econometrics

MDPI

Article

The Specification of Dynamic Discrete-Time Two-State Panel Data Models

Tue Gørgens [1,*] and Dean Robert Hyslop [2,*]

1 Research School of Economics, The Australian National University, Acton ACT 2601, Australia
2 Motu Economic and Public Policy Research, P.O. Box 24390, Wellington 6142, New Zealand
* Correspondence: tue.gorgens@startmail.com (T.G.); dean.hyslop@motu.org.nz (D.R.H.)

Received: 1 November 2018; Accepted: 18 December 2018; Published: 24 December 2018

check for updates

Abstract: This paper compares two approaches to analyzing longitudinal discrete-time binary outcomes. Dynamic binary response models focus on state occupancy and typically specify low-order Markovian state dependence. Multi-spell duration models focus on transitions between states and typically allow for state-specific duration dependence. We show that the former implicitly impose strong and testable restrictions on the transition probabilities. In a case study of poverty transitions, we show that these restrictions are severely rejected against the more flexible multi-spell duration models.

Keywords: panel data; transition data; binary response; duration analysis; event history analysis; dynamic models; censored data; initial conditions; random effects

JEL Classification: C33; C35; C41; C51

1. Introduction

This paper is about modeling discrete-time two-state panel data, where outcomes indicate which of two states an individual is occupying in each period, and where transitions between states occur between periods. Often individuals' outcomes are characterized by a degree of persistence, and an important part of the analysis is to discover the extent to which persistence is due to heterogeneity across individuals or to true state dependence (e.g., Heckman 1978, 1981c; Heckman and Borjas 1980). Analysis of such data is central to many empirical studies in economics and other social sciences.[1]

There are two conceptually distinct approaches to analyzing such two-state panel data in the literature. First, dynamic binary response (DBR) approaches focus on the probability of occupying one of the two states in each period, and usually assume Markovian state dependence, in which the current period's occupancy depends on the occupancy of previous periods. In contrast, multi-spell duration (MSD) approaches focus on the probability of a transition between states occurring in each period, and usually assume current spell duration dependence (or semi-Markovian state dependence), in which the transition probability depends on the elapsed duration in the current state.

This paper compares the DBR and MSD approaches. In the first part of the paper, we present prototype DBR and MSD models and discuss the relationship between them. The DBR models are simpler and more restrictive than the MSD models, involving fewer equations and parameters, and require less data and information about past outcomes. We show that typical DBR models embody

[1] Typical topics include employment (e.g., Heckman 1981a; Hyslop 1999), unemployment (e.g., Arulampalam et al. 2000), poverty (e.g., Stevens 1999; Cappellari and Jenkins 2004), welfare dependency (e.g., Bane and Ellwood 1983), health (e.g., Halliday 2008), and peace and conflict between national states (e.g., Beck and Katz 1997; Beck et al. 2001).

strong restrictions on the corresponding probabilities of transitioning between states (the hazard rates). First, they restrict the effects of observed and unobserved heterogeneity to have the same magnitude but opposite signs on the implied transition probabilities. Second, rth order DBR models restrict the implied transition probabilities to be constant after r periods into a spell. In fact, first and second-order DBR models are special cases of a particularly simple MSD model, where duration dependence is limited to one or two periods.

In the second part of the paper, we use an empirical case study to illustrate the two approaches. We analyze data from the US Panel Study of Income Dynamics (PSID) on individual poverty experiences, previously analyzed by Stevens (1999) using an MSD approach. We fit a range of DBR and MSD model specifications. The estimation results show MSD models dominate the more restrictive DBR models on several dimensions. In particular, the patterns of state dependence in these data are more complicated than allowed for in simple DBR models, and the restriction of opposite effects of heterogeneity on poverty entry and exit is also strongly rejected. Consequently, the MSD models provide better within-sample predictions than do the DBR models. Consistent with recent literature (e.g., Bhuller et al. 2017), we conclude that the standard dynamic binary response model is unacceptably restrictive in this context.

In practice, simple first- or at most second-order DBR models are far more widely used than MSD models. Our theoretical and empirical results suggest that data analysis can benefit from considering less restrictive models and, in particular, from considering the implications of model specifications on both probabilities of occupying a particular state and probabilities of transitioning between states.

To the best of our knowledge, the relationship between the DBR and MSD models has not been formally recognized, and very few studies have discussed the implications of the DBR models for transition probabilities and spell durations. Cappellari et al. (2007) compared a duration and a Markov model for employment transitions, and Bhuller et al. (2017) analyzed the adequacy of first-order dynamic binary response models against more general models that allow for duration and occurrence dependence. Gørgens and Hyslop (2018) showed that the DBR and MSD approaches are equivalent in a nonparametric context, and that nonparametric DBR models of order $r \leq 2$ are nested within a simple MSD model. Among other things, this implies that either model can be used to estimate both probabilities of occupying a state and probabilities of transitioning between states. The present paper complements that analysis by showing that the equivalence and the nesting property carries over to commonly used parametric model specifications, and by documenting that the DBR restrictions are rejected in an empirical case study.

The paper is organized as follows. Section 2 introduces the two approaches and present prototypical DBR and MSD models. Section 3 presents the empirical analysis. The paper concludes with a discussion in Section 4.

2. Modeling Discrete-Time Two-State Panel Data

In this section we provide context for our analysis, present prototype DBR and MSD models, and discuss how they are related.

2.1. Context, Data, and Likelihood

Our interest in this paper is processes that are well represented in discrete time. In a typical application, time is divided into periods of equal length, an individual occupies one of two states during each period, and transitions between states occur between periods. This framework is particularly well suited for studies where a time scale is determined by convention or by law. For example, in some countries receiving welfare is determined on a weekly or monthly basis. The framework is also applicable when an outcome indicates the state an individual is occupying at a point in time and transitions take place during the period between these observations, provided it is reasonable to assume that at most one transition takes place in each period and that the precise timing of the transition within this period can be ignored. The framework is not suitable for data where an outcome indicates state occupancy at a point in time and multiple transitions are likely between the observation times. For example, an analysis of employment status at the time of an annual interview must deal with unobserved transitions between interviews.

The data available for analysis are indicators of state occupancy, indicators of transition between states, and covariates for N individuals observed over T periods. We assume the sample is an independent random sample from a given population. The indicators of the state occupied by individual i at time t are denoted Y_{it} with $Y_{it} \in \{0,1\}$. The indicators of whether or not individual i makes a transition between times $t-1$ and t are denoted C_{it} with $C_{it} \in \{0,1\}$. The covariates for individual i at time t are denoted X_{it} with $X_{it} \in \mathbb{R}^{\dim(x)}$. The covariate time reference is for modeling purposes, and X_{it} may include contemporaneous values, lags and leads, of underlying variables. Let $H_{it} = (Y_{i1} \ldots, Y_{it})$ denote the outcome history up to time t. Lower case letters with a subscript i represent observed values of the corresponding upper case random variables.

For simplicity, we assume the data constitute a balanced panel, and the data for each individual are both left- and right-censored. For identification and estimation purposes, we also assume that the data censoring is independent of the underlying process.

Following the literature (e.g., Heckman and Singer 1984), we allow for unobserved heterogeneity in the form of random effects in the equations which make up the DBR and MSD models discussed below. Let V_i denote a random vector representing unobserved heterogeneity for individual i. We assume that V_i has a discrete distribution with support $\{\nu_1, \ldots, \nu_K\}$ and probability distribution $\pi_1, \ldots, \pi_K$ with $\sum_{k=1}^{K} \pi_k = 1$. Each ν_k is a vector with as many element as there are equations in the model.

As mentioned, the DBR approach focuses on the probabilities of occupying one of the states in each period given previous history while the MSD approach focuses on the probabilities of making a transition between states in each period. For convenience, define

$$\theta_k^I(x_{i1}) = \mathsf{P}(Y_{i1} = 1 | X_{i1} = x_{i1}, V_i = \nu_k), \tag{1}$$
$$\theta_k^Y(h_{it-1}, x_{it}) = \mathsf{P}(Y_{it} = 1 | H_{it-1} = h_{it-1}, X_{it} = x_{it}, V_i = \nu_k), \quad t = 2, \ldots, T, \tag{2}$$
$$\theta_k^C(h_{it-1}, x_{it}) = \mathsf{P}(C_{it} = 1 | H_{it-1} = h_{it-1}, X_{it} = x_{it}, V_i = \nu_k), \quad t = 2, \ldots, T. \tag{3}$$

Assuming independent sampling and independent censoring, the conditional likelihood function for the outcome data given covariates has the form

$$L^W = \prod_{i=1}^{N} \left\{ \sum_{k=1}^{K} \pi_k \left[\theta_k^I(x_{i1}, \nu_k)^{y_{i1}} [1 - \theta_k^I(x_{i1}, \nu_k)]^{1-y_{i1}} \times \prod_{t=2}^{T} \theta_k^W(h_{it-1}, x_{it})^{w_{it}} [1 - \theta_k^W(h_{it-1}, x_{it})]^{1-w_{it}} \right] \right\}, \tag{4}$$

where $W \in (Y, C)$, and (θ_k^W, w_{it}) is either (θ_k^Y, y_{it}) or (θ_k^C, c_{it}). Typical modeling involves parameterizing the probabilities and estimating the unknown parameters by maximizing this likelihood function.

Gørgens and Hyslop (2018) showed that the DBR and MSD approaches are equivalent in a nonparametric context. In particular, one can transform data on state occupancy, $(Y_{i1}, Y_{i2}, \ldots, Y_{it})$, to data on transitions between states, $(Y_{i1}, C_{i2}, \ldots, C_{it})$, and vice versa, and one can transform probabilities of state occupancy, θ_k^Y, to probabilities of transition between states, θ_k^C, and vice versa. In the present paper, we show that prototype parametric DBR and MSD models are also equivalent (barring initial conditions and left-censoring as explained later).

2.2. Prototype DBR Models

The DBR approach assumes that the probabilities of being in a given state depend on the history only through the r most recent outcomes (Markovian state dependence of order r). Formally, for $r \geq 1$ it is assumed that

$$\theta_k^Y(h_{it-1}, x_{it}) = \mathrm{P}(Y_{it} = 1 | Y_{it-r} = y_{it-r}, \ldots, Y_{it-1} = y_{it-1}, X_{it} = x_{it}, V_i = \nu_k), \quad t = r+1, \ldots, T, \tag{5}$$

where $(y_{it-r}, \ldots, y_{it-1})$ are the r-most recent outcomes in h_{it-1}. Equation (5) is the DBR model's main equation of interest, which we refer to as the "structural" equation. The simplest and most common DBR model used empirically adopts $r = 1$, although $r = 2$ is sometimes used in cases of either higher-frequency and/or longer-period data (e.g., Chay et al. 1999; Card and Hyslop 2005, 2009; Andrén and Andrén 2013).

Equation (5) does not restrict the probability distribution for the initial r outcomes, $(Y_{i1}, \ldots, Y_{ir})$, referred to as the "initial conditions" of the process. When the data are left-censored, there is little interest in the probabilities associated with the initial conditions, but it is important they are dealt with unless they can be considered to be exogenous. Adapting the ideas of Heckman (1981b), we shall model the initial conditions using r "approximate reduced form" equations.

In the empirical case study in Section 3 we consider models with $r = 1$ and $r = 2$, labeled DBR1 and DBR2 respectively. The DBR1 model has two equations:

$$\theta_k^I(x_{i1}) = G(\nu_{k1} + \beta_1' x_{i1}), \tag{6}$$

$$\theta_k^Y(h_{it-1}, x_{it}) = G(\nu_{k2} + \beta_2' x_{it} + \gamma_2 y_{it-1}), \quad t = 2, \ldots, T, \tag{7}$$

where G denotes the logistic function.

In the DBR2 model we allow second-order Markovian state dependence. This model extends the first-order model to include two equations for the first two outcomes, while the structural equation includes two lags of the outcome variable as well as their interaction term. Thus, the DBR2 model has three equations:

$$\theta_k^I(x_{i1}) = G(\nu_{k1} + \beta_1' x_{i1}), \tag{8}$$

$$\theta_k^Y(h_{i1}, x_{i2}) = G(\nu_{k2} + \beta_2' x_{i2} + \gamma_2 y_{i1}), \tag{9}$$

$$\theta_k^Y(h_{it-1}, x_{it}) = G(\nu_{k3} + \beta_3' x_{it} + \gamma_{31} y_{it-1} + \gamma_{32} y_{it-2} + \gamma_{33} y_{it-1} y_{it-2}), \quad t = 3, \ldots, T. \tag{10}$$

The prototype DRB models can be generalized. In addition to allowing for higher-order state dependence, several authors have pointed out the possibility of interacting covariates with the lagged outcome variables (e.g., Heckman 1981c; Barmby 1998; Beck et al. 2001). Such an extension leads to a intermediate specification between the prototypical DBR and MSD models, discussed in Section 2.4 and in the case study in Section 3.

2.3. Prototype MSD Models

The MSD approach assumes that the probabilities of a transition between states depends only on the state currently occupied and the elapsed time spent in the current state (duration dependence, or semi-Markovian state dependence). Let F_i denote the time of the first observed transition for individual i, with $F_i > T$ if no transitions are observed. Also, let D_{it} denote the observed elapsed time in the spell observed in period t: $D_{i1} = 1$ for the left-censored spells, and $D_{iF_i} = 1$ when $F_i \leq T$. Then the prototype MSD model assumes that the transition probabilities for the "fresh" (i.e., non-left-censored) spells that begin during the observation period satisfy

$$\theta_k^C(h_{it-1}, x_{it}) = \mathrm{P}(C_{it} = 1 | Y_{it-1} = y_{it-1}, D_{it-1} = d_{it-1}, X_{it} = x_{it}, V_i = \nu_k), \quad t = f_i + 1, \ldots, T. \tag{11}$$

Conditioning on $Y_{it-1} = y_{it-1}$ captures first-order Markovian state dependence, and conditioning on $D_{it-1} = d_{it-1}$ captures duration dependence. We refer to (11) as the MSD model's "structural" equation of interest.

In practice, the usefulness of assumption (11) will depend on the number of transitions observed within the observation period. In applications with substantial persistence in state occupancy and few transitions, there may be relatively few fresh-spell observations for which Equation (11) applies. This is alleviated if the effect of elapsed duration becomes constant after some time. For this purpose, we consider the additional assumption that the effect of elapsed duration is constant after m periods within the spell. (For example, if $m = 1$, the transition probabilities are constant throughout each spell.) To write this compactly, define $D_{it}^m = \min(D_{it}, m)$. Then, for $m \geq 1$,

$$\theta_k^C(h_{it-1}, x_{it}) = \mathsf{P}(C_{it} = 1 | Y_{it-1} = y_{it-1}, D_{it-1}^m = d_{it-1}^m, X_{it} = x_{it}, V_i = \nu_k), t = \min(f_i, m) + 1, \ldots, T. \quad (12)$$

The main advantage of assumption (12) is that only parameters for times $1, \ldots, m$ depend on unobserved outcome variables. In other words, Equation (12) implies that all data after the earlier of the first observed transition at f_i or time m contribute to identifying and estimating the structural parameters of interest.

Neither (11) nor (12) restrict the probabilities of making a transition out of the left-censored initial spells,

$$\theta_k^C(h_{it-1}, x_{it}), \quad t = 1, \ldots, \min(f_i, m). \quad (13)$$

As in the DBR case, these are typically nuisance parameters, and again we use Heckman's (1981b) ideas in modeling them. A fully flexible specification of the approximate reduced form would involve separate equations for each probability in (13). Depending on the extent of left-censoring, this may be prohibitive in practice. We adopt a more parsimonious approach and specify three equations for the probability of the initial outcome, and the initial spell transitions from each state.

In the empirical study in Section 3 we consider two models which both assume (11) and (12), and capture duration dependence through a flexible specification with separate parameters for the first m potential transition times in each state ($m = 6$). The first model, MSD1, uses all the data available but does not fully exploit Assumption (12). Specifically, the duration dependence is assumed to be constant after duration m, but this restriction is not imposed across the structural equations and probabilities for left-censored spells. This model has five equations: a reduced form equation for the initial state, two reduced-form equations for modeling transitions from the initial spells, and two structural equations for modeling the transitions from the fresh spells. The model specification (for $m > 1$) is

$$\theta_k^I(x_{i1}) = G(\nu_{k1} + \beta_1' x_{i1}), \quad (14)$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k2} + \beta_2' x_{it} + \textstyle\sum_{j=2}^m \lambda_{2j} 1(d_{it-1}^m \geq j)), \quad y_{i1} = 0, \quad t = 2, \ldots, \min(f_i, m), \quad (15)$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k3} + \beta_3' x_{it} + \textstyle\sum_{j=2}^m \lambda_{3j} 1(d_{it-1}^m \geq j)), \quad y_{i1} = 1, \quad t = 2, \ldots, \min(f_i, m), \quad (16)$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k4} + \beta_4' x_{it} + \textstyle\sum_{j=2}^m \lambda_{4j} 1(d_{it-1}^m \geq j)), \quad y_{it-1} = 0, \quad t = \min(f_i, m) + 1, \ldots, T, \quad (17)$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k5} + \beta_5' x_{it} + \textstyle\sum_{j=2}^m \lambda_{5j} 1(d_{it-1}^m \geq j)), \quad y_{it-1} = 1, \quad t = \min(f_i, m) + 1, \ldots, T. \quad (18)$$

The second model, MSD2, utilizes assumption (12) together with a second implication, namely that we can ignore observed transitions in the first m periods, and still obtain consistent estimates of the structural equation parameters of interest. This may entail some loss of precision in estimating the structural parameters, but simplifies modeling the nuisance parameters. Specifically, for $d_{it-1} \geq m$ we restrict Equations (15) and (17) to be the same and Equations (16) and (18) to be the same. Furthermore, we ignore the likelihood contributions for the first $m - 1$ time periods. This allows us to estimate a three-equation model, which represent the approximate reduced form specification for the probability distribution of the state at time m, and the two transition probabilities out of the state-specific spells. The MSD2 model specification (for $m > 1$) is

$$\theta_k^C(h_{im-1}, x_{im}) = G(\nu_{k1} + \beta_1' x_{im}), \tag{19}$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k2} + \beta_2' x_{it} + \textstyle\sum_{j=2}^{m} \lambda_{2j} 1(d_{it-1}^m \geq j)), \quad y_{it-1} = 0, \quad t = m+1, \ldots, T, \tag{20}$$

$$\theta_k^C(h_{it-1}, x_{it}) = G(\nu_{k3} + \beta_3' x_{it} + \textstyle\sum_{j=2}^{m} \lambda_{3j} 1(d_{it-1}^m \geq j)), \quad y_{it-1} = 1, \quad t = m+1, \ldots, T. \tag{21}$$

The prototype MSD models can be generalized, for example, by allowing occurrence dependence (i.e., the effects of the number of previous transitions), and/or lagged duration dependence (i.e., completed durations of previous spells) (e.g., Heckman and Borjas 1980; Doiron and Gørgens 2008). In practice, the higher data demands mean that estimation of such models may only be feasible when the data are non-left-censored for all individuals.

2.4. The Relationship between DBR and MSD Models

As mentioned, the data representations for occupancies and transitions are equivalent. The probabilities for state occupancy and for transitions between states are also equivalent in a nonparametric context. Furthermore, Gørgens and Hyslop (2018) also showed that the Markov assumption (5) with $r = 1$ or $r = 2$ implies the semi-Markov assumption (11). That is, the low-order DBR model is a special case of the MSD model in nonparametric setting.

The prototype parametric models differ in the details of how they deal with initial conditions and left-censoring. However, in this section we show that both the DBR1 and the DBR2 structural equations are special cases of the structural MSD1/MSD2 equations. That is, the nesting property survives typical parameterizations.

To see this for the DBR1 model, note that by symmetry of the logistic function the DBR1 Equation (7) implies

$$\theta_k^C(h_{it-1}, x_{it}) = \begin{cases} G(\nu_{k2} + \beta_2' x_{it}) & \text{if } y_{it-1} = 0, \\ G(-\nu_{k2} - \beta_2' x_{it} - \gamma_2) & \text{if } y_{it-1} = 1, \end{cases} \quad t = 2, \ldots, T, \tag{22}$$

whereas the MSD1 equations can be written (the MSD2 model is similar)

$$\theta_k^C(h_{it-1}, x_{it}) = \begin{cases} G(\nu_{k4} + \beta_4' x_{it} + \sum_{j=2}^{m} \lambda_{4j} 1(d_{it-1}^m \geq j)) & \text{if } y_{it-1} = 0, \\ G(\nu_{k5} + \beta_5' x_{it} + \sum_{j=2}^{m} \lambda_{5j} 1(d_{it-1}^m \geq j)) & \text{if } y_{it-1} = 1, \end{cases} \quad t = m+1, \ldots, T. \tag{23}$$

Matching coefficients shows that the DBR1 model arises as a special case of the MSD1 models when there is no duration dependence ($m = 1$), and the effects of observed and unobserved heterogeneity are opposite in the two transition probabilities. Specifically, the restrictions are $\lambda_{4j} = 0$ and $\lambda_{5j} = 0$ for $j = 2, \ldots, m$; $\beta_4 + \beta_5 = 0$; and $\nu_{k4} - \nu_{14} + \nu_{k5} - \nu_{15} = 0$ for $k = 2, \ldots, K$. Under these restrictions, the state dependence parameter in the DBR1 model can be recovered from $\gamma_2 = -\nu_{14} - \nu_{15}$, with $\nu_{k2} = \nu_{k4}$.

To show that the DBR2 model is also a special case, note that the DBR2 structural Equation (10) implies

$$\theta_k^C(h_{it-1}, x_{it}) = \begin{cases} G(\nu_{k3} + \beta_3' x_{it} + \gamma_{32}) & \text{if } y_{it-1} = 0, y_{it-2} = 1, \\ G(\nu_{k3} + \beta_3' x_{it}) & \text{if } y_{it-1} = 0, y_{it-2} = 0, \\ G(-\nu_{k3} - \beta_3' x_{it} - \gamma_{31}) & \text{if } y_{it-1} = 1, y_{it-2} = 0, \\ G(-\nu_{k3} - \beta_3' x_{it} - \gamma_{31} - \gamma_{32} - \gamma_{33}) & \text{if } y_{it-1} = 1, y_{it-2} = 1, \end{cases} \quad t = 3, \ldots, T. \tag{24}$$

Notice that unequal values of y_{it-1} and y_{it-2} means that the spell has lasted exactly one period at time t, while equal values means that the spell has lasted two or more periods. Since $d^m_{it-1} \geq 2$ if and only if $y_{it-1} = y_{it-2}$, the MSD1 structural equations for $m \geq 2$ can be re-written (the MSD2 model is similar)

$$\theta^C_k(h_{it-1}, x_{it}) = \begin{cases} G(\nu_{k4} + \beta_4' x_{it} + \lambda_{42}(1 - y_{it-2}) + \sum_{j=3}^m \lambda_{4j} 1(d^m_{it-1} \geq j)) & \text{if } y_{it-1} = 0, \\ G(\nu_{k5} + \beta_5' x_{it} + \lambda_{52} y_{it-2} + \sum_{j=3}^m \lambda_{5j} 1(d^m_{it-1} \geq j)) & \text{if } y_{it-1} = 1, \end{cases} \quad t = m+1, \ldots, T. \quad (25)$$

Matching coefficients shows that the DBR2 model arises as a special case when there is no duration dependence after one period ($m = 2$), and the effects of observed and unobserved heterogeneity are opposite in the two transition probabilities. Specifically, the restrictions are $\lambda_{4j} = 0$ and $\lambda_{5j} = 0$ for $j = 3, \ldots, m$; $\beta_4 + \beta_5 = 0$; and $\nu_{k4} - \nu_{14} + \nu_{k5} - \nu_{15} = 0$ for $k = 2, \ldots, K$. Under there restrictions, the state dependence parameters in the DBR1 model can be recovered from $\gamma_{31} = -\nu_{14} - \nu_{15} - \lambda_{42}$, $\gamma_{32} = -\lambda_{42}$, and $\gamma_{33} = \lambda_{42} - \lambda_{52}$, with $\nu_{k3} = \nu_{k4} + \lambda_{42}$.

Higher-order DBR models are generally not nested within the MSD framework. However, it is easy to show that a rth-order DBR model has transition probabilities that are constant after r periods in a spell. Moreover, prototype DBR models implicitly assume that the effects of covariates and unobserved heterogeneity on the probability of being in state 1 at time t are the same whether or not the individual is in state 0 or state 1 at time $t-1$. In other words, the effects on the entry and exit transition probabilities have the same magnitude but opposite signs. From the perspective of duration analysis, the restrictions on the transition probabilities embodied in typical DBR models appear strong.

Several authors working with DBR models have considered the possibility of interacting the covariates with the lagged occupancy indicator (e.g., Heckman 1981c; Barmby 1998; Beck et al. 2001; Card and Hyslop 2009; Browning and Carro 2010; Cappellari and Jenkins 2014). Although the intent was to allow for heterogeneity in state dependence rather than a specific consideration of transition probabilities, Equation (25) shows that the extension to permit $\beta_4 \neq -\beta_5$ is an intermediate case with the restrictions on duration dependence and the effects of unobserved heterogeneity maintained.

The nesting of DBR1 and DBR2 models within the MSD model means that it is possible to test the former against the latter using a simple Wald statistic. Obviously, it is also possible to test intermediate models against the MSD model to investigate separately the validity of the restrictions on duration dependence and on the heterogeneity effects. In the case study below, we investigate whether rejection of the DBR models is mainly due to inadequate modeling of duration dependence or to restricting the heterogeneity effects to be opposite by estimating and testing a hierarchy of model specifications.

3. Case Study

We now apply each of the methods discussed above to an analysis of poverty persistence, previously analyzed using multi-spell duration models by Stevens (1999). Our focus here is to estimate and compare the DBR and MSD models, rather than to replicate or critique Stevens' original analysis. So, for example, we select a different analytical extract from the data provided to us than that used by Stevens.

3.1. Data

Our analytical sample consists of a balanced panel of 5248 individuals over the 20 years 1970–89 from the Panel Study of Income Dynamics (PSID).[2] Each individual's poverty status is determined by

[2] The data we use come from the PSID survey years 1970–89, with the income and poverty observations corresponding to calendar years 1969–88. Years mentioned in the text refer to survey years. We use a balanced panel in order to abstract from attrition issues that may differentially affect the estimation methods.

whether their family's annual income is below or above a needs threshold which depends on family size and composition, so that all individuals in a family have the same poverty status in that year.[3] The main sample selection criteria we apply is that all individuals experience at least one year in poverty over the extended period 1968–89, and are observed and have no missing outcome or covariate information over the analysis period 1970–89.[4] The covariates include dummy variables for age groups 0–5, 6–17, 18–24, and 55+, and dummy variables for whether the household head is female and/or black. The sample characteristics are summarized in Table 1.

Table 1. Descriptive statistics.

	Mean (Standard Error)
Person-years	
Aged 0–5	0.025 (0.0005)
Aged 6–17	0.225 (0.001)
Aged 18–24	0.204 (0.001)
Aged 25–54	0.420 (0.002)
Aged 55+	0.126 (0.001)
Female head	0.336 (0.001)
Black head	0.582 (0.002)
Poor (y_{it})	0.353 (0.001)
Transition (c_{it})	0.177 (0.001)
No. person-years	104,960
Persons	
Transitions	3.35 (0.032)
No. persons	5248
Spells	
Duration of all spells	4.59 (0.033)
Duration of initial spells	7.11 (0.084)
Duration of fresh spells	3.84 (0.032)
No. spells	22,849

Notes: No adjustments for censoring.

3.2. Estimation Results

We present the estimates of the DBR and MSD models in Tables 2 and 3. Table 2 contains estimates for the first- and second-order DBR models with two discrete points of unobserved heterogeneity, DBR1 and DBR2. Similarly, Table 3 contains the estimates for two MSD models with two random effects mass points: a five-equation model, MSD1, estimated using the full sample; and a three-equation model, MSD2, which exploits the assumption of constant transition probabilities after 6 years to estimate common equations for initial and fresh spells, excluding the first five years of initial spells. As the data do not constitute a random sample, we report robust standard errors with clustering at the level of the households originally selected for the survey. Note that, although the initial state equation is the same in each model, since each model is jointly estimated as a system of different equations, the initial state equation parameter estimates will vary across the models.

The MSD models relax restrictions implied by the DBR models in two important respects. First, the DBR model specifications imply that covariate coefficients should be equal in magnitude and opposite in sign in the entry and exit equations. In contrast, although the covariate coefficients are predominantly positive in the entry equation and negative in the exit equation (in line with the DBR

3 See Stevens (1999) for more details of this and other data issues.
4 This criteria is used as a proxy to identify the poverty at-risk population, and follows Stevens (1999).

estimates), we find that there are some exceptions with the signs of the young and old age coefficients. In addition, there are also substantial differences in magnitudes of the coefficients in these equations.

Table 2. Dynamic binary response model estimates.

	DBR1		**DBR2**		
	IC	**Strl**	**IC1**	**IC2**	**Strl**
Variables					
y_{it-1}		2.191 (0.040)		2.426 (0.147)	1.859 (0.051)
y_{it-2}					0.942 (0.050)
$y_{it-1}y_{it-2}$					0.039 (0.075)
Aged 0–5	0.253 (0.104)	0.549 (0.093)	0.218 (0.100)	0.163 (0.136)	0.328 (0.104)
Aged 6–17	0.601 (0.075)	0.560 (0.043)	0.558 (0.073)	0.165 (0.077)	0.449 (0.037)
Aged 18–24	0.397 (0.113)	0.186 (0.030)	0.349 (0.111)	−0.13 (0.126)	0.109 (0.029)
Aged 55+	−0.131 (0.172)	0.272 (0.048)	−0.105 (0.170)	−0.236 (0.180)	0.288 (0.043)
Female Head	1.076 (0.139)	0.935 (0.047)	1.085(0.138)	0.744 (0.156)	0.874 (0.045)
Black Head	1.429 (0.145)	0.620 (0.055)	1.48 (0.144)	0.855 (0.155)	0.527 (0.048)
Random effects (mass points and probabilities)					
ν_1	−2.178 (0.157)	−3.186 (0.057)	−2.121 (0.163)	−2.686 (0.178)	−3.206 (0.057)
ν_2	−0.654 (0.150)	−1.604 (0.073)	−0.767 (0.155)	−1.607 (0.171)	−1.926 (0.076)
π_1	0.638 (0.023)		0.640 (0.032)		
Statistics					
No. persons		5248			5248
No. years		20			20
Log QL		−45,060.2			−44,110.7

Notes: y_{it} indicates poverty in year t; IC: initial conditions equation; Strl: structural equation; Log QL: logarithm of quasi-likelihood value. Standard errors in parentheses (clustered at the original 1968 household level).

Table 3. Multi-spell duration model estimates.

	MSD1					**MSD2**		
	Initial	**Initial Spells**		**Fresh Spells**		**Initial**	**Fresh Spells**	
	State	**Entry**	**Exit**	**Entry**	**Exit**	**State**	**Entry**	**Exit**
Variables								
$1(d_{it} \geq 2)$		−0.070	−0.425	−0.507	−0.562		−0.509	−0.544
		(0.161)	(0.168)	(0.070)	(0.068)		(0.075)	(0.069)
$1(d_{it} \geq 3)$		−0.490	0.234	−0.365	−0.188		−0.373	−0.163
		(0.185)	(0.191)	(0.093)	(0.094)		(0.096)	(0.097)
$1(d_{it} \geq 4)$		0.007	−0.046	−0.186	−0.290		−0.228	−0.321
		(0.192)	(0.202)	(0.115)	(0.119)		(0.114)	(0.120)
$1(d_{it} \geq 5)$		0.003	−0.028	−0.051	−0.169		−0.068	−0.258
		(0.209)	(0.236)	(0.130)	(0.142)		(0.117)	(0.135)
$1(d_{it} \geq 6)$		0.257	−0.051	−0.309	−0.056		0.029	−0.055
		(0.162)	(0.203)	(0.113)	(0.136)		(0.091)	(0.114)
Aged 0–5	0.156	0.332	−0.068	−0.340	−0.049	0.036	−0.206	−1.124
	(0.100)	(0.121)	(0.120)	(0.200)	(0.226)	(0.168)	(0.429)	(0.646)
Aged 6–17	0.512	−0.098	−0.440	0.511	−0.188	0.497	0.431	−0.272
	(.071)	(.063)	(.061)	(.059)	(.057)	(.069)	(.053)	(.055)
Aged 18–24	0.314	0.513	0.401	0.111	0.169	0.073	0.310	0.197
	(0.111)	(0.060)	(0.067)	(0.044)	(0.044)	(0.100)	(0.040)	(0.040)
Aged 55+	−0.078	0.047	−0.289	0.366	−0.287	0.150	0.327	−0.289
	(0.170)	(0.080)	(0.157)	(0.061)	(0.060)	(0.153)	(0.052)	(0.060)

Table 3. *Cont.*

	MSD1					MSD2		
	Initial	**Initial Spells**		**Fresh Spells**		**Initial**	**Fresh Spells**	
	State	**Entry**	**Exit**	**Entry**	**Exit**	**State**	**Entry**	**Exit**
Female Head	1.086	0.906	−0.732	0.881	−0.817	1.195	0.838	−0.802
	(0.140)	(0.087)	(0.121)	(0.067)	(0.065)	(0.135)	(0.058)	(0.066)
Black Head	1.529	0.246	−0.871	0.713	−0.498	1.379	0.407	−0.493
	(0.142)	(0.084)	(0.118)	(0.072)	(0.063)	(0.144)	(0.059)	(0.063)
Random effects (mass points and probabilities)								
ν_1	−2.148	−2.403	0.332	−2.610	1.161	−2.592	−2.154	1.003
	(0.181)	(0.126)	(0.193)	(0.102)	(0.094)	(0.179)	(0.099)	(0.109)
ν_2	−0.901	−1.678	−0.615	−1.143	0.121	−1.180	−0.870	0.029
	(0.147)	(0.134)	(0.147)	(0.093)	(0.076)	(0.180)	(0.130)	(0.088)
π_1	0.590					0.677		
	(0.042)					(0.049)		
Statistics								
No. persons					5248			5248
No. years					20			16
Log QL					−43,444.5			−34,621.8

Notes: d_{it} indicates elapsed duration in current spell at the end of year t; Entry: structural equation for entering poverty; Exit: structural equation for exiting poverty; Log QL: logarithm of quasi-likelihood value. Standard errors in parentheses (clustered at the original 1968 household level).

Second, the order of state dependence in DBR models implies that the impact on the probability of a transition occurring should be zero for spell-durations longer than that order. That the estimates of the elapsed-duration variables in both the entry and exit equations are statistically significant up to 6-plus years, strongly rejects both the first- and second-order state dependence specifications in the estimated DBR models. Also, all of the coefficients are negative, which implies that the probability of a transition either into or out of poverty occurring declines with the duration of the spell. However, the coefficient magnitudes vary across the entry and exit equations.

This suggests, perhaps not surprisingly, that the MSD model provides a substantially better fit to the data than the DBR models. This is true in terms of the overall fit of the model; and also in terms of the more specific heterogeneity and duration-dependence restrictions implied by the DBR models.

To explore the respective contributions of relaxing the DBR models' strict state dependence and heterogeneity restrictions, we estimated a variety of model specifications that vary between the extremes of the DBR1 model (A) and the unrestricted MSD1 model (G). The intermediate model specifications are as follows: model (B) relaxes the heterogeneity restriction on both the covariates and unobserved heterogeneity in the DBR1 model; model (C) is the DRB2 model which relaxes the first-order Markov assumption, but imposes the entry and exit heterogeneity restrictions; model (D) relaxes the heterogeneity restrictions in the DBR2 model; and models (E) and (F) allow for further duration dependence ($m = 6$), with and without the heterogeneity effect restrictions in the entry and exit equations. Assumption (12) is fully imposed in these models; that is, for $t \geq m$ the transition probabilities out of the left-censored initial spells are modeled by the structural equations.[5]

The results are summarized in Table 4, including Wald tests of the hypotheses associated with the various sets of restrictions. Each of the hypotheses is severely rejected, implying both the Markovian state dependence and the heterogeneity restrictions on entry and exit are rejected against the MSD

5 Models (F) and (G) are conceptually the same, except that assumption (12) is not fully exploited in model (G), and the initial spells continue to be modeled separately from the structural spells even after m periods. Because model (G) does not nest the other specifications, a Wald test is not possible; however, the improvement in the log quasi-likelihood value from model (F) to (G) is huge, suggesting assumption (12) is also problematic.

alternatives in all cases. The most significant gain appears to be associated with relaxing the first-order Markov assumption to allow second-order effects.

Table 4. Differences between the DBR1 and MSD1 models.

Model	Duration *p* Dependence	Heterogeneity	Log QL	No. Parms	H_0/H_A	Wald Statistic	df
A (DBR1)	1	Opposite	−45,060.2	18			
B	1	Flexible	−44,968.7	25	A / B	118.6	7
C (DBR2)	2	Opposite	−44,110.7	29	A / C	817.5	11
D	2	Flexible	−44,003.7	43	C / D	167.2	14
E	6	Opposite	−43,832.9	45	C / E	184.9	16
F	6	Flexible	−43,709.2	59	E / F	199.8	14
G (MSD1)	6	Flexible	−43,444.5	61			

Notes: p: constant hazard rates from period p; Log QL: logarithm of quasi-likelihood value; No. parms: number of parameters; df: degrees of freedom; Opposite: parameters in entry and exit equations have opposite signs. For all models $K = 2$ and $N = 104,960$.

3.3. In Sample Prediction

We also compare in-sample predictions of the estimated DBR and MSD models against actual outcome, using alternative measures. Table 5 uses two-way frequency tables to summarize the number of years individuals are poor and the number of transitions in- and out-of poverty they experience; while the frequency distributions of the number of spells individuals experience is summarized in Table 6, and the average durations of those spells in Table 7.

The first panel of Table 5 summarizes the actual poverty experience of individuals over the 20 year observation period. This shows substantial variation in poverty experience in the sample around the 7.06 year average: about 5 percent of the sample had no spells of poverty, and 18 percent had a single year poor, while at the other extreme, 3 percent of individuals were always in poverty, and the remaining three-quarters experienced between 2 and 19 years of poverty. Similarly, there was substantial heterogeneity in poverty transitions around the 3.35 average, with 8 percent of individuals experiencing no transitions, while over 40 percent had at least 4 transitions.

The next three panels in Table 5 present analogous summaries of the predictions from the first- and second-order DBR models, and the MSD1 model respectively. Each model closely predicts both the average incidence of poverty (between 7.00 and 7.03 years), and the average number of transitions (between 3.36 and 3.41 transitions). However, they each overpredict the frequency of zero transitions, and underpredict the frequencies of 1 and 2 transitions, and also underpredict the frequency of persistent poverty (more the 15 years). The prediction errors are lower for the MSD model than the DBR models (and for the DBR2 compared to DBR1), consistent with the estimation results discussed above.

We have constructed Pearson goodness-of-fit statistics for each of the models based on the tables of actual and predicted frequencies in Table 5.[6] The relative magnitudes of these statistics are consistent with the MSD1 model fitting substantially better than the two DBR models (and the DBR2 model fits better than the DBR1 model). However, the value of MSD1 model's goodness-of-fit statistic (98.9 with 23 degrees of freedom) suggests that it also does not provide an adequate statistical fit to the data using conventional significance levels.

[6] These statistics are intended for indicative purposes, as their distribution is unclear, and using critical values from a chi-square distribution with "df" degrees of freedom is likely to result in a conservative test (under-rejection). For models estimated by maximizing the complete likelihood function, Chernoff and Lehmann (1954) and Moore (1977) among others show that the critical value is somewhere between chi-squares with q and $q - l$ degrees of freedom, where q is the number of free terms in the test statistic and l is the number of estimated parameters. Andrews (1988) extends this to non-dynamic models estimated by maximizing the conditional likelihood function given covariates. However, these results do not apply to dynamic models estimated by maximizing a quasi-likelihood function using clustered samples. For convenience, we report the "maximum degrees of freedom" (i.e., q).

Table 5. Predictions of years in poverty and transitions.

No. Years Poor	No. Transitions 0	1	2	3	4+ Even	5+ Odd	Total
Actual data							
0	255	0	0	0	0	0	255
1	0	201	735	0	0	0	936
2–5	0	232	315	291	526	168	1532
6–10	0	88	68	209	325	322	1012
11–15	0	55	38	95	299	290	777
16–19	0	88	155	92	190	62	587
20	149	0	0	0	0	0	149
Total	404	664	1311	687	1340	842	5248
DBR1 predictions							
0	473.2	0	0	0	0	0	473.2
1	0	123.3	388.6	0	0	0	511.8
2–5	0	131.4	320.1	369.0	580.0	205.1	1605.5
6–10	0	31.0	62.4	194.6	475.5	458.6	1221.9
11–15	0	23.0	53.7	131.8	365.8	274.0	848.2
16–19	0	49.9	179.2	79.1	172.2	37.4	517.7
20	69.9	0	0	0	0	0	69.9
Total	543.0	358.5	1003.8	774.4	1593.4	975.0	5248.0
GOF = 973.5 (23df)							
DBR2 predictions							
0	527.5	0	0	0	0	0	527.5
1	0	113.8	440.7	0	0	0	554.5
2–5	0	166.8	228.2	334.0	588.0	212.9	1529.9
6–10	0	68.0	84.0	208.0	406.1	402.6	1168.6
11–15	0	44.4	72.8	136.4	315.5	265.0	834.1
16–19	0	58.6	179.1	76.7	179.6	45.1	539.0
20	94.6	0	0	0	0	0	94.6
Total	622.1	451.5	1004.7	755.0	1489.1	925.6	5248.0
GOF = 613.6 (23df)							
MSD1 predictions							
0	336.4	0	0	0	0	0	336.4
1	0	156.4	625.0	0	0	0	781.4
2–5	0	217.0	318.9	306.5	582.9	167.3	1592.5
6–10	0	105.2	73.4	187.8	341.6	350.5	1058.4
11–15	0	68.5	54.4	121.5	288.2	262.0	794.5
16–19	0	67.6	175.0	75.2	188.1	43.1	548.9
20	136.1	0	0	0	0	0	136.1
Total	472.5	614.5	1246.6	690.9	1400.8	822.8	5248.0
GOF = 98.9 (23df)							

Notes: GOF: Pearson goodness of fit statistic; df: degrees of freedom. Poverty rates and incidence rates can be computed by taking the number of years in poverty and the number of transitions and divide by the number of years under observation.

Table 6 summarizes the actual and predicted frequency distributions of the number of distinct poor and non-poor spells experienced over the observation period.[7] Consistent with overpredicting zero-transitions, both models overpredict the frequency of single spells and underpredict the frequencies of 2 and 3 poverty spells, especially for individuals whose initial state is not-poor.

[7] We exclude the DBR2 model, and show just the DBR1 and MSD1 model predictions here, as the DBR model predictions are comparatively similar.

The Pearson goodness-of-fit statistics for these predictions again indicate the MSD model provides substantially better fit to this characterization of the data than the DBR model.

Finally, Table 7 summarizes the average durations of the actual and predicted spells for the DBR1 and MSD1 models. The averages of the MSD1 model predictions are again closer to the actual spell average durations than those of the DBR1 model.

Table 6. Predictions of spells by initial state.

No. Spells	Actual Data		DBR1 Predictions		MSD1 Predictions	
	Initial State		Initial State		Initial State	
	Not Poor	Poor	Not Poor	Poor	Not Poor	Poor
1	255	149	473.2	69.9	336.4	136.1
2	159	505	99.2	259.3	132.7	481.8
3	1089	222	739.0	264.8	989.7	257.0
4	214	473	230.4	544.0	203.0	487.9
5	455	271	600.0	370.8	518.5	265.3
6	129	398	211.2	458.3	158.7	323.3
7	219	155	261.8	225.5	227.4	175.4
8	81	139	84.4	174.2	79.3	155.6
9	119	73	60.9	61.1	81.8	79.5
10	31	47	14.9	28.5	31.9	51.0
11	21	16	6.2	6.2	21.2	22.2
12	6	10	1.2	2.5	8.2	11.7
13	7	3	0.5	0.5	3.5	4.8
14	0	1	0.0	0.0	1.1	1.9
15	0	1	0.0	0.0	0.4	0.8
16	0	0	0.0	0.0	0.1	0.2
17	0	0	0.0	0.0	0.0	0.1
Total	2785	2463	2782.7	2465.4	2793.7	2454.3
GOF			564.6	481.9	66.1	32.6
(df)			(11)	(11)	(11)	(11)

Notes: GOF: Pearson goodness-of-fit statistic conditional on the initial state, with cells 12–17 combined; df: degrees of freedom.

Table 7. Predictions of spell type and poverty status.

	Actual Data		DBR1 prd		MSD1 prd	
	Initial Spells	Fresh Spells	Initial Spells	Fresh Spells	Initial Spells	Fresh Spells
Status: not poor						
Avg spell duration	8.23	4.85	8.43	4.70	8.10	4.92
No. spells	2785	9277	2783	9510	2794	9266
Status: Poor						
Avg spell duration	5.86	2.72	4.48	2.96	5.88	2.67
No. spells	2463	8324	2465	8684	2454	8368

Notes: Prd: predictions; avg: average.

3.4. Out of Sample Prediction

Finally, to gauge whether policy recommendations are sensitive to the model specification, we examine the results of a simple simulation exercise using the DBR1 and MSD1 models. In this exercise, we consider a hypothetical policy intervention that moves each person out of poverty in a year, and use each model to simulate their subsequent poverty experience and transitions over the following decade. The intervention date is randomly selected within each person's years in poverty, or the first year for the 5 percent of the sample who don't experience poverty over the observation period. Each person's covariates correspond to their characteristics in the intervention year, which are held constant except for subsequent aging.

Table 8 summarizes the results of this exercise. The first row contains the full-sample summary, and subsequent rows summarize various demographic subsamples. The first two columns show the actual first-year exit rates from non-poor (i.e., transitions back into poverty) for all fresh spells, and the person-average respectively.[8] The next two columns show the exit rates associated with the simulations of each of the DBR1 and MSD1 models. The DBR1 model is lower than the person-average exit rate, while the MSD1 model is higher by about the same amount. Both the actual and simulated exit rates vary across subsamples, generally with the DBR1 model lower and the MSD1 model higher than the sample average.[9]

Table 8. Results from policy intervention simulation.

Sample	First-Year Exit Rate from Non-Poor				No. Years Poor Next Decade		No. Transitions Next Decade	
	Data: [a] Spells	Data: [b] Persons	DBR1	MSD1	DBR1	MSD1	DBR1	MSD1
All	0.34	0.25	0.21	0.30	3.19	2.38	1.82	2.86
Female head	0.40	0.28	0.30	0.41	4.58	3.00	2.16	4.01
Black head	0.38	0.30	0.25	0.36	3.91	2.79	2.01	3.52
Aged 0–5	0.24	0.19	0.24	0.20	3.81	2.32	1.99	2.83
Aged 6–17	0.37	0.24	0.26	0.37	3.86	2.79	2.00	3.35
Aged 18–24	0.29	0.24	0.20	0.29	3.07	2.31	1.80	2.71
Aged 25–54	0.34	0.25	0.16	0.25	2.72	2.12	1.69	2.52
Aged 55+	0.39	0.31	0.20	0.32	3.37	2.50	1.86	3.19

Notes: The policy intervention is to move each person out of poverty in a year (randomly selected within their poverty years); for the (255) people who have no poverty spell, they are selected in the first year. The first-year exit rate is calculated as the fraction who re-enter poverty immediately after the first year. [a] Average across all (9277) new non-poor spells; [b] average (person-average first-year exit rate) across all (4993) persons who have a new non-poor spell.

In the next pair of columns, we present the average number of years poor over the 10 year simulation time frame predicted by the models, and in the final pair of columns we present the average number of transitions predicted by each model. Perhaps surprisingly given the higher first-year exit rate back into poverty, the MSD1 model predicts about 25 percent lower poverty experience over the following 10 years than the DBR1 model (2.4 versus 3.2 years) on average, explained by over 50 percent more transitions (2.9 versus 1.8). Again, the relative differences in the models' predictions are broadly similar across the various subsamples.

The results from this simple experiment indicate that the policy predictions provided by the alternative models are different. In particular, the DBR1 model predicts substantially smaller long-term benefits of the intervention than does the MSD1 model. Given the MSD1 model's generally superior fit, the policy implications of the DBR1 model may be misleading.

4. Concluding Remarks

There are two main approaches to modeling longitudinal discrete-time binary outcomes in the literature, each emphasizing different aspects of the persistence properties of the data. The DBR approach focuses on Markovian state dependence and usually restricts the effects of observed and unobserved heterogeneity on the probability of transitioning into and out of a state to have the same magnitude but opposite signs. The MSD approach focuses on duration dependence as well

[8] The latter is the average person-average, which gives equal weight to each person who has a fresh spell.
[9] There are exceptions across the age subsamples, and the MSD1 predicted rate is close to the actual for three of the five age groups. Note that because of differences in the outcome history and covariate values at the exit times, the models are not designed to fit the sample averages.

as Markovian state dependence and allows the transition probabilities to vary flexibly. Generally, DBR models are tightly specified and parsimonious, while MSD models are comparatively flexible and more demanding. In this paper we formally show that typical (first and second order) DBR models are nested within an MSD model specification, so that the two approaches should not be viewed as separate.

The case study analysis of poverty experiences demonstrates the two approaches. We conclude that the estimated MSD models fit the data far better than the first- and second-order DBR models, with each of the DBR model restrictions being severely rejected. Given that the MSD models are more flexibly specified than the parsimonious DBR models, it is perhaps not surprising that the log quasi-likelihood values are much higher for MSD models and that the DBR models are formally rejected in statistical tests. However, this conclusion also holds in terms of the model predictions: the MSD model's within-sample predictions were substantially better than the predictions from the DBR model. Finally, we showed that the choice of model specification also matters for deriving policy implications from the fitted models. These findings underscore the potential limitations of the popular DBR model approach in this case. Of course the DBR model may be sufficient in some empirical situations, but the results here emphasize the importance of considering more flexible alternatives.

While all commonly used DBR model specifications are special cases of a relatively simple MSD model specification, it is of course possible to consider more general specifications. As mentioned, phenomena such as occurrence dependence and lagged duration dependence may be important in practice. Furthermore, it is straightforward to allow for covariate effects to vary with elapsed time spent in the current state. Lastly, the influence of unobserved heterogeneity can be extended by allowing for random coefficients more generally.

Author Contributions: Both authors contributed equally to the paper.

Funding: T.G.'s research was supported in part by Australian Research Council Grant DP1096862. D.R.H.'s research was supported in part by the Royal Society of New Zealand Marsden Fund Grant MEP1301.

Acknowledgments: We thank Ann Huff Stevens for providing us the data, Sanghyeok Lee for excellent programming assistance, and Colin Cameron, Stephen Jenkins, Dave Maré and two referees for discussion and comments on earlier drafts of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

Andrén, Thomas, and Daniela Andrén. 2013. Never give up? The persistence of welfare participation in Sweden. *IZA Journal of European Labor Studies* 2: 1–21. [CrossRef]

Andrews, Donald. 1988. Chi-square diagnostic tests for econometric models. *Journal of Econometrics* 37: 135–56. [CrossRef]

Arulampalam, Wiji, Alison L. Booth, and Mark P. Taylor. 2000. Unemployment persistence. *Oxford Economic Papers* 52: 24–50. [CrossRef]

Bane, Mary J., and David T. Ellwood. 1983. *The Dynamics of Dependence: The Routes to Self Sufficiency*. Technical Report. Washington, DC: The U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Barmby, Tim. 1998. The relationship between event history and discrete time duration models: An application to the analysis of personnel absenteeism. *Oxford Bulletin of Economics and Statistics* 60: 261–65. [CrossRef]

Beck, Nathaniel, David Epstein, Simon Jackman, and Sharyn O'Halloran. 2001. Alternative models of dynamics in binary time-series-cross-section models: The example of state failure. Paper presented at the Annual Meeting of the Political Methodology Group, Emory University, Atlanta, GA, USA, July 20.

Beck, Nathaniel, and Jonathan Katz. 1997. The analysis of binary time-series-cross-section data and/or the democratic peace. Paper presented at the Annual Meeting of the Political Methodology Group, Ohio State University, Columbus, OH, USA, July 26.

Bhuller, Manudeep, Christian N. Brinch, and Sebastian Königs. 2017. Time aggregation and state dependence in welfare receipt. *Economic Journal* 127: 1833–73. [CrossRef]

Browning, Martin, and Jesus M. Carro. 2010. Heterogeneity in dynamic discrete choice models. *Econometrics Journal* 13: 1–39. [CrossRef]

Cappellari, Lorenzo, Richard Dorsett, and Getinet Haile. 2007. *State Dependence, Duration Dependence and Unobserved Heterogeneity in the Employment Transitions of the Over-50s*. Working Paper 2007-16. Colchester, UK: Institute for Social and Economic Research.

Cappellari, Lorenzo, and Stephen P. Jenkins. 2004. Modelling low income transitions. *Journal of Applied Econometrics* 19: 593–610. [CrossRef]

Cappellari, Lorenzo, and Stephen P. Jenkins. 2014. The dynamics of social assistance benefit receipt in Britain. *Research in Labor Economics* 39: 39–77.

Card, David, and Dean R. Hyslop. 2005. Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica* 73: 1723–70. [CrossRef]

Card, David, and Dean R. Hyslop. 2009. The dynamic effects of an earnings subsidy for long-term welfare recipients: Evidence from the self sufficiency project applicant experiment. *Journal of Econometrics* 153: 1–20. [CrossRef]

Chay, Kenneth Y., Hillary Hoynes, and Dean Hyslop. 1999. A non-experimental analysis of "true" state dependence in monthly welfare participation sequences. In *American Statistical Association Proceedings of the Business and Economic Statistics Section*. Alexandria: American Statistical Association, August, pp. 9–17.

Chernoff, Herman, and Erich L. Lehmann. 1954. The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Annals of Mathematical Statistics* 25: 579–86. [CrossRef]

Doiron, Denise, and Tue Gørgens. 2008. State dependence in youth labor market experiences, and the evaluation of policy interventions. *Journal of Econometrics* 145: 81–97. [CrossRef]

Gørgens, Tue, and Dean Hyslop. 2018. Equivalent representations of discrete-time two-state panel data models. *Economics Letters* 163: 65–67. [CrossRef]

Halliday, Timothy J. 2008. Heterogeneity, state dependence and health. *Econometrics Journal* 11: 499–516. [CrossRef]

Heckman, James J. 1978. Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE* 30–31: 227–69. [CrossRef]

Heckman, James J. 1981a. Heterogeneity and state dependence. In *Studies in Labor Markets*. Edited by Sherwin Rosen. Chicago and London: University of Chicago Press, pp. 91–139.

Heckman, James J. 1981b. The incidental parameter problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In *Structural Analysis of Discrete Data with Econometric Applications*. Edited by Charles F. Manski and Daniel McFadden. Cambridge: MIT Press, pp. 179–95.

Heckman, James J. 1981c. Statistical models for discrete panel data. In *Structural Analysis of Discrete Data with Economic Applications*. Edited by Charles F. Manski and Daniel McFadden. Cambridge: MIT Press, pp. 114–78.

Heckman, James J., and George J. Borjas. 1980. Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica* 47: 247–83. [CrossRef]

Heckman, James J., and Burton Singer. 1984. Econometric duration analysis. *Journal of Econometrics* 24: 63–132. [CrossRef]

Hyslop, Dean R. 1999. State dependence, serial correlation and heterogeneity in intertemporal labor force participation behavior of married women. *Econometrica* 67: 1255–94. [CrossRef]

Moore, David. 1977. Generalized inverses, Wald's method and the construction of chi-squared tests of fit. *Journal of the American Statistical Association* 72: 131–37. [CrossRef]

Stevens, Ann H. 1999. Climbing out of poverty, falling back in: Measuring the persistence of poverty over multiple spells. *Journal of Human Resources* 34: 557–88. [CrossRef]