Dynamic Ordered Panel Logit Models*

Bo E. Honoré[‡] Chris Muris[§] Martin Weidner[¶]

April 3, 2025

Abstract

This paper studies a dynamic ordered logit model for panel data with fixed effects. The main contribution of the paper is to construct a set of valid moment conditions that are free of the fixed effects. The moment functions can be computed using four or more periods of data, and the paper presents sufficient conditions for the moment conditions to identify the common parameters of the model, namely the regression coefficients, the autoregressive parameters, and the threshold parameters. The availability of moment conditions suggests that these common parameters can be estimated using the generalized method of moments, and the paper documents 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. The data used in this paper come from the British Household Panel Survey (BHPS, University of Essex, Institute for Social and Economic Research (2021)), which was accessed via the UK Data Service. The BHPS data was originally collected by the Institute for Social and Economic Research (ISER) and funded by the Economic and Social Research Council (ESRC). 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 Riccardo D'Adamo, Geert Dhaene, Sharada Dharmasankar and Runtong Ding for useful comments and discussion. We also thank Limor Golan and four anonymous referees for constructive feedback and suggestions.

[‡]Princeton University, 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 is available.

For other types of outcome 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.

The challenge of accommodating unobserved heterogeneity in nonlinear models is well understood, especially when the researcher also wants to allow for lagged dependent variables. For example, while recent developments (Kitazawa 2021 and Honoré and Weidner 2020) relax these requirements, early work on the dynamic binary logit model with fixed effects either assumed no regressors, or restricted their joint distribution (cf. Chamberlain 1985 and Honoré and Kyriazidou 2000). The challenge of accommodating unobserved heterogeneity in the ordered logit model seems even greater than in the binary model. The reason is that

[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.

even the static version of the model is not in the exponential family (Hahn 1997). As a result, one cannot directly appeal to a sufficient statistic approach. An alternative approach in the static ordered logit model is to reduce it to a set of binary choice models (cf. Das and van Soest 1999, Johnson 2004b, Baetschmann, Staub, and Winkelmann 2015, Muris 2017, and Botosaru, Muris, and Pendakur 2023). Unfortunately, the dynamic ordered logit model cannot be similarly reduced to a dynamic binary choice model (see Muris, Raposo, and Vandoros 2023). Therefore, a new approach is needed. The contribution of this paper is to develop such an approach. To do this, 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 (2023) 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 (2023), 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 places no restrictions on the dependence between fixed effects and regressors, requires only 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 (2004b), 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 (2022) 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 both a Monte Carlo study and an empirical application to self-reported health status in the British Household Panel Study.

The remainder of this paper is organized as follows. Section 2 introduces the model and the moment conditions that are free of fixed effects. Section 3 presents identification results for the common model parameters based on those moment conditions. Section 4 discusses how to use the moment conditions for estimation and inference. Section 5 explores the practical performance of the resulting estimation method through Monte Carlo simulations, including comparison with a correlated random effects approach. That section also provides an empirical illustration to health data. Section 6 concludes the paper. The appendix provides proofs and further computational details.

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 parameters 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 $\lambda = (\lambda_1, \ldots, \lambda_{Q-1}) \in \mathbb{R}^{Q-1}$ is a vector of 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 dependent variable $Y_{i,t-1}$ is allowed to 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.[2]

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

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

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

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

[2] If the observed Y_{it} is a discretized version of a continuous variable with a natural economic interpretation, then it would be more natural to model the state dependence in (2) as $Y_{i,t-1}^{\star}\gamma$. A numerical investigation suggests that it is not possible to develop moment conditions for such a model. This suggests that the common parameters in this model are not point-identified, or are only point-identified under strong support assumptions on the covariates and hence not generally $\sqrt{n}$ estimable. We explore this alternative model in Appendix A.4.

Below, 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

In the next subsection, we present moment functions for the ordered logit model discussed above. These are 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 $\alpha \in \mathbb{R} \cup \{\pm\infty\}$. 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 $A = \alpha$, the distribution of $Y = (Y_1, \ldots, Y_T) \in \{1, \ldots, Q\}^T$ is given by (5). The model assumptions in the last subsection are therefore sufficient to evaluate the conditional expectation in (6).

If we can establish the conditional moment condition (6) then, 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). Such an 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. Notice, however, that if one is interested in (functions of) the individual-specific effects such as 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 this 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 (with 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 important for finding solutions of this infinite-dimensional linear system in a finite number of variables, and for most choices of error distributions (e.g. normally distributed error), we do not expect such solutions to exist. It seems likely that for the error distributions in Johnson (2004a) and Davezies, D'Haultfoeuille, and Mugnier (2022), and also for a mixture of logistics (briefly discussed in Honoré and Weidner 2020), one could also find valid moment conditions, if a sufficient number of time periods are available, but we focus purely on logistic errors in this paper.

In the following, we present moment functions $m_{Y_0}(Y, X, \theta^0)$ that satisfy (6). We derive 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 expressions for the moment functions, their derivation is no longer relevant and we can focus on showing that they are valid – i.e. that (6) holds – and on their implications for the identification and estimation of θ.

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 $\lambda_{qr} := \lambda_q - \lambda_r$.

We find multiple moment functions $m_{y_0,q_1,q_2,q_3}(y, x, \theta)$ which are distinguished by the additional indices $q_1 \in \{1, \ldots, Q-1\}$, $q_2 \in \{1, \ldots, Q\}$, $q_3 \in \{1, \ldots, Q-1\}$. For the moment function labelled by q_1, q_2, q_3, the dependence on y is only through the coarser outcome $\widetilde{y}_{q_1,q_2,q_3}(y) \in \{0, 1\} \times \{1, 2, 3\} \times \{0, 1\}$, which is a vector with three components $\widetilde{y}_{t,q_1,q_2,q_3}(y)$, $t \in \{1, 2, 3\}$, given by

$$\widetilde{y}_{2,q_1,q_2,q_3}(y) := \begin{cases} 1 & \text{if } y_2 < q_2, \\ 2 & \text{if } y_2 = q_2, \\ 3 & \text{if } y_2 > q_2. \end{cases} \qquad \widetilde{y}_{t,q_1,q_2,q_3}(y) := \mathbb{1}\{y_r > q_r\}, \quad \text{for } t \in \{1, 3\}.$$

The moment functions presented below have the property that

$$m_{y_0,q_1,q_2,q_3}(y, x, \theta) = m_{y_0,q_1,q_2,q_3}(\widetilde{y}_{q_1,q_2,q_3}(y), x, \theta). \tag{8}$$

Thus, for given q_1, q_2, q_3, it would be sufficient to observe the outcome $\widetilde{Y} = \widetilde{y}_{q_1,q_2,q_3}(Y)$ to implement this moment function. Notice that $\widetilde{y}_{t,q_1,q_2,q_3}(y)$, for time periods $t = 1$ and $t = 3$ is just a binarization, as in Muris (2017) and Muris, Raposo, and Vandoros (2023), but for $t = 2$ we crucially deviate from those existing papers, because for $q_2 \in \{2, \ldots, Q-1\}$ the coarser outcome $\widetilde{y}_{2,q_1,q_2,q_3}(y)$ is a trinarization of the second period outcome, not a binarization. It turns out that this is a crucial extension to obtain all the valid moment conditions in our model.

The moment functions presented below were obtained using Mathematica by following the methods described in Section 2 of Honoré and Weidner (2020), which builds on ideas in Bonhomme (2012). Once derived, we can prove by hand that the moment functions are valid (proof of Theorem 1 in the appendix), but we do not have any useful explanation or intuition for the detailed functional form of these moment functions. However, the binarized/trinarized outcome $\widetilde{Y}$ and equation (8) help to appreciate some aspects of the structure of the moment functions (see also Lemma 1 in the appendix). Furthermore, while the functional form of the moment functions is mysterious, one can show the existence of valid moment functions much more easily, see Appendix A.3.

For $q_1, q_3 \in \{1, \ldots, Q-1\}$ and $q_2 \in \{2, \ldots, Q-1\}$, we define

$$m_{y_0,q_1,q_2,q_3}(y, x, \theta)$$
$$:= \begin{cases} \exp\left(x_{13}'\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) \dfrac{\exp(x_{32}'\beta + \gamma_{q_2,y_1} + \lambda_{q_2,q_3}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_1 \leq q_1,\ y_2 = q_2, y_3 \leq q_3, \\ \exp\left(x_{13}'\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) \dfrac{1 - \exp(x_{23}'\beta + \gamma_{y_1,q_2} + \lambda_{q_3,q_2})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_1 \leq q_1,\ y_2 = q_2,\ y_3 > q_3, \\ \exp\left(x_{13}'\beta + \gamma_{y_0,q_2} + \lambda_{q_3,q_1}\right) & \text{if } y_1 \leq q_1,\ y_2 > q_2, \\ -1 & \text{if } y_1 > q_1,\ y_2 < q_2, \\ -\dfrac{1 - \exp(x_{32}'\beta + \gamma_{q_2,y_1} + \lambda_{q_2-1,q_3})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_1 > q_1,\ y_2 = q_2,\ y_3 \leq q_3, \\ -\dfrac{\exp(x_{23}'\beta + \gamma_{y_1,q_2} + \lambda_{q_3,q_2-1}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_1 > q_1,\ y_2 = q_2,\ y_3 > q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Any valid moment function satisfying (6) can be multiplied by an arbitrary constant and remain a valid moment function. In (9) 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 $(y_1 \leq q_1, y_2 > q_2)$ 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 (9).

Equation (9) 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 (9), 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

$$m_{y_0,q_1,1,q_3}(y,x,\theta) := \begin{cases} \exp\left(x_{23}'\,\beta + \gamma_{y_1,1} + \lambda_{q_3,1}\right) - 1 & \text{if } y_1 \leq q_1,\ y_2 = 1,\ y_3 > q_3, \\ -1 & \text{if } y_1 \leq q_1,\ y_2 > 1, \\ \exp\left(x_{31}'\,\beta + \gamma_{1,y_0} + \lambda_{q_1,q_3}\right) & \text{if } y_1 > q_1,\ y_2 = 1,\ y_3 \leq q_3, \\ \exp\left(x_{21}'\,\beta + \gamma_{y_1,y_0} + \lambda_{q_1,1}\right) & \text{if } y_1 > q_1,\ y_2 = 1,\ y_3 > q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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

$$m_{y_0,q_1,Q,q_3}(y,x,\theta) := \begin{cases} \exp\left(x_{12}'\,\beta + \gamma_{y_0,y_1} + \lambda_{Q-1,q_1}\right) & \text{if } y_1 \leq q_1,\ y_2 = Q,\ y_3 \leq q_3, \\ \exp\left(x_{13}'\,\beta + \gamma_{y_0,Q} + \lambda_{q_3,q_1}\right) & \text{if } y_1 \leq q_1,\ y_2 = Q,\ y_3 > q_3, \\ -1 & \text{if } y_1 > q_1,\ y_2 < Q, \\ \exp\left(x_{32}'\,\beta + \gamma_{Q,y_1} + \lambda_{Q-1,q_3}\right) - 1 & \text{if } y_1 > q_1,\ y_2 = Q,\ y_3 \leq q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Notice that the moment functions for $q_2 = 1$ and $q_2 = Q$ also satisfy (8), but here $\widetilde{y}_{q_1,q_2,q_3}(y)$ corresponds to a binarization of the outcome in all time periods, because $\widetilde{y}_{t,q_1,q_2,q_3}(y)$ also only takes two values for those values of q_2. Those moment functions are therefore conceptually much closer to Muris (2017) and Muris, Raposo, and Vandoros (2023), and they also incorporate the moment functions for the dynamic binary choice model in Honoré and Weidner (2020) as special cases.

Together, the formulas (9), (10), and (11) 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 $T = 3$.[3] 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*

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

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.[4]

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 the completeness of those binary choice moment conditions is discussed in Kruiniger

[3] 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 (9) alone.

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

(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 than 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 first consider moment conditions that are based on the outcome in 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 $y_0 \in \{1, \ldots, Q\}$, $q_1, q_3 \in \{1, \ldots, Q-1\}$, $q_2 \in \{2, \ldots, Q-1\}$, and $t, s \in \{1, 2, \ldots, T-1\}$ with $t < s$ we define

$$m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta) := \begin{cases} \exp\left(z_{t,s+1} + \lambda_{q_3,q_1}\right) \dfrac{\exp(z_{s+1,s} + \lambda_{q_2,q_3}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_t \le q_1,\ y_s = q_2, y_{s+1} \le q_3, \\ \exp\left(z_{t,s+1} + \lambda_{q_3,q_1}\right) \dfrac{1 - \exp(z_{s,s+1} + \lambda_{q_3,q_2})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_t \le q_1,\ y_s = q_2,\ y_{s+1} > q_3, \\ \exp\left(z_{t,s+1} + \gamma_{y_s,q_2} + \lambda_{q_3,q_1}\right) & \text{if } y_t \le q_1,\ y_s > q_2, \\ -1 & \text{if } y_t > q_1,\ y_s < q_2, \\ -\dfrac{1 - \exp(z_{s+1,s} + \lambda_{q_2-1,q_3})}{1 - \exp\left(\lambda_{q_2-1,q_2}\right)} & \text{if } y_t > q_1,\ y_s = q_2,\ y_{s+1} \le q_3, \\ -\dfrac{\exp(z_{s,s+1} + \lambda_{q_3,q_2-1}) - 1}{\exp\left(\lambda_{q_2,q_2-1}\right) - 1} & \text{if } y_t > q_1,\ y_s = q_2,\ y_{s+1} > q_3, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

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 (12) equals the moment function in equation (9). For larger values of T, the moment function in (12) 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 (12) 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

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

$$m^{(t,s,r)}_{y_0,q_1,Q,q_3}(y, x, \theta) := \begin{cases} \exp\left(z_{ts} + \lambda_{Q-1,q_1}\right) & \text{if } y_t \leq q_1,\ y_s = Q,\ y_r \leq q_3, \\ \exp\left(z_{tr} + \lambda_{q_3,q_1}\right) & \text{if } y_t \leq q_1,\ y_s = Q,\ y_r > q_3, \\ -1 & \text{if } y_t > q_1,\ y_s < Q, \\ \exp\left(z_{rs} + \lambda_{Q-1,q_3}\right) - 1 & \text{if } y_t > q_1,\ y_s = Q,\ y_r \leq q_3, \\ 0 & \text{otherwise}. \end{cases} \tag{13}$$

When T equals 3 and $(t, s, r) = (1, 2, 3)$, these moment functions agree with the ones in equations (10) and (11), where all the arguments were made explicit. For $r = s+1$, analogous to (10) and (11) for $T = 3$, the two moment conditions in (13) for $T \geq 3$ can be obtained from (12) 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 (12) as our master equation, which summarizes all the moment conditions provided in this paper. In (13) we can choose more general $r \geq s+1$, but otherwise the structure of (13) can be derived from (12).

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.[5] 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

[5]For example, an estimator that allows for selection to be correlated with (Y_{i0}, X_i, A_i) can be constructed using the results on GMM estimation with incomplete data in Muris (2020).

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 (13).

The following theorem establishes that the moment functions in (12) and (13) 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*

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

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 (13) 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 (12) 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 ($Y_s < q_2$, $Y_s = q_2$, $Y_s > q_2$). The proof, in this case, 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 (12) and (13) 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 let the true parameters $\theta^0 = (\beta^0, \gamma^0, \lambda^0)$ be 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

$$m_{y_0}(y, x, \theta^0) = \sum_{q_1=1}^{Q-1} \sum_{q_2=1}^{Q} \sum_{q_3=1}^{Q-1} \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} w_{y_0}(q_1, q_2, q_3, t, s, y_1, \ldots, y_{t-1}, x, \theta^0)\, m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0), \tag{14}$$

where $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ are the moment functions defined in (12) and (13). In other words, we conjecture that 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.

In equation (14), the function $m^{(t,s,s+1)}_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ is multiplied with an arbitrary function of $y_1, \ldots, y_{t-1}$. Those functions of $y_1, \ldots, y_{t-1}$ constitute a Q^{t-1} dimensional space. Thus, (14) suggests that the total number of available moment conditions for each value of the covariates x and initial conditions y_0 is equal to

$$\ell = \sum_{q_1=1}^{Q-1} \sum_{q_2=1}^{Q} \sum_{q_3=1}^{Q-1} \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} Q^{t-1} = (Q-1)\, Q\, (Q-1) \sum_{t=1}^{T-2} (T-t-1)\, Q^{t-1}$$
$$= Q^T - (T-1)\, Q^2 + (T-2)\, Q. \tag{15}$$

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 (14). For fixed values of y_0, x, θ_0, T, Q, one can numerically verify the dimension of the solution space of the system of linear equations (6), and thereby check (15) numerically. In Appendix A.3 we furthermore show that the total number of linearly independent conditional moment conditions for our model is at least the number obtained in (A.3), but that argument in the appendix still allows for the possibility that there could be more, although we do not believe that there are.

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 the 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 (12) and (13) only depend on the regressors and the parameters β and γ through the single index $z_t = z(y_{t-1}, x_t, \theta)$.[6] So far, we have only explicitly discussed the linear specification in (3) for this single index, but Theorem 2 is valid independently of the functional form of $z(y_{t-1}, x_t, \theta)$.[7] 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 (9), (10), (11), 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:

$$z(Y_{i,t-1}, X_{it}, \theta) := X_{it}'\,\beta + \sum_{q=1}^{Q} \gamma_q \left[\mathbb{1}\left\{Y_{i,t-1} = q\right\} + \mathbb{1}\left\{Y_{i,t-1} = q\right\} X_{it}'\,\delta_q\right], \tag{16}$$

where $\delta_q \in \mathbb{R}^K$ are the 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

[6] As written, the moment condition in (12) 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 artefact, because in that line of the moment condition we 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.

[7] 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

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

The moment functions in (12) and (13) 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 (12)) 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.

paper could be employed in applications with the more general state dependence as in (16), we do not consider these more general models further.

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* $\mathcal{X}^{\text{reg}}_{y_0} \subset \mathbb{R}^{K \times 3}$ *such that for all* $x \in \mathcal{X}^{\text{reg}}_{y_0}$*, 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 in order to ensure 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 $Y_t = Q$ would be 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 $X \in \mathcal{X}^{\text{reg}}_{y_0}$ 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 (10) reads

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

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*

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

Then, if $\gamma \in \mathbb{R}^Q$ satisfies[8]

$$\mathbb{E}\left[m_{y_0}(Y, \gamma) \,\middle|\, Y_0 = y_0,\ X = (x_1, x_1, x_1)\right] = 0 \qquad \textit{for all } y_0 \in \{1, \ldots, Q\}, \tag{18}$$

for $m_{y_0}(y, \gamma)$ as defined in (17), 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 $y_0 \in \{1, \ldots, Q\}$. 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,

[8] Here, we implicitly assume that $\mathbb{E}\left[m_{y_0}(Y, \gamma) \,\middle|\, Y_0 = y_0,\ X = (x_1, x_1, x_1)\right]$ is uniquely defined. This can be guaranteed, for example, by demanding that this conditional expectation is continuous in x_1.

guarantees that the conditional expectation in (18) 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 (10) with $q_1 = q_2 = q_3 = 1$, but now for general regressor values

$$m_{y_0,1,1,1}(y, x, \beta, \gamma) := m_{y_0,1,1,1}(y, x, \theta) = \begin{cases} \exp\left(x_{23}' \beta\right) - 1 & \text{if } y_1 = 1,\ y_2 = 1,\ y_3 > 1, \\ -1 & \text{if } y_1 = 1,\ y_2 > 1, \\ \exp\left(x_{31}' \beta + \gamma_1 - \gamma_{y_0}\right) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 = 1, \\ \exp\left(x_{21}' \beta + \gamma_{y_1} - \gamma_{y_0}\right) & \text{if } y_1 > 1,\ y_2 = 1,\ y_3 > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

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 $x \in \mathcal{X}_{k,+}$, then $m_{y_0,1,1,1}(y, x, \beta, \gamma)$ is strictly increasing in β_k, and if $x \in \mathcal{X}_{k,-}$, 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 $x \in \mathcal{X}_s$, 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 $s_k = -$). These 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*

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

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

$$\mathbb{E}\left[m_{y_0,1,1,1}(Y, X, \beta, \gamma^0) \,\middle|\, Y_0 = y_0,\ X \in \mathcal{X}_s \right] = 0 \qquad \text{for all } s \in \{-,+\}^K \text{ with } s_K = +, \tag{20}$$

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 (20) is well-defined.

Identification of λ

Having identified γ and β, we now turn to the problem of identifying λ up to a normalization. The moment function in (10) with $q_2 = q_3 = 1$ and $q_1 \in \{2, \ldots, Q-1\}$ can be written as

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

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 to all $q_1 \in \{2, \ldots, Q-1\}$, we can therefore identify λ up to an additive constant. This is summarized in the following proposition.

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

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

we have $\lambda = \lambda^0 + c$ for some $c \in \mathbb{R}$. Thus, if we normalize $\lambda^0_1 = 0$, 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. Hence, in our Monte Carlo simulations and empirical application, we 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. It is tempting to try to mimic the identification argument in Section 3 in order to turn these moment conditions into an estimator. The problem with such an approach is that the conditioning set in Proposition 1 will often have probability 0. Alternatively, one can form a set of unconditional moment functions 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 (9), (10), and (11), and g is a vector-valued function of the initial condition Y_{i0} and the strictly exogenous X_i. Let $\theta = (\beta', \gamma', \lambda')'$. A generalized method of moments (GMM) estimator can then be defined by[9]

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

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.

One limitation of the GMM approach is that it is often difficult to know whether the moment condition $\mathbb{E}\left[M(Y_{i0}, Y_i, X_i, \theta)\right] = 0$ is uniquely satisfied at the true parameter value. When the strictly exogenous explanatory variables, X_i, are discrete, sufficient conditions for this can be obtained from the identification results in Section 3 by defining $g(Y_{i0}, X_i)$ 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 calculate valid confidence sets for θ_0 based on moment conditions like $\mathbb{E}\left[M(Y_{i0}, Y_i, X_i, \theta)\right] = 0$. See, for example, Chen, Christensen, and Tamer (2018).

A second limitation of the GMM approach 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 (9), (10), and (11). Moreover, the most natural ad hoc ways to

[9] 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.

do this, such as considering all interactions between the conditional moment and the explanatory variables and the initial condition, can lead to a very large number of moment conditions, which in turn can result in poor small sample performance. It is in principle known how to most efficiently turn a set of conditional moment conditions into a set of moment conditions 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 $\mathbb{E}\left[\left. m_{Y_0}(Y,X,\theta)\right| X,Y_0\right] = 0$ when θ takes its true value, θ_0, the optimal unconditional moment function is $A(X,Y_0)\, m_{Y_0}(Y,X,\theta)$, where $A(X,Y_0) = \mathbb{E}\left[\left.\nabla_\theta m_{Y_0}(Y,X,\theta_0)\right| X,Y_0\right]' V\left[\left. m_{Y_0}(Y,X,\theta_0)\right| X,Y_0\right]^{-1}$. Unfortunately, the construction of estimators of these efficient moments depends heavily on the distribution of Y given (X,Y_0). On the other hand, the moment conditions are still valid if $A(X,Y_0)$ is misspecified. One approach therefore is to estimate a flexible reduced form model for the distribution of Y given (X,Y_0), and then use this reduced form for the distribution of Y given X to construct an estimate of $A(X,Y_0)$. In the simulations and the empirical illustration below, we take this approach using a correlated random effects approach to obtain the reduced form for the distribution of Y given (X,Y_0). The dimensionality of the moment function $A(X,Y_0)\, m_{Y_0}(Y,X,\theta)$ is the same as that of the parameter vector, and the asymptotic distribution of the estimator therefore follows from the theory of nonlinear method of moments estimators.

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, $\widehat{\theta}$, 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, 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 a method of moments estimator

In the next subsection, we first present the results from a small Monte Carlo experiment designed to illustrate the performance of the method of moments estimator based on the discussion in Section 4. We then compare this estimator's performance with that of a correlated random effects estimator. Subsequently, we illustrate the application of the method of moments estimator through an empirical example.

In Appendix A.5 we investigate the potential usefulness of the moment conditions further. Specifically, we compare the asymptotic distribution of various GMM estimators based on the moment conditions to the asymptotic distribution of a correlated random effects estimator. The advantage of this exercise is that it allows one to abstract from the particular implementation of the GMM estimators.

5.1 Monte Carlo illustration

We illustrate the performance of the GMM estimator described above through a Monte Carlo study that considers three data generating processes, two with a fixed effect and one without a fixed effect. The two data generating processes that include a fixed effect are chosen such that one satisfies the assumptions underpinning the correlated random effects estimator proposed by Wooldridge (2005), while the other does not. We consider sample sizes of $N = 500$, 1000, and 2000 with five time periods for each individual. This includes the initial observations, so $T = 4$ using the notation above. There are $k = 3$ explanatory variables and the dependent variable can take $Q = 4$ values. The true parameters are $\beta = (1, 0, 0)'$, $\gamma = (-1, 0, 0, 1)'$ and $\lambda = (-2, 0, 2)'$ and we normalize $\gamma_2 = \lambda_2 = 0$.

The explanatory variables are drawn as follows. First, let $\tilde{A}_i$ be a discrete random variable with $E\left[\tilde{A}_i\right] = 0$ and $V\left[\tilde{A}_i\right] = 3$. The exact distribution of $\tilde{A}_i$ differs across specifications. Secondly, let Z_{ijt} ($j = 1, ..., k,\ t = 0, ..., 4$) be independent normal random variables with mean 0 and variance 3, and let the first explanatory variable be $X_{i1t} = \left(Z_{i1t} + \tilde{A}_i\right)/\sqrt{2}$. The second through k'th explanatory variables are given by $X_{ijt} = (Z_{ijt} + X_{i1t})/\sqrt{2}$. This implies that all the explanatory variables and $X_{it}'\beta$ have mean 0 and variance 3. This is comparable to the magnitude of the logistic distribution, which has mean 0 and variance $\pi^2/3$.

For the two data generating processes with a fixed effect, one (Design B) has $\tilde{A}_i$ normally distributed while the other (Design C) has $P\left(\tilde{A}_i = \sqrt{6}\right) = \frac{1}{3}$ and $P\left(\tilde{A}_i = -\sqrt{6}/2\right) = \frac{2}{3}$. For both of these specifications, the fixed effect, A_i, equals $\tilde{A}_i$. The data generating process without fixed effects (Design A) has the same distribution of $\tilde{A}_i$ as Design C, but here $A_i = 0$. For Design C, the initial dependent variable, Y_{i0} is generated from the ordered logit model where the only explanatory variable is A_i. For Designs A and B, Y_{i0} is generated from the ordered logit model without explanatory variables. From a fixed effects perspective, this makes Design B a little special, but it makes it fit the assumptions for the correlated random effects approach. Design A is without fixed effects, so it also satisfies the assumptions for the correlated random effects approach, but in this case the true parameter value of one of the parameters (the variance of the error in the specification of the fixed effect) is at the boundary of the parameter space, which could render standard inference problematic.

We perform 400 Monte Carlo replications. The results are presented in Tables 1, 2 and 3. For comparison, we also include the results for the correlated random effects estimator (cf. Wooldridge 2005) that specifies the distribution of the unobserved heterogeneity as $A = \sum_{t=0}^{4} X_t'\theta(t) + \sum_{q=1}^{4} 1\{Y_0 = q\}\theta(q) + \sigma Z$, where $Z \sim N(0,1)$. Design C violates the implicit assumption behind the correlated random effects approach. Most importantly, the distribution of A_i is discrete. On the other hand, the relationship between the explanatory variables and A_i is linear, so the violation is not extreme.

For each design and for each sample size, Tables 1, 2 and 3 report the true values of the parameters, the median bias of the method of moments estimator and the correlated random effects estimator, the interquartile range of the estimators, and the median absolute errors of the estimators. For each parameter, we also report the ratio of the median absolute error of the correlated random effects estimator relative to the method of moment estimator. Values of this ratio greater than one suggest that the method of moments estimator is more precise than the correlated random effects estimator. In Designs A and B, the correlated random effects estimator is the correctly specified maximum likelihood estimator. It is therefore not surprising that the median absolute error ratio is less than one for all parameters and all sample sized in this case. On the other hand, the ratio is above 0.65 in all cases, suggesting that the loss of efficiency from the method of moments estimator is not too large for this design. For these designs, both estimators appear to be close to median unbiased, and the relative performance of the estimators is driven by the difference in their variability. For Design C with non-normal heterogeneity, the relative performance of the two estimators is different for the different parameters. The correlated random effects estimator is always less variable in terms of interquartile range, but the biases in the estimates of the γ's and δ's are large enough that the method of moments estimator tends to be more precise when the sample size is large. For the coefficients on the explanatory variables, β, the correlated random effects estimator is almost unbiased in Design C despite the misspecification of the model for the fixed effect. This makes sense, because the specified model for the unobserved heterogeneity will tend to control for any linear dependence between explanatory variables and the level of the fixed effect. The specific results for each of the γ's and for each of the λ's should be considered with some care since the calculations are done under the specific normalization that $\gamma_2 = 0$ and $\lambda_2 = 0$. With different normalizations, the pattern of the results would have been different. However, it is clear from Table 3 that in the design with heterogeneity, the misspecification embedded in the correlated random effects approach generally speaking leads to biased estimates of the γ's and the λ, and that these biases will make the method of moments estimator more precise for large sample sizes.

Tables 4-6 illustrate that the smaller bias of the method of moments estimator can make it more reliable for inference than the correlated random effects estimator when the assumptions underpinning the correlated random effects estimator are violated. Specifically, the table presents the fraction of times that 80, 90 and 95 percent confidence intervals based on the correlated random effects estimator and on the method of moments estimator cover the true unknown parameter. As one would expect from the results in Table 3, the bias in the correlated random effects estimator combined with its low variability can make it unlikely that a confidence interval covers the true parameter value. Tables 4 and 5 show the same

Table 1: Design A

Correlated Random Effects. Sample Size: 500 (400 replications).

	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
Bias	0.008	0.004	0.001	0.014	0.000	-0.036	-0.027	-0.027	0.000	0.024
IQR	0.096	0.057	0.055	0.181	0.000	0.174	0.201	0.097	0.000	0.108
MAE	0.046	0.028	0.028	0.097	0.000	0.092	0.104	0.051	0.000	0.060

Method of Moments. Sample Size: 500 (400 replications).

	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
Bias	-0.001	0.004	-0.002	0.019	0.000	-0.018	-0.010	-0.003	0.000	0.009
IQR	0.112	0.068	0.066	0.249	0.000	0.232	0.252	0.135	0.000	0.144
MAE	0.056	0.034	0.033	0.124	0.000	0.117	0.123	0.067	0.000	0.072
MAE ratio	0.832	0.833	0.830	0.780	—	0.787	0.845	0.760	—	0.835

Correlated Random Effects. Sample Size: 1000 (400 replications).

	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
Bias	0.006	0.000	0.002	0.010	0.000	-0.014	-0.023	-0.013	0.000	0.013
IQR	0.066	0.038	0.039	0.118	0.000	0.112	0.139	0.082	0.000	0.082
MAE	0.032	0.019	0.020	0.057	0.000	0.060	0.072	0.038	0.000	0.041

Method of Moments. Sample Size: 1000 (400 replications).

	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
Bias	-0.001	-0.001	0.002	0.005	0.000	-0.009	-0.005	0.008	0.000	0.005
IQR	0.076	0.046	0.049	0.155	0.000	0.159	0.195	0.107	0.000	0.107
MAE	0.037	0.023	0.023	0.077	0.000	0.081	0.097	0.055	0.000	0.053
MAE ratio	0.876	0.814	0.859	0.737	—	0.736	0.741	0.701	—	0.775

Correlated Random Effects. Sample Size: 2000 (400 replications).

	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
Bias	0.002	-0.000	-0.001	0.013	0.000	-0.010	-0.011	-0.010	0.000	0.012
IQR	0.046	0.032	0.025	0.100	0.000	0.084	0.098	0.049	0.000	0.059
MAE	0.022	0.016	0.013	0.052	0.000	0.041	0.051	0.025	0.000	0.032

Method of Moments. Sample Size: 2000 (400 replications).

	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
Bias	-0.001	0.001	-0.002	0.005	0.000	-0.001	0.008	-0.004	0.000	0.001
IQR	0.046	0.033	0.030	0.123	0.000	0.101	0.129	0.067	0.000	0.072
MAE	0.023	0.016	0.015	0.061	0.000	0.050	0.066	0.033	0.000	0.036
MAE ratio	0.955	0.979	0.842	0.839	—	0.810	0.775	0.768	—	0.898

MAE is the median absolute error, IQR is the interquartile range. γ_2 and λ_2 are normalized.

Table 2: Design B

Correlated Random Effects. Sample Size: 500 (400 replications).

	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
Bias	0.003	-0.003	-0.003	-0.007	0.000	-0.010	0.007	-0.007	0.000	0.009
IQR	0.102	0.057	0.064	0.224	0.000	0.203	0.262	0.144	0.000	0.131
MAE	0.051	0.028	0.032	0.112	0.000	0.100	0.130	0.073	0.000	0.062

Method of Moments. Sample Size: 500 (400 replications).

	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
Bias	-0.001	-0.003	-0.004	0.010	0.000	-0.007	-0.006	-0.000	0.000	0.002
IQR	0.118	0.078	0.084	0.323	0.000	0.300	0.323	0.187	0.000	0.172
MAE	0.059	0.039	0.043	0.163	0.000	0.154	0.162	0.096	0.000	0.087
MAE ratio	0.867	0.718	0.742	0.688	—	0.650	0.802	0.754	—	0.715

Correlated Random Effects. Sample Size: 1000 (400 replications).

	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
Bias	-0.001	-0.000	-0.000	-0.008	0.000	0.006	0.006	-0.005	0.000	0.000
IQR	0.073	0.047	0.048	0.173	0.000	0.141	0.183	0.106	0.000	0.098
MAE	0.037	0.023	0.023	0.086	0.000	0.070	0.089	0.054	0.000	0.049

Method of Moments. Sample Size: 1000 (400 replications).

	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
Bias	0.001	-0.002	0.001	0.004	0.000	-0.007	0.006	-0.002	0.000	0.004
IQR	0.086	0.060	0.055	0.205	0.000	0.196	0.231	0.126	0.000	0.128
MAE	0.043	0.030	0.028	0.104	0.000	0.098	0.119	0.062	0.000	0.063
MAE ratio	0.867	0.782	0.852	0.829	—	0.720	0.751	0.883	—	0.775

Correlated Random Effects. Sample Size: 2000 (400 replications).

	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
Bias	0.004	-0.000	0.000	0.010	0.000	0.011	0.005	-0.004	0.000	-0.001
IQR	0.049	0.034	0.032	0.121	0.000	0.096	0.119	0.066	0.000	0.065
MAE	0.024	0.017	0.016	0.059	0.000	0.049	0.062	0.034	0.000	0.033

Method of Moments. Sample Size: 2000 (400 replications).

	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
Bias	0.003	-0.002	0.001	0.016	0.000	0.008	0.004	-0.002	0.000	0.001
IQR	0.058	0.040	0.042	0.145	0.000	0.122	0.156	0.084	0.000	0.077
MAE	0.029	0.020	0.021	0.073	0.000	0.061	0.078	0.042	0.000	0.039
MAE ratio	0.853	0.847	0.753	0.807	—	0.799	0.793	0.802	—	0.854

MAE is the median absolute error, IQR is the interquartile range. γ_2 and λ_2 are normalized.

Table 3: Design C

Correlated Random Effects. Sample Size: 500 (400 replications).

	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
Bias	0.006	0.007	0.003	0.168	0.000	0.127	0.674	-0.049	0.000	-0.308
IQR	0.108	0.073	0.070	0.216	0.000	0.260	0.365	0.156	0.000	0.173
MAE	0.054	0.037	0.037	0.170	0.000	0.156	0.674	0.081	0.000	0.308

Method of Moments. Sample Size: 500 (400 replications).

	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
Bias	-0.018	0.005	0.009	0.074	0.000	0.007	0.303	0.026	0.000	-0.146
IQR	0.185	0.112	0.106	0.327	0.000	0.479	0.702	0.271	0.000	0.376
MAE	0.097	0.058	0.054	0.173	0.000	0.232	0.445	0.139	0.000	0.227
MAE ratio	0.556	0.638	0.677	0.983	—	0.671	1.515	0.586	—	1.355

Correlated Random Effects. Sample Size: 1000 (400 replications).

	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
Bias	0.004	-0.004	-0.000	0.189	0.000	0.137	0.691	-0.045	0.000	-0.308
IQR	0.081	0.052	0.055	0.151	0.000	0.192	0.272	0.098	0.000	0.110
MAE	0.040	0.026	0.028	0.189	0.000	0.156	0.691	0.058	0.000	0.308

Method of Moments. Sample Size: 1000 (400 replications).

	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
Bias	-0.013	-0.006	-0.000	0.017	0.000	0.035	0.202	0.025	0.000	-0.075
IQR	0.144	0.084	0.092	0.290	0.000	0.457	0.638	0.176	0.000	0.322
MAE	0.070	0.044	0.045	0.148	0.000	0.227	0.354	0.098	0.000	0.169
MAE ratio	0.574	0.594	0.626	1.277	—	0.687	1.950	0.597	—	1.822

Correlated Random Effects. Sample Size: 2000 (400 replications).

	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
Bias	0.007	-0.003	-0.002	0.179	0.000	0.133	0.712	-0.043	0.000	-0.315
IQR	0.052	0.037	0.034	0.099	0.000	0.125	0.171	0.080	0.000	0.079
MAE	0.025	0.019	0.017	0.179	0.000	0.133	0.712	0.051	0.000	0.315

Method of Moments. Sample Size: 2000 (400 replications).

	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
Bias	-0.003	-0.003	-0.008	0.006	0.000	0.026	0.090	0.024	0.000	-0.028
IQR	0.107	0.066	0.058	0.200	0.000	0.336	0.559	0.143	0.000	0.226
MAE	0.052	0.033	0.030	0.101	0.000	0.163	0.276	0.075	0.000	0.122
MAE ratio	0.483	0.561	0.560	1.781	—	0.819	2.577	0.686	—	2.589

MAE is the median absolute error, IQR is the interquartile range. γ_2 and λ_2 are normalized.

results for Design A and Design B. In these case, the confidence intervals based on the correlated random effects estimator do very well. For Design B, this is not surprising as this is a correctly specified maximum likelihood setting. It is interesting that the correlated random effects estimator also does well in Design A. Although it is the maximum likelihood estimator of a correctly specified model, the estimation is made non-standard by the fact that the true value of one of the parameters (the variance of the error in the specification for the random effect) is on the boundary of the parameter space and asymptotic normality would therefore not follow from textbook asymptotic theory. As a general statement, Tables 4, 5 and 6 also illustrate that the confidence interval based on the method of moments estimator can be somewhat erratic even with relatively large sample sizes.[10]

5.2 Empirical illustration

In this section, we illustrate the value of the moment conditions derived in this paper in an empirical illustration inspired by Contoyannis, Jones, and Rice (2004). The dependent variable is self-reported health status. We use data from the first five waves of the British Household Panel Survey[11], and we restrict the sample to individuals who are between 26 and 70 years old in the first wave. This yields a data set with 5093 individuals observed in 5 time periods, including the initial observation (so $T = 4$). In the original data set, the dependent variable can take five values. We aggregate these into "Poor or Very Poor" (8.1% of the observations), "Fair" (18.6%), "Good" (47.6%), 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 $(Age - 45)^2/100$, respectively, where Age is measured in years). In the second, we also include log-income.

The results are presented in Table 7, which also presents the estimates from a correlated

[10] On the other hand, in simulations not reported here, we have found that the natural estimator of the variance of the method of moments estimator can perform poorly. The results reported here therefore use bootstrap standard errors based on the interquartile range of 1000 bootstrap replications. The standard errors reported for the correlated random effects are based on the "robust" expression for the asymptotic variance of extremum estimators.

[11] University of Essex, Institute for Social and Economic Research (2021).

Table 4: Design A

Correlated Random Effects. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.797	0.812	0.807	0.790	—	0.770	0.780	0.807	—	0.787
90% CI	0.912	0.897	0.902	0.892	—	0.892	0.875	0.912	—	0.900
95% CI	0.948	0.953	0.950	0.938	—	0.940	0.938	0.960	—	0.968

Method of Moments. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.870	0.850	0.850	0.902	—	0.870	0.875	0.930	—	0.953
90% CI	0.935	0.932	0.950	0.950	—	0.945	0.930	0.973	—	0.988
95% CI	0.965	0.975	0.983	0.970	—	0.978	0.968	0.995	—	0.995

Correlated Random Effects. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.785	0.815	0.800	0.818	—	0.775	0.795	0.772	—	0.782
90% CI	0.912	0.915	0.887	0.895	—	0.883	0.875	0.873	—	0.900
95% CI	0.953	0.950	0.940	0.938	—	0.935	0.932	0.935	—	0.950

Method of Moments. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.890	0.897	0.883	0.900	—	0.885	0.905	0.910	—	0.940
90% CI	0.958	0.953	0.955	0.960	—	0.960	0.963	0.988	—	0.978
95% CI	0.980	0.973	0.978	0.978	—	0.975	0.990	0.998	—	0.988

Correlated Random Effects. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.765	0.772	0.850	0.777	—	0.815	0.805	0.782	—	0.807
90% CI	0.863	0.892	0.917	0.890	—	0.927	0.897	0.883	—	0.895
95% CI	0.927	0.940	0.945	0.940	—	0.963	0.927	0.932	—	0.953

Method of Moments. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.863	0.865	0.892	0.892	—	0.920	0.885	0.905	—	0.890
90% CI	0.915	0.935	0.958	0.950	—	0.973	0.955	0.965	—	0.958
95% CI	0.958	0.968	0.985	0.980	—	0.990	0.985	0.975	—	0.988

Based on 1000 bootstrap replications. γ_2 and λ_2 are normalized.

Table 5: Design B

Correlated Random Effects. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.828	0.843	0.815	0.777	—	0.792	0.795	0.785	—	0.820
90% CI	0.900	0.925	0.915	0.900	—	0.892	0.895	0.880	—	0.925
95% CI	0.943	0.955	0.955	0.945	—	0.938	0.945	0.935	—	0.968

Method of Moments. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.887	0.880	0.860	0.875	—	0.870	0.873	0.883	—	0.943
90% CI	0.958	0.945	0.943	0.935	—	0.938	0.943	0.940	—	0.983
95% CI	0.985	0.965	0.983	0.965	—	0.965	0.970	0.973	—	0.995

Correlated Random Effects. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.785	0.780	0.800	0.792	—	0.795	0.782	0.772	—	0.795
90% CI	0.907	0.897	0.895	0.885	—	0.905	0.897	0.902	—	0.885
95% CI	0.953	0.948	0.955	0.943	—	0.955	0.935	0.965	—	0.943

Method of Moments. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.892	0.890	0.895	0.900	—	0.875	0.897	0.943	—	0.920
90% CI	0.965	0.950	0.955	0.948	—	0.968	0.968	0.973	—	0.960
95% CI	0.985	0.960	0.983	0.975	—	0.983	0.985	0.990	—	0.985

Correlated Random Effects. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.800	0.810	0.802	0.805	—	0.820	0.785	0.787	—	0.802
90% CI	0.880	0.897	0.905	0.920	—	0.922	0.887	0.887	—	0.907
95% CI	0.945	0.948	0.950	0.973	—	0.960	0.935	0.932	—	0.943

Method of Moments. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.860	0.902	0.880	0.895	—	0.883	0.883	0.910	—	0.932
90% CI	0.948	0.965	0.960	0.950	—	0.953	0.948	0.963	—	0.983
95% CI	0.968	0.988	0.978	0.978	—	0.983	0.985	0.988	—	0.995

Based on 1000 bootstrap replications. γ_2 and λ_2 are normalized.

Table 6: Design C

Correlated Random Effects. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.780	0.825	0.787	0.593	—	0.728	0.103	0.742	—	0.113
90% CI	0.892	0.922	0.897	0.695	—	0.835	0.200	0.850	—	0.218
95% CI	0.960	0.963	0.953	0.787	—	0.900	0.328	0.940	—	0.295

Method of Moments. Sample Size: 500.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.792	0.820	0.792	0.797	—	0.762	0.682	0.828	—	0.730
90% CI	0.915	0.935	0.915	0.895	—	0.867	0.792	0.922	—	0.830
95% CI	0.955	0.958	0.958	0.940	—	0.915	0.883	0.963	—	0.877

Correlated Random Effects. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.792	0.802	0.775	0.345	—	0.560	0.005	0.720	—	0.007
90% CI	0.910	0.895	0.887	0.478	—	0.718	0.015	0.863	—	0.030
95% CI	0.963	0.963	0.943	0.590	—	0.823	0.045	0.925	—	0.060

Method of Moments. Sample Size: 1000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.845	0.815	0.792	0.810	—	0.785	0.720	0.833	—	0.758
90% CI	0.915	0.920	0.877	0.907	—	0.873	0.833	0.917	—	0.833
95% CI	0.950	0.950	0.932	0.950	—	0.900	0.900	0.963	—	0.917

Correlated Random Effects. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.780	0.802	0.795	0.135	—	0.480	0.000	0.645	—	0.000
90% CI	0.883	0.885	0.900	0.230	—	0.637	0.000	0.765	—	0.000
95% CI	0.940	0.948	0.945	0.353	—	0.735	0.000	0.875	—	0.002

Method of Moments. Sample Size: 2000.

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3
80% CI	0.805	0.807	0.843	0.843	—	0.815	0.745	0.860	—	0.762
90% CI	0.915	0.927	0.945	0.930	—	0.920	0.830	0.932	—	0.885
95% CI	0.965	0.973	0.965	0.958	—	0.960	0.877	0.960	—	0.922

Based on 1000 bootstrap replications. γ_2 and λ_2 are normalized.

random effects specification. We have normalized the γ-coefficient associated with "Good Health" and the threshold (λ) just below "Good Health" to be zero.

The most consistent result presented in Table 7 is that the coefficient on age is negative across all specifications and that the coefficient on age-squared is insignificantly different from 0 in all specifications. The point estimates for the effect of income on self-reported health are positive and statistically significant for all the specifications. The most puzzling aspect of Table 7 is that the standard error of the correlated random effects estimator of the coefficient on age gets much more precise when log-income is included. The method of moments estimator does not display this pattern. A comparison of the other standard errors reveals that the correlated random effects estimator is less variable than the method of moments estimator. This is in line with the interquartile ranges reported in Tables 1, 2 and 3.

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 (2023). 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 for 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 on panel data estimation of non-linear models with fixed effects. In this context, the main contribution is to illustrate the

Table 7: Empirical Results: Application to Self-Reported Health Status

	Three Outcomes				Four Outcomes			
	MoM	CRE	MoM	CREs	MoM	CRE	MoM	CRE
$Age/10$	−1.070	−1.110	−1.271	−0.179	−1.136	−0.961	−1.078	−0.147
	(0.272)	(0.145)	(0.410)	(0.035)	(0.208)	(0.137)	(0.335)	(0.032)
$(Age-45)^2/100$	0.396	−0.251	0.804	0.109	0.278	−0.226	−0.293	0.025
	(0.920)	(0.548)	(2.031)	(0.544)	(0.740)	(0.503)	(1.131)	(0.499)
log-income			0.054	0.283			0.130	0.217
			(0.196)	(0.048)			(0.088)	(0.045)
γ_1					−1.477	−1.594	−1.427	−1.674
					(0.174)	(0.109)	(0.239)	(0.111)
γ_2	0.269	−0.810	−0.753	−0.824	−0.759	−0.618	−0.691	−0.657
	(0.135)	(0.067)	(0.335)	(0.068)	(0.099)	(0.063)	(0.122)	(0.064)
γ_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
γ_4	0.147	0.484	0.559	0.590	0.395	0.557	0.635	0.619
	(0.119)	(0.063)	(0.502)	(0.064)	(0.102)	(0.064)	(0.244)	(0.065)
λ_1					−2.464	−2.560	−2.510	−2.555
					(0.056)	(0.050)	(0.080)	(0.050)
λ_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
λ_3	3.707	3.818	3.304	3.780	3.374	3.758	3.340	3.728
	(0.076)	(0.054)	(0.173)	(0.054)	(0.057)	(0.053)	(0.082)	(0.054)

The dependent variable is self-reported health status and the data is a balanced panel from the first five waves of the British Household Panel Survey. We restrict the sample to individuals who are between 26 and 70 years old in the first wave. In columns 1-4, the dependent variable takes the values "Poor, Very Poor, or Fair", "Good", and "Excellent". The dependent variable in columns 5-8 takes values "Poor or Very Poor", "Fair", "Good" and "Excellent". MoM refers to the method of moments estimator and CRE is the correlated random effects estimator.

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," unpublished, pp. 1–31.

Aguirregabiria, V., J. Gu, and Y. Luo (2021): "Sufficient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models," Journal of Econometrics, 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 (2023): "Identification of Time-Varying Transformation Models with Fixed Effects, with an Application to Unobserved Heterogeneity in Resource Shares," Journal of Econometrics, 232(2), 576–597.

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.

Chen, X., T. Christensen, and E. Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," Econometrica, 86(6), 1965–2018.

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.

Davezies, L., X. D'Haultfoeuille, and M. Mugnier (2022): "Fixed Effects Binary Choice Models with Three or More Periods," Quantitative Economics (forthcoming).

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.

Galama, T., and A. Kapteyn (2011): "Grossman's missing health threshold," Journal of Health Economics, 30, 1044–1056.

Grossman, M. (1972): "On the concept of health capital and the demand for health," Journal of Political Economy, 80, 223–255.

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. (2004a): "Identification in discrete choice models with fixed effects," in Working paper, Bureau of Labor Statistics. Citeseer.

JOHNSON, E. G. (2004b): "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.

——— (2020): "Efficient GMM Estimation with Incomplete Data," The Review of Economics and Statistics, 102(3), 518–530.

MURIS, C., P. RAPOSO, AND S. VANDOROS (2023): "A dynamic ordered logit model with fixed effects," Review of Economics and Statistics, forthcoming.

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 (2022): "Unobserved Heterogeneity, State Dependence, and Health Plan Choices," Working paper. Revised Sept. 2022.

Rigobon, R. (2002): "The curse of non-investment grade countries," Journal of Development Economics, 69, 423–449.

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

University of Essex, Institute for Social and Economic Research (2021): "British Household Panel Survey: Waves 1-18, 1991-2009," [data collection]. 8th Edition. UK Data Service. SN: 5151, DOI: http://doi.org/10.5255/UKDA-SN-5151-2.

Wooldridge, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," Journal of Applied Econometrics, 20(1), 39–54.

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

$$p(\widetilde{y}, w) = p_3(\widetilde{y}_3 \,|\, \widetilde{y}_2, w) \ \ p_2(\widetilde{y}_2 \,|\, w) \ \ f(w \,|\, \widetilde{y}_1) \ \ p_1(\widetilde{y}_1), \tag{22}$$

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:

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

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

$$p_3(\widetilde{y}_3 \,|\, 2,\, w) = \begin{cases} 1 - \Lambda(\pi_3) & \text{if } \widetilde{y}_3 = 0, \\ \Lambda(\pi_3) & \text{if } \widetilde{y}_3 = 1, \end{cases} \tag{23}$$

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

$$m(\widetilde{y}, w) := \begin{cases} \exp(\pi_1 - \pi_3) \dfrac{\exp[\pi_3 - \pi_{2,2}(w)] - 1}{\exp[\pi_{2,1}(w) - \pi_{2,2}(w)] - 1} & \text{if } \widetilde{y}_1 = 0,\ \widetilde{y}_2 = 2,\ \widetilde{y}_3 = 0, \\ \exp(\pi_1 - \pi_3) \dfrac{1 - \exp[\pi_{2,2}(w) - \pi_3]}{1 - \exp[\pi_{2,2}(w) - \pi_{2,1}(w)]} & \text{if } \widetilde{y}_1 = 0,\ \widetilde{y}_2 = 2,\ \widetilde{y}_3 = 1, \\ \exp(\pi_1 - \pi_3) & \text{if } \widetilde{y}_1 = 0,\ \widetilde{y}_2 = 3, \\ -1 & \text{if } \widetilde{y}_1 = 1,\ \widetilde{y}_2 = 1, \\ -\dfrac{1 - \exp[\pi_3 - \pi_{2,1}(w)]}{1 - \exp[\pi_{2,2}(w) - \pi_{2,1}(w)]} & \text{if } \widetilde{y}_1 = 1,\ \widetilde{y}_2 = 2,\ \widetilde{y}_3 = 0, \\ -\dfrac{\exp[\pi_{2,1}(w) - \pi_3] - 1}{\exp[\pi_{2,1}(w) - \pi_{2,2}(w)] - 1} & \text{if } \widetilde{y}_1 = 1,\ \widetilde{y}_2 = 2,\ \widetilde{y}_3 = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

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 (22) and (23), and let $m : \{0, 1\} \times \{1, 2, 3\} \times \{0, 1\} \times \{1, 2, \ldots, Q\} \to \mathbb{R}$ be defined by (24). 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 (23) and (24) one finds that for $\widetilde{y}_1 = 1$ we have

$$
\begin{aligned}
g(1, w) = &- \{1 - \Lambda[\pi_{2,1}(w)]\} - \{\Lambda[\pi_{2,1}(w)] - \Lambda[\pi_{2,2}(w)]\} \times \\
&\times \underbrace{\left(\frac{[1 - \Lambda(\pi_3)]\,\{1 - \exp[\pi_3 - \pi_{2,1}(w)]\}}{1 - \exp\left[\pi_{2,2}(w) - \pi_{2,1}(w)\right]} + \frac{\Lambda(\pi_3)\,\{\exp[\pi_{2,1}(w) - \pi_3] - 1\}}{\exp\left[\pi_{2,1}(w) - \pi_{2,2}(w)\right] - 1} \right)}_{= \frac{\Lambda[\pi_{2,1}(w)] - \Lambda(\pi_3)}{\Lambda[\pi_{2,1}(w)] - \Lambda[\pi_{2,2}(w)]}} \\
= &- \left[1 - \Lambda(\pi_3)\right], \qquad (25)
\end{aligned}
$$

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

$$
g(0, w) = \exp\left(\pi_1 - \pi_3\right) \Lambda(\pi_3). \qquad (26)
$$

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 (25), (26), and the expression for $p_1(\widetilde{y}_1)$ in (23) 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*

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

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

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

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 (22) and (23), as long as we choose

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

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 (24), we find that this moment function exactly coincides with $m_{y_0,q_1,q_2,q_3}(y, x, \theta^0)$ in equation (9) 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 (11) 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 (11) into the one in (10), implying that this is also a valid moment function. ■

Proof of Theorem 2. As was 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

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

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 (22) and (23), 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 (24) 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 (12) 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 (13) 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

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

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 $y_0, q \in \{1, \dots, Q\}$ 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 probabilities that enter into the definition of $B_{y_0,q}$ are non-negative, and we therefore have

$$B_{y_0,q} > 0, \qquad \text{for all } y_0 \neq q. \tag{27}$$

Applying Theorem 1 we find that the moment function in (17) satisfies

$$\mathbb{E}\left[m_{y_0}(y, \gamma^0) \mid Y_0 = y_0,\, X = (x_1, x_1, x_1)\right] = 0, \qquad \text{for all } y_0 \in \{1, \dots, Q\}, \tag{28}$$

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 $g_q^0 := \exp\left(\gamma_q^0\right) > 0$ 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 (28) and (18) as

$$B\, g^0 = 0, \qquad\qquad B\, g = 0. \tag{29}$$

Since we have (27) and (29) 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

$$g_s(\beta) = \mathbb{E}\left[m_{y_0,1,1,1}(Y, X, \beta, \gamma_0) \,\middle|\, Y_0 = y_0, \; X \in \mathcal{X}_s \right], \tag{30}$$

where $m_{y_0,1,1,1}(y, x, \beta, \gamma)$ is the moment function in (19). Our assumptions guarantee that the conditioning sets in (30) 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 (20) 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 (21) 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 Lower bound on the number of moment conditions

Dobronyi, Gu, and Kim (2021) point out that it is sometimes possible to easily derive a lower bound on the number of linearly independent valid moment conditions in a given panel data model. The key insight (to be verified below) is that one can write the model probabilities defined in equation (5) as

$$p_{y_0}(y, x, \theta, \alpha) = \kappa\,(a) \sum_{k=1}^{K} a^{k-1} c_k\,(y)\,, \tag{31}$$

for some constant $K \in \{1, 2, \ldots\}$, some positive function κ of $a = \exp{(\alpha)}$ that does not depend on y, and some functions c_k of y that do not depend on a. Here, the functions κ and c_k also depend on y_0, x, θ, but those arguments are dropped to focus more clearly on the dependence on α and y. In other words, y_0, x, θ are simply assumed fixed here. The dependence of all functions and constants on T is also not made explicit (here or anywhere else in the paper).

Once we have shown (31), then a a valid moment function must satisfy

$$\sum_{y \in \mathcal{Y}} m(y) \sum_{k=1}^{K} a^{k-1} c_k\,(y) = 0, \qquad \text{for all } a \in (0, \infty)\,, \tag{32}$$

which is equivalent to

$$\sum_{y \in \mathcal{Y}} m(y)\, c_k(y) = 0, \qquad \text{for all } k \in \{1, \ldots, K\}.$$

These are K linear conditions in Q^T unknown parameters $m(y)$. We, therefore, have at least $Q^T - K$ linearly independent solutions $m(y)$. In other words, the model must have at least $Q^T - K$ linearly independent conditional moment conditions.

What is left to do now is to show that (31) indeed holds for our the dynamic order choice model, for some $K < Q^T$. For that purpose, remember that $\Lambda(\varepsilon) = [1 + \exp(-\varepsilon)]^{-1}$, and also define $\omega_{t,q} = X_t' \beta - \lambda_q$, and

$$a = \exp(\alpha), \qquad w_q = \exp(\omega_{t,q}), \qquad g_q = \exp(\gamma_q).$$

Also remember that $\lambda_0 = -\infty$ and $\lambda_Q = \infty$, which implies that

$$w_0 = \infty, \qquad w_Q = 0.$$

According to (5), the probability of observing $Y = y$, conditional on $Y_0 = y_0$, $X = x$, $A = \alpha$, $\theta = \theta_0$, is then given by

$$\begin{aligned} p_{y_0}(y, x, \theta, \alpha) &= \prod_{t=1}^{T} \left\{ \Lambda\Big[\omega_{t,(y_t-1)} + \gamma_{y_{t-1}} + \alpha\Big] - \Lambda\Big[\omega_{t,y_t} + \gamma_{y_{t-1}} + \alpha\Big] \right\} \\ &= \prod_{t=1}^{T} \left\{ \frac{w_{t,(y_t-1)}\, g_{y_{t-1}}\, a}{1 + w_{t,(y_t-1)}\, g_{y_{t-1}}\, a} - \frac{w_{t,y_t}\, g_{y_{t-1}}\, a}{1 + w_{t,y_t}\, g_{y_{t-1}}\, a} \right\} \\ &= \kappa\,(a)\, \widetilde{p}(y, a), \end{aligned}$$

where

$$\kappa\,(a) = \prod_{q=1}^{Q-1} \frac{1}{1 + w_{t,q}\, g_{y_0}\, a} \prod_{t=2}^{T} \prod_{r=1}^{Q} \frac{1}{1 + w_{t,q}\, g_r\, a},$$

and

$$\widetilde{p}(y,a) = \Bigg\{ \mathbb{1}\{y_1 \neq 1\}\, w_{1,(y_1-1)}\, g_{y_0}\, a \prod_{q \in \{1,\ldots,Q-1\} \setminus \{y_1 - 1\}} (1 + w_{1,q}\, g_{y_0}\, a)$$

$$- \mathbb{1}\{y_1 = Q\} - \mathbb{1}\{y_1 \neq Q\}\, w_{1,y_1}\, g_{y_0}\, a \prod_{q \in \{1,\ldots,Q-1\} \setminus \{y_1\}} (1 + w_{1,q}\, g_{y_0}\, a) \Bigg\}$$

$$\times \prod_{t=2}^{T} \Bigg\{ \mathbb{1}\{y_t \neq 1\}\, w_{t,(y_t-1)}\, g_{y_{t-1}}\, a \prod_{(q,r) \in \{1,\ldots,Q-1\} \times \{1,\ldots,Q\} \setminus \{(y_t - 1, y_{t-1}\}} (1 + w_{t,q}\, g_r\, a)$$

$$- \mathbb{1}\{y_t = Q\} - \mathbb{1}\{y_t \neq Q\}\, w_{t,y_t}\, g_{y_{t-1}}\, a \prod_{(q,r) \in \{1,\ldots,Q-1\} \times \{1,\ldots,Q\} \setminus \{(y_t, y_{t-1}\}} (1 + w_{t,q}\, g_{y_r}\, a) \Bigg\}.$$

Here, $\kappa(a)$ does not depend on y, and $\widetilde{p}(y,a)$ is a polynomial in a of order

$$K - 1 = (Q-1) + (T-1)Q(Q-1).$$

This implies that a lower bound on the number of linearly independent valid moment conditions in this model is given by

$$\begin{aligned} Q^T - K &= Q^T - Q - (T-1)Q(Q-1) \\ &= Q^T - (T-1)Q^2 + (T-2)Q, \end{aligned}$$

which is exactly the number of linearly independent valid moment conditions we found in the main text.

A.4 An alternative specification

The model analyzed in this paper specifies that the latent variable evolves according to

$$Y_{it}^* = X_{it}'\,\beta + \sum_{q=1}^{Q} \gamma_q\, \mathbb{1}\{Y_{i,t-1} = q\} + A_i + \varepsilon_{it},$$

see (2). An alternative specification would have the dynamics in the lagged latent variable, for example

$$Y_{it}^{*} = X_{it}' \beta + \gamma Y_{i,t-1}^{*} + A_i + \varepsilon_{it}, \tag{33}$$

In many cases, one is interested in the dynamics in the actual discrete outcome as in our specification (2). For example, the outcome of a soccer game is clearly ordered (win, draw, defeat) and there may be strategic or physiological effects of past results on actual outcomes. In such a case, the dynamics are likely to be through the actual outcome rather than some underlying index.

A second example is the empirical application in Muris, Raposo, and Vandoros (2023) on self-reported health status: "While there are multiple levels of self-reported health, being above or below a threshold at which medical care is demanded in period $t-1$ is essential in determining health status in period t. Galama and Kapteyn (2011) formalized this important issue by building on the seminal work by Grossman (1972)." A third example comes from the analysis of government bonds (see also the discussion in Muris, Raposo, and Vandoros (2023). Government bonds are ordered ratings (20 tiers), which can be grouped into two categories: Investment Grade (IG) and Non-Investment Grade (NIG). IG Governments can easily raise credit, whereas NIG Governments cannot. This has implication for next period's credit rating for these governments. Thus, the actual lagged value of the ordered variable – not the latent index – affects a country's current credit rating, cf. Rigobon (2002). Additional examples may include the analysis of demand for ordered goods. For example, goods could be available in discrete quantities (8 oz, 12 oz, or 16 oz). In that case, it is more reasonable to model today's choice in terms of the actual discrete choice last period. Goods can also be ordered in terms of quality. In that case, one might model this period's choice (low, medium, high quality variety of a product) in terms of the actual choice in the previous periods. In conclusion, there are many settings of applied interest where interest is in the relationship between the current and lagged actual value of the discrete variable.

However, in some applications, (33) may be a more natural specification. With that in mind, we performed a numerical experiment to explore what happens when the techniques in our paper are applied to data from a DGP with dynamics that are linear in the lagged latent variable $Y^*_{i,t-1}$.

Specifically, we consider the case of binary choice, which would be the worst case in terms of approximation error. Specifically, we consider the dynamic binary choice model with fixed effects, with $Q = 2$ and $T = 3$, with $(X_1, X_2, X_3, Y_0) \in \{0, 1\}^4$ with equal probabilities, and set $\beta_{1,0} = 1$. We let the autoregressive parameter range over

$$\gamma_0 \in \{-0.9, -0.75, \cdots, 0.6, 0.75, 0.9\},$$

and let the fixed effects takes on the values $A \in \{-2, -1, 0, 1, 2\}$.

We generate data from a DGP1 that is consistent with our model (2), and from a DGP2 with outcome equation (33). For DGP2, we draw $Y^*_0 \sim \mathcal{N}(0, 1)$.

We construct optimal instruments under DGP1 evaluated at the true value of the parameters,

$$Z^*(y_0, x_1, x_2, x_3; \beta_{10}, \gamma_0) = D'(y_0, x_1, x_2, x_3, \beta_{10}, \gamma_0)\Omega(y_0, x_1, x_2, x_3; \beta_{10}, \gamma_0)$$

and form the unconditional moments

$$h(b, c) = E\left[Z^*(Y_0, X_1, X_2, X_3; \beta_{10}, \gamma_0)m(Y_0, X_1, X_2, X_3; b, c)\right].$$

To explore the sensitivity of our results to DGP2, we approximate $E(h(b, c))$ using 500000 draws for each value of the conditioning variables. We determine the values of (b, c) that set the moments equal to zero under both DGP1 and DGP2.

Figure 1 reports the results for the regression coefficient. Each panel corresponds to a value of A. The autoregressive parameter γ_0 varies along the horizontal axis. Results for DGP1 are covered by the horizontal line at $\beta_{10} = 1$, as expected. Results for DGP2 are given by the dashed line. Even for large values of γ_0, the pseudo-true values of b are close to $\beta_{10} = 1$, deviating by at most 0.25. Thus, the moment conditions seem to be robust against severe misspecification.

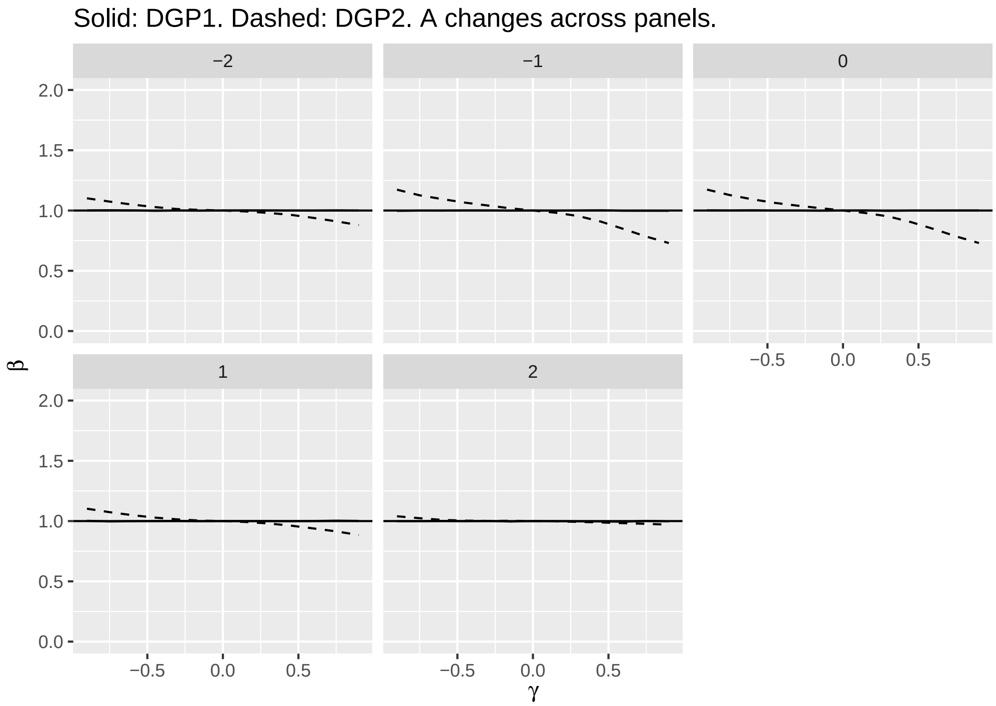


Figure 1: Results sensitivity analysis, β.

We also report results for γ, in Figure 2. Some care is required in interpreting these results, because the interpretation of γ differs across the two DGPs. When the true γ (horizontal axis) is 0, the (pseudo-)true values of γ are 0 for both DGPs. Results for DGP1 (solid line) are on the 45 degree line, as expected. Results for DGP2 always have the right sign, and are typically close to the true γ.

A.5 The usefulness of the moment conditions

The Monte Carlo results in Section 5 suggest that GMM estimators based on the moment conditions in this paper can exhibit fairly large biases even when the sample size is moderately large. In this subsection, we attempt to disentangle the potential value of the moment condition proposed in this paper from the finite sample performance of a particular implementation of a GMM estimator. We do this by calculating the asymptotic variance of four GMM estimators for three generating processes. We then compare the mean squared error implied by the (first-order) asymptotic distribution for the four estimators to each other and

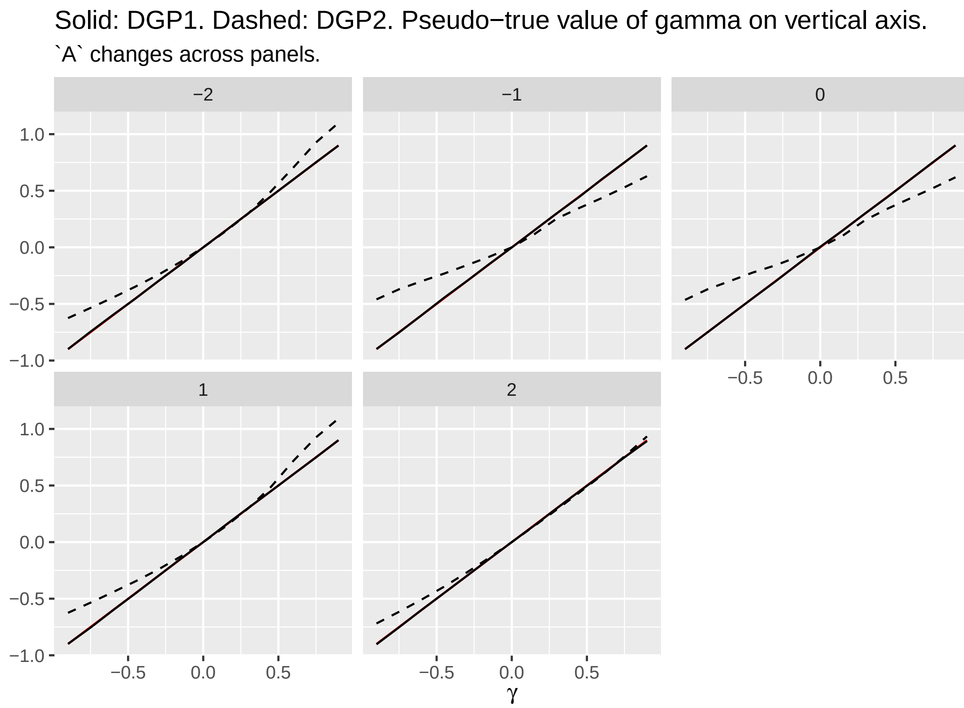


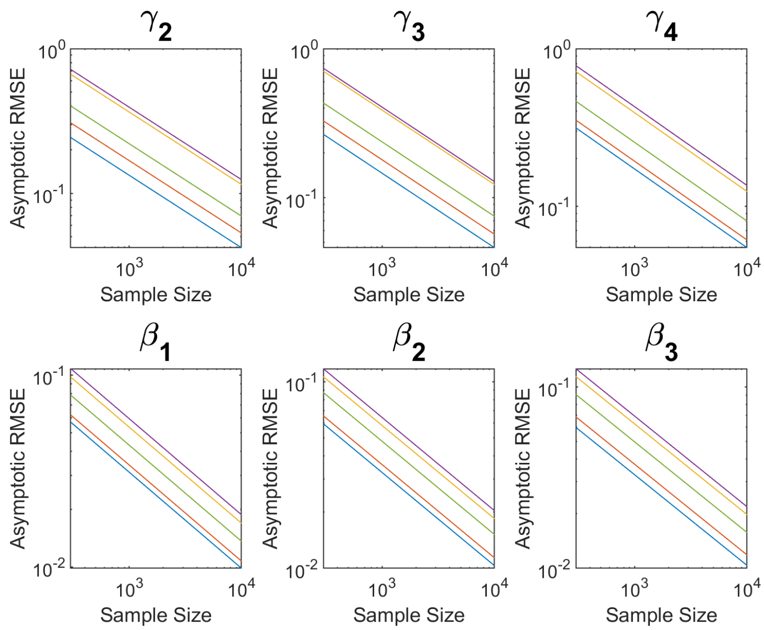
Figure 2: Results sensitivity analysis, γ.

to that of a simple correlated random effects estimator described below.

The correlated random effects estimator will typically have some bias when the claimed model for the individual heterogeneity is misspecified. We calculate this bias by maximizing the expected value of the objective function for the estimator. In order to make this computationally feasible, we consider very simple designs in which the explanatory variables are discrete. We set $T = 3$, $Q = 4$, and the number of explanatory variables to three. The joint distribution of the explanatory variables and the initial condition can take 100 different value with equal probability. These 100 values are constructed as follows: The initial condition, y_0, is drawn are random from $\{1, 2, 3, 4\}$, and 12 random variables, ξ_0, ξ_{jt} ($j = 1, ..., 3$, $t = 0, ..., 3$) are drawn from a standard normal distribution. These are all independent. The explanatory variables and the specification for the heterogeneity are then defined in two steps. For $t = 0, ..., 3$, the first explanatory variable is first defined as $x_{1t} = \xi_{1t} + \xi_0 + 1\{t = 1, y_0 = 2\} + 1\{t = 1, y_0 = 3\} - 0.4$, the second explanatory variable is first defined as $x_{2t} = 1\{\xi_{2t} + \xi_0 > 0\}$, and the third explanatory variable is first defined as $x_{3t} = (\xi_{3t} + \xi_0)^2$. This is not meant to be a realistic distribution, but it captures the idea that there is likely to be some correlation between the explanatory variables and that some of them are likely to be correlated with the initial condition. For each of the 100 points of support for the initial condition and the explanatory variables, we then initially define one of the following for each of the 100 points of support for the initial condition and explanatory variables: $\mu = 0$, $\sigma^2 = 2$ (Design 1), $\mu = 0$, $\sigma^2 = \exp(\xi_{11})$ (Design 2), or $\mu = \exp\left(\frac{1}{4}\sum_{t=0}^{3}\xi_{1t}\right)$, $\sigma^2 = \exp\left(\frac{1}{4}\sum_{t=0}^{3}\xi_{1t}\right)$ (Design 3). In the second step of the definition of the explanatory variables and the specification for the heterogeneity, we normalize each of the three explanatory variables to have mean 0 and variance 3. For Design 3 we also normalize μ to have mean 0 and variance 1, and for Designs 2 and 3, we normalize σ^2 to have mean equal to 2. For each of the 100 points of support for the initial condition and explanatory variables, the unobserved heterogeneity is normally distributed with mean μ and variance σ. The coefficients on the three explanatory variables, β, are all 0.5, the coefficients on the lagged dependent variables, γ, are 0, 1, 1 and 2, and the thresholds, λ are 1, 1 and 3.

The specifications above allow us to calculate the probability of each sequence of outcomes for each of the 100 points. With this, we calculate the asymptotic variance of GMM estimators as well as probability limits and asymptotic variances of correlated random effects estimators. Assuming that the moment conditions identify the parameters of interest, we can then calculate the root mean squared error implied by the asymptotic distribution of the estimators. In these calculations, we assume that γ_0 has been normalized to 0. In Figures 3-5, we plot the log of this root mean squared error against the log of the sample size for the parameters β and γ. The five estimators are (1, in blue) a correlated random effects estimator that specifies the distribution of the unobserved heterogeneity as $A = \sum_{t=0}^{3} x_t' \theta_{3x} + \sum_{q=2}^{4} 1\{y_0 = q\}\theta_q + \sigma Z$, where $Z \sim N(0,1)$, (2, in green) the efficient GMM estimator based on the unconditional moments that result from interaction all the moments in (9)-(11) with $(x_2 - x_1)$, $(x_3 - x_2)$ and with each value of the initial condition, (3, in yellow) the efficient GMM estimator that is based on all the moments in (9)-(11) first aggregated over all values of q_1 and q_3 and then interacted with $(x_2 - x_1)$, $(x_3 - x_2)$ and each value of the initial condition, (4, in purple) the inefficient GMM estimator that uses the same moment conditions as (3) but uses the inverse of the diagonal of the variance of the moments as weighting matrix, and (5, in red) the GMM

Figure 3: RMSEs Implied by the Asymptotic Distribution for Design 1



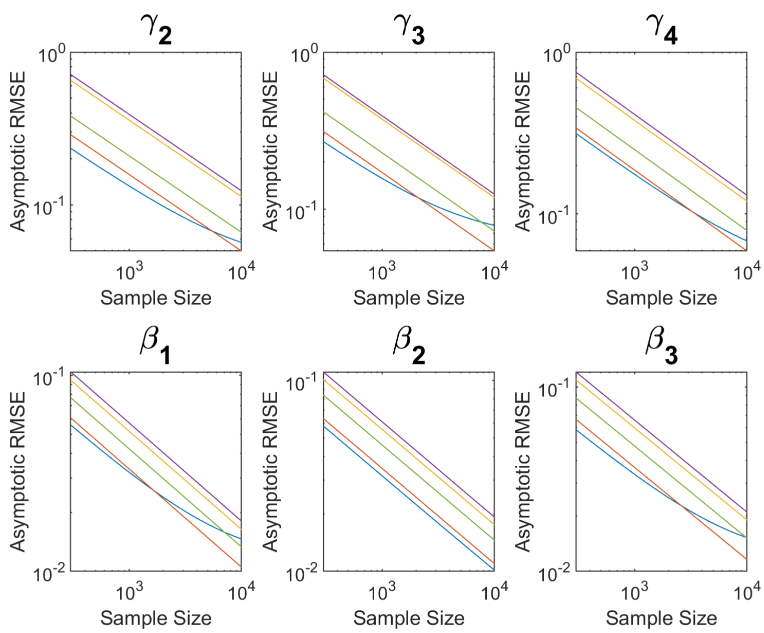
The figure shows the root mean squared error as a function of sample size predicted by asymptotic theory for five estimators: The correlated random effects estimator (blue), the estimator that efficiently exploits all the conditional moment conditions (red), efficient GMM based on many unconditional moments (green), efficient GMM based on few unconditional moments (yellow), and inefficient GMM based on few unconditional moments (purple).

estimator that efficiently combines the conditional moment conditions in (9)-(11).

The first three GMM estimators are deliberately somewhat arbitrary. A comparison between them and the fifth estimator is meant to illustrate the value of efficiently combining the conditional moment conditions.

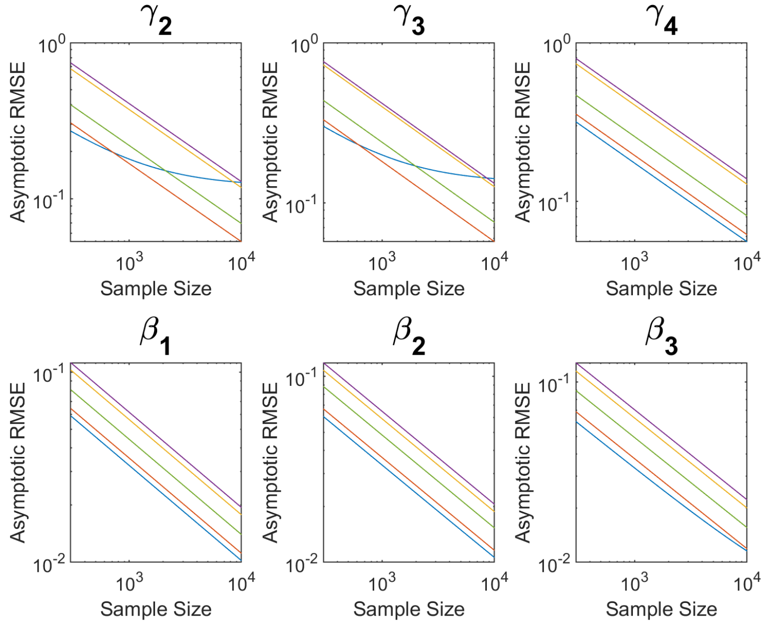
Design 1 is a pure random effects specification. Since this is a special case of the correlated random effects estimator, it is not surprising that the correlated random effects estimator is superior to all the GMM estimators in that case. Perhaps surprisingly, the GMM estimator that efficiently combines the conditional moment conditions in (9)-(11) is almost as efficient as the maximum likelihood estimator in this case. Design 2 and 3 demonstrate that the correlated random effects estimator will generally be misspecified and that this will lead to biased estimates. In these designs, the bias is low enough to make the estimator better than

Figure 4: RMSEs Implied by the Asymptotic Distribution for Design 2



The figure shows the root mean squared error as a function of sample size predicted by asymptotic theory for five estimators: The correlated random effects estimator (blue), the estimator that efficiently exploits all the conditional moment conditions (red), efficient GMM based on many unconditional moments (green), efficient GMM based on few unconditional moments (yellow), and inefficient GMM based on few unconditional moments (purple).

Figure 5: RMSEs Implied by the Asymptotic Distribution for Design 3



The figure shows the root mean squared error as a function of sample size predicted by asymptotic theory for five estimators: The correlated random effects estimator (blue), the estimator that efficiently exploits all the conditional moment conditions (red), efficient GMM based on many unconditional moments (green), efficient GMM based on few unconditional moments (yellow), and inefficient GMM based on few unconditional moments (purple).

the inefficient GMM estimators for moderate sample sizes. Overall, we conclude that the moment conditions proposed in this paper can in principle be very informative about the common parameters in a panel data ordered logit model.