanovic 1978.) For a more complete discussion of alternative specifications of this model see Heckman (1981a).

Heterogeneity arises in this model from $\varepsilon(i, t)$, an unmeasured disturbance due to essential uncertainty (as perceived by the consumer) as well as to factors unknown to the observing economist but known to the consumer. The assumption that disturbances across individuals are uncorrelated is an implication of the random sampling scheme used to generate the data analyzed below.[11]

It is plausible that $\sigma_{tt'} \neq 0$ for $t \neq t'$, i.e., that unmeasured variables like ability are correlated over time for a consumer. Even if the only source of randomness in the model arose from variables that operate on the consumer at a point in time, and are themselves uncorrelated over time, the disturbances are serially correlated. This is so because the difference in utilities in periods t and t' depend on some of the same set of unmeasured expected future variables that determine remaining *lifetime* utility. The empirical work presented below suggests that the unobservables obey a first-order stationary autoregression (i.e., first-order Markov process).

The model of equation 3 can be used to characterize all of the urn models previously considered. The first urn scheme, in which all women face identical urns, and successive drawings are independent, is given by a specialization of equation 3 in which $Z(i, t) = 1$, $\sigma = \lambda = 0$, and $\varepsilon(i, t)$ is distributed independently of all other disturbances. Under these assumptions the probability that $V(i, t)$ is positive is the same for all women at all times, and is independent of any past events.

McFadden's (1976) random utility model corresponds to a special case of equation 3 in which $Z(i, t)$ is not restricted, $\delta = \lambda = 0$, and $\varepsilon(i, t)$ is a mean zero random variable which is distributed independently of other disturbances.

An urn scheme in which a woman's work status is perfectly correlated over time is a special case of equation 3 in which $Z(i, t) = 1$, $\delta = \lambda = 0$, and $\varepsilon(i, t)$ is perfectly correlated over time.[12]

The second urn scheme in which each woman in a population makes independent drawings from her own (distinctive) urn is a special case of equation 3 in which $Z(i, t) = Z(i)$ (regressors are constant over time for a given person but may vary among people), $\delta = \lambda = 0$, and $\varepsilon(i, t)$ has a components of variance structure, i.e., $\varepsilon(i, t) = \phi(i) + U(i, t)$, where $\phi(i)$ and $U(i, t)$ are realizations of mean zero random variables, $\phi(i)$ is a person effect that does not change over the life cycle, and $U(i, t)$ is an independently identically distributed random variable with zero mean. In this model, the term $Z(i)\beta + \phi(i)$ corresponds to the idiosyncratic person-specific loading of balls in the second urn scheme. For each woman, successive draws are independent, but women differ in the composition of red and black balls in their urns. In essential detail this model is that of Heckman and Willis (1977).

The third urn scheme, in which all women start life alike but receive a red ball each time they work, corresponds to a special case of equation 3

in which $Z(i, t) = 1$, $\delta(t, t') = \delta > 0$, $\lambda = 0$, and $\varepsilon(i, t)$ is an independently identically distributed random variable with zero mean.[13] Setting $\delta(t, t') = 0$, but letting λ be nonzero, generates a renewal process version of the third urn scheme.

The fourth urn scheme corresponds to a special case of equation 3 in which $Z(i, t) = 1$, $\delta = \lambda = 0$, and $\varepsilon(i, t)$ is a mean zero random variable following a first-order Markov process, i.e., $\varepsilon(i, t) = \rho\varepsilon(i, t-1) + U(i, t)$, where $U(i, t)$ is independently identically distributed with mean zero.

The general model that is estimated below contains all of these schemes as a special case of a more general model in which the exogenous variables $Z(i, t)$ are permitted to change over time, δ and λ are permitted to be nonzero, and $\varepsilon(i, t)$ is permitted to have a very general serial correlation pattern.

3.2.1   The Econometric Specification

The model of equation 3 is estimated by the method of maximum likelihood. The disturbance terms are assumed to be jointly normally distributed so that the statistical model is a "multivariate probit model with structural shift." A formal analysis of this model is presented elsewhere (Heckman, 1978a, 1981a b).[14] In estimating the model, special care is taken to avoid bias that arises from the correlation of $\varepsilon(i, t)$ with previous work experience $d(i, t')$, $t > t'$. Such bias would arise in estimating the coefficients of the model if values of $\varepsilon(i, t)$ are serially correlated, which is the plausible case. In the presence of serial correlation, the work experience variables are correlated with the disturbance term for period t since prior work experience is determined by prior values of the disturbances, and prior disturbances are correlated with the disturbance in period t.

The statistical model used here avoids large-sample bias in estimating the structural coefficients by correcting the distribution of $\varepsilon(i, t)$ for the effect of previous work experience using the model of equation 3 to form the correct conditional distribution. The distribution of $\varepsilon(i, t)$ conditional on previous work experience may be written as

(4) $$g[\varepsilon(i, t) \mid d(i, t-1), d(i, t-2), \ldots]$$

For details on constructing this distribution see Heckman (1978a, 1981a) or appendix B. The probability that $d(i, t)$ is unity (i.e., that woman i works in period t) is the probability that $V(i, t)$ is positive. This probability is computed with respect to the appropriate *conditional* distribution of $\varepsilon(i, t)$. Defining $P(i, t)$ as the probability of participation in period t by woman i conditional on previous work experience,

$$P(i, t) = \Pr[V(i, t) > 0 \mid d(i, t-1), d(i, t-2), \ldots]$$

$$= \Pr[\varepsilon(i, t) > -Z(i, t)\beta - \sum_{t' < t} \delta(t, t') d(i, t') - \sum_j \lambda(t, t-j) \prod_{\ell=1}^{j} d(i, t-\ell)]$$

$$= \int_{K}^{\infty} g(\varepsilon(i, t) \mid d(i, t-1), \ldots) d\varepsilon(i, t)$$

where
$$K = [-Z(i, t)\beta - \sum_{t' < t} \delta(t, t') d(i, t')$$
$$- \sum_{j} \lambda(t, t-j) \prod_{\ell = 1}^{j} d(i, t-\ell)].$$

Conditioning the distribution of $\varepsilon(i, t)$ on previous experience using the model of equation 3 to construct the correct conditional distribution avoids large-sample bias in the estimated coefficients.

The same likelihood function for random variables $d(i, t)$, where $t = 1, \ldots, T$, and $i = 1, \ldots, I$, is

$$\pounds = \prod_{i=1}^{I} \prod_{t=1}^{T} [P(i, t)]^{d(i, t)} [1 - P(i, t)]^{1 - d(i, t)}$$

This function is maximized with respect to the parameters of the model. The properties of the maximum likelihood estimators are discussed in Heckman (1978a). Under standard conditions, they possess desirable large-sample properties.

The information that woman i works in period t reveals that $V(i, t) > 0$. The inequality is not reversed if both sides are divided by the standard deviation of the unobservables $\sigma_{tt}^{1/2}$. This implies that from sample information about a sequence of work patterns it is possible to estimate the coefficients β, $\delta(t, t')$, and $\lambda(t, t-j)$ in equation 3 only up to a factor of proportionality. However, if there are regressors in equation 3, and β is invariant across periods, it is possible to estimate the ratios $\sigma_{tt}/\sigma_{t't'}$ among variances (Heckman, 1981a). Normalizing σ_{11} to unity, it is possible to estimate $\sigma_{22}, \ldots, \sigma_{TT}$.

If the latent variables $\varepsilon(i, t)$ are covariance stationary (Koopmans 1974), $\sigma_{tt'} = \sigma_{t+k}\sigma_{t'+k}$ for all t, t', and k. Since it is possible to estimate $\sigma_{tt}/\sigma_{t't'}$, it is possible to test for stationarity in the disturbances of equation 3. This test is performed below.

To facilitate computations, it is assumed that the disturbances in equation 3 can be one-factor analyzed. This means that it is possible to represent the correlation matrix in the following fashion. Define the correlation coefficient between disturbance $\varepsilon(i, t)$ and disturbance $\varepsilon(i, t')$ as $r_{tt'}$. If the disturbances can be one-factor analyzed

$$r_{tt'} = \alpha_t \alpha_{t'} \text{ for } t \neq t', t' \text{ and } t = 1, \ldots, T$$

Since the number T of panel observations per person is three in the empirical analysis of this paper, this restriction is not serious.[15]

Because of computational considerations the number of panel observations per person is small. Thus it is impossible to estimate all of the models of structural state dependence that could be generated by equation 3. Instead, in the empirical work reported below attention is con-

fined to a model with $\delta(t, t') = \delta$ and $\lambda(t, t-j) = 0$. This specification assumes that prior work experience has the same impact on labor force decisions in period t independent of the time period in which it occurred. In fact, this rigid specification is relaxed to a certain degree in the empirical work.

Two types of prior work experience are considered: presample experience and within-sample experience. It is likely that presample experience exerts a weaker measured effect on current participation decisions than more recent experience because of depreciation and also because the data on presample experience, which are based on a retrospective question, are likely to be measured with error.

Moreover, the data source utilized in the empirical analysis is not sufficiently rich to correctly adjust conditional distribution 4 using the model of equation 3. As demonstrated in appendix B and Heckman (1981b), appropriate conditioning requires, in general, the entire life cycle history of individuals including presample values of exogenous variables. Elsewhere (Heckman 1981b) exact and approximate solutions to this problem of correctly initializing the process are proposed. One estimator, which is shown to work well especially for testing the null hypothesis of no structural state dependence, predicts presample experience by a set of regressors and utilizes the predicted value as another element of $Z(i, t)$. This estimator is utilized to generate the empirical estimates reported in this paper. Within-sample work experience is treated in the manner described in the preceding paragraphs. Thus, conditional distribution 4 is constructed using actual within-sample realizations of prior work experience, and predicted values of presample work experience.

3.3   Empirical Results

This section presents evidence from an empirical analysis of the dynamics of married female labor supply. Empirical results are presented for two groups of white women: women of age 30–44 in 1968 and women of age 45–59 in 1968. Both groups of women were continuously married to the same spouse in seven years of panel data drawn from the probability sample of the Michigan Panel Survey of Income Dynamics.[16] For the sake of brevity, we focus on the results for older women and the contrasts in the empirical findings between the two age groups. Our discussion focuses on the central empirical issue of distinguishing heterogeneity and structural state dependence.

The major finding reported here is that for older women there is some evidence of structural state dependence in individual probabilities. For younger women, there is much less evidence of structural state dependence. The results reported here question the validity of the simple

"permanent-transitory" or "convolution" scheme commonly used to characterize heterogeneity in much applied work in social science. A first-order Markov process for the disturbances describes the data better. Tests for nonstationarity in the unobservables reject that hypothesis.

A mostly conventional set of variables is used to explain employment. These are (1) the woman's education; (2) family income excluding the wife's earnings; (3) number of children younger than six; (4) number of children at home; (5) presample work experience; (6) within-sample work experience; (7) unemployment rate in the county in which the woman resides; (8) the wage of unskilled labor in the county—a measure of the availability of substitutes for the woman's time in the home; and (9) the national unemployment rate for prime-age males—a measure of aggregate labor market tightness. Mean values for each of these variables in both samples are presented in table 3.A.2 in appendix C. A woman is defined to be a market participant if she worked for money any time in the sample year. This definition departs from the standard census definition in two respects. First, participation is defined as work, and excludes unemployment. The second way in which the definition used here departs from the standard one is that the time unit of definition of the event is the year and not the usual census week. For both reasons, our results are not directly comparable with previous cross-section empirical work by Cain (1966) and Mincer (1962). Our definitions are comparable with those of Heckman and Willis (1977).

A noteworthy feature of the data is that roughly 80 percent of the women in the sample of older women either work all of the time or do not work at all. (See table 3.1A) The corresponding figure for younger women is 75 percent. (See table 3.1B) Both samples are roughly evenly divided between full-time workers and full-time nonworkers. There is little evidence of frequent turnover in these data, nor is there much evidence of turnover in the full seven years of data.[17]

We first present results for the older group of women. Then results for the younger women are briefly discussed.

Coefficient estimates of equation 3 for women aged 45–59 for the most general model estimated in this paper are presented in column 1 of table 3.2. A positive value for a coefficient means that an increase in the associated variable increases the probability that a woman works, while a negative value for a coefficient means that an increase in the associated variable decreases the probability. Inspection of the coefficients arrayed in column 1 reveals that more children and a higher family income (excluding wife's earnings) depress the probability of female employment.

Higher rates of unemployment (both local and national) tend to depress the probability of female participation.[18] This finding suggests that the net impact of labor market unemployment is to discourage female

Table 3.1　　　**Runs Patterns in the Data (1 corresponds to work in the year, 0 corresponds to no work)**

Runs Pattern (1968, 1969, 1970)			No. of Observations	Runs Pattern (1971, 1972, 1973)			No. of Observations
A. Women Aged 45–59 in 1968							
0	0	0	87	0	0	0	96
0	0	1	5	0	0	1	5
0	1	0	5	0	1	0	4
1	0	0	4	1	0	0	8
1	1	0	8	1	1	0	5
0	1	1	10	0	1	1	2
1	0	1	1	1	0	1	2
1	1	1	78	1	1	1	76
B. Women Aged 30–44 in 1968							
0	0	0	126	0	0	0	133
0	0	1	16	0	0	1	13
0	1	0	4	0	1	0	5
1	0	0	12	1	0	0	16
1	1	0	24	1	1	0	8
0	1	1	20	0	1	1	19
1	0	1	5	1	0	1	8
1	1	1	125	1	1	1	130

employment. The estimated effect of the wage of unskilled labor in the county on participation is statistically insignificant.

The estimated values of the ratios of the second and third-period variances in the disturbances to the first-period disturbance variance (i.e., σ_{22} and σ_{33} respectively) are close to one. Utilizing conventional test criteria, one cannot reject the hypothesis that both of these estimated coefficients equal one. Thus the variance in the unobservables is the same in each period. When the model is recomputed constraining σ_{22} and σ_{33} to unity (see the results reported in column 2), the decrease in log likelihood for the model is trivial (.82), and well below the variation that would arise solely from chance fluctuations. The remaining coefficients in the model are unaffected by the imposition of this restriction, lending further support to the assumption of constant variances.

Coefficients α_1, α_2 and α_3 are normalized factor-loading coefficients which, when multiplied, yield estimates of the correlation coefficients among the unobservable variables $\varepsilon(i, t)$. Utilizing the estimates reported in column 1 of table 3.2, the estimated correlation between disturbances in year 1 and year 2 is $(.922) \times (.992) = .915$ while the estimated correlation between disturbances in year 2 and year 3 is $(.992) \times (.926) = .918$. The estimated two-year correlation is $(.922) \times (.926) = .854$. Note that the product of the estimated one-year correlation coefficients (.840) is very close to the estimated two-year correlation

Table 3.2 Estimates of the Model for Women Aged 45–59 in 1968 (asymptotic normal test statistics in parentheses; these statistics are obtained from the estimated information matrix)

Variable	(1)	(2)	(3)	(4)	(5)
Intercept	−2.498(4.1)	−2.576(4.6)	−1.653(2.5)	−2.325(5.1)	−2.367(6.4)
No. of Children aged less than 6	−.803(2.8)	−.816(2.7)	−.840(2.3)	−.741(2.6)	−.742(2.6)
County unemployment rate (%)	−.039(1.4)	−.035(1.5)	−.027(1.0)	−.030(1.4)	−.030(1.5)
County wage rate ($/hr.)	.106(.96)	.104(.91)	.104(.91)	.100(.93)	.099(.93)
Total no. of children	−.141(3.2)	−.146(4.3)	−.117(2.2)	−.127(3.5)	−.124(4.9)
Wife's education (yrs.)	.157(5.0)	.162(6.5)	.105(2.8)	.145(5.3)	.152(7.3)
Family income excluding wife's earnings	$−.36 \times 10^{-4}(4.1)$	$−.363 \times 10^{-4}(4.8)$	$−.267 \times 10^{-4}(2.7)$	$−.32 \times 10^{-4}(4.3)$	$−.312 \times 10^{-4}(5.2)$
National unemployment rate	−.098(.5)	−.106(.51)	−.254(1.4)	−.035(.34)	−.003(.38)
Current experience (δ)	.136(.97)	.143(.95)	.273(1.5)	—	—
Predicated presample experience	.069(4.0)	.072(5.8)	.059(3.4)	.062(4.3)	.062(6.2)
α_1	.922(33)	.921(3.)	—	.922(35)	.920(35)
α_2	.922(124)	.991(116)	—	.996(164)	.997(196)
α_3	.926(19)	.919(14)	—	.948(41)	.949(42)
σ_{22}	.935(4.8)	1	1	.895(4.4)	1
σ_{33}	1.114(4.5)	1	1	1.079(4.7)	1
ρ	—	—	.873(14.0)	—	—
η	—	—	—	—	—
ln likelihood	−236.92	−237.74	−240.32	−238.401	−239.81

Table 3.2 (continued)

Variable	(6)	(7)	(8)	(9)	(10)
Intercept	−2.011(3.4)	−1.5(0)	−2.37(5.5)	.227(.4)	−3.53(4.6)
No. of children aged less than 6	−.793(2.1)	−.69(1.2)	−.70(2.0)	−.814(2.1)	−1.42(2.3)
County unemployment rate (%)	−.027(1.2)	.046(11)	−.03(1.6)	−.018(.57)	−.059(1.3)
County wage rate ($/hr.)	.139(1.5)	.105(.68)	.13(1.4)	.004(.02)	.27(1.1)
Total no. of children	−.116(2.2)	−.160(6.1)	−.161(4.9)	−.090(2.4)	−.203(3.9)
Wife's education (yrs.)	.095(2.5)	.105(3.3)	.077(3)	.104(3.7)	.196(4.8)
Family income excluding wife's earnings	$−.207 \times 10^{-4}$(2.3)	$−.385 \times 10^{-4}$(20)	$−.2 \times 10^{-4}$(2.6)	$−.32 \times 10^{-4}$(3.6)	$−.65 \times 10^{-4}$(5.1)
National unemployment rate	−.021(.26)	−.71(0)	.02(.3)	−1.30(6)	1.03(.14)
Current experience (δ)	—	—	—	1.46(12.2)	—
Predicted presample experience	.062(3.5)	.095(11.0)	.091(7.0)	.045(3.4)	.101(5.4)
α_1	—	—	—	—	—
α_2	—	—	—	—	—
α_3	—	—	—	—	—
σ_{22}	1	1	1	1	1
σ_{33}	1	1	1	1	1
ρ	.942(50)	—	—	—	—
η	—	.941(4.1)	.92(4.5)	—	—
ln likelihood	−243.11	−242.37	−244.7	−263.65	−367.3

(*See following page for source notes*)

coefficient, a result that strongly suggests that the disturbances obey a first-order stationary Markov process, i.e., that

$$\varepsilon(i, t) = \rho\varepsilon(i, t-1) + U(i, t) \ i = 1, \ldots, I$$

$$t = t, \ldots, T$$

where $U(i, t)$ is independently identically distributed across people and time.

Column 3 of table 3.2 reports estimates of a model that constrains the disturbances to follow a first-order Markov scheme. The Markov model is a special case of the general model in which $\alpha_1 = \alpha_3$, and $\alpha_2 = 1$. The empirical results appear to support the hypothesis of a Markov error process. Comparing the value of the likelihood function presented in column 2 with the value presented in column 3, one cannot reject the null hypothesis that the Markov model describes the distribution of disturbances. Twice the difference in log likelihood (5.25) is to be compared with a value of the χ^2 statistic with two degrees of freedom, 5.99, for a five percent significance level.[19] Most of the estimated coefficients presented in column 3 are essentially the same as the corresponding coefficients presented in the two preceding columns of the table so the Markov restriction appears to be innocuous. However, the coefficient on presample work experience drops slightly while the coefficient on recent experience almost doubles, and almost becomes statistically significant using conventional test statistics.

For reasons already presented, the measured effect of previous work experience on current employment is broken into two components: (a) the effect of work experience acquired prior to the first year of the sample (1968), and (b) the effect of more recent experience measured in the sample. The coefficient of recent experience is roughly twice the size of the coefficient on presample experience. Both coefficients are positive, as expected, but only the coefficient on predicted presample experience is statistically significant using conventional asymptotic "t" test statistics.

Table 3.2 Source Notes
Source: First three years of the Michigan Panel Survey of Income Dynamics, 1968, 1969, 1970.
Note: Increase in variables with positive coefficients increase the probability of employment, while increases in variables with negative coefficients decrease the probability of employment.
Presample experience is predicted from all of the other regressors in $Z(i,t)$, a set of dummy variables for education, city size variables, regional variables and family background variables (mother's and father's education). Note that utilizing predicted experience in place of actual experience in the probit model assumes that the errors from prediction are approximately normally distributed. Even if this is not a correct assumption, this procedure permits a valid statistical test of the important null hypothesis that past labor force experience has no effect on current experience. The standard errors of the estimated coefficients must be interpreted as *conditional* on the predicted values of presample experience variables.

Testing the null hypothesis that work experience does not affect the probability of employment is a central issue in this paper. It is important to proceed cautiously before any final conclusions are reached on this matter.

Testing this hypothesis raises a technical issue that cannot be evaded. There are a variety of asymptotically equivalent test statistics available to test the same hypothesis (Rao 1973). These alternative test statistics lead to the same inference in large samples but may lead to conflicting inferences in small samples. In the model considered in this paper, there is no theoretical basis for preferring one statistic over another. A recent Monte Carlo study of a nonlinear model somewhat similar (in its degree of nonlinearity) to the one estimated in this paper that compares the asymptotic "t" statistics of the sort presented in table 3.2 with the likelihood ratio statistics obtained from likelihood functions evaluated at restricted and unrestricted values concludes with the advice, "use the likelihood ratio test when the hypothesis is an important aspect of the study" (Gallant 1975).

Following this advice, the statistical models displayed in columns 1, 2, and 3 are reestimated deleting the recent work experience variable from each model. The empirical results from this procedure are reported in columns 4, 5, and 6 which correspond, respectively, to the models associated with the estimates reported in columns 1, 2, and 3.

At a ten percent significance level, one would reject the hypothesis that within-sample work experience does not affect the probability of employment in each of the models. Maintaining the assumption of stationarity in the unobserved variables (see columns 5 and 6) leads to rejection of the hypothesis at five percent significance levels. From these tests we provisionally conclude that recent work experience determines the probability of employment. However, it must be acknowledged that with these data, if the stationarity assumption is not maintained this inference is not strong.

We tentatively conclude that the most appropriate model is one with both recent and presample work experience as determinants of employment and with the disturbances in the equations generated by a stationary first-order Markov process. In order to place these empirical results in perspective, it is useful to compare the model just selected with a recent model presented by Heckman and Willis (1977). Their model is a special case of the general model of equation 3 in which (a) the impact of past participation on the current probability of participation is ignored; (b) the disturbances obey a "permanent-transistory" model so that

$$\varepsilon(i, t) = \phi(i) + U(i, t)$$

where $U(i, t)$ is independently identically distributed, and $\phi(i)$ is a person effect that is not assumed to change over the course of the sample;[20] and

(c) no variation is permitted in the regressors $Z(i, t)$ for an individual during the same period, although differences among individuals are permitted.[21]

Empirical estimates of their model are presented in column 7, table 3.2. Strictly speaking, the model displayed there is "too generous" to Heckman and Willis because it includes predicted presample experience in the model, deleting the effect of *recent* experience on participation. The only innovation in notation is the symbol "η" defined as the ratio of the variance in $\phi(i)$ to the variance in $\varepsilon(i, t)$. This parameter is important in the Heckman-Willis analysis because as they show in their appendix, a value of η in excess of ½ implies that the distribution of participation probabilities among a group of women with identical observed characteristics is U-shaped with most of the mass of the distribution concentrated near zero or one (i.e., most women work nearly all of the time or not at all). Since η is estimated to be .94, the implied distribution of probabilities is strongly U-shaped.

A direct comparison between the new model with estimated coefficients reported in column 3 and the Heckman-Willis model is not possible using conventional testing criteria since neither model nests the other as a special case. However, as previously noted, examination of the empirical results for the general models of columns 1 or 4 suggests that the correlation pattern for the unobservables favors the first-order Markov structure and not the "permanent-transistory" structure which imposes the restriction of equicorrelation among disturbances ($\alpha_1 = \alpha_2 = \alpha_3$). Moreover, as previously noted, there is evidence that recent work experience determines employment.

One way to compare the two models is to examine their predictive power on fresh data. Table 3.3 displays the results of such a comparison. In column 1, the actual numbers in each pattern of labor force activity are recorded. In the remaining columns, the numbers predicted from the model described at the top of the column are recorded. In particular, the predicted numbers from the new model are recorded in column 2 while predicted numbers from the Heckman-Willis model are recorded in column 3. The bottom row of the table records the χ^2 goodness of fit test. A lower value of the χ^2 statistic implies a better fit for a model.

A major difference in predictive power between the two models comes in the interior cells of the table that register labor market turnover. The model developed in this paper is more accurate in predicting labor force turnover than is the Heckman-Willis model, especially for the turnover pattern of women who work most of the time.

Given that Heckman and Willis ignore the impact of past participation on current participation, and hence relegate this effect to the disturbance term in their model, it is plausible that their disturbance terms exhibit a greater degree of intertemporal correlation than is present in the model

Table 3.3   Comparisons of Models Using Runs Data—Women Aged 45–59 in 1968 (Data for 1971, 1972, 1973, Three years following the sample data used to estimate the model)

Runs Pattern[a]	(1) Actual Number	(2) Number Predicted from the New Model (Column 3, Table 3.2)	(3) Number Predicted from Heckman-Willis Model (Column 7, Table 3.2)	(4) Probit Model That Ignores Heterogeneity (Column 9, Table 3.2)	(5) Probit Model That Ignores Heterogeneity and Recent Sample State Dependence (Column 10, Table 3.2)
0, 0, 0	96	94.2	139.5	145.3	36.1
0, 0, 1	5	17.6	4.1	38.5	20.5
0, 1, 0	4	1.8	4.1	1.9	20.2
1, 0, 0	8	2.6	4.1	.35	20.6
1, 1, 0	5	1.4	3.6	.02	21.2
1, 0, 1	2	2.4	3.6	1.38	21.1
0, 1, 1	2	16.4	3.6	8.51	21.7
1, 1, 1	76	61.5	34.9	2.05	36.6
χ^{2b}	. . .	48.5	66.3	4,419	221.8

[a]"0" corresponds to not working, "1" corresponds to working. Thus "1, 1, 0" corresponds to a woman who worked the first two years of the sample and did not work in the final year.
[b]This is the standard chi-square statistic for goodness of fit. The higher the value of the statistic, the worse the fit.

estimated in this paper. Moreover, their analysis overstates the amount of heterogeneity in probabilities at a point in time, and ignores the life cycle evolution of the distribution of employment probabilities that arises from the impact of past employment on current employment. Their estimate of heterogeneity, and the U shape in the distribution of employment probabilities, overstate the extent of heterogeneity, especially at the youngest ages, and overstate the degree of intertemporal correlation in error terms, and the persistence over time in the correlation of unobservables. The evolutionary view of the participation process offered in this paper is considerably more dynamic than the view offered by Heckman and Willis.

It is of some interest to estimate the Heckman-Willis model relaxing their assumption that the regressors do not change over the sample period while retaining their other assumptions. Estimates of this model are presented in column 8 of table 3.2.

This model is a special case of the model estimated in column 2 in which the α_j are restricted to equality and current experience is deleted from the model. The model under consideration thus imposes three restrictions on the parameters ($\alpha_1 = \alpha_2 = \alpha_3$, and $\delta = 0$). Twice the difference in log likelihood between the two models is 14 which is to be compared with a χ^2 statistic of 7.8 with three degrees of freedom using a five percent significance level. Accordingly, one would reject the null hypothesis that the model with estimates reported in column 8 explains the data better than the model of column 2.[22]

These tests suggest that the principal defect in the Heckman and Willis scheme is not the assumption that the regressors are fixed over the sample period. The real problem comes in the permanent-transitory error structure for unobservables, and the neglect of "true" state dependence.

In order to illustrate the importance of treating heterogeneity correctly in estimating dynamic models, we consider models that are estimated ignoring heterogeneity. The model presented in column 9 of table 3.2 is the model of column 1 in which no heterogeneity is permitted ($\alpha_1 = \alpha_2 = \alpha_3 = 0$) so that the unobservables in different periods are assumed to be uncorrelated. A likelihood ratio test clearly rejects the hypothesis of no heterogeneity. The effect of recent market experience on employment is dramatically overstated in a model that neglects heterogeneity. Compare the estimated effect of recent market experience on current employment status that is recorded in column 9 (1.46) with the estimated effect reported in column 1 (.136). Ignoring heterogeneity in estimating this effect would lead to an overstatement of the impact of past work experience on current employment by a factor of ten! Too much credit would be attributed to past experience as a determinant of employment if intertemporal correlation in the unobservables is ignored. Moreover, a comparison of estimated effects of national unemployment

on employment suggests that the model that ignores heterogeneity dramatically overstates the impact of this variable on employment. The effect of children on employment is *understated* in a model that ignores heterogeneity.

Another way to gauge the importance of heterogeneity in the unobservable variables is to consider how well a model that utilizes past work experience as a regressor but ignores unmeasured heterogeneity predicts sample runs patterns. It is plausible to conjecture that "lagged employment" might serve as a good "proxy" for the effect of heterogeneity. To explore this conjecture, consider the numbers displayed in column 4 of table 3.3.

As is familiar from a reading of the literature on the "mover-stayer" problem, a model that ignores unmeasured heterogeneity underpredicts the number of individuals who either work all of the time or do not work at all. A dynamic model estimated without controlling for heterogeneity will overstate the estimated frequency of turnover in the labor force. In table 3.3 the overstatement is dramatic. The overall "goodness of fit" statistic is decidedly inferior to the goodness of fit statistics for the preceding models. A simple lagged work status "proxy" for heterogeneity does not adequately substitute for a more careful treatment of heterogeneity.

Next, consider a model that ignores both heterogeneity and the effect of *recent* employment on current employment. Estimates of such a model are presented in column 10 of table 3.2. A likelihood ratio test strongly rejects this specification of the general model. The simulations reported in column 5 of table 3.3 suggest that introducing "lagged employment status" into the model as a substitute for a more careful treatment of heterogeneity is an imperfect procedure and is worse than using no proxy at all. Moreover, a model that does not allow for heterogeneity or state dependence dramatically overestimates the extent of labor market turnover.

The two models just discussed have one feature in common: they can be estimated from a single cross section of data.[23] Accordingly, comparisons between the performance of models that ignore heterogeneity, and models that account for heterogeneity, reveal the potential value of panel data for estimating models that can accurately forecast labor market dynamics. Labor supply functions fit on cross-section data overstate the true extent of turnover in the labor force. Ad hoc "proxies" for heterogeneity generate models that yield misleading forecasts of the true microdynamics of the labor market.

We now turn to the empirical results for younger women. Table 3.4 is identical in format to table 3.2. No estimates are given for the models of columns 1, 4, and 5 of the table. The reason for this is that estimated values of α_2 strongly tend to unity leading to numerical instability in

Table 3.4　　　　　Estimates of the Model for Women Aged 30–44 in 1968 (asymptotic normal statistics in parentheses)

	(1)	(2)	(3)	(4)	(5)
Intercept	—	.257(.66)	−.15(.27)	—	—
Number of children younger than 6	—	.293(2.7)	−.31(3.1)	—	—
County unemployment rate (%)	—	-1.21×10^{-3}(.06)	-1.03×10^{-3}(.05)	—	—
County wage rate ($/hr)	—	−.074(.8)	−.080(.71)	—	—
Number of children	—	9.1×10^{-3}(.2)	−.0125(.3)	—	—
Wife's education (yrs)	—	.0324(1.9)	.056(1.9)	—	—
Family income excluding wife's earnings	—	-1.62×10^{-5}(2.1)	-1.50×10^{-5}(2.3)	—	—
National unemployment rate	—	−.378(3)	−.139(1.1)	—	—
Current experience	—	.366(3.7)	.116(1.0)	—	—
Predicted presample experience	—	.057(2.1)	.0324(1.0)	—	—
α_1	—	.93(23)	—	—	—
α_2	—	$.9998(6 \times 10^3)$	—	—	—
α_3	—	.926(17.8)	—	—	—
σ_{22}	—	1	—	—	—
σ_{33}	—	1	—	—	—
ρ	—	—	.844(16)	—	—
η	—	—	—	—	—
ln likelihood	—	−479.5	−481.1	—	—

Table 3.4 (continued)

	(6)	(7)	(8)	(9)	(10)
Intercept	$-.387(.7)$	$-.91(0)$	$-.397(.77)$	$1.4(2.9)$	$.0379(.08)$
Number of children younger than 6	$-.277(3.0)$	$-.28(2.1)$	$-.270(3.2)$	$-.34(3.3)$	$-.37(3.9)$
County unemployment rate (%)	$-3.5 \times 10^{-3}(.19)$	$-4.4 \times 10^{-3}(.15)$	$-9.58 \times 10^{-4}(.05)$	$.0107(.45)$	$-4.51 \times 10^{-3}(.2)$
County wage rate ($/hr)	$-7.89 \times 10^{-2}(.76)$	$-.21(1.9)$	$-.089(.84)$	$-.265(2.4)$	$-.336(3.3)$
Number of children	$-8.32 \times 10^{-3}(.19)$	$.01(6.1)$	$-9.077 \times 10^{-3}(.22)$	$3.8 \times 10^{-3}(.12)$	$2.2 \times 10^{-3}(.07)$
Wife's education (yrs)	$.055(1.8)$	$.08(3.0)$	$.052(1.8)$	$.065(2.8)$	$.075(3.6)$
Family income excluding wife's earnings	$-1.21 \times 10^{-5}(2.0)$	$-2.3 \times 10^{-5}(2.4)$	$-1.159 \times 10^{-5}(2.0)$	$-2.3 \times 10^{-5}(3.6)$	$-2.76 \times 10^{-5}(4.)$
National unemployment rate	$-.048(.74)$	$-.041(.69)$	$-.045(.7)$	$-1.06(7.2)$	$5.8 \times 10^{-3}(.05)$
Current experience	—	—	—	$1.14(14)$	—
Predicted presample experience	$.042(1.2)$	$.052(1.7)$	$.044(1.4)$	$.02(.69)$	$.038(1.7)$
α_1	—	—	—	—	—
α_2	—	—	—	—	—
α_3	—	—	—	—	—
σ_{22}	—	—	—	—	—
σ_{33}	—	—	—	—	—
ρ	$.886(39)$	$.86$	—	—	—
η	—	—	$.846$	—	—
ln likelihood	-482.05	-489.8	-491.92	-520.4	-652.5

(See following page for source notes)

evaluating the sample likelihood function. At $\alpha_2 = 1$, the likelihood function assumes a limiting functional form that is mathematically different from the model with $\alpha_2 \neq 1$. Because of the instability, the estimates recorded in column 2 of table 3.4 are somewhat suspect.

Since the estimated values of α_1 and α_3 are virtually identical, and since the estimated value of α_2 tends to unity, the data appear to be consistent with a first-order Markov scheme for the unobservables. This result is in accord with the analysis for older women. When the first-order Markov scheme is imposed (see column 3 of table 3.4) the decrease in log likelihood is negligible and well within sampling variation. Moreover, since the computational procedure is much more stable when the first-order Markov restriction for unobservables is imposed, the estimates (and test statistics) reported in column 3 are to be preferred to the estimates reported in column 2.

The major difference in the results for younger women as compared with the results for older women comes in the importance of recent experience for current employment decisions. The "t" statistic on recent experience is 1.0, and would lead to acceptance of the null hypothesis that recent experience is not a determinant of current employment. Anticipating a potential conflict between the statistical inference based on this statistic and the likelihood ratio test, along the lines previously discussed, the model is reestimated deleting current experience (results are reported in column 6). The change in log likelihood is trivial. This suggests that recent work experience does not determine current employment. These tests lead to adoption of the model with estimates reported in column 6 of table 3.4 as the appropriate model for younger married women, i.e., a model without any effect of recent experience on current participation decisions.

Columns 7 through 10 of table 3.4 record estimates of models directly comparable to the models with the corresponding headings in table 3.2. Estimates of the Heckman-Willis analogue are reported in column 7.

Table 3.4 Source Notes
Note: Increases in variables with positive coefficients increase the probability of employment, while increases in variables with negative coefficients decrease the probability of employment.
Presample experience is predicted from all of the other regressors in $Z(i, t)$, a set of dummy variables for education, city size variables, regional variables, and family background variables (mother's and father's education). Note that utilizing predicted experience in place of actual experience in the probit model assumes that the errors from prediction are approximately normally distributed. Even if this is not a correct assumption, this procedure permits a valid statistical test of the important null hypothesis that past labor force experience has no effect on current experience. The standard errors of the estimated coefficients must be interpreted as *conditional* on the predicted values of presample experience variables.

Estimates of the model that retains the permanent-transitory structure, and ignores the impact of recent participation on current choices, but permits the regressors to change over the sample period, are reported in column 8. That model is clearly rejected by the data.

Columns 9 and 10 of table 3.4 report estimates of models that ignore heterogeneity. As in the case of older women, neglect of heterogeneity leads to a systematic overstatement of the effect of recent experience on employment choices. (See the estimate reported in column 9 and compare with the estimates reported in columns 2 and 3.) Assuming that the model of column 6 is the correct model, the misspecified model of column 9 dramatically overstates (in absolute value) the negative effect of aggregate unemployment on employment, leads to an overstatement of the effect of income on employment (again, in absolute value), and overstates the effect of education on employment. Similar remarks apply to the empirical results reported in column 10. Tests of the predictive power of the alternative models are similar to the tests reported for older women and so are not discussed here. (They are available on request from the author.)

The main conclusions of the empirical analysis are as follows. (1) For older women, there is evidence that recent labor market experience is a determinant of current employment decisions. There is no such evidence for younger women. (2) There is considerable evidence that the unobservables determining employment choices follow a first-order Markov process. The estimated correlation coefficients for both age groups are comfortably close. (3) Dynamic models that neglect heterogeneity overestimate labor market turnover. "Proxy methods" for solving the problems raised by heterogeneity such as ad hoc introduction of lagged work experience variables lead to dynamic models that yield exceedingly poor forecast equations for labor force turnover. Models that neglect recent market experience and heterogeneity actually perform better in forecasting turnover on fresh data, but these forecasts are still poor, and considerably overestimate the amount of turnover in the labor market. (4) Models that neglect heterogeneity lead to biased estimates of the effect of all variables on labor force participation probabilities in models that include past employment as a determinant of current employment.

Since the unobservables that determine employment probabilities follow a first-order Markov process, standard procedures for introducing heterogeneity into dynamic models do not work, and may lead to erroneous estimates of structural parameters, especially in models that explicitly allow for state dependence. A simple "components of variance" scheme gives misleading estimates of structural parameters and generates forecasts of work force turnover that are quite erroneous. While the assumption of the "convolution property" or "components of variance" scheme is mathematically convenient, its application in empir-

ical work may result in a misleading characterization of population heterogeneity. Given that heterogeneity arises, in part, from omitted variables that plausibly change over time, it is reasonable to expect that there is decay in the correlation between unobservable variables that determine choices the more distant in time the choices are.

3.3.1 Qualifications and Suggested Extensions of the Empirical Analysis

As is always the case in empirical work, there is considerable room for improvement in the data and in the approach taken to analyze the data. In this paper six improvements seem especially warranted.

First, the rigid separation of presample from within-sample work experience is a crude way to allow for depreciation in the effect of past work experience on current labor supply. A more appropriate procedure would use the general models of equation 3 and Heckman (1981a) on longer panels of data to estimate less ad hoc depreciation schemes.

Second, the procedure used to "solve" the initial conditions problem in this paper is exact only for the construction of test statistics for the null hypothesis of no structural state dependence, although Monte Carlo work presented elsewhere (Heckman 1981b) suggests that the procedure performs reasonably well in generating estimates. The exact solution proposed in Heckman (1981b) remains to be empirically implemented.

Third, more explicit economic models should be estimated. The procedures proposed here are useful exploratory tools but are no substitute for an explicit dynamic economic model.

Fourth, a normality assumption has been employed in the empirical work although it is not essential to the approach. The one-factor model is an especially flexible format within which to relax this assumption. It would be of great interest to examine the sensitivity of the estimates and the accuracy of model forecasts under alternative distributional assumptions. Especially interesting would be an examination of distributions that allow for more general forms of nonstationarity in the unobservables than are permitted in the multivariate normal.

Fifth, the entire empirical analysis has been conducted in discrete time yet employment decisions are more suitably modeled in continuous time. The empirical treatment of the time unit is largely a consequence of the availability of the data. Approximating a continuous time model by a discrete time model results in a well-known time aggregation bias (see Bergstrom 1976). For one continuous time model of heterogeneity and state dependence, see Heckman and Borjas (1980).

Sixth, unemployment has been treated as a form of "leisure" or non-market time. A more general approach consistent with much recent work treats measured unemployment as a separate decision variable. While estimation of the more general model is more costly it is also more

informative. Estimates of such a model would enable analysts to determine whether being unemployed is, in fact, a separate activity distinct from being out of the labor force. For one approach, see Flinn and Heckman (1981).

3.4   What Does Structural State Dependence Mean?

Granting the validity of the preceding evidence in support of structural state dependence in the employment decisions of older women, it remains to interpret it. This section presents a brief menu of behavioral models that generate structural state dependence. Before these models are presented, however, it is useful to restate the key statistical assumption used to secure this evidence.

In the discussion surrounding equation 3, a distinction is made between the effects of unmeasured variables—the $\varepsilon(i, t)$—and prior work experience—lagged $d(i, t)$—on choices made in period t. The crucial assumption not subject to test in this paper is that the unmeasured variables cause but are not caused by prior choices. "Cause" is used in the sense of Sims (1977) suitably modified for a discrete data model. That is, the conditional distribution of $\varepsilon(i, t)$ given all lagged values of $\varepsilon(i, t)$ and all lagged values of $d(i, t)$ is the same as the conditional distribution of $\varepsilon(i, t)$ given all lagged values of $\varepsilon(i, t)$.

Structural state dependence is defined to exist if the conditional distribution of $d(i, t)$ given all past values of $\varepsilon(i, t)$ and lagged $d(i, t)$ is a nontrivial function of the latter set of variables. Structural dependence is tested in this paper by a discrete data analogue of time series causality tests. Correctly conditioning the distribution of current $\varepsilon(i, t)$ "controls" for the effect of past $\varepsilon(i, t)$ on current $d(i, t)$.

The validity of the estimates of and tests for structural state dependence presented here depends on the validity of this untested assumption. If, in fact, the unmeasured variables are caused by lagged $d(i, t)$, the statistical procedures discussed in section 3.2 and implemented in section 3.3 are inappropriate. Evidence of serial correlation derived from these procedures may, in fact, be evidence of structural state dependence. Evidence for or against structural state dependence derived from the procedures presented in this paper will necessarily be inconclusive.

In any empirical application of our procedures, the maintained hypothesis will be controversial. In our analysis of the employment decisions of women this is the case. Following the analysis of Ryder, Stafford, and Stephan (1976), women may devote more time to human capital investment in periods in which they work than in other periods if the cost of investing is lower on the job than off. Since investment time is not observed, and is not an exact function of employment status, controlling for prior work experience as we have done only imperfectly accounts for

human capital investment. Estimated heterogeneity will arise from human capital investment. The unobservables are "caused" by past employment.[24] However, there is also structural state dependence as we have defined it in this paper.

Granting the validity of the maintained assumption, at least as a first approximation, it is of some interest to consider how well-defined economic models can generate structural state dependence. Apart from the model just discussed, three further examples are presented: a model of stimulus-response conditioning of the sort developed by mathematical psychologists, a model of decision making under uncertainty, and a model of decision making under perfect foresight.

In the stimulus-response model developed by behavioral psychologists (e.g., Bush and Mosteller 1955; Restle and Greeno 1970; Johnson and Kotz 1977), the individual who makes a given "correct" response is rewarded, so that he is more likely to make the response in the future. Decision making is myopic. This model closely resembles the generalized Polya process discussed above. Models that resemble the stimulus-response model have been proposed by dual labor market economists who assume that individuals who are randomly allocated to one market are rewarded for staying in the market and are conditioned by institutions in that market so that their preferences are altered. The more time one has spent in a particular type of market, the more likely one is to stay in it (Cain 1976).

The model of myopic sequential decision making just discussed is unlikely to prove attractive to many economists. Nonmyopic sequential models of decision making under imperfect information also generate structural state dependence. Such models have been extensively developed in the literature on dynamic programing (e.g., Dreyfus 1965, pp. 213–15; Astrom 1970). An example is a model in which an agent at time t maximizes expected utility over the remaining horizon, given all the information at his disposal and given his constraints as of time t. Transition to a state may be uncertain. As a consqeuence of being in a state, costs may be incurred or information may be acquired that alters the information set or opportunity set or both relevant for future decisions. In such cases the outcome of the process affects subsequent decision making, and structural state dependence is generated.

The disturbance in this model consists of unmeasured variables known to the agent but unknown to the observing economist as well as unanticipated random components unknown to both the agent and the observing economist.

Structural state dependence can also be generated as *one representation* of a model of decision making under perfect certainty. In such a model there are no surprises. Given the initial conditions of the process, the full outcome of the process is perfectly predictable from information

available to the agent (but not necessarily available to the observing economist).

To illustrate this point in the most elementary way, consider the following three-period model of consumer decision making under perfect certainty with indivisibility in purchase quantities: a consumer's strictly concave utility function is specified as

$$U[a(1)d(1), \ a(2)d(2), \ a(3)d(3)]$$

where the $a(i)$ are the fixed amounts that can be consumed in each period. The consumer purchases amount $a(i)$ if $d(i) = 1$, otherwise $a(i) = 0$. Resources M are fixed so that

$$\Sigma a(i)d(i) \le M$$

The agent has full information and selects the $d(i)$ optimally. Optimal solutions are denoted by $d^*(i)$.

An alternative characterization of the problem is the following sequential interpretation. Given $d^*(1)$, maximize utility with respect to remaining choices.
Thus

$$\underset{d(2),d(3)}{\text{Max}} \ U[a(1)d^*(1), \ a(2)d(2), \ a(3)d(3)]$$

$$\text{subject to} \ \overset{3}{\underset{i=2}{\Sigma}} \ a(i)d(i) \le M - a(1)d^*(1)$$

The demand functions (really the demand inequalities) for $d(2)$ and $d(3)$ may be written in terms of $d^*(1)$ and available resources $M - a(1)d^*(1)$. This characterization is a discrete choice analogue of the Hotelling (1935), Samuelson (1960), Pollak (1969) treatment of ordinary consumer choice and demonstrates that the demand function for a good can be expressed as a function of quantities consumed of some goods, the prices of the remaining goods, and income. (Pollak's term, "conditional demand function," is felicitous.)[25]

The choice of which characterization of the decision problem to use is a matter of convenience. When the analyst knows current disposable resources $M - a(1)d^*(1)$ and past choices $d^*(1)$ but not $a(1)$ or M, the second form of the problem is econometrically more convenient. The conditional demand function gives rise to structural state dependence in the sense that past choices influence current decisions. The essential point in this example is that past choices serve as a legitimate proxy for missing M and $a(1)$ variables known to the consumer but unknown to the observing econometrician. The conditional demand function is a legitimate structural equation.[26]

Both a model of decision making under uncertainty and a model of decision making under perfect foresight may be brought into sequential form so that past outcomes of the choice process *may* determine future outcomes. In principle one can distinguish between a certainty model and an uncertainty model if one has access to all the relevant information at the agent's disposal. In a model of decision making under perfect certainty, if all past prices are known and entered as explanatory variables for current choices, past outcomes of the choice process contribute no new information relevant to determining current choices. In a model of decision making under uncertainty, past outcomes would contribute information on current choices not available from past prices since uncertainty necessarily makes the prediction of past outcomes from past prices inexact, and the unanticipated components of past outcomes alter the budget set and cause a revision of initial plans.[27] In practice it is difficult to distinguish between the two models given limitations of data. The observing economist usually has less information at his disposal than the agent being analyzed has at his disposal when he makes his decisions.

The key point to extract from these examples is that structural state dependence as defined in this paper may be generated from a variety of models. It is not necessary to assume myopic decision making to generate structural dependence. Nor does empirical evidence in support of structural state dependence prove that agents make their decisions myopically. The divergence in estimated state dependence effects for the two age groups of women *may* be reconciled, in part, by an appeal to human capital theory. Under this interpretation previous work experience may be viewed as a proxy for investment in market human capital. The higher the stock of market-oriented human capital, the more likely is the event that a woman works (ceteris paribus). It is likely that the investment content of recent work experience is lower for women in their child-rearing years, when there are many competing demands for their time, than it is for older women past the child-rearing period who are reentering the work force in earnest. However, the empirical evidence on structural state dependence presented here is consistent with a variety of interpretations, and without further structure imposed we cannot be precise about which source of state dependence explains our empirical results.

Appendix A presents a first attempt at a decomposition of estimated structural state dependence effects into wage and nonwage components. It is estimated that forty-nine percent of an estimated structural state dependence effect arises from the effect of work experience on raising wage rates and the subsequent effect of higher wage rates on employment. A full fifty-one percent of estimated structural state dependence arises from other sources.

2.5 Summary

This paper presents a statistical model of discrete dynamic choice and applies the model to address the problem of distinguishing heterogeneity from structural state dependence. The concept of heterogeneity is generalized, and the concept of structural state dependence is given an economic interpretation. The methodology developed here is applied to analyze the dynamics of female labor supply. Evidence of structural state dependence is found for older women. There is little evidence that recent work experience determines the labor supply of younger women once heterogeneity is properly controlled. Heterogeneity arising from unobservables is found in the data for both groups of women. However, the traditional permanent-transitory model of heterogeneity is found to be inappropriate; for women, a first-order Markov model is a better description of the error process. Ad hoc shortcut procedures for controlling for heterogeneity are shown to produce erroneous estimates and forecasts.

Appendix A

A Procedure for Identifying Separate Components of the Effect of Previous Work Experience on Current Participation, and Some Preliminary Empirical Results

In the text, a procedure for estimating the impact of work experience on employment is proposed and implemented. As noted in the text, evidence for the existence of structural state dependence is consistent with several different hypotheses. One hypothesis is that work experience raises wage rates and that wage rates in turn influence employment. A second hypothesis is that fixed costs of entry into and exit out of the work force cause women to bunch their employment spells. A third hypothesis is that household-specific capital is acquired by women who do not participate in the market and that this nonmarket capital causes women who have not worked in the past to be less likely to work in the future. Closely related to this hypothesis is the hypothesis of "reinforcement" of work or nonwork activity of the sort considered by mathematical learning theorists. Other hypotheses have been advanced, and each hypothesis can be further specialized (Heckman 1981a), but for the purposes of the present discussion it will be assumed that these hypotheses are exhaustive explanations of structural state dependence.

In this appendix, a method for isolating these three effects is proposed, and some preliminary empirical evidence with this method is presented. Forty-nine percent of the estimated effect of work experience on employ-

ment is estimated to be due to the effect of market experience on wage growth.

The basic idea underlying the methodology is very simple. If measures of wage rates and nonmarket capital are available, it is possible to estimate the effect of work experience on these measures, as well as the direct effect of work experience on employment. If one can determine how nonmarket capital, fixed costs, and wage rates determine employment, one can apportion an estimated structural state dependence effect among these three sources.

The following equation system underlies the analysis in this appendix.

(A1)
$$V(i, t) = Z(i, t)\tilde{\beta} + \delta_W W(i, t) + \delta_H H(i, t)$$
$$+ \delta_F F(i, t) + \bar{\varepsilon}(i, t)$$
$$\text{and } d(i, t) = 1 \text{ iff } V(i, t) > 0$$
$$d(i, t) = 0 \text{ otherwise}$$

(A2)
$$W(i, t) = Z_W(i, t)\gamma_W + \eta_W \sum_{t > t'} d(i, t') + U_1(i, t)$$

(A3)
$$H(i, t) = Z_H(i, t)\gamma_H + \eta_H \sum_{t > t'} d(i, t') + U_2(i, t)$$

(A4)
$$F(i, t) = Z_F(i, t)\gamma_F + \eta_F \sum_{t > t'} d(i, t') + U_3(i, t)$$

$$\text{for } i = 1, \ldots, I \quad t = 1, \ldots, T$$

Equation A1 corresponds to equation 3 in the text except that it distinguishes a wage effect on employment $\delta_W W(i, t)$, a nonmarket capital effect on employment $\delta_H H(i, t)$, and a fixed cost effect on employment $\delta_F F(i, t)$. Equations A2–A4 are, respectively, equations explaining wage rates $W(i, t)$, nonmarket capital $H(i, t)$, and fixed costs $F(i, t)$. $Z_W(i, t)$, $Z_H(i, t)$, and $Z_F(i, t)$ are the exogenous explanatory variables in the equations in which they appear. Substituting these equations into equation A1, leads to a specialization of equation 3 in the text

$$V(i, t) = Z(i, t) \beta + \delta \sum_{t > t'} d(i, t') + \varepsilon(i, t)$$

where $\delta = \delta_W \eta_W + \delta_H \eta_H + \delta_F \eta_F$ and $\varepsilon(i, t) = \delta_W U_1(i, t) + \delta_H U_2(i, t) + \delta_F U_3(i, t) + \bar{\varepsilon}(i, t)$.

If one can estimate $\delta_W \eta_W$, $\delta_H \eta_H$, and $\delta_F \eta_F$, one can allocate the structural state dependence effect δ into wage sources, nonmarket sources, and fixed costs sources.

Equation system A1–A4 is a special case of a dummy endogenous variable model (see Heckman 1978a, especially appendix B). If the disturbances $\bar{\varepsilon}(i, t)$, $U_1(i, t)$, $U_2(i, t)$, $U_3(i, t)$ are jointly normally dis-

tributed with mean zero and variance covariance matrix Σ, the analysis of estimation and identification developed for the dummy endogenous variable model carries over to the model discussed in this appendix. In particular, without using covariance restrictions, and relying solely on classical exclusion restrictions, if one variable appears in $Z_W(i, t)$ that does not appear in $Z(i, t)$, $Z_H(i, t)$, and $Z_F(i, t)$, it is possible to estimate δ_W up to a factor of proportionality given by the standard deviation of $\varepsilon(i, t)$.[28] Permuting the subscripts W, H, and F generates necessary conditions for identifiability of normalized values of δ_H and δ_F.

In order to estimate the contribution of each of the three components of structural state dependence to the (normalized) total effect, δ, only two of the three left-hand-side variables that appear in equations A2–A4 need be observed, and the exclusion restrictions must be satisfied for the two equations for which observations on the dependent variable are available. To see why this is so, assume that data are available on $W(i, t)$ and $H(i, t)$ but not on $F(i, t)$. Given exclusion restrictions, this information can be used to estimate γ_W, η_W, γ_H, η_H and hence (normalized) δ_W and δ_H.[29] From the reduced form equation 3 it is possible to estimate (normalized) δ. Thus one can estimate (normalized) $\delta_F \eta_F = \delta - \delta_W \eta_W - \delta_H \eta_H$. If data are available on only one of the three variables and the exclusion restrictions are satisfied for the equation for which the dependent variable is available, one can estimate the fraction of the (normalized) total effect, δ, due to the variable that is observed.

In the analysis of female labor supply, a direct measure of market capital is available: the market wage rate, $W(i, t)$. Direct measures of fixed costs or household capital are not available, although children variables might be used to "proxy" household capital. Accordingly, with the available data, it is possible only to estimate the fraction of structural state dependence due to the effect of work experience on wage rates and the effect of wage rates on employment.

The specific econometric model used to derive the estimates presented in this appendix is based on a *fixed effect*–multiple equation Tobit model developed by the author and T. MaCurdy. That model is a conditional version of the general dummy endogenous variable model, and is described in detail elsewhere (Heckman and MaCurdy 1980, appendix A). In this paper, only the probit wage equation component of that model is used to estimate equation 3 and equation A2.[30] To achieve identification of a wage effect on employment, it is assumed that the local unemployment rate affects labor force participation only through its effect on wage rates—an assumption that could easily be challenged, and which would be counterfactual in a model of labor supply under uncertainty of employment in which local unemployment rates affect expectations of em-

ployment. This assumption is maintained here. Then (normalized) δ_W is just identified.

The variables that appear in the wage equation and affect employment are the same as those used by Heckman and MaCurdy (1980), except that in place of their market experience variable (age minus schooling minus six), actual work experience is used. Log wage rates are assumed to depend on (a) local labor market unemployment; (b) work experience; and (c) schooling variables. Labor force participation is assumed to depend (in reduced form) on these three types of variables used in the wage equation, and on (d) variables representing family composition; (e) family income exclusive of the wife's earnings; (f) the wife's age; and (g) variables representing the head's health status. A sample of 672 white women aged 25–65 in 1968 interviewed in the Michigan Panel Survey of Income Dynamics who were continuously married to the same spouse during the sample period 1968–75 was used to generate the estimates.

In order to focus the discussion on the main topic of this appendix, only the key wage and state dependence parameters are presented in 3.A.1.

For the full sample of women of all ages, the estimated normalized total state dependence effect δ is .163—a number that is between the estimates for the two age groups presented in the text. As is apparent from table 3.A.1, only 49 percent of the estimated effect of market experience on employment is due to the effect of market experience on wage growth and the effect of wage rates on market participation. Fifty-one percent of the estimated structural state dependence effect is due to the acquisition of nonmarket capital (including psychological reinforcement effects) and the effect of fixed costs. This estimate, though clearly tentative, suggests that a considerable part of the effect of work experience on employment is due to factors other than the wage-rate-enhancing

Table 3.A.1.	Estimates of the Total State Dependence Effect and Its Wage Component (asymptotic normal test statistics derived from the estimated information matrix in parentheses)		
(Normalized) Total State Dependence Effect δ	Effect of Experience on ℓn Wage Rates η_W	(Normalized) Effect of Wage Rates on Employment $\delta_W{}^*$	Fraction of Total State Dependence Effect Due to the Effect of Experience on Wage Growth**
.163	.032	2.45	.49
(1.9)	(4.9)	(2.9)	(3.6)

*This is obtained by dividing the estimated effect of local unemployment rates on participation by the estimated effect of local unemployment rates on ℓn wage rates.
**This is obtained by multiplying η_W and δ_W, and taking the ratio of this product to δ.

Appendix B

The derivation of the likelihood function used in this paper is presented in Heckman (1981a). Here we present the essential features of the derivation and the problem of initial conditions discussed extensively in Heckman (1981b).

Write equation 3 in the text in shorthand notation as

$$V(i, t) = \bar{V}(i, t) + \varepsilon(i, t)$$

for $i = 1, \ldots, I$ and $t = 1, \ldots, T$ where

$$\bar{V}(i, t) = Z(i, t)\beta + \sum_{t > t'} \delta(t, t')\, d(i, t')$$

$$+ \sum_{j} \lambda(t, t-j) \prod_{\ell = 1}^{j} d(i, t-\ell)$$

Let $\varepsilon(i, t)$ be arrayed in a $1 \times T$ vector $\varepsilon(i)$, and array $\bar{V}(i, t)$ in a $1 \times T$ vector $\bar{V}(i)$. The initial conditions of the process are assumed to be fixed nonstochastic constants $d(i, 0), d(i, -1), \ldots$; $V(i, t) > 0$ iff $d(i, t) = 1$; $V(i, t) \leq 0$ otherwise.

The disturbances are assumed to be joint normally distributed

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

Define diagonal matrix D as the square root of the diagonal elements of Σ. Normalize $\sigma_{11} = 1$. Define the correlation matrix by

$$\tilde{\Sigma} = D^{-1} \Sigma\, D^{-1}$$

and define the normalized $\bar{V}(i)$ by

$$\tilde{\bar{V}}(i) = \bar{V}(i)\, D^{-1}$$

and the normalized $\varepsilon(i)$ vector by

$$\tilde{\varepsilon}(i) = \varepsilon(i)\, D^{-1}$$

Conditional density (equation 4 in the text) is most conveniently defined in recursive fashion. Here we simply start the recursion. The remaining steps are obvious and hence are deleted. Define the joint density of $\tilde{\varepsilon}(i, 2)$ and $\tilde{\varepsilon}(i, 1)$ as

$$f_{21}[\tilde{\varepsilon}(i, 2), \tilde{\varepsilon}(i, 1)]$$

The conditional density of $\bar{\varepsilon}(i, 2)$ given $d(i, 1)$ is

$$g[\varepsilon(i, 2) \mid d(i, 1)] =$$

$$\left[\frac{\int_{-\bar{V}(i, 1)}^{\infty} f_{21}[\bar{\varepsilon}(i, 2), \bar{\varepsilon}(i, 1)] \, d\bar{\varepsilon}(i, 1)}{\int_{-\bar{V}(i, 1)}^{\infty} f_1[\bar{\varepsilon}(i, 1)] \, d\bar{\varepsilon}(i, 1)} \right]^{d(i, 1)}$$

$$\left[\frac{\int_{-\infty}^{-\bar{V}(i, 1)} f_{21}[\bar{\varepsilon}(i, 2), \bar{\varepsilon}(i, 1)] \, d\bar{\varepsilon}(i, 1)(1 - d(i, 1))}{\int_{-\infty}^{-\bar{V}(i, 1)} f_1[\bar{\varepsilon}(i, 1)] \, d\bar{\varepsilon}(i, 1)} \right]^{1 - d(i, 1)}$$

where f_1 is the marginal density of $\bar{\varepsilon}(i, 1)$. $\bar{\varepsilon}(i, 2)$ *conditioned* on $d(i, 1)$ is independent of $d(i, 1)$. Thus the probability of $d(i, 2) = 1$ conditional on $d(i, 1)$ is, in the notation of the text,

$$P(i, 2) = \int_{-\bar{V}(i, 2)}^{\infty} g[\varepsilon(i, 2) \mid d(i, 1)] \, d\varepsilon(i, 2)$$

Recall that $\bar{V}(i, 2)$ contains $d(i, 1)$. This creates no simultaneity problem in forming $P(i, 2)$ because $\varepsilon(i, 2)$ is conditionally independent of $d(i, 1)$ by construction.

Define $P(i, 1)$ as

$$P(i, 1) = \int_{-\bar{V}(i, 1)}^{\infty} f_1[\varepsilon(i, 1)] d\varepsilon(i, 1)$$

The joint density of $d(i, 1)$ and $d(i, 2)$ is

$$k[d(i, 1), d(i, 2)] = [P(i, 1)]^{d(i, 1)} [1 - P(i, 1)]^{1 - d(i, 1)}$$
$$\cdot [P(i, 2)]^{d(i, 2)} [1 - P(i, 2)]^{1 - d(i, 2)}$$

The procedure to be used to derive the full distribution of $d(i)$ should now be clear.

A convenient representation of the probability of $d(i)$ that exploits the symmetry of the normal around its mean is the following. Define F as the multivariate cumulative normal integral. The probability that $d(i) = \underline{d(i)}$ given the values of the exogenous variables, the parameters, and the initial conditions is

$$\text{Prob} \, [d(i) = \underline{d(i)}] = F[\bar{V}(i) * \underline{(2d(i))} - \iota;$$
$$\bar{\Sigma} * (2d(i) - \iota)'(2d(i) - \iota)]$$

where "*" denotes a Hadamard product (Rao 1973), and ι is a 1 X T vector of ones. Maximizing the sample product of these probabilities with respect to the parameters of the model produces the maximum likelihood estimator. Modifying this expression for nonnormal symmetric and non-symmetric densities is straightforward.

A crucial assumption in writing down the expression is that presample values of $d(i, 0)$, $d(i, -1)$, . . . are fixed nonstochastic constants. If they are not, correct conditioning for the process requires treating the presample values in the same fashion as the sample values (i.e., conditioning to correct for simultaneity). In life cycle models of the sort considered in the text, this requires the entire history of the process. For a more complete discussion of this problem and for some exact and approximate solutions see Heckman (1981b).

Appendix C   Data Summaries

Table 3.A.2   **Mean Value of Variables**

	1968	1969	1970	1971	1972	1973	1974
			A. Women Aged 45–59				
No. of children aged less than 6	.035	.020	.040	.010	.015	.015	.015
County unemployment rate (%)	4.01	3.61	4.92	5.89	5.85	4.87	5.54
County wage rate ($/hr)	1.90	1.99	2.05	2.09	2.21	2.29	2.48
Total no. of children	2.68	2.68	2.74	2.79	2.86	2.93	2.97
Presample work experience (yr.)[a]	11.22	11.22	11.22	11.22	11.22	11.22	11.22
Wife's education	11.71	11.71	11.71	11.71	11.71	11.71	11.71
Family income excluding wife's earnings ($)	1.26×10^4	1.30×10^4	1.34×10^4	1.39×10^4	1.42×10^4	1.53×10^4	1.56×10^4
Cumulated sample experience[b]	0	.459	.969	1.44	1.91	2.34	2.77
National unemployment rate prime-age males 35–44	1.4	1.4	2.3	2.9	2.5	1.8	2.4
Participation rate	.46	.51	.47	.46	.44	.43	.41
No. of observations	198						

Table 3.A.2 (continued)

	1968	1969	1970	1971	1972	1973	1974
			B. Women Aged 30–44				
No. of children aged less than 6	.392	.325	.280	.250	.204	.135	.123
County unemployment rate (%)	3.83	3.53	4.92	5.81	5.74	4.71	5.40
County wage rate ($/hr)	1.94	2.0	2.05	2.12	2.25	2.30	2.51
Total no. of children	3.20	3.23	3.31	3.37	3.42	3.49	3.54
Presample work experience (yr)[a]	6.51	6.51	6.51	6.51	6.51	6.51	6.51
Wife's education	12.24	12.24	12.24	12.24	12.24	12.24	12.24
Family income excluding wife's earnings ($)	1.28×10^4	1.39×10^4	1.46×10^4	1.59×10^4	1.70×10^4	1.92×10^4	2.08×10^4
Cumulated sample experience[b]	0	.50	1.02	1.52	2.00	2.49	3.00
National unemployment rate prime-age males 35–44	1.4	1.4	2.3	2.9	2.5	1.8	2.4
Participation rate	.5	.52	.50	.49	.49	.51	.54
No. of observations	332						

Source: University of Michigan Panel on Income Dynamics 1968–74.
[a]Defined as the number of years since age 18 that the woman has worked prior to 1968.
[b]Defined as the number of years the woman has worked in the sample years.

Notes

1. For a complete description of the Polya process and its generalizations see Johnson and Kotz (1977, Chapter 4). They note (pp. 180–81) that in the special case in which a person draws a ball and receives the *same number* of the balls of the color drawn whether a black or red ball is drawn, urn model three (in this case a strict Polya model) generates sequences of outcomes identical in probability with the same sequences generated from urn model two provided that the population distribution of the proportion of red and black balls in the urn is *Beta*. In this case, panel data cannot be used to distinguish between the two urn models. In a stationary environment, in which urn contents are not exogenously changed, as long as the number of red balls placed in the urn differs from the number of black balls placed in the urn when a black ball is drawn, it is possible to use panel data to discriminate between the two models. This observation is one of the key insights in the Bates-Neyman paper (1951). A similar nonidentification result appears in the multivariate probit model discussed below. (See note 13).

2. Throughout this section, variables such as children, family income, etc., are ignored solely to simplify the presentation of the main ideas, *not* because they are deemed to be unimportant determinants of participation. Clearly, in actual empirical work it is necessary to control for the effects of such variables on participation probabilities. Note further that the proposed test assumes that model four does not generate the data.

3. Thus this model abstracts from the serial correlation process of the fourth urn scheme.

4. $U(i, t)$ is defined by $U(i, t) = d(i, t) - \phi(i) - \delta \sum_{t > t'} d(i, t')$. The requirement that $U(i, t)$ has a zero mean implies that the probability that $d(i, t) = 1$ conditional on $d(i, t-1), \ldots, d(i, 0)$, and $\phi(i)$ is

$$\Pr[d(i, t) = 1 \mid \phi(i), d(i, t-1), \ldots, d(i, 0)] = \phi(i) + \delta \sum_{t > t'} d(i, t').$$

The conditional probability that $d(i, t) = 0$ is the complement of the preceding probability. Clearly $U(i, t)$ is uncorrelated with $U(i, t')$, $t \neq t'$.

5. The asymptotic test statistics required to perform the test would have to adjust for the heteroscedasticity in the $U(i, t)$. GLS could be used to compute more efficient estimators and test statistics.

6. Readers familiar with the Work of Balestra and Nerlove (1966) will recognize that the model in the text closely resembles the Balestra-Nerlove scheme.

7. Thus, letting β be the effect of children on current participation (assumed to be negative), the probability limits of least squares estimators of β and δ in regression 2 (augmented to include children) are

$$\text{plim} \begin{bmatrix} \hat{\beta} \\ \tilde{\delta} \end{bmatrix} = \begin{bmatrix} \beta \\ \delta \end{bmatrix} + \frac{1}{D} \begin{bmatrix} m_{cc} & -m_{ch} \\ -m_{ch} & m_{hh} \end{bmatrix} \begin{bmatrix} 0 \\ m_{h\varepsilon} \end{bmatrix}$$

where D is the probability limit of the regressor matrix, m_{cc} is the population variance in children, m_{ch} is the population covariance between the number of children and the history of participation $h = \sum_{t > t'} d(i, t')$, and m_{hh} is the population variance participation histories. Finally, $m_{h\varepsilon}$ is the covariance between the disturbance $\varepsilon(i, t) = U(i, t) + \phi(i) - \bar{\phi}(i)$ and the participation history h. D is positive. Under the assumptions made in the text, $m_{h\varepsilon} > 0$ and $m_{ch} < 0$. Thus, least squares estimators overstate the true value of δ and overstate the true value of β.

8. For discussion of this generalized linear probability model, see Heckman and Ma-Curdy (1978).

9. In addition, the small sample properties of the least squares estimator are very undesirable. For example, for $T=2$, least squares estimators of δ in equation 1 are either undefined or nonpositive, even when the true value of δ is positive.

10. Thus we abstract from unemployment. Unemployment is viewed as one form of nonmarket activity, on the same footing as leisure.

11. However, unmeasured business cycle factors and local labor market variables may induce correlation in the errors across individuals.

12. These conditions on equation 3 are sufficient but not necessary. In particular, it is possible for $\delta > 0$.

13. Note the asymmetry. If a woman receives a black ball each time she does not work, the model would have to be augmented to include a time effect. In this case, we obtain a nonidentification result similar to the result mentioned in note 1. For example, consider the following alteration of equation 3. Let

$$V(i, t) = \beta_0 t + \delta \sum_{j=1}^{t} d(i, t-j) + \delta' \sum_{j=1}^{t} [1 - d(i, t-j)] + \varepsilon(i, t)$$

$$= \beta_0 t + (\delta - \delta') \sum_{j=1}^{t} d(i, t-j) + \delta' t + \varepsilon(i, t)$$

$$= (\beta_0 + \delta')t + (\delta - \delta') \sum_{j=1}^{t} d(i, t-j) + \varepsilon(i, t).$$

In this model, if individual i does not experience the event in time period t, she receives a dose of "nonmarket capital" δ'. This model is a special case of equation 3 with t contained among the regressors $Z(i, t)$ (so long as $\beta_0 + \delta' \neq 0$). This underscores the point that the structural state dependence parameters δ in the text are measured only *relative to* nonmarket alternatives.

If $\delta - \delta' = 0$, there is no structural state dependence as defined in the text even though the woman receives a dose of δ when she works and a dose of δ' when she does not. If the doses are of equal strength there is no way to measure the dose, provided $\beta_0 \neq 0$. The nonidentification in this model corresponds to the nonidentification result for the classical Polya urn scheme mentioned in note 1.

14. As discussed in Heckman (1981a), the normality assumption is not critical to this approach. Indeed, the one-factor scheme proposed below can be readily adapted to handle nonnormal disturbances. The advantage of the normality assumption in the general case is that it permits one to control for nonstationarity in the environment. An arbitrary nonstationary distribution would generate an unidentified model. The normal distribution permits the analyst to control for the sorts of nonstationarity in first and second moments usually considered in econometrics.

15. For discussion of one-factor probit model see Heckman (1981a). There it is noted that if $T > 3$, the one-factor assumption imposes nonstationarity onto the error process, except in certain special cases. However, for $T \leq 3$, the one-factor model is consistent with a wide variety of interesting error processes. The principal advantage of the one-factor model is that a multivariate normal integral can be written as one numerical integration of a product of simply computed univariate cumulative normal distributions. This representation greatly facilitates computation.

16. For a complete description of these data, see Morgan, et al. (1974). The restriction to seven years was made to facilitate comparability of samples with other studies, in particular, that of Heckman and MaCurdy (1980).

17. In the complete seven-year sample, fully two-thirds of the women either work all of the time or do not work at all. Again, the numbers of full-time workers and nonworkers are roughly the same.

18. While the coefficient associated with each unemployment variable is not statistically significant at conventional levels, the joint set is statistically significant.

19. The Markov model imposes two restrictions on the α_t coefficients, $\alpha_1 = \alpha_3$, and $\alpha_2 = 1$. Thus two degrees of freedom are lost when this restriction is imposed.

20. This implies that all disturbances have the same correlation with other disturbances ($\alpha_1 = \alpha_2 = \alpha_3$) and the idiosyncratic component is not permitted to decay.

21. Heckman and Willis discuss this model but actually estimate a "beta logistic" approximation to it. The model estimated in the text is the analogue of the Heckman-Willis model estimated in a multivariate probit framework.

22. Note that it is not possible to test the model directly with estimates reported in column 3 against the model with estimates reported in column 8 since neither model nests the other as a special case.

23. Of course, a model in which past participation affects current participation requires retrospective data on participation.

24. To capture this model, a random-coefficient simultaneous equation multivariate probit model is required. Development of such a model is deferred to another occasion.

25. In this example, if the utility function is additive $d^*(1)$ would have no effect on future choices except through its effect on current resources $M - a(1)d^*(1)$. Thus a test of structural state dependence in this model is a test of intertemporal independence in preferences.

26. Another model that generates structural state dependence in an environment of perfect certainty is a model with fixed costs. In some dynamic models of labor supply, training costs are assumed to be incurred by labor force entrants. Once these costs are incurred, they are not incurred again until reentry occurs. Labor force participation decisions taken by labor force participants take account of such costs. In this way structural state dependence is generated.

27. If the uncertainty comes in the form of price uncertainty, ex ante prices are required to perform the test.

28. This condition implicitly assumes that the disturbances $\bar{\varepsilon}(i, t)$ are stationary. If they are not, and variances are normalized relative to first-period disturbances as in the text, the factor of proportionality is the standard deviation of the first-period disturbance.

29. The most direct way to verify this statement is to substitute equations A2–A4 into equation A1. Assuming that the necessary conditions for identifiability are sufficient, from the exogenous variables unique to $Z_W(i, t)$ and $Z_H(i, t)$ one can estimate δ_W and δ_H (up to a common factor of proportionality), since direct estimates of γ_W and γ_H can be achieved from (A2) and (A3). The factor of proportionality is the standard deviation of $\varepsilon(i, t)$.

30. The model that is estimated is

(3) $$V(i, t) = Z(i, t)\beta + \delta \sum_{t > t'} d(i, t') + \varepsilon(i, t) \text{ for}$$

$i = 1, \ldots, I, \text{ t} = 1, \ldots, T$ and

$d(i, t) = 1 \text{ iff } V(i, t) > 0$

$d(i, t) = 0$ otherwise

(A1) $$W(i, t) = Z_W(i, t)\gamma_W + \eta_W \sum_{t > t'} d(i, t')$$

$$+ U_1(i, t) \text{ for } i = 1, \ldots, I, t = 1, \ldots, T$$

where $\varepsilon(i, t) = \phi_1(i) + V_1(i, t)$ and $U_1(i, t) = \phi_2(i) + V_2(i, t)$. $\phi_1(i)$ and $\phi_2(i)$ are permanent components, and $V_1(i, t)$ and $V_2(i, t)$ are mean zero, jointly normally distributed, temporally independent components with variances σ_{11} and σ_{22}, respectively, and contemporaneous covariance σ_{12}.

The model is estimated treating $\phi_1(i)$ and $\phi_2(i)$ as parameters. Statistical properties of this model are discussed in Heckman and MaCurdy (1980). However, an intuitive justification of the procedure may be of interest, and is given here.

The discussion in the text has already demonstrated that consistent estimates of (normalized) β and σ can be achieved. The procedure for estimating the parameters of equation A1 requires further argument because the wage $W(i, t)$ is observed only for working women

(and hence there may be selection bias in estimating equation A1 on a subsample of working women for whom wage data are available) and because $\sum_{t > t'} d(i, t')$ is likely to be correlated with the error term $U_1(i, t)$ because previous wage rates determine previous labor force participation.

The likelihood function conditions the distribution of $V_2(i, t)$ on the fixed effects $\phi_1(i)$, $\phi_2(i)$ and the event $d(i, t) = 1$ (the condition required to be able to observe a wage rate for woman i at time t). The conditional distribution of $V_2(i, t)$ *given* $\phi_1(i)$, $\phi_2(i)$, and $d(i, t) = 1$ is

(A5)
$$g(V_2(i, t) \mid \phi_1(i), \phi_2(i), d(i, t) = 1) = \frac{\int_{-\infty}^{-[\phi_1(i) + Z(i, t)\beta + \delta \sum_{t > t'} d(i, t')]} f(V_1(i, t), V_2(i, t)) dV_1(i, t)}{\int_{-\infty}^{-[\phi_1(i) + Z(i, t)\beta + \delta \sum_{t > t'} d(i, t')]} f(V_1(i, t)) dV_1(i, t)}$$

where $f[V_1(i, t), V_2(i, t)]$ is the joint distribution of $V_1(i, t)$ and $V_2(i, t)$, and $f_1[V_1(i, t)]$ is the marginal distribution of $V_1(i, t)$. The mean of this conditional distribution is $\lambda(i, t)$.

Given $\lambda(i, t)$, the conditional mean of $W(i, t)$ given $Z_W(i, t)$, $\sum_{t > t'} d(i, t')$, $\phi_1(i)$, $\phi_2(i)$, and $d(i, t) = 1$, is

$$E[W(i, t) \mid Z_W(i, t), \sum_{t > t'} d(i, t'), \phi_1(i), \phi_2(i), d(i, t) = 1] =$$
$$Z_W(i, t)\gamma_W + \eta_W \sum_{t > t'} d(i, t') + \phi_2(i) + \lambda(i, t)$$

The difference between the conditional mean of $W(i, t)$ and $W(i, t)$, $\omega(i, t)$, is uncorrelated with $\sum_{t > t'} d(i, t')$ and $\phi_2(i)$. Thus, given $\lambda(i, t)$ and $\phi_2(i)$, consistent estimates of γ_W and η_W can be achieved by a conventional fixed effect regression of $W(i, t)$ on $Z_W(i, t)$, $\sum_{t > t'} d(i, t')$, and $\lambda(i, t)$. Estimates of the (normalized) parameters of equation 3 can be used to estimate $\lambda(i, t)$. It is thus possible to enter $\lambda(i, t)$ as a regressor in the wage equation, and hence it is possible to secure consistent estimates of the parameters of equation A2 using regression methods.

The likelihood approach used by Heckman and MaCurdy (1980) essentially corrects for selection bias and simultaneous equation bias by adjusting the conditional distribution of $V_2(i, t)$ to account for selectivity and simultaneity. Their estimation procedure is more efficient than the two-step estimator discussed in this note because it incorporates all available prior information. The essential principle underlying the two methods is the same.

An advantage of the fixed effect procedure is that it provides a more satisfactory solution to the problem of "initial conditions" or the problem of endogeneity of presample experience than the procedure used in the text of the paper (see Heckman 1981b).

References

Astrom, K. *Introduction to Stochastic Control Theory.* New York: Academic Press, 1970.

Atkinson, R.; Bower, G.; and Crothers, E. *An Introduction to Mathematical Learning Theory.* New York: Wiley, 1965.

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

Bates, G., and Neyman, J. "Contributions to the Theory of Accident Proneness II: True or False Contagion." *University of California Publications in Statistics* 1 (1951): 215–53.

Ben Porath, Y. "Labor Force Participation Rates and the Supply of Labor." *Journal of Political Economy* 81 (May/June 1973): 697–704.

Bergstrom, A. R. *Statistical Inference in Continuous Time Economic Models.* North Holland, 1976.

Bush, R., and Mosteller, F. *Stochastic Models for Learning.* New York: Wiley, 1955.

Cain, G., *Labor Force Participation of Married Women.* Chicago: University of Chicago Press, 1966.

———. "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey." *Journal of Economic Literature* 14 (December 1976): 1215–57.

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

Cripps, T., and Tarling, R. "An Analysis of the Duration of Male Unemployment in Great Britain, 1932-1973." *Economic Journal* 84 (June 1974): 289–316.

Domencich, T., and McFadden, D. *Urban Travel Demand: A Behavioral Analysis.* Amsterdam: North Holland, 1975.

Dreyfus, S. *Dynamic Programming and the Calculus of Variations.* New York: Academic Press, 1965.

Feller, W. "On a General Class of Contagious Distributions." *Annals of Mathematical Statistics* 14 (May 1943): 389–400.

Flinn, C., and Heckman, J. "Models for the Analysis of Labor Force Dynamics," in R. Basmann and G. Rhodes, eds., *Advances in Econometrics.* JAI Press, 1981.

Gallant, A. R. "Testing a Subset of Parameters of Nonlinear Regression Models." *Journal of the American Statistical Association* 70 (December 1975): 927–32.

Goodman, L. "Statistical Methods for the Mover-Stayer Model." *Journal of the American Statistical Association* 56 (1961): 841–68.

Griliches, Z. "Distributed Lags: A Survey." *Econometrica*, vol. 35 (1967).

Heckman, J. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (1978a) 931–59.

———. "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." *Annales de l'Insee* (Paris), nos. 30–31 (1978b), pp. 227–70.

———. "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, eds., *The Structural Analysis of Discrete Data.* Cambridge: MIT Press, 1981a.

———. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in C. Manski and D. McFadden, eds., *The Structural Analysis of Discrete Data.* Cambridge: MIT Press, 1981b.

Heckman, J., and Borjas, G. "Does Employment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* (1980): 247–83.

Heckman, J., and MaCurdy, T. "Estimation of a Simultaneous Equation Linear Probability Model." Unpublished Ms., University of Chicago, 1978.

————. "A Dynamic Model of Female Labor Supply." *Review of Economic Studies*, January, 1980.

Heckman, J., and Willis, R. "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women." *Journal of Political Economy* 85 (February 1977): 27–58.

Hotelling, H. "Demand Functions with Limited Budgets." *Econometrica* 3 (1935): 66–78.

Johnson, N., and Kotz, S. *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory.* New York: Wiley and Sons, 1977.

Jovanovic, B. "State Dependence in a Continuous Time Stochastic Model of Worker Behavior." Mimeograph, Columbia University, New York, 1978.

Karlin, S., and Taylor, H. *A First Course in Stochastic Processes.* 2d ed. New York: Academic Press, 1975.

Koopmans, L. H. *The Spectral Analysis of Time Series.* New York: Academic Press, 1974.

Layton, L. "Unemployment Over the Work History." Ph.D. dissertation, Department of Economics, Columbia University, New York, 1978.

Lucas, R., and Rapping, L. "Real Wages, Employment and Inflation," in E. Phelps, ed., *Microfoundations of Employment and Inflation Theory.* New York: W. W. Norton, 1970.

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics.* New York: Academic Press, 1973a.

————. "Analysis of Qualitative Choice Behavior," in Zarembka, ed., *Frontiers in Econometrics,* 1973b.

————. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5 (December 1976): 363–90.

Malinvaud, E. *Statistical Methods of Econometrics.* 2d ed. Amsterdam: North-Holland, 1970.

Mincer, J. "Labor Force Participation of Married Women," in H. G. Lewis, ed., *Aspects of Labor Economics.* Princeton: Princeton University Press, 1962.

Morgan, J., et al. *Five Thousand American Families: Patterns of Economic Progress.* Vols. 1 and 2. Ann Arbor: ISR Press, 1974.

Nerlove, M. "Econometric Analysis of Longitudinal Data," in *Annales de l'Insee* (Paris), nos. 30–31 (1978).

Phelps, E. *Inflation Policy and Unemployment Theory*. New York: Norton, 1972.

Pollak, R. "Conditional Demand Functions and Consumption Theory." *Quarterly Journal of Economics* 83 (April 1969): 209–27.

Rao, C. R. *Linear Statistical.Inference and Its Applications*. 2d ed. New York: Wiley, 1973.

Restle, F., and Greeno, J. *Introduction to Mathematical Psychology*. Reading, Mass.: Addison-Wesley, 1970.

Ryder, H.; Stafford, F.; and Stephan, P. "Labor, Leisure and Training over the Life Cycle." *International Economic Review* 17 (October 1976): 651–74.

Samuelson, P. "Structure of Minimum Equilibrium Systems," in R. Pfouts, ed., *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*. Chapel Hill: University of North Carolina Press, 1960.

Sims, C. "Exogeneity and Causal Ordering in Macroeconomic Models," in C. Sims, ed., *New Methods in Business Cycle Research: Proceedings from a Conference*. Minneapolis: Federal Reserve Bank of Minneapolis, 1977.

Singer, B., and Spilerman, S. "Some Methodological Issues in the Analysis of Longitudinal Surveys." *Annals of Economic and Social Measurement* 5 (December 1976): 447–74.

This Page Intentionally Left Blank