The specification of dynamic discrete-time two-state panel data models

Tue Gørgens and Dean Hyslop

Motu Working Paper 16-01
Motu Economic and Public Policy Research

February 2016

Author contact details
Tue Gørgens
Research School Economics
The Australian National University
Canberra 0200, Australia
tue.gorgens@anu.edu.au

Dean Hyslop
Motu Economic and Public Policy Research Trust
PO Box 24390, Wellington 6142
dean.hyslop@motu.org.nz

Acknowledgements
Tue Gørgens' research was supported in part by Australian Research Council Grant DP1096862. Dean Hyslop's research was supported in part by the Royal Society of New Zealand Marsden Fund Grant MEP1301. We thank Ann Huff Stevens for providing us the data, Sanghyeok Lee for excellent programming assistance, and Colin Cameron, Stephen Jenkins and Dave Maré for discussion and comments on an early draft of the paper.

Motu Economic and Public Policy Research
PO Box 24390
Wellington
New Zealand

Email	info@motu.org.nz
Telephone	+64 4 9394250
Website	www.motu.org.nz

Abstract
This paper examines dynamic binary response and multi-spell duration model approaches to analyzing longitudinal discrete-time binary outcomes. Prototypical dynamic binary response models specify low-order Markovian state dependence and restrict the effects of observed and unobserved heterogeneity on the probability of transitioning into and out of a state to have the same magnitude and opposite signs. In contrast, multi-spell duration models typically allow for state-specific duration dependence, and allow the probability of entry into and exit from a state to vary flexibly. We show that both of these approaches are special cases within a general framework. We compare specific dynamic binary response and multi-spell duration models empirically using a case study of poverty transitions. In this example, both the specification of state dependence and the restrictions on the state-specific transition probabilities imposed by the simpler dynamic binary response models are severely rejected against the more flexible multi-spell duration models. Consistent with recent literature, we conclude that the standard dynamic binary response model is unacceptably restrictive in this context.

JEL codes
C33, C35, C41, C51

Keywords
Panel data, transition data, binary response, duration analysis, event history analysis, initial conditions, random effects.

Contents

List of Tables

List of Figures

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 can only take place on the boundary between periods.[1] Analysis of such data is central to many empirical studies in economics and other social sciences. Typical topics include employment (Heckman, 1981a; Hyslop, 1999), unemployment (Arulampalam et al., 2000), poverty (Stevens, 1999; Capellari and Jenkins, 2004), welfare dependency (Bane and Ellwood, 1983), health (Halliday, 2008), and peace and conflict between national states (Beck and Katz, 1997; Beck et al., 2002).

The aim of a two-state panel data analysis is understanding the factors which influence either which state is occupied, or the times of transition between states. 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.[2] There are two conceptually distinct approaches to analyzing two-state panel data in the literature, which differ primarily in their choice of outcome variable and in the type of state dependence they focus on.

The first approach, which we refer to as the dynamic binary response (DBR) approach, focuses on the probability of occupying one of the two states in each period. State dependence is modeled in terms of the effects of previous periods' state occupancy on the probability distribution for the current period's state occupancy (Markovian state dependence). Usually, DBR models implicitly assume that the effects of heterogeneity have the same magnitude but opposite signs on the implied probabilities of transitioning

[1] These are also known as discrete-time transition data (Lancaster, 1990). Two-state data can be contrasted with "one-state" data where a binary observation indicates a recurring event of short duration such as a heart attack or an earth quake, and the main interest is understanding the time until the next event. Discrete-time data representing an underlying continuous-time process may involve interval censoring and are outside the scope of this paper (Huang and Wellner, 1997; Chen et al., 2013).

[2] Heckman (1978, 1981c) first proposed the latent variable threshold model, with a general intertemporal covariance matrix for the disturbances. Heckman distinguishes three sources of persistence: time-invariant unobserved heterogeneity, persistent shocks, and true state dependence. Heckman and Borjas (1980) distinguish four types of state dependence: (first-order) Markovian state dependence, (current-spell) duration dependence, lagged duration dependence, and occurrence dependence.

into and out of a state. A simple first-order DBR model specification is

$$\mathsf{P}(Y_{it} = 1 | \mathbf{H}_{it-1} = \mathbf{h}_{it-1}, X_{it} = x_{it}, V_i = v_i) = G(\alpha + \gamma_1 y_{it-1} + x_{it}'\beta + \delta v_i), \tag{1}$$

where Y_{it} is a binary indicator of state occupancy for individual i in period t, X_{it} is vector of observable covariates, V_i is a random variable representing time-constant unobservable factors, $\mathbf{H}_{it}$ denotes the entire history of covariates and outcomes to period t, and G is the logistic function. The complete model requires specification of probability distributions for Y_{i1} and for V_i. We postpone those issues till later.

The second approach, which we refer to as the multi-spell duration (MSD) approach, focuses on the probability that the current-state spell ends or, equivalently, that a transition between states occurs. State dependence is usually modeled in terms of the effects of the current-spell elapsed duration since last entering the current state on the probability of a transition occurring (duration dependence). Moreover, because MSD models usually allow the transition probabilities to vary by state, they also allow for (first-order) Markovian state dependence.[3] A simple MSD model specification is

$$\begin{aligned} &\mathsf{P}(C_{it} = 1 | \mathbf{H}_{it-1} = \mathbf{h}_{it-1}, X_{it} = x_{it}, V_i = v_i) \\ &\qquad = \begin{cases} G(\alpha_0 + \lambda_0 d_{it-1} + x_{it}'\beta_0 + \delta_0 v_i) & \text{if } y_{it-1} = 0, \\ G(\alpha_1 + \lambda_1 d_{it-1} + x_{it}'\beta_1 + \delta_1 v_i) & \text{if } y_{it-1} = 1, \end{cases} \end{aligned} \tag{2}$$

where C_{it} is a binary indicator for whether individual i makes a transition between periods $t-1$ and t, and D_{it} is elapsed duration at time t. Again, we postpone the discussion of the complete model till later.

The objective of this paper is to compare these two approaches to modeling discrete-time two-state panel data. While the focus is on state occupancy probabilities in DBR models and on transition probabilities in MSD models, either model can be used to estimate both probabilities, as well as mean spell durations etc. Analysts therefore have

[3] Some studies also consider the effects of the number of previous transitions (occurrence dependence), and/or of completed durations of previous spells (lagged duration dependence), see e.g. Doiron and Gørgens (2008).

a choice of which approach to use. In practice, DBR models are far more widely used than MSD models, but very few studies discuss the implications of the DBR models for hazard rates and spell durations.[4] One of the main conclusions of this paper is that commonly used first-order and second-order specifications of DBR models are nested within a simple MSD model and, when seen from a duration analysis perspective, embody strong and questionable restrictions on both state dependence and the effects of observed and unobserved heterogeneity.[5]

The paper begins by discussing theoretical issues using a general framework for modeling two-state panel data that encompasses the DBR and MSD models. We first show that a sequence of binary state occupancy indicators can be equivalently represented by the initial state and a sequence of transition indicators associated with subsequent periods, and that modeling the probabilities of state occupancy or the probabilities of transitions between states is nonparametrically equivalent. We then consider issues associated with incorporating observed covariates and unobserved heterogeneity in the models, and handling left-censored spells and initial outcomes at the start of the observation period. The discussion demonstrates that the approaches differ in their specification of state dependence, and how flexibly they model observed and unobserved heterogeneity. Furthermore, because alternative specifications differ in their implications for the initial outcomes and left-censoring, they vary in their demands on the data. While DBR and MSD models are not nested in general, we conclude the theoretical discussion by showing that commonly used DBR model specifications are indeed special cases of simple MSD models.

We then use an empirical case study to illustrate potential limitations of the popular DBR approach. 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

[4]To the best of our knowledge little comparative analysis has been conducted of these alternative approaches. Exceptions include Cappellari et al. (2007) who compare a duration and Markov model for employment transitions, and Bhuller et al. (2014) who analyze the adequacy of first-order dynamic binary response models against more general models that allow for duration and occurrence dependence.

[5]For example, (1) is the special case of (2) with $\alpha_0 = \alpha$, $\alpha_1 = \alpha + \gamma$, $\lambda_0 = 0$, $\lambda_1 = 0$, $\beta_0 = \beta$, $\beta_1 = -\beta$, $\delta_0 = \delta$, and $\delta_1 = -\delta$.

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 rejected. Consequently, the MSD models provide better within-sample predictions than do the DBR models.

The paper is organized as follows. Section 2 introduces the general framework and discusses the DBR and MSD special cases. Section 3 presents the empirical analysis. The paper concludes with a discussion in Section 4. The appendix provides a link between the representations given in the main text and continuous-time duration analysis.

2 Modeling discrete-time two-state panel data

In this section we present a general framework for handling discrete-time two-state panel data, that encompasses the DBR and MSD models commonly used in analysis.

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 can only take place at the boundary 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 eligibility for 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 1 transition takes place in each period and that the precise timing of the transition within this period can be ignored. For example, outcomes indicating whether or not an individual is married at the time of an annual interview might be adequately described in discrete time. 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/unemployment status at the time of an annual interview must deal with unobserved transitions between interviews.

2.1 Equivalent data representations

In general, two-state panel data consists of a sequence of binary outcomes ordered in time for each individual (or entity). The outcomes indicate which of two states, labeled 0 and 1, the individual is occupying. The times are evenly spaced and may refer to periods in time or points in time. There can be at most 1 transition between two consecutive times. Individuals are indexed by $i = 1, \ldots, N$, and time by $t = 1, \ldots, T$. For simplicity, we assume the data constitute a balanced panel. The indicators of the state occupied by individual i at time t are denoted Y_{it} for $t = 1, \ldots, T$ 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} for $t = 2, \ldots, T$ with $C_{it} \in \{0, 1\}$. We assume the individuals constitute an independent random sample from a given population.[6] In the following, we consider a representative individual i and suppress the range $i = 1, \ldots, N$. Lower case letters with a subscript i represent observed or realized values.

The data may be incomplete. We shall refer to potential outcomes, whether they are actually observed or not, as the "process" and reserve the word "observations" for the actual outcomes available for analysis. Outcomes may be missing at the beginning, during, or at the end of the process. If the process begins before the first observation takes place, then the data are said to be left-censored. We assume that the individual time origin is not known if the data are left-censored. Although observation ends at T, we make no assumption that the process has ended at time T. Therefore, we assume that the data are always right-censored at time T. Throughout the paper we assume that left- and right-censoring are independent of the underlying process. Essentially this means that the start and the end of the observation period are not determined such that certain outcome sequences are favored. For simplicity, our general framework allows only for left- and right-censoring; middle-censoring can be handled using similar methods.

The different modeling approaches require different organization of the data. The DBR approach models $(Y_{i1}, Y_{i2}, \ldots, Y_{iT})$ while the MSD approach models $(Y_{i1}, C_{i2}, \ldots, C_{iT})$. However, it is easy to show that the two representations are equivalent in the sense the

[6]In the empirical case study we work with a cluster sample.

one can be recovered from the other. Specifically, the Y_{it}s and C_{it}s are related by

$$C_{it} = 1(Y_{it-1} \neq Y_{it}), \quad t = 2, \ldots, T, \tag{3}$$

and

$$Y_{it} = \left(Y_{i1} + \sum_{k=2}^{t} C_{ik} \right) \bmod 2, \quad t = 2, \ldots, T. \tag{4}$$

Therefore, from a data perspective focusing on the sequence of states occupied or on the transitions between states is equivalent.

In duration analysis, the data are often represented as transition times or spell durations, instead of a sequence of state or transition indicators. (A spell is a period between consecutive transitions during which the individual stays in the same state.) We focus here on time-based representations, since they are most convenient if covariates are time-varying. However, we show in Appendix A that spell-based and time-based representations are also equivalent.

2.2 Equivalent parameterizations

Random sampling of individuals identifies the probability distributions of the sequences $(Y_{i1}, Y_{i2}, \ldots, Y_{iT})$ and $(Y_{i1}, C_{i2}, \ldots, C_{iT})$. Since these distributions are discrete, they can be characterized by a finite number of probabilities (2^T). These probabilities can be nonparametrically estimated given a sufficiently large sample. However, the parameters of interest in most applications are not the probabilities associated with these unconditional joint distributions, but rather conditional probabilities of current outcomes given past outcomes.

The DBR approach focuses on the conditional probabilities of being in one of the states given the sequence of states previously occupied. Let $\mathbf{Y}_{it} = (Y_{i1} \ldots, Y_{it})$ denote the outcome history up to time t, for $t = 1, \ldots, T$, and let $\mathbf{y}_{it} = (y_{i1}, \ldots, y_{it})$ denote the observed history. Let $\mathbf{y}_t$ with no subscript i denote a generic element of $\{0, 1\}^t$. Then the conditional probability of being in state 1 at time t, given the outcome history prior to time t, is

$$\begin{aligned} &\chi = \mathsf{P}(Y_{i1} = 1), \\ &\zeta_t(\mathbf{y}_{t-1}) = \mathsf{P}(Y_{it} = 1|\mathbf{Y}_{it-1} = \mathbf{y}_{t-1}), \quad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad t = 2, \ldots, T. \end{aligned} \tag{5}$$

Note that there is the initial probability, and 2, 4, ..., 2^{T-1} conditional probabilities in the equations in (5), depending on the conditioning set; adding them up yields $2^T - 1$ total probabilities.

Assuming there is no left-censoring and that right-censoring is independent of outcomes, the probabilities in (5) are fundamental parameters of interest. With left-censoring, they may or may not be, depending on whether there is interest in the effect of past unobserved outcomes; i.e. the parameters of interest may depend on unobserved outcomes prior to the observation period. We return to this issue in Section 2.5.

In any case, treating each of the probabilities in (5) as a parameter to be estimated, the likelihood contribution for individual i is[7]

$$L_i^Y = \chi^{y_{i1}}(1-\chi)^{1-y_{i1}} \prod_{t=2}^{T} \zeta_t(\mathbf{y}_{it-1})^{y_{it}}(1-\zeta_t(\mathbf{y}_{it-1}))^{1-y_{it}}. \tag{6}$$

Combining the contributions of all N individuals yields a likelihood function which is valid for inference under the assumptions stated above. It is straightforward to show that the maximum likelihood estimates are simply the sample analogs.

In contrast to the DBR approach, the MSD approach focuses on the conditional probabilities of changing state at time t given the prior history; that is, on the hazard rates. In addition, there is the probability distribution of the initial state. The conditional probability of beginning in state 1 given the outcome history prior to time t and the hazard

[7]Obviously, the likelihood can be equivalently parameterized in terms of being in state 0 rather than state 1 (i.e. $Y_{it} = 0$ versus $Y_{it} = 1$).

rates are defined as[8]

$$
\begin{aligned}
&\chi = \mathsf{P}(Y_{i1} = 1), \\
&\xi_t(\mathbf{y}_{t-1}) = \mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{t-1}), \quad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad t = 2, \ldots, T.
\end{aligned} \tag{7}
$$

Again, there is 1 initial probability, and 2, 4, …, 2^{T-1} conditional probabilities in the equations in (7), giving $2^T - 1$ distinct probabilities in this representation.

Treating each of the probabilities in (7) as a parameter to be estimated and using c_{it} for the observed transition indicator, the likelihood contribution for individual i is[9]

$$
L_i^C = \chi^{y_{i1}} (1-\chi)^{1-y_{i1}} \prod_{t=2}^{T} \xi_t(\mathbf{y}_{it-1})^{c_{it}} (1 - \xi_t(\mathbf{y}_{it-1}))^{1-c_{it}}. \tag{8}
$$

The comments following (6) apply here as well. In the absence of left-censoring, the probabilities in (7) are fundamental parameters of interest. However, the question of whether the probabilities are parameters of interest with left-censored data is complicated. We return to this issue in Section 2.6.

To emphasize that (6) and (8) are simply reparameterizations of the same likelihood, note that

$$
\begin{aligned}
\zeta_t(\mathbf{y}_{t-1}) &= \xi_t(\mathbf{y}_{t-1})^{1-y_{t-1}} (1 - \xi_t(\mathbf{y}_{t-1}))^{y_{t-1}} \\
&= 1 - \xi_t(\mathbf{y}_{t-1})^{y_{t-1}} (1 - \xi_t(\mathbf{y}_{t-1}))^{1-y_{t-1}}, \quad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad t = 2, \ldots, T,
\end{aligned} \tag{9}
$$

and

$$
\begin{aligned}
\xi_t(\mathbf{y}_{t-1}) &= \zeta_t(\mathbf{y}_{t-1})^{1-y_{t-1}} (1 - \zeta_t(\mathbf{y}_{t-1}))^{y_{t-1}} \\
&= 1 - \zeta_t(\mathbf{y}_{t-1})^{y_{t-1}} (1 - \zeta_t(\mathbf{y}_{t-1}))^{1-y_{t-1}}, \quad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad t = 2, \ldots, T,
\end{aligned} \tag{10}
$$

where y_{t-1} denotes the final element of $\mathbf{y}_{t-1}$. Thus, the likelihood functions are equivalent,

[8] It is possible to define $\mathbf{C}_{it} = (Y_{i1}, C_{i2}, \ldots, C_{it})$ for $t \geq 1$, but since $\mathbf{Y}_{it}$ and $\mathbf{C}_{it}$ are equivalent we do not need $\mathbf{C}_{it}$.

[9] Obviously, the likelihood can be equivalently parameterized in terms of the absence rather than the presence of a transition (i.e. $C_{it} = 0$ versus $C_{it} = 1$).

since both the data representations and the parameters are in one-to-one relationships.[10]

A comparison of spell-based and time-based parameterizations of the likelihood function is given in Appendix A.

2.3 Covariates

Introducing (predetermined) covariates has little effect on the theoretical analysis; however, in an empirical analysis there are two practical issues regarding the time reference for covariates. First, surveys often collect retrospective information relating to different periods, so the time of observation may not be the same as the logical time reference for the information. (This is of course true for the outcome variables as well.) Second, some covariates which logically relate to time t are inappropriate conditioning variables because of simultaneity issues. For example, it is probably not interesting to condition a person's employment status in a given month on wage income earned in that month. In the discussion here, we assume that covariates are lagged or led so that it is sensible to condition outcomes at t on covariates with (notational) time reference t.

In preparation for the statement of the general likelihood functions later, we briefly present the conditional likelihood contributions for individual i given their covariate history. Let X_{it} denote a vector of covariates for individual i with reference to time t. For $t = 1, \ldots, T$, let $\mathbf{X}_{it} = (X_{i1}, \ldots, X_{it})$ denote the covariate history at time t, and let $\mathbf{x}_{it} = (x_{i1}, \ldots, x_{it})$ denote the observed history. Let $\mathcal{X}_t$ denote the support of $\mathbf{X}_{it}$, and let $\mathbf{x}_t$ denote a generic element of $\mathcal{X}_t$.

For the DBR approach, let the conditional probability of being in state 1 at time t

[10] Let c_t denote whether a transition occurs between $t-1$ and t according to $\mathbf{y}_t$. Substituting (10) into (8) gives (6) since, for given t, we have

$$\begin{aligned}\xi_t(\mathbf{y}_{t-1})^{c_t}(1-\xi_t(\mathbf{y}_{t-1}))^{1-c_t} &= \zeta_t(\mathbf{y}_{t-1})^{(1-y_{t-1})c_t+y_{t-1}(1-c_t)}(1-\zeta_t(\mathbf{y}_{t-1}))^{y_{t-1}c_t+(1-y_{t-1})(1-c_t)} \\ &= \zeta_t(\mathbf{y}_{t-1})^{c_t+y_{t-1}-2c_ty_{t-1}}(1-\zeta_t(\mathbf{y}_{t-1}))^{1-c_t-y_{t-1}+2c_ty_{t-1}} = \zeta_t(\mathbf{y}_{t-1})^{y_t}(1-\zeta_t(\mathbf{y}_{t-1}))^{1-y_t}.\end{aligned}$$

The final step follows because $y_t = c_t + y_{t-1} - 2c_ty_{t-1}$. Similarly, plugging (9) into (6) gives (8).

given the history to time t be represented by

$$\begin{aligned}
&\chi(\mathbf{x}_1) = \mathsf{P}(Y_{i1} = 1 | \mathbf{X}_{i1} = \mathbf{x}_1), \quad \mathbf{x}_1 \in \mathcal{X}_1, \\
&\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \mathsf{P}(Y_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{t-1}, \mathbf{X}_{it} = \mathbf{x}_t), \\
&\qquad\qquad\qquad\qquad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T.
\end{aligned} \tag{11}$$

Then the likelihood contribution for individual i, conditional on right-censoring at T and on the covariate history, is

$$L_i^Y = \chi(\mathbf{x}_{i1})^{y_{i1}} (1 - \chi(\mathbf{x}_{i1}))^{1-y_{i1}} \prod_{t=2}^{T} \zeta_t(\mathbf{y}_{it-1}, \mathbf{x}_{it})^{y_{it}} (1 - \zeta_t(\mathbf{y}_{it-1}, \mathbf{x}_{it}))^{1-y_{it}}. \tag{12}$$

For the MSD approach, let the conditional probability of being in state 1 at time 1 and the hazard rates given the history to time t be

$$\begin{aligned}
&\chi(\mathbf{x}_1) = \mathsf{P}(Y_{i1} = 1 | \mathbf{X}_{i1} = \mathbf{x}_1), \quad \mathbf{x}_1 \in \mathcal{X}_1, \\
&\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{t-1}, \mathbf{X}_{it} = \mathbf{x}_t), \\
&\qquad\qquad\qquad\qquad \mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T.
\end{aligned} \tag{13}$$

Then the likelihood contribution for individual i, conditional on right-censoring at T, is

$$L_i^C = \chi(\mathbf{x}_{i1})^{y_{i1}} (1 - \chi(\mathbf{x}_{i1}))^{1-y_{i1}} \prod_{t=2}^{T} \xi_t(\mathbf{y}_{it-1}, \mathbf{x}_{it})^{c_{it}} (1 - \xi_t(\mathbf{y}_{it-1}, \mathbf{x}_{it}))^{1-c_{it}}. \tag{14}$$

If all covariates are discrete, so that X_{it} can take only a finite, say k, number of values, then there are $(2k)^{T-1}$ unknown probabilities in the likelihood contributions. Since this is also a finite number, the probabilities are in principle nonparametrically identified. If some covariates are continuous, the probabilities may be nonparametrically identified and estimable using standard smoothing techniques such as kernel regression, series estimation, or maximum penalized likelihood.

The definitions in (11) and (13) are in parallel with (5) and (7). Expressions analogous to (9) and (10) also hold, namely

$$\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t)^{1-y_{t-1}}(1 - \xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t))^{y_{t-1}},$$
$$\mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T, \quad (15)$$

and

$$\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t)^{1-y_{t-1}}(1 - \zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t))^{y_{t-1}},$$
$$\mathbf{y}_{t-1} \in \{0,1\}^{t-1}, \quad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T. \quad (16)$$

Again these equations imply that (12) and (14) are reparameterizations of the same likelihood function, so the DBR and MSD approaches remain nonparametrically equivalent after controlling for covariates.

In linear models, if the outcome variable depends on the level of covariates, then *changes* in the outcome variable over time depends on *changes* in covariates. This relationship does not carry over to the present nonlinear context. Equations (15) and (16) show that the parameters $\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ and $\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ are simple transformations of each other. Therefore, if the covariates affect $\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ in a certain way, say by their contemporary levels or by their time change or by some single-index restriction, then they affect $\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ in essentially the same way (and vice versa). It follows that the question of whether the level of a covariate or changes over time matters is an issue distinct from whether the focus is on the probability of state occupancy or transitions between states. If changes in covariates matter, then both probabilities depends on the changes. If levels matter, both probabilities depends on the levels of the covariate.

The property that the covariate relationship is the same for the probability of state occupancy and the probability of transitioning depends critically on the conditioning on the previous state occupancy, Y_{it-1}. This can be seen in an example. Suppose that the covariates are strictly exogenous and that $\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ depends only on contemporary covariates; that is, $\zeta_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \zeta_0(x_t)$ for $t = 2, \ldots, T$, where ζ_0 is some function. Then by (16), $\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t)$ depends on the covariates in the same way; namely $\xi_t(\mathbf{y}_{t-1}, \mathbf{x}_t) = \zeta_0(x_t)^{1-y_{t-1}}(1-\zeta_0(x_t))^{y_{t-1}}$ for $t = 2, \ldots, T$. However, without conditioning on Y_{it-1} the probabilities are

$$\mathsf{P}(Y_{it} = 1 | \mathbf{X}_{it} = \mathbf{x}_t) = \zeta_0(x_t), \quad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T, \tag{17}$$

and

$$\mathsf{P}(C_{it} = 1 | \mathbf{X}_{it} = \mathbf{x}_t) = (1-\zeta_0(x_t))\zeta_0(x_{t-1}) + \zeta_0(x_t)(1-\zeta_0(x_{t-1})), \qquad \mathbf{x}_t \in \mathcal{X}_t, \quad t = 2, \ldots, T. \tag{18}$$

It follows that if the state occupancy probabilities are static, then these transition probabilities depend on both contemporary and lagged covariates, although not necessarily in a simple first-difference form.

2.4 Unobserved heterogeneity

Unobserved heterogeneity may be a concern if average probabilities do not represent outcomes for specific individuals, even conditional on X_{it}. For example, a population may contain some people with strong immune systems and others who easily get sick. The average probability of coming down with the flu in a particular week given that a person is not already sick may reflect a near-zero probability for the former and near-one probability for the latter group. The parameters of interest are the individual-specific probabilities of becoming sick, rather than the average probability. If characteristics of a person's immune system were observed and available in the data, they could simply be included as covariates and there would be no problem. However, if data are not available, there is important unobserved heterogeneity in the population.

Unobserved heterogeneity precludes nonparametric identification of parameters of interest. Untestable assumptions such as parametric functional-form specifications are necessary if the data are to be used for inference. In the literature, unobserved heterogeneity is treated as equivalent to an omitted covariate. Usually it is assumed to be predetermined for each individual, independent of covariates (past and future), and independent of the observation scheme. While these assumptions are strong and perhaps implausible in most applications, they are still not sufficient to ensure identification. We proceed here by presenting the general form of the likelihood contributions in the presence of (independent) unobserved heterogeneity, without being explicit about identifying assumptions. Specific cases are discussed in detail later.

Let V_i denote a random variable (or vector) representing unobserved heterogeneity for individual i. Let $\mathcal{V}$ denote the support of V_i, and let Ψ denote the distribution function of V_i. To keep the expressions simple and compact, the likelihood functions are stated in terms of probabilities rather than Greek-letter parameters.[11] In the DBR framework, the likelihood contribution for individual i becomes

$$L_i^Y = \int_{\mathcal{V}} \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}, V_i = v) \times \left(\prod_{t=2}^{T} \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = v) \right) d\Psi(v). \quad (19)$$

Similarly, in the MSD framework we have

$$L_i^C = \int_{\mathcal{V}} \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}, V_i = v) \times \left(\prod_{t=2}^{T} \mathsf{P}(C_{it} = c_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = v) \right) d\Psi(v). \quad (20)$$

An important implication is that the likelihood contributions are no longer separable across time. As we have seen, the likelihood contributions can be broken into multiplicative time-specific components when there is no unobserved heterogeneity. If unobserved heterogeneity needs to be integrated out, this is no longer the case.

In practice, there are different ways of incorporating unobserved heterogeneity in the literature. A common DBR approach is to specify V_i as a normally distributed scalar random variable and include V_i as a regressor with a loading similar to the covariates (e.g. Hyslop, 1999; Chay and Hyslop, 2014); cf. equations (1) and (2). In the MSD

[11] For example, we write $\mathsf{P}(Y_{i1} = y_{i1})$ instead of $\chi^{y_{i1}}(1-\chi)^{1-y_{i1}}$ as in (6) and (8).

framework, it is natural to specify different marginal distributions for the entry and exit hazard rates and allow for correlations between them. Following Heckman and Singer (1984), an alternative which we adopt here is to assume that unobserved heterogeneity has a discrete distribution in a multidimensional space. The discrete distribution can be thought of either as an approximation to a true underlying continuous distribution or as a distribution of a finite number, K, of "types". If a model has a number, Q, of "equations" each representing a different aspect, then we assume each type is characterized by a Q-vector of constants, one for each equation. Formally, we assume that V_i is a discrete random Q-vector with support $\{\nu_1, \ldots, \nu_K\}$, where $\nu_k = (\nu_{k1}, \nu_{k2}, \ldots, \nu_{kQ}) \in \mathbb{R}^Q$ for $k = 1, \ldots, K$, and probability distribution $\pi_1, \ldots, \pi_K$ with $\sum_{k=1}^{K} \pi_k = 1$.

2.5 The DBR approach

In the DBR approach, the model focuses on low-order (p) Markovian state dependence, and assumes that the conditional probabilities of being in a given state depend only on the p most recent outcomes. For simplicity, assume also that only contemporary covariates matter.[12] Thus, for fixed $p \geq 1$ it is assumed that

$$\begin{aligned}
&\mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = v_i) \\
&\qquad = \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}^p_{it-1} = \mathbf{y}^p_{it-1}, X_{it} = x_{it}, V_i = v_i), \quad t = p+1, \ldots, T, \quad (21)
\end{aligned}$$

where $\mathbf{Y}^p_{it-1} = (Y_{it-p}, \ldots, Y_{it-1})$ and $\mathbf{y}^p_{it-1} = (y_{it-p}, \ldots, y_{it-1})$. Equation (21) is the DBR model's main equation of interest, which we refer to as the "structural" equation. This equation does not restrict the probability distribution for the initial p outcomes, $(Y_{i1}, \ldots, Y_{ip})$, referred to as the "initial conditions" of the process. There is typically less substantive 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 (Heckman, 1981b).

Assuming (21) and the discrete distribution of unobserved heterogeneity, the likelihood

[12] In practice, the specification of covariates may include either contemporaneous and/or leads and lags of exogenous covariates.

contribution (19) can be written

$$L_i^Y = \sum_{k=1}^{K} \pi_k \Bigg[A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k) \times \Bigg(\prod_{t=p+1}^{T} \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1}^p = \mathbf{y}_{it-1}^p, X_{it} = x_{it}, V_i = \nu_k) \Bigg) \Bigg], \quad (22)$$

where

$$A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k) = \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}, V_i = \nu_k) \times \Bigg(\prod_{t=2}^{p} \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \Bigg). \quad (23)$$

Thus the likelihood contribution for individual i has two main components: the contribution of the structural equations in big parentheses in (22), and the contribution of the initial outcomes, $(y_{i1}, \ldots, y_{ip})$, represented by the term $A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k)$.[13]

If the process is observed from the beginning, so the data are not left-censored, the initial outcomes will be relevant (unless all individuals have the same states for the first p periods). In this case, the probabilities in (23) may be of substantive interest in terms of the underlying process. However, typically the process is ongoing prior to the observation period, so the data are left-censored, and the probabilities in (23) are determined partly by the unknown true origins of the process and partly by the mechanisms described in the structural equations. Without knowledge of (or interest in) the former, A^Y has unknown functional form and is treated as a nuisance parameter. Below we adopt Heckman's (1981b) suggestion of modeling A^Y flexibly and separately from the structural equations.

If there is no unobserved heterogeneity, say $\mathsf{P}(V_i = \nu_1) = 1$, then the sum over K-types in (22) effectively disappears. In this case, the term in big parentheses in (22) involve only observed variables, and these probabilities are nonparametrically identified. (They can be estimated by their sample analogs.) Thus the term $A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_1)$ in (22) can be ignored when maximizing the likelihood, and valid inference obtained conditional on $\mathbf{Y}_{ip}$

[13]Note that $A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k)$ is defined in (23) using the most general specification for the probabilities assuming the covariates are predetermined. In general $A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k) \neq \mathsf{P}(\mathbf{Y}_{ip} = \mathbf{y}_{ip} | \mathbf{X}_{ip} = \mathbf{x}_{ip}, V_i = \nu_k)$.

and $\mathbf{X}_{ip}$.

Parametric DBR models

Adapting the ideas of Heckman (1981b), we shall model $A^Y(\mathbf{y}_{ip}, \mathbf{x}_{ip}, \nu_k)$, representing the probabilities of the first p outcomes in (23), using p "approximate reduced form" equations.[14] The probabilities associated with the structural equations in (22) are represented by equation (21). The simplest and most common DBR model used empirically adopts $p = 1$, although $p = 2$ 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).

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

$$\mathsf{P}(Y_{it} = 1 | \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) = G(\nu_{k1} + \beta_1' x_{it}) \equiv G_{it}^{11}(\nu_{k1}, \beta_1), \quad t = 1, \tag{24}$$

and

$$\begin{aligned}&\mathsf{P}(Y_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\ &\qquad = G(\nu_{k2} + \beta_2' x_{it} + \gamma_2 y_{it-1}) \equiv G_{it}^{12}(\nu_{k2}, \beta_2, \gamma_2), \quad t = 2, \ldots, T. \end{aligned} \tag{25}$$

The corresponding likelihood contribution, cf. (22), for individual i is

$$\begin{aligned}&L_i^{DBR1}(\nu_1, \ldots, \nu_K, \pi_1, \ldots, \pi_K, \beta_1, \beta_2, \gamma_2) \\ &\quad = \sum_{k=1}^{K} \pi_k \Bigg[G_{it}^{11}(\nu_{k1}, \beta_1)^{y_{it}} \big(1 - G_{it}^{11}(\nu_{k1}, \beta_1)\big)^{1-y_{it}} \\ &\qquad \times \left(\prod_{t=2}^{T} G_{it}^{12}(\nu_{k2}, \beta_2, \gamma_2)^{y_{it}} \big(1 - G_{it}^{12}(\nu_{k2}, \beta_2, \gamma_2)\big)^{1-y_{it}} \right) \Bigg]. \end{aligned} \tag{26}$$

The dimension of the parameters are $\beta_q \in \mathbb{R}^{\dim(x)}$ for $q = 1, 2$, and $\gamma_2 \in \mathbb{R}$. Unobserved heterogeneity is represented by a probability π_k and a 2-vector $\nu_k = (\nu_{k1}, \nu_{k2}) \in \mathbb{R}^2$ for

[14] An alternative approach is to condition on the initial conditions (Wooldridge, 2005), resulting in a single-equation model. This approach commingles the underlying process and the sampling scheme.

$k = 1, \ldots, K$.

In the DBR2 model we relax the assumption of first-order Markovian state dependence and consider second-order Markovian state dependence. This model naturally 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:

$$\mathsf{P}(Y_{it} = 1 | \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) = G(\nu_{k1} + \beta_1' x_{it}) \equiv G_{it}^{21}(\nu_{k1}, \beta_1), \quad t = 1, \tag{27}$$

$$\begin{aligned} &\mathsf{P}(Y_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\ &\qquad = G(\nu_{k2} + \beta_2' x_{it} + \gamma_2 y_{i1}) \equiv G_{it}^{22}(\nu_{k2}, \beta_2, \gamma_2), \quad t = 2, \end{aligned} \tag{28}$$

and

$$\begin{aligned} &\mathsf{P}(Y_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\ &\quad = G(\nu_{k3} + \beta_3' x_{it} + \gamma_{31} y_{it-1} + \gamma_{32} y_{it-2} + \gamma_{33} y_{it-1} y_{it-2}) \equiv G_{it}^{23}(\nu_{k3}, \beta_3, \gamma_3), \\ &\qquad t = 3, \ldots, T. \end{aligned} \tag{29}$$

The corresponding likelihood contribution, c.f. (22), for individual i is

$$\begin{aligned} &L_i^{DBR2}(\nu_1, \ldots, \nu_K, \pi_1, \ldots, \pi_K, \beta_1, \beta_2, \beta_3, \gamma_2, \gamma_3) \\ &\quad = \sum_{k=1}^{K} \pi_k \Big[G_{it}^{21}(\nu_{k1}, \beta_1)^{y_{it}} \big(1 - G_{it}^{21}(\nu_{k1}, \beta_1)\big)^{1-y_{it}} \\ &\quad \times G_{it}^{22}(\nu_{k2}, \beta_2, \gamma_2)^{y_{it}} \big(1 - G_{it}^{22}(\nu_{k2}, \beta_2, \gamma_2)\big)^{1-y_{it}} \\ &\quad \times \Big(\prod_{t=3}^{T} G_{it}^{23}(\nu_{k3}, \beta_3, \gamma_3)^{y_{it}} \big(1 - G_{it}^{23}(\nu_{k3}, \beta_3, \gamma_3)\big)^{1-y_{it}} \Big) \Big]. \end{aligned} \tag{30}$$

The dimension of the parameters are $\beta_q \in \mathbb{R}^{\dim(x)}$ for $q = 1, 2, 3$, $\gamma_2 =\in \mathbb{R}$, and $\gamma_3 = (\gamma_{31}, \gamma_{32}, \gamma_{33}) \in \mathbb{R}^3$. Unobserved heterogeneity is represented by a probability π_k and a 3-vector $\nu_k = (\nu_{k1}, \nu_{k2}, \nu_{k3}) \in \mathbb{R}^3$ for $k = 1, \ldots, K$.

Several points are in order. First, in practice G is either the logistic or the standard normal distribution function, although in principle G could be any function compatible with probability and does not even need to be monotone.

Second, it is instructive to consider the hazard rates implied by low-order DBR models. These are shown in Figures 1 and 2 for models with no covariates and no unobserved heterogeneity. The figures graph the time profiles of the hazard rates out of state 0 and state 1 assuming a new spell begins in period t. For the DBR1 model they are everywhere constant, and for the DBR2 model they are constant from period $t + 2$. In general, a pth-order DBR model has hazard rates that are constant from period $t + p$ onwards (not shown). From the perspective of duration analysis, the hazard rates embodied in these DBR models are very restrictive.

Third, it is assumed that the effects of covariate 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 hazard rates have the same magnitude but opposite signs. Several authors have pointed out the possibility of interacting covariates with the lagged outcome variables (e.g. Barmby, 1998; Beck et al., 2002). Although the intent is often to allow for heterogeneity in state dependence rather than a specific consideration of transition probabilities, this was done by e.g. Card and Hyslop (2009), Browning and Carro (2010), and Capellari and Jenkins (2014).

2.6 The MSD approach

The main interest in the MSD approach is usually the dependence of the transition probabilities on the elapsed time spent in the current state (i.e. duration dependence). In addition, by modeling the transition probabilities out of each state using separate equations, the MSD approach obviously includes first-order Markovian state dependence. For simplicity, we ignore other forms of state dependence (e.g. lagged duration dependence and occurrence dependence) in the models discussed in this paper.

To present the models, let $F_i = \min\{t : Y_{it-1} \neq Y_{it} \text{ and } 1 \leq t \leq T\}$ denote the time of the first observed transition, and define $F_i = T$ if no transitions are observed for individual i. Also, let $D_{it} = t + 1 - \min\{s : Y_{ik} = Y_{it} \text{ for } k = s, \ldots, t\}$ for $t = 1, \ldots, T$ denote observed elapsed duration in the spell ongoing at time t. With this definition we have $D_{i1} = 1$ and $D_{iF_i} = 1$.

The absence of lagged duration dependence and occurrence dependence implies that outcomes prior to entering the current spell do not influence the transition probabilities. With the additional assumption that only contemporaneous covariates matter, the transition probability for the "fresh" (i.e. non-left-censored) spells that start during the observation period is

$$\begin{aligned}
&\mathsf{P}(C_{it} = c_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = v_i) \\
&\quad = \mathsf{P}(C_{it} = c_{it} | Y_{it-1} = y_{it-1}, D_{it-1} = d_{it-1}, X_{it} = x_{it}, V_i = v_i), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t = f_i + 1, \ldots, T. \quad (31)
\end{aligned}$$

Conditioning on $Y_{it-1} = y_{it-1}$ means that the transition probabilities may depend on the state occupied (Markovian state dependence), and conditioning on $D_{it-1} = d_{it-1}$ means that the transition probabilities may depend on elapsed duration in the spell at time $t-1$ (duration dependence). We refer to (31) as the MSD model's "structural" equation of interest. Assumption (31) leaves unrestricted the transition probabilities associated with the left-censored initial spells that are ongoing at the beginning of the observation period (i.e. those for $t = 1, \ldots, f_i$).

Under assumption (31) and imposing the discrete distribution of unobserved heterogeneity, the likelihood contribution (20) can be written

$$\begin{aligned}
L_i^C = \sum_{k=1}^{K} \pi_k \Bigg[& A^C(f_i, y_{if_i}, \mathbf{x}_{if_i}, \nu_k) \\
& \times \left(\prod_{t=f_i+1}^{T} \mathsf{P}(C_{it} = c_{it} | Y_{it-1} = y_{it-1}, D_{it-1} = d_{it-1}, X_{it} = x_{it}, V_i = \nu_k) \right) \Bigg], \quad (32)
\end{aligned}$$

where

$$A^C(f_i, y_{if_i}, \mathbf{x}_{if_i}, \nu_k) = \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}, V_i = \nu_k)$$
$$\times \left(\prod_{t=2}^{f_i} \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \right). \quad (33)$$

Similarly to the DBR model, the likelihood contribution in (32) has two components: the structural equations contribution of the fresh spells in big parentheses, and the term $A^C(f_i, y_{if_i}, \mathbf{x}_{if_i}, \nu_k)$ which represents the contribution of the left-censored initial spell in progress at the start of the observation period.[15]

As for DBR models, if the data are not left-censored, then the probabilities in (33) may be of interest; otherwise, they are simply necessary nuisance parameters. Again, we shall deal with A^C using Heckman's (1981b) method. Also, if there is no unobserved heterogeneity, the integration in (32) effectively disappears, so the probabilities associated with the fresh spells are nonparametrically identified, and can be estimated by sample analogs. Maximum likelihood estimation based on the fresh spells is therefore consistent for those parameters.

The usefulness of assumption (31) 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 (31) applies. To help alleviate the problem we consider a second assumption that, in addition to (31), the effect of elapsed duration in the spell is constant from time p onwards.[16] (For example, if $p = 1$, the hazard rate is everywhere constant.) To write this

[15] In (33), it is understood that $y_{it} = 1 - y_{if_i}$ for $t = 1, \ldots, f_i - 1$ and $\mathbf{y}_{it-1}$ takes compatible values; i.e. $\mathbf{y}_{it-1} = (1 - y_{if_i}, \ldots, 1 - y_{if_i}, y_{if_i}, y_{if_i+1}, \ldots, y_{it-1})$.

[16] For example, see Stevens (1999).

compactly, define $D^p_{it} = \min(D_{it}, p)$ for $t = 1, \ldots, T$. Then, for fixed $p \geq 1$,[17]

$$\begin{aligned}
&\mathsf{P}(C_{it} = c_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = v_i) \\
&\quad = \mathsf{P}(C_{it} = c_{it} | Y_{it-1} = y_{it-1}, D^p_{it-1} = d^p_{it-1}, X_{it} = x_{it}, V_i = v_i), \\
&\qquad\qquad\qquad t = \min(f_i, p) + 1, \ldots, T. \quad (34)
\end{aligned}$$

The power of constant hazard rates from time p onwards is two-fold. First, only parameters for times $1, \ldots, p$ depend on unobserved variables. In other words, equation (34) implies that all data after either the first observed transition at f_i or time p contribute to identifying and estimating the parameters of interest (as opposed to nuisance parameters). Second, assumption (34) implies we can ignore observed transitions in the first p periods, and still obtain consistent estimates of the structural equation parameters of interest. This entails some loss of precision in estimating the structural parameters, but simplifies modeling the nuisance parameters.

The likelihood function under assumption (34) and disregarding the timing of the first observed transition at time f_i is[18]

$$\begin{aligned}
L^C_i = \sum_{k=1}^{K} \pi_k \Bigg[& A^D(y_{ip}, \mathbf{x}_{ip}, \nu_k) \\
& \times \left(\prod_{t=p+1}^{T} \mathsf{P}(C_{it} = c_{it} | Y_{it-1} = y_{it-1}, D^p_{it-1} = d^p_{it-1}, X_{it} = x_{it}, V_i = \nu_k) \right) \Bigg], \quad (35)
\end{aligned}$$

where $A^D(y_{ip}, \mathbf{x}_{ip}, \nu_k)$ represent the likelihood contribution of y_{ip}; that is, the outcome at time p. This can be expressed as

$$\begin{aligned}
A^D(y_{ip}, \mathbf{x}_{ip}, \nu_k) = {} & \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}, V_i = \nu_k) \\
& \times \left(\prod_{t=2}^{p} \mathsf{P}(Y_{it} = y_{it} | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \right). \quad (36)
\end{aligned}$$

Thus, ignoring transitions in the first p observations means that A^D depends on a single

[17]For simplicity, we assume p is the same for transitions into and out of a given state.

[18]If transitions before time p are included, $A^D(y_{ip}, \mathbf{x}_{ip}, \nu_k)$ becomes $A^D(f_i, y_{i\min(f_i,p)}, \mathbf{x}_{i\min(f_i,p)}, \nu_k)$ and the product is over $t = \min(f_i, p) + 1, \ldots, T$ in (35) and over $t = 2, \ldots, \min(f_i, p)$ in (36).

binary variable, y_{ip} where p is exogenous, rather than the complicated duo $(f_i, y_{i\min(f_i,p)})$. As shown below, this permits a three-equation specification.

Parametric MSD models

As discussed above, we focus on models which assume (31) that the transition probabilities depend on the current state and the elapsed time in the current spell, but not on the history prior to entering that state. We also assume (34) that duration dependence is constant from time p onwards. In particular, to capture duration dependence, we adopt a flexible specification with separate parameters for the first p potential transition times in each state. A commonly used alternative is to specify a quadratic relationship for duration dependence (e.g. Ham and LaLonde, 1996; Beck et al., 1998, 2002).[19] In the case study in Section 3 we set $p = 6$. Within this framework, we consider two models based on the likelihood functions (32)–(33) and (35)–(36), respectively.

The first model, MSD1, is based on the likelihood function (32)–(33) and uses all the data available. Assumption (34) is maintained, but not fully exploited. Specifically, the duration dependence is assumed to be constant after duration p, but this restriction is not imposed across the structural equations and A^C. Again we use Heckman's (1981b) ideas in modeling the A^C component. A fully flexible specification of the approximate reduced form for A^C would involve separate equations for each probability on the right-hand side of (33). Depending on the extent of left-censoring, this may be prohibitive in practice. As a compromise between flexibility and feasibility, we model A^C using three equations representing the probability distribution of the initial outcome, and the initial spell transition probabilities from each state. Thus, the MSD1 model has five equations: a reduced form equation for the initial state, two separate reduced form equations for modeling transitions from the initial spells, and two separate structural equations for

[19] Note that Brown (1975) and Heckman and Borjas (1980) made the point that one can interact covariates with elapsed time in duration models; Jenkins and García-Serrano (2004) provide a rare example.

modeling the transitions from the fresh spells.[20] The model specification is

$$\mathsf{P}(Y_{it}=1|\mathbf{X}_{it}=\mathbf{x}_{it},V_i=\nu_k)=G(\nu_{k1}+\beta_1' x_{it})\equiv G_{it}^{11}(\nu_{k1},\beta_1),\quad t=1, \tag{37}$$

$$\begin{aligned}&\mathsf{P}(C_{it}=1|Y_{it-1}=0,D_{it-1}^p=d_{it-1}^p,X_{it}=x_{it},V_i=\nu_k)\\ &\qquad=G(\nu_{k2}+\beta_2' x_{it}+\textstyle\sum_{r=2}^p\lambda_{2r}\,1(d_{it-1}^p\geq r))\equiv G_{it}^{12}(\nu_{k2},\beta_2,\lambda_2),\\ &\qquad\qquad t=2,\ldots,f_i,\end{aligned} \tag{38}$$

$$\begin{aligned}&\mathsf{P}(C_{it}=1|Y_{it-1}=1,D_{it-1}^p=d_{it-1}^p,X_{it}=x_{it},V_i=\nu_k)\\ &\qquad=G(\nu_{k3}+\beta_3' x_{it}+\textstyle\sum_{r=2}^p\lambda_{3r}\,1(d_{it-1}^p\geq r))\equiv G_{it}^{13}(\nu_{k3},\beta_3,\lambda_3),\\ &\qquad\qquad t=2,\ldots,f_i,\end{aligned} \tag{39}$$

$$\begin{aligned}&\mathsf{P}(C_{it}=1|Y_{it-1}=0,D_{it-1}^p=d_{it-1}^p,X_{it}=x_{it},V_i=\nu_k)\\ &\qquad=G(\nu_{k4}+\beta_4' x_{it}+\textstyle\sum_{r=2}^p\lambda_{4r}\,1(d_{it-1}^p\geq r))\equiv G_{it}^{14}(\nu_{k4},\beta_4,\lambda_4),\\ &\qquad\qquad t=f_i+1,\ldots,T,\end{aligned} \tag{40}$$

$$\begin{aligned}&\mathsf{P}(C_{it}=1|Y_{it-1}=1,D_{it-1}^p=d_{it-1}^p,X_{it}=x_{it},V_i=\nu_k)\\ &\qquad=G(\nu_{k5}+\beta_5' x_{it}+\textstyle\sum_{r=2}^p\lambda_{5r}\,1(d_{it-1}^p\geq r))\equiv G_{it}^{15}(\nu_{k5},\beta_5,\lambda_5),\\ &\qquad\qquad t=f_i+1,\ldots,T.\end{aligned} \tag{41}$$

[20] Even with these restrictions imposed, a five-equation specification can be difficult to estimate, and MSD models are rarely fully specified in practice. In particular, to our knowledge, the initial conditions problem is rarely considered in MSD models, and initial spells are often either dropped from the analysis or modeled using the same specifications as fresh spells. Biewen (2006) and Devicienti (2011) model the initial state but ignore the left-censored spells; Lacroix and Brouillette (2011), Ham and LaLonde (1996), and Eberwein et al. (1997) model the initial and fresh spells separately and have no initial conditions problem to deal with. Also, although Stevens (1999) does not model the initial conditions, she carefully considers the duration dependence associated with fresh spells, and this enables her to (partially) include initial spells within the two-equation specification for fresh spells. Note however, that we did not encounter any problems in maximizing this likelihood function ($K=2$) in the case study.

The corresponding likelihood contribution for individual i is (c.f. (32))

$$
\begin{aligned}
&L_i^{\mathit{MSD1}}(\nu_1,\dots,\nu_K,\pi_1,\dots,\pi_K,\beta_1,\dots,\beta_5,\lambda_2,\dots,\lambda_5)\\
&\quad=\sum_{k=1}^{K}\pi_k\Bigg[G_{i1}^{11}(\nu_{k1},\beta_1)^{y_{i1}}\big(1-G_{i1}^{11}(\nu_{k1},\beta_1)\big)^{1-y_{i1}}\\
&\quad\times\left(\prod_{t=2}^{f_i}G_{it}^{12}(\nu_{k2},\beta_2,\lambda_2)^{c_{it}}\big(1-G_{it}^{12}(\nu_{k2},\beta_2,\lambda_2)\big)^{1-c_{it}}\right)^{1-y_{it-1}}\\
&\quad\times\left(\prod_{t=2}^{f_i}G_{it}^{13}(\nu_{k3},\beta_3,\lambda_3)^{c_{it}}\big(1-G_{it}^{13}(\nu_{k3},\beta_3,\lambda_3)\big)^{1-c_{it}}\right)^{y_{it-1}}\\
&\quad\times\left(\prod_{t=f_i+1}^{T}G_{it}^{14}(\nu_{k4},\beta_4,\lambda_4)^{c_{it}}\big(1-G_{it}^{14}(\nu_{k4},\beta_4,\lambda_4)\big)^{1-c_{it}}\right)^{1-y_{it-1}}\\
&\qquad\qquad\times\left(\prod_{t=f_i+1}^{T}G_{it}^{15}(\nu_{k5},\beta_5,\lambda_5)^{c_{it}}\big(1-G_{it}^{15}(\nu_{k5},\beta_5,\lambda_5)\big)^{1-c_{it}}\right)^{y_{it-1}}\Bigg]. \qquad (42)
\end{aligned}
$$

The dimensions of the parameters are $\beta_q \in \mathbb{R}^{\dim(x)}$, and $\lambda_q = (\lambda_{q2},\dots,\lambda_{qp}) \in \mathbb{R}^{p-1}$ for $q = 1,\dots,5$. Unobserved heterogeneity is represented by a probability π_k and a 5-vector $\nu_k = (\nu_{k1},\dots,\nu_{k5}) \in \mathbb{R}^5$ for $k = 1,\dots,K$. The MSD1 model is relatively flexible and utilizes all available data, but comes at the cost of having to estimate five equations and a large number of parameters.

The second model, MSD2, is based on the likelihood function (35)–(36). We utilize the full power of assumption (34); in particular, for $d_{it-1} \geq p$ we restrict equations (38) and (40) to be the same and equations (39) and (41) to be the same. Furthermore, we ignore the likelihood contributions for the first $p-1$ time periods. This allows us to estimate a three-equation model, which represent the approximate reduced form specification for the probability distribution of the initial state, A^D, and the two transition probabilities out of the state-specific spells. Thus, the likelihood contribution for individual i is

$$
\begin{aligned}
& L_i^{MSD2}(\nu_1,\ldots,\nu_K,\pi_1,\ldots,\pi_K,\beta_1,\beta_4,\beta_5,\lambda_2,\lambda_4,\lambda_5) \\
& \quad = \sum_{k=1}^{K} \pi_k \Bigg[G_{ip}^{11}(\nu_{k1},\beta_1)^{y_{ip}} \big(1 - G_{ip}^{11}(\nu_{k1},\beta_1)\big)^{1-y_{ip}} \\
& \quad \times \left(\prod_{t=p+1}^{T} G_{it}^{14}(\nu_{k4},\beta_4,\lambda_4)^{c_{it}} \big(1 - G_{it}^{14}(\nu_{k4},\beta_4,\lambda_4)\big)^{1-c_{it}} \right)^{1-y_{it-1}} \\
& \qquad\qquad\qquad \times \left(\prod_{t=p+1}^{T} G_{it}^{15}(\nu_{k5},\beta_5,\lambda_5)^{c_{it}} \big(1 - G_{it}^{15}(\nu_{k5},\beta_5,\lambda_5)\big)^{1-c_{it}} \right)^{y_{it-1}} \Bigg], \qquad (43)
\end{aligned}
$$

where G_{it}^{11} is defined as in (37) but with $t = p$, and G_{it}^{14} and G_{it}^{15} are defined as in (40) and (41) but with $t = p+1,\ldots,T$.[21]

2.7 Nesting of DBR and MSD models

The structural DBR1 equation is obviously a special case of the structural DBR2 equation. Moreover, both the DBR1 and the DBR2 equations are special cases of the structural MSD1 and MSD2 equations. To see this for the DBR1 model, note that by symmetry of the logistic function the DBR1 equation (25) implies

$$
\begin{aligned}
& \mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\
& \qquad\qquad = \begin{cases} G(\nu_{k2} + \beta_2' x_{it}) & \text{if } y_{it-1} = 0, \\ G(-\nu_{k2} - \beta_2' x_{it} - \gamma_2) & \text{if } y_{it-1} = 1, \end{cases} \qquad t = 2,\ldots,T, \qquad (44)
\end{aligned}
$$

whereas the MSD1/MSD2 equations can be written

$$
\begin{aligned}
& \mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\
& \quad = \begin{cases} G(\nu_{k4} + \beta_4' x_{it} + \sum_{r=2}^{p} \lambda_{4r}\, 1(d_{it-1}^{p} \geq r)) & \text{if } y_{it-1} = 0, \\ G(\nu_{k5} + \beta_5' x_{it} + \sum_{r=2}^{p} \lambda_{5r}\, 1(d_{it-1}^{p} \geq r)) & \text{if } y_{it-1} = 1, \end{cases} \qquad t = p+1,\ldots,T. \qquad (45)
\end{aligned}
$$

[21] If $p = 1$, implying that there is no duration dependence, the likelihood function for the MSD2 model involves all data. Furthermore, equations (38) and (40), and equations (39) and (41), are the same.

Matching coefficients shows that the DBR1 model arises as a special case of the MSD models when there is no duration dependence ($p = 1$), and the effects of observed and unobserved heterogeneity are opposite in the two hazard rates; i.e. $\beta_4 + \beta_5 = 0$, and $\nu_{k4} + \nu_{k5} = \textit{constant}$ for $k = 1, \dots, K$. The constant here corresponds to the state dependence parameter, γ_2, in the DBR1 model.

To show that the DBR2 model is also a special case, note that the DBR2 structural equation (29) implies

$$\begin{aligned}
&\mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\
&= \begin{cases}
G(\nu_{k3} + \beta_3' x_{it} + \gamma_{32}) & \text{if } y_{it-1} = 0,\ y_{it-2} = 1, \\
G(\nu_{k3} + \beta_3' x_{it}) & \text{if } y_{it-1} = 0,\ y_{it-2} = 0, \\
G(-\nu_{k3} - \beta_3' x_{it} - \gamma_{31}) & \text{if } y_{it-1} = 1,\ y_{it-2} = 0, \\
G(-\nu_{k3} - \beta_3' x_{it} - \gamma_{31} - \gamma_{32} - \gamma_{33}) & \text{if } y_{it-1} = 1,\ y_{it-2} = 1,
\end{cases} \\
&\hspace{20em} t = 2, \dots, T. \quad (46)
\end{aligned}$$

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 $d^p_{it-1} \geq 2$ if and only if $y_{it-1} = y_{it-2}$, the MSD1/MSD2 structural equations for $p \geq 2$ can be written

$$\begin{aligned}
&\mathsf{P}(C_{it} = 1 | \mathbf{Y}_{it-1} = \mathbf{y}_{it-1}, \mathbf{X}_{it} = \mathbf{x}_{it}, V_i = \nu_k) \\
&= \begin{cases}
G(\nu_{k4} + \beta_4' x_{it} + \lambda_{42}(1 - y_{it-2}) + \sum_{r=3}^{p} \lambda_{4r}\, 1(d^p_{it-1} \geq r)) & \text{if } y_{it-1} = 0, \\
G(\nu_{k5} + \beta_5' x_{it} + \lambda_{52} y_{it-2} + \sum_{r=3}^{p} \lambda_{5r}\, 1(d^p_{it-1} \geq r)) & \text{if } y_{it-1} = 1,
\end{cases} \\
&\hspace{20em} t = p+1, \dots, T. \quad (47)
\end{aligned}$$

Matching coefficients shows that the DBR2 model arises as a special case when there is no duration dependence after one period ($p = 2$), and the effects of observed and unobserved heterogeneity are opposite in the two hazard rates.[22]

Third- and higher-order DBR models are generally not nested within the MSD framework as they involve a kind of lagged duration dependence.

3 Case study

We now apply each of the methods above to an empirical context, that of poverty persistence analyzed using multi-spell duration models by Stevens (1999). Our focus is to use this case study as an empirical setting to compare the results obtained from reasonably standard prototypical DBR and MSD models, and illustrate the differences associated with them, 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 analysis data consists of an extract of 5,248 individuals over the 20 years 1970–89 from the Panel Study of Income Dynamics (PSID).[23] Each individual's poverty status is determined by 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.[24] 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.[25] 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.

[22] Specifically, in terms of (46) and (47) the four intercepts satisfy $\nu_{k3} + \gamma_{32} = \nu_{k4}$, $\nu_{k3} = \nu_{k4} + \lambda_{42}$, $-\nu_{k3} - \gamma_{31} = \nu_{k5}$, $-\nu_{k3} - \gamma_{31} - \gamma_{32} - \gamma_{33} = \nu_{k5} + \lambda_{52}$; the slope parameters satisfy $\beta_3 = \beta_4 = -\beta_5$; and $\lambda_{4r} = 0$ and $\lambda_{5r} = 0$ for $r = 3, \ldots, p$.

[23] 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.

[24] See Stevens (1999) for more details of this and other data issues.

[25] This criteria is used as a proxy to identify the poverty at-risk population, and follows Stevens (1999).

Tables 1 and 2 present descriptive statistics of the sample, summarized along three dimensions: at the person-year level, at the person-level, and at the person-spell level. The first panel of Table 1 shows the means of the covariates used in the models for the full sample of observations as well as for the subsamples defined by individuals' initial poverty state. On average, individuals were in poverty 35 percent of the time, and had an 18 percent chance of a poverty transition in any year. Individuals in poverty in the first year experienced poverty about 52 percent of the time, compared to 21 percent for individuals not initially in poverty. The second panel of Table 1 shows that, on average, individuals experienced 3.35 poverty transitions (equivalently, 4.35 poverty and non-poverty spells) over the observation period.

In Table 2, we briefly summarize the numbers of poverty and non-poverty spells, and the average durations of these spells. The first rows show that non-poverty spells are more prevalent than poverty spells and, on average, the non-poverty spells are longer (5.6 years compared to 3.4). The subsequent rows show these relative patterns hold for both initial (left-censored) and fresh (including right-censored) spells. The average observed durations of initial spells are roughly twice that of fresh spells (7.1 years compared to 3.8 years for poverty and non-poverty spells).

3.2 Estimation results

We present the estimates of the DBR and MSD models in Tables 3 and 4. Table 3 contains estimates for three DBR models: a first-order DBR model without unobserved heterogeneity, DBR0; and the first- and second-order DBR models with two discrete points of unobserved heterogeneity, DBR1 and DBR2. Table 4 contains the estimates for two MSD models: a five-equation MSD model with two random effects mass points, MSD1, and a three-equation MSD model which exploits the assumption of constant hazard rates after 6 years to estimate common equations for initial and fresh spells, MSD2. 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.

We briefly discuss the DBR model results in Table 3. First, the estimates of the coefficients in the structural equations are consistent across the models: there is strong evidence of positive state dependence associated with poverty status; and non-prime aged (25–54) individuals, and those in female and black headed households, are more likely to be in poverty than other households. Second, although the approximate reduced form equations for the initial conditions do not have obvious interpretation, the coefficients on the covariates in these equations are generally of the same signs as those in the structural equations. (The only exceptions being some of the age-variables coefficients, but these are statistically insignificant in all cases.)

We next discuss the MSD model results, presented in Table 4. The main (structural) estimates of interest are those in the fresh-spell equations for poverty entry and exit, and these indicate some substantial differences with the DBR models. The parameter estimates for the MSD1 and MSD2 models are broadly similar.[26] Our subsequent discussion of the MSD models will focus on the MSD1 specification.

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

The second important restriction in the DBR models is that the order of state dependence implies that, for spell-durations longer than that order, the impact on the probability of a transition occurring should be zero. 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

[26]Eyeballing the estimates across the two specifications, the coefficients on 6+ years duration, for being aged 18–24, and for having a black head of household in the poverty entry equations, and the coefficients on being aged 0–5 in the exit equations are noticeably different. The difference in the 6+ years duration coefficients may be an artifact of the sample selection which requires nearly all individuals to be poor for at least one period.

that the hazard 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.[27]

To explore the respective contributions of relaxing the DBR models' strict state dependence and heterogeneity restrictions, we estimated a variety of model specifications between the DBR1 and MSD1 models. Table 5 presents a summary of the results. First, as discussed above, the DBR1 model is equivalent to an MSD model in which there is no duration dependence, and the effects of both the observed covariates and unobserved heterogeneity are opposite on poverty entry and exit. This is the first model (A) summarized in Table 5.

The second model (B) summarized in Table 5 relaxes the heterogeneity restriction on both the covariates and unobserved heterogeneity, but maintain the constant hazard rate restriction. This essentially introduces a third equation to allow separate specifications for the entry and exit transitions. The Wald statistic for the hypothesis that these heterogeneity restrictions are valid clearly rejects that hypothesis (118.6 with 7 degrees of freedom).[28]

The third model (C) summarized in Table 5 is the DRB2 model, which allows for limited duration dependence ($p = 2$) but (re-)imposes the entry and exit heterogeneity restrictions. The improvement in the quasi-likelihood value is huge. The Wald statistic overwhelmingly rejects the DBR1 model against the DBR2 (817.5 with 11 degrees of freedom). The fourth model (D) relaxes the heterogeneity restrictions in the DBR2 model; these are again very clearly rejected by a Wald test (167.2 with 14 degrees of freedom).

[27] The Vuong (1989) test statistic for comparing the DBR1 and MSD1 models is 32.1 in favor of the MSD1 model. (The null is that both models are misspecified but fit equally well, and the statistic is asymptotically standard normally distributed.)

[28] Browning and Carro (2010) similarly reject the restriction of opposite effects in a first-order DBR model, although their focus is on showing heterogeneous state dependence effects that vary with the observable characteristics.

Models (E) and (F) allow for further duration dependence ($p = 6$), with and without restricting the heterogeneity effects to be opposite in the entry and exit equations. Assumption (34) is fully imposed in these models; that is, for $t \geq p$ the transition probabilities out of the left-censored initial spells are modeled by the structural equations. Wald tests strongly reject the simpler DBR2 model (184.9 with 16 degrees of freedom) and the heterogeneity restrictions (199.8 with 14 degrees of freedom).

Finally, the last row in Table 5 summarizes the MSD1 model, which is the same specification as model (F) except that Assumption (34) is not fully exploited. (The initial spells continue to be modeled separately from the structural spells even after p periods.) The MSD1 model does not nest the other specifications and hence a Wald test is not possible. However, the improvement in the log quasi-likelihood value is huge and, as mentioned, the Vuong test also strongly prefers the MSD1 over the DBR1 specification.

In summary, the estimation results presented here clearly show that the DBR models are severely rejected against the MSD alternatives. First, the Markovian state dependence implied by the DBR models is too strong against the duration dependence alternative of the MSD models. Second, the restriction of opposite heterogeneity effects into and out of poverty transitions in the standard DBR models is also strongly rejected. These conclusions are consistent with the results by Bhuller et al. (2014).

3.3 In sample prediction

As well as comparing the estimates of the models, we also compare how they fit the data in the sense of their respective within-sample predictions. For example, it may be that, although the DBR models are rejected in favor of the MSD alternative, the predictive fit of the models may be substantially similar. For this purpose, we compare summary statistics of the actual data and model predictions, presented in Tables 6, 7, and 8.

In Table 6, we present summaries using two-way frequency tables of the number of years that a person is poor and the number of transitions that occur between poverty and non-poverty, for each of the actual data and the predictions based on the first- and second-order DBR models with unobserved heterogeneity and the five-equation MSD model.[29]

The first panel of Table 6 summarizes the actual poverty experience of individuals over the observation period. This shows that there is substantial variation in poverty experience around the seven-year average: about 5 percent of the sample had no spells of poverty,[30] 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: about 8 percent experience no transitions (corresponding to those who were either never or always poor), while over 40 percent had at least 4 transitions.

The next three panels in Table 6 present analogous summaries of the predictions from the first- and second-order DBR models, and the MSD1 model respectively. First, the average incidence of poverty, or equivalently the number of years poor, is predicted well by each of the models (not shown in the table): compared to the actual average of 0.353 (7.06 years over the 20 year sample), the DBR1 model average prediction is 0.350 (7.00 years), while the DBR2 model's average is 0.351 (7.03 years) and the MSD1 model's average is 0.350 (7.01 years). The frequency distribution of the MSD1 model's predicted years poor is noticeably closer to the actual distribution than those of the DBR models. However, the models all substantially overpredict the number of individuals who have no poverty experience, and underpredict the number with a single year; in addition, the DBR models also substantially underpredict the number of individuals who are always in poverty.

Second, the models also accurately predict the average number of transitions (not shown in the table): the DBR1, DBR2 and MSD1 model average predicted transitions are 3.36, 3.41 and 3.36 respectively, compared to 3.35 actual transitions. However, associated

[29]In order to obtain manageable summaries and limit the extent of small cell frequencies, we group the number of years poor as 0, 1, 2–5, 6–10, 11–15, 16–19, and 20 years, and the number of transitions 0, 1, 2, 3, 4+ even (so the initial and final states are the same), and 5+ odd (so the initial and final states are different). In this two-way tabulation, some cells are necessarily null; e.g. individuals who are either *never* or *always* poor experience no transitions; similarly, individuals who are poor in only 1 year must experience either 1 or 2 transitions. The predictions from each of the models are based on 20 simulations per individual taking the covariates as given.

[30]Recall this is in the at-risk population who experienced some poverty between survey years 1968 and 1989.

with overpredicting the zero poverty incidence, the models overpredict the number of cases with zero transitions and dramatically underpredict the incidence of 1–3 transition cases.

For indicative purposes, we have constructed Pearson goodness-of-fit statistics for each of the models based on the tables of actual and predicted frequencies in Table 6. Note that the distribution of these statistics 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).[31] The MSD1 model's goodness-of-fit statistic (98.9 with 23 degrees of freedom) is substantially lower than those of the two DBR models. Thus, although this implies the MSD1 model does not provide an adequate statistical fit to the data using conventional significance levels, the relative magnitudes are consistent with the MSD1 model fitting substantially better than the two DBR models; in turn, the DBR2 model fits better than the DBR1 model.

In Table 7 we present a different summary of the actual and predicted poverty experiences from the DBR1 and MSD1 models.[32] The table shows the frequency distribution of the number of distinct spells experienced over the observation period separately by the initial state. The actual experiences include up to 15 separate spells, while the maximum number of spells predicted by the DBR1 and the MSD1 models is 17. Table 7 shows that the tendency for the models to overpredict the frequency of single spells and underpredict the frequencies of 2 and 3 poverty spells is especially strong for those whose initial state is not-in-poverty. On the other hand, the models fit a little better for those who are poor initially; this is particularly true for the MSD1 model.

In Table 8 we present the average durations of the actual and predicted spells. The averages of the MSD1 model predictions are again closer to the actual spell average

[31] 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 m and $m-l$ degrees of freedom, where m 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. m).

[32] We exclude the DBR2 model predictions here as these are comparatively similar to those of the DBR1 model.

durations than those of the DBR1 model. Thus, these prediction results are consistent with the estimation results indicating the MSD model fits better than the DBR models.

3.4 Out of sample prediction

To gauge whether policy recommendations are sensitive to the model specification, we discuss the results a simple experiment 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 next 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 9 summarizes the results of this exercise. The first row contains the full-sample summary, and subsequent rows the summary for 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.[33] 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.[34]

In the next pair of columns, we present the average number of years poor over the 10 year simulation time frame for the DBR1 and MSD1 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

[33] The latter is the average person-average, which gives equal weight to each person who has a fresh spell.

[34] 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.

predicts about 25 percent lower poverty experience over the following 10 years than the DBR1 model (2.4 versus 3.2 years) on average, but 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 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.

This paper argues that the two approaches should not be viewed as separate. We show that they both can be viewed as special cases of a general analytical framework. Analysts therefore have a choice of which approach to use, or may choose to combine features from either. In particular, as DBR models are more widely used than MSD models, this paper suggests that DBR modeling can benefit from explicitly considering the implications of the specifications for the probabilities of transitioning between states.

The case study analysis of poverty experiences provided some clear conclusions regarding the relative efficacy of the models. The MSD models fit the data far better than the first- and second-order DBR models. 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 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 limitations of the popular DBR model approach and highlight the importance of considering more flexible alternatives.

Appendix A Spell-based representations

As mentioned, in duration analysis and event history approaches, the data are often represented as transition times or spells instead of a sequence of state indicators. In this literature, the data are often considered as observations in continuous time, in which case the spell-based representations are the only ones practical. In discrete time, spell-based representations may be computationally efficient if covariates are constant within each spell. Time-based representations may be more convenient if the covariates are time-varying rather than spell-varying.

In this appendix, we show that spell-based and time-based data representations are equivalent. The nonparametric likelihood functions are also equivalent in the sense that there is a one-to-one relationship between the different parameter sets. The final subsection discusses covariates.

A.1 Data representation

Let J_i denote the number of transitions observed between time 1 and time T; that is,

$$J_i = \sum_{t=2}^{T} C_{it}. \tag{48}$$

Let Z_{ij} for $j = 1, \ldots, J_i$ denote the observed times of change of state (spell endings). Then $(Y_{i1}, Z_{i1}, \ldots, Z_{iJ_i}, T)$ is an equivalent complete representation of the data. To see this, note that the transition times can be defined recursively (assuming $J_i > 0$) by[35]

$$\begin{aligned} Z_{i1} &= \inf\{t \in \mathbb{N} : 1 \le t < T,\ Y_{i1} \ne Y_{it+1}\}, \\ Z_{ij} &= \inf\{t \in \mathbb{N} : Z_{j-1} + 1 \le t < T,\ Y_{iZ_{j-1}+1} \ne Y_{it+1}\}, \quad j = 2, \ldots, J_i. \end{aligned} \tag{49}$$

Conversely, the state indicators can be recovered from the transition times by

$$Y_{it} = \begin{cases} 1 - Y_{it-1} & \text{if } \exists j \in \mathbb{N} : 1 \le j \le J_i,\ Z_{ij} = t - 1, \\ Y_{it-1} & \text{otherwise,} \end{cases} \qquad t = 2, \ldots, T, \tag{50}$$

or by

$$Y_{it} = \left(Y_{i1} + \sum_{j=1}^{J_i} 1(Z_{ij} < t) \right) \bmod 2, \quad t = 2, \ldots, T. \tag{51}$$

For simplicity, in this paper we assume all histories are right-censored at time T. Since we do not know the state at time $T + 1$, we therefore do not know whether or not there is a transition at time T.

The data can also be represented as a panel of durations. Let D_{ij} for $j = 1, \ldots, J_i$ denote the duration of the jth spell. Formally,

$$D_{ij} = Z_{ij} - Z_{ij-1}, \quad j = 2, \ldots, J_i. \tag{52}$$

If $J_i > 0$, we may also define the (possibly left-censored) duration at the beginning of the observation period by $D_{i1} = Z_{i1}$ and the (possibly right-censored) duration at the end of the observation period by $D_{iJ_i+1} = T - Z_{iJ_i}$. If $J_i = 0$, define $D_{i1} = T$. Then $(Y_{i1}, D_{i1}, \ldots, D_{iJ_i}, D_{iJ_i+1})$ is an equivalent representation of the data.

Example Suppose $T = 4$ and the state occupancy indicators are $Y_{i1} = 0$, $Y_{i2} = 0$, $Y_{i3} = 1$, and $Y_{i4} = 1$. Then there is one transition and two spells, which can be represented in four different ways: the first representation is $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}) = (0, 0, 1, 1)$, the second

[35]Note that for $j = 1, \ldots, J_i$ we have $Z_{ij} = t \Rightarrow C_{it+1} = 1$, and similarly for $t = 2, \ldots, T$, we have $C_{it} = 1 \Rightarrow \exists j \in \mathbb{N} : 1 \le j \le J_i,\ Z_{ij} = t - 1$.

is $(Y_{i1}, C_{i2}, C_{i3}, C_{i4}) = (0, 0, 1, 0)$, the third is $(Y_{i1}, Z_{i1}, Z_{i2}) = (0, 2, 4)$, and the fourth representation is $(Y_{i1}, D_{i1}, D_{i2}) = (0, 2, 2)$.

A.2 Parameterization

The likelihood contribution in the MSD approach can also be written in a spell-based form instead of the time-based form given in (8).[36] The notation becomes slightly more involved, since the number of spells may vary across individuals. First, let $\mathbf{Z}_{ij}$ denote the random outcome history at the time of the jth transition (the end of the jth spell) for individual i; that is, define $\mathbf{Z}_{i0} = (Y_{i1})$ and $\mathbf{Z}_{ij} = (Y_{i1}, Z_{i1}, \ldots, Z_{ij})$ for $j = 1, \ldots, J_i$. Let $\mathbf{z}_{i0} = (y_{i1})$ and $\mathbf{z}_{ij} = (y_{i1}, z_{i1}, \ldots, z_{ij})$ denote the observed history. Second, let $\mathcal{Z}_{t-1}^{j-1}$ denote the space of possible prior histories when spell j is in progress at time t. If there has been no transitions before time t, we have $\mathcal{Z}_{t-1}^{0} = \{0, 1\}$ for $t = 1, \ldots, T-1$, and if there has been one previous transition, then $\mathcal{Z}_{t-1}^{1} = \{0, 1\} \times \{1, \ldots, t-1\}$ for $t = 2, \ldots, T-1$. For $j = 2, \ldots, t-1$ and $t = 2, \ldots, T-1$, the space of possible prior histories is $\mathcal{Z}_{t-1}^{j-1} = \{0, 1\} \times \{1, \ldots, t-j\} \times \cdots \times \{j-1, \ldots, t-1\}$. Let $\mathbf{z}_{j-1}$ with no subscript i denote a generic element of $\mathcal{Z}_{t-1}^{j-1}$. Then the conditional probability of beginning in state 1 given the history and the hazard rates at each time t are defined as[37]

$$
\begin{aligned}
&\chi = \mathsf{P}(Y_{i1} = 1), \\
&\varphi_t(\mathbf{z}_{j-1}) = \mathsf{P}(Z_{ij} = t | Z_{ij} \geq t, \mathbf{Z}_{ij-1} = \mathbf{z}_{j-1}), \\
&\qquad\qquad\qquad\qquad \mathbf{z}_{j-1} \in \mathcal{Z}_{t-1}^{j-1}, \quad j = 1, \ldots, t, \quad t = 1, \ldots, T-1.
\end{aligned} \tag{53}
$$

Of course, there are also $2^T - 1$ distinct probabilities in this representation. This can be verified as follows: given T and j with $0 \leq j \leq T-1$, there is $T-1$ choose j possible transition times; by the binomial formula there are then a total of $\sum_{j=0}^{T-1} \binom{T-1}{j} = 2^{T-1}$ possible transition times; each of which can begin in either of the two states, so 2^T possible outcomes; and one probability is determined by the adding-up constraint, so

[36] The time-based form is standard in the continuous-time literature, see e.g. Honoré (1993) and Horowitz and Lee (2004).

[37] Allison (1982, p92) defined the discrete-time hazard rate for repeated events, but did not provide the likelihood function.

$2^T - 1$ probabilities.

Example For $\tau = 3$, there are 7 parameters depending on time t and the history prior to t; namely,

$$\begin{aligned}
\chi &= \mathsf{P}(Y_{i1} = 1), \\
\varphi_1(0) &= \mathsf{P}(Z_{i1} = 1 | Z_{i1} \geq 1, (Y_{i1}) = (0)) \text{ with } \mathbf{z}_0 = (0) \in \mathcal{Z}_0^0, \\
\varphi_1(1) &= \mathsf{P}(Z_{i1} = 1 | Z_{i1} \geq 1, (Y_{i1}) = (1)) \text{ with } \mathbf{z}_0 = (1) \in \mathcal{Z}_0^0, \\
\varphi_2(0) &= \mathsf{P}(Z_{i1} = 2 | Z_{i1} \geq 2, (Y_{i1}) = (0)) \text{ with } \mathbf{z}_0 = (0) \in \mathcal{Z}_1^0, \\
\varphi_2(1) &= \mathsf{P}(Z_{i1} = 2 | Z_{i1} \geq 2, (Y_{i1}) = (1)) \text{ with } \mathbf{z}_0 = (1) \in \mathcal{Z}_1^0, \\
\varphi_2(0,1) &= \mathsf{P}(Z_{i2} = 2 | Z_{i2} \geq 2, (Y_{i1}, Z_{i1}) = (0,1)) \text{ with } \mathbf{z}_1 = (0,1) \in \mathcal{Z}_1^1, \\
\varphi_2(1,1) &= \mathsf{P}(Z_{i2} = 2 | Z_{i2} \geq 2, (Y_{i1}, Z_{i1}) = (1,1)) \text{ with } \mathbf{z}_1 = (1,1) \in \mathcal{Z}_1^1.
\end{aligned}$$

For $\tau = 4$, there are additionally 8 parameters; namely,

$$\begin{aligned}
\varphi_3(0) &= \mathsf{P}(Z_{i1} = 3 | Z_{i1} \geq 3, (Y_{i1}) = (0)) \text{ with } \mathbf{z}_0 = (0) \in \mathcal{Z}_2^0, \\
\varphi_3(1) &= \mathsf{P}(Z_{i1} = 3 | Z_{i1} \geq 3, (Y_{i1}) = (1)) \text{ with } \mathbf{z}_0 = (1) \in \mathcal{Z}_2^0, \\
\varphi_3(0,1) &= \mathsf{P}(Z_{i2} = 3 | Z_{i2} \geq 3, (Y_{i1}, Z_{i1}) = (0,1)) \text{ with } \mathbf{z}_1 = (0,1) \in \mathcal{Z}_2^1, \\
\varphi_3(1,1) &= \mathsf{P}(Z_{i2} = 3 | Z_{i2} \geq 3, (Y_{i1}, Z_{i1}) = (1,1)) \text{ with } \mathbf{z}_1 = (1,1) \in \mathcal{Z}_2^1, \\
\varphi_3(0,2) &= \mathsf{P}(Z_{i2} = 3 | Z_{i2} \geq 3, (Y_{i1}, Z_{i1}) = (0,2)) \text{ with } \mathbf{z}_1 = (0,2) \in \mathcal{Z}_2^1, \\
\varphi_3(1,2) &= \mathsf{P}(Z_{i2} = 3 | Z_{i2} \geq 3, (Y_{i1}, Z_{i1}) = (1,2)) \text{ with } \mathbf{z}_1 = (1,2) \in \mathcal{Z}_2^1, \\
\varphi_3(0,1,2) &= \mathsf{P}(Z_{i3} = 3 | Z_{i3} \geq 3, (Y_{i1}, Z_{i1}, Z_{i2}) = (0,1,2)) \text{ with } \mathbf{z}_2 = (0,1,2) \in \mathcal{Z}_2^2, \\
\varphi_3(1,1,2) &= \mathsf{P}(Z_{i3} = 3 | Z_{i3} \geq 3, (Y_{i1}, Z_{i1}, Z_{i2}) = (1,1,2)) \text{ with } \mathbf{z}_2 = (1,1,2) \in \mathcal{Z}_2^2.
\end{aligned}$$

The likelihood contribution of individual i can be built up by considering the initial state and each subsequent transition separately. To simplify the notation, it is customary to state the likelihood using a "latent" variable, Z_{iJ_i+1}, which represents the $(J_i + 1)$th transition that would have been observed, except for right-censoring. We then have[38]

$$L_i^Z = \chi^{y_{i1}}(1-\chi)^{1-y_{i1}}\left(\prod_{j=1}^{j_i} \mathsf{P}(Z_{ij}=z_{ij}|\mathbf{Z}_{ij-1}=\mathbf{z}_{ij-1})\right)\mathsf{P}(Z_{ij_i+1}\geq T|\mathbf{Z}_{ij_i}=\mathbf{z}_{ij_i}), \quad (54)$$

where

$$\mathsf{P}(Z_{ij}=z_{ij}|\mathbf{Z}_{ij-1}=\mathbf{z}_{ij-1}) = \varphi_{z_{ij}}(\mathbf{z}_{ij-1})\prod_{t=z_{ij-1}+1}^{z_{ij}-1}(1-\varphi_t(\mathbf{z}_{ij-1})), \quad j=1,\ldots,j_i, \quad (55)$$

and

$$\mathsf{P}(Z_{ij_i+1}\geq T|\mathbf{Z}_{ij_i}=\mathbf{z}_{ij_i}) = \prod_{t=z_{ij_i}+1}^{T-1}(1-\varphi_t(\mathbf{z}_{ij_i})). \quad (56)$$

In (54), the first term on the right-hand side is the contribution of the initial state, the term in large parentheses is the contribution of the j_i observed transitions, and the last term is the contribution of the fact that no event took place between z_{ij_i} and T.

The two representations of the MSD likelihood contributions (8) and (54) are of course equivalent. In particular, the parameters are one-to-one and the likelihood values are identical. To verify the first claim, given t and $\mathbf{y}_t \in \{0,1\}^t$, arguments similar to those given in Section 2.1 can be used to deduce $\mathbf{z}_{j-1} \in \mathcal{Z}_{t-1}^{j-1}$, where either $j=1$ and $0<t\leq z_1$ or $j>1$ and $z_{j-1}<t\leq z_j$. Conversely, given t and $\mathbf{z}_{j-1}\in\mathcal{Z}_{t-1}^{j-1}$, it is straightforward to deduce $\mathbf{y}_t\in\{0,1\}^t$. Therefore, given t and j and compatible histories $\mathbf{y}_t\in\{0,1\}^t$ and $\mathbf{z}_{j-1}\in\mathcal{Z}_{t-1}^{j-1}$, we have the one-to-one relationship between parameters

$$z_{j-1}<t\leq z_j \;\Rightarrow\; \varphi_t(\mathbf{z}_{j-1}) = \xi_{t+1}(\mathbf{y}_t). \quad (57)$$

Intuitively, the conditional probability of spell j ending at t given prior history is the same as the conditional probability of a transition between t and $t+1$; or in other words the hazard rates can be expressed in terms of C_{it}s or Z_{ij}s.

[38]It is possible to state the likelihood without the use of a latent variable, by noting that the probability of no events taking place between z_{ij_i} and T is the same as $\mathsf{P}(J_i=j_i|\mathbf{Z}_{ij_i}=\mathbf{z}_{ij_i}, T=T)$.

To verify that the likelihood values are identical, $L_i^C = L_i^Z$, note that (assuming $j_i > 0$)

$$
\begin{aligned}
L_i^Z &= \chi^{y_{i1}}(1-\chi)^{1-y_{i1}}\Bigg(\prod_{j=1}^{j_i}\prod_{t=z_{ij-1}+1}^{z_{ij}} \varphi_t(\mathbf{z}_{ij-1})^{1(t=z_{ij})}(1-\varphi_t(\mathbf{z}_{ij-1}))^{1(t\neq z_{ij})}\Bigg)\\
&\qquad\qquad \times\Bigg(\prod_{t=z_{ij_i}+1}^{T-1} 1-\varphi_t(\mathbf{z}_{ij_i})\Bigg)\\
&= \chi^{y_{i1}}(1-\chi)^{1-y_{i1}}\\
&\qquad\qquad \times\Bigg(\prod_{j=1}^{j_i}\prod_{t=1}^{T-1} 1(z_{ij-1} < t \le z_{ij})\varphi_t(\mathbf{z}_{ij-1})^{1(t=z_{ij})}(1-\varphi_t(\mathbf{z}_{ij-1}))^{1(t\neq z_{ij})}\Bigg)\\
&\qquad\qquad \times\Bigg(\prod_{t=1}^{T-1} 1(z_{ij_i} < t \le T-1)(1-\varphi_t(\mathbf{z}_{ij_i}))\Bigg)\\
&= \chi^{y_{i1}}(1-\chi)^{1-y_{i1}}\Bigg(\prod_{t=1}^{T-1}\prod_{j=1}^{j_i} 1(z_{ij-1} < t \le z_{ij})\xi_{t+1}(\mathbf{y}_{it})^{1(t=z_{ij})}(1-\xi_{t+1}(\mathbf{y}_{it}))^{1(t\neq z_{ij})}\Bigg)\\
&\qquad\qquad \times\Bigg(\prod_{t=1}^{T-1} 1(z_{ij_i} < t \le T-1)(1-\xi_{t+1}(\mathbf{y}_{it}))\Bigg)\\
&= \chi^{y_{i1}}(1-\chi)^{1-y_{i1}}\prod_{t=1}^{T-1}\xi_{t+1}(\mathbf{y}_{it})^{c_{it+1}}(1-\xi_{t+1}(\mathbf{y}_{it}))^{1-c_{it+1}}\\
&= L_i^C.
\end{aligned} \tag{58}
$$

This shows that the likelihood contribution in the multi-spell duration approach can be expressed equivalently either in a j- or a t-dimension.

Example Suppose $Y_{i1} = 0$, $Y_{i2} = 0$, $Y_{i3} = 1$, and $Y_{i4} = 1$ with $T = 4$. Then

$$
\begin{aligned}
L_i^Y &= \mathsf{P}(Y_{i1}=0)\mathsf{P}(Y_{i2}=0|\mathbf{Y}_{i1}=(0))\\
&\qquad\qquad \times \mathsf{P}(Y_{i3}=1|\mathbf{Y}_{i2}=(0,0))\mathsf{P}(Y_{i4}=1|\mathbf{Y}_{i3}=(0,0,1)),
\end{aligned} \tag{59}
$$

while

$$
\begin{aligned}
L_i^Z &= \mathsf{P}(Y_{i1}=0)\big(1-\mathsf{P}(Z_{i1}=1|Z_{i1}\ge 1,\mathbf{Z}_{i0}=(0))\big)\\
&\qquad \times \mathsf{P}(Z_{i1}=2|Z_{ij}\ge 2,\mathbf{Z}_{i0}=(0))\big(1-\mathsf{P}(Z_{i2}=3|Z_{i2}\ge 3,\mathbf{Z}_{i1}=(0,1))\big).
\end{aligned} \tag{60}
$$

Note that

$$\mathsf{P}(Y_{i2} = 0 | \mathbf{Y}_{i1} = (0)) = 1 - \mathsf{P}(Z_{i1} = 1 | Z_{i1} \geq 1, \mathbf{Z}_{i0} = (0)), \tag{61}$$

$$\mathsf{P}(Y_{i3} = 1 | \mathbf{Y}_{i2} = (0, 0)) = \mathsf{P}(Z_{i1} = 2 | Z_{ij} \geq 2, \mathbf{Z}_{i0} = (0)), \tag{62}$$

$$\mathsf{P}(Y_{i4} = 1 | \mathbf{Y}_{i3} = (0, 0, 1)) = 1 - \mathsf{P}(Z_{i2} = 3 | Z_{i2} \geq 3, \mathbf{Z}_{i1} = (0, 1)). \tag{63}$$

The two nonparametric likelihood representations are therefore equivalent.

A.3 Likelihood contribution with covariates

Suppose the covariates are constant within each spell and only vary between spells. For $j = 0, \ldots, j_i + 1$, let X^*_{ij} denote the vector of spell-constant covariates, and let $\mathbf{X}^*_{ij}$ and $\mathbf{x}^*_{ij}$ denote the random and observed covariate histories up to (and including) spell j.[39] Then the likelihood contribution for individual i in the spell-based form becomes

$$\begin{aligned} L_i^Z = \chi^{y_{i1}}(1-\chi)^{1-y_{i1}} & \left(\prod_{j=1}^{j_i} \mathsf{P}(Z_{ij} = z_{ij} | \mathbf{Z}_{ij-1} = \mathbf{z}_{ij-1}, \mathbf{X}^*_{ij-1} = \mathbf{x}^*_{ij-1}) \right) \\ & \times \mathsf{P}(Z_{ij_i+1} \geq T | \mathbf{Z}_{ij_i} = \mathbf{z}_{ij_i}, \mathbf{X}^*_{ij_i} = \mathbf{x}^*_{ij_i}), \end{aligned} \tag{64}$$

It can be shown that this is equivalent to (12) and (14).

Conceptually, it makes no difference if the covariates are not spell-constant, but time-varying; however, the expression for the likelihood contribution is more complicated. With time-varying covariates, the likelihood contribution for individual i becomes

$$\begin{aligned} L_i^Z = \mathsf{P}(Y_{i1} = y_{i1} | X_{i1} = x_{i1}) & \\ \times \Bigg(\prod_{j=1}^{j_i} & \mathsf{P}(Z_{ij} = z_{ij} | Z_{ij} \geq z_{ij}, \mathbf{Z}_{ij-1} = \mathbf{z}_{ij-1}, \mathbf{X}_{iz_{ij}} = \mathbf{x}_{iz_{ij}}) \\ \times & \prod_{t=z_{ij-1}+1}^{z_{ij}-1} \left(1 - \mathsf{P}(Z_{ij} = t | Z_{ij} \geq t, \mathbf{Z}_{ij-1} = \mathbf{z}_{ij-1}, \mathbf{X}_{it} = \mathbf{x}_{it})\right) \Bigg) \\ \times \Bigg(& \prod_{t=z_{ij_i}+1}^{T-1} \left(1 - \mathsf{P}(Z_{ij_i+1} = t | Z_{ij_i+1} \geq t, \mathbf{Z}_{ij_i} = \mathbf{z}_{ij_i}, \mathbf{X}_{it} = \mathbf{x}_{it})\right) \Bigg). \end{aligned} \tag{65}$$

[39]Endogenous spell-varying covariates is beyond the scope of this paper.

In general, it is not possible to simplify further. If the covariates remain constant within each spell, then (65) simplifies to (64). In terms of computing time and memory requirements, (64) is likely to be more efficient than (65).

References

Allison, P. (1982). Discrete time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological Methodology 1982*, Chapter 2, pp. 61–97. San Francisco: Jossey-Bass.

Andrén, T. and D. Andrén (2013). Never give up? The persistence of welfare participation in Sweden. *IZA Journal of European Labor Studies 2*(1), 1–21.

Andrews, D. (1988). Chi-square diagnostic tests for econometric models. *Journal of Econometrics 37*, 135–156.

Arulampalam, W., A. L. Booth, and M. P. Taylor (2000). Unemployment persistence. *Oxford Economic Papers 52*, 24–50.

Bane, M. J. and D. T. Ellwood (1983). The dynamics of dependence: The routes to self sufficiency. Technical report, The U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Barmby, T. (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*(2), 261–265.

Beck, N., D. Epstein, S. Jackman, and S. O'Halloran (2002). 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.

Beck, N. and J. 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.

Beck, N., J. N. Katz, and R. Tucker (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Political Science Review 42*(4), 1260–1288.

Bhuller, M., C. N. Brinch, and S. Königs (2014). Time aggregation and state dependence in welfare receipt. Discussion paper 3, Statistics Norway Research Department.

Biewen, M. (2006). Who are the chronic poor? An econometric analysis of chronic poverty in Germany. In *Dynamics of Inequality and Poverty*, Volume 13 of *Research on Economic Inequality*. Emerald Group Publishing Limited.

Brown, C. C. (1975). On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics 31*, 863–872.

Browning, M. and J. M. Carro (2010). Heterogeneity in dynamic discrete choice models. *Econometrics Journal 13*, 1–39.

Capellari, L. and S. P. Jenkins (2004). Modelling low income transitions. *Journal of Applied Econometrics 19*(5), 593–610.

Capellari, L. and S. P. Jenkins (2014). The dynamics of social assistance benefit receipt in Britain. *Research in Labor Economics 39*, 39–77.

Cappellari, L., R. Dorsett, and G. Haile (2007). State dependence, duration dependence and unobserved heterogeneity in the employment transitions of the over-50s. Working paper 2007-16, Institute for Social and Economic Research.

Card, D. and D. R. Hyslop (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica 73*(6), 1723–1770.

Card, D. and D. 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), 1–20.

Chay, K. Y., H. Hoynes, and D. 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*, pp. 9–17.

Chay, K. Y. and D. R. Hyslop (2014). Identification and estimation of dynamic binary response panel data models: Empirical evidence using alternative approaches. *Research in Labor Economics 39*, 1–39.

Chen, D.-G. D., J. Sun, and K. E. Peace (2013). *Interval-Censored Time-to-Event Data: Methods and Applications*. Boca Raton, Florida: Chapman & Hall/CRC Press.

Chernoff, H. and E. L. Lehmann (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Annals of Mathematical Statistics 25*(3), 579–586.

Devicienti, F. (2011). Estimating poverty persistence in Britain. *Empirical Economics 40*, 657–686.

Doiron, D. and T. Gørgens (2008). State dependence in youth labor market experiences, and the evaluation of policy interventions. *Journal of Econometrics 145*, 81–97.

Eberwein, C., J. C. Ham, and R. J. Lalonde (1997). The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data. *The Review of Economic Studies 64*(4), 655–682.

Halliday, T. J. (2008). Heterogeneity, state dependence and health. *Econometrics Journal 11*(3), 499–516.

Ham, J. C. and R. J. LaLonde (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica 64*(1), 175–205.

Heckman, J. 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*, 227–269.

Heckman, J. J. (1981a). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in Labor Markets*, pp. 91–139. Chicago and London: University of Chicago Press.

Heckman, J. J. (1981b). The incidental parameter problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–195. Cambridge, Massachusetts: MIT Press.

Heckman, J. J. (1981c). Statistical models for discrete panel data. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Economic Applications*, pp. 114–178. Cambridge, Massachusetts: MIT Press.

Heckman, J. J. and G. J. Borjas (1980). Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica 47*, 247–283.

Heckman, J. J. and B. Singer (1984). Econometric duration analysis. *Journal of Econometrics 24*, 63–132.

Honoré, B. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies 60*, 241–246.

Horowitz, J. L. and S. Lee (2004). Semiparametric estimation of a panel data proportional hazards model with fixed effects. *Journal of Econometrics 119*(1), 155–198.

Huang, J. and J. A. Wellner (1997). Interval censored survival data: A review of recent progress. In D. Y. Lin and T. R. Fleming (Eds.), *Proceedings of the First Seattle Symposium in Biostatistics*, Volume 123 of *Lecture Notes in Statistics*, pp. 123–169. Springer US.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation behavior of married women. *Econometrica 67*(6), 1255–1294.

Jenkins, S. P. and C. García-Serrano (2004). The relationship between unemployment benefits and re-employment probabilities: Evidence from Spain. *Oxford Bulletin of Economics and Statistics 66*(2), 239–260.

Lacroix, G. and D. Brouillette (2011). Assessing the impact of a wage subsidy for single parents on social assistance. *Canadian Journal of Economics 44*(4), 1195–1221.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Econometric Society Monographs No. 17. Cambridge; New York: Cambridge University Press.

Moore, D. (1977). Generalized inverses, Wald's method and the construction of chi-squared tests of fit. *Journal of the American Statistical Association 72*, 131–137.

Stevens, A. H. (1999). Climbing out of poverty, falling back in: Measuring the persistence of poverty over multiple spells. *Journal of Human Resources 34*, 557–588.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), 307–333.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*, 39–54.

DBR1: $\mathsf{P}(Y_{it} = 1|\mathbf{Y}_{it-1} = \mathbf{y}_{it-1}) = G(\gamma_0 + \gamma_1 y_{it-1})$

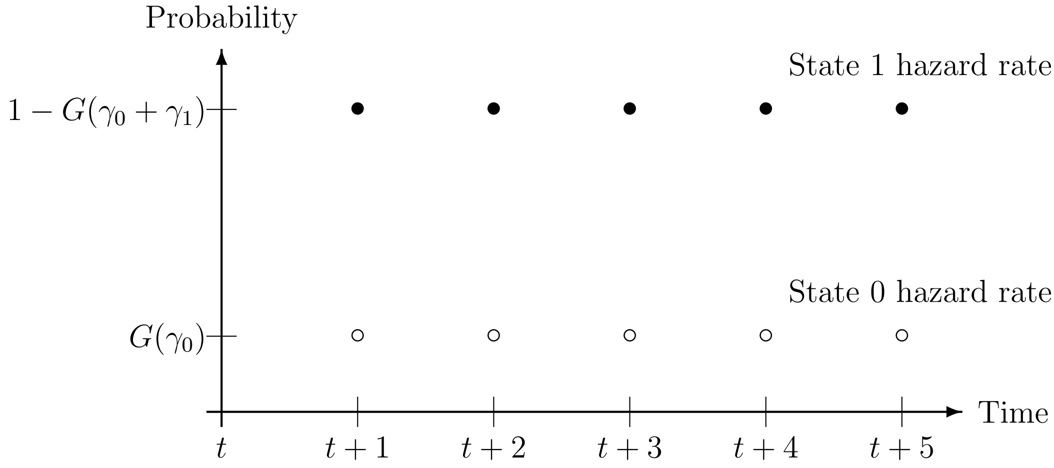

Figure 1: Hazard rates for a DBR1 model

DBR2: $\mathsf{P}(Y_{it} = 1|\mathbf{Y}_{it-1} = \mathbf{y}_{it-1}) = G(\gamma_0 + \gamma_1 y_{it-1} + \gamma_2 y_{it-2} + \gamma_3 y_{it-1} y_{it-2})$

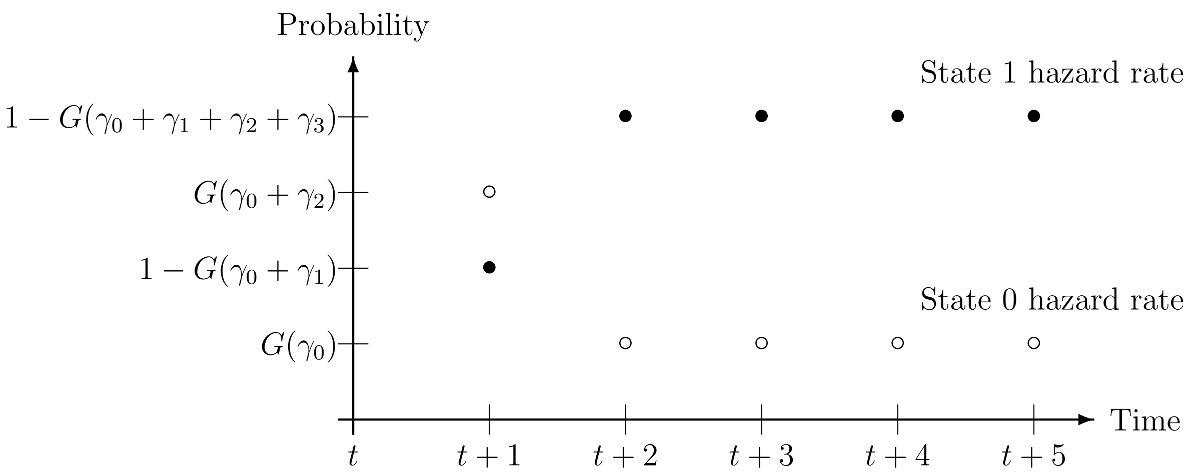

Figure 2: Hazard rates for a DBR2 model

Table 1: Descriptive statistics by initial state

	Full sample	Initial state	
		Not poor	In-poverty
Person-years: means (and standard errors)			
Aged 0–5	0.025	0.026	0.025
	(.0005)	(.0005)	(.0005)
Aged 6–17	0.225	0.213	0.239
	(.001)	(.002)	(.002)
Aged 18–24	0.204	0.188	0.223
	(.001)	(.002)	(.002)
Aged 25–54	0.420	0.432	0.406
	(.002)	(.002)	(.002)
Aged 55+	0.126	0.143	0.107
	(.001)	(.001)	(.001)
Female head	0.336	0.270	0.411
	(.001)	(.002)	(.002)
Black head	0.582	0.411	0.775
	(.002)	(.002)	(.002)
Poor (y_{it})	0.353	0.208	0.517
	(.001)	(.002)	(.002)
Transition (c_{it})	0.177	0.168	0.186
	(.001)	(.002)	(.002)
No. person-years	104,960	55,700	49,260
Persons: means (and standard errors)			
Transitions	3.35	3.20	3.53
	(.032)	(.043)	(.048)
No. persons	5,248	2,785	2,463

Table 2: Descriptive statistics by poverty status

	Full sample	Poverty status	
		Not poor	In-poverty
All spells: means (and standard errors)			
Duration	4.59	5.63	3.44
	(.033)	(.050)	(.038)
No. spells	22,849	12,062	10,787
Initial spells: means (and standard errors)			
Duration	7.11	8.23	5.86
	(.084)	(.121)	(.111)
No. spells	5,248	2,785	2,463
Fresh spells: means (and standard errors)			
Duration	3.84	4.85	2.72
	(.032)	(.051)	(.032)
No. spells	17,601	9,277	8,324

Notes: No adjustments for censoring.

Table 3: Dynamic binary response model estimates

	DBR0		DBR1		DBR2		
	IC	Strl	IC	Strl	IC1	IC2	Strl
Variables							
y_{it-1}		2.708		2.191		2.426	1.859
		(.040)		(.040)		(.147)	(.051)
y_{it-2}							0.942
							(.050)
$y_{it-1}y_{it-2}$							0.039
							(.075)
Aged 0–5	0.183	0.357	0.253	0.549	0.218	0.163	0.328
	(.091)	(.076)	(.104)	(.093)	(.100)	(.136)	(.104)
Aged 6–17	0.463	0.395	0.601	0.560	0.558	0.165	0.449
	(.065)	(.030)	(.075)	(.043)	(.073)	(.077)	(.037)
Aged 18–24	0.220	0.070	0.397	0.186	0.349	–0.13	0.109
	(.101)	(.024)	(.113)	(.030)	(.111)	(.126)	(.029)
Aged 55+	0.020	0.314	–0.131	0.272	–0.105	–0.236	0.288
	(.168)	(.042)	(.172)	(.048)	(.170)	(.180)	(.043)
Female Head	0.957	0.717	1.076	0.935	1.085	0.744	0.874
	(.129)	(.042)	(.139)	(.047)	(.138)	(.156)	(.045)
Black Head	1.446	0.607	1.429	0.620	1.48	0.855	0.527
	(.131)	(.044)	(.145)	(.055)	(.144)	(.155)	(.048)
Random effects (mass points and probabilities)							
ν_1	–1.519	–2.599	–2.178	–3.186	–2.121	–2.686	–3.206
	(.121)	(.039)	(.157)	(.057)	(.163)	(.178)	(.057)
ν_2			–0.654	–1.604	–0.767	–1.607	–1.926
			(.150)	(.073)	(.155)	(.171)	(.076)
π_1			0.638		0.640		
			(.023)		(.032)		
Statistics							
No. persons		5,248		5,248			5,248
No. years		20		20			20
Log QL		–46,520.1		–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 4: Multi-spell duration model estimates

	MSD1					MSD2		
	IS	Initial spells		Fresh spells		IS	Fresh spells	
		Entry	Exit	Entry	Exit		Entry	Exit
Variables								
$1(d_{it} \geq 2)$		–0.070	–0.425	–0.507	–0.562		–0.509	–0.544
		(.161)	(.168)	(.070)	(.068)		(.075)	(.069)
$1(d_{it} \geq 3)$		–0.490	0.234	–0.365	–0.188		–0.373	–0.163
		(.185)	(.191)	(.093)	(.094)		(.096)	(.097)
$1(d_{it} \geq 4)$		0.007	–0.046	–0.186	–0.290		–0.228	–0.321
		(.192)	(.202)	(.115)	(.119)		(.114)	(.120)
$1(d_{it} \geq 5)$		0.003	–0.028	–0.051	–0.169		–0.068	–0.258
		(.209)	(.236)	(.130)	(.142)		(.117)	(.135)
$1(d_{it} \geq 6)$		0.257	–0.051	–0.309	–0.056		0.029	–0.055
		(.162)	(.203)	(.113)	(.136)		(.091)	(.114)
Aged 0–5	0.156	0.332	–0.068	–0.340	–0.049	0.036	–0.206	–1.124
	(.100)	(.121)	(.120)	(.200)	(.226)	(.168)	(.429)	(.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
	(.111)	(.060)	(.067)	(.044)	(.044)	(.100)	(.040)	(.040)
Aged 55+	–0.078	0.047	–0.289	0.366	–0.287	0.150	0.327	–0.289
	(.170)	(.080)	(.157)	(.061)	(.060)	(.153)	(.052)	(.060)
Female Head	1.086	0.906	–0.732	0.881	–0.817	1.195	0.838	–0.802
	(.140)	(.087)	(.121)	(.067)	(.065)	(.135)	(.058)	(.066)
Black Head	1.529	0.246	–0.871	0.713	–0.498	1.379	0.407	–0.493
	(.142)	(.084)	(.118)	(.072)	(.063)	(.144)	(.059)	(.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
	(.181)	(.126)	(.193)	(.102)	(.094)	(.179)	(.099)	(.109)
ν_2	–0.901	–1.678	–0.615	–1.143	0.121	–1.180	–0.870	0.029
	(.147)	(.134)	(.147)	(.093)	(.076)	(.180)	(.130)	(.088)
π_1	0.590					0.677		
	(.042)					(.049)		
Statistics								
No. persons					5,248			5,248
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; IS: initial state equation; 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).

Table 5: Differences between the DBR1 and MSD1 models

#		Duration p dependence	Hetero-geneity	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 / G	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 $N = 104,960$.

Table 6: 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	1,532
6–10	0	88	68	209	325	322	1,012
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	1,311	687	1,340	842	5,248
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	1,605.5
6–10	0	31.0	62.4	194.6	475.5	458.6	1,221.9
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	1,003.8	774.4	1,593.4	975.0	5,248.0
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	1,529.9
6–10	0	68.0	84.0	208.0	406.1	402.6	1,168.6
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	1,004.7	755.0	1,489.1	925.6	5,248.0
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	1,592.5
6–10	0	105.2	73.4	187.8	341.6	350.5	1,058.4
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	1,246.6	690.9	1,400.8	822.8	5,248.0
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 7: Predictions of spells by initial state

No. spells	Actual data Initial state		DBR1 predictions Initial state		MSD1 predictions Initial state	
	Not poor	In-poverty	Not poor	In-poverty	Not poor	In-poverty
1	255	149	473.2	69.9	336.4	136.1
2	159	505	99.2	259.3	132.7	481.8
3	1,089	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	2,785	2,463	2,782.7	2,465.4	2,793.7	2,454.3
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 8: 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	2,785	9,277	2,783	9,510	2,794	9,266
Status: in-poverty						
Avg spell duration	5.86	2.72	4.48	2.96	5.88	2.67
No. spells	2,463	8,324	2,465	8,684	2,454	8,368

Notes: Prd: predictions; avg: average.

Table 9: 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 (9,277) new non-poor spells; [b]average (person-average first-year exit rate) across all (4,993) persons who have a new non-poor spell.

Recent Motu Working Papers

All papers in the Motu Working Paper Series are available on our website www.motu.org.nz, or by contacting us on info@motu.org.nz or +64 4 939 4250.

15-20 Maré David C., Ruth M. Pinkerton and Jacques Poot. 2015. “Residential Assimilation of Immigrants: A Cohort Approach.”

15-19 Timar, Levente, Arthur Grimes and Richard Fabling. 2015. “Before a Fall: Impacts of Earthquake Regulation and Building Codes on the Commercial Market”

15-18 Maré David C., Dean R. Hyslop and Richard Fabling. 2015. “Firm Productivity Growth and Skill.”

15-17 Fabling, Richard and David C. Maré. 2015. “Addressing the absence of hours information in linked employer-employee data.”

15-16 Thirkettle, Matt and Suzi Kerr. 2015. “Predicting harvestability of existing *Pinus radiata* stands: 2013-2030 projections of stumpage profits from pre-90 and post-89 forests”

15-15 Fabling, Richard and David C. Maré. 2015. “Production function estimation using New Zealand’s Longitudinal Business Database.”

15-14 Grimes, Arthur, Robert MacCulloch and Fraser McKay. 2015. “Indigenous Belief in a Just World: New Zealand Maori and other Ethnicities Compared.”

15-13 Apatov, Eyal, Richard Fabling, Adam Jaffe, Michele Morris and Matt Thirkettle. 2015. “Agricultural Productivity in New Zealand: First estimates from the Longitudinal Business Database.”

15-12:Laws, Athene, Jason Gush, Victoria Larsen and Adam B Jaffe. 2015. “The effect of public funding on research output: The New Zealand Marsden Fund.”

15-11 Dorner, Zachary and Suzi Kerr. 2015. “Methane and Metrics: From global climate policy to the NZ farm.”

15-10 Grimes, Arthur and Marc Reinhardt. 2015. “Relative Income and Subjective Wellbeing: Intra-national and Inter-national Comparisons by Settlement and Country Type”

15-09 Grimes, Arthur and Sean Hyland. 2015. “A New Cross-Country Measure of Material Wellbeing and Inequality: Methodology, Construction and Results.”

15-08 Jaffe, Adam and Trinh Le. 2015. “The impact of R&D subsidy of innovation: a study of New Zealand firms.”

15-07 Duhon, Madeline, Hugh McDonald and Suzi Kerr. 2015 “Nitrogen Trading in Lake Taupo: An Analysis and Evaluation of an Innovative Water Management Policy.

15-06 Allan, Corey, Suzi Kerr and Campbell Will. 2015. “Are we turning a brighter shade of green? The relationship between household characteristics and greenhouse gas emissions from consumption in New Zealand” (forthcoming)

15-05 Fabling, Richard and Lynda Sanderson. 2015. "Exchange rate fluctuations and the margins of exports”

15-04 Fabling, Richard, Richard Kneller and Lynda Sanderson. 2015. "The impact of tax changes on the short-run investment behaviour of New Zealand firms"

15-03 Sin, Isabelle, Steven Stillman. 2015. “Economic Liberalisation and the Mobility of Minority Groups: Evidence for Māori in New Zealand”

15-02 Grimes, Arthur, Ian Mitchell. 2015. “Impacts of Planning Rules, Regulations, Uncertainty and Delay on Residential Property Development”

15-01 De Rassenfosse, Gaétan, and Adam B. Jaffe. 2015. “Are Patent Fees Effective at Weeding Out Low-Quality Patents?”

14-15 Sin, Isabelle, Richard Fabling, Adam Jaffe, David C. Maré and Lynda Sanderson. 2014. “Exporting, Innovation and the Role of Immigrants.”