cemmap

centre for microdata methods and practice

Dynamic ordered panel logit models

Bo E. Honoré
Chris Muris
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP44/21

Economic and Social Research Council

Dynamic Ordered Panel Logit Models*

Bo E. Honoré‡ Chris Muris§ Martin Weidner¶

December 19, 2021

Abstract

We study a dynamic ordered logit model for panel data with fixed effects. We establish the validity of a set of moment conditions that are free of the fixed effects and that can be computed using four or more periods of data. We establish sufficient conditions for these moment conditions to identify the regression coefficients, the autoregressive parameters, and the threshold parameters. The parameters can be estimated using the generalized method of moments. We document the performance of this estimator using Monte Carlo simulations and an empirical illustration to self-reported health status using the British Household Panel Survey.

*This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University, by the National Science Foundation (Grant Number SES-1530741), by the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice (grant numbers RES-589-28-0001, RES-589-28-0002 and ES/P008909/1), by the Social Sciences and Humanities Research Council of Canada (grant number IG 435-2021-0778), and by the European Research Council grants ERC-2014-CoG-646917-ROMIA and ERC-2018-CoG-819086-PANEDA. We are grateful to seminar participants at the Universities of Leuven and Toronto, to Paul Contoyannis for helpful conversations about the BHPS data used in Contoyannis, Jones, and Rice (2004), and to Geert Dhaene and Sharada Dharmasankar for useful comments and discussion.

‡Princeton University and The Dale T. Mortensen Centre at the University of Aarhus, honore@princeton.edu

§McMaster University, muerisc@mcmaster.ca

¶University of Oxford, martin.weidner@economics.ox.ac.uk

1 Introduction

Panel surveys routinely collect data on an ordinal scale. For example, many nationally representative surveys ask respondents to rate their health or life satisfaction on an ordinal scale.[1] Other examples include test results in longitudinal data sets gathered for studying education.

We are interested in regression models for ordinal outcomes that allow for lagged dependent variables as well as fixed effects. In the model that we propose, the ordered outcome depends on a fixed effect, a lagged dependent variable, regressors, and a logistic error term. We study identification and estimation of the finite-dimensional parameters in this model when only a small number (≥ 4) of time periods are available.

For other types of outcomes variables (continuous outcomes in linear models, binary and multinomial outcomes), results for regression models with fixed effects and lagged dependent variables are already available. Such results are of great importance for applied practice, as they allow researchers to distinguish unobserved heterogeneity from state dependence, and to control for both when estimating the effect of regressors. The demand for such methods is evidenced by the popularity of existing approaches for the linear model, such as those proposed by Arellano and Bond (1991) and Blundell and Bond (1998). In contrast, for ordinal outcomes, almost no results are available.

From the analysis of other nonlinear models with fixed effects, we know that it is challenging to accommodate unobserved heterogeneity in nonlinear models, especially when also allowing for lagged dependent variables. For example, early work on the dynamic binary choice model with fixed effects assumed no regressors, or restricted their joint distribution (cf. Chamberlain 1985 and Honoré and Kyriazidou 2000). Recent developments by Honoré and Weidner (2020) and Kitazawa (2021) relax this requirement.

This challenge is even greater for the dynamic ordered logit model. The ordered logit

[1] One example is the British Household Panel Survey in our empirical application. Others include the U.S. Health and Retirement Study and Medical Expenditure Panel Survey, the Canadian Longitudinal Study on Ageing and the National Longitudinal Survey of Children and Youth, the Australian Longitudinal Study on Women's health, the European Union Statistics on Income and Living Conditions, the Survey on Health, Ageing, and Retirement in Europe, among many others.

model is not in the exponential class (Hahn 1997), so even for the static version we cannot directly appeal to a sufficient statistic approach. However, the static ordered logit model can be reduced to a set of binary choice models (cf. Das and van Soest 1999, Johnson 2004, Baetschmann, Staub, and Winkelmann 2015, Muris 2017, and Botosaru, Muris, and Pendakur 2021). Unfortunately, the dynamic ordered logit model cannot be similarly reduced to a dynamic binary choice model (see Muris, Raposo, and Vandoros 2020). Therefore, a new approach is needed.

We follow the functional differencing approach in Bonhomme (2012) to obtain moment conditions for the finite-dimensional parameters in this model, namely the autoregressive parameters (one for each level of the lagged dependent variable), the threshold parameters in the underlying latent variable formulation, and the regression coefficients. Our approach is closely related to Honoré and Weidner (2020), and can be seen as the extension of their method to the case of an ordered response variable.

This paper contributes to the literature on dynamic ordered logit models. We are aware of only one paper that studies a fixed-T version of this model while allowing for fixed effects. The approach in Muris, Raposo, and Vandoros (2020) builds on methods for dynamic binary choice models in Honoré and Kyriazidou (2000) by restricting how past values of the dependent variable enter the model. In particular, in Muris, Raposo, and Vandoros (2020), the lagged dependent variable $Y_{i,t-1}$ enters the model only via $\mathbb{1}\{Y_{i,t-1} \geq k\}$ for some known k. We do not impose such a restriction, and allow the effect of $Y_{i,t-1}$ to vary freely with its level.

Other existing work on dynamic panel models for ordered outcomes uses a random effects approach (Contoyannis, Jones, and Rice 2004, Albarran, Carrasco, and Carro 2019) or requires a large number of time periods for consistency (Carro and Traferri 2014, Fernández-Val, Savchenko, and Vella 2017). An earlier version of Aristodemou 2021 contained partial identification results for a dynamic ordered choice model without logistic errors. Our approach requires no restrictions on the dependence between fixed effects and regressors, requires four periods of data for consistency, and delivers point identification and estimates.

More broadly, this paper contributes to the literature on fixed-T identification and estimation in nonlinear panel models with fixed effects (see Honoré 2002, Arellano 2003, and Arellano and Bonhomme 2011 for overviews). The literature contains results for several models adjacent to ours. For example, the static panel ordered logit model with fixed effects was studied by Das and van Soest (1999), Johnson (2004), Baetschmann, Staub, and Winkelmann (2015), and Muris (2017); results for static and dynamic binomial and multinomial choice models are in Chamberlain (1980), Honoré and Kyriazidou (2000), Magnac (2000), Shi, Shum, and Song (2018), Aguirregabiria, Gu, and Luo (2021), Aguirregabiria and Carro (2021), Pakes, Porter, Shepard, and Calder-Wang (2021) and Khan, Ouyang, and Tamer (2021).

Our main contribution is to obtain novel moment conditions for the common parameters in the dynamic ordered logit model with fixed effects. Additionally, we obtain conditions under which these moment conditions identify those parameters. Finally, we discuss the implied generalized method of moments (GMM) estimator and demonstrate its performance in a Monte Carlo study and in an empirical application to self-reported health status in the British Household Panel Study.

2 Model and moment conditions

In this section we first describe the panel ordered logit model that is used throughout the paper, and then present moment conditions for the model that can be used to estimate the common parameter of the model without imposing any knowledge of the individual specific effects.

2.1 Model and notation

We consider panel data with cross-sectional units $i = 1, \ldots, n$ and time periods $t = 0, \ldots, T$. For each pair (i, t) we observe the discrete outcome $Y_{it} \in \{1, 2, \ldots, Q\}$, which can take $Q \in \{2, 3, 4, \ldots\}$ different values, and the strictly exogenous regressors $X_{it} \in \mathbb{R}^K$. We discuss unbalanced panels in Section 2.4, but for now we assume a balanced panel where outcomes are observed for all $t \geq 0$ and regressors for all $t \geq 1$. Thus the total number of time periods for which outcomes are observed is $T+1$. For $t \geq 1$ the observed discrete outcomes depend on an unobserved latent variable $Y^*_{it} \in \mathbb{R}$ as follows:

$$Y_{it} = \begin{cases} 1 & \text{if} \quad Y^*_{it} \in (-\infty\,, \lambda_1], \\ 2 & \text{if} \quad Y^*_{it} \in (\lambda_1\,, \lambda_2], \\ \vdots & \\ Q & \text{if} \quad Y^*_{it} \in (\lambda_{Q-1}\,, \infty), \end{cases} \tag{1}$$

where the $\lambda = (\lambda_1, \ldots, \lambda_{Q-1}) \in \mathbb{R}^{Q-1}$ are unknown parameters with $\lambda_1 < \lambda_2 < \ldots < \lambda_{Q-1}$. The latent variable is generated by the model

$$Y^*_{it} = X'_{it}\,\beta + \sum_{q=1}^{Q} \gamma_q\, \mathbb{1}\left\{Y_{i,t-1} = q\right\} + A_i + \varepsilon_{it}, \tag{2}$$

with unknown parameters $\beta \in \mathbb{R}^K$ and $\gamma = (\gamma_1, \ldots, \gamma_Q) \in \mathbb{R}^Q$. Here, $A_i \in \mathbb{R} \cup \{\pm\infty\}$ is an unobserved individual specific effect whose distribution is not specified, and A_i is allowed to be arbitrarily correlated with the regressors X_{it} and the initial conditions Y_{i0}. Let $X_i := (X_{i1}, \ldots, X_{iT})$. Conditional on Y_{i0}, X_i, and A_i, the idiosyncratic error term ε_{it} is assumed to be independent and identically distributed over t with cumulative distribution function $\Lambda(\varepsilon) := [1 + \exp(-\varepsilon)]^{-1}$. Thus, ε_{it} is a logistic error term. For the cross-sectional sampling, we assume that $(Y_{i0}, X_{i1}, \ldots, X_{iT}, A_i, \varepsilon_{i1}, \ldots, \varepsilon_{iT})$ are independent and identically distributed across i.

The model described by (1) and (2) is a dynamic ordered panel logit model, where an arbitrary function $\gamma_{Y_{i,t-1}}$ of the lagged depend variable $Y_{i,t-1}$ is allowed enter additively into the latent variable Y^*_{it}. This model strikes a balance between a general functional form and a parsimonious parameter structure. We discuss possible generalizations of the model for Y^*_{it} in Section 2.5, but otherwise impose (2) throughout the paper.

Our ultimate goal is to estimate the unknown parameters $\theta = (\beta, \gamma, \lambda) \in \Theta := \mathbb{R}^{K+2Q-1}$ without imposing any assumptions on the individual-specific effect A_i. This requires two normalizations, because common additive shifts of all the parameters γ_q or of all the parameters λ_q can be absorbed into A_i. For example, we could impose the normalizations $\gamma_1 = 0$

and $\lambda_1 = 0$, but in this section there is no need to specify such normalizations.

It is convenient to define $\lambda_0 := -\infty$, and $\lambda_Q := \infty$, and

$$z(Y_{i,t-1}, X_{it}, \theta) := X_{it}' \beta + \sum_{q=1}^{Q} \gamma_q \, \mathbb{1}\left\{Y_{i,t-1} = q\right\}. \tag{3}$$

With this notation, the model assumptions imposed so far imply that the distribution of Y_{it} conditional on the regressors X_i, past outcomes $Y_i^{t-1} = (Y_{i,t-1}, Y_{i,t-2}, \ldots)$, and fixed effects A_i, is given by

$$\Pr\left(Y_{it} = q \,\middle|\, Y_i^{t-1}, X_i, A_i, \theta\right) = \Lambda\Big[z(Y_{i,t-1}, X_{it}, \theta) + A_i - \lambda_{q-1}\Big] - \Lambda\Big[z(Y_{i,t-1}, X_{it}, \theta) + A_i - \lambda_q\Big] \tag{4}$$

for all $i \in \{1, \ldots, n\}$, $t \in \{1, 2, \ldots, T\}$, and $q \in \{1, 2, \ldots, Q\}$. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})$, and let the true model parameters be denoted by $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$. In the following, all probabilistic statements are for the model distribution generated under θ^0. For example, we have $\Pr\big(Y_i = y_i \,\big|\, Y_{i0} = y_{i0},\, X_i = x_i,\, A_i = \alpha_i\big) = p_{y_{i0}}(y_i, x_i, \theta^0, \alpha_i)$, where

$$p_{y_{i0}}(y_i, x_i, \theta, \alpha_i) := \prod_{t=1}^{T} \left\{ \Lambda\Big[z(y_{i,t-1}, x_{it}, \theta) + \alpha_i - \lambda_{y_{it}-1}\Big] - \Lambda\Big[z(y_{i,t-1}, x_{it}, \theta) + \alpha_i - \lambda_{y_{it}}\Big] \right\}. \tag{5}$$

From now on we drop the index i until we discuss estimation; instead of Y_{i0}, Y_i, X_i, A_i, we just write Y_0, Y, X, A for those random variables and random vectors.

2.2 Moment condition approach

For the model just introduced, we want to find moment functions $m : \{1, \ldots, Q\} \times \{1, \ldots, Q\}^T \times \mathbb{R}^{T \times K} \times \Theta \to \mathbb{R}$ such that

$$\mathbb{E}\left[m_{Y_0}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha \right] = 0 \tag{6}$$

for all $y_0 \in \{1, \ldots, Q\}$, $x \in \mathbb{R}^{T \times K}$, and $\alpha \in \mathbb{R} \cup \{\pm\infty\}$. Here, we write the first argument y_0 of the moment function as an index, but that is purely for notational convenience. Conditional on $Y_0 = y_0$, $X = x$, and $A = \alpha$, only the outcome $Y = (Y_1, \ldots, Y_T) \in \{1, \ldots, Q\}^T$ remains random, and according to the ordered logit model its distribution is given by (5). The model assumptions in the last subsection are therefore completely sufficient to evaluate the conditional expectation in (6).

Once we have established the conditional moment conditions, by the law of iterated expectations we also have the unconditional moment conditions

$$\mathbb{E}\left[h(Y_0, X, \theta^0)\, m_{Y_0}(Y, X, \theta^0)\right] = 0 \tag{7}$$

for any function $h : \{1, \ldots, Q\} \times \mathbb{R}^{T \times K} \times \Theta \to \mathbb{R}$ such that the expectation is well-defined. Those unconditional moment conditions can then be used to estimate the model parameters θ^0 by the generalized method of moments (GMM).

This estimation approach solves the incidental parameter problem (Neyman and Scott 1948), because the moment condition (7) does not feature the individual-specific effect A at all. No assumptions are imposed on the distribution of those nuisance parameters, and they need not be estimated. On the flip-side, this implies that we do not learn anything about the distribution of A, which is why in this paper we focus exclusively on inference for θ. Notice, however, that even if one is interested in (functions of) the individual specific effects, like average partial effects, then the estimation of the common parameters θ will always be a key first step in any inference procedure.

The moment condition approach just described eliminates the individual-specific effect A from the estimation, because (6) is assumed to hold for all $\alpha \in \mathbb{R} \cup \{\pm\infty\}$, but the moment function $m_{Y_0}(Y, X, \theta^0)$ does not depend on A at all. The existence of moment functions with that property is quite miraculous: for any given values of $Y_0 = y_0$, $X = x$, and θ^0, the moment function $m_{y_0}(\cdot, x, \theta^0) : \{1, \ldots, Q\}^T \to \mathbb{R}$ can be viewed as a finite-dimensional vector (Q^T real numbers), but (6) imposes an infinite number of linear constraints – one for each $\alpha \in \mathbb{R} \cup \{\pm\infty\}$. The logistic assumption on ε_{it} is crucial for finding solutions of this infinite dimensional linear system in a finite number of variables. For other distributional assumptions on ε_{it} we do not expect such solutions to exist.

In the following, we present moment functions $m_{Y_0}(Y, X, \theta^0)$ that satisfy (6). We have derived those moment functions for the dynamic panel ordered logit model analogously to the results for the dynamic panel binary choice logit model in Honoré and Weidner (2020). Indeed, for the binary choice case ($Q = 2$) our moment functions below exactly coincide with those in Honoré and Weidner (2020), and we refer to that paper for more details on the derivation, which is closely related to the functional differencing method in Bonhomme (2012).

Once we have obtained the expressions presented below for the moment functions, then we can completely forget about their derivation and focus on showing that they are valid moment functions – i.e. that (6) holds – and on their implications for identification and estimation of θ. This is the focus of this paper.

2.3 Moment conditions for $T = 3$

We first introduce our moment functions for $T = 3$. In our convention this means that outcomes Y_t are observed for the four time periods $t = 0, 1, 2, 3$ (including the initial conditions Y_0). We have verified numerically that no moment functions satisfying (6) for general parameter and regressor values exist for $T < 3$, and for the binary choice case ($Q = 2$) a proof of this non-existence is given in Honoré and Weidner (2020). Thus, $T = 3$ is the smallest number of time periods that we can consider.

We use lower case letters for generic arguments (as opposed to random variables) of the moment function $m_{y_0}(y, x, \theta)$, where $y_0 \in \{1, \ldots, Q\}$, $y \in \{1, \ldots, Q\}^T$, $x \in \mathbb{R}^{T \times K}$ and $\theta = (\beta, \gamma, \lambda) \in \Theta$. The t'th row of x is denoted by $x_t' \in \mathbb{R}^K$, and we define $x_{ts} := x_t - x_s$, $\gamma_{qr} := \gamma_q - \gamma_r$ and $\lambda_{qr} := \lambda_q - \lambda_r$.

The approach in Honoré and Weidner (2020) results in multiple moment functions $m_{y_0,q_1,q_2,q_3}(y, x, \theta)$ which are distinguished by the additional indices q_1, q_2, q_3. For $q_1, q_3 \in \{1, \ldots, Q-1\}$

and $q_2 \in \{2, \ldots, Q-1\}$ we define

$$
\begin{aligned}
& m_{y_0,q_1,q_2,q_3}(y,x,\theta) \\
& := \begin{cases}
\exp\left(x_{13}'\,\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) \dfrac{\exp(x_{32}'\,\beta + \gamma_{q_2,y_1} + \lambda_{q_2,q_3}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_1 \le q_1,\ y_2 = q_2, y_3 \le q_3, \\
\exp\left(x_{13}'\,\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) \dfrac{1 - \exp(x_{23}'\,\beta + \gamma_{y_1,q_2} + \lambda_{q_3,q_2})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_1 \le q_1,\ y_2 = q_2,\ y_3 > q_3, \\
\exp\left(x_{13}'\,\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) & \text{if } y_1 \le q_1,\ y_2 > q_2, \\
-1 & \text{if } y_1 > q_1,\ y_2 < q_2, \\
-\dfrac{1 - \exp(x_{32}'\,\beta + \gamma_{q_2,y_1} + \lambda_{q_2-1,q_3})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_1 > q_1,\ y_2 = q_2,\ y_3 \le q_3, \\
-\dfrac{\exp(x_{23}'\,\beta + \gamma_{y_1,q_2} + \lambda_{q_3,q_2-1}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_1 > q_1,\ y_2 = q_2,\ y_3 > q_3, \\
0 & \text{otherwise.}
\end{cases}
\end{aligned}
\tag{8}
$$

Any valid moment function satisfying (6) can be multiplied by an arbitrary constant and remain a valid moment function. In (8) we used that rescaling freedom to normalize the entry for the case $(y_1 > q_1, y_2 < q_2)$ to be equal to -1. If, alternatively, we normalize the entry for $(y_1 \le q_1, y_2 > q_2)$ to be equal to -1, then we obtain the equally valid moment function

$$
\widetilde{m}_{y_0,q_1,q_2,q_3}(y,x,\theta) = -\frac{m_{y_0,q_1,q_2,q_3}(y,x,\theta)}{\exp\left(x_{13}'\,\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right)}.
$$

This rescaling is interesting, because if we reverse the order of the outcome labels (i.e. $Y_t \mapsto Q + 1 - Y_t$), the model remains unchanged except for the parameter transformations $\beta \mapsto -\beta$, $\gamma_q \mapsto -\gamma_{Q+1-q}$, and $\lambda_q \mapsto -\lambda_{Q-q}$. Under this transformation, the moment function $m_{y_0,q_1,q_2,q_3}(y,x,\theta)$ becomes $\widetilde{m}_{\widetilde{y}_0,\widetilde{q}_1,\widetilde{q}_2,\widetilde{q}_3}(y,x,\theta)$ with $\widetilde{y}_0 = Q+1-y_0$ and $(\widetilde{q}_1,\widetilde{q}_2,\widetilde{q}_3) = (Q - q_1, Q+1-q_2, Q-q_3)$. This transformation therefore does not deliver any new moment functions, which are not already (up to rescaling) given by (8).

Equation (8) does not define $m_{y_0,q_1,q_2,q_3}(y,x,\theta)$ for $q_2 = 1$ and $q_2 = Q$. If we plug those values of q_2 into (8), then various undefined terms appear since $\lambda_0 = -\infty$ and $\lambda_Q = \infty$.

However, if for $q_2 = 1$ we properly evaluate the limit of $\widetilde{m}_{y_0,q_1,q_2,q_3}(y,x,\theta)$ as $\lambda_0 \to -\infty$, then we obtain

$$m_{y_0,q_1,1,q_3}(y,x,\theta) := \begin{cases} \exp\left(x_{23}'\,\beta + \gamma_{y_1,1} + \lambda_{q_3,1}\right) - 1 & \text{if } y_1 \leq q_1,\ y_2 = 1,\ y_3 > q_3, \\ -1 & \text{if } y_1 \leq q_1,\ y_2 > 1, \\ \exp\left(x_{31}'\,\beta + \gamma_{1,y_0} + \lambda_{q_1,q_3}\right) & \text{if } y_1 > q_1,\ y_2 = 1,\ y_3 \leq q_3, \\ \exp\left(x_{21}'\,\beta + \gamma_{y_1,y_0} + \lambda_{q_1,1}\right) & \text{if } y_1 > q_1,\ y_2 = 1,\ y_3 > q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Similarly, if for $q_2 = Q$ we properly evaluate the limit of $m_{y_0,q_1,q_2,q_3}(y,x,\theta)$ as $\lambda_Q \to \infty$, then we obtain

$$m_{y_0,q_1,Q,q_3}(y,x,\theta) := \begin{cases} \exp\left(x_{12}'\,\beta + \gamma_{y_0,y_1} + \lambda_{Q-1,q_1}\right) & \text{if } y_1 \leq q_1,\ y_2 = Q,\ y_3 \leq q_3, \\ \exp\left(x_{13}'\,\beta + \gamma_{y_0,Q} + \lambda_{q_3,q_1}\right) & \text{if } y_1 \leq q_1,\ y_2 = Q,\ y_3 > q_3, \\ -1 & \text{if } y_1 > q_1,\ y_2 < Q, \\ \exp\left(x_{32}'\,\beta + \gamma_{Q,y_1} + \lambda_{Q-1,q_3}\right) - 1 & \text{if } y_1 > q_1,\ y_2 = Q,\ y_3 \leq q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Together, the formulas (8), (9), and (10) provide one moment function for every value of $(y_0, q_1, q_2, q_3) \in \{1, \ldots, Q\} \times \{1, \ldots, Q-1\} \times \{1, \ldots, Q\} \times \{1, \ldots, Q-1\}$, and these constitute all our moment functions for the dynamic ordered logit model with $T = 3$. [2] The following theorem states that these are indeed valid moment functions for the dynamic panel ordered logit model, independent of the value of the fixed effect A.

Theorem 1 *If the outcomes $Y = (Y_1, Y_2, Y_3)$ are generated from model (4) with $Q \geq 2$, $T = 3$ and true parameters $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$, then we have for all $y_0 \in \{1, \ldots, Q\}$, $q_1, q_3 \in \{1, \ldots, Q-1\}$, $q_2 \in \{1, \ldots, Q\}$, $x \in \mathbb{R}^{K \times 3}$, and $\alpha \in \mathbb{R} \cup \{\pm\infty\}$ that*

$$\mathbb{E}\left[m_{y_0,q_1,q_2,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\ X = x,\ A = \alpha\right] = 0.$$

[2] By the limiting arguments ($\lambda_0 \to -\infty$ and $\lambda_Q \to \infty$) described above, all of those moment functions are already implicitly defined via (8) alone.

The proof of the theorem is given in the appendix. For any fixed value of Q one could, in principle, show by direct calculation that

$$\sum_{y \in \{1,2,\ldots,Q\}^3} p_{y_0}(y, x, \theta^0, \alpha)\, m_{y_0,q_1,q_2,q_3}(y, x, \theta^0) = 0$$

for the model probabilities $p_{y_0}(y, x, \theta^0, \alpha)$ given by (5), but our proof in the appendix does not rely on such a brute force calculation and is valid for any $Q \geq 2$.

For each initial condition y_0 we thus have $\ell = Q(Q-1)^2$ available moment conditions. For example, for $Q = 2, 3, 4, 5$ there are respectively $\ell = 2$, 12, 36, 80 available moment conditions for each initial condition. For those values of Q we have verified numerically that our ℓ moment conditions are linearly independent, and that they constitute all the valid moment conditions that are available for the dynamic panel ordered logit model with $T = 3$, for generic values of γ.[3] We believe that this is true for all $Q \geq 2$, but a proof of this completeness result is beyond the scope of this paper. For the special case of dynamic binary choice ($Q = 2$), the moment conditions here are identical to those in Honoré and Weidner (2020) and Kitazawa (2021), and completeness of those binary choice moment conditions is discussed in Kruiniger (2020) and Dobronyi, Gu, and Kim (2021).

2.4 Moment conditions for $T > 3$

We now consider the case where the econometrician has data for more that three time periods (in addition to the period that gives the initial condition). Obviously, all the moment conditions above for $T = 3$ are still valid when applied to three consecutive periods, but additional moment conditions become available for $T > 3$. We consider first moment conditions that are based on the outcome on three periods, where the last two are consecutive. Let $z_t := z(y_{t-1}, x_t, \theta)$, with $z(\cdot,\cdot,\cdot)$ defined in (3), and define $z_{ts} := z_t - z_s$. For $y_0 \in \{1, \ldots, Q\}$,

[3] If some of the parameters γ_q are equal to each other, then additional moment conditions become available.

$q_1, q_3 \in \{1, \ldots, Q-1\}$, $q_2 \in \{2, \ldots, Q-1\}$, and $t, s \in \{1, 2, \ldots, T-1\}$ with $t < s$ we define

$$m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta) := \begin{cases} \exp\left(z_{t,s+1} + \lambda_{q_3,q_1}\right) \dfrac{\exp(z_{s+1,s} + \lambda_{q_2,q_3}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_t \leq q_1,\ y_s = q_2, y_{s+1} \leq q_3, \\ \exp\left(z_{t,s+1} + \lambda_{q_3,q_1}\right) \dfrac{1 - \exp(z_{s,s+1} + \lambda_{q_3,q_2})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_t \leq q_1,\ y_s = q_2,\ y_{s+1} > q_3, \\ \exp\left(z_{t,s+1} + \gamma_{y_s,q_2} + \lambda_{q_3,q_1}\right) & \text{if } y_t \leq q_1,\ y_s > q_2, \\ -1 & \text{if } y_t > q_1,\ y_s < q_2, \\ -\dfrac{1 - \exp(z_{s+1,s} + \lambda_{q_2-1,q_3})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_t > q_1,\ y_s = q_2,\ y_{s+1} \leq q_3, \\ -\dfrac{\exp(z_{s,s+1} + \lambda_{q_3,q_2-1}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_t > q_1,\ y_s = q_2,\ y_{s+1} > q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

For $T = 3$, $t = 1$, and $s = 2$, it is straightforward to verify that $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta)$ in equation (11) equals the moment function in equation (8). For larger values for T, the moment function in (11) can be implemented as long as outcomes are observed for the time periods $\{t-1, t, s-1, s, s+1\}$ and covariates are observed for time periods $\{t, s, s+1\}$.

Since $\lambda_0 = -\infty$ and $\lambda_Q = \infty$, equation (11) can not be used to define a moment function when q_2 equals 1 or Q. We next define moment functions for these cases. For $y_0 \in \{1, \ldots, Q\}$,

$q_1, q_3 \in \{1, \ldots, Q-1\}$, $t, s, r \in \{1, 2, \ldots, T\}$, and $t < s < r$, we define

$$m^{(t,s,r)}_{y_0,q_1,1,q_3}(y,x,\theta) := \begin{cases} \exp\left(z_{sr} + \lambda_{q_3,1}\right) - 1 & \text{if } y_t \leq q_1,\ y_s = 1,\ y_r > q_3, \\ -1 & \text{if } y_t \leq q_1,\ y_s > 1, \\ \exp\left(z_{rt} + \lambda_{q_1,q_3}\right) & \text{if } y_t > q_1,\ y_s = 1,\ y_r \leq q_3, \\ \exp\left(z_{st} + \lambda_{q_1,q_2}\right) & \text{if } y_t > q_1,\ y_s = 1,\ y_r > q_3, \\ 0 & \text{otherwise,} \end{cases}$$

$$m^{(t,s,r)}_{y_0,q_1,Q,q_3}(y,x,\theta) := \begin{cases} \exp\left(z_{ts} + \lambda_{Q-1,q_1}\right) & \text{if } y_t \leq q_1,\ y_s = Q,\ y_r \leq q_3, \\ \exp\left(z_{tr} + \lambda_{q_3,q_1}\right) & \text{if } y_t \leq q_1,\ y_s = Q,\ y_r > q_3, \\ -1 & \text{if } y_t > q_1,\ y_s < Q, \\ \exp\left(z_{rs} + \lambda_{Q-1,q_3}\right) - 1 & \text{if } y_t > q_1,\ y_s = Q,\ y_r \leq q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

When T equals 3 and $(t, s, r) = (1, 2, 3)$, these moment functions agree with the ones in equations (9) and (10), where all the arguments were made explicit. For $r = s+1$, analogous to (9) and (10) for $T = 3$, the two moment conditions in (12) for $T \geq 3$ can be obtained from (11) by setting $q_2 = 1$ and carefully evaluating the limit $\lambda_0 \to -\infty$ (after normalizing the value for $y_t \leq q_1$, $y_s > 1$ to be -1), or setting $q_2 = Q$ and taking the limit $\lambda_Q \to \infty$. It is therefore appropriate to think of (11) as our master equation which summarizes all the moment conditions provided in this paper. In (12) we can choose more general $r \geq s+1$, but otherwise the structure of (12) can be derived from (11).

It turns out that the moment functions with $r > s+1$ are not actually needed to span all possible valid moment functions of the dynamic ordered choice logit model (see our discussion of independence and completeness below). However, since implementation of these moment functions requires only that we observe three pairs (y_{t-1}, y_t), (y_{s-1}, y_s), (y_{r-1}, y_r) of consecutive outcomes, they may be empirically relevant for the case where observations for some time periods are (exogenously) missing. We also include $r > s+1$ in our discussion here to ensure that our results in this paper contain those for the dynamic binary choice logit model studied in Honoré and Weidner (2020) as a special case — notice that for $Q = 2$

we always have $q_2 = 1$ or $q_2 = Q$, that is, for the binary choice case all available moment functions are stated in (12).

The following theorem establishes that the moment function in (11) and (12) do indeed deliver valid moment conditions.

Theorem 2 *If the outcomes $Y = (Y_1, \ldots, Y_T)$ are generated from model (4) with $Q \geq 2$, $T \geq 3$ and true parameters $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$, then we have for all $t, s, r \in \{1, 2, \ldots, T\}$ with $t < s < r$, $y_0 \in \{1, \ldots, Q\}$, $q_1, q_3 \in \{1, \ldots, Q-1\}$, $x \in \mathbb{R}^{K \times T}$, $\alpha \in \mathbb{R} \cup \{\pm\infty\}$, and $w : \{1, \ldots, Q\}^{t-1} \to \mathbb{R}$ that*

$$\mathbb{E}\left[w(Y_1, \ldots, Y_{t-1})\, m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha\right] = 0, \quad \textit{for } q_2 \in \{2, \ldots, Q-1\},$$
$$\mathbb{E}\left[w(Y_1, \ldots, Y_{t-1})\, m^{(t,s,r)}_{y_0,q_1,q_2,q_3}(Y, x, \theta) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha\right] = 0, \quad \textit{for } q_2 \in \{1, Q\}.$$

The proof is provided in the appendix. Notice that for $q_2 \in \{1, Q\}$ we can choose the time indices $t < s < r$ freely. By contrast, for $q_2 \in \{2, \ldots, Q-1\}$ we can only choose $t < s$ freely, but the third time index that appears in the definition of the moment function needs to be equal to $s + 1$, otherwise we do not obtain a valid moment function for those values of q_2.

This distinction between $q_2 \in \{1, Q\}$ and $q_2 \in \{2, \ldots, Q-1\}$ is also reflected in the proof of Theorem 2. The moment functions in (12) for $q_2 \in \{1, Q\}$ only depend on Y_1, Y_2, Y_3 through the binarized variables $\widetilde{Y}_1 = \mathbb{1}\{Y_1 > q_1\}$, $\widetilde{Y}_2 = \mathbb{1}\{Y_2 = q_2\}$, $\widetilde{Y}_3 = \mathbb{1}\{Y_3 > q_3\}$, and the proof relies on Lemma 2 in the appendix, which provides a general set of valid moment functions for such binary variables, very closely related to the dynamic binary choice results in Honoré and Weidner (2020). By contrast, the moment functions in (11) for $q_2 \in \{2, \ldots, Q-1\}$ cannot be expressed through binarized variables only, because there the dependence on Y_2 requires distinguishing three cases ($Y_s < q_2$, $Y_s = q_2$, $Y_s > q_2$), and the proof relies on Lemma 1 in the appendix which is completely novel to the current paper. However, that proof strategy for $q_2 \in \{2, \ldots, Q-1\}$ does not work for $s > r + 1$, and we have also numerically verified that our moment conditions for $q_2 \in \{2, \ldots, Q-1\}$ indeed do not generalize to $s > r + 1$.

Conjecture on the completeness of the moment conditions

Theorem 2 states that the moment functions in (11) and (12) are valid, but it is natural to ask whether they are also linearly independent, and whether they constitute all possible valid moment functions of the dynamic panel ordered logit model. We do not aim to formally prove such a linear independence and completeness result in this paper, and the following statement should accordingly be understood as a conjecture, which we have numerically confirmed for various combinations of Q and T and for many different numerical values of the regressors and model parameters:

Let the outcomes $Y = (Y_1, \ldots, Y_T)$ be generated from model (4) with $Q \geq 2$, $T \geq 3$, and true parameters $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$ such that $\gamma^0_{q_1} \neq \gamma^0_{q_2}$ for all $q_1 \neq q_2$. For given $y_0 \in \{1, \ldots, Q\}$ and $x \in \mathbb{R}^{K \times T}$, let $m_{y_0}(y, x, \theta^0) \in \mathbb{R}$ be a moment function that satisfies (6) for all $\alpha \in \mathbb{R} \cup \{\pm\infty\}$. Our calculations suggest that there exist unique weights $w_{y_0}(q_1, q_2, q_3, s, y_1, \ldots, y_{t-1}, x, \theta^0) \in \mathbb{R}$ such that for all $y \in \{1, \ldots, Q\}^T$ we have

$$m_{y_0}(y, x, \theta^0) = \sum_{q_1=1}^{Q-1} \sum_{q_2=1}^{Q} \sum_{q_3=1}^{Q-1} \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} w_{y_0}(q_1, q_2, q_3, t, s, y_1, \ldots, y_{t-1}, x, \theta^0) \, m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0), \tag{13}$$

where $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ are the moment functions defined in (11) and (12). In other words, every valid moment condition in this model is a unique linear combination of the moment conditions in Theorem 2 with $r = s + 1$. Notice that the uniqueness of the linear combination implies that the moment functions involved in this linear combination are linearly independent.

The right hand side of (13) is not quite a standard basis expansion, because we have not chosen a basis for the $(y_1, \ldots, y_{t-1})$ dependence of $w_{y_0}(q_1, q_2, q_3, s, y_1, \ldots, y_{t-1}, x, \theta^0)$. This $(y_1, \ldots, y_{t-1})$ dependence implies that for any given (q_1, q_2, q_3, t, s), the function $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ gives rise to Q^{t-1} linearly independent moment functions. Thus, we conjecture that the total number of available moment conditions for each value of the covariates x and initial conditions y_0 is equal to

$$\ell = \sum_{q_1=1}^{Q-1} \sum_{q_2=1}^{Q} \sum_{q_3=1}^{Q-1} \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} Q^{t-1} = (Q-1)\, Q\, (Q-1) \sum_{t=1}^{T-2} (T-t-1)\, Q^{t-1}$$
$$= Q^T - (T-1)\, Q^2 - (T-2)\, Q.$$

As explained in Section 2.2, the function $m_{y_0}(\cdot, x, \theta^0) : \{1, \ldots, Q\}^T \to \mathbb{R}$ is a vector in a Q^T dimensional space. The condition (6), for all α, then imposes $Q^T - \ell = (T-1)\, Q^2 + (T-2)\, Q$ linear restrictions on this vector, leaving an ℓ-dimensional linear subspace of valid moment functions, a basis representation of which is given by (13).

The condition $\gamma^0_{q_1} \neq \gamma^0_{q_2}$ for all $q_1 \neq q_2$ is important for this result. For example, if all the γ^0_q are the same, then the parameter γ^0 can be absorbed into the fixed effects, and we are left with a static ordered logit model as in Muris (2017), for which one finds an additional $(T-1)(Q-1)^2$ moment conditions to be available, bringing the total number of linearly independent valid moment conditions (for each value of covariates and parameters) in the static model to $\ell = Q^T - T(Q-1) - 1$.

We reiterate that the discussion of linear independence and completeness of the moment functions presented above are conjectures which we do not aim to prove in this paper. A proof for the special case $Q = 2$ (dynamic binary choice logit models) is provided in Kruiniger (2020) and Dobronyi, Gu, and Kim (2021). We also note that counting of moment conditions as above does not consider whether the resulting moment conditions actually contain information about (all) the parameters θ. Some of the valid moment functions may not depend on (all of) those model parameters. Identification of the model parameters through the moment conditions is discussed in Section 3.

2.5 More general regressors

The model probabilities in (5) and the moment functions in (11) and (12) only depend on the regressors and the parameters β and γ through the single index $z_t = z(y_{t-1}, x_t, \theta)$.[4] So

[4]As written, the moment condition in (11) depends explicitly on the model parameter γ for the case that $y_t \leq q_1$ and $y_s > q_2$. However, that is a notational artifact, because in that line of the moment condition we

far, we have only explicitly discussed the linear specification in (3) for this single index, but Theorem 2 is valid completely independently of the functional form of $z(y_{t-1}, x_t, \theta)$.[5] In other words, if we replace the latent variable specification in (2) by

$$Y_{it}^* = z\left(Y_{i,t-1}, X_{it}, \theta\right) + A_i + \varepsilon_{it}$$

for an arbitrary function $z(\cdot,\cdot,\cdot)$, then the moment functions (8), (9), (10), and Theorem 2 remain fully valid.

We believe that the linear specification in (2) is the most relevant in practice, but one could certainly consider other specifications as well. In particular, it is possible to include regressors that are interactions between the observed regressors and the lagged dependent variable:

$$z(Y_{i,t-1}, X_{it}, \theta) := X_{it}'\,\beta + \sum_{q=1}^{Q} \gamma_q \left[\mathbb{1}\left\{Y_{i,t-1} = q\right\} + \mathbb{1}\left\{Y_{i,t-1} = q\right\} X_{it}'\,\delta_q\right], \tag{14}$$

where the $\delta_q \in \mathbb{R}^K$ are additional unknown parameters to be included in θ. This specification allows the effect of the regressors X_{it} on the outcome Y_{it} to be arbitrarily dependent on the current state $Y_{i,t-1}$. While a GMM estimator based on moment functions developed in this paper could be employed in applications with the more general state dependence as in (14), we not consider these more general models further.

could have written $\exp\left[z(y_{t-1}, x_t, \theta) - z(q_2, x_{s+1}, \theta) + \lambda_{q_3,q_1}\right]$ instead of $\exp\left(z_{t,s+1} + \gamma_{y_s,q_2} + \lambda_{q_3,q_1}\right)$; that is, the explicit dependence on γ can be fully absorbed into the single index, but one needs to evaluate $z_{s+1} = z(y_s, x_{s+1}, \theta)$ at q_2 instead of y_s.

[5]The parameters λ can also be absorbed into the single index. One just needs to define $\widetilde{z}_q(y_{t-1}, x_t, \theta) := z(y_{t-1}, x_t, \theta) - \lambda_q$ and rewrite (5) as

$$p_{y_0}(y, x, \theta, \alpha) = \prod_{t=1}^{T} \left\{ \Lambda\Big[\widetilde{z}_{y_t - 1}(y_{t-1}, x_t, \theta) + \alpha\Big] - \Lambda\Big[\widetilde{z}_{y_t}(y_{t-1}, x_t, \theta) + \alpha\Big] \right\}.$$

The moment functions in (11) and (12) then remain valid for arbitrary functional forms of $\widetilde{z}_q(y_{t-1}, x_t, \theta)$. We just need to replace $z_t - \lambda_{q_1}$, $z_s - \lambda_{q_2}$, and $z_r - \lambda_{q_3}$ (with $r = s+1$ in (11)) by $\widetilde{z}_{q_1}(y_{t-1}, x_t, \theta)$, $\widetilde{z}_{q_2}(y_{s-1}, x_s, \theta)$, and $\widetilde{z}_{q_3}(y_{r-1}, x_r, \theta)$, respectively. The proof of Theorem 2 remains valid under that replacement.

3 Identification

This section presents identification results for the parameters $\theta = (\beta, \gamma, \lambda)$ based on the moment conditions for $T = 3$ in Theorem 1. All results in this section impose the following model assumption.

Assumption ID *The outcomes* $Y = (Y_1, Y_2, Y_3)$ *are generated from model* (4) *with* $z(\cdot, \cdot, \cdot)$ *defined in* (3)*,* $Q \geq 2$*,* $T = 3$*, and true parameters* $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$*. Furthermore, for all* $y_0 \in \{1, \ldots, Q\}$ *there exists a non-empty set* $\mathcal{X}_{y_0}^{\text{reg}} \subset \mathbb{R}^{K \times 3}$ *such that for all* $x \in \mathcal{X}_{y_0}^{\text{reg}}$ *the conditional probability* $\Pr(A \in \{\pm\infty\} \mid Y_0 = y_0, \, X = x)$ *is well-defined and smaller than one.*

We impose the assumption $\Pr(A \in \{\pm\infty\} \mid Y_0 = y_0, \, X = x) < 1$ for some x to make sure that the model probabilities in (5) are strictly positive for all possible outcomes. If $\Pr(A \in \{\pm\infty\} \mid Y_0 = y_0, \, X = x) = 1$, for all x, then only the outcomes $Y_t = 1$ and $Y_t = Q$ are generated by the model. A violation of this assumption on the fixed effects A would therefore be readily observable from the data. All the propositions below also impose that $X \in \mathcal{X}_{y_0}^{\text{reg}}$ occurs with non-zero probability.

The aim is to identify the parameter vector θ^0 from the distribution of Y conditional on Y_0 and X under Assumption ID. The model for that conditional distribution is semi-parametric: The distribution of Y conditional on Y_0, X, and A is specified parametrically, but only weak regularity conditions are imposed on the unknown distribution of A conditional on Y_0 and X. The main challenge in the identification problem is how to deal with the unspecified conditional distribution of A, which is an infinite-dimensional component of the parameter space of the model. Fortunately, the moment conditions in Theorem 1 already partly solve this challenge, because they give us implications of the model that do not depend on A. The remaining question is whether θ^0 is point-identified from those moment conditions.

Identification of γ

In order to identify the parameters $\gamma = (\gamma_1, \ldots, \gamma_Q)$, up to normalization, we condition on the event $X_1 = X_2 = X_3$. For $x = (x_1, x_1, x_1)$ and $q_1 = q_2 = q_3 = 1$, the moment function in

(9) reads

$$m_{y_0}(y,\gamma) := \exp(\gamma_{y_0})\, m_{y_0,1,1,1}(y,x,\theta) = \begin{cases} -\exp(\gamma_{y_0}) & \text{if } y_1 = 1,\ y_2 > 1, \\ \exp(\gamma_1) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 = 1, \\ \exp\left(\gamma_{y_1}\right) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Theorem 1 implies that $\mathbb{E}\left[m_{y_0}(Y,\gamma^0) \,\middle|\, Y_0 = y_0,\ X = (x_1,x_1,x_1)\right] = 0$. The following lemma states that these moment conditions are sufficient to uniquely identify γ up to a normalization.

Proposition 1 *Let Assumption ID hold, and let $x_1 \in \mathbb{R}$ be such that*

$$\Pr\left(Y_0 = y_0 \ \&\ X \in \mathcal{X}^{\text{reg}}_{y_0} \ \&\ \|X - (x_1,x_1,x_1)\| \le \epsilon\right) > 0 \qquad \textit{for all } y_0 \in \{1,\dots,Q\} \textit{ and } \epsilon > 0.$$

Then, if $\gamma \in \mathbb{R}^Q$ satisfies

$$\mathbb{E}\left[m_{y_0}(Y,\gamma) \,\middle|\, Y_0 = y_0,\ X = (x_1,x_1,x_1)\right] = 0 \qquad \textit{for all } y_0 \in \{1,\dots,Q\}, \tag{16}$$

for $m_{y_0}(y,\gamma)$ as defined in (15), we have $\gamma = \gamma^0 + c$ for some $c \in \mathbb{R}$. Thus, if we normalize $\gamma^0_1 = 0$, then γ^0 is uniquely identified from the data.

The proof is given in the appendix. This identification result requires observed data for every initial condition $y_0 \in \{1,\dots,Q\}$. If this is not available, but we observe $T = 4$ time periods of data after the initial condition, then we can instead apply Proposition 1 to the data shifted by one time period.

In addition to Assumption ID, the proposition demands that covariate values $X \in \mathcal{X}^{\text{reg}}_{y_0}$ in any ϵ-ball around (x_1,x_1,x_1) occur with positive probability. This condition, in particular, guarantees that the conditional expectation in (16) is well-defined, and that conditional on $X = (x_1,x_1,x_1)$ the event $A \in \{\pm\infty\}$ occurs with probability less than one for every value of the initial condition Y_0.

Identification of β

Taking the identification result for γ as given, we now turn to the problem of identifying β. We again consider the moment function in (9) with $q_1 = q_2 = q_3 = 1$, but now for general regressor values

$$m_{y_0,1,1,1}(y,x,\beta,\gamma) := m_{y_0,1,1,1}(y,x,\theta) = \begin{cases} \exp\left(x_{23}'\beta\right) - 1 & \text{if } y_1 = 1,\ y_2 = 1,\ y_3 > 1, \\ -1 & \text{if } y_1 = 1,\ y_2 > 1, \\ \exp\left(x_{31}'\beta + \gamma_1 - \gamma_{y_0}\right) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 = 1, \\ \exp\left(x_{21}'\beta + \gamma_{y_1} - \gamma_{y_0}\right) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

For $k \in \{1,\ldots,K\}$ we define

$$\mathcal{X}_{k,+} := \{x \in \mathcal{X}_{y_0}^{\text{reg}} \,:\, x_{k,1} \leq x_{k,3} < x_{k,2} \ \text{ or } \ x_{k,1} < x_{k,3} \leq x_{k,2}\},$$
$$\mathcal{X}_{k,-} := \{x \in \mathcal{X}_{y_0}^{\text{reg}} \,:\, x_{k,1} \geq x_{k,3} > x_{k,2} \ \text{ or } \ x_{k,1} > x_{k,3} \geq x_{k,2}\}.$$

Here, the set $\mathcal{X}_{k,+}$ is the set of possible regressor values $x \in \mathbb{R}^{K\times 3}$ such that $x_{k,1} \leq x_{k,3} \leq x_{k,2}$ with at least one of the inequalities being strict. For the set $\mathcal{X}_{k,-}$ those inequalities are reversed. Therefore, if $\beta \in \mathcal{X}_{k,+}$, then $m_{y_0,1,1,1}(y,x,\beta,\gamma)$ is strictly increasing in β_k, and if $\beta \in \mathcal{X}_{k,-}$, then $m_{y_0,1,1,1}(y,x,\beta,\gamma)$ is strictly decreasing in β_k.

For any vector $s \in \{-,+\}^K$ we furthermore define the set $\mathcal{X}_s = \bigcap_{k\in\{1,\ldots,K\}} \mathcal{X}_{k,s_k}$. If $\beta \in \mathcal{X}_s$, then for all $k \in \{1,\ldots,K\}$ we have that β_k is strictly increasing (or strictly decreasing) in $m_{y_0,1,1,1}(y,x,\beta,\gamma)$ if $s_k = +$ (or $s_k = -$). Those monotonicity properties allow us to uniquely identify β from the moment conditions $\mathbb{E}\left[m_{y_0,1,1,1}(y,x,\beta^0,\gamma^0) \,\middle|\, Y_0 = y_0,\ X \in \mathcal{X}_s\right] = 0$, which are valid moment conditions according to Theorem 1. The following proposition formalizes this.

Proposition 2 *Let Assumption ID hold and let $y_0 \in \{1,\ldots,Q\}$ be such that*

$$\Pr\left(Y_0 = y_0 \;\&\; X \in \mathcal{X}_s\right) > 0 \qquad \textit{for all } s \in \{-,+\}^K \textit{ with } s_K = +.$$

Then, if $\beta \in \mathbb{R}^K$ *satisfies*

$$\mathbb{E}\left[m_{y_0,1,1,1}(y,x,\beta,\gamma^0)\,\middle|\,Y_0=y_0,\ X\in\mathcal{X}_s\right]=0 \qquad \textit{for all } s\in\{-,+\}^K \textit{ with } s_K=+, \quad (18)$$

we have $\beta = \beta^0$*. Thus, since* γ^0 *is already identified from Proposition 1, we find that* β^0 *is also uniquely identified from the data.*

The proof is given in the appendix. Again, in addition to Assumption ID the additional condition in Proposition 2 simply guarantees that the conditional expectation in (18) is well-defined.

Identification of λ

Having identified γ and β already, we now turn to the problem of identifying λ, up to a normalization. The moment function in (9) with $q_2 = q_3 = 1$ and $q_1 \in \{2,\dots,Q-1\}$ can be written as

$$m_{y_0,q_1,1,1}(y,x,\beta,\gamma,\lambda)=\begin{cases}\exp\left(x_{23}'\,\beta+\gamma_{y_1,1}\right)-1 & \text{if } y_1\le q_1,\ y_2=1,\ y_3>1,\\ -1 & \text{if } y_1\le q_1,\ y_2>1,\\ \exp\left(x_{31}'\,\beta+\gamma_{1,y_0}+\lambda_{q_1}-\lambda_1\right) & \text{if } y_1>q_1,\ y_2=1,\ y_3=1,\\ \exp\left(x_{21}'\,\beta+\gamma_{y_1,y_0}+\lambda_{q_1}-\lambda_1\right) & \text{if } y_1>q_1,\ y_2=1,\ y_3>1,\\ 0 & \text{otherwise.}\end{cases} \quad (19)$$

The expected value of this moment function only depends on λ through $\lambda_{q_1}-\lambda_1$, and is strictly increasing in $\lambda_{q_1}-\lambda_1$. This implies that this moment function identifies $\lambda_{q_1}-\lambda_1$ uniquely. By applying this argument for all $q_1 \in \{2,\dots,Q-1\}$ we can therefore identify λ up to an additive constant. This is summarized in the following proposition.

Proposition 3 *Let Assumption ID hold. Let* $y_0 \in \{1,\dots,Q\}$ *be such that* $\Pr\big(Y_0 = y_0 \ \&\ X \in \mathcal{X}_{y_0}^{\text{reg}}\big) > 0$*. Then, if* λ *satisfies*

$$\mathbb{E}\left[m_{y_0,q_1,1,1}(Y,X,\beta^0,\gamma^0,\lambda)\,\middle|\,Y_0=y_0\right]=0 \qquad \textit{for all } q_1\in\{2,\dots,Q-1\},$$

we have $\lambda = \lambda^0 + c$ for some $c \in \mathbb{R}$. Thus, if we normalize $\lambda_1^0 = 0$, and since γ^0 and β^0 are already identified from Proposition 1 and 2, we find that λ^0 is also uniquely identified from the data.

The proof is given in the appendix.

Combining Proposition 1, 2, and 3, we find that θ^0 is uniquely identified from the data. Under the regularity conditions of those propositions we can recover $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$ uniquely from the distribution of Y conditional on Y_0 and X.

Our identification arguments in this section are constructive. However, they condition on special values of the regressors. In particular, Proposition 1 conditions on the event $X_1 = X_2 = X_3$, which is a zero-probability event if X is continuously distributed (and may happen rarely even for discrete X). An estimator based on the identification strategy in this section would therefore in general be quite inefficient. In our Monte Carlo simulations and empirical application we therefore construct more general GMM estimators based on our moment conditions.

4 Implication for estimation and specification testing

The moment conditions in Section 2 are conditional on the initial condition Y_{i0} and the strictly exogenous explanatory variables X_i. A set of unconditional moment functions can be formed by constructing

$$M(Y_{i0}, Y_i, X_i, \beta, \gamma, \lambda) = g\left(Y_{i0}, X_i\right) \otimes m_{Y_{i0}}\left(Y_i, X_i, \beta, \gamma, \lambda\right)$$

where the vector-valued function, $m_{Y_{i0}}$, is composed of linear combinations of the moment functions in (8), (9), and (10), and g is a vector valued function of the initial conditions and the strictly exogenous X_i. Let $\theta = (\beta', \gamma', \lambda')'$. A generalized method of moments (GMM) estimator can then be defined by[6]

$$\widehat{\theta}=\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\\ \widehat{\lambda}\end{pmatrix}=\underset{\beta\in\mathbb{R}^K,\gamma\in\mathbb{R}^{Q-1},\lambda\in\mathbb{R}^{Q-2}}{\operatorname{argmin}}\left(\sum_{i=1}^{n}M(Y_{i0},Y_i,X_i,\beta,\gamma,\lambda)\right)'$$
$$\widehat{W}_n\left(\sum_{i=1}^{n}M(Y_{i0},Y_i,X_i,\beta,\gamma,\lambda)\right),$$

where the weighting matrix $\widehat{W}_n$ converges to a positive definite matrix, W_0. Assuming that $\mathbb{E}\left[M(Y_{i0},Y_i,X_i,\theta)\right]=0$ is *uniquely* satisfied at $\theta=\theta_0$, and that mild regularity conditions (see Hansen 1982) are satisfied, $\widehat{\theta}$ will be consistent and asymptotically normally distributed. Specifically, with a random sample, $\{Y_{i0},Y_i,X_i\}_{i=1}^n$,

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\overset{d}{\longrightarrow}\mathcal{N}\left(0,\left(\Gamma'W_0\Gamma\right)^{-1}\Gamma'W_0SW_0\Gamma\left(\Gamma'W_0\Gamma\right)^{-1}\right)$$

with $\Gamma=\mathbb{E}\left[\frac{\partial M(Y_{i0},Y_i,X_i,\theta_0)}{\partial\theta}\right]$, $W_0=\operatorname{plim}\widehat{W}_n$ and $S=V\left[M(Y_{i0},Y_i,X_i,\theta_0)\right]$.

The main limitation of the GMM approach is that it is often difficult to know whether $\mathbb{E}\left[M(Y_{i0},Y_i,X_i,\theta)\right]=0$ is uniquely satisfied at the true parameter value. When the strictly exogenous explanatory variables are discrete, sufficient conditions for this can be obtained from the identification results in Section 3 be defining $g\left(Y_{i0},X_i\right)$ to be a vector of indicator functions for values in the support of (Y_{i0},X_i). If X_i is not discrete, it may be possible to define a root-n consistent estimator by combining nonparametrically estimated conditional moment conditions with the unconditional moment conditions. See, for example, Honoré and Hu (2004) for such an approach. Whether or not $\mathbb{E}\left[M(Y_{i0},Y_i,X_i,\theta)\right]=0$ is uniquely satisfied at the true parameter value, one can estimate confidence sets for θ_0 by inverting tests for the hypothesis that $\mathbb{E}\left[M(Y_{i0},Y_i,X_i,\theta)\right]=0$.

The moment conditions derived in this paper can also be used for specification testing. Suppose that a researcher has estimated the parameters of interest, $\theta_0=(\beta_0,\gamma_0,\lambda_0)$, by an estimator that solves a moment condition of the type $\frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i,X_i,\widehat{\theta}\right)=0$. For example,

[6] As mentioned in Section 2, it is necessary to normalize one of the Q elements of γ and one of the $Q-1$ elements of λ.

she might have estimated a model without individual-specific heterogeneity or a model in which the heterogeneity is captured parametrically by a random effects approach, and she might be interested in testing her parametric assumptions against the less parametric fixed effects model. Let $\widehat{M} = \frac{1}{n}\sum_{i=1}^{n} M\left(Y_i, X_i, \widehat{\theta}\right)$ where M is defined as above. $\widehat{M}$ is then a standard two-step estimator, and it is straightforward to test whether $\widehat{M}$ is statistically different from 0.

5 Practical Performance of GMM Estimator

In the next subsection, we present the results from a small Monte Carlo experiment designed to assess the performance of a GMM estimator based on the discussion in Section 4. Following that, we illustrate its use in an empirical example. Section A.3 provides details about the implementation of the GMM estimator.

5.1 Monte Carlo Results

We illustrate the performance of the GMM estimator described above through a small Monte Carlo study. We consider sample sizes of $N = 1000$, 3000, and 9000 with four time periods for each individual. This includes the initial observations, so $T = 3$ using the notation above. There are $k = 3$ explanatory variables and the dependent variable can take 4 values.

The explanatory variables are drawn as follows. Let Z_i and Z_{ijt} ($j = 1, ..., k$, $t = 0, ..., 3$) be independent standard normal random variables. The second through k'th explanatory variables are given by $X_{ijt} = \sqrt{3}\left(Z_{ijt} + Z_{i1t}\right)/\sqrt{2}$, while the first explanatory variable is $X_{i1t} = \sqrt{3}\left(Z_i + Z_{i1t}\right)/\sqrt{2}$. This implies that in each time period, the explanatory variables will all have variance equal to 3 and their pairwise correlations are all 0.5. The correlation in the first explanatory variable in any two time periods is 0.5, while the remaining covariates are independent over time. We consider one specification without heterogeneity, in which case $A_i = 0$, and one with heterogeneity in which case $A_i = \sqrt{3}Z_i$. The first element of β is 1, and the remaining elements are 0. This makes the variance of $X_{it}'\beta$ close to that of the logistic distribution as well as close to that of the fixed effect in the specification with heterogeneity. We use $\gamma = (-1, 0, 0, 1)'$ and $\lambda = (-2, 0, 2)'$ and normalize $\gamma_2 = \lambda_2 = 0$. The dependent variables are generated from the model with the lagged dependent variables in period 0 set to 0.

We perform 400 Monte Carlo replications. The results are presented in Tables 1 and 2.

Table 1: Without Heterogeneity

$N = 1000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.030	0.002	0.013	-0.460	0.000	-0.186	0.498	-2.331	0.000	2.266
MAE	0.118	0.065	0.061	0.547	0.000	0.229	0.509	0.332	0.000	0.268
IQR	0.231	0.128	0.125	0.432	0.000	0.374	0.406	0.298	0.000	0.235

$N = 3000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.011	0.002	-0.002	-0.658	0.000	-0.180	0.683	-2.151	0.000	2.119
MAE	0.067	0.036	0.036	0.345	0.000	0.203	0.323	0.157	0.000	0.123
IQR	0.132	0.071	0.074	0.329	0.000	0.269	0.321	0.151	0.000	0.128

$N = 9000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.006	-0.002	-0.002	-0.871	0.000	-0.086	0.858	-2.043	0.000	2.046
MAE	0.036	0.021	0.022	0.153	0.000	0.107	0.153	0.059	0.000	0.051
IQR	0.072	0.043	0.044	0.193	0.000	0.174	0.234	0.088	0.000	0.069

The results in Tables 1 and 2 suggest that our implementation of the GMM estimator underestimates the degree of state dependence, although it does improve with sample size. The estimator of the coefficient on the strictly exogenous explanatory variables are close to median unbiased whether or not the data generating process includes fixed effects.

5.2 Empirical Illustration

In the section, we illustrate the value of the moment conditions derived in the paper in an empirical example inspired by Contoyannis, Jones, and Rice (2004). The dependent variable is self-reported health status, and we use data from the first four waves of the British Household Panel Survey. This yields a data set with 7233 individuals observed in 4 time periods including the initial observation (so $T = 3$). In the original data set, the dependent

Table 2: With Heterogeneity

$N = 1000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.049	0.016	0.019	-0.369	0.000	-0.131	0.360	-2.421	0.000	2.385
MAE	0.153	0.081	0.077	0.634	0.000	0.244	0.656	0.426	0.000	0.391
IQR	0.287	0.160	0.159	0.487	0.000	0.471	0.579	0.366	0.000	0.343

$N = 3000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.032	0.009	0.004	-0.549	0.000	-0.194	0.448	-2.263	0.000	2.240
MAE	0.095	0.049	0.052	0.459	0.000	0.237	0.552	0.265	0.000	0.242
IQR	0.186	0.098	0.101	0.404	0.000	0.375	0.388	0.241	0.000	0.220

$N = 9000, k = 3$

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
True	1.000	0.000	0.000	-1.000	0.000	0.000	1.000	-2.000	0.000	2.000
Median	1.019	0.003	-0.000	-0.711	0.000	-0.148	0.635	-2.134	0.000	2.139
MAE	0.059	0.033	0.033	0.302	0.000	0.176	0.369	0.144	0.000	0.147
IQR	0.117	0.066	0.064	0.299	0.000	0.247	0.315	0.168	0.000	0.187

variable can take five values. We aggregate these into "Poor or Very Poor" (8.0% of the observations), "Fair" (19.1%), "Good" (47.2%), and "Excellent" (25.7%). We also consider specifications where the first two are merged into one outcome.

We use two sets of explanatory variables. In the first, we use age and age-squared (measured as $Age/10$ and $(Age - 45)^2/1000$, respectively, where Age is measured in years). In the second, we also include log-income. The results are presented in Table 3. We have normalized the δ-coefficient associated with "Good Health" to be 0 and the threshold (λ) just below "Good Health" to be 0. The most consistent result presented in Table 3 is a concave relationship between age and self-reported age. For all of the specifications, this relationship is decreasing after the age of 30.

The point estimates for the effect of income on self-reported health are positive in columns two and four, but neither is statistically significant.

Table 3: Empirical Results

	Four Outcomes		Three Outcomes	
$Age/10$	-1.214	-1.829	-0.980	-1.261
	(0.306)	(0.309)	(0.488)	(0.475)
$(Age-45)^2/1000$	-0.200	-0.261	-0.304	-0.236
	(0.074)	(0.080)	(0.108)	(1.411)
log-income		0.063		0.196
		(0.070)		(0.117)
δ_1	-0.310	-0.339		
	(0.166)	(0.146)		
δ_2	-0.313	-0.338	-0.535	-0.489
	(0.130)	(0.129)	(0.338)	(0.239)
δ_3	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
δ_4	-0.180	-0.263	0.127	0.021
	(0.128)	(0.113)	(0.146)	(0.115)
λ_1	-2.756	-3.050		
	(0.086)	(0.101)		
λ_2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
λ_3	3.693	4.137	3.395	3.440
	(0.126)	(0.128)	(0.155)	(0.120)

6 Conclusions

This paper has extended the analysis in Honoré and Weidner (2020) to provide conditional moment conditions for panel data fixed effects versions of the dynamic ordered logit models like the one considered in Muris, Raposo, and Vandoros (2020). The moment conditions are interesting in their own right, and the paper also illustrates the potential for systematically deriving moment conditions for nonlinear panel models. The moment conditions presented here can be used for estimation as well as for testing more parametric specifications of the individual-specific effects in dynamic ordered logits. For point-identification, it is important to investigate whether the moment conditions are *uniquely* satisfied at the true parameter values. The paper presents conditions under which this is the case. The paper also proposes a practical strategy for turning the derived conditional moment conditions into unconditional moment conditions that can be used to GMM estimation, and it illustrates the use of the resulting estimator in a small Monte Carlo study as well as in an empirical application.

More broadly, this paper contributes to the literature that is concerned with panel data estimation of nonlinear models with fixed effects. In this context, the main contribution is to illustrate the potential for applying the functional differencing insights of Bonhomme (2012) to logit-type models.

References

Aguirregabiria, V., and J. M. Carro (2021): "Identification of Average Marginal Effects in Fixed Effects Dynamic Discrete Choice Models," pp. 1–31.

Aguirregabiria, V., J. Gu, and Y. Luo (2021): "Sufficient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models," 223(2), 280–311.

Albarran, P., R. Carrasco, and J. M. Carro (2019): "Estimation of Dynamic Nonlinear Random Effects Models with Unbalanced Panels," Oxford Bulletin of Economics and Statistics, 81(6), 1424–1441.

Arellano, M. (2003): "Discrete choices with panel data," Investigaciones económicas, 27(3), 423–458.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," The Review of Economic Studies, 58(2), 277.

Arellano, M., and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," Annual Review of Economics, 3(1), 395–424.

Aristodemou, E. (2021): "Semiparametric Identification in Panel Data Discrete Response Models," Journal of Econometrics, 220(2), 253–271.

Baetschmann, G., K. E. Staub, and R. Winkelmann (2015): "Consistent estimation of the fixed effects ordered logit model," Journal of the Royal Statistical Society A, 178(3), 685–703.

Blundell, R., and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, 87(1), 115–143.

Bonhomme, S. (2012): "Functional differencing," Econometrica, 80(4), 1337–1385.

Botosaru, I., C. Muris, and K. Pendakur (2021): "Identification of Time-Varying Transformation Models with Fixed Effects, with an Application to Unobserved Heterogeneity in Resource Shares," Journal of Econometrics, forthcoming.

Carro, J. M., and A. Traferri (2014): "State Dependence and Heterogeneity in Health Using a Bias-Corrected Fixed-Effects Estimator," Journal of Applied Econometrics, 29(2), 181–207.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," The Review of Economic Studies, 47(1), 225–238.

——— (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in Longitudinal Analysis of Labor Market Data, ed. by J. J. Heckman, and B. Singer, no. 10 in Econometric Society Monographs series, pp. 3–38. Cambridge University Press, Cambridge, New York and Sydney.

Contoyannis, P., A. M. Jones, and N. Rice (2004): "The dynamics of health in the British Household Panel Survey," Journal of Applied Econometrics, 19(4), 473–503.

Das, M., and A. van Soest (1999): "A panel data model for subjective information on household income growth," Journal of Economic Behavior & Organization, 40(4), 409–426.

Dobronyi, C., J. Gu, and K. i. Kim (2021): "Identification of Dynamic Panel Logit Models with Fixed Effects," arXiv preprint arXiv:2104.04590.

Fernández-Val, I., Y. Savchenko, and F. Vella (2017): "Evaluating the role of income, state dependence and individual specific heterogeneity in the determination of subjective health assessments," Economics & Human Biology, 25, 85–98.

Hahn, J. (1997): "A Note on the Efficient Semiparametric Estimation of Some Exponential Panel Models," Econometric Theory, 13(4), 583–588.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50(4), pp. 1029–1054.

Honoré, B. E. (2002): "Nonlinear models with panel data," Portuguese Economic Journal, 1(2), 163.

Honoré, B. E., and L. Hu (2004): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," Journal of Econometrics, 122(2), 293–316.

Honoré, B. E., and E. Kyriazidou (2000): "Panel data discrete choice models with lagged dependent variables," Econometrica, 68(4), 839–874.

Honoré, B. E., and M. Weidner (2020): "Moment Conditions for Dynamic Panel Logit Models with Fixed Effects," arXiv preprint arXiv:2005.05942.

Johnson, E. G. (2004): "Panel Data Models With Discrete Dependent Variables," Ph.D. thesis, Stanford University.

Khan, S., F. Ouyang, and E. Tamer (2021): "Inference on Semiparametric Multinomial Response Models," Quantitative Economics, 12, 743–777.

Kitazawa, Y. (2021): "Transformations and moment conditions for dynamic fixed effects logit models," Journal of Econometrics.

Kruiniger, H. (2020): "Further results on the estimation of dynamic panel logit models with fixed effects," arXiv preprint arXiv:2010.03382.

Magnac, T. (2000): "Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories," The Economic Journal, 110(466), 805–837.

Muris, C. (2017): "Estimation in the Fixed-Effects Ordered Logit Model," The Review of Economics and Statistics, 99(3), 465–477.

Muris, C., P. Raposo, and S. Vandoros (2020): "A dynamic ordered logit model with fixed effects," arXiv preprint arXiv:2008.05517.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, ed. by R. F. Engle, and D. L. McFadden, no. 4 in Handbooks in Economics,, pp. 2111–2245. Elsevier, North-Holland, Amsterdam, London and New York.

Neyman, J., and E. L. Scott (1948): "Consistent estimates based on partially consistent observations," Econometrica, 16, 1–32.

Pakes, A., J. Porter, M. Shepard, and S. Calder-Wang (2021): "Unobserved Heterogeneity, State Dependence, and Health Plan Choices," .

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," Econometrica, 86(2), 737–761.

A Appendix

A.1 Proof of Theorem 1 and 2

We first want to establish Lemma 1 below, which is key to proving the main text theorems. In order to state the lemma, we require some additional notation. Recall that $Q \in \{2, 3, \ldots\}$ is the number of values that the observed outcomes Y_{it} can take. Let $\widetilde{Y}_1, \widetilde{Y}_3 \in \{0, 1\}$, $\widetilde{Y}_2 \in \{1, 2, 3\}$, and $W \in \{1, 2, \ldots, Q\}$ be random variables, and let $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$. For the joint distribution of $\widetilde{Y}$ and W we write

$$p(\widetilde{y}, w) := \Pr\left(\widetilde{Y} = \widetilde{y} \;\;\&\;\; W = w\right)$$

and we assume that

$$p(\widetilde{y}, w) = p_3(\widetilde{y}_3 \,|\, \widetilde{y}_2, w)\;\; p_2(\widetilde{y}_2 \,|\, w)\;\; f(w \,|\, \widetilde{y}_1)\;\; p_1(\widetilde{y}_1), \tag{20}$$

where $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$, and

$$\begin{aligned} p_1(\widetilde{y}_1) &:= \Pr\left(\widetilde{Y}_1 = \widetilde{y}_1\right), \\ f(w \,|\, \widetilde{y}_1) &:= \Pr\left(W = w \,\middle|\, \widetilde{Y}_1 = \widetilde{y}_1\right), \\ p_2(\widetilde{y}_2 \,|\, w) &:= \Pr\left(\widetilde{Y}_2 = \widetilde{y}_2 \,\middle|\, W = w\right), \\ p_3(\widetilde{y}_3 \,|\, \widetilde{y}_2, w) &:= \Pr\left(\widetilde{Y}_3 = \widetilde{y}_3 \,\middle|\, \widetilde{Y}_2 = \widetilde{y}_2,\, W = w\right). \end{aligned}$$

We do not impose any assumptions on the transition probabilities $f(w \,|\, \widetilde{y}_1)$, $p_3(\widetilde{y}_3 \,|\, 1, w)$, and $p_3(\widetilde{y}_3 \,|\, 3, w)$. All the other transition probabilities are assumed to follow an (ordered) logit model:

$$p_1(\widetilde{y}_1) = \begin{cases} 1 - \Lambda(\pi_1) & \text{if } \widetilde{y}_1 = 0, \\ \Lambda(\pi_1) & \text{if } \widetilde{y}_1 = 1, \end{cases}$$

$$p_2(\widetilde{y}_2 \,|\, w) = \begin{cases} 1 - \Lambda[\pi_{2,1}(w)] & \text{if } \widetilde{y}_2 = 1, \\ \Lambda[\pi_{2,1}(w)] - \Lambda[\pi_{2,2}(w)] & \text{if } \widetilde{y}_2 = 2, \\ \Lambda[\pi_{2,2}(w)] & \text{if } \widetilde{y}_2 = 3, \end{cases}$$

$$p_3(\widetilde{y}_3 \,|\, 2,\, w) = \begin{cases} 1 - \Lambda(\pi_3) & \text{if } \widetilde{y}_3 = 0, \\ \Lambda(\pi_3) & \text{if } \widetilde{y}_3 = 1, \end{cases} \tag{21}$$

where $\Lambda(\xi) := [1 + \exp(-\xi)]^{-1}$ is the cumulative distribution function of the logistic distribution, $\pi_1, \pi_3 \in \mathbb{R}$ are constants, and $\pi_{2,1}, \pi_{2,2} : \{1, 2, \ldots, Q\} \to \mathbb{R}$ are functions such that $\pi_{2,1}(w) \geq \pi_{2,2}(w)$ for all $w \in \{1, 2, \ldots, Q\}$. Notice that $p_3(\widetilde{y}_3 \,|\, 2,\, w)$ does not depend on w. Finally, we define $m : \{0,1\} \times \{1,2,3\} \times \{0,1\} \times \{1,2,\ldots,Q\} \to \mathbb{R}$ by

$$m(\widetilde{y}, w) := \begin{cases} \exp(\pi_1 - \pi_3) \dfrac{\exp[\pi_3 - \pi_{2,2}(w)] - 1}{\exp[\pi_{2,1}(w) - \pi_{2,2}(w)] - 1} & \text{if } \widetilde{y}_1 = 0,\, \widetilde{y}_2 = 2,\, \widetilde{y}_3 = 0, \\ \exp(\pi_1 - \pi_3) \dfrac{1 - \exp[\pi_{2,2}(w) - \pi_3]}{1 - \exp[\pi_{2,2}(w) - \pi_{2,1}(w)]} & \text{if } \widetilde{y}_1 = 0,\, \widetilde{y}_2 = 2,\, \widetilde{y}_3 = 1, \\ \exp(\pi_1 - \pi_3) & \text{if } \widetilde{y}_1 = 0,\, \widetilde{y}_2 = 3, \\ -1 & \text{if } \widetilde{y}_1 = 1,\, \widetilde{y}_2 = 1, \\ -\dfrac{1 - \exp[\pi_3 - \pi_{2,1}(w)]}{1 - \exp[\pi_{2,2}(w) - \pi_{2,1}(w)]} & \text{if } \widetilde{y}_1 = 1,\, \widetilde{y}_2 = 2,\, \widetilde{y}_3 = 0, \\ -\dfrac{\exp[\pi_{2,1}(w) - \pi_3] - 1}{\exp[\pi_{2,1}(w) - \pi_{2,2}(w)] - 1} & \text{if } \widetilde{y}_1 = 1,\, \widetilde{y}_2 = 2,\, \widetilde{y}_3 = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Lemma 1 *Let $\pi_1, \pi_3 \in \mathbb{R}$, and $\pi_{2,1}, \pi_{2,2} : \{1, 2, \ldots, Q\} \to \mathbb{R}$ be such that $\pi_{2,1}(w) \geq \pi_{2,2}(w)$, for all $w \in \{1, 2, \ldots, Q\}$. Let the random variables $\widetilde{Y} \in \{0,1\} \times \{1,2,3\} \times \{0,1\}$ and $W \in \{1, 2, \ldots, Q\}$ be such that their distributions satisfy (20) and (21), and let $m : \{0,1\} \times \{1,2,3\} \times \{0,1\} \times \{1,2,\ldots,Q\} \to \mathbb{R}$ be defined by (22). Then we have*

$$\mathbb{E}\left[m(\widetilde{Y}, W)\right] = 0.$$

Proof. Define

$$
\begin{aligned}
g(\widetilde{y}_1, w) &:= \mathbb{E}\left[m(\widetilde{Y}, W) \,\middle|\, \widetilde{Y}_1 = \widetilde{y}_1,\, W = w \right] \\
&= \sum_{\widetilde{y}_2 \in \{1,2,3\}} \sum_{\widetilde{y}_3 \in \{0,1\}} m(\widetilde{y}, w) \;\; p_3(\widetilde{y}_3 \,|\, \widetilde{y}_2, w) \;\; p_2(\widetilde{y}_2 \,|\, w),
\end{aligned}
$$

where $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$. Using the expressions for $p_2(\widetilde{y}_2 \,|\, w)$, $p_3(\widetilde{y}_3 \,|\, \widetilde{y}_2, w)$, and $m(\widetilde{y}, w)$ in (21) and (22) one finds that for $\widetilde{y}_1 = 1$ we have

$$
\begin{aligned}
g(1, w) = & - \{1 - \Lambda[\pi_{2,1}(w)]\} - \{\Lambda[\pi_{2,1}(w)] - \Lambda[\pi_{2,2}(w)]\} \times \\
& \times \underbrace{\left(\frac{[1 - \Lambda(\pi_3)] \, \{1 - \exp[\pi_3 - \pi_{2,1}(w)]\}}{1 - \exp\left[\pi_{2,2}(w) - \pi_{2,1}(w)\right]} + \frac{\Lambda(\pi_3) \, \{\exp[\pi_{2,1}(w) - \pi_3] - 1\}}{\exp\left[\pi_{2,1}(w) - \pi_{2,2}(w)\right] - 1} \right)}_{= \frac{\Lambda[\pi_{2,1}(w)] - \Lambda(\pi_3)}{\Lambda[\pi_{2,1}(w)] - \Lambda[\pi_{2,2}(w)]}} \\
= & - [1 - \Lambda(\pi_3)] \,, \qquad (23)
\end{aligned}
$$

and analogously one calculates for $\widetilde{y}_1 = 0$ that

$$
g(0, w) = \exp\left(\pi_1 - \pi_3\right) \Lambda(\pi_3). \qquad (24)
$$

Notice that $g(\widetilde{y}_1, w)$ therefore does not depend on w, so we can simply write $g(\widetilde{y}_1) := g(\widetilde{y}_1, w)$ in the following. Using (23) (24), and the expression for $p_1(\widetilde{y}_1)$ in (21) we obtain that

$$
\sum_{\widetilde{y}_1 \in \{0,1\}} g(\widetilde{y}_1) \, p_1(\widetilde{y}_1) = 0.
$$

Together with $\sum_{w \in \{1,\ldots,Q\}} f(w \,|\, \widetilde{y}_1) = 1$, this gives

$$
\begin{aligned}
\mathbb{E}\left[m(\widetilde{Y}, W) \right] &= \sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{w \in \{1,\ldots,Q\}} \sum_{\widetilde{y}_2 \in \{1,2,3\}} \sum_{\widetilde{y}_3 \in \{0,1\}} m(\widetilde{y}, w) \, p(\widetilde{y}, w) \\
&= \sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{w \in \{1,\ldots,Q\}} g(\widetilde{y}_1) \, f(w \,|\, \widetilde{y}_1) \; p_1(\widetilde{y}_1)
\end{aligned}
$$

$$= \sum_{\widetilde{y}_1 \in \{0,1\}} g(\widetilde{y}_1) \underbrace{\left[\sum_{w \in \{1,\ldots,Q\}} f(w \,|\, \widetilde{y}_1) \right]}_{=1} p_1(\widetilde{y}_1)$$

$$= \sum_{\widetilde{y}_1 \in \{0,1\}} g(\widetilde{y}_1)\, p_1(\widetilde{y}_1) = 0,$$

which is what we wanted to show. ■

The following lemma is similar to Lemma 1 above, but the random variables $\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3$ and their distributional assumptions are now different, and the lemma should be understood independently from any notation established above.

Lemma 2 *Let $\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3 \in \{0,1\}$ and $W, V \in \{1,\ldots,Q\}$ be random variables such that the joint distribution of $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$, W, and V satisfies*

$$\Pr\Big(\widetilde{Y} = \widetilde{y} \;\;\&\;\; W = w \;\;\&\;\; V = v\Big) = p_3(\widetilde{y}_3 \,|\, v)\;\; g(v \,|\, \widetilde{y}_2, w)\;\; p_2(\widetilde{y}_2 \,|\, w)\;\; f(w \,|\, \widetilde{y}_1)\;\; p_1(\widetilde{y}_1),$$

where $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$, and the functions p_3, g, p_2, and f are appropriate conditional probabilities, while $p_1(\widetilde{y}_1)$ is the marginal distribution of $\widetilde{Y}_1$. For $p_1(\widetilde{y}_1)$, $p_2(\widetilde{y}_2 \,|\, w)$, and $p_3(\widetilde{y}_3 \,|\, v)$ we assume logistic binary choice models:

$$p_1(\widetilde{y}_1) = \Lambda\left[(2\,\widetilde{y}_1 - 1)\,\pi_1\right], \quad p_2(\widetilde{y}_2 \,|\, w) = \Lambda\left[(2\,\widetilde{y}_2 - 1)\,\pi_2(w)\right], \quad p_3(\widetilde{y}_3 \,|\, v) = \Lambda\left[(2\,\widetilde{y}_3 - 1)\,\pi_3(v)\right],$$

where $\pi_1 \in \mathbb{R}$ is a constant, and $\pi_2, \pi_3 : \{1,\ldots,Q\} \to \mathbb{R}$ are functions. The only assumption that we impose on $f(w \,|\, \widetilde{y}_1)$ and $g(v \,|\, \widetilde{y}_2, w)$ is that $g(v \,|\, 1, w) = g(v \,|\, 1)$; that is, conditional on $\widetilde{Y}_2 = 1$ the distribution of V is independent of W. Furthermore, let $m : \{0,1\}^3 \times \{1,\ldots,Q\}^2 \to \mathbb{R}$ be given by

$$m(\widetilde{y}, w, v) := \begin{cases} \exp\left[\pi_1 - \pi_2(w)\right] & \text{if } \widetilde{y} = (0,1,0), \\ \exp\left[\pi_1 - \pi_3(v)\right] & \text{if } \widetilde{y} = (0,1,1), \\ -1 & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,0), \\ \exp\left[\pi_3(v) - \pi_2(w)\right] - 1 & \text{if } \widetilde{y} = (1,1,0), \\ 0 & \text{otherwise.} \end{cases}$$

We then have

$$\mathbb{E}\left[m(\widetilde{Y}, W, V)\right] = 0.$$

Proof. This lemma is a restatement of Lemma 6 in the 2021 arXiv version of Honoré and Weidner (2020), and the proof can be found there. ∎

Using Lemma 1 and 2, we are now ready to prove the two main text theorems.

Proof of Theorem 1. We consider the three cases $q_2 \in \{2, \ldots, Q-1\}$, $q_2 = Q$, and $q_2 = 1$ separately.

Case $q_2 \in \{2, \ldots, Q-1\}$: In this case, we define

$$W := Y_1, \qquad \widetilde{Y}_1 := \mathbb{1}\{Y_1 > q_1\}, \qquad \widetilde{Y}_3 := \mathbb{1}\{Y_3 > q_3\}, \qquad \widetilde{Y}_2 := \begin{cases} 1 & \text{if } Y_2 < q_2, \\ 2 & \text{if } Y_2 = q_2, \\ 3 & \text{if } Y_2 > q_2. \end{cases}$$

Our ordered logit model in (4) then implies that the joint distribution of $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$ and W conditional on $A = \alpha$, $Y_0 = y_0$, $X = (x_1, x_2, x_3)$, and $\theta = \theta^0$ satisfies (20) and (21), as long as we choose

$$\begin{aligned}
f(y_1 \,|\, 1) &= \Pr\left(Y_1 = y_1 \,\middle|\, Y_1 > q_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right), \\
f(y_1 \,|\, 0) &= \Pr\left(Y_1 = y_1 \,\middle|\, Y_1 \le q_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right), \\
p_3(\widetilde{y}_3 \,|\, 1,\, y_1) &= \begin{cases} \Pr\left(Y_3 \le q_3 \,\middle|\, Y_2 < q_2,\, Y_1 = y_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 0, \\ \Pr\left(Y_3 > q_3 \,\middle|\, Y_2 < q_2,\, Y_1 = y_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 1, \end{cases} \\
p_3(\widetilde{y}_3 \,|\, 3,\, y_1) &= \begin{cases} \Pr\left(Y_3 \le q_3 \,\middle|\, Y_2 > q_2,\, Y_1 = y_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 0, \\ \Pr\left(Y_3 > q_3 \,\middle|\, Y_2 > q_2,\, Y_1 = y_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 1, \end{cases}
\end{aligned}$$

and

$$\pi_1 = \alpha + z(y_0, x_1, \theta^0) - \lambda_{q_1} = \alpha + x_1'\,\beta^0 + \gamma^0_{y_0} - \lambda_{q_1},$$

$$\pi_{2,1}(y_1) = \alpha + z(y_1, x_2, \theta^0) - \lambda_{q_2-1} = \alpha + x_2' \beta^0 + \gamma^0_{y_1} - \lambda_{q_2-1},$$
$$\pi_{2,2}(y_1) = \alpha + z(y_1, x_2, \theta^0) - \lambda_{q_2} = \alpha + x_2' \beta^0 + \gamma^0_{y_1} - \lambda_{q_2},$$
$$\pi_3 = \alpha + z(q_2, x_3, \theta^0) - \lambda_{q_3} = \alpha + x_3' \beta^0 + \gamma^0_{q_2} - \lambda_{q_3},$$

where $w = y_1$, and $z(y_{t-1}, x_t, \theta)$ is defined in (3). Plugging those expressions for π_1, $\pi_{2,1}(w)$, $\pi_{2,2}(w)$ and π_3 into the moment function $m(\widetilde{y}, w)$ in (22), we find that this moment function exactly coincides with $m_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ in equation (8) of the main text. Thus, by applying Lemma 1 to the distribution of Y conditional on $A = \alpha$, $Y_0 = y_0$, and $X = x$ (the lemma does not feature those conditioning variables, which is why we are applying the lemma to the conditional distribution), we obtain

$$\mathbb{E}\left[m_{y_0,q_1,q_2,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha\right] = 0,$$

which concludes the proof for the case $q_2 \in \{2, \ldots, Q-1\}$.

Case $q_2 = Q$: In this case, we choose

$$W := Y_1, \quad V := Y_2, \quad \widetilde{Y}_1 := \mathbb{1}\{Y_1 > q_1\}, \quad \widetilde{Y}_2 := \mathbb{1}\{Y_2 = Q\}, \quad \widetilde{Y}_3 := \mathbb{1}\{Y_3 > q_3\}.$$

Our ordered logit model in (4) then implies that the joint distribution of $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$, W, and V conditional on $A = \alpha$, $Y_0 = y_0$, $X = (x_1, x_2, x_3)$, and $\theta = \theta^0$ satisfies the assumptions of Lemma 2, as long as we choose

$$f(y_1 \,|\, 1) = \Pr\left(Y_1 = y_1 \,\middle|\, Y_1 > q_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right),$$
$$f(y_1 \,|\, 0) = \Pr\left(Y_1 = y_1 \,\middle|\, Y_1 \leq q_1,\, Y_0 = y_0,\, X = x,\, A = \alpha\right),$$
$$g(y_2 \,|\, 1) = \mathbb{1}\{y_2 = Q\},$$
$$g(y_2 \,|\, 0, y_1) = \Pr\left(Y_2 = y_2 \,\middle|\, Y_2 < Q,\, Y_1 = y_1,\, X = x,\, A = \alpha\right),$$

and

$$\pi_1 = \alpha + z(y_0, x_1, \theta^0) - \lambda_{q_1} = \alpha + x_1' \beta^0 + \gamma^0_{y_0} - \lambda_{q_1},$$

$$\pi_2(y_1) = \alpha + z(y_1, x_2, \theta^0) - \lambda_{Q-1} = \alpha + x_2' \beta^0 + \gamma^0_{y_1} - \lambda_{Q-1},$$
$$\pi_3(y_2) = \alpha + z(y_2, x_3, \theta^0) - \lambda_{q_3} = \alpha + x_3' \beta^0 + \gamma^0_{y_2} - \lambda_{q_3},$$

where $w = y_1$ and $v = y_2$, and $z(y_{t-1}, x_t, \theta)$ is defined in (3). Plugging those expressions for π_1, $\pi_2(y_1)$, and $\pi_3(y_2)$ into the moment function $m(\widetilde{y}, w, v)$ in Lemma 2, we find that this moment function exactly coincides with $m_{y_0,q_1,Q,q_3}(y, x, \theta^0)$ in equation (10) of the main text. Thus, by applying Lemma 2 to the distribution of Y conditional on $A = \alpha$, $Y_0 = y_0$, and $X = x$ (the lemma does not feature those conditioning variables, which is why we are applying the lemma to the conditional distribution), we obtain

$$\mathbb{E}\left[m_{y_0,q_1,Q,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\ X = x,\ A = \alpha\right] = 0,$$

which concludes the proof for the case $q_2 = Q$.

Case $q_2 = 1$: The result for this case follows from the result for $q_2 = Q$ by applying the transformation $Y_t \mapsto Q + 1 - Y_t$, $\lambda_q \mapsto -\lambda_{Q-q}$, $\beta \mapsto -\beta$, $\gamma_q \mapsto -\gamma_{Q+1-Y_t}$, $A_i \mapsto -A_i$. This transformation leaves the model probabilities in (5) unchanged but transforms the moment function in (10) into the one in (9), implying that this is also a valid moment function. ■

Proof of Theorem 2. As was the the case in the proof of Theorem 1, we consider the three cases $q_2 \in \{2, \ldots, Q-1\}$, $q_2 = Q$, and $q_2 = 1$ separately.

Case $q_2 \in \{2, \ldots, Q-1\}$: In this case, we define

$$W := Y_{s-1}, \qquad \widetilde{Y}_1 := \mathbb{1}\{Y_t > q_1\}, \qquad \widetilde{Y}_3 := \mathbb{1}\{Y_{s+1} > q_3\}, \qquad \widetilde{Y}_2 := \begin{cases} 1 & \text{if } Y_s < q_2, \\ 2 & \text{if } Y_s = q_2, \\ 3 & \text{if } Y_s > q_2. \end{cases}$$

Let $Y^{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_0)$. Our ordered logit model in (4) then implies that the joint distribution of $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$ and W conditional on $A = \alpha$, $Y^{t-1} = y^{t-1}$, $X = (x_1, x_2, x_3)$, and $\theta = \theta^0$ satisfies (20) and (21), as long as we choose

$$f(y_1 \,|\, 1) = \Pr\left(Y_t = y_1 \,\middle|\, Y_t > q_1,\ Y^{t-1} = y^{t-1},\ X = x,\ A = \alpha\right),$$

$$f(y_1 \,|\, 0) = \Pr\left(Y_t = y_1 \,\middle|\, Y_t \le q_1,\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right),$$

$$p_3(\widetilde{y}_3 \,|\, 1,\, y_{s-1}) = \begin{cases} \Pr\left(Y_{s+1} \le q_3 \,\middle|\, Y_s < q_2,\, Y_{s-1} = y_{s-1},\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 0, \\ \Pr\left(Y_{s+1} > q_3 \,\middle|\, Y_s < q_2,\, Y_{s-1} = y_{s-1},\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 1, \end{cases}$$

$$p_3(\widetilde{y}_3 \,|\, 3,\, y_{s-1}) = \begin{cases} \Pr\left(Y_{s+1} \le q_3 \,\middle|\, Y_s > q_2,\, Y_{s-1} = y_{s-1},\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 0, \\ \Pr\left(Y_{s+1} > q_3 \,\middle|\, Y_s > q_2,\, Y_{s-1} = y_{s-1},\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right) & \text{if } \widetilde{y}_3 = 1, \end{cases}$$

and

$$\begin{aligned} \pi_1 &= \alpha + z(y_{t-1}, x_t, \theta^0) - \lambda_{q_1} = \alpha + x_t'\,\beta^0 + \gamma^0_{y_{t-1}} - \lambda_{q_1}, \\ \pi_{2,1}(y_1) &= \alpha + z(y_{s-1}, x_s, \theta^0) - \lambda_{q_2-1} = \alpha + x_s'\,\beta^0 + \gamma^0_{y_{s-1}} - \lambda_{q_2-1}, \\ \pi_{2,2}(y_1) &= \alpha + z(y_{s-1}, x_s, \theta^0) - \lambda_{q_2} = \alpha + x_s'\,\beta^0 + \gamma^0_{y_{s-1}} - \lambda_{q_2}, \\ \pi_3 &= \alpha + z(q_s, x_{s+1}, \theta^0) - \lambda_{q_3} = \alpha + x_{s+1}'\,\beta^0 + \gamma^0_{q_2} - \lambda_{q_3}, \end{aligned}$$

where $w = y_{s-1}$, and $z(y_{t-1}, x_t, \theta)$ is defined in (3). Plugging those expressions for π_1, $\pi_{2,1}(w)$, $\pi_{2,2}(w)$ and π_3 into the moment function $m(\widetilde{y}, w)$ in (22) we find that this moment function exactly coincides with $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ in equation (11) of the main text. Thus, by applying Lemma 1 to the distribution of Y conditional on $A = \alpha$, $Y^{t-1} = y^{t-1}$, and $X = x$ (the lemma does not feature those conditioning variables, which is why we are applying the lemma to the conditional distribution), we obtain

$$\mathbb{E}\left[m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(Y, X, \theta^0) \,\middle|\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right] = 0.$$

Applying the law of iterated expectations, we thus also find that

$$\mathbb{E}\left[w(Y_1, \ldots, Y_{t-1})\, m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha\right] = 0,$$

which concludes the proof for the case $q_2 \in \{2, \ldots, Q-1\}$.

Case $q_2 = Q$: In this case, we choose

$$W := Y_{s-1}, \quad V := Y_{r-1}, \quad \widetilde{Y}_t := \mathbb{1}\{Y_1 > q_1\}, \quad \widetilde{Y}_s := \mathbb{1}\{Y_2 = Q\}, \quad \widetilde{Y}_r := \mathbb{1}\{Y_3 > q_3\}.$$

Our ordered logit model in (4) then implies that the joint distribution of $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$, W, and V conditional on $A = \alpha$, $Y^{t-1} = y^{t-1}$, $X = (x_1, x_2, x_3)$, and $\theta = \theta^0$ satisfies the assumptions of Lemma 2, as long as we choose

$$\begin{aligned} f(y_{s-1} \,|\, 1) &= \Pr\left(Y_{s-1} = y_{s-1} \,\middle|\, Y_t > q_1,\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right), \\ f(y_{s-1} \,|\, 0) &= \Pr\left(Y_{s-1} = y_{s-1} \,\middle|\, Y_t \leq q_1,\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right), \\ g(y_{r-1} \,|\, 1) &= \mathbb{1}\left\{y_{r-1} = Q\right\}, \\ g(y_{r-1} \,|\, 0, y_{s-1}) &= \Pr\left(Y_{r-1} = y_{r-1} \,\middle|\, Y_s < Q,\, Y_{s-1} = y_{s-1},\, X = x,\, A = \alpha\right), \end{aligned}$$

and

$$\begin{aligned} \pi_1 &= \alpha + z(y_{t-1}, x_t, \theta^0) - \lambda_{q_1} = \alpha + x_t'\,\beta^0 + \gamma^0_{y_{t-1}} - \lambda_{q_1}, \\ \pi_2(y_{s-1}) &= \alpha + z(y_{s-1}, x_s, \theta^0) - \lambda_{Q-1} = \alpha + x_s'\,\beta^0 + \gamma^0_{y_{s-1}} - \lambda_{Q-1}, \\ \pi_3(y_{r-1}) &= \alpha + z(y_{r-1}, x_r, \theta^0) - \lambda_{q_3} = \alpha + x_r'\,\beta^0 + \gamma^0_{y_{r-1}} - \lambda_{q_3}, \end{aligned}$$

where $w = y_{s-1}$ and $v = y_{r-1}$, and $z(y_{t-1}, x_t, \theta)$ is defined in (3). Plugging those expressions for π_1, $\pi_2(y_{s-1})$, and $\pi_3(y_{r-1})$ into the moment function $m(\widetilde{y}, w, v)$ in Lemma 2 we find that this moment function exactly coincides with $m_{y_0,q_1,Q,q_3}(y, x, \theta^0)$ in equation (12) of the main text. Thus, by applying Lemma 2 to the distribution of Y conditional on $A = \alpha$, $Y^{t-1} = y^{t-1}$, and $X = x$ (the lemma does not feature those conditioning variables, which is why we are applying the lemma to the conditional distribution) we obtain

$$\mathbb{E}\left[m^{(t,s,r)}_{y_0,q_1,Q,q_3}(Y, X, \theta^0) \,\middle|\, Y^{t-1} = y^{t-1},\, X = x,\, A = \alpha\right] = 0.$$

Applying the law of iterated expectations, we thus also find that

$$\mathbb{E}\left[w(Y_1, \ldots, Y_{t-1})\, m^{(t,s,r)}_{y_0,q_1,Q,q_3}(Y, X, \theta^0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha\right] = 0,$$

which concludes the proof for the case $q_2 = Q$.

Case $q_2 = 1$: The result for this case again follows from the result for $q_2 = Q$ by applying the transformation $Y_t \mapsto Q + 1 - Y_t$, $\lambda_q \mapsto -\lambda_{Q-q}$, $\beta \mapsto -\beta$, $\gamma_q \mapsto -\gamma_{Q+1-Y_t}$, $A_i \mapsto -A_i$. ■

A.2 Proof of Proposition 1, 2, and 3

The following lemma is useful for the proof of Proposition 1.

Lemma 3 *Let $Q \geq 2$. Let B be a $Q \times Q$ matrix for which all non-diagonal elements are positive (i.e. $B_{q,r} > 0$ for $q \neq r$). Let $g^0, g \in (0, \infty)^Q$ be two vectors with only positive entries. Assume that $Bg^0 = 0$ and $Bg = 0$. Then there exists $\kappa > 0$ such that $g = \kappa\, g^0$.*

Proof. This is a proof by contradiction. Let all assumptions of the lemma be satisfied, and assume that there does not exist a $\kappa > 0$ such that $g = \kappa\, g^0$. Define the vector $h \in [0, \infty)^Q$ and the two sets $\mathcal{Q}_+, \mathcal{Q}_0 \subset \{1, \ldots, Q\}$ by

$$h := g^0 - \left(\min_{q \in \{1,\ldots,Q\}} \frac{g_q^0}{g_q} \right) g, \qquad \mathcal{Q}_+ := \{q \,:\, h_q > 0\}, \qquad \mathcal{Q}_0 := \{q \,:\, h_q = 0\}.$$

All elements of h are non-negative by construction, and we have $h \neq 0$, because otherwise we would have $g = \kappa\, g^0$ for some $\kappa > 0$. Therefore, neither $\mathcal{Q}_+$ nor $\mathcal{Q}_0$ are empty sets. Furthermore, since h is a linear combination of g^0 and g, and we have $Bg^0 = Bg = 0$, we also have $Bh = 0$. This can equivalently be written as

$$\sum_{r \in \mathcal{Q}_+} B_{q,r}\, h_r = 0, \qquad \text{for all } q \in \{1, \ldots, Q\},$$

where we dropped the indices r from the sum for which we have $h_r = 0$.

Now, let $q \in \mathcal{Q}_0$. We then have $q \notin \mathcal{Q}_+$, and therefore $B_{q,r} > 0$ for all $r \in \mathcal{Q}_+$, according to our assumption on B. We have argued that $\mathcal{Q}_+$ is non-empty, and by construction we have $h_q > 0$ for $q \in \mathcal{Q}_+$. We therefore have

$$\sum_{r \in \mathcal{Q}_+} B_{q,r}\, h_r > 0.$$

The last two displays are the contradiction that we wanted to derive here. ■

Proof of Proposition 1. Let $x_{(1)} = (x_1, x_1, x_1)$. For $y_0, q \in \{1, \ldots, Q\}$ we define

$$B_{y_0,q} := \begin{cases} \Pr\left(y_1 > 1 \;\&\; y_2 = 1 \;\&\; y_3 = 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \text{if } y_0 \neq q \text{ and } q = 1, \\ \Pr\left(y_1 = q \;\&\; y_2 = 1 \;\&\; y_3 > 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \text{if } y_0 \neq q \text{ and } q > 1. \\ \Pr\left(y_1 > 1 \;\&\; y_2 = 1 \;\&\; y_3 = 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \\ \qquad\qquad -\Pr\left(y_1 = 1 \;\&\; y_2 > 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \text{if } y_0 = q \text{ and } q = 1, \\ \Pr\left(y_1 = q \;\&\; y_2 = 1 \;\&\; y_3 > 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \\ \qquad\qquad -\Pr\left(y_1 = 1 \;\&\; y_2 > 1 \mid Y_0 = y_0,\, X = x_{(1)}\right) & \text{if } y_0 = q \text{ and } q > 1. \end{cases}$$

Let B be the $Q \times Q$ matrix with entries $B_{y_0,q}$. Our assumptions guarantee that all the conditional the probabilities that enter into the definition of $B_{y_0,q}$ are non-negative, and we therefore have

$$B_{y_0,q} > 0, \qquad \text{for all } y_0 \neq q. \tag{25}$$

Applying Theorem 1 we find that the moment function in (15) satisfies

$$\mathbb{E}\left[m_{y_0}(y, \gamma^0) \mid Y_0 = y_0,\, X = (x_1, x_1, x_1)\right] = 0, \qquad \text{for all } y_0 \in \{1, \ldots, Q\}, \tag{26}$$

where γ^0 is the true parameter that generates the data. In the proposition we assume that $\gamma \in \mathbb{R}^Q$ is an alternative parameter that satisfies the same moment conditions. Let g^0 and g be the Q-vectors with entries $g^0_q := \exp\left(\gamma^0_q\right) > 0$ and $g_q := \exp\left(\gamma_q\right)$. Using the definition of the matrix B we can rewrite the two systems of Q equations in (26) and (16) as

$$B\, g^0 = 0, \qquad\qquad B\, g = 0. \tag{27}$$

Since we have (25) and (27) we can apply Lemma 3 to find that there exists $\kappa > 0$ such that $g = \kappa\, g^0$. Taking logarithms we thus have $\gamma = \gamma^0 + c$, where $c = \log(\kappa)$. This is what we wanted to show. ■

The following lemma is useful for the proof of Proposition 2.

Lemma 4 *Let $K \in \mathbb{N}$. For every $s = (s_1, \ldots, s_{K-1}, +) \in \{-, +\}^K$ let $g_s : \mathbb{R}^K \to \mathbb{R}$ be a continuous function such that for all $\beta \in \mathbb{R}^K$ we have*

(i) $g_s(\beta)$ is strictly increasing in β_K.

(ii) For all $m \in \{1, \ldots, K-1\}$: If $s_m = +$, then $g_s(\beta)$ is strictly increasing in β_m.

(iii) For all $m \in \{1, \ldots, K-1\}$: If $s_k = -$, then $g_s(\beta)$ is strictly decreasing in β_m.

Then, the system of 2^{K-1} equations in K variables

$$g_s(\beta) = 0 \qquad \text{for all } s \in \{-, +\}^K \text{ with } s_K = +,$$

has at most one solution.

Proof. This is the same as Lemma 2 in Honoré and Weidner (2020), only presented using slightly different notation here. ■

Proof of Proposition 2. For $s = (s_1, \ldots, s_{K-1}, +) \in \{-, +\}^K$ we define

$$g_s(\beta) = \mathbb{E}\left[m_{y_0,1,1,1}(y, x, \beta, \gamma_0) \,\middle|\, Y_0 = y_0,\ X \in \mathcal{X}_s \right], \tag{28}$$

where $m_{y_0,1,1,1}(y, x, \beta, \gamma)$ is the moment function in (17). Our assumptions guarantee that the conditioning sets in (28) have positive probability, which together with the definition of $m_{y_0,1,1,1}(y, x, \beta, \gamma_0)$ and $\mathcal{X}_s$ guarantee that the functions $g_s(\beta)$ satisfy the monotonicity requirements (i), (ii) and (iii) of Lemma 4. Theorem 1 guarantees that

$$g_s(\beta^0) = 0, \qquad \text{for all } s \in \{-, +\}^K \text{ with } s_K = +,$$

where β^0 is the true parameter value that generates the data. Equation (18) in the proposition can equivalently be written as

$$g_s(\beta) = 0, \qquad \text{for all } s \in \{-, +\}^K \text{ with } s_K = +.$$

According to Lemma 4 the system of equations in the last two displays can have at most one solution, and we therefore must have $\beta = \beta_0$. ■

Proof of Proposition 3. The definition of $m_{y_0,q_1,1,1}(y, x, \beta, \gamma, \lambda)$ in (19) together with the assumptions of the proposition guarantee that $g(\lambda) := \mathbb{E}\left[m_{y_0,q_1,1,1}(Y, X, \beta^0, \gamma^0, \lambda) \,\middle|\, Y_0 = y_0 \right]$ is strictly increasing in $\lambda_{q_1} - \lambda_1$ for all $q_1 \in \{2, \ldots, Q-1\}$. Theorem 1 guarantees that $g(\lambda^0) = 0$ for the true parameter λ^0 that generates the data. For any $\lambda \in \mathbb{R}^{Q-1}$ that satisfies $g(\lambda) = 0$ we therefore must have $\lambda_{q_1} - \lambda_1 = \lambda^0_{q_1} - \lambda^0_1$, which implies $\lambda = \lambda^0 + c$ for $c = \lambda_1 - \lambda^0_1$. ■

A.3 Computational details

From an estimation point of view, it is natural to estimate (β, γ, λ) by applying generalized methods of moments to a finite dimensional vector of unconditional moment conditions derived from (8), (9), and (10). There are at least two problems with this. The first is that, as discussed in Section 3 above, one needs to worry about whether the moment conditions actually identify the parameters of interest. The second problem is that even if one ignores the issue of identification, there are many ways to form a finite set of unconditional moment conditions from the expressions in (8), (9), and (10). Of course, it is in principle known how to most efficiently turn a set of conditional moment conditions into a set of moment condition of the same dimensionality as the parameter to be estimated. See, for example, the discussion in Newey and McFadden (1994). Specifically, with a conditional moment condition $\mathbb{E}\left[m\left(Y, X, \theta\right) \middle| X \right] = 0$ when θ takes it true value, θ_0, the optimal unconditional moment function is $A\left(X\right) m\left(Y, X, \theta\right)$, where $A\left(X\right) = \mathbb{E}\left[\nabla_\theta m\left(Y, X, \theta_0\right) \middle| X \right]' V\left[m\left(Y, X, \theta_0\right) \middle| X \right]$. Unfortunately, the construction of estimators of these efficient moments depends heavily on the distribution of Y given X. In the fixed effects context, this will depend on the distribution of the fixed effects conditional on all the explanatory variables. This prevents a simple two-step procedure for efficiently estimating (β, γ, λ) from the conditional moment conditions. To turn the conditional moment conditions in Theorem 1 into an estimator, we therefore use a slightly different approach.

We first estimate (β, γ, λ) by GMM using a fairly arbitrary set of moment conditions. In practice, we consider the moments constructed by interacting each of the conditional moments in Theorem 1 with each element of $(1, X_{i1} - X_{i2}, X_{i1} - X_{i3}, X_{i2} - X_{i3})$. We then first calculate the GMM estimator based on the following simple diagonal weight matrix. We set β and γ to vectors of 0's and estimate λ by maximum likelihood estimation under the assumption that there is no individual specific heterogeneity. We then calculate the empirical variance of each moment and use the inverse of those as the diagonal elements of the weight matrix (except when the sample variance is 0, in which case the weight is set to 0). This yields an initial GMM estimator of (β, γ, λ). We use this to calculate the empirical variance of each moment. The inverse of those are then used as diagonal elements of a diagonal weight matrix, which is in turn used to form a second GMM estimator. Finally, we use this estimator of (β, γ, λ) to calculate the sample covariance matrix of the moments. The inverse of this is used as the weight matrix for a final GMM estimator, except that we inflate the diagonal elements of the covariance matrix by 10% before inverting it. This is done to overcome numerical problems associated with collinear moments.

We next estimate a flexible reduced form model for the distribution of Y given X. One may think of this as a nonparametric sieve estimator in which case our approach will be an attempt to construct the efficient estimator. Alternatively, one can acknowledge that the model for Y given X is most likely incorrect, in which case the approach can be interpreted as an attempt to construct unconditional moments that are close to the efficient ones. In practice, we estimate the reduced form model for the distribution of Y given X by a period-by-period ordered logit model where the explanatory variables are all observed lagged Y, the contemporaneous X, and the average X over all time periods.

We then construct an estimate of $A(X)$ and use that to calculate the method moments estimator.