Moment conditions for dynamic panel logit models with fixed effects

Bo E. Honoré
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP22/22

Moment Conditions for Dynamic Panel Logit Models with Fixed Effects*

Bo E. Honoré‡ Martin Weidner§

November 30, 2022

Abstract

This paper investigates the construction of moment conditions in discrete choice panel data with individual specific fixed effects. We describe how to systematically explore the existence of moment conditions that do not depend on the fixed effects, and we demonstrate how to construct them when they exist. Our approach is closely related to the numerical "functional differencing" construction in Bonhomme (2012), but our emphasis is to find explicit analytic expressions for the moment functions. We first explain the construction and give examples of such moment conditions in various models. Then, we focus on the dynamic binary choice logit model and explore the implications of the moment conditions for identification and estimation of the model parameters that are common to all individuals.

*We thank Francesca Molinari and three anonymous referees for constructive feedback and suggestions. We are also grateful to Stephane Bonhomme, Geert Dhaene, Luojia Hu, Yoshitsugu Kitazawa, Shakeeb Khan, Chris Muris, Whitney Newey, and numerous seminar participants for useful comments and discussions. Sharada Dharmasankar provided excellent research assistance. This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University, by the National Science Foundation (Award Numbers 1824131 and 2116630), 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), and by the European Research Council grants ERC-2014-CoG-646917-ROMIA and ERC-2018-CoG-819086-PANEDA.

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

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

1 Introduction

This paper is concerned with estimation of the common parameters in nonlinear panel data models with individual-specific fixed effects in situations where the relevant asymptotics is an increasing number of cross-sectional units (n) observed over a fixed number of time periods (T). We discuss a general approach for constructing conditional moment conditions when the dependent variable can take a finite number of values, and we demonstrate how the approach can be used to construct moment conditions for logit models with strictly exogenous explanatory variables as well as lagged dependent variables.

The economic motivation for the main econometric model investigated in this paper is the question of whether persistence in economic data is due to unobserved heterogeneity or state dependence. This question dates back to papers by Heckman (1978, 1981c,b,a) and can be formulated as a distinction between individual-specific fixed effects and lagged dependent variables. This distinction has been found interesting in many areas of economics. For example, Pakes, Porter, Shepard, and Calder-Wang (2022) have used the same general framework to study the importance of switching costs in a model of health insurance plan choice.[1]

Econometrically, this paper makes a contribution to the literature on estimation of nonlinear econometric models with fixed effects. The challenge in this literature is that when the fixed effects enter in a way that is not additive or multiplicative, then one cannot simply difference or quasi-difference away the fixed effects as one would in a linear or multiplicative model. At the same time, treating the fixed effects as parameters to be estimated in a nonlinear model will generally lead to an estimator of the common parameters that is inconsistent as the number of cross sectional units increases while the number of time periods is fixed.[2] This is what is known as the incidental parameters problem. See Neyman and Scott (1948). One solution to the incidental parameters problem in parametric models is to look for sufficient statistics for the fixed effects. By

[1]See also the references in that paper.

[2]The Poisson regression model is a notable exception to this, see Blundell, Griffith, and Windmeijer (2002) and Lancaster (2002).

definition, the conditional likelihood, conditional on these sufficient statistics, will not depend on the fixed effects, so if it depends on the parameters of interest, then it can be used for estimation. Building on Cox (1958), this approach was pioneered by Rasch (1960b) and Andersen (1970). Unfortunately, there are relatively few models for which one can find sufficient statistics, so an alternative approach is to try to construct moment conditions that depend on the parameters of interest, but not on the individual-specific fixed effects. The papers by Honoré (1992), Hu (2002), Johnson (2004) and Davezies, D'Haultfoeuille, and Mugnier (2022) are specific examples of this.

Bonhomme (2012) developed a general approach for constructing moment conditions via "functional differencing." Our paper makes the proposal in Bonhomme (2012) concrete when the dependent variable is discrete and can take a finite number of values. Specifically, we provide a recipe for how to first investigate numerically whether informative moment conditions can be constructed, and how to subsequently get analytic expressions for the moment conditions. We then apply this machinery to create moment conditions for the case that we consider the most interesting: the fixed effects logit model with strictly exogenous explanatory variables and lagged dependent variables. For models with one lag, we present explicit expressions for all available moment conditions when $T \geq 3$, where T is the number of time period in addition to those that give the initial conditions for the dependent variable. We also derive all the moment conditions for the case with two lagged dependent variables and $T = 4$ and 5, as well as with three lags and $T = 5$. In the case of one lag and three time periods (in addition to the one that delivers the initial condition), our moment conditions are the same as those previously found by Kitazawa (2013, 2016). We also illustrate how the approach can be used to investigate other nonlinear panel data models for which it was not previously known how to approach estimation.

This paper relates to a number of strands of the literature. Estimation of panel data binary response models dates back to Rasch (1960b), who noticed that in a logit model with strictly exogenous explanatory variables, one can make inference regarding the remaining parameters by conditioning on the sums of the dependent variable for each individual, which are the sufficient statistics for the fixed effects. Manski (1987) showed that it is possible to identify and consistently estimate a semiparametric version of the model which relaxes the logistic assumption. Manski (1987)'s estimator is not root-n consistent, and Chamberlain (2010) showed that regular root-n consistent estimation is only possible in a logit setting when only two time periods are observed. See also Jochmans and Magnac (2017).

A number of papers have attempted to relax the assumption that the explanatory variables are strictly exogenous by including lagged dependent variables. Building on Cox (1958), Chamberlain (1985) and Magnac (2000) demonstrated that it is possible to find sufficient statistics for the individual-specific fixed effects in logit models where the only explanatory variables are lagged outcomes. Conditioning on these sufficient statistics leads to a likelihood function that does not depend on the fixed effects, but they typically depend on some of the unknown parameters of the model. The parameters can then be estimated by maximizing the conditional likelihood. The resulting estimator is root-n consistent and asymptotically normal. The conditional likelihood approach referenced above does not generally carry over to logit models that have both lagged dependent variables and strictly exogenous explanatory variables. However, as shown in Honoré and Kyriazidou (2000), this approach does apply if one is also willing to condition on the vector of covariates being equal across certain time periods. This leads to an estimator that is asymptotically normal under suitable regularity conditions, but the rate of convergence is slow when there are continuous covariates.

Papers by Honoré and Kyriazidou (2000), Aristodemou (2018) and Khan, Ponomareva, and Tamer (2019) relax the logistic assumption. This literature suggests that point estimation is sometimes possible and that informative bounds can be constructed when it is not. On the other hand, the impossibility result in Chamberlain (2010) suggests that the most fruitful way to achieve regular root-n consistent estimation is by imposing a logistic assumption.[3] This motivates the estimation approach based on a logistic distribution assumption that we follow in the current paper.

[3]As an alternative to this, Bartolucci and Nigro (2010) and Al-Sadoon, Li, and Pesaran (2017) have proposed models that are outside the framework considered by Chamberlain (2010).

Our results also relate to a larger literature on estimating nonlinear panel data models with fixed effects and short panels.[4] Censored regression was studied by Honoré (1992) for the static model and by Honoré (1993) and Hu (2002) for models with lagged dependent variables. Kyriazidou (1997) constructed an estimator for the static panel data sample selection model and Kyriazidou (2001) for models with lagged dependent variables. Hausman, Hall, and Griliches (1984) developed a conditional likelihood approach for static panel data Poisson regression models and Blundell, Griffith, and Windmeijer (1997, 2002) considered models with lagged dependent variables. Other contributions to the literature on estimating nonlinear panel data models with fixed effects and short panels include Abrevaya (1999, 2000), and more recently, Botosaru and Muris (2018), Muris (2017), and Abrevaya and Muris (2020).

Another set of papers relies on asymptotics as both n and T increase to infinity (possibly at different rates). This includes Hahn and Newey (2004), Arellano and Bonhomme (2009), Bonhomme and Manresa (2015), and Dhaene and Jochmans (2015). In this setting, it becomes important to also be concerned about the possibility of an increasing number of time dummies as in Fernández-Val and Weidner (2016).

The paper is organized as follows: Section 2 introduces our approach for systematically investigating the existence of moment conditions and for constructing them when they exist. The approach is generic and can be uses whenever the dependent variable takes a finite number of values. Building on Dobronyi, Gu, and Kim (2021), we also discuss the number of moment conditions that one should expect to find. In Section 3, we illustrate the approach in the context of a panel data logit AR(1) model with $T = 3$ time periods (in addition to the one that gives the initial condition for y). Section 4 follows up on this by presenting moment conditions for a number of other models. Some of these, most notably the dynamic ordered logit model and the dynamic multinomial model are of independent interest, and we follow up on these models in future work. Other examples, like the panel data logit AR(1) model with heterogeneous time trend and the generalization of our results to AR(p) logit models

[4]Reviews of this literature can be found in Arellano (2003) and Arellano and Bonhomme (2011).

with $p > 1$ are probably mostly interesting because they demonstrate the power of the approach. The remaining example in this section is a static binary response model in which the logistic distribution of the logit model is replaces by a mixture of logits. The results here are interesting because they illustrate how access to panels of length greater than two allows one to construct moment conditions for models that are more general than the logit model. Section 5 discussed conditions under which the moment conditions found in Section 3 are guaranteed to identify the common parameters in the AR(1) logit model with strictly exogenous explanatory variables, while Section 6 discusses and demonstrates how to find moment conditions for the AR(1) logit model with strictly exogenous explanatory variables when the number of time periods is not three. That section also discussed the connection between the moment conditions found in this paper and those in Kitazawa (2013, 2016) as well as the connection between our moment conditions and the first order conditions for the estimator in Honoré and Kyriazidou (2000) (under much stronger assumptions). Section 7 illustrates the usefulness of the approach by estimating a simple model for labor force participation and Section 8 concludes the paper.

2 Incidental parameter free moment conditions

2.1 Model and moment conditions

In this paper, we consider a panel data setting with $i = 1, \ldots, n$ cross-sectional units and $t = 1, \ldots, T$ time periods. An Econometrician models a sequence of discrete outcomes, $Y_i = (Y_{i1}, \ldots, Y_{iT})$, as a function of explanatory variables, $X_i = (X_{i1}, \ldots, X_{iT})$, initial conditions, $Y_i^{(0)} = (Y_{it} : t \leq 0)$, and time invariant "fixed effects", A_i, as

$$\Pr\left(Y_i = y_i \,\middle|\, Y_i^{(0)} = y_i^{(0)},\, X_i = x_i,\; A_i = \alpha_i\right) = f\big(y_i \,\big|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\big). \qquad (1)$$

The function f is assumed to be known up to the finite dimensional parameter θ. The variables Y_i, $Y_i^{(0)}$ and X_i are observed, but the A_i are unobserved. The corresponding conditional probabilities that can be identified from the observed data are

$$\Pr\left(Y_i = y_i \,\middle|\, Y_i^{(0)} = y_i^{(0)},\, X_i = x_i\right) = \int f\left(y_i \,|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\right) g\left(\alpha_i \,\middle|\, y_i^{(0)},\, x_i\right) d\alpha_i,$$

where the probability mass or density function, $g\left(\alpha_i \,\middle|\, x_i,\, y_i^{(0)}\right)$, of A_i conditional on X_i and $Y_i^{(0)}$ is left unspecified. We use $\mathcal{Y}$ to denote the set of possible values of $Y_i = (Y_{i1}, \ldots, Y_{iT})$, which will be a finite set in all the models considered in this paper.

Throughout this paper we assume that that $(Y_i^{(0)}, Y_i, X_i, A_i)$ are independent and identically distributed across $i = 1, \ldots, n$, and our goal is to estimate the common parameters θ from the observed data as $n \to \infty$ and T is fixed. The difficulty in identifying and estimating θ is that the individual specific fixed effects $(\alpha_1, \ldots, \alpha_n)$, or equivalently their unknown conditional distribution $g\left(\alpha_i \,\middle|\, y_i^{(0)},\, x_i\right)$, constitute a high-dimensional nuisance parameter, that is, we are faced with a classic Neyman and Scott (1948) incidental parameter problem.

The leading example considered throughout most of this paper is the binary choice logit AR(1) model, where $Y_{it} \in \{0, 1\}$, $X_{it} \in \mathbb{R}^K$, $A_i \in \mathbb{R}$, and the model restriction reads

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i, A_i\right) = \frac{\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)}{1 + \exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)}, \tag{2}$$

with $Y_i^{t-1} = (Y_{i,t-1}, Y_{i,t-2}, \ldots)$, $\beta \in \mathbb{R}^K$ and $\gamma \in \mathbb{R}$. In this example, we have $\theta = (\beta, \gamma)$, $Y_i^{(0)} = Y_{i0}$, and

$$\begin{aligned} f\left(y_i \,|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\right) &= \prod_{t=1}^{T} \frac{\exp\left(y_{it}\left(x_{it}'\,\beta + y_{i,t-1}\,\gamma + \alpha_i\right)\right)}{1 + \exp\left(x_{it}'\,\beta + y_{i,t-1}\,\gamma + \alpha_i\right)} \\ &=: p(y_i, y_i^{(0)}, x_i, \beta, \gamma, \alpha_i). \end{aligned} \tag{3}$$

In this binary choice example, the set of possible outcomes $\mathcal{Y} = \{0, 1\}^T$ has cardinality $|\mathcal{Y}| = 2^T$.

A very general method to overcome the incidental parameter problem in the models considered here is to find moment functions $m(y_i, y_i^{(0)}, x_i, \theta)$ (different from zero) such that the model restriction (1) implies for any true parameter value θ,

$$\mathbb{E}\left[m(Y_i, Y_i^{(0)}, X_i, \theta)\right] = 0. \tag{4}$$

If we can find such valid moment functions, then they can typically be used to study identification of the parameter θ and to estimate it using generalized method of moments (GMM). The main challenge in this process is to find such valid moment functions for a given panel model of interest.

In the absence of any further restriction on the distribution of $(Y_i^{(0)}, X_i, A_i)$, the unconditional moment restriction (4) can only be a consequence of the model (1) if the conditional moment restriction

$$\mathbb{E}\left[m(Y_i, Y_i^{(0)}, X_i, \theta) \,\middle|\, Y_i^{(0)} = y_i^{(0)},\, X_i = x_i,\; A_i = \alpha_i\right] = 0 \tag{5}$$

holds for all possible realizations $y_i^{(0)}$, x_i, α_i. Under weak regularity conditions, (4) then follows from (5) by the law of iterated expectations. Furthermore, (5) can be rewritten as

$$\sum_{y_i \in \mathcal{Y}} m(y_i, y_i^{(0)}, x_i, \theta)\, f\big(y_i \,\big|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\big) = 0, \tag{6}$$

which shows that knowledge of $f\big(y_i \,\big|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\big)$ is sufficient to verify (5), and therefore (4), for a given moment function, m.

Consider a single moment function $m(y_i, y_i^{(0)}, x_i, \theta) \in \mathbb{R}$ and fixed values of $y_i^{(0)}$, x_i, θ. Then, for every value of α_i, the condition (6) constitutes one linear restriction on the vector $[m(y_i, y_i^{(0)}, x_i, \theta) \;:\; y_i \in \mathcal{Y}] \in \mathbb{R}^{|\mathcal{Y}|}$. Finding $m(y_i, y_i^{(0)}, x_i, \theta) \in \mathbb{R}$ for fixed $y_i^{(0)}$, x_i, θ then requires solving an infinite number of linear equations in $|\mathcal{Y}|$ variables. Depending on the choice of $f\big(y_i \,\big|\, y_i^{(0)},\, x_i,\, \alpha_i;\, \theta\big)$ no solution may exist to this infinite dimensional system of equations. The key finding of this paper is that dynamic logits model do generally have solutions to this system, that is, moment conditions of the form (4) are generally available in such models.

2.2 Strategy for exploring and using such moment conditions

In this subsection, we briefly outline a three-step strategy for obtaining valid moment conditions of the form (4) for a model $f\left(y \,\middle|\, y^{(0)},\, x,\, \alpha;\, \theta\right)$ with T time periods. We drop all indices i unless they are explicitly required.

The first step is to determine numerically whether it seems likely that one can find moment functions that satisfy (5). To do this, we choose numerical values for $y^{(0)}$, x, θ, and also choose $Q > |\mathcal{Y}|$ different numerical values for the fixed effects $(\alpha_1, \ldots, \alpha_Q) \subset \mathcal{A}^Q$. We can then check numerically whether for those values, the system

$$\sum_{y \in \mathcal{Y}} m(y)\, f\left(y \,\middle|\, y^{(0)},\, x,\, \alpha_q;\, \theta\right) = 0, \qquad q = 1, \ldots, Q, \tag{7}$$

of Q equations in $|\mathcal{Y}|$ unknowns $[m(y) : y \in \mathcal{Y}] \in \mathbb{R}^{|\mathcal{Y}|}$ has a solution other than $m = 0$ (and if so, how many). If the Q equations do have a solutions, then one could repeat this exercise for multiple randomly chosen numerical values of $y^{(0)}$, x, θ and of the fixed effects. In this step it is important to use sufficient numerical precision in those calculations, see Appendix A.4 for more details.

If the conclusion of the first step is that moment functions seem to exist, then the next step is to find them. One way to proceed is by choosing specific numerically values for $(\alpha_1, \ldots, \alpha_Q)$, but now solve the system (7) analytically for arbitrary values of $y^{(0)}$, x, θ.[5] The corresponding solution for $m(y)$ will depend on $y^{(0)}$, x, and θ, and we therefore denote the solution by $m(y, y^{(0)}, x, \theta)$. The solution will not depend on $\alpha_1, \ldots, \alpha_Q$ if we truly found a valid moment condition for the chosen model. See Section 3 for a concrete example.

Since the moment functions, $m(y, y^{(0)}, x, \theta)$, obtained in the second step are obtained using a set of specific numerical values of $\alpha_1, \ldots, \alpha_Q$, the third step is to verify analytically that they satisfy the condition (6) for all $\alpha \in \mathbb{R}$.

Once one has constructed moment functions, $m(y, y^{(0)}, x, \theta)$, using the strategy out-

[5] Since we consider discrete choice, the initial condition $y^{(0)}$ takes a finite number of discrete values, and we can perform the analysis separately for each of value of $y^{(0)}$. But for x and θ we need to allow for arbitrary general values in this step.

lined so far, the next step is to study the implications of those moment functions for identification and estimation of θ. It is also useful to study the properties of the moment functions to obtain a better understanding of their structure and origin. In particular, the first two steps can only be implemented for a given number of time periods T. However, by studying the moment functions obtained for specific choices of T, one may be able to draw general conclusions that make it possible to write down all moment functions for a given model for all possible values of T.

In the next section, we follow the steps outlined above to construct moment conditions for the binary choice logit AR(1) model in equation (2). However, there are many other interesting semi-parametric discrete choice panel models $f\big(y \,\big|\, y^{(0)},\, x,\, \alpha;\, \theta\big)$ for which moment conditions of the form (4) exist, but have not yet been studied systematically — see Section 4 below for some concrete examples. The above work program can therefore be seen as a blueprint for an extensive research agenda beyond the current paper. We have recently implemented this blueprint in Honoré, Muris, and Weidner (2021) for the case of dynamic ordered choice panel models. Davezies, D'Haultfoeuille, and Mugnier (2022) can be seen as another example that implements the above program for static binary choice panel model with idiosyncratic error distributions that generalize the logistic case in a particular way.

2.3 Lower bound on the number of moment conditions

Dobronyi, Gu, and Kim (2021) point out that it is sometimes possible to determine a lower bound on the number of moment conditions that can be derived for a given model. Specifically, for many of the models considered below we have $A \in \mathbb{R}$, and one can write the probability distribution in (1) as

$$f\big(y \,\big|\, y^{(0)},\, x,\, \alpha;\, \theta\big) = \kappa\,(a) \sum_{k=1}^{K} a^{k-1} c_k\,(y)$$

for some $K \in \{1, 2, \dots\}$, 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 analogous to our discussion in the last subsection those arguments are dropped to focus more clearly on the dependence on α and y.

A moment function must then satisfy

$$\sum_{y \in \mathcal{Y}} m(y) \sum_{k=1}^{K} a^{k-1} c_k(y) = 0, \qquad \text{for all } a \in (0, \infty) \tag{8}$$

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 $|\mathcal{Y}|$ unknown parameters $m(y)$. We therefore have at least $|\mathcal{Y}| - K$ linearly independent solutions $m(y)$. In other words, the model must have at least $|\mathcal{Y}| - K$ conditional moment conditions (conditional on the initial conditions and on the explanatory variables). Of course, there is no guarantee that all of these moment conditions are functions of the common parameter, θ.

3 Moment functions for the $T = 3$ logit AR(1) model

In this section, we apply the strategy outlined in Section 2.2 to construct moment conditions for the binary choice logit AR(1) model in equation (2) when T is three. In most applications, this corresponds to a total of four time periods: three for which the models is assumed to apply, plus one that delivers the initial condition, y_0.

3.1 Verifying existence of moment functions numerically

By numerically evaluating whether solutions to (7) exist for this model, one finds that $T = 3$ is the smallest number of time periods for which non-zero valid moment functions are available. Our discussion in Section 6.1 below formally shows that it is indeed not possible to derive moment conditions when $T = 2$. This is the reason why we focus on $T = 3$ in this section. For $T = 3$ and $\gamma \neq 0$, one furthermore finds by numerical experimentation that for each value of the initial condition y_0, there exist exactly two linearly independent moment functions that satisfy (7).

3.2 Finding analytical solutions for the moment functions

Having verified the existence of moment functions numerically, the next goal is to find analytic formulas for them. That is, we want to find functions $m(y, y_0, x, \beta, \gamma)$ that satisfy (6).

Since $T = 3$, we have $|\mathcal{Y}| = 2^T = 8$. We define vectors in $\mathbb{R}^8$ for the model probabilities and for the candidate moment functions:

$$\mathbf{p}(y_0,x,\beta,\gamma,\alpha)=\begin{pmatrix} p((0,0,0),y_0,x,\beta,\gamma,\alpha) \\ p((0,0,1),y_0,x,\beta,\gamma,\alpha) \\ p((0,1,0),y_0,x,\beta,\gamma,\alpha) \\ p((0,1,1),y_0,x,\beta,\gamma,\alpha) \\ p((1,0,0),y_0,x,\beta,\gamma,\alpha) \\ p((1,0,1),y_0,x,\beta,\gamma,\alpha) \\ p((1,1,0),y_0,x,\beta,\gamma,\alpha) \\ p((1,1,1),y_0,x,\beta,\gamma,\alpha) \end{pmatrix}, \quad \mathbf{m}(x,y_0,\beta,\gamma)=\begin{pmatrix} m((0,0,0),y_0,x,\beta,\gamma) \\ m((0,0,1),y_0,x,\beta,\gamma) \\ m((0,1,0),y_0,x,\beta,\gamma) \\ m((0,1,1),y_0,x,\beta,\gamma) \\ m((1,0,0),y_0,x,\beta,\gamma) \\ m((1,0,1),y_0,x,\beta,\gamma) \\ m((1,1,0),y_0,x,\beta,\gamma) \\ m((1,1,1),y_0,x,\beta,\gamma) \end{pmatrix}.$$

For simplicity, we drop the arguments y_0, x, β, and γ for the rest of this subsection. They are all kept fixed in the following derivation, and they are the same in the probability vector $\mathbf{p}(\alpha) = \mathbf{p}(x, y_0, \beta, \gamma, \alpha)$ and in the moment vector $\mathbf{m} = \mathbf{m}(x, y_0, \beta, \gamma)$. The probability vector $\mathbf{p}(\alpha)$ as a function of α is given by the model specification. A moment vector $\mathbf{m} \in \mathbb{R}^8$ with $\mathbf{m} \neq 0$ is valid if it satisfies $\mathbf{m}'\,\mathbf{p}(\alpha) = 0$ for all $\alpha \in \mathbb{R}$; that is, a valid moment vector needs to be orthogonal to $\mathbf{p}(\alpha)$ for all values of α. If we can find such a valid moment vector, then its entries will provide moment functions that satisfy equation (5), because $\mathbf{m}'\,\mathbf{p}(\alpha)$ is equal to $\mathbb{E}\left[m(Y, Y_0, X, \beta_0, \gamma_0) \,\middle|\, Y_0 = y_0,\ X = x,\ A = \alpha\right]$.

Any valid moment vector also satisfies $\lim_{\alpha\to\pm\infty} \mathbf{m}'\,\mathbf{p}(\alpha) = 0$. Moreover, the model probabilities $\mathbf{p}(\alpha)$ are continuous functions of α with $\lim_{\alpha\to-\infty} \mathbf{p}(\alpha) = \mathbf{e}_1 = (1, 0, 0, 0, 0, 0, 0, 0)'$ and $\lim_{\alpha\to+\infty} \mathbf{p}(\alpha) = \mathbf{e}_8 = (0, 0, 0, 0, 0, 0, 0, 1)'$, where $\mathbf{e}_k$ denotes the k'th standard unit vector in eight dimensions. From this, we conclude:

(1) Any valid moment vector $\mathbf{m}$ satisfies $\mathbf{e}_1' \, \mathbf{m} = 0$ and $\mathbf{e}_8' \, \mathbf{m} = 0$.

Furthermore, from our "step 1" analysis with concrete numerical values we already know that:[6]

(2) There are two linearly independent solutions to the equations $\mathbf{m}' \, \mathbf{p}(\alpha) = 0$ for all $\alpha \in \mathbb{R}$. This implies that the set of valid moment vectors $\mathbf{m}$ is two-dimensional.

Motivated by hypothesis (2), the aim is to find two linearly independent moment vectors $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ for each $y_0 \in \{0, 1\}$. To distinguish $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ from each other, we impose the condition $\mathbf{e}_7' \, \mathbf{m}^{(0)} = 0$ for the first vector and the condition $\mathbf{e}_2' \, \mathbf{m}^{(1)} = 0$ for the second vector. In addition, we require a normalization for each of these vectors, because an element of the nullspace can be multiplied by an arbitrary nonzero constant to obtain another element of the nullspace. We choose the normalizations $\mathbf{e}_4' \, \mathbf{m}^{(0)} = -1$ and $\mathbf{e}_5' \, \mathbf{m}^{(1)} = -1$. Together with the conditions in (1), this specifies four affine restrictions on each of the vectors $\mathbf{m}^{(0)}, \mathbf{m}^{(1)} \in \mathbb{R}^8$. To define $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ uniquely, we require four more affine conditions for each. We therefore choose four numeric values α_q and impose the orthogonality between $\mathbf{p}(\alpha_q)$ and $\mathbf{m}^{(0/1)}$. Thus, motivated by (1) and (2), we need to solve the following two linear systems of equations:

(0) $\mathbf{e}_1' \, \mathbf{m}^{(0)} = 0, \quad \mathbf{e}_8' \, \mathbf{m}^{(0)} = 0, \quad \mathbf{e}_7' \, \mathbf{m}^{(0)} = 0, \quad \mathbf{e}_4' \, \mathbf{m}^{(0)} = -1,$
$\mathbf{p}'(\alpha_q) \, \mathbf{m}^{(0)} = 0, \quad \text{for } q = 1, 2, 3, 4.$

(1) $\mathbf{e}_1' \, \mathbf{m}^{(1)} = 0, \quad \mathbf{e}_8' \, \mathbf{m}^{(1)} = 0, \quad \mathbf{e}_2' \, \mathbf{m}^{(1)} = 0, \quad \mathbf{e}_5' \, \mathbf{m}^{(1)} = -1,$
$\mathbf{p}'(\alpha_q) \, \mathbf{m}^{(1)} = 0, \quad \text{for } q = 1, 2, 3, 4.$

If it is indeed possible to find such moment functions $\mathbf{m}^{(0/1)}$, then it must be possible for the four values of α to be chosen arbitrarily without affecting the solutions $\mathbf{m}^{(0/1)}$. For example, $\alpha_q = q$ is a valid choice. Note that the two-dimensional span of the vectors

[6] The numerical experiment does not provide a proof of this, but we still take this as an input in our moment condition derivation, with the final justification given by Lemma 1.

$\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ and the potential of the moment conditions to identify and estimate β and γ is not affected by the normalizations $\mathbf{e}_4' \, \mathbf{m}^{(0)} = -1$, $\mathbf{e}_5' \, \mathbf{m}^{(1)} = -1$, $\mathbf{e}_7' \, \mathbf{m}^{(0)} = 0$, and $\mathbf{e}_2' \, \mathbf{m}^{(1)} = 0$.

The systems of linear equations (0) and (1) above uniquely determine $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$. By defining the 8×8 matrices $\mathbf{B}^{(0)} = [\mathbf{e}_1, \mathbf{e}_8, \mathbf{e}_7, \mathbf{e}_4, \mathbf{p}(\alpha_1), \mathbf{p}(\alpha_2), \mathbf{p}(\alpha_3), \mathbf{p}(\alpha_4)]'$, and $\mathbf{B}^{(1)} = [\mathbf{e}_1, \mathbf{e}_8, \mathbf{e}_2, \mathbf{e}_5, \mathbf{p}(\alpha_1), \mathbf{p}(\alpha_2), \mathbf{p}(\alpha_3), \mathbf{p}(\alpha_4)]'$, we can rewrite those systems of equations as $\mathbf{B}^{(0)} \, \mathbf{m}^{(0)} = -\mathbf{e}_4$, and $\mathbf{B}^{(1)} \, \mathbf{m}^{(1)} = -\mathbf{e}_4$. Solving this gives

$$
\begin{aligned}
\mathbf{m}^{(0)} &= - \left(\mathbf{B}^{(0)}\right)^{-1} \mathbf{e}_4 \, , \\
\mathbf{m}^{(1)} &= - \left(\mathbf{B}^{(1)}\right)^{-1} \mathbf{e}_4 \, .
\end{aligned} \tag{9}
$$

Plugging the analytical expression for $\mathbf{p}(\alpha) = \mathbf{p}(x, y_0, \beta, \gamma, \alpha)$ into the definitions $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(0)}$, we thus obtain analytical expressions for $\mathbf{m}^{(0)} = \mathbf{m}^{(0)}(x, y_0, \beta, \gamma)$ and $\mathbf{m}^{(1)} = \mathbf{m}^{(1)}(x, y_0, \beta, \gamma)$.

To report the results, we denote the components of the solutions $\mathbf{m}^{(0/1)} = \mathbf{m}^{(0/1)}(x, y_0, \beta, \gamma) \in \mathbb{R}^8$ by $m^{(0/1)}(y, x, y_0, \beta, \gamma) \in \mathbb{R}$, for $y \in \mathcal{Y}$. Furthermore, let $x_{ts} = x_t - x_s$. Then, the solutions are

$$
m^{(0)}(y, y_0, x, \beta, \gamma) = \begin{cases}
\exp\left(x_{23}'\beta\right) - 1 & \text{if } y = (0,0,1), \\
-1 & \text{if } (y_1, y_2) = (0,1), \\
\exp\left(x_{31}'\beta - y_0\,\gamma\right) & \text{if } y = (1,0,0), \\
\exp\left(x_{21}'\beta + (1-y_0)\,\gamma\right) & \text{if } y = (1,0,1), \\
0 & \text{otherwise,}
\end{cases}
$$

$$
m^{(1)}(y, y_0, x, \beta, \gamma) = \begin{cases}
\exp\left(x_{12}'\beta + y_0\,\gamma\right) & \text{if } y = (0,1,0), \\
\exp\left(x_{13}'\beta - (1-y_0)\,\gamma\right) & \text{if } y = (0,1,1), \\
-1 & \text{if } (y_1, y_2) = (1,0), \\
\exp\left(x_{32}'\beta\right) - 1 & \text{if } y = (1,1,0), \\
0 & \text{otherwise.}
\end{cases} \tag{10}
$$

The two solutions in (10) are closely related: If Y_t is generated according to (2), then

$Z_t = 1 - Y_t$ is also generated according to (2), but with X_t replaced by $-X_t$ and A replaced by $A - \gamma$. The solutions $m^{(0)}$ and $m^{(1)}$ are symmetric in the sense that $m^{(0)}(y, y_0, x, \beta, \gamma) = m^{(1)}(1-y, 1-y_0, -x, \beta, \gamma)$.

3.3 Verifying that the moment functions are valid

The following lemma establishes that the moment functions, $m^{(0/1)}(y, y_0, x, \beta, \gamma)$, displayed in (10) are indeed valid. For this, it is not relevant how the moment functions were derived.

Lemma 1 *If the outcomes $Y = (Y_1, Y_2, Y_3)$ are generated from model (2) with $T = 3$ and true parameters β_0 and γ_0, then we have for all $q \in \{0,1\}$, $y_0 \in \{0,1\}$, $x \in \mathbb{R}^{K\times 3}$, $\alpha \in \mathbb{R}$ that*

$$\mathbb{E}\left[m^{(q)}(Y, Y_0, X, \beta_0, \gamma_0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha \right] = 0.$$

This Lemma is a special case of Theorem 2 below. However, one can prove this lemma more easily by direct calculation: just plug-in the definition of the probabilities $p(y, y_0, x, \beta_0, \gamma_0, \alpha)$ and moments $m^{(0/1)}(y, y_0, x, \beta_0, \gamma_0)$ to show that

$$\sum_{y\in\{0,1\}^3} p(y, y_0, x, \beta_0, \gamma_0, \alpha)\; m^{(0/1)}(y, y_0, x, \beta_0, \gamma_0) = 0.$$

The details of this calculation are provided in Appendix B.2.1.

3.4 On the number of moment conditions

As explained in Section 2.3, Dobronyi, Gu, and Kim (2021) provide a method for deriving (a lower bound on) the number of moment conditions for a given model. For the panel logit AR(1) model, the probability distribution for $Y_i = (Y_{i1}, \ldots, Y_{iT})$

(conditional on Y_{i0}, Y_i, A_i) is given by

$$f\big(y \,|\, y^{(0)},\, x,\, \alpha;\, \theta\big) \;= \prod_{t=1}^{T} \frac{[\exp{(x_t' \, \beta + y_{t-1} \, \gamma + \alpha)}]^{y_{it}}}{1 + \exp{(x_t' \, \beta + y_{t-1} \, \gamma + \alpha)}},$$

with $a = \exp(\alpha)$ and $\pi_t(y_{t-1}) = \exp[x_i' \, \beta + y_{t-1} \, \gamma]$, we then have

$$\begin{aligned} f\big(y \,|\, y^{(0)},\, x,\, \alpha;\, \theta\big) \;&= \prod_{t=1}^{T} \frac{[a \, \pi_t(y_{t-1})]^{y_t}}{1 + a \, \pi_t(y_{t-1})} = \frac{[a \, \pi_1(y_0)]^{y_1}}{1 + a \, \pi_1(y_0)} \prod_{t=2}^{T} \frac{[a \, \pi_t(y_{t-1})]^{y_t}}{1 + a \, \pi_t(y_{t-1})} \\ &= \frac{[a \, \pi_1(y_0)]^{y_1}}{1 + a \, \pi_1(y_0)} \prod_{t=2}^{T} \frac{[1 + a \, \pi_t(1 - y_{t-1})] \; [a \, \pi_t(y_{t-1})]^{y_t}}{[1 + a \, \pi_t(0)][1 + a \, \pi_t(1)]} = \kappa(a) \, \cdot \widetilde{p}(y, a), \end{aligned}$$

where we defined

$$\begin{aligned} \kappa(a) &= \frac{1}{1 + a \, \pi_1(y_0)} \cdot \prod_{t=2}^{T} \frac{1}{[1 + a \, \pi_t(0)][1 + a \, \pi_t(1)]}, \\ \widetilde{p}(y, a) &= [a \, \pi_1(y_0)]^{y_1} \prod_{t=2}^{T} \left\{ [1 + a \, \pi_t(1 - y_{t-1})] \; [a \, \pi_t(y_{t-1})]^{y_t} \right\} = \sum_{k=1}^{2T} a^{k-1} \, c_k(y). \end{aligned}$$

This has the exact structure of equation (8) with $K = 2T$ and $|\mathcal{Y}| = 2^T$, so there must be at least $2^T - 2T$ moment conditions. When $T = 3$, the lower bound on the number of conditional moment conditions is $2^T - 2T = 2$, which is exactly the number of moment conditions we found in Lemma 1 above.

4 Examples of moment functions in other models

In this section, we briefly discuss some other fixed effects panel data models with discrete outcomes, for which it is possible to use the approach outlined in Section 2.2 to derive moment conditions. The goal of this section is to illustrate the broad applicability of the moment condition approach, and it can be skipped by a reader interested in the binary choice AR(1) panel model only.

4.1 Static binary choice models

In a static panel binary response model with strictly exogenous regressors $X_i = (X_{i1}, \dots, X_{iT})$ and fixed effects A_i the conditional distribution of the outcomes $Y_i = (Y_{i1}, \dots, Y_{iT})$ is given by

$$f\big(y_i \,\big|\, x_i,\, \alpha_i;\, \beta\big) = \prod_{t=1}^{T} \left[F\left(x_{it}'\,\beta + \alpha_i\right)\right]^{y_{it}} \left[1 - F\left(x_{it}'\,\beta + \alpha_i\right)\right]^{1-y_{it}}, \tag{11}$$

where $F(\cdot)$ is a cumulative distribution function. The distribution in (11) is a special case of (1). For the logistic case, $F(\varepsilon) = [1+\exp(-\varepsilon)]^{-1}$, one can use that $S_i = \sum_{t=1}^{T} Y_{it}$ is a sufficient statistic for A_i to estimate β via the conditional maximum likelihood estimator (CMLE) that conditions on S_i, see Rasch (1960b) and Andersen (1970). In fact, Chamberlain (2010) showed that for $T = 2$, and subject to weak regularity conditions, root-n-consistent estimation of β is only possible if $F(\varepsilon)$ is logistic.[7] This implies that non-trivial moment functions $m(y_i, x_i, \beta)$ are available for the $T = 2$ static panel model if and only if $F(\varepsilon) = [1 + \exp(-\kappa\,\varepsilon + \mu)]^{-1}$, for some constants $\kappa > 0$ and $\mu \in \mathbb{R}$.

Interestingly, for the static panel model with $T = 3$ one can allow for distributions $F(\cdot)$ that are not logistic and still estimate the parameter β at $\sqrt{n}$ rate, that is, the $T = 2$ result of Chamberlain (2010) does not apply in that case. In particular, for $T = 3$, Johnson (2004) and Davezies, D'Haultfoeuille, and Mugnier (2022) consider distributions of the form $F(\varepsilon) = \left[1 + w_1 \exp(-\lambda_1\,\varepsilon) + w_2 \exp(-\lambda_2\,\varepsilon)\right]^{-1}$, with non-negative real-valued parameters w_1, w_2, λ_1, λ_2, and derive moment conditions for (11). One can use the machinery in Section 2 to show that the specification considered by Johnson (2004) and Davezies, D'Haultfoeuille, and Mugnier (2022) is not the only extension of the logistic distribution that provides moment conditions for the $T = 3$ static model. For example, consider the case where $F(\varepsilon)$ is a mixture of two logistic distributions

[7] This result is for ε_{it} independent across t. Generalizations that allow for dependence across t are derived in Magnac (2004).

with the same variance:

$$F(\varepsilon) = \frac{\omega}{1 + \exp(-\lambda\,\varepsilon + \mu_1)} + \frac{1-\omega}{1 + \exp(-\lambda\,\varepsilon + \mu_2)}, \tag{12}$$

where $\omega \in [0,1]$ is a mixture weight, $\lambda > 0$ parametrizes the common variance of the logistic components, and $\mu_1, \mu_2 \in \mathbb{R}$ parametrize the mean of the two components. When plugging (12) into (11) and then applying the procedure described in Section 2.2 to derive valid moment functions, one finds that for $\mu_1 \neq \mu_2$ and $\omega \in (0,1)$ exactly one moment function exists for general values of β and x when $T = 3$. This moment condition is given by

$$\begin{aligned} m(y, x, \beta) = \sum_{(t,s,r)\in\mathcal{P}} \mathbb{1}\left\{(y_t, y_s) = (0,1)\right\}\; (1 - 2\,y_r)\, \operatorname{sgn}(t,s,r)\; \exp[\lambda\,(x_t - x_s)'\beta] \\ \times \Big\{\omega\, \exp[(1-y_r)(\mu_2 - \mu_1)] + (1-\omega)\, \exp[y_r\,(\mu_2 - \mu_1)]\Big\}, \end{aligned} \tag{13}$$

where $\mathcal{P}$ is the set of all six permutations of $(1,2,3)$, and for $(t,s,r) \in \mathcal{P}$ the signature of that permutation is denoted by $\operatorname{sgn}(t,s,r)$.

In addition, we have found numerically that one can allow for more general finite mixtures of logistic distributions when T exceeds 3. For example, it appears that one can allow for a mixture of three logistics when T is 4, six when $T = 5$, ten when $T = 6$, and eighteen when T is 7. Calculations like the ones in Section 2.3 suggest that if the number of mixtures is Q, then there are $2^T - TQ - 1$ non-trivial conditional moment conditions this model. For example, if T is 9 then there will be seven moment conditions when F is a mixture of 56 logistic cumulative distribution functions. We leave it to future research to derive these and to investigate the extent to which they identify the common parameters of the model.

4.2 Fixed Effect Logit AR(p) Models With $p > 1$

The analysis of the dynamic panel data logit model in Section 3 generalizes to a model with more than one lag. Specifically, consider the model

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i, A_i, \beta, \gamma\right) = \frac{\exp\left(X_{it}'\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}{1 + \exp\left(X_{it}'\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}, \tag{14}$$

where $\gamma = (\gamma_1, \ldots, \gamma_p)'$. We assume that the autoregressive order $p \in \{2, 3, 4, \ldots\}$ is known, and that outcomes Y_{it} are observed for time periods $t = t_0, \ldots, T$, with $t_0 = 1 - p$. Thus, the total number of time periods for which outcomes are observed is $T_{\text{obs}} = T + p$, consisting of T periods for which the model applies and p periods to observe the initial conditions. We maintain the definition $Y_i = (Y_{i1}, \ldots, Y_{iT})$, but the initial conditions are now described by the vector $Y_i^{(0)} = (Y_{i,t_0}, \ldots, Y_{i0})$.

Numeric calculations similar to those for the model with one lag suggest that for a given value of p, one requires $T \geq 2 + p$ (i.e. $T_{\text{obs}} \geq 2 + 2p$) time periods to find conditional moment conditions that hold without restrictions on the parameters or on the support on the explanatory variables.[8] For example, a model with $p = 3$ lags requires a total of eight time periods; three that provide the initial conditions for Y_{it}, and five for which the model is assumed to apply. Numerical calculations also suggest that the number of linearly independent moment conditions available for each initial condition, $y^{(0)}$, is equal to[9] $2^T - (T + 1 - p)\, 2^p$. In Appendix A.3, we provide analytic formulas for all the moment functions that can be obtained with $T \leq 5$. Specifically, when $p = 2$, we provide four moment functions for $T = 4$ and sixteen for $T = 5$. For $p = 3$, there are eight moment functions, while there are no general moment functions when $p \geq 4$ and $T \leq 5$. Identification of the parameters β and γ for $p \leq 3$ is shown in Appendix Section B.3.

The special case of an AR(2) logit model with fixed effects and no explanatory

[8] In addition to those general moment conditions, there are additional ones that only become available for special values of the parameters and of the regressors.

[9] We have verified this for $p \in \{0, \ldots, 6\}$ and $T \in \{2 + p, \ldots, 8\}$, but believe that this formula for the number of linearly independent moments holds for all integers p, T with $T \geq 2 + p$. However, a general proof of this conjecture is beyond the scope of this paper.

variables was considered in Honoré and Kyriazidou (2019). Numerical calculations in that paper suggested that the common parameters in such a model are point identified for $T = 3$ (i.e. $T_{\text{obs}} = 5$), but no proof of identification was provided. Evaluating the moment functions in Appendix A.3 at $\beta = 0$ makes it clear why $\gamma = (\gamma_1, \gamma_2)$ is identified and how one would estimate it. Specifically, if the outcomes $Y = (Y_1, Y_2, Y_3)$ are generated from the AR(2) panel logit model without explanatory variables, then we have, for all $y^{(0)} \in \{0,1\}^2$ and $\alpha \in \mathbb{R}$, that

$$\mathbb{E}\left[m_{y^{(0)}}(Y,\gamma_0) \,\middle|\, Y^{(0)} = y^{(0)},\, A = \alpha\right] = 0,$$

with moment functions given by

$$m_{(y_0,y_0)}(y,\gamma) = \begin{cases} 1 & \text{if } y = (y_0, 1-y_0, y_0), \\ e^{-\gamma_1} & \text{if } y = (y_0, 1-y_0, 1-y_0), \\ -1 & \text{if } (y_1, y_2) = (1-y_0, y_0), \\ 0 & \text{otherwise}, \end{cases} \qquad m_{(1-y_0,y_0)}(y,\gamma) = \begin{cases} -1 & \text{if } (y_1, y_2) = (1-y_0, y_0), \\ e^{\gamma_2-\gamma_1} & \text{if } y = (y_0, 1-y_0, 1-y_0), \\ e^{\gamma_2} & \text{if } y = (y_0, 1-y_0, y_0), \\ 0 & \text{otherwise}, \end{cases}$$

where $y_0 \in \{0, 1\}$.

The moment functions $m_{(0,0)}$ and $m_{(1,1)}$ are strictly monotone in γ_1 and do not depend on γ_2. Each of them therefore identify the parameter γ_1. For a given value of γ_1, the moment function $m_{(0,1)}$ and $m_{(1,0)}$ are strictly monotone in γ_2, and they therefore each identify the parameter γ_2 once γ_1 has been identified. A GMM estimator based on those moment will be root-n consistent under standard regularity conditions.

4.3 Panel logit AR(1) with heterogeneous time trends

For arbitrary regressors, no valid moment functions seem to exists when some of the elements of β are replaced by fixed effects β_i. However, if one of the explanatory variables is a linear time trend, then it is possible to allow for the coefficient on this variable to differ arbitrarily across observations. Specifically, consider the generalization of the model in equation (2) to

$$\Pr\left(Y_{it}=1 \,\middle|\, Y_i^{t-1}, X_i, A_i\right) = \frac{\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + tD_i + A_i\right)}{1+\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + tD_i + A_i\right)}.$$

By mimicking the calculations in Section 3.4, one finds that at least $\ell_{\min} = 2^T - \frac{T}{3}\left(2T^2 - 3T + 7\right)$ moment conditions need to exist in this model. For $T \geq 9$ we have $\ell_{\min} > 0$, that is, it must be the case that moment conditions for this model exist. See Appendix Section A.5 for details. However, such calculations only yield a lower bound on the number of moment conditions.[10] Numerically, we do not find any moment conditions for $T \leq 8$ for general parameter values with $x_{it} \neq 0$, but we do find two valid moment conditions for $T = 8$ if $x_{it} = 0$ (so there are no additional regressors in the model, and γ is the only common parameter). Both of these moment conditions depend on the parameter γ. In Appendix Section A.5, we discuss these moment conditions.

4.4 Extensions to dynamic ordered logit model and dynamic multinomial logit models

The results derived in Section 3 extend to dynamic panel data versions of other "text-book" logit models. For example, Honoré, Muris, and Weidner (2021) use the procedure outlined in Section 2.2 to generalize the moment conditions in this paper to dynamic panel data ordered logit models of the type

$$Y_{it} = q \quad \text{if} \quad Y_{it}^* \;\in\; (\lambda_{q-1}, \lambda_q], \tag{15}$$

where

$$Y_{it}^* = X_{it}'\,\beta + \sum_{q=1}^{Q} \gamma_q\, \mathbb{1}\left\{Y_{i,t-1} = q\right\} + A_i + \varepsilon_{it}, \tag{16}$$

[10] Numerically, we find 126 linearly independent moment conditions for the model with additional strictly exogenous explanatory variables when $T = 9$. This is larger than the lower bound of $\ell_{\min} = 86$ moment conditions derived in Appendix Section A.5. This illustrates that exploring the polynomial structure of the model probabilities (as described in Section 3.4) does not always give the exact number of available moment conditions in binary logit models.

and ε_{it} is i.i.d. logistically distributed and A_i is an unobserved individual specific fixed effect. In this model, $\lambda_0 = -\infty$, $\lambda_Q = \infty$, and the unknown parameters are β, $\{\gamma_q\}_{q=1}^{Q}$ and $\{\lambda_q\}_{q=1}^{Q-1}$. Honoré, Muris, and Weidner (2021) show that for this model, it is possible to construct $(Q-1)^2 Q$ moment conditions when one has access to data on (Y_{it}, X_{it}) for three time periods in addition to the initial conditions for Y_{it}. When $Q = 2$, these moment conditions coincide with the ones presented in Section 3 of this papers.

In the same spirit, Honoré and Weidner (2022) consider generalizations of the binary model considered in Section 3 to multinomial models. Specifically, assume that there are $Q \in \{2, 3, 4, \ldots\}$ possible for a variable $Y_{it} \in \{1, 2, \ldots, Q\}$ and that Y_{it} is generated by the model

$$Y_{it} = \underset{q \in \{1,2,\ldots,Q\}}{\operatorname{argmax}} \; U_{qit},$$

for $t \in \{1, 2, \ldots, T\}$, with latent variable given by

$$U_{qt} = \sum_{r=1}^{Q} \gamma_{qr} \, \mathbb{1}\{Y_{t-1} = r\} + X_t' \beta_q + A_q + \varepsilon_{qt}.$$

Here, $A = (A_1, \ldots, A_Q)$ are fixed effects, $X = (X_1, X_2, \ldots, X_T)$ are strictly exogenous regressors, and the errors ε_{qt} are independent of A, X, and past outcomes, and are i.i.d. with extreme value distribution. The γ_{qr}'s and β_q are unknown parameters. This model is a dynamic panel data version of the multinomial logit model, and it implies the choice probabilities:

$$P\left(Y_t = q \,\middle|\, Y^{t-1}, X, A\right) = \frac{\exp\left(\sum_{r=1}^{Q} \gamma_{qr} \, \mathbb{1}\{Y_{t-1} = r\} + X_t' \beta_q + A_q\right)}{\sum_{\ell=1}^{Q} \exp\left(\sum_{r=1}^{Q} \gamma_{\ell r} \, \mathbb{1}\{Y_{t-1} = r\} + X_t' \beta_\ell + A_\ell\right)}.$$

Honoré and Weidner (2022) show that when $T = 3$, there are Q linearly independent conditional moment conditions that eliminate the fixed effects for each initial condition. This generalizes the result for the binary logit model in this paper. The calculations in Honoré and Weidner (2022) also demonstrate the number of available moment conditions increases dramatically with larger values of T.

5 Identification

This section shows that the moment conditions for the panel logit AR(1) model in Lemma 1 can be used to uniquely identify the parameters β and γ under appropriate support conditions on the regressor X. The following technical lemma turns out to be very useful in showing this.

Lemma 2 *Let $K \in \mathbb{N}_0$. For every $s = (s_1, \ldots, s_K) \in \{-,+\}^K$ let $g_s : \mathbb{R}^K \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that for all $(\beta, \gamma) \in \mathbb{R}^K \times \mathbb{R}$ we have*

(i) $g_s(\beta, \gamma)$ is strictly increasing in γ.

(ii) For all $k \in \{1, \ldots, K\}$: If $s_k = +$, then $g_s(\beta, \gamma)$ is strictly increasing in β_k.

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

Then, the system of 2^K equations in $K + 1$ variables

$$g_s(\beta, \gamma) = 0, \qquad \text{for all } s \in \{-,+\}^K, \tag{17}$$

has at most one solution.

To explain the lemma, consider the case $K = 1$,[11] when we have two scalar parameters $\beta, \gamma \in \mathbb{R}$. The lemma then requires that the two functions $g_+(\beta, \gamma)$ and $g_-(\beta, \gamma)$ are both strictly increasing in γ, and g_+ is also strictly increasing in β, while g_- is strictly decreasing in β. Thus, $g_+(\beta, \gamma) = 0$ gives a solution for $\gamma = \gamma(\beta)$ that is strictly decreasing in β, while $g_-(\beta, \gamma) = 0$ gives a solution $\gamma(\beta)$ that is strictly increasing, implying that the joint solution must be unique.

Combining the moment conditions in Lemma 1 with the result of Lemma 2 allows us to provide sufficient conditions for point identification in panel logit AR(1) models with $T = 3$. For that purpose we define the sets

$$\mathcal{X}_{k,+} = \{x \in \mathbb{R}^{K \times 3} \, : \, x_{k,1} \leq x_{k,3} < x_{k,2} \ \text{ or } \ x_{k,1} < x_{k,3} \leq x_{k,2}\},$$

[11] Notice that $K = 0$ is trivially allowed in Lemma 2. We then have $s = \emptyset$ and $g_\emptyset : \mathbb{R} \to \mathbb{R}$ is a single increasing function, implying that $g_\emptyset(\gamma) = 0$ can at most have one solution.

$$\mathcal{X}_{k,-} = \{x \in \mathbb{R}^{K\times 3} \, : \, x_{k,1} \geq x_{k,3} > x_{k,2} \;\text{ or }\; x_{k,1} > x_{k,3} \geq x_{k,2}\},$$

for $k \in \{1,\ldots,K\}$. The set $\mathcal{X}_{k,+}$ is the set of possible regressor values $x \in \mathbb{R}^{K\times 3}$ such that either $x_{k,1} \leq x_{k,3} < x_{k,2}$ or $x_{k,1} < x_{k,3} \leq x_{k,2}$; that is, the k'th regressor takes its smallest value in time period $t = 1$ and its largest value in time period $t = 2$. Conversely, the set $\mathcal{X}_{k,-}$ is the set of possible regressor values $x \in \mathbb{R}^{K\times 3}$ for which the k'th regressor takes its largest value in time period $t = 1$ and its smallest value in time period $t = 2$.

The motivation behind the definition of those sets is that for $x \in \mathcal{X}_{k,\pm}$ our moment functions $m^{(0/1)}(y, y_0, x, \beta, \gamma)$ defined in Section 3.2 have convenient monotonicity properties in the parameters β_k. For example, for $m^{(0)}$ we have

$$\frac{\partial\, \mathbb{E}\left[m^{(0)}(Y, Y_0, X, \beta, \gamma) \,\middle|\, Y_0 = y_0,\; X = x\right]}{\partial \beta_k} \begin{array}{ll} > 0, & \text{for } x \in \mathcal{X}_{k,+}\,, \\ < 0, & \text{for } x \in \mathcal{X}_{k,-}\,. \end{array} \tag{18}$$

This is because the parameter β appears in $m^{(0)}(y, y_0, x, \beta, \gamma)$ only through $\exp\left(x'_{23}\beta\right)$, $\exp\left(x'_{23}\beta\right)$ and $\exp\left(x'_{31}\beta\right)$; $x \in \mathcal{X}_{k,+}$ (or $x \in \mathcal{X}_{k,-}$) guarantees that the differences $x_{k,2} - x_{k,3}$ and $x_{k,3} - x_{k,1}$ and $x_{k,2} - x_{k,1}$ are all ≥ 0 (≤ 0), with some of them strictly positive (negative). The moment functions $m^{(1)}(y, y_0, x, \beta, \gamma)$ has exactly the opposite monotonicity properties in β.

Next, for any vector $s = (s_1, \ldots, s_K) \in \{-,+\}^K$ we define the set $\mathcal{X}_s = \bigcap_{k\in\{1,\ldots,K\}} \mathcal{X}_{k,s_k}$ and the corresponding expected moment functions, for $q, y_0 \in \{0,1\}$,

$$\overline{m}^{(q)}_{y_0,s}(\beta,\gamma) = \mathbb{E}\left[m^{(q)}(Y, Y_0, X, \beta, \gamma) \,\middle|\, Y_0 = y_0,\; X \in \mathcal{X}_s\right].$$

Because $\mathcal{X}_s$ is the intersection of the sets $\mathcal{X}_{k,\pm}$, the expected moment functions $\overline{m}^{(0/1)}_{y_0,s}(\beta,\gamma)$ have monotonicity properties with respect to all the elements of β specified by the sign vector $s = (s_1, \ldots, s_K)$. For example, (18) implies that $\overline{m}^{(0)}_{y_0,s}(\beta,\gamma)$ is strictly increasing in β_k if $s_k = +$, and strictly decreasing in β_k if $s_k = -$, for all $k \in \{1,\ldots,K\}$.

Theorem 1 *Let $q, y_0 \in \{0,1\}$. Let the outcomes $Y = (Y_1, Y_2, Y_3)$ be generated from model (2) with $T = 3$ and true parameters β_0 and γ_0. Furthermore, for all $s \in \{-,+\}^K$ assume that*

$$\Pr\left(Y_0 = y_0,\ X \in \mathcal{X}_s\right) > 0,$$

and that the expected moment function $\overline{m}^{(q)}_{y_0,s}(\beta,\gamma)$ is well-defined.[12] *Then, the solution to*

$$\overline{m}^{(q)}_{y_0,s}(\beta,\gamma) = 0 \qquad \text{for all } s \in \{-,+\}^K \tag{19}$$

is unique and given by (β_0, γ_0). Thus, the parameters β_0 and γ_0 are point-identified

The proof of the theorem is provided in the appendix. Notice that only one of the moment functions $m^{(0)}$ or $m^{(1)}$ is required to derive identification in Theorem 1, and only one of the initial conditions $y_0 \in \{0,1\}$ needs to be observed. The key assumption in Theorem 1 is that we have enough variation in the observed regressor values $X = (X_1, X_2, X_3)$ to satisfy the condition $\Pr\left(Y_0 = y_0,\ X \in \mathcal{X}_s\right) > 0$, for all $s \in \{+,-\}^K$.

Theorem 1 achieves identification of β and γ via conditioning on the sets of regressor values $\mathcal{X}_s$, which all have positive Lebesgue measure. By contrast, the conditional likelihood approach in Honoré and Kyriazidou (2000) conditions on the set $x_2 = x_3$, which has zero Lebesgue measure, and therefore also often zero probability measure, implying that the resulting estimates for β and γ usually converge at a rate slower than root-n. In our approach here, using the sample analogs of the moment conditions $\overline{m}^{(0/1)}_{y_0,s}(\beta,\gamma) = 0$ for $s \in \{+,-\}^K$, we immediately obtain GMM estimates for β and γ that are root-n consistent under standard regularity conditions.

However, in practice, we do not actually recommend estimation via the moment conditions in Theorem 1, because by conditioning on $X \in \mathcal{X}_s$ these moment conditions still only use a small subset of the available information in the data. Instead, many more unconditionally valid moment conditions for β and γ can be obtained from Lemma 1

[12] We could always guarantee $\overline{m}^{(q)}_{y_0,s}(\beta,\gamma)$ to be well-defined by modifying the definition the set $\mathcal{X}_s$ to only contain bounded regressor values.

(or from Theorem 2 below for $T > 3$), resulting in potentially much more efficient estimators for β and γ, and Section 7 describes how we implement such estimators in practice. Nevertheless, from a theoretical perspective, the identification result in Theorem 1 is important, because it comprises a significant improvement over existing results for dynamic panel logit models with explanatory variables.

6 Panel logit AR(1) model for general $T \geq 3$

In Section 3.2 above we already found analytic formulas for valid moment functions that are free of the fixed effects for the panel logit AR(1) model with three time periods. In this section we discuss various generalizations of this result, most importantly to $T > 3$ time periods. Before presenting those positive results, we first briefly discuss a negative result for $T = 2$ time periods.

6.1 Impossibility of moment conditions when $T = 2$

Here, we argue that it is not possible to derive moment conditions for model (2) on the basis of two time periods plus the initial condition, y_0. If one could construct such moment conditions for $T = 2$ that hold conditional on the individual specific effects A, then the corresponding moment functions $m(y, y_0, x, \beta, \gamma)$ would satisfy

$$\sum_{y \in \{0,1\}^2} p(y, y_0, x, \beta, \gamma, \alpha)\, m(y, y_0, x, \beta, \gamma) = 0, \tag{20}$$

for all $\alpha \in \mathbb{R}$. In the limit $\alpha \to \infty$ the model probabilities become zero, except for $p((1,1), y_0, x, \beta, \gamma, \alpha) \to 1$. This implies $m((1,1), y_0, x, \beta, \gamma) = 0$. Analogously, in the limit $\alpha \to -\infty$ we have $p((0,0), y_0, x, \beta, \gamma, \alpha) \to 1$, which implies $m((0,0), y_0, x, \beta, \gamma) = 0$. Thus, only $m((0,1), y_0, x, \beta, \gamma)$ and $m((1,0), y_0, x, \beta, \gamma)$ can be non-zero, and (20) therefore implies that

$$\frac{m((0,1), y_0, x, \beta, \gamma)}{m((1,0), y_0, x, \beta, \gamma)} = -\frac{p((1,0), y_0, x, \beta, \gamma, \alpha)}{p((0,1), y_0, x, \beta, \gamma, \alpha)}$$

$$= -\exp\left((x_1 - x_2)' \beta + \gamma y_0\right) \frac{1 + \exp(x_2'\beta + \alpha)}{1 + \exp(x_2'\beta + \gamma + \alpha)}. \tag{21}$$

Unless $\gamma = 0$, the right hand side of (21) will always have a non-trivial dependence on α, implying that no moment conditions can be constructed for $T = 2$ (that are valid conditional on arbitrary $A = \alpha$). For $\gamma = 0$ equation (21) yields the moment conditions implied by Rasch (1960a)'s conditional likelihood.

6.2 Master lemma for obtaining all moment conditions

One can work out analytic moment functions for model (2) with $T = 4$ and $T = 5$ using the same derivation method described in Section 3.2 for $T = 3$. These are relatively brute force calculations that only require limited human input and creativity. However, once those analytic moment functions are obtained, one can move on to study their common structure, which leads to the following lemma that allows us to derive all the valid moment conditions for the panel logit AR(1) model for an arbitrary number of time periods.

Before presenting the lemma, we introduce some additional notation: The cumulative distribution function of the logistic distribution is given by $\Lambda(\xi) := [1+\exp(-\xi)]^{-1}$. In addition, we define the cyclical decrement function $\delta : \{1, 2, 3\} \to \{1, 2, 3\}$ by

$$\delta(t) := \begin{cases} 3 & \text{for } t = 1, \\ 1 & \text{for } t = 2, \\ 2 & \text{for } t = 3. \end{cases}$$

Lemma 3 *Let $\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3 \in \{0, 1\}$ be binary random variables, and W_1, W_2, W_3 be random variables (or vectors) such that $W_1 \to \widetilde{Y}_1 \to W_2 \to \widetilde{Y}_2 \to W_3 \to \widetilde{Y}_3$ is a Markov chain, conditional on the random vector (X, A).[13] Assume furthermore that $p_t(\widetilde{y}_t \,|\, w_t, x, \alpha) := \Pr\big(\widetilde{Y}_t = \widetilde{y}_t \,\big|\, W_t = w_t,\, X = x,\, A = \alpha\big)$ satisfies $0 < p_t(\widetilde{y}_t \,|\, w_t, x, \alpha) <$*

[13] The Markov chain assumption means that the density of $(W_1, \widetilde{Y}_1, W_2, \widetilde{Y}_2, W_3, \widetilde{Y}_3)$, conditional on (X, A), can be written as a product $f_{\widetilde{Y}_3|W_3,X,A}\, f_{W_3|\widetilde{Y}_2,X,A}\, f_{\widetilde{Y}_2|W_2,X,A}\, f_{W_2|\widetilde{Y}_1,X,A} f_{\widetilde{Y}_1|W_1,X,A}\, f_{W_1|X,A}$.

1, *for all* $\widetilde{y}_t$, w_t, x, α, *and* $t \in \{1,2,3\}$. *Then, for* $q \in \{0,1\}$, *the function*

$$\begin{aligned} m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x, \alpha) &:= -\mathbb{1}\{\widetilde{y}_1 = q\} \\ &+ \mathbb{1}\{\widetilde{y}_2 = q\} \exp\left(\frac{1}{2} \sum_{t=1}^{3} \left\{ \Lambda^{-1}\Big[p_{\delta(t)}\left(\widetilde{y}_t \,\middle|\, w_{\delta(t)}, x, \alpha\right) \Big] - \Lambda^{-1}\Big[p_t\left(\widetilde{y}_t \,\middle|\, w_t, x, \alpha\right) \Big] \right\} \right) \end{aligned}$$

satisfies $\mathbb{E}\left[m^{(q)}(W_1, \widetilde{Y}_1, W_2, \widetilde{Y}_2, W_3, \widetilde{Y}_3, X, A) \,\middle|\, W_1,\ X,\ A \right] = 0.$

The proof is given in Appendix A.1. Notice that the vector of conditioning variables (X, A) is only included in Lemma 3 to better connect the lemma to our panel AR(1) model, but for the mathematical result of the lemma this vector (X, A) is actually irrelevant (no assumptions are imposed on these conditioning variables, all probability statements are conditional on (X, A)), and the lemma may be easier read and understood by initially ignoring all occurrences of (X, A) and (x, α). Furthermore, when applying Lemma 3 to the $T = 3$ panel AR(1) model of Section 3, then we simply have $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3) = (Y_1, Y_2, Y_3)$ and $(W_1, W_2, W_3) = (Y_0, Y_1, Y_2)$, but the more general notation in the lemma is convenient when generalizing the results to models with $T > 3$.

The assumptions imposed in Lemma 3 are relatively weak. In particular, the outcomes $\widetilde{Y}_1$, $\widetilde{Y}_2$, $\widetilde{Y}_3$ are *not* assumed to be generated from a logit model. However, the result of Lemma 3 is in general equally weak, because the moment functions $m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x, \alpha)$ provided by the lemma still depend on the individual specific effects α, that is, the lemma in general does *not* deliver the type of moment conditions (5) that we are interested in this paper. We find it nevertheless useful to state the lemma in this weak form, because it provides some understanding for why the logit assumption is important here to obtain valid moment conditions that are free of the fixed effects.

For a binary choice model with single index $z_t(W_t, X) \in \mathbb{R}$ and additive fixed effects $A \in \mathbb{R}$ we have $\widetilde{Y}_t = \mathbb{1}\{z_t(W_t, X) + A + \varepsilon_t \geq 0\}$, for $t \in \{1,2,3\}$. If, in addition, we assume a logistic distribution for the random shock ε_t, then we obtain, for $\widetilde{y} \in \{0,1\}$,

$$p_t(\widetilde{y} \,|\, w_t, x, \alpha) = \Lambda\Big\{ (2\widetilde{y} - 1) \left[z_t(w_t, x) + \alpha \right] \Big\}, \tag{22}$$

which implies that for all $s, t \in \{1, 2, 3\}$,

$$\Lambda^{-1}\Big[p_s\left(\widetilde{y} \,\middle|\, w_s, x, \alpha\right)\Big] - \Lambda^{-1}\Big[p_t\left(\widetilde{y} \,\middle|\, w_t, x, \alpha\right)\Big] = (2\widetilde{y} - 1)\left[z_s(w_s, x) - z_t(w_t, x)\right]$$

does not depend on the fixed effects α. For this logistic specification with additive fixed effects we therefore find that the moment functions $m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x, \alpha)$ in Lemma 3 do not depend on the fixed effects α, and can be written as[14]

$$\begin{aligned} & m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x) \\ & \qquad = -\mathbb{1}\left\{\widetilde{y}_1 = q\right\} + \mathbb{1}\left\{\widetilde{y}_2 = q\right\} \exp\left\{\sum_{t=1}^{3} \widetilde{y}_t \left[z_{\delta(t)}(w_{\delta(t)}, x) - z_t(w_t, x)\right]\right\}. \end{aligned} \tag{23}$$

For the case $T = 3$, $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3) = (Y_1, Y_2, Y_3)$, $(W_1, W_2, W_3) = (Y_0, Y_1, Y_2)$, and $z_t(w_t, x) = y_{t-1}\,\gamma_0 + x_t'\,\beta_0$ it is easy to verify that $m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x)$ in (23) is equal to $m^{(q)}(y, y_0, x, \beta_0, \gamma_0)$ in display (10) above, that is, Lemma 3 delivers the moment functions derived for the $T = 3$ dynamic logit model in Section 3.2 as a special case.

6.3 Moment conditions for $T \geq 3$

We now discuss how the moment functions for $T = 3$ generalize to more than three time periods (after the initial y_0). We have already argued above that Lemma 3 is useful for our purposes for logit models of the form (22) where it delivers the moment functions in (23) that do not depend on the fixed effects. We now apply those results to the fixed effect logit AR(1) model with an arbitrary number of time periods $T \geq 3$ by setting $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3) = (Y_t, Y_s, Y_r)$ and $(W_1, W_2, W_3) = (Y_{t-1}, Y_{s-1}, Y_{r-1})$, for any triplet of time periods $t, s, r \in \{1, 2, \ldots, T\}$ that satisfy $t < s < r$. Notice that for this choice the Markov chain assumption in Lemma 3 is satisfied, that is, conditional on (X, A), $Y_{t-1} \to Y_t \to Y_{s-1} \to Y_s \to Y_{r-1} \to Y_r$ indeed constitutes a Markov chain according to model (2). Furthermore, in that model, the distribution of Y_t conditional on Y_{t-1}, X, A is indeed of the logistic form (22) with $z_t(w_t, x) = y_{t-1}\,\gamma_0 + x_t'\,\beta_0$ for all time periods t.

[14] Here we also use that $\sum_{t=1}^{3}\left[z_{\delta(t)}(w_{\delta(t)}, x) - z_t(w_t, x)\right] = 0$.

Making the unknown parameter dependence explicit, we now define the single index for time period t as $z_t(y, y_0, x, \beta, \gamma) = x_t' \beta + y_{t-1} \gamma$, and we also define the corresponding pairwise differences $z_{ts}(y, y_0, x, \beta, \gamma) = z_t(y, y_0, x, \beta, \gamma) - z_s(y, y_0, x, \beta, \gamma)$. Then, for triples of time periods $t, s, r \in \{1, 2, \ldots, T\}$ with $t < s < r$, the moment function in (23) can be written more explicitly as

$$m^{(0)(t,s,r)}(y, y_0, x, \beta, \gamma) = \begin{cases} \exp\left[z_{sr}(y, y_0, x, \beta, \gamma)\right] - 1 & \text{if } (y_t, y_s, y_r) = (0, 0, 1), \\ -1 & \text{if } (y_t, y_s) = (0, 1), \\ \exp\left[z_{rt}(y, y_0, x, \beta, \gamma)\right] & \text{if } (y_t, y_s, y_r) = (1, 0, 0), \\ \exp\left[z_{st}(y, y_0, x, \beta, \gamma)\right] & \text{if } (y_t, y_s, y_r) = (1, 0, 1), \\ 0 & \text{otherwise,} \end{cases}$$

$$m^{(1)(t,s,r)}(y, y_0, x, \beta, \gamma) = \begin{cases} \exp\left[z_{ts}(y, y_0, x, \beta, \gamma)\right] & \text{if } (y_t, y_s, y_r) = (0, 1, 0), \\ \exp\left[z_{tr}(y, y_0, x, \beta, \gamma)\right] & \text{if } (y_t, y_s, y_r) = (0, 1, 1), \\ -1 & \text{if } (y_t, y_s) = (1, 0), \\ \exp\left[z_{rs}(y, y_0, x, \beta, \gamma)\right] - 1 & \text{if } (y_t, y_s, y_r) = (1, 1, 0), \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

For $T = 3$ and $(t, s, r) = (1, 2, 3)$, these moment functions are exactly those calculated in Section 3.2 above. For general $T \geq 3$ and triplets (t, s, r) we can apply Lemma 3, conditional also on $Y_0, Y_1, \ldots, Y_{t-1}$, to obtain the following theorem.

Theorem 2 *If the outcomes Y are generated from the panel logit AR(1) model with $T \geq 3$ and true parameters β_0 and γ_0, then we have for all $t, s, r \in \{1, 2, \ldots, T\}$ with $t < s < r$, and for all $q \in \{0, 1\}$, $y^{(t)} \in \{0, 1\}^t$, $x \in \mathbb{R}^{K \times T}$, $\alpha \in \mathbb{R}$, that*

$$\mathbb{E}\left[m^{(q)(t,s,r)}(Y, Y_0, X, \beta_0, \gamma_0) \,\middle|\, (Y_0, Y_1, \ldots, Y_{t-1}) = y^{(t)}, \, X = x, \, A = \alpha\right] = 0.$$

The proof is given in Appendix A.1, but as argued above, the theorem really is an immediate corollary of Lemma 3. Instead of conditioning on $Y_1, \ldots, Y_{t-1}$, we can also multiply the moment function with an arbitrary function of $Y_1, \ldots, Y_{t-1}$. Namely, by applying Theorem 2 and the law of iterated expectations, we find, for any function

$w : \{0,1\}^{t-1} \to \mathbb{R}$, that

$$\mathbb{E}\left[w(Y_1,\ldots,Y_{t-1})\, m^{(q)(t,s,r)}(Y,Y_0,X,\beta_0,\gamma_0) \,\middle|\, Y_0 = y_0,\, X = x,\, A = \alpha \right] = 0. \tag{25}$$

From Section 3.4 we know that, for any fixed value of the initial condition y_0, there are at least $\ell = 2^T - 2T$ linearly independent moment conditions available for our AR(1) logit model with T time periods. It turns out for $\gamma \neq 0$ this is exactly the correct number of linearly independent moment conditions in this model. In a previous version of the current paper (Honoré and Weidner 2020) we conjectured this, and subsequent papers by Kruiniger (2020) and Dobronyi, Gu, and Kim (2021) have shown that this is indeed the case.

Equation (25) provides all of the $\ell = 2^T - 2T$ available valid moment functions for this model, but not all those moment functions $w(Y_1,\ldots,Y_{t-1})\, m^{(q)(t,s,r)}(Y,Y_0,X,\beta_0,\gamma)$ are linearly independent, that is, some of them can be written as linear combinations (with coefficients that depend on x, β, γ) of the others. However, if we restrict ourselves to $r = T$, then we have verified numerically that a linearly independent basis is obtained. Notice that once we fix $r = T$, then, for given y_0, we can still choose $q \in \{0,1\}$, $w : \{0,1\}^{t-1} \to \mathbb{R}$, and (t,s), with $1 \leq t < s < T$. The total number of basis elements is therefore equal to

$$\ell = \underbrace{2}_{q\,\in\,\{0,1\}} \times \underbrace{\sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1}}_{\text{allowed } t,s \text{ values}} \underbrace{2^{t-1}}_{\substack{\text{number of linearly} \\ \text{independent functions} \\ w(y_1,\ldots,y_{t-1})}} = \sum_{t=1}^{T-2} 2^t \,(T-t-1) = 2^T - 2T,$$

as claimed above.[15]

[15]We consider $\gamma \neq 0$ here. For $\gamma = 0$ we have a static panel logit model, and in that case $T-1$ additional moment conditions become available, bringing the total number of available moments to $2^T - T - 1$. The first-order conditions of the conditional likelihood in Rasch (1960b) and Andersen (1970) are linear combinations of these moment functions.

6.3.1 Unbalanced panels and missing time periods

The only regressor and outcome values that enter into the moment functions $m^{(q)(t,s,r)}$ are (x_t, x_s, x_r) and $(y_{t-1}, y_t, y_{s-1}, y_s, y_{r-1}, y_t)$.[16] Thus, as long as those variables are observed we can evaluate $m^{(q)(t,s,r)}$. The moment conditions for $T > 3$ can therefore also be applied to unbalanced panels where regressors and outcomes are not observed in all time periods, provided that the occurrence of missing values is independent of the outcomes Y, conditional on the regressors X and the individual-specific effects A. The data in our empirical illustration are indeed unbalanced, and in Section 7 we discuss how to combine the moment functions for unbalanced panels.

6.3.2 Relation to Kitazawa (2013, 2016)

The first paper to obtain moment conditions for the dynamic panel logit model without imposing restrictions on the covariate values is the working paper by Kitazawa (2013), which was recently published (Kitazawa 2022). That paper defines

$$
\begin{aligned}
U_t &= y_t + (1-y_t)y_{t+1} - (1-y_t)y_{t+1}\exp(-\beta\Delta x_{t+1}) - \delta y_{t-1}(1-y_t)y_{t+1}\exp(-\beta\Delta x_{t+1}),\\
\hbar U_t &= U_t - y_{t-1} - \tanh\left[\frac{-\gamma y_{t-2} + \beta(\Delta x_t + \Delta x_{t+1})}{2}\right](U_t + y_{t-1} - 2U_t y_{t-1}),\\
\Upsilon_t &= y_t y_{t+1} + y_t(1-y_{t+1})\exp(\beta\Delta x_{t+1}) + \delta(1-y_{t-1})y_t(1-y_{t+1})\exp(\beta\Delta x_{t+1}),\\
\hbar\Upsilon_t &= \Upsilon_t - y_{t-1} - \tanh\left[\frac{\gamma(1-y_{t-2}) + \beta(\Delta x_t + \Delta x_{t+1})}{2}\right](\Upsilon_t + y_{t-1} - 2\Upsilon_t y_{t-1}),
\end{aligned}
\tag{26}
$$

where $\delta = e^{\gamma} - 1$ and $\Delta x_t = x_t - x_{t-1}$. The paper then shows that, for $t \in \{2, \ldots, T-1\}$,[17] the functions $\hbar U_t$ and $\hbar\Upsilon_t$ are valid moment functions, in the sense of (5). Kitazawa (2016) uses the same moment conditions, but also includes time dummies in the model, which in our notation are included in the parameter vector β (one just needs to define the regressors x_t as appropriate dummy variables).

Those definitions look quite different to our moment functions above, but one can

[16] Of course, y_t coincides with y_{s-1} if $t = s-1$, and y_s coincides with y_{r-1} if $s = r-1$.

[17] This is written here in our conventions for t and T.

show that

$$\hbar U_2 = \left\{ \tanh\left[\frac{-\gamma y_0 + \beta(\Delta x_2 + \Delta x_3)}{2}\right] - 1 \right\} m^{(0)}(y, y_0, x, \beta, \gamma),$$
$$\hbar \Upsilon_2 = \left\{ \tanh\left[\frac{\gamma(1 - y_0) + \beta(\Delta x_2 + \Delta x_3)}{2}\right] + 1 \right\} m^{(1)}(y, y_0, x, \beta, \gamma).$$

Thus, apart from a rescaling (with a non-zero function of the parameters and conditioning variables), the moment functions of Kitazawa (2013) coincide with our moment functions for AR(1) models with $T = 3$. However, the complete set of moment conditions for $T > 3$ in Theorem 2 is new.

6.3.3 Relation to other existing results

Honoré and Kyriazidou (2000) observe that with (2) and $T = 3$, the conditional likelihood function that conditions on $Y = y_0$, $Y_3 = y_3$, and $Y_1 + Y_2 = 1$,

$$\ell_{y_0,y_3}(y, x, \beta, \gamma) = \Pr\left(Y = y \,\middle|\, Y_0 = y_0,\, Y_1 + Y_2 = 1,\, Y_3 = y_3,\, X = x, \beta, \gamma\right),$$

does not depend on α, when $x = (x_1, x_2, x_2)$ (so the explanatory variables are the same in the last two periods). The corresponding scores are

$$\begin{aligned}
\frac{\partial \ell_{0,0}(y, x, \beta, \gamma)}{\partial \gamma} &= 0, \\
\frac{\partial \ell_{0,0}(y, x, \beta, \gamma)}{\partial \beta} &= \frac{x_{12}}{1 + \exp\left(x_{12}'\beta\right)} \left[\frac{m^{(1)}(y, 0, x, \beta, \gamma) + \exp\left(x_{12}'\beta - \gamma\right)\, m^{(0)}(y, 0, x, \beta, \gamma)}{\exp\left(-\gamma\right) - 1}\right],
\end{aligned} \tag{27}$$

and

$$\begin{pmatrix} \dfrac{\partial \ell_{0,1}(y, x, \beta, \gamma)}{\partial \gamma} \\ \dfrac{\partial \ell_{0,1}(y, x, \beta, \gamma)}{\partial \beta} \end{pmatrix} = \begin{pmatrix} -1 \\ x_{12} \end{pmatrix} \frac{1}{1 + \exp\left(x_{12}'\beta - \gamma\right)}$$

$$\times \left[\frac{m^{(1)}(y,0,x,\beta,\gamma) + \exp\left(x_{12}'\beta\right) \, m^{(0)}(y,0,x,\beta,\gamma)}{\exp(\gamma)-1} \right], \tag{28}$$

where $m^{(0)}$ and $m^{(1)}$ are defined in (10). The results for $y_0 = 1$ are analogous. Thus, the score functions of the conditional likelihood in Honoré and Kyriazidou (2000) are linear combinations of our moment conditions when $x_2 = x_3$. The conditional likelihood estimation discussed in Cox (1958) and Chamberlain (1985) are special cases of this without regressors ($x_1 = x_2 = x_3 = 0$).

Hahn (2001) considers model (2) with $T = 3$, initial condition $y_0 = 0$, and time dummies as regressors, that is, $x_t'\beta = \beta_t$, with the normalization $\beta_1 = 0$. The common parameters in that model are $(\beta_2, \beta_3, \gamma)$. Hahn shows that these parameters cannot be estimated at root-n-rate. This is not in conflict with our results here, because Lemma 1 only provides two moment conditions for $y_0 = 0$. However, there are three model parameters in the setup of Hahn (2001), so just from counting parameters and moment conditions, we know that our moments cannot identify $(\beta_2, \beta_3, \gamma)$.[18] Thus, our moment conditions cannot be used to estimate the parameters $(\beta_2, \beta_3, \gamma)$ at root-n-rate. This is in agreement with Hahn's calculation of the information bound for this model.

The main reason why we can identify and estimate β and γ is that we consider non-constant regressors $X = (X_1, X_2, X_3)$, which gives us two moment conditions for each initial condition and each support point of the regressors, and thus many more moment conditions than parameters — see our formal results on point-identification of β and γ in Section 5 above.

6.4 More general dynamic panel models

As explained in Section 6.2, Lemma 3 delivers valid moment functions for any model with logistic conditional probabilities of the form (22). This means that the single index of the model need not be of the form $x_t' \beta + y_{t-1} \gamma$ that is linear in x_t and y_{t-1}, but it can

[18] Including moment conditions that use the initial condition $y_0 = 1$ will give two additional moments. From the point of view of counting moments, this will result in a model which is over-identified.

actually be any function of the strictly exogenous regressors, lagged depend variable, and parameters. In particular, if we replace the model specification (2) by

$$\Pr\left(Y_{it}=1 \,\middle|\, Y_i^{t-1}, X_i, A_i\right) = \frac{\exp\left[(1-Y_{i,t-1})X_{it}'\,\beta_0 + Y_{i,t-1}X_{it}'\,\beta_1 + Y_{i,t-1}\,\gamma + A_i\right)}{1+\exp\left((1-Y_{i,t-1})X_{it}'\,\beta_0 + Y_{i,t-1}X_{it}'\,\beta_1 + Y_{i,t-1}\,\gamma + A_i\right]}, \tag{29}$$

then the moment functions (24) and Theorem 2 remain fully valid, as long as we replace the parameters (β,γ) by (β_0,β_1,γ), and define the single index by $z_t(y,y_0,x,\beta_0,\beta_1,,\gamma) = (1-y_{t-1})\,x_t'\,\beta_0 + y_{t-1}\,x_t'\,\beta_1 + y_{t-1}\,\gamma$.

The generalized model (29) is interesting, because it allows the effect of the regressors X_{it} on Y_{it} to depend on the current "state" of the process, $Y_{i,t-1}$, with $\beta_{0/1}$ measuring the effect of X_{it} on Y_{it} if $Y_{i,t-1} = 0/1$. We do not consider this more general model structure further in this paper, but it is noteworthy that the regressors $(1-Y_{i,t-1})X_{it}$ and $Y_{i,t-1}X_{it}$ are pre-determined regressors that are more general than just the lagged dependent variable $Y_{i,t-1}$ we have considered so far. Further comments on more general regressors structures are given in Appendix A.2.

Another interesting generalization of the model that still allows for the construction of moment conditions, is to make the AR(1) coefficient γ individual specific. Equation (2) then reads

$$\Pr\left(Y_{it}=1 \,\middle|\, Y_i^{t-1}, X_i, A_i, C_i\right) = \frac{\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,C_i + A_i\right)}{1+\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,C_i + A_i\right)}. \tag{30}$$

We can then treat (C_i, A_i) as a two-dimensional fixed-effect and employ the methods of Section 2 and 3 to explore moment conditions for β that are free of (C_i, A_i). For general covariate and parameter values, we find that no such moment conditions exist for $T=3$, but they do exist for $T \geq 4$. For example, for $T=4$ a valid moment condition in this model (i.e. satisfying $\mathbb{E}\left[m(Y_i, Y_{0,i}, X_i, \beta_0) \middle| Y_{0,i}, X_i, C_i, A_i\right] = 0$) is

$$m(y, y_0, x, \beta) = -\mathbb{1}\Big\{(y_1, y_3, y_4) = (1,0,0)\Big\} + \begin{cases} \exp\left(x_{12}'\beta\right) & \text{if } y = (0,1,0,0), \\ \exp\left(x_{14}'\beta\right) & \text{if } y = (0,1,0,1), \\ -\exp\left(x_{34}'\beta\right) & \text{if } y = (1,0,0,1), \\ \exp\left(x_{32}'\beta\right) & \text{if } y = (1,1,0,0), \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

We have found numerically that there is one additional moment condition for $T = 4$, a total of ten for $T = 5$, and thirty-two for $T = 6$. The moment function in the last display was derived using the ideas described in Section 3. Alternatively, one could also obtain those moment functions by finding a linear combination of the moment functions in equation (25) (where the γ parameter appears as a common parameter) such that the linear combination is free of γ. Under appropriate support assumptions on the covariates X_i, the identification of β in (30) follows from the moment function in (31) by arguments analogous to those in Section 5.

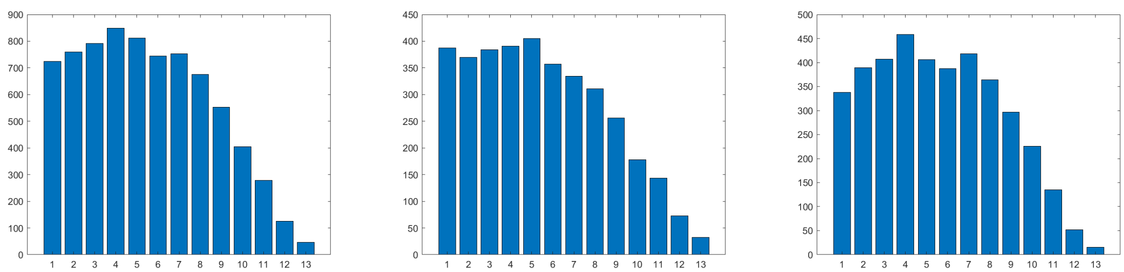
7 Empirical illustration

In this section, we illustrate how to use the conditional moment functions in this paper to implement a GMM approach to estimation. We use data from the National Longitudinal Survey of Youth 1997[19] (NLSY97) covering the years 1997 to 2010, and the dependent variable is a binary variable indicating employment status by whether the respondent reported working $\geq$1000 hours in the past year. We estimate fixed effects logit AR(1) and AR(2) models using the number of biological children the respondent

[19]The analysis is restricted to the years in which the survey was conducted annually, from 1997-2011. For years in which the respondent was not interviewed, all time-varying variables (e.g., employment status, school enrollment status, age, income, marital variables, etc) are marked as missing. Otherwise, unless the raw data was marked as missing in some capacity (e.g., due to non-response, the interviewee not knowing the answer to the question), no other entries had missings imposed upon them.

has (Children), a dummy variable for being married (Married), a transformation[20] of the spouse's income (Sp.Inc.), and a full set of time dummies as the explanatory variables. There are a total of 8,274 individuals aged 16 to 32, resulting in 54,166 observations. For the estimation, we consider the full sample, as well as females and males separately. Figure 1 displays the number of observations, T_i, per individual in each of the three samples.

Figure 1: Histogram of Number of Observations Per Individual.



(a) All (b) Females (c) Males

The moment conditions for the fixed effects logit AR(1) in (10) and (24) are all indexed by three time-periods, and they are conditional on the strictly exogenous variables and the initial conditions. One could in principle construct seperate conditional moment conditions for each value of T_i and each triplet $1 \leq t < s < r \leq T_i$, and then use them to construct efficient unconditional moment functions. See, for example, the discussion in Newey and McFadden (1994). Unfortunately, the construction of these moment functions depends on the conditional expectation of the derivative of the conditional moment function as well as on the conditional variance of the conditional moment function. We therefore pursue a different approch to obtaing unconditional moment functions. We do not claim that the resulting GMM estimator has any optimality properties, but we have found that it performs well in our Monte Carlo simulations even for relatively small sample sizes, see Appendix B.1.

We first normalize all moment functions such that $\sup_{y,x,\beta,\gamma} |\widetilde{m}(y, x, \beta, \gamma)| < \infty$. For example, the rescaled versions of our $T = 3$ moment functions in Section 3.2 are given

[20] The spouse's income can be zero or negative. This prevents us from using the logarithm of the income as an explanatory variable. We therefore use the signed fourth root.

by

$$
\begin{aligned}
\widetilde{m}^{(0)}(y,y_0,x,\beta,\gamma) &= \frac{m^{(0)}(y,y_0,x,\beta,\gamma)}{1+\exp\left(x_{23}'\beta\right)+\exp\left(x_{31}'\beta-y_0\,\gamma\right)+\exp\left(x_{21}'\beta+(1-y_0)\,\gamma\right)} \\
\widetilde{m}^{(1)}(y,y_0,x,\beta,\gamma) &= \frac{m^{(2)}(y,y_0,x,\beta,\gamma)}{1+\exp\left(x_{12}'\beta+y_0\,\gamma\right)+\exp\left(x_{13}'\beta-(1-y_0)\,\gamma\right)+\exp\left(x_{32}'\beta\right)}
\end{aligned}
$$

Here, each moment function is divided by the sum of the absolute values of all the different positive summands that appear in that moment function. We have found that this rescaling improves the performance of the resulting GMM estimators, particularly for small samples, because it bounds the moment functions and its gradients uniformly over the parameters β and γ. Interestingly, the score functions of the conditional likelihood in Honoré and Kyriazidou (2000) are essentially rescaled in this way.

The rescaled moment functions are valid conditional on any realization of the regressors. We can therefore form unconditional moment functions by multiplying them with arbitrary functions of the regressors and the initial conditions. In our example, we multiply them by 1, the initial condition, and the explanatory variables for the three time periods that index the moment function. For example, for $T=3$, we use

$$
M(y,y_0,x,\beta,\gamma) = \left(1,\ y_0,\ x_1',\ x_2',\ x_3'\right)' \otimes \begin{pmatrix} \widetilde{m}^{(1)}(y,y_0,x,\beta,\gamma) \\ \widetilde{m}^{(0)}(y,y_0,x,\beta,\gamma) \end{pmatrix},
$$

where $\otimes$ is the tensor product.

One could in principle construct a moment function for each (t,s,r) which indexes a moment function. However, this would create a very large number of moment conditions. For a given individual, we therefore add up all the moment functions over all triplets, $t<s<r$. Observations with $T=T_i$ time periods will then contribute $\binom{T_i}{3}$ terms to the sample analog of the moment. This gives very large weight to observations with large T_i. We therefore weigh the triplets (t,s,r) for an observation with T_i time periods by $(T_i-1)/\binom{T_i}{3}$. This yields sample moments of the form

$\frac{1}{n}\sum_{i=1}^{n} M(Y_i, Y_{i,0}, X_i, \beta, \gamma)$, and the corresponding GMM estimator is given by

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \underset{\beta \in \mathbb{R}^K, \gamma \in \mathbb{R}}{\operatorname{argmin}} \left(\sum_{i=1}^{n} M(Y_i, Y_{i,0}, X_i, \beta, \gamma) \right)' W \left(\sum_{i=1}^{n} M(Y_i, Y_{i,0}, X_i, \beta, \gamma) \right),$$

where W is a symmetric positive-definite weight matrix. We use a diagonal weight matrix with the inverse of the moment variances on the diagonal. The motivation stems from Altonji and Segal (1996) who demonstrate that estimating the optimal weighting matrix can result in poor finite sample performance of GMM estimators. They suggest equally weighted moments (i.e., $W = I$) as an alternative. Of course, using equal weights will not be invariant to changes in units, which explains the practice we have adopted.[21]

Under standard regularity conditions we have

$$\sqrt{n} \left[\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} \right] \Rightarrow \mathcal{N}\left(0,\ (G'WG)^{-1}\, G'W\,\Omega\, WG\, (G'WG)^{-1}\right),$$

with $\Omega = \mathrm{Var}[m(Y_{i,0}, Y_i, X_i, \beta_0, \gamma_0)]$ and $G = \mathbb{E}\left[\frac{\partial m(Y_{i,0},Y_i,X_i,\beta_0,\gamma_0)}{\partial \beta'}, \frac{\partial m(Y_{i,0},Y_i,X_i,\beta_0,\gamma_0)}{\partial \gamma'}\right]$.

Table 1 reports the estimation results. As expected, and consistent with the Monte Carlo results in Appendix Section B.1, the standard logit maximum likelihood estimator of the coefficient on the lagged dependent variable is much larger than the one that estimates a fixed effect for each individual: the estimated fixed effects will be "'over-fitted", leading to a downward bias in the estimated state dependence. Moreover, the standard logit estimator that ignores fixed effects will capture the presence of persistent heterogeneity by the lagged dependent variable, leading to an upwards bias if such heterogeneity is present in the data. The GMM estimator gives a much smaller coefficient than the standard logit maximum likelihood estimator, suggesting that heterogeneity plays a big role in this application.

To estimate the AR(2) version of the model, we apply the moment conditions pro-

[21] Our choice of weight matrix is quite common in empirical work. See, for example, Gayle and Shephard (2019) for a recent example.

Table 1: Empirical Results (AR(1)).

	Females			Males			All		
	Logit	Logit w FE	GMM	Logit	Logit w FE	GMM	Logit	Logit w FE	GMM
Lagged y	2.585	0.780	1.512	2.947	0.709	1.454	2.797	0.768	1.417
	(0.038)	(0.050)	(0.076)	(0.040)	(0.063)	(0.088)	(0.027)	(0.039)	(0.060)
Children	−0.335	−0.444	−0.244	−0.153	0.018	−0.275	−0.278	−0.252	−0.214
	(0.016)	(0.052)	(0.196)	(0.021)	(0.067)	(0.133)	(0.012)	(0.043)	(0.102)
Married	0.082	−0.044	0.637	0.335	0.332	0.038	0.349	0.173	0.707
	(0.084)	(0.159)	(0.890)	(0.071)	(0.171)	(0.295)	(0.053)	(0.111)	(0.397)
SP.Inc.	−0.010	−0.050	−0.104	0.033	0.003	0.019	−0.017	−0.044	−0.089
	(0.006)	(0.011)	(0.068)	(0.007)	(0.016)	(0.026)	(0.004)	(0.009)	(0.033)

The estimation also includes 12 time dummies. Standard error for the GMM and Logit Fixed Effects Estimators are calculated as the interquartile range of 1,000 bootstrap replications divided by 1.35.

vided in Appendix A.3.1 to all consecutive sequences of six outcomes (treating the first two as initial conditions). The moment functions are scaled as described in the Monte Carlo simulations in Appendix Section B.1. The results are presented in Table 2. The most interesting finding is that for all three samples, the GMM estimator of (γ_1, γ_2) is between the maximum likelihood estimator that ignores the fixed effects, and the one that estimates a fixed effect for each individual. This suggests that unobserved individual-specific heterogeneity is important in this example. Economically, it is also interesting that for each estimation method, the estimates of (γ_1, γ_2) are quite similar across the three samples.

8 Conclusion

Bonhomme (2012) proposed a general approach for constructing moment restrictions in nonlinear panel data models that do not depend on individual specific effects. In this paper, we have operationalized this in models with discrete outcomes by first presenting

Table 2: Empirical Results (AR(2)).

	Females			Males			All		
	Logit	Logit w FE	GMM	Logit	Logit w FE	GMM	Logit	Logit w FE	GMM
y_{t-1}	2.259	0.742	1.356	2.422	0.514	1.116	2.361	0.665	1.297
	(0.047)	(0.069)	(0.162)	(0.052)	(0.083)	(0.131)	(0.035)	(0.053)	(0.092)
y_{t-2}	0.917	−0.379	0.678	1.332	−0.286	0.558	1.137	−0.319	0.648
	(0.048)	(0.072)	(0.081)	(0.053)	(0.080)	(0.066)	(0.036)	(0.054)	(0.046)
Children	−0.260	−0.410	−1.926	−0.143	0.112	−0.188	−0.223	−0.192	−1.209
	(0.018)	(0.069)	(0.282)	(0.024)	(0.100)	(0.251)	(0.014)	(0.051)	(0.219)
Married	0.136	0.022	−0.193	0.411	0.534	−0.019	0.393	0.269	−0.121
	(0.095)	(0.184)	(0.028)	(0.083)	(0.203)	(0.025)	(0.061)	(0.145)	(0.022)
Sp.Inc	−0.015	−0.050	0.255	0.023	−0.008	0.179	−0.022	−0.046	0.093
	(0.007)	(0.014)	(0.210)	(0.008)	(0.019)	(0.316)	(0.005)	(0.011)	(0.183)

The estimation also includes 11 time dummies. Standard error for the GMM and Logit Fixed Effects Estimators are calculated as the interquartile range of 1,000 bootstrap replications divided by 1.35.

a blueprint for deciding whether such moment conditions exist, and then an approach for actually finding analytic expressions for the moment conditions.

We have used our approach to derive all the moment conditions for the panel logit AR(1) model that are free of the fixed effects, and we have employed those moment conditions to show identification of the common model parameters and to obtain a GMM estimator that is useful and performs well in practice. The immediate practical relevance of this paper is therefore for the dynamic panel logit model (both AR(1) and AR(2) models are estimated in an empirical application).

While part of this paper emphasises binary logit models, the methods explained in Section 2 and 3 for exploring and deriving moment conditions are applicable for more general panel models, as illustrated by the examples provided in Section 4. Exploring such moment conditions in other interesting models is a research agenda that has only started (e.g. Honoré, Muris, and Weidner 2021, Davezies, D'Haultfoeuille, and Mugnier 2022), and a lot more future work should be done to provide useful new estimation methods in various discrete choice panel models.

References

Abrevaya, J. (1999): "Leapfrog Estimation of a Fixed-Effects Model with Unknown Transformation of the Dependent Variable," Journal of Econometrics, 93, 203–228.

——— (2000): "Rank estimation of a generalized fixed-effects regression model," Journal of Econometrics, 95(1), 1 – 23.

Abrevaya, J., and C. Muris (2020): "Interval censored regression with fixed effects," Journal of Applied Econometrics, 35(2), 198–216.

Al-Sadoon, M. M., T. Li, and M. H. Pesaran (2017): "Exponential class of dynamic binary choice panel data models with fixed effects," Econometric Reviews, 36(6-9), 898–927.

Altonji, J. G., and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," Journal of Business & Economic Statistics, 14(3), 353–366.

Andersen, E. B. (1970): "Asymptotic properties of conditional maximum-likelihood estimators," Journal of the Royal Statistical Society: Series B (Methodological), 32(2), 283–301.

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

Arellano, M., and S. Bonhomme (2009): "Robust Priors in Nonlinear Panel Data Models," Econometrica, 77(2), 489–536.

——— (2011): "Nonlinear Panel Data Analysis," Annual Review of Economics, 3(1), 395–424.

Aristodemou, E. (2018): "Semiparametric Identification in Panel Data Discrete Response Models," Tinbergen Institute Discussion Paper 2018-065/III.

Bartolucci, F., and V. Nigro (2010): "A Dynamic Model for Binary Panel Data With Unobserved Heterogeneity Admitting a $\sqrt{n}$-Consistent Conditional Estimator," Econometrica, 78(2), 719–733.

Blundell, R., R. Griffith, and F. Windmeijer (1997): "Individual Effects and Dynamics in Count Data Models," University College, London.

——— (2002): "Individual effects and dynamics in count data models," Journal of econometrics, 108(1), 113–131.

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

Bonhomme, S., K. Dano, and B. Graham (2022): "Sequential Moment Restrictions in Nonlinear Panel Models with Fixed Effects," Working paper.

Bonhomme, S., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," Econometrica, 83(3), 1147–1184.

Botosaru, I., and C. Muris (2018): "Binarization for Panel Models with Fixed Effects," University of Bristol.

Chamberlain, G. (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.

——— (2010): "Binary response models for panel data: Identification and information," Econometrica, 78(1), 159–168.

Cox, D. R. (1958): "The regression analysis of binary sequences," Journal of the Royal

Statistical Society: Series B (Methodological), 20(2), 215–232.

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

Dhaene, G., and K. Jochmans (2015): "Split-panel Jackknife Estimation of Fixed-effect Models," The Review of Economic Studies, 82(3), 991–1030.

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., and M. Weidner (2016): "Individual and time effects in non-linear panel models with large N, T," Journal of Econometrics, 192(1), 291 – 312.

Gayle, G.-L., and A. Shephard (2019): "Optimal Taxation, Marriage, Home Production, and Family Labor Supply," Econometrica, 87(1), 291–326.

Hahn, J. (2001): "The Information Bound of a Dynamic Panel Logit Model With Fixed Effects," Econometric Theory, 17(5), 913–932.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Econometrica, 72(4), 1295–1319.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," Econometrica, 52(4), 909–38.

Heckman, J. J. (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Tests of the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence," Annales de l'INSEE, pp. 227–269.

——— (1981a): "Heterogeneity and State Dependence," Studies in Labor Markets, S. Rosen (ed).

——— (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," Structural Analysis of Discrete Panel Data with Econometric Applications, C. F. Manski and D. McFadden (eds), pp. 179–195.

——— (1981c): "Statistical Models for Discrete Panel Data," Structural Analysis of Discrete Panel Data with Econometric Applications, C. F. Manski and D. McFadden (eds), pp. 114–178.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," Econometrica, 60, 533–565.

——— (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," Journal of Econometrics, 59(1-2), 35–61.

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 E. Kyriazidou (2019): "Identification in Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?," Annals of Economics and Statistics, (134), 207–226.

Honoré, B. E., C. Muris, and M. Weidner (2021): "Dynamic Ordered Panel Logit Models," arXiv preprint arXiv:2107.03253.

Honoré, B. E., and M. Weidner (2020): "Moment conditions for dynamic panel logit models with fixed effects," arXiv preprint arXiv:2005.05942v2.

——— (2022): "Dynamic Multinomial Panel Logit Models," In Progress.

Hu, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," Econometrica, 70(6), pp. 2499–2517.

Jochmans, K., and T. Magnac (2017): "A note on sufficiency in binary panel models," The Econometrics Journal, 20(2), 259–269.

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

Khan, S., M. Ponomareva, and E. Tamer (2019): "Identification of Dynamic Panel Binary Response Models," Discussion paper, Boston College Department of Economics.

Kitazawa, Y. (2013): "Exploration of dynamic fixed effects logit models from a traditional angle," Discussion paper, No. 60, Kyushu Sangyo University Faculty of Economics.

——— (2016): "Root-N consistent estimations of time dummies for the dynamic fixed effects logit models: Monte Carlo illustrations," Discussion paper, No. 72, Kyushu Sangyo University Faculty of Economics.

——— (2022): "Transformations and moment conditions for dynamic fixed effects logit models," Journal of Econometrics, 229(2), 350–362.

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

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," Econometrica, 65, 1335–1364.

——— (2001): "Estimation of Dynamic Panel Data Sample Selection Models," Review of Economic Studies, 68, 543–572.

Lancaster, T. (2002): "Orthogonal Parameters and Panel Data," Review of Economic Studies, 69, 647–66.

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.

——— (2004): "Panel binary variables and sufficiency: generalizing conditional logit," Econometrica, 72(6), 1859–1876.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," Econometrica, 55, 357–62.

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

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: Journal of the Econometric Society, pp. 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.

Rasch, G. (1960a): Probabilistic Models for Some Intelligence and Attainment Tests. Denmarks Pædagogiske Institut, Copenhagen.

Rasch, G. (1960b): Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.

A Appendix

A.1 Proofs of main text results

Lemma 1 is a special case of Theorem 2, which is proven below. Alternatively, Lemma 1 can be proved more directly by "brute-force calculations", see Section B.2.1 of the supplementary appendix.

Proof of Lemma 2. The lemma holds trivially for $K = 0$ when $s = \emptyset$ and $g_\emptyset : \mathbb{R} \to \mathbb{R}$ is a single increasing function, implying that $g_\emptyset(\gamma) = 0$ can at most have one solution. Consider $K \geq 1$ in the following. We follow a proof by contradiction. Assume that $(\beta_1, \gamma_1) \in \mathbb{R}^K \times \mathbb{R}$ and $(\beta_2, \gamma_2) \in \mathbb{R}^K \times \mathbb{R}$ both solve $g_s(\beta_1, \gamma_1) = 0$ and $g_s(\beta_2, \gamma_2) = 0$, for all $s \in \{-, +\}^K$, with $(\beta_1, \gamma_1) \neq (\beta_2, \gamma_2)$. Our goal is to derive a contradiction between this and the assumptions of the lemma. Without loss of generality, we assume that $\gamma_1 \leq \gamma_2$. Define $s^* \in \{-, +\}^K$ by

$$s_k^* = \begin{cases} + & \text{if } \beta_{1,k} \leq \beta_{2,k}, \\ - & \text{otherwise,} \end{cases}$$

for all $k \in \{1, \ldots, K\}$. By the monotonicity assumptions on $g_s(\beta, \gamma)$ in the lemma, we have that $g_{s^*}(\beta, \gamma)$ is strictly increasing in γ and we have $\gamma_1 \leq \gamma_2$; if $s_k^* = +$, then $g_{s^*}(\beta, \gamma)$ is strictly increasing in β_k and we have $\beta_{1,k} \leq \beta_{2,k}$; and if $s_k^* = -$, then $g_{s^*}(\beta, \gamma)$ is strictly decreasing in β_k and we have $\beta_{1,k} > \beta_{2,k}$. Furthermore, one of these inequalities on the parameters must be strict, because we have $(\beta_1, \gamma_1) \neq (\beta_2, \gamma_2)$. We therefore conclude that

$$g_{s^*}(\beta_1, \gamma_1) < g_{s^*}(\beta_2, \gamma_2).$$

This violates the assumption that $g_s(\beta_1, \gamma_1) = 0$ and $g_s(\beta_2, \gamma_2) = 0$. Thus, under the assumptions of the lemma there cannot be two solutions of the system (17). ∎

Proof of Theorem 1. Consider $y_0 = 0$ and $q = 0$. Using the definition of the moment function $m^{(0)}(y, y_0, x, \beta, \gamma)$ in Section 3.2 and the distribution of $Y|X, A$ implied by model (2) we find

$$\begin{aligned}
\frac{\partial\, \overline{m}^{(0)}_{0,s}(\beta,\gamma)}{\partial \gamma} &= \mathbb{E}\left[\frac{\partial\, m^{(0)}(Y,0,X,\beta,\gamma)}{\partial \gamma} \,\middle|\, Y_0 = 0,\ X \in \mathcal{X}_s \right] \\
&= \mathbb{E}\left[\frac{\partial\, \exp\left(\gamma + X_{21}'\beta\right)}{\partial \gamma} \operatorname{Pr}\left[Y = (1,0,1) \,\middle|\, Y_0 = 0, X, A \right] \,\middle|\, Y_0 = 0,\ X \in \mathcal{X}_s \right] \\
&= \mathbb{E}\left[\exp\left(\gamma + X_{21}'\beta\right) p((1,0,1),0,X,\beta_0,\gamma_0,A) \,\middle|\, Y_0 = 0,\ X \in \mathcal{X}_s \right] > 0,
\end{aligned}$$

where in the last step (to conclude that the expression is positive) we used that $p((1,0,1),0,x,\beta_0,\gamma_0,\alpha) > 0$ for all $x \in \mathbb{R}^{K\times 3}$ and $\alpha \in \mathbb{R}$.[22] We have thus shown that $\overline{m}^{(0)}_{0,s}(\beta,\gamma)$ is strictly increasing in γ. Analogously, one can show that $\overline{m}^{(0)}_{0,s}(\beta,\gamma)$ is strictly increasing in β_k if $s_k = +$, and strictly decreasing in β_k if $s_k = -$, for all $k \in \{1,\ldots,K\}$, because of the result in (18) above.

We can therefore apply Lemma 2 with $g_s(\beta,\gamma)$ equal to $\overline{m}^{(0)}_{0,s}(\beta,\gamma)$ to find that the system of equations in (19) has at most one solution. Using Lemma 1 we find that such a solution exists and is given by (β_0,γ_0).

For the other values of $q, y_0 \in \{0,1\}$ we can analogously apply Lemma 2 with $g_s(\beta,\gamma)$ equal to $\overline{m}^{(0)}_{1,s}(\beta,-\gamma)$, $\overline{m}^{(1)}_{0,s}(-\beta,-\gamma)$, $\overline{m}^{(1)}_{1,s}(-\beta,\gamma)$, respectively. ∎

Key intermediate result for the proof of Lemma 3 and Theorem 2

Lemma 4 below is a slight reformulation and generalization of Lemma 3 in the main text. Once we have established Lemma 4, then the proof of both Lemma 3 and Theorem 2 are straightforward. To present Lemma 4, it is useful to first introduce some additional notation:

Let $\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3 \in \{0,1\}$, $W_2 \in \mathcal{W}_2$, and $W_3 \in \mathcal{W}_3$ be random variables. Let $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$, and let $p(\widetilde{y}, w_2, w_3) \in [0,\infty)$ describe the joint distribution of $(\widetilde{Y}, W_2, W_3)$,

[22] Notice that we have introduced the domain of A to be $\mathbb{R}$. If we had introduced the domain of A to be $\mathbb{R} \cup \{\pm\infty\}$, then all other results in the paper hold completely unchanged, but here we would need to impose the additional regularity condition that A does not take values $\pm\infty$ with probability one, conditional on $Y_0 = 0$ and $X \in \mathcal{X}_s$, since otherwise we can have $p((1,0,1),0,X,\beta_0,\gamma_0,A) = 0$.

that is, for all measurable subsets $\mathcal{Y}_* \subset \{0,1\}^3$, $\mathcal{W}_2^* \subset \mathcal{W}_2$ and $\mathcal{W}_3^* \subset \mathcal{W}_3$ we have

$$\Pr\left(\widetilde{Y} \in \mathcal{Y}_* \ \ \& \ \ W_2 \in \mathcal{W}_2^* \ \ \& \ \ W_3 \in \mathcal{W}_3^*\right) = \sum_{\widetilde{y} \in \mathcal{Y}_*} \int_{w_2 \in \mathcal{W}_2^*} \int_{w_3 \in \mathcal{W}_3^*} p(\widetilde{y}, w_2, w_3)\, \mu(dw_2)\, \nu(dw_3),$$

for appropriate probability measures μ and ν on $\mathcal{W}_2$ and $\mathcal{W}_3$. We assume that we can decompose the joint distribution of $\widetilde{Y}$, W_2, W_3 as follows,

$$p(\widetilde{y}, w_2, w_3) = p_3(\widetilde{y}_3 \,|\, w_3)\ \ g(w_3 \,|\, \widetilde{y}_2, w_2)\ \ p_2(\widetilde{y}_2 \,|\, w_2)\ \ f(w_2 \,|\, \widetilde{y}_1)\ \ p_1(\widetilde{y}_1), \tag{32}$$

where $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$, the functions p_3, g, p_2, f are appropriate transition probabilities/densities, and $p_1(\widetilde{y}_1) = \Pr\left(\widetilde{Y}_1 = \widetilde{y}_1\right)$ is the marginal distribution of $\widetilde{Y}_1$. For $p_1(\widetilde{y}_1)$, $p_2(\widetilde{y}_2 \,|\, w_2)$, $p_3(\widetilde{y}_3 \,|\, w_3)$ we impose:

$$\begin{aligned} p_1(\widetilde{y}_1) &= \Lambda\left[(2\,\widetilde{y}_1 - 1)\,\pi_1\right], \\ p_2(\widetilde{y}_2 \,|\, w_2) &= \Lambda\left[(2\,\widetilde{y}_2 - 1)\,\pi_2(w_2)\right], \\ p_3(\widetilde{y}_3 \,|\, w_3) &= \Lambda\left[(2\,\widetilde{y}_3 - 1)\,\pi_3(w_3)\right], \end{aligned} \tag{33}$$

where $\Lambda(\xi) := [1 + \exp(-\xi)]^{-1}$ is the cumulative distribution function of the logistic distribution, $\pi_1 \in \mathbb{R}$ is a constant, and $\pi_2 : \mathcal{W}_2 \to \mathbb{R}$ and $\pi_3 : \mathcal{W}_3 \to \mathbb{R}$ are functions.[23] The only assumption that we impose on $f(w_2 \,|\, \widetilde{y}_1)$ and $g(w_3 \,|\, \widetilde{y}_2, w_2)$ is that

$$g(w_3 \,|\, 1, w_2) = g(w_3 \,|\, 1), \tag{34}$$

that is, conditional on $\widetilde{Y}_2 = 1$, the distribution of W_3 is independent of W_2. Apart from that, we only require that $f(w_2 \,|\, \widetilde{y}_1)$ and $g(w_3 \,|\, \widetilde{y}_2, w_2)$ are conditional probability distributions, which sum to one:

$$\int_{w_2 \in \mathcal{W}_2} f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2) = 1, \qquad\qquad \int_{w_3 \in \mathcal{W}_3} g(w_3 \,|\, \widetilde{y}_2, w_2)\, \nu(dw_3) = 1. \tag{35}$$

[23] Since π_1, $\pi_2(w_2)$, $\pi_3(w_3)$ are unrestricted, the only substantial assumption that is actually imposed by (33) is that the conditional probabilities should not be zero or one.

Notice that if we would strengthen (34) to $g(w_3 \,|\, \widetilde{y}_2, w_2) = g(w_3 \,|\, \widetilde{y}_2)$, for $\widetilde{y}_2 \in \{0,1\}$, then (32) would be equivalent to the Markov chain condition imposed in Lemma 3 ($\widetilde{Y}_1 \xrightarrow{f} W_2 \xrightarrow{p_2} \widetilde{Y}_2 \xrightarrow{g} W_3 \xrightarrow{p_3} \widetilde{Y}_3$). However, since here we only impose (34), we also allow for dependence of W_2 and W_3, conditional on $\widetilde{Y}_2 = 0$. This generalization to a non-Markovian structure is crucial for the ordered logit model in Honoré, Muris, and Weidner (2021), but is less important for the binary logit model discussed in this paper (See Appendix A.2 below for further discussion). Finally, we define $m : \{0,1\}^3 \times \mathcal{W}_2 \times \mathcal{W}_3 \to \mathbb{R}$ by

$$m(\widetilde{y}, w_2, w_3) := \begin{cases} \exp\left[\pi_1 - \pi_2(w_2)\right] & \text{if } \widetilde{y} = (0,1,0), \\ \exp\left[\pi_1 - \pi_3(w_3)\right] & \text{if } \widetilde{y} = (0,1,1), \\ -1 & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,0), \\ \exp\left[\pi_3(w_3) - \pi_2(w_2)\right] - 1 & \text{if } \widetilde{y} = (1,1,0), \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

Lemma 4 *Let $\pi_1 \in \mathbb{R}$, $\pi_2 : \mathcal{W}_2 \to \mathbb{R}$ and $\pi_3 : \mathcal{W}_3 \to \mathbb{R}$. Let the random variables $\widetilde{Y} \in \{0,1\}^3$, $W_2 \in \mathcal{W}_2$, $W_3 \in \mathcal{W}_3$ be such that their distributions satisfy (32), (33), (34), (35), and let $m : \{0,1\}^3 \times \mathcal{W}_2 \times \mathcal{W}_3 \to \mathbb{R}$ be defined by (36). Then we have*

$$\mathbb{E}\left[m(\widetilde{Y}, W_2, W_3)\right] = 0.$$

Proof. Define

$$h(\widetilde{y}_1, \widetilde{y}_2, w_2, w_3) := \sum_{\widetilde{y}_3 \in \{0,1\}} m(\widetilde{y}, w_2, w_3)\; p_3(\widetilde{y}_3 \,|\, w_3)\; p_2(\widetilde{y}_2 \,|\, w_2)\; p_1(\widetilde{y}_1),$$

where $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$. By using the expressions for the functions p_1, p_2, p_3, and m in

(33) and (36) we find that

$$h(\widetilde{y}_1, \widetilde{y}_2, w_2, w_3) = \begin{cases} \Lambda(\pi_1)\, \Lambda[-\pi_3(w_3)] & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (0,1), \\ -p_2(\widetilde{y}_2 \,|\, w_2)\, p_1(\widetilde{y}_1) & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,0), \\ \Lambda(\pi_1)\, \Lambda[\pi_3(w_3)] - p_2(\widetilde{y}_2 \,|\, w_2)\, p_1(\widetilde{y}_1) & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,1), \\ 0 & \text{otherwise,} \end{cases}$$

$$= \underbrace{\mathbb{1}\{\widetilde{y}_2 = 1\}\ \Lambda(\pi_1)\, \Lambda[(2\widetilde{y}_1 - 1)\pi_3(w_3)]}_{=:\, h_1(\widetilde{y}_1, \widetilde{y}_2, w_3)} \underbrace{-\, \mathbb{1}\{\widetilde{y}_1 = 1\}\ p_2(\widetilde{y}_2 \,|\, w_2)\, p_1(\widetilde{y}_1)}_{=:\, h_2(\widetilde{y}_1, \widetilde{y}_2, w_2)},$$

where we have decomposed $h(\widetilde{y}_1, \widetilde{y}_2, w_2, w_3)$ into the sum of $h_1(\widetilde{y}_1, \widetilde{y}_2, w_3)$, which does not depend on w_2, and of $h_2(\widetilde{y}_1, \widetilde{y}_2, w_2)$, which does not depend on w_3. Notice that the term $\Lambda[(2\widetilde{y}_1 - 1)\pi_3(w_3)]$ in $h_1(\widetilde{y}_1, \widetilde{y}_2, w_3)$ is identical to $p_3(\widetilde{y}_3 \,|\, w_3)$, but with $\widetilde{y}_3$ replaced by $\widetilde{y}_1$. Also using (34) and (35) we find

$$\sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{\widetilde{y}_2 \in \{0,1\}} \int_{w_2 \in \mathcal{W}_2} \int_{w_3 \in \mathcal{W}_3} h_1(\widetilde{y}_1, \widetilde{y}_2, w_3)\ g(w_3 \,|\, \widetilde{y}_2, w_2)\ f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2)\, \nu(dw_3)$$

$$= \Lambda(\pi_1) \sum_{\widetilde{y}_1 \in \{0,1\}} \int_{w_2 \in \mathcal{W}_2} \int_{w_3 \in \mathcal{W}_3} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(w_3)]\ g(w_3 \,|\, 1)\ f(w_2 \,|\, \widetilde{y}_1)\ \mu(dw_2)\, \nu(dw_3)$$

$$= \Lambda(\pi_1) \sum_{\widetilde{y}_1 \in \{0,1\}} \int_{w_3 \in \mathcal{W}_3} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(w_3)]\ g(w_3 \,|\, 1) \underbrace{\int_{w_2 \in \mathcal{W}_2} f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2)}_{=1} \nu(dw_3)$$

$$= \Lambda(\pi_1) \int_{w_3 \in \mathcal{W}_3} \underbrace{\sum_{\widetilde{y}_1 \in \{0,1\}} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(w_3)]}_{=1}\ g(w_3 \,|\, 1)\, \nu(dw_3)$$

$$= \Lambda(\pi_1) \underbrace{\int_{w_3 \in \mathcal{W}_3} g(w_3 \,|\, 1)\, \nu(dw_3)}_{=1}$$

$$= \Lambda(\pi_1).$$

Similarly, we calculate

$$\sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{\widetilde{y}_2 \in \{0,1\}} \int_{w_2 \in \mathcal{W}_2} \int_{w_3 \in \mathcal{W}_3} h_2(\widetilde{y}_1, \widetilde{y}_2, w_2)\ g(w_3 \,|\, \widetilde{y}_2, w_2)\ f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2)\, \nu(dw_3)$$

$$
\begin{aligned}
&= \sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{\widetilde{y}_2 \in \{0,1\}} \int_{w_2 \in \mathcal{W}_2} h_2(\widetilde{y}_1, \widetilde{y}_2, w_2) \underbrace{\int_{w_3 \in \mathcal{W}_3} g(w_3 \,|\, \widetilde{y}_2, w_2)\, \nu(dw_3)}_{=1} f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2) \\
&= -\, p_1(1) \int_{w_2 \in \mathcal{W}_2} \underbrace{\sum_{\widetilde{y}_2 \in \{0,1\}} p_2(\widetilde{y}_2 \,|\, w_2)}_{=1} \; f(w_2 \,|\, 1)\, \mu(dw_2) \\
&= -\, p_1(1) \underbrace{\int_{w_2 \in \mathcal{W}_2} f(w_2 \,|\, 1)\, \mu(dw_2)}_{=1} \\
&= -\, \Lambda(\pi_1).
\end{aligned}
$$

Combining the results in the last two displays gives

$$
\sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{\widetilde{y}_2 \in \{0,1\}} \int_{w_2 \in \mathcal{W}_2} \int_{w_3 \in \mathcal{W}_3} h(\widetilde{y}_1, \widetilde{y}_2, w_2, w_3) \;\; g(w_3 \,|\, \widetilde{y}_2, w_2) \;\; f(w_2 \,|\, \widetilde{y}_1)\, \mu(dw_2)\, \nu(dw_3) = 0,
$$

and by the definition of $h(\widetilde{y}_1, \widetilde{y}_2, w_2, w_3)$ this is equivalent to $\mathbb{E}\left[m(\widetilde{Y}, W_2, W_3)\right] = 0$, which is what we wanted to show. ■

We are now ready to prove the remaining main text results.

Proof of Lemma 3. Let the assumptions of Lemma 3 hold. We are going to verify that this implies that all the assumptions of Lemma 4 hold, conditional on (W_1, X, A). We condition on $(W_1, X, A) = (w_1, x, a)$ in all the following stochastic statements, and when verifying the assumptions of Lemma 4, we write $p_t(\widetilde{y}_t \,|\, w_t, x, \alpha)$, $t \in \{1, 2, 3\}$, instead of $p_1(\widetilde{y}_1)$, $p_2(\widetilde{y}_2 \,|\, w_2)$, $p_3(\widetilde{y}_3 \,|\, w_3)$. The additional arguments w_1, x, a do not matter here, since they are held constant (i.e. conditioned on). By the same argument, we write $\pi_t(w_t, x, \alpha)$, $t \in \{1, 2, 3\}$, here instead of π_1, $\pi_2(w_2)$, $\pi_3(w_3)$. The Markov chain assumption in Lemma 3 implies that conditions (32) and (34) hold.[24] For $t \in \{1, 2, 3\}$ we define

$$
\pi_t(w_t, x, \alpha) := \Lambda^{-1}\Big[p_t\left(1 \,\middle|\, w_t, x, \alpha\right)\Big], \tag{37}
$$

[24]In fact, we have $g(w_3 \,|\, \widetilde{y}_2, w_2) = g(w_3 \,|\, \widetilde{y}_2)$ not only for $\widetilde{y}_2 = 1$ but also for $\widetilde{y}_2 = 0$ here, which is stronger than required for applying Lemma 4.

which guarantees that (33) holds. Finally, (35) is satisfied for any conditional probability.[25] We can thus apply Lemma 4 to find that

$$\mathbb{E}\left[m^{(1)}(w_1,\widetilde{Y}_1,W_2,\widetilde{Y}_2,W_3,\widetilde{Y}_3,x,a)\,\middle|\,W_1=w_1,\;X=x,\;A=a\right]=0,$$

where $m^{(1)}(w_1,\widetilde{y}_1,w_2,\widetilde{y}_2,w_3,\widetilde{Y}_3,x,a)$ is equal to $m(\widetilde{y},w_2,w_3)$ defined in (36), with $\widetilde{y}=(\widetilde{y}_1,\widetilde{y}_2,\widetilde{y}_3)$, and with the additional arguments w_1, x, a added through $\pi_t(w_t,x,\alpha)$, as explained above, that is,

$$\begin{aligned}
&m^{(1)}(w_1,\widetilde{y}_1,w_2,\widetilde{y}_2,w_3,\widetilde{y}_3,x,\alpha)\\
&\quad=\begin{cases}
\exp\left[\pi_1(w_1,x,\alpha)-\pi_2(w_2,x,\alpha)\right] & \text{if } \widetilde{y}=(0,1,0),\\
\exp\left[\pi_1(w_1,x,\alpha)-\pi_3(w_3,x,\alpha)\right] & \text{if } \widetilde{y}=(0,1,1),\\
-1 & \text{if } (\widetilde{y}_1,\widetilde{y}_2)=(1,0),\\
\exp\left[\pi_3(w_3,x,\alpha)-\pi_2(w_2,x,\alpha)\right]-1 & \text{if } \widetilde{y}=(1,1,0),\\
0 & \text{otherwise},
\end{cases}\\
&\quad=-\mathbb{1}\{\widetilde{y}_1=1\}+\begin{cases}
\exp\left[\pi_1(w_1,x,\alpha)-\pi_2(w_2,x,\alpha)\right] & \text{if } \widetilde{y}=(0,1,0),\\
\exp\left[\pi_1(w_1,x,\alpha)-\pi_3(w_3,x,\alpha)\right] & \text{if } \widetilde{y}=(0,1,1),\\
\exp\left[\pi_3(w_3,x,\alpha)-\pi_2(w_2,x,\alpha)\right] & \text{if } \widetilde{y}=(1,1,0),\\
1 & \text{if } \widetilde{y}=(1,1,1),\\
0 & \text{otherwise},
\end{cases}\\
&\quad=-\mathbb{1}\{\widetilde{y}_1=1\}\\
&\qquad+\mathbb{1}\{\widetilde{y}_2=1\}\,\exp\left\{\sum_{t=1}^{3}\widetilde{y}_t\left[\pi_{\delta(t)}\left(w_{\delta(t)},x,\alpha\right)-\pi_t(w_t,x,\alpha)\right]\right\}\\
&\quad=-\mathbb{1}\{\widetilde{y}_1=1\}\\
&\qquad+\mathbb{1}\{\widetilde{y}_2=1\}\,\exp\left\{\frac{1}{2}\sum_{t=1}^{3}(2\widetilde{y}_t-1)\left[\pi_{\delta(t)}\left(w_{\delta(t)},x,\alpha\right)-\pi_t(w_t,x,\alpha)\right]\right\}
\end{aligned}$$

[25] In other words, explicitly imposing (35) is redundant in Lemma 4 since it is always satisfied in the setup described. We list this condition explicitly in the statement of Lemma 4 since it is used in the proof. For example, non-negativity of $f(w_2\,|\,\widetilde{y}_1)$ and $g(w_3\,|\,\widetilde{y}_2,w_2)$ is also implied by the setup described, but is not actually required for the proof.

$$= -\mathbb{1}\left\{\widetilde{y}_1 = 1\right\}$$
$$+ \mathbb{1}\left\{\widetilde{y}_2 = 1\right\} \exp\left(\frac{1}{2}\sum_{t=1}^{3}\left\{\Lambda^{-1}\Big[p_{\delta(t)}\left(\widetilde{y}_t \,\middle|\, w_{\delta(t)}, x, \alpha\right)\Big] - \Lambda^{-1}\Big[p_t\left(\widetilde{y}_t \,\middle|\, w_t, x, \alpha\right)\Big]\right\}\right),$$

where in the last step we used that (37) together with $\sum_{y\in\{0,1\}} p_t\left(y\middle|w_t, x, \alpha\right) = 1$ implies that $p_t\left(y\middle|w_t, x, \alpha\right) = \Lambda[(2y-1)\,\pi\left(w_t, x, \alpha\right)]$, for all $y \in \{0,1\}$ and $t \in \{1,2,3\}$. We have therefore shown the result of Lemma 3 for $q = 1$. The result for $q = 0$ directly follows from this by applying the symmetry transformation $\widetilde{y}_t \leftrightarrow 1 - \widetilde{y}_t$, and $\pi_t(w_t, x, \alpha) \leftrightarrow -\pi_t(w_t, x, \alpha)$. ■

Proof of Theorem 2. Let the assumptions of Theorem 2 be satisfied. By choosing $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3) = (Y_t, Y_s, Y_r)$ and $(W_1, W_2, W_3) = (Y_{t-1}, Y_{s-1}, Y_{r-1})$, it is then easy to verify that the conditions of Lemma 3 are satisfied, when conditioning on $(X, A, Y_0, Y_1, \ldots, Y_{t-1})$. The only substantial assumption to verify here is the Markov chain condition, which immediately follows from the AR(1) structure of Y_t. The panel AR(1) model restriction in (2) (after dropping the index i) then implies that the moment function $m^{(q)}(w_1, \widetilde{y}_1, w_2, \widetilde{y}_2, w_3, \widetilde{y}_3, x, \alpha)$ in the lemma now reads

$$\begin{aligned}
&m^{(q)(t,s,r)}(y, y_0, x, \beta_0, \gamma_0)\\
&\quad= -\mathbb{1}\left\{y_t = q\right\} + \mathbb{1}\left\{y_s = q\right\} \exp\left\{\frac{1}{2}\left[(2y_t - 1)z_{rt} + (2y_s - 1)z_{ts} + (2y_r - 1)z_{sr}\right]\right\}\\
&\quad= -\mathbb{1}\left\{y_t = q\right\} + \mathbb{1}\left\{y_s = q\right\} \exp\left(y_t\, z_{rt} + q\, z_{ts} + y_r\, z_{sr}\right),
\end{aligned}$$

where $z_{ts} = z_{ts}(y, y_0, x, \beta_0, \gamma_0) = (x_t - x_s)'\beta_0 + (y_{t-1} - y_{s-1})\gamma_0$ was already defined in the main text. It is easy to verify that the moment functions in the last display coincide with those defined in (24) of the main text. Lemma 3 therefore guarantees that the conditional moment condition in Theorem 2 holds. ■

A.2 Remarks on more general predetermined regressors

Everywhere in the main text of the paper we have assumed that X_{it} is strictly exogenous (because the model for the outcomes was specified conditional on X). However, our results for the panel AR(1) model can still be applied to certain types of predetermined regressors. In particular, arbitrary feedback from last period outcomes $Y_{i,t-1}$ into X_{it} can be allowed for, as long as X_{it} otherwise only depends on strictly exogenous variables. That is, for an arbitrary function $h(\cdot,\cdot)$ and an unobserved strictly exogenous variable $\widetilde{X}_{it}$ we consider

$$X_{it} = h(Y_{i,t-1}, \widetilde{X}_{it}). \tag{38}$$

Notice that (29) is a special case of this where $\widetilde{X}_{it} = X_{it}$ is observed and the regressors that actually enter into the model are given by the functions $(1-Y_{i,t-1})\, X_{it}$ and $Y_{i,t-1}\, X_{it}$. The novel point here is that $\widetilde{X}_{it}$ can be unobserved. In this setting, the model in (2) needs to be changed to

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, \widetilde{X}_i, A_i\right) = \frac{\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)}{1 + \exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)},$$

which by the law of iterated expectations implies

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i^t, A_i\right) = \frac{\exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)}{1 + \exp\left(X_{it}'\,\beta + Y_{i,t-1}\,\gamma + A_i\right)},$$

where $X_i^t = (X_{it}, X_{i,t-1}, X_{i,t-2}, \ldots)$. The last two displays formalize what we mean by $\widetilde{X}_{it}$ being strictly exogenous and X_{it} being predetermined.

The moment functions in (24) can be shown to be valid in this setting by applying Lemma 3 with $W_t = (Y_{t-1}, X_t)$, and with the conditioning variables X replaced by $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_T)$. For covariates of the form (38), the conclusion of Theorem 2 then gets modified to

$$\mathbb{E}\left[m^{(q)(t,s,r)}(Y, Y_0, X, \beta_0, \gamma_0) \,\middle|\, Y_0,\, Y_1, \ldots, Y_{t-1},\, X_1,\, X_2, \ldots, X_t,\, \widetilde{X},\, A\right] = 0,$$

which by the law of iterated expectations implies

$$\mathbb{E}\left[m^{(q)(t,s,r)}(Y, Y_0, X, \beta_0, \gamma_0) \,\middle|\, Y_0,\, Y_1, \ldots, Y_{t-1},\, X_1,\, X_2, \ldots, X_t\right] = 0. \tag{39}$$

Here, as in Theorem 2, we consider $t < s < r$, but the difference is that we only condition on covariates up to time period t in this moment condition, while in the main text we always conditioned on the whole X. The identification arguments and the GMM estimator in the main text would have to be adjusted accordingly, but we do not explore this generalization further here.

Another possible extension of our results to predetermined covariates is as follows: Consider the model in (29), but set $\beta_1 = 0$ and make X_{it} predetermined, that is,

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i^t, A_i\right) = \frac{\exp\left[(1 - Y_{i,t-1})X_{it}'\,\beta_0 + Y_{i,t-1}\,\gamma + A_i\right)}{1 + \exp\left((1 - Y_{i,t-1})X_{it}'\,\beta_0 + Y_{i,t-1}\,\gamma + A_i\right]}. \tag{40}$$

It turns out that the moment conditions in (39) remain valid in this model for $q = 1$ even for general pre-determined regressors X_{it}. The reason for this is that Lemma 4 in the appendix is more general than Lemma 3 in the main text. Specifically, in Lemma 4, the condition (34) only demands $g(w_3 \,|\, 1, w_2) = g(w_3 \,|\, 1)$, which rules out direct feedback from W_2 into W_3 whenever $\widetilde{Y}_2 = 1$, but this still allows for arbitrary feedback from W_2 into W_3 whenever $\widetilde{Y}_2 = 0$. By following the proof of Theorem 2 (where $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3) = (Y_t, Y_s, Y_r)$), but setting $(W_1, W_2, W_3) = ((Y_{t-1}, (1 - Y_{t-1})X_t), (Y_{s-1}, (1 - Y_{s-1})X_s), (Y_{r-1}, (1 - Y_{r-1})X_r))$, and conditioning on W_1, we find that for general pre-determined X_{it} the moment condition in (39) still holds for $q = 1$, as long as X_t only enters into the model for Y_t through $(1 - Y_{t-1})X_t$, as in (40).

Of course, model (40) and pre-determined regressors of the form (38) are both quite restrictive. This is why we only briefly discuss those possible extensions to predetermined regressors here in the appendix. Nevertheless, the possibility of using functional differencing ideas to make progress on non-linear panel model with more general predetermined regressors is quite exciting, see also Bonhomme, Dano, and Graham (2022).

A.3 Fixed effect logit AR(p) models with $p > 1$

In this appendix section, we consider logit AR(p) models, that is, we generalize the model in (2) to

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i, A_i, \beta, \gamma\right) = \frac{\exp\left(X_{it}'\,\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}{1 + \exp\left(X_{it}'\,\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}, \tag{41}$$

where $\gamma = (\gamma_1, \ldots, \gamma_p)'$. We assume that the autoregressive order $p \in \{2, 3, 4, \ldots\}$ is known, and that outcomes Y_{it} are observed for time periods $t = t_0, \ldots, T$, with $t_0 = 1 - p$. Thus, the total number of time periods for which outcomes are observed is $T_{\text{obs}} = T + p$, consisting of T periods for which the model applies and p periods to observe the initial conditions. We maintain the definition $Y_i = (Y_{i1}, \ldots, Y_{iT})$, but the initial conditions are now described by the vector $Y_i^{(0)} = (Y_{i,t_0}, \ldots, Y_{i0})$. Analogous to (3), we define

$$p_{y_i^{(0)}}(y_i, x_i, \beta, \gamma, \alpha_i) = \prod_{t=1}^{T} \frac{\exp\left(x_{it}'\,\beta + \sum_{\ell=1}^{p} y_{i,t-\ell}\,\gamma_\ell + \alpha_i\right)}{1 + \exp\left(x_{it}'\,\beta + \sum_{\ell=1}^{p} y_{i,t-\ell}\,\gamma_\ell + \alpha_i\right)}, \tag{42}$$

and we continue to denote the true model parameters by β_0 and γ_0. We again drop the cross-sectional indices i unless they are required.

Our main focus is again on finding moment conditions that are applicable to all values of the parameters and all realizations of the regressors. For a given autoregressive order p, one requires $T \geq 2 + p$ (i.e. $T_{\text{obs}} \geq 2 + 2p$) time periods to find such general moment conditions, and the number of linearly independent moment conditions available for each initial condition $y^{(0)}$ is then equal to

$$\ell = 2^T - (T + 1 - p)\, 2^p. \tag{43}$$

Analogous to Section 3.4 one can use the polynomial structure of (42) in $\exp(\alpha_i)$ to show that at least that many moment conditions have to exist. That (43) actually holds with equality can then be verified numerically.

In addition to those general moment conditions, which exist for all possible values of β, γ and x_{it}, there are additional ones that only become available for special values of the parameters and of the regressors. For example, for $T = 4$ the AR(2) model has $\ell = 4$ general moments for each initial condition, but if $\gamma_2 = 0$ then the model becomes an AR(1) model with $\ell = 8$ available moments for each initial condition.[26] Furthermore, there are additional moment conditions available for special realizations of the regressors, and those can provide identifying information for the parameters β and γ even when $T < 2 + p$,[27] see Section B.3.1.

A.3.1 AR(2) models with $T = 4$

Consider model (14) with $p = 2$ and $T = 4$. Again write $z_t(y, y_0, x, \beta, \gamma) = x_t' \beta + y_{t-1}\gamma_1 + y_{t-2}\gamma_2$ for the single index that describes how the parameters β and γ enter into the model at time period t, and let $z_{ts}(y, y_0, x, \beta, \gamma) = z_t(y, y_0, x, \beta, \gamma) - z_s(y, y_0, x, \beta, \gamma)$ be its time-differences. To save space, we drop the arguments of the differences in the following formulas and write z_{ts} instead of $z_{ts}(y, y_0, x, \beta, \gamma)$. One valid moment function for any initial condition $y^{(0)} \in \{0, 1\}^2$ and general covariate values $x \in \mathbb{R}^{K \times T}$ is then given by

$$m^{(a,2,4)}_{y^{(0)}}(y, x, \beta, \gamma) = \begin{cases} \exp(z_{23}) - \exp(z_{43}) & \text{if } y = (0,0,1,0), \\ \exp(z_{24}) - 1 & \text{if } y = (0,0,1,1), \\ -1 & \text{if } (y_1, y_2) = (0,1), \\ \exp(z_{41} + \gamma_1) & \text{if } (y_1, y_2, y_3) = (1,0,0), \\ \exp(z_{41})\big[1 + \exp(z_{23}) - \exp(z_{43})\big] & \text{if } y = (1,0,1,0), \\ \exp(z_{21}) & \text{if } y = (1,0,1,1), \\ 0 & \text{otherwise.} \end{cases}$$

[26]Setting $\gamma_k = 0$ gives additional moments not only for $k = p$. For example, for the AR(2) model with $T = 4$ one finds nine valid moment conditions for each initial condition when $\gamma_1 = 0$ and $\gamma_2 \neq 0$.

[27]Interestingly, this is not the case for AR(1) models, where for $\gamma \neq 0$ we have always found the same number of available moment conditions, completely independent of the regressor value $x \in \mathbb{R}^{K \times T}$.

There are three additional moment functions $m_{y^{(0)}}^{(b,2,4)}(y,x,\beta,\gamma)$, $m_{y^{(0)}}^{(c,2,4)}(y,x,\beta,\gamma)$ and $m_{y^{(0)}}^{(d,2,4)}(y,x,\beta,\gamma)$, and they are provided in Appendix B.3.2.

Lemma 5 *If the outcomes $Y = (Y_1, Y_2, Y_3, Y_4)$ are generated from model (14) with $p = 2$, $T = 4$ and true parameters β_0 and γ_0, then we have for all $y^{(0)} \in \{0,1\}^2$, $x \in \mathbb{R}^{K\times 4}$, $\alpha \in \mathbb{R}$, and $\xi \in \{a,b,c,d\}$ that*

$$\mathbb{E}\left[m_{y^{(0)}}^{(\xi,2,4)}(Y,X,\beta_0,\gamma_0) \,\Big|\, (Y_{-1},Y_0) = y^{(0)},\ X = x,\ A = \alpha \right] = 0.$$

The proof of the lemma is discussed in Appendix B.3.3. Analogous to our discussion for the AR(1) model, one can apply GMM to those moment conditions. Under suitable regularity conditions, this allows to estimate the parameters β and γ of the panel AR(2) logit model at root-n-rate.

A.3.2 AR(3) models with $T = 5$

To illustrate the applicability of this general approach to constructing moment conditions further, we next consider a panel logit AR(3) model with $T = 5$. Since the model contains three lags, one needs to observe three time periods as initial conditions, which gives a total number of observations required of $T_{\text{obs}} = 8$. Let $z_t(y, y_0, x, \beta, \gamma) = x_t'\beta + y_{t-1}\gamma_1 + y_{t-2}\gamma_2 + y_{t-3}\gamma_3$, and let $z_{ts}(y, y_0, x, \beta, \gamma) = z_t(y, y_0, x, \beta, \gamma) - z_s(y, y_0, x, \beta, \gamma)$. We again drop the arguments of z_{ts} in the following. There are then eight valid moment functions for any initial condition $y^{(0)} \in \{0,1\}^3$ and general covariate values $x \in \mathbb{R}^{K\times 5}$.

The first of those is given by

$$m_{y^{(0)}}^{(a,3,5)}(y,x,\beta,\gamma)=\begin{cases}\exp\left(z_{23}\right)-\exp\left(z_{53}\right) & \text{if } y=(0,0,1,0,0),\\ \left[\exp\left(z_{43}\right)-\exp\left(z_{45}\right)+1\right] & \\ \qquad\qquad\qquad \times\left[\exp\left(z_{25}\right)-1\right] & \text{if } y=(0,0,1,0,1),\\ \exp\left(\gamma_1+z_{25}\right)-1 & \text{if } (y_1,\ldots,y_4)=(0,0,1,1),\\ -1 & \text{if } (y_1,y_2)=(0,1),\\ \exp\left(\gamma_2+z_{51}\right) & \text{if } (y_1,\ldots,y_4)=(1,0,0,0),\\ \exp\left(-\gamma_1+\gamma_2+z_{51}\right) & \text{if } (y_1,\ldots,y_4)=(1,0,0,1),\\ \exp\left(z_{51}\right)\left(\exp\left(z_{23}\right)-\exp\left(z_{53}\right)+1\right) & \text{if } y=(1,0,1,0,0),\\ \exp\left(z_{21}\right)+\exp\left(z_{41}\right)+\exp\left(z_{21}+z_{43}\right) & \\ \quad -\exp\left(z_{21}+z_{45}\right)-\exp\left(z_{41}+z_{53}\right) & \text{if } y=(1,0,1,0,1),\\ \exp\left(z_{21}\right) & \text{if } (y_1,\ldots,y_4)=(1,0,1,1),\\ 0 & \text{otherwise,}\end{cases}$$

where the additional superscripts indicate $p=3$ and $T=5$. The remaining moment functions $m_{y^{(0)}}^{(b,3,5)}$, $m_{y^{(0)}}^{(c,3,5)}$, $m_{y^{(0)}}^{(d,3,5)}$, $m_{y^{(0)}}^{(e,3,5)}$, $m_{y^{(0)}}^{(f,3,5)}$, $m_{y^{(0)}}^{(g,3,5)}$, and $m_{y^{(0)}}^{(h,3,5)}$ are displayed in Appendix B.3.4.

Lemma 6 *If the outcomes $Y=(Y_1,Y_2,Y_3,Y_4,Y_5)$ are generated from model (14) with $p=3$, $T=5$ and true parameters β_0 and γ_0, then we have for all $y^{(0)}\in\{0,1\}^3$, $x\in\mathbb{R}^{K\times 5}$, $\alpha\in\mathbb{R}$, and $\xi\in\{a,b,c,d,e,f,g,h\}$ that*

$$\mathbb{E}\left[m_{y^{(0)}}^{(\xi,3,5)}(Y,X,\beta_0,\gamma_0)\,\middle|\,(Y_{-2},Y_{-1},Y_0)=y^{(0)},\ X=x,\ A=\alpha\right]=0.$$

The proof of this lemma is analogous to that of Lemma 5. In fact, the structure and derivation of $m_{y^{(0)}}^{(a,3,5)}(y,x,\beta,\gamma)$ is very similar to that of of $m_{y^{(0)}}^{(a,2,4)}(y,x,\beta,\gamma)$.

AR(2) models with $T=5$: The AR(2) model is a special case of the AR(3) model, that is, the moment conditions in Lemma 6 are also applicable to AR(2) models with $T=5$, we just need to set $\gamma_3=0$. In addition, we can construct valid moment functions for the AR(2) model with $T = 5$ by using (time-shifted versions of) the moment functions in Section A.3.1. Using the results presented so far then gives a total of twenty valid moment functions for AR(2) models with $T = 5$. However, there are four linear dependencies between those, so the total number of linearly independent moment conditions available for $p = 2$ and $T = 5$ is equal to $\ell = 16$, in agreement with equation (43). See Appendix B.3.6 for details.

A.4 Computational remarks

In Section 2.2 we discussed a strategy for numerically exploring the existence of moment conditions. For this, we need to verify whether solutions to (7) exist for some fixed numerical values for $y^{(0)}$, x, θ, and $\alpha_1,\ldots,\alpha_Q$, for some $Q > |\mathcal{Y}|$. For those fixed values, let $\mathbf{F}$ be the $Q \times |\mathcal{Y}|$ matrix with entries $f\big(y \,\big|\, y^{(0)},\, x,\, \alpha_q;\, \theta\big)$, where rows are labeled by $q \in \{1, \ldots, Q\}$ and columns are labeled by $y \in \mathcal{Y}$. Finding a non-trivial solution to (7) is then equivalent to finding a $|\mathcal{Y}|$-vector $\mathbf{m} \neq 0$ such that

$$\mathbf{F}\,\mathbf{m} = 0,$$

that is, finding an element of the nullspace of $\mathbf{F}$. Another equivalent rewriting of this condition is

$$\mathbf{F}'\,\mathbf{F}\,\mathbf{m} = 0.$$

A solution $\mathbf{m} \neq 0$ exists if and only if the symmetric semi-definite $|\mathcal{Y}| \times |\mathcal{Y}|$ matrix $\mathbf{F}'\,\mathbf{F}$ has a zero eigenvalue. Furthermore, the number of linearly independent valid moment functions $\mathbf{m}$ is equal to the number of zero eigenvalues of $\mathbf{F}'\,\mathbf{F}$.

Calculating eigenvalues numerically is of course straightforward on modern computers. However, one needs to be careful here to distinguish true zero eigenvalues from merely very small eigenvalues. For the numerical counting of linearly independent moment conditions in this paper we have used Mathematica to calculate the eigenvalues of $\mathbf{F}'\,\mathbf{F}$ with a working precision of 1,000 digits, and we only counted an eigenvalue as equal to zero when it was smaller than 10^{-100}. See the accompanying Mathematica notebook “BinaryChoiceMoments.nb” for further details.

In Section 3.2 we explain how to find analytical expressions for the moment functions of the panel logit AR(1) model with $T = 3$. For this, one needs to construct the 8×8 matrices $\mathbf{B}^{(0)}$ and $\mathbf{B}^{(1)}$ and the unit vector $\mathbf{e}_4$, and then calculate the moment vectors $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ according to (9). This calculation and the subsequent simplification of the results for $\mathbf{m}^{(0)}$ and $\mathbf{m}^{(1)}$ only takes a few seconds on a modern symbolic computation system like Mathematica. However, the analogous calculation for more complicated models and larger number of time periods T can be slower and may require some additional human input to speed up the computation.

For example, for the derivation of the panel AR(2) moment functions for $T = 4$ in Appendix A.3.1 above, we again followed the approach described in Section 3.2, but to keep the symbolic computation manageable, it was important to first gain some intuition about the structure of the moment functions by numerical experimentation (i.e. using concrete numerical values to calculate numerical solutions for the moment functions when imposing various constraints and normalizations on them). For example, for $m^{(a,2,4)}_{y^{(0)}}$ above, one can first form the conjecture, based on the numerical calculations, that there should exist a valid moment function of the form

$$m^{(a,2,4)}_{y^{(0)}} = \begin{cases} \mu_1 & \text{if } y = (0,0,1,0), \\ \mu_2 & \text{if } y = (0,0,1,1), \\ -1 & \text{if } (y_1, y_2) = (0,1), \\ \mu_3 & \text{if } (y_1, y_2, y_3) = (1,0,0), \\ \mu_4 & \text{if } y = (1,0,1,0), \\ \mu_5 & \text{if } y = (1,0,1,1), \\ 0 & \text{otherwise}, \end{cases} \tag{44}$$

and one can then solve for the unknown $\mu_1, \ldots, \mu_5$ analytcially by solving the linear system

$$\sum_{y \in \{0,1\}^4} p_{y^{(0)}}(y, x, \beta_0, \gamma_0, \alpha_q)\, m^{(a,2,4)}_{y^{(0)}}(y, x, \beta_0, \gamma_0) = 0, \quad q \in \{1,2,3,4,5\}, \tag{45}$$

where we can choose five mutually different values $\alpha_q \in \mathbb{R}$ arbitrarily (the solution for $m^{(a,2,4)}_{y^{(0)}}$ will not depend on that choice). Once the structure of the moment function in (44) is found (i.e. guessed correctly via numerical experimentation), then solving (45) to find analytic solutions for (45) is again very quick.

A.5 More results for the model with heterogeneous time trends

In Section 4.3 we have already introduced and discussed the panel logit AR(1) with heterogeneous time trends. Analogous to Section 3.4, we now derive a lower bound on the number of valid moment conditions in this model with heterogeneous time trends. The probability distribution for $Y_i = (Y_{i1}, \ldots, Y_{iT})$ (conditional on Y_{i0}, X_i, D_i, A_i) is given by

$$f\big(y \,\big|\, y^{(0)},\, x,\, \alpha, \delta;\, \beta, \gamma\big) \;=\prod_{t=1}^{T} \frac{\exp\left(x_t'\,\beta + y_{t-1}\,\gamma + t\,\delta + \alpha\right)^{y_t}}{1+\exp\left(x_t'\,\beta + y_{t-1}\,\gamma + t\,\delta + \alpha\right)},$$

Defining $a = \exp(\alpha)$, $\pi_t(y_{t-1}) = \exp[x_{it}'\,\beta + y_{i,t-1}\,\gamma]$ and $d = \exp(\delta)$ and noticing that $\exp[\alpha + \delta\,(t-1)] = a\,d^{t-1}$, we have

$$\begin{aligned} f\big(y \,\big|\, y^{(0)},\, x,\, \alpha, \delta;\, \beta, \gamma\big) &= [a\,\pi_1(y_0)]^{y_1} \prod_{t=2}^{T} \left\{ [1 + a\,d^{t-1}\,\pi_t(1-y_{t-1})] \,\left[a\,d^{t-1}\,\pi_t(y_{t-1})\right]^{y_t} \right\} \\ &=: \sum_{k=0}^{2T-1} \sum_{\ell=\ell_{\min}(k)}^{\ell_{\max}(k)} a^k\, d^\ell\, c_{k\ell}(y) \end{aligned}$$

where $c_{k\ell}(y)$ is implicitly defined by the last equation, and

$$\begin{aligned} \ell_{\min}(k) &:= \lfloor k/2 \rfloor \big(\lfloor k/2 \rfloor + \mathrm{mod}(k,2)\big), \\ \ell_{\max}(k) &:= T\,k - \lfloor (k+1)/2 \rfloor \big(\lfloor (k+1)/2 \rfloor + \mathrm{mod}(k+1,2)\big) \end{aligned}$$

are the minimum and maximum power d^ℓ that are possible in the above polynomial for a given power a^k.

Finding a function $m(y)$ that satisfies $\sum_{y\in\mathcal{Y}} m(y)\, f\big(y \,\big|\, y^{(0)},\, x,\, \alpha, \delta;\, \beta, \gamma\big) = 0$ for all

$\alpha \in \mathbb{R}$ and $\delta \in \mathbb{R}$ (or equivalently for all $a > 0$ and $d > 0$) is therefore equivalent to finding a function $m(y)$ that satisfies $\sum_{y \in \mathcal{Y}} m(y)\, c_{k\ell}(y) = 0$ for all the values that the indices k and ℓ can take in the above expressions for $f\big(y \,\big|\, y^{(0)},\, x,\, \alpha, \delta;\, \beta, \gamma\big)$. The total number of different values that k and ℓ can take is given by

$$
\begin{aligned}
&\sum_{k=0}^{2T-1} \left[1 + \ell_{\max}(k) - \ell_{\min}(k)\right] \\
&= 2T + T \sum_{k=0}^{2T-1} k \\
&\qquad - \sum_{k=0}^{2T-1} \underbrace{\Big\{ \lfloor k/2 \rfloor \Big(\lfloor (k+1)/2 \rfloor \Big(\lfloor (k+1)/2 \rfloor + \operatorname{mod}(k+1, 2) \Big) + \lfloor k/2 \rfloor + \operatorname{mod}(k, 2) \Big) \Big\}}_{= k(k+1)/2} \\
&= 2T + T^2\,(2T-1) + \frac{1}{3}\, T\,(2T+1)\,(2T-1) \\
&= \frac{T}{3}\,\big(2T^2 - 3T + 7\big) =: r_T
\end{aligned}
$$

Thus, the condition $\sum_{y \in \mathcal{Y}} m(y)\, c_{k\ell}(y) = 0$ is imposing r_T linear restrictions on the 2^T variables $m(y) \in \mathbb{R}$. If $2^T > r_T$, then a solution needs to exist. For $T \geq 9$ one finds that indeed $2^T > r_T$, which implies that for $T \geq 9$ it must be the case that moment conditions for this model exist.

We have not calculated any of those moment conditions for $T \geq 9$. However, as already mentioned in the main text, for the special case $\beta = 0$ (or equivalently $x_t = 0$ for all t), where the only remaining common model parameter is $\gamma \in \mathbb{R}$, we have calculated valid moment functions for this model when $T = 8$. There are two such valid moment functions $m(y, \gamma)$ and we have obtained analytical expressions for both of them. The resulting expressions are reported in the accompanying Mathematica program "BinaryChoiceMoments.nb". All the entries of both of these moment functions are polynomials in $\exp(\gamma)$, but the polynomial order can be as high as 79, making it impossible to report them here. The fact that expressions for these moment functions could still be obtained analytically nevertheless illustrates the wide applicability of the methods described in this paper.

B Supplementary Appendix

B.1 Monte Carlo simulations

In this section, we present the results of a small Monte Carlo study for the panel logit AR(1) and panel data AR(2) models. Our aim is to illustrate the possibility of our moment conditions to estimate the parameters in cases with moderate sample sizes and a realistic number of explanatory variables. A secondary aim is to gauge the efficiency loss of the GMM procedure relative to maximum likelihood when the data generating process does not contain individual-specific fixed effects.

B.1.1 Fixed effects logit AR(1)

We consider first the fixed effect logit AR(1) model in equation (2). In addition to the fixed effects, there are $K = 3$ or $K = 10$ explanatory variables. These are independent and identically distributed over time. We consider $T = 3$ (i.e. $T_{\text{obs}} = 4$ observed time periods), and in each time period ($t = 0, 1, 2, 3$) we generate the random variables as follows: $X_{1it} \sim \mathcal{N}(0, 1)$, while for $k = 2, \ldots, K$ we set $X_{kit} = (X_{1it} + Z_{itk}) / \sqrt{2}$, with $Z_{itk} \sim \mathcal{N}(0, 1)$. A_i is either 0 or $\frac{1}{2} \sum_{t=0}^{3} x_{1it}$ — we refer to the latter as "A_i varies". Y_{it} is generated according to the model with the lagged Y in period $t = 0$ set to 0. The parameters are $\gamma = 1$, $\beta_1 = \beta_2 = 1$, and $\beta_k = 0$ for $k = 3, \ldots, K$. The number of replications is $2,500$. The weight matrix is diagonal with the diagonals being the inverses of the variances of each moment evaluated at the logit maximum likelihood estimator that ignores the fixed effects.

The mean of $(Y_{i0}, Y_{i1}, Y_{i2}, Y_{i3})$ is approximately $(0.500, 0.577, 0.589, 0.590)$ for the designs with $\alpha_i = 0$ and $(0.500, 0.561, 0.570, 0.571)$ with non-constant fixed effects.

The approximate probability of each sequence based on 100,000 draws is displayed in Table 3 in Section B.1.3. We note that for the design with a fixed effect that varies across observations, approximately half of the observations do not contribute to any of the moments discussed above.

We compare the performance of the GMM estimator to the logit maximum likelihood estimator that ignores the fixed effect (but includes a common constant) and to the maximum likelihood estimator that estimates a constant for each individual.

Tables 5 through 8 in Section B.1.3 display the median bias and median absolute error for each estimator for sample sizes 500, 2,000, and 8,000. When $K = 10$, the last eight explanatory variables enter symmetrically. We therefore report the average median bias and the average median absolute error over the coefficients of those eight variables.

The estimation results for the logit model without fixed effects are as expected. When the fixed effect is 0 for all observations, this is the correctly specified maximum likelihood estimator, and it performs very well. However, when the data-generating process includes a fixed effect that varies across observations, the logit model will attempt to capture all of the persistence via the lagged dependent variable, leading to an upwards bias in that parameter. On the other hand, when we treat the fixed effects as parameters to be estimated, these fixed effect are estimated on the basis of three observations each, leading to severe overfitting. This, in turn, leads to large downwards biases in the estimate of γ. The GMM estimator does well in terms of bias when the sample size is large. This is true whether or not the data generating process includes a fixed effect that varies across individuals. Not surprisingly, this estimator is less precise than the logit maximum likelihood estimate when the fixed effect is constant.

The GMM estimator does suffer from moderate small sample bias when the number of explanatory variables is large. This is not surprising, since estimation is based on a very large number of moments (124 when $K = 10$)[28].

Figure 2 shows the densities for the estimators of γ for the three estimators and the four designs. As predicted by asymptotic theory, all three estimators have a distribution which is well-approximated by a normal centered around some (pseudo-true) value when the sample is large (2,000 or 8,000). The logit estimator which estimates a fixed effect for each i appears to have a slight asymmetry when $n = 500$. There are three striking feature of Figure 2. The first is the obvious inconsistency of the maximum likelihood estimator when it is misspecified (the left column of Figure 2). The second striking feature of Figure 2 is the importance of the incidental parameters problem (the middle column). Finally, the right column suggests that the GMM estimator is approximately centered on the true value when the sample size is relatively large. As mentioned above, it does have some bias when the sample is small.

[28]This is also the reason why we do not pursue efficient GMM estimation.

B.1.2 Fixed effects logit AR(2)

The Monte Carlo design for the AR(2) model resembles that for the AR(1) model. In addition to the fixed effects, there are $K = 3$ or $K = 10$ explanatory variables. These are independent and identically distributed over time. We consider $T = 4$ (i.e. $T_{\text{obs}} = 6$ observed time periods), and in each time period ($t = -1, 0, 1, 2, 3, 4$) we generate the random variables as follows:

$X_{1it} \sim \mathcal{N}(0,1)$, while for $k = 2, \ldots, K$ we set $X_{kit} = (X_{1it} + Z_{itk})/\sqrt{2}$, with $Z_{itk} \sim \mathcal{N}(0,1)$. A_i is either 0 or $\frac{1}{2}\sum_{t=-1}^{4} x_{1it}$ — we again refer to the latter as "A_i varies". Y_{it} is generated according to the model with the lagged Y's in period $t = -1$ set to 0. The parameters are $\gamma_1 = \gamma_2 = 1$, $\beta_1 = \beta_2 = 1$, and $\beta_k = 0$ for $k = 3, \ldots, K$, and the number of replications is $2,500$. The weight matrix is diagonal with the diagonals being the inverses of the variances of each moment evaluated at the logit maximum likelihood estimator that ignores the fixed effects.

For the designs with $\alpha_i = 0$, the mean of $(Y_{i,-1}Y_{i0}, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ is approximately $(0.500, 0.577, 0.625, 0.638, 0.644, 0.646)$. It is $(0.500, 0.561, 0.595, 0.603, 0.606, 0.607)$ with non-constant fixed effects. Table 3 in Section B.1.3 displays the distribution of $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ (averaged over the initial conditions). For the design with a fixed effect that varies across observations, approximately 44% of the observations do not contribute to any of the moments discussed above.

We again compare the performance of the GMM estimator to the logit maximum likelihood estimator that ignores the fixed effect (but includes a common constant) and to the maximum likelihood estimator that estimates a constant for each individual. As mentioned in Section 7, we have found that it is important to scale the moment conditions in order to limit the influence of any one observation. In order to limit the influence of the explanatory variables on the moment functions, we therefore scale the four moment conditions in Section A.3.1 by the sum of the absolute values of the possible seven non-zero terms in the moment conditions in Sections A.3.1 and B.3.3 The resulting conditional moment functions are interacted with dummy variables for each of the four initial conditions as well as $x_{i2} - x_{i1}$, $x_{i3} - x_{i2}$, and $x_{i4} - x_{i3}$. This leads to a total of $4 \cdot (4 + 3k)$ moment conditions.

Tables 9 through 12 in Section B.1.3 display the median bias and median absolute error for each estimator for sample sizes 500, 2,000, and 8,000. When $K = 10$, the last eight explanatory variables enter symmetrically, and report the average statistics of those eight variables.

The estimation results for the logit model without fixed effects are as expected. When the fixed effect is 0 for all observations, this estimator performs very well. However, when the data-generating process includes a fixed effect that varies across observations, the logit model will attempt to capture all of the persistence via the lagged dependent variables, leading to an upwards bias in those parameters. For the designs considered here, this leads to approximately equal bias in γ_1 and γ_2. On the other hand, estimating a fixed effect for each individual again leads to severe overfitting because each fixed effect is estimated on the basis of five observations. This, in turn, leads to large downwards biases in the estimate of the γ's.

As was the case for the AR(1) model, the GMM estimator always performs significantly better than the logit estimator that treats the fixed effects as parameters to be estimated. It also outperforms the logit model that ignores the fixed effects when the data-generating model includes fixed effects. Not surprisingly, this estimator is less precise than the logit maximum likelihood estimate when the fixed effect is constant across individuals. The GMM estimator does suffer from a small sample bias when the number of explanatory variables is large. Again, this is not surprising, since estimation is based on a very large number of moments (136 when $K = 10$)

B.1.3 Monte Carlo Tables

Table 3: Frequency of (y_0, y_1, y_2, y_3) for the two distributions of α_i in the AR(1) Design

$A_i = 0$		A_i varies	
Sequence	Probability	Sequence	Probability
0000	6.266%	0000	13.974%
0001	6.273%	0001	5.763%
0010	4.305%	0010	4.323%
0011	8.175%	0011	5.780%
0100	4.316%	0100	4.334%
0101	4.314%	0101	2.997%
0110	5.656%	0110	4.030%
0111	10.661%	0111	8.764%
1000	4.331%	1000	4.367%
1001	4.323%	1001	3.018%
1010	3.000%	1010	2.120%
1011	5.657%	1011	4.526%
1100	5.621%	1100	4.018%
1101	5.671%	1101	4.544%
1110	7.464%	1110	5.741%
1111	13.967%	1111	21.701%

Table 4: Frequency of (y_1, y_2, y_3, y_4) for the two distributions of α_i in the AR(2) Design

$A_i = 0$		A_i varies	
Sequence	Probability	Sequence	Probability
0000	4.330 %	0000	13.351 %
0001	4.349 %	0001	4.519 %
0010	2.996 %	0010	3.476 %
0011	5.657 %	0011	3.853 %
0100	2.929 %	0100	3.419 %
0101	4.013 %	0101	2.731 %
0110	3.626 %	0110	2.599 %
0111	9.521 %	0111	6.536 %
1000	3.980 %	1000	4.267 %
1001	3.959 %	1001	2.621 %
1010	3.784 %	1010	2.605 %
1011	7.156 %	1011	5.029 %
1100	5.086 %	1100	3.532 %
1101	6.981 %	1101	4.929 %
1110	8.717 %	1110	6.028 %
1111	22.916 %	1111	30.505 %

Table 5: AR(1). No Fixed Effects. $K = 3$. 2500 replications.

Logit MLE

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	−0.002	0.007	0.001	0.000	0.001	−0.001	−0.000	0.002	0.001	0.001	0.000	0.001
MAE	0.094	0.081	0.070	0.063	0.048	0.039	0.035	0.032	0.023	0.020	0.017	0.016

Logit MLE with Estimated Fixed Effects

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	−2.202	0.764	0.751	−0.002	−2.201	0.741	0.747	−0.002	−2.193	0.739	0.742	0.000
MAE	2.202	0.764	0.751	0.169	2.201	0.741	0.747	0.084	2.193	0.739	0.742	0.040

GMM

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.055	0.057	0.046	0.028	−0.001	0.001	0.008	0.014	0.001	0.000	0.003	0.003
MAE	0.254	0.284	0.211	0.199	0.127	0.131	0.098	0.092	0.065	0.058	0.044	0.042

Note: “Bias” refers to median bias.

Table 6: AR(1). With Fixed Effects. $K = 3$. 2500 replications.

Logit MLE

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.756	0.323	−0.082	0.002	0.746	0.314	−0.083	0.001	0.745	0.314	−0.084	0.000
MAE	0.756	0.323	0.099	0.067	0.746	0.314	0.083	0.033	0.745	0.314	0.084	0.017

Logit MLE with Estimated Fixed Effects

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	−2.402	0.787	0.757	−0.015	−2.382	0.751	0.755	0.002	−2.368	0.744	0.750	0.001
MAE	2.402	0.787	0.757	0.183	2.382	0.751	0.755	0.096	2.368	0.744	0.750	0.048

GMM

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3	γ	β_1	β_2	β_3
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.147	0.111	0.053	0.028	0.027	0.015	0.009	0.012	0.002	0.000	0.003	0.006
MAE	0.350	0.327	0.234	0.220	0.157	0.150	0.113	0.103	0.077	0.066	0.053	0.049

Note: “Bias” refers to median bias.

Table 7: AR(1). No Fixed Effects. $K = 10$. 2500 replications.

Logit MLE

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.008	0.019	0.012	−0.001	0.002	0.007	0.003	−0.001	0.000	0.001	0.000	−0.000
MAE	0.091	0.151	0.072	0.065	0.047	0.073	0.034	0.032	0.023	0.035	0.017	0.016

Logit MLE with Estimated Fixed Effects

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	−2.239	0.811	0.805	−0.002	−2.209	0.760	0.752	−0.000	−2.190	0.740	0.740	0.000
MAE	2.239	0.813	0.805	0.172	2.209	0.760	0.752	0.083	2.190	0.740	0.740	0.041

GMM

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.177	0.044	0.111	0.024	0.022	−0.060	−0.003	0.021	0.004	−0.063	−0.001	0.010
MAE	0.333	0.647	0.271	0.235	0.142	0.375	0.110	0.108	0.068	0.206	0.050	0.051

Note: "Bias" refers to median bias. The last column for each entry averages over the last eight explanatory variables

Table 8: AR(1). With Fixed Effects. $K = 10$. 2500 replications.

Logit MLE

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.755	0.327	−0.072	−0.001	0.747	0.323	−0.082	−0.000	0.744	0.314	−0.084	0.000
MAE	0.755	0.328	0.092	0.069	0.747	0.323	0.082	0.034	0.744	0.314	0.084	0.017

Logit MLE with Estimated Fixed Effects

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	−2.449	0.833	0.839	−0.002	−2.388	0.770	0.767	−0.001	−2.374	0.741	0.751	0.001
MAE	2.449	0.848	0.839	0.200	2.388	0.770	0.767	0.096	2.374	0.741	0.751	0.046

GMM

	$n = 500$				$n = 2000$				$n = 8000$			
	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$	γ	β_1	β_2	$\beta_{k\geq 3}$
True	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000
Bias	0.372	0.246	0.199	0.018	0.064	−0.041	0.003	0.019	0.010	−0.046	−0.004	0.010
MAE	0.490	0.780	0.347	0.289	0.172	0.417	0.128	0.120	0.084	0.226	0.060	0.057

Note: “Bias” refers to median bias. The last column for each entry averages over the last eight explanatory variables

Table 9: AR(2). No Fixed Effects. $K = 3$. 2500 replications.

Logit MLE

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.00	−0.00	0.00	0.01	0.00	0.00	−0.00	0.00	0.00	−0.00	0.00	−0.00	−0.00	0.00	0.00
MAE	0.08	0.08	0.07	0.06	0.06	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01

Logit MLE with Estimated Fixed Effects

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	−2.67	−1.73	0.06	0.07	−0.00	−2.66	−1.72	0.06	0.06	−0.00	−2.66	−1.72	0.06	0.06	0.00
MAE	2.67	1.73	0.16	0.13	0.12	2.66	1.72	0.09	0.07	0.06	2.66	1.72	0.06	0.06	0.03

GMM

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.10	0.06	0.08	0.07	−0.01	0.03	0.02	0.01	0.00	−0.00	0.01	0.01	−0.00	0.00	0.00
MAE	0.30	0.27	0.24	0.22	0.19	0.17	0.14	0.13	0.11	0.09	0.09	0.07	0.06	0.05	0.05

Note: "Bias" refers to median bias.

Table 10: AR(2). With Fixed Effects. $K = 3$. 2500 replications.

Logit MLE

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.72	0.70	0.24	−0.10	−0.00	0.71	0.70	0.23	−0.10	0.00	0.71	0.70	0.23	−0.10	0.00
MAE	0.72	0.70	0.24	0.10	0.06	0.71	0.70	0.23	0.10	0.03	0.71	0.70	0.23	0.10	0.02

Logit MLE with Estimated Fixed Effects

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	−2.82	−1.84	0.05	0.06	0.00	−2.79	−1.83	0.05	0.05	0.00	−2.79	−1.82	0.04	0.05	0.00
MAE	2.82	1.84	0.18	0.15	0.14	2.79	1.83	0.09	0.08	0.07	2.79	1.82	0.05	0.05	0.03

GMM

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.52	0.40	0.21	0.07	−0.04	0.11	0.09	0.03	−0.01	−0.02	0.02	0.02	0.00	0.00	−0.00
MAE	0.59	0.49	0.36	0.27	0.24	0.27	0.21	0.16	0.13	0.12	0.12	0.09	0.08	0.06	0.05

Note: “Bias” refers to median bias.

Table 11: AR(2). No Fixed Effects. $K = 10$. 2500 replications.

Logit MLE

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	−0.00	0.00	0.00	0.00	0.00	−0.00
MAE	0.08	0.08	0.13	0.06	0.06	0.04	0.04	0.06	0.03	0.03	0.02	0.02	0.03	0.01	0.01

Logit MLE with Estimated Fixed Effects

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	−2.71	−1.74	0.09	0.08	−0.00	−2.67	−1.72	0.07	0.07	0.00	−2.66	−1.72	0.06	0.06	−0.00
MAE	2.71	1.74	0.28	0.13	0.12	2.67	1.72	0.15	0.08	0.06	2.66	1.72	0.08	0.06	0.03

GMM

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.17	0.17	0.15	0.21	0.01	0.05	0.05	0.02	0.04	0.00	0.02	0.02	−0.01	0.01	0.00
MAE	0.35	0.32	0.44	0.28	0.22	0.18	0.15	0.26	0.12	0.11	0.10	0.08	0.15	0.05	0.05

Note: “Bias” refers to median bias.

Table 12: AR(2). With Fixed Effects. $K = 10$. 2500 replications.

Logit MLE

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.72	0.71	0.24	−0.10	0.00	0.71	0.71	0.23	−0.10	0.00	0.71	0.70	0.23	−0.10	0.00
MAE	0.72	0.71	0.24	0.10	0.06	0.71	0.71	0.23	0.10	0.03	0.71	0.70	0.23	0.10	0.02

Logit MLE with Estimated Fixed Effects

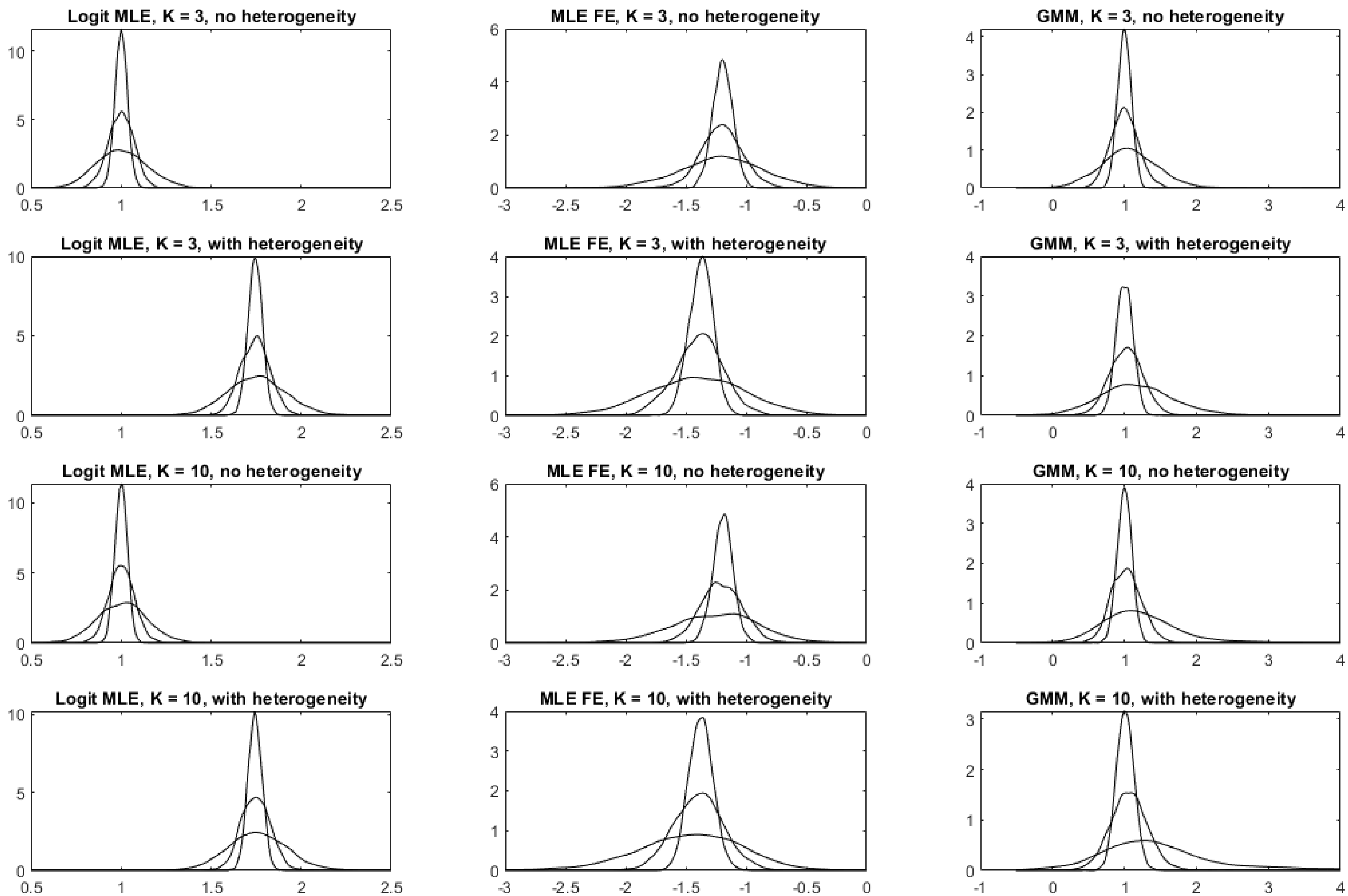
	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	−2.87	−1.86	0.10	0.07	−0.00	−2.81	−1.83	0.06	0.05	0.00	−2.79	−1.82	0.05	0.05	−0.00
MAE	2.87	1.86	0.33	0.16	0.14	2.81	1.83	0.16	0.08	0.07	2.79	1.82	0.09	0.05	0.03

GMM

	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	0.77	0.67	0.48	0.30	−0.01	0.26	0.23	0.19	0.04	−0.02	0.04	0.04	0.10	0.00	−0.01
MAE	0.79	0.69	0.64	0.38	0.29	0.32	0.27	0.36	0.15	0.13	0.15	0.11	0.24	0.07	0.07

Note: “Bias” refers to median bias.

Figure 2: Densities of Estimators of γ for $n = 500$, $n = 2000$, and $n = 8000$. The true value is 1



B.2 Additional proofs

B.2.1 Proof of Lemma 1

Lemma 1 is a special case of Theorem 2, which was already proven in Appendix A.1. Here, we provide an alternative proof of Lemma 1 by direct calculation. By plugging in the definition of the model probabilities $p(y, y_0, x, \beta_0, \gamma_0, \alpha)$ and the moment functions $m^{(q)}(y, y_0, x, \beta_0, \gamma_0)$ we want to verify that

$$\sum_{y \in \{0,1\}^3} p(y, y_0, x, \beta_0, \gamma_0, \alpha)\, m^{(q)}(y, y_0, x, \beta_0, \gamma_0) = 0,$$

for $q \in \{0, 1\}$. For $y_0 = 0$ and $q = 1$ we obtain

$$\begin{aligned}
\sum_{y \in \{0,1\}^3} p(y, 0, x, \beta_0, \gamma_0, \alpha)\, m^{(1)}(y, 0, x, \beta_0, \gamma_0) &= \frac{\exp\left(\zeta_1 - \zeta_2\right)}{[1+\exp(\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]} \\
&+ \frac{\exp\left(\zeta_1 - \zeta_3 - \gamma_0\right)}{[1+\exp(\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(-\zeta_3-\gamma_0)]} - \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]} \\
&+ \frac{\exp\left[(\zeta_3+\gamma_0) - (\zeta_2+\gamma_0)\right] - 1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2-\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]},
\end{aligned}$$

where $\zeta_t := x_t'\beta_0 + \alpha$, and we used that $\zeta_t - \zeta_s = x_{ts}'\beta_0$ and $(\zeta_3 + \gamma_0) - (\zeta_2 + \gamma_0) = x_{32}'\beta_0$. Simplifying the expression in the last display we obtain

$$\begin{aligned}
\sum_{y \in \{0,1\}^3} p(y, 0, x, \beta_0, \gamma_0, \alpha)\, m^{(1)}(y, 0, x, \beta_0, \gamma_0) &= \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]} \\
&+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]} - \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]} \\
&+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]\,[1+\exp(-\zeta_3-\gamma_0)]} \\
&- \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2-\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]},
\end{aligned}$$

where we used multiple times that $\exp(c)/[1 + \exp(c)] = 1/[1 + \exp(-c)]$, for $c \in \mathbb{R}$. The first two summands on the right hand side of the last display add up to

$$\frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_3+\gamma_0)]},$$

because $1/\left[1+\exp(\zeta_2)\right]+1/\left[1+\exp(-\zeta_2)\right]=1$. Subtracting the very last term in that right hand side expression gives

$$\begin{aligned}
&\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(\zeta_3+\gamma_0)\right]}-\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(-\zeta_2-\gamma_0)\right]\left[1+\exp(\zeta_3+\gamma_0)\right]}\\
&=\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(\zeta_2+\gamma_0)\right]\left[1+\exp(\zeta_3+\gamma_0)\right]},
\end{aligned}$$

because $1-1/\left[1+\exp(-\zeta_2-\gamma_0)\right]=1/\left[1+\exp(\zeta_2+\gamma_0)\right]$. We thus obtain

$$\begin{aligned}
\sum_{y\in\{0,1\}^3} & p(y,0,x,\beta_0,\gamma_0,\alpha)\, m^{(1)}(y,0,x,\beta_0,\gamma_0)=-\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(\zeta_2+\gamma_0)\right]}\\
&+\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(\zeta_2+\gamma_0)\right]\left[1+\exp(-\zeta_3-\gamma_0)\right]}\\
&+\frac{1}{\left[1+\exp(-\zeta_1)\right]\left[1+\exp(\zeta_2+\gamma_0)\right]\left[1+\exp(\zeta_3+\gamma_0)\right]}\\
&=0,
\end{aligned}$$

where we used that $-1+1/\left[1+\exp(-\zeta_3-\gamma_0)\right]+1/\left[1+\exp(\zeta_3+\gamma_0)\right]=0$. We have thus explicitly shown the statement of Lemma 1 for $y_0=0$ and $q=1$.

The results for other values of $y_0,q\in\{0,1\}$ can be derived analogously. However, once the result for $y_0=0$ and $q=1$ is derived, then there is actually no need for any further calculation. Instead, it suffices to notice that the model probabilities $p(y,y_0,x,\beta,\gamma,\alpha)$ are unchanged under the symmetry transformation

- $y_t \leftrightarrow 1-y_t, \quad x_t \leftrightarrow -x_t, \quad \beta \leftrightarrow \beta, \quad \gamma \leftrightarrow \gamma, \quad \alpha \leftrightarrow -\alpha-\gamma,$

and the same transformation applied to the moment function $m^{(q)}(y,y_0,x,\beta,\gamma)$ gives the moment function $m^{(1-q)}(y,1-y_0,x,\beta,\gamma)$. Furthermore, the model probabilities $p(y,y_0,x,\beta,\gamma,\alpha)$ are also unchanged under the transformation

- $y_0 \to 1-y_0\ , \quad x_1'\beta \to x_1'\beta+(2y_0-1)\gamma\ , \quad y_{t-1}$ and x_t unchanged for $t\geq 2$,

with parameters β, γ, α otherwise unchanged. Notice that for this symmetry transformation we need to consider $p(y,y_0,x,\beta,\gamma,\alpha)$ as a function of the product $x_t'\beta$, instead of x_t and β individually.

Applying this transformation to $m^{(q)}(y, y_0, x, \beta, \gamma)$ gives the moment function $m^{(q)}(y, 1-y_0, x, \beta, \gamma)$. By applying these symmetry transformation to our known result for $y_0 = 0$ and $q = 1$ we therefore obtain the result for all $y_0, q \in \{0, 1\}$. ■

B.3 Additional Material and Omitted Proofs for Section A.3

B.3.1 AR(p) models with $p \geq 2$, $T = 3$, and $x_2 = x_3$

Consider model (14) with $p \geq 2$ and $T = 3$ (i.e., $T_{\text{obs}} = 3 + p$ total periods). In this case, there are no moment conditions available that are valid for all possible realizations of the regressors $x \in \mathbb{R}^{K \times 3}$. However, for regressor realizations $x = (x_1, x_2, x_3)$ with $x_2 = x_3$, one finds valid moment conditions for the p-vectors of initial conditions $y^{(0)} = (y_{t_0}, \dots, y_0)$ that are constant over their last $p - 1$ elements. It is interesting that the condition $x_2 = x_3$ appears, since this is exactly the kind of condition that was used in Honoré and Kyriazidou (2000).

For $r \in \{1, 2, \dots\}$, let $0_r = (0, 0, \dots, 0)$ and $1_r = (1, 1, \dots, 1)$ be r-vectors with all entries equal to zero or one, respectively. Then, for $p \geq 2$, $T = 3$, and $x_2 = x_3$, we have one valid moment function $m_{y^{(0)}}(y, x, \beta, \gamma)$ for each of the initial conditions $y^{(0)} = 0_p$, $y^{(0)} = (0, 1_{p-1})$, $y^{(0)} = (1, 0_{p-1})$, and $y^{(0)} = 1_p$. They read

$$m_{(0_p)}(y, x, \beta, \gamma) = \begin{cases} \exp\left(x_{12}'\beta\right) & \text{if } y = (0, 1, 0), \\ \exp\left(x_{12}'\beta - \gamma_1\right) & \text{if } y = (0, 1, 1), \\ -1 & \text{if } (y_1, y_2) = (1, 0), \\ 0 & \text{otherwise}, \end{cases}$$

$$m_{(0,1_{p-1})}(y, x, \beta, \gamma) = \begin{cases} -1 & \text{if } (y_1, y_2) = (0, 1), \\ \exp\left(x_{21}'\beta - \gamma_1 + \gamma_p\right) & \text{if } y = (1, 0, 0), \\ \exp\left(x_{21}'\beta + \gamma_p\right) & \text{if } y = (1, 0, 1), \\ 0 & \text{otherwise}, \end{cases}$$

$$m_{(1,0_{p-1})}(y,x,\beta,\gamma) = \begin{cases} \exp\left(x_{12}'\beta + \gamma_p\right) & \text{if } y = (0,1,0), \\ \exp\left(x_{12}'\beta - \gamma_1 + \gamma_p\right) & \text{if } y = (0,1,1), \\ -1 & \text{if } (y_1,y_2) = (1,0), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_{(1_p)}(y,x,\beta,\gamma) = \begin{cases} -1 & \text{if } (y_1,y_2) = (0,1), \\ \exp\left(x_{21}'\beta - \gamma_1\right) & \text{if } y = (1,0,0), \\ \exp\left(x_{21}'\beta\right) & \text{if } y = (1,0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Here, the subscripts on $m_{y^{(0)}}$ denote the corresponding initial condition. Thus, for $p = 2$ we have one moment function available for each possible initial condition $y^{(0)} \in \{0,1\}^2$, and these moment functions together deliver information on all of the model parameters β and (γ_1, γ_2). By contrast, for $p > 2$ we only have moment functions for four out of 2^p possible initial conditions $y^{(0)} \in \{0,1\}^p$.

Lemma 7 *If the outcomes $Y = (Y_1, Y_2, Y_3)$ are generated from model (14) with $p \geq 2$, $T = 3$, and true parameters β_0 and γ_0, then we have for all $y^{(0)} \in \left\{0_p, (0, 1_{p-1}), (1, 0_{p-1}), 1_p\right\}$, $(x_1, x_2) \in \mathbb{R}^{K\times 2}$, and $\alpha \in \mathbb{R}$ that*

$$\mathbb{E}\left[m_{y^{(0)}}(Y, X, \beta_0, \gamma_0) \,\middle|\, Y^{(0)} = y^{(0)},\, X = (x_1, x_2, x_2),\, A = \alpha\right] = 0.$$

The proof of the lemma is given in Appendix B.3.7.

Identification: We now want to provide identification results for the parameters β and γ using the moment conditions in Lemma 7, analogous to the identification results in Theorem 1 for AR(1) models. For $p = 2$ all the parameters can be identified in this way. However, for $p \geq 3$ the moment conditions only contain the parameters β, γ_1 and γ_p, and we therefore only obtain an identification result for those parameters. To obtain moment conditions and identification results for $\gamma_2, \ldots, \gamma_{p-1}$, we need $T \geq 4$ (see Appendix B.3.8).

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

$$\mathcal{X}_{k,+} = \{x \in \mathbb{R}^{K\times 3} \,:\, x_{k,1} < x_{k,2}\}, \qquad \mathcal{X}_{k,-} = \{x \in \mathbb{R}^{K\times 3} \,:\, x_{k,1} > x_{k,2}\},$$

and for $s = (s_1, \ldots, s_K) \in \{-,+\}^K$ define $\mathcal{X}_s = \bigcap_{k\in\{1,\ldots,K\}} \mathcal{X}_{k,s_k}$.

Theorem 3 *Let the outcomes $Y = (Y_1, Y_2, Y_3)$ be generated from model (14) with $p \geq 2$, $T = 3$ and true parameters β_0 and γ_0.*

(i) Identification of β and γ_1: Let $y^{(0)} \in \{0_p, 1_p\}$. For all $\epsilon > 0$ and $s \in \{-,+\}^K$, assume that

$$\Pr\left(Y^{(0)} = y^{(0)},\ X \in \mathcal{X}_s,\ \|X_2 - X_3\| \leq \epsilon\right) > 0.$$

Also assume that the expectation in the following display is well-defined. Then,

$$\forall s \in \{-,+\}^K: \ \ \mathbb{E}\left[m_{y^{(0)}}(Y, X, \beta, \gamma) \,\middle|\, Y^{(0)} = y^{(0)},\ X \in \mathcal{X}_s,\ X_2 = X_3\right] = 0$$

if and only if $\beta = \beta_0$ and $\gamma_1 = \gamma_{0,1}$. Thus, the parameters β and γ_1 are point-identified under the assumptions provided here.

(ii) Identification of γ_p: Let $y^{(0)} \in \{(0, 1_{p-1}), (1, 0_{p-1})\}$. For all $\epsilon > 0$ assume that

$$\Pr\left(Y^{(0)} = y^{(0)} \quad \& \quad \|X_2 - X_3\| \leq \epsilon\right) > 0.$$

Also assume that the expectation in the following display is well-defined. Then,

$$\mathbb{E}\left[m_{y^{(0)}}(Y, X, \beta_0, (\gamma_{0,1}, \gamma_2, \ldots, \gamma_p)) \,\middle|\, Y^{(0)} = y^{(0)},\ X_2 = X_3\right] = 0$$

if and only if $\gamma_p = \gamma_{0,p}$. Thus, if the parameters β and γ_1 are point-identified, then γ_p is also point-identified under the assumptions provided here.

Proof. The proof is analogous to the proof of Theorem 3. Part (i) again follows by an application of Lemma 2. Part (ii) holds, because the expectations of $m_{(0,1_{p-1})}$ and $m_{(1,0_{p-1})}$ are strictly increasing in γ_p. ■

Comments on estimation: Because we only have moment conditions for $x_2 = x_3$, we are generally unable to estimate the model parameters at root-n rate here. However, for a model without regressors ($K = 0$), we can estimate γ_1 and γ_p at root-n rate using the moment conditions provided, as discussed in Section 4.2 for $p = 2$. More generally, for suitable discrete regressors one may also estimate β and γ at root-n rate.

B.3.2 Moment functions for the AR(2) model with $T = 4$

In Section A.3.1 we already presented $m^{(a,2,4)}_{y^{(0)}}(y, x, \beta, \gamma)$ as a valid moment function for the panel logit AR(2) model and $T = 4$. There are three additional such moment functions:

$$m^{(b,2,4)}_{y^{(0)}}(y, x, \beta, \gamma) = \begin{cases} \exp(z_{12}) & \text{if } y = (0,1,0,0), \\ \exp(z_{14})\big[1 + \exp(z_{32}) - \exp(z_{34})\big] & \text{if } y = (0,1,0,1), \\ \exp(z_{14} + \gamma_1) & \text{if } (y_1, y_2, y_3) = (0,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ \exp(z_{42}) - 1 & \text{if } y = (1,1,0,0), \\ \exp(z_{32}) - \exp(z_{34}) & \text{if } y = (1,1,0,1), \\ 0 & \text{otherwise,} \end{cases}$$

$$m^{(c,2,4)}_{y^{(0)}}(y, x, \beta, \gamma) = \begin{cases} [\exp(z_{24}) - 1]\,[1 - \exp(z_{34})] & \text{if } y = (0,0,0,1), \\ \exp(z_{24} + \gamma_1) - 1 & \text{if } (y_1, y_2, y_3) = (0,0,1), \\ -1 & \text{if } (y_1, y_2) = (0,1), \\ \exp(z_{41}) & \text{if } y = (1,0,0,0), \\ \exp(z_{21})\big[1 + \exp(z_{32}) - \exp(z_{34})\big] & \text{if } y = (1,0,0,1), \\ \exp(z_{21}) & \text{if } (y_1, y_2, y_3) = (1,0,1), \\ 0 & \text{otherwise,} \end{cases}$$

$$m^{(d,2,4)}_{y^{(0)}}(y, x, \beta, \gamma) = \begin{cases} \exp(z_{12}) & \text{if } (y_1, y_2, y_3) = (0,1,0), \\ \exp(z_{12})\big[1 + \exp(z_{23}) - \exp(z_{43})\big] & \text{if } y = (0,1,1,0), \\ \exp(z_{14}) & \text{if } y = (0,1,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ \exp(z_{42} + \gamma_1) - 1 & \text{if } (y_1, y_2, y_3) = (1,1,0), \\ [\exp(z_{42}) - 1]\,[1 - \exp(z_{43})] & \text{if } y = (1,1,1,0), \\ 0 & \text{otherwise.} \end{cases}$$

B.3.3 Proof of Lemma 5

Consider the initial conditions $y^{(0)} = (0,0)$. Then, by plugging in the definition of $z_{ts} = (x_{ts})' \beta + (y_{t-1} - y_{s-1}) \gamma_1 + (y_{t-2} - y_{s-2}) \gamma_2$, where $x_{ts} = x_t - x_s$, we obtain expressions for the moment functions that feature the parameters β and γ directly:

$$
m^{(a,2,4)}_{(0,0)}(y,x,\beta,\gamma) = \begin{cases}
\exp(x_{23}'\beta) - \exp(x_{43}'\beta + \gamma_1) & \text{if } y = (0,0,1,0), \\
\exp(x_{24}'\beta - \gamma_1) - 1 & \text{if } y = (0,0,1,1), \\
-1 & \text{if } (y_1, y_2) = (0,1), \\
\exp(x_{41}'\beta + \gamma_1) & \text{if } (y_1, y_2, y_3) = (1,0,0), \\
\exp(x_{41}'\beta + \gamma_1)\big[1 + \exp(x_{23}'\beta + \gamma_1 - \gamma_2) & \\
\qquad\qquad - \exp(x_{43}'\beta + \gamma_1 - \gamma_2)\big] & \text{if } y = (1,0,1,0), \\
\exp(x_{21}'\beta + \gamma_1) & \text{if } y = (1,0,1,1), \\
0 & \text{otherwise},
\end{cases}
$$

$$
m^{(b,2,4)}_{(0,0)}(y,x,\beta,\gamma) = \begin{cases}
\exp(x_{12}'\beta) & \text{if } y = (0,1,0,0), \\
\exp(x_{14}'\beta - \gamma_2)\big[1 + \exp(x_{32}'\beta + \gamma_1) & \\
\qquad\qquad - \exp(x_{34}'\beta + \gamma_1 - \gamma_2)\big] & \text{if } y = (0,1,0,1), \\
\exp(x_{14}'\beta - \gamma_2) & \text{if } (y_1, y_2, y_3) = (0,1,1), \\
-1 & \text{if } (y_1, y_2) = (1,0), \\
\exp(x_{42}'\beta - \gamma_1 + \gamma_2) - 1 & \text{if } y = (1,1,0,0), \\
\exp(x_{32}'\beta + \gamma_2) - \exp(x_{34}'\beta + \gamma_1) & \text{if } y = (1,1,0,1), \\
0 & \text{otherwise},
\end{cases}
$$

$$
m^{(c,2,4)}_{(0,0)}(y,x,\beta,\gamma) = \begin{cases}
[\exp(x_{24}'\beta)-1]\,[1-\exp(x_{34}'\beta)] & \text{if } y=(0,0,0,1),\\
\exp(x_{24}'\beta)-1 & \text{if } (y_1,y_2,y_3)=(0,0,1),\\
-1 & \text{if } (y_1,y_2)=(0,1),\\
\exp(x_{41}'\beta) & \text{if } y=(1,0,0,0),\\
\exp(x_{21}'\beta+\gamma_1)\big[1-\exp(x_{34}'\beta+\gamma_2) & \\
\qquad\qquad +\exp(x_{32}'\beta-\gamma_1+\gamma_2)\big] & \text{if } y=(1,0,0,1),\\
\exp(x_{21}'\beta+\gamma_1) & \text{if } (y_1,y_2,y_3)=(1,0,1),\\
0 & \text{otherwise,}
\end{cases}
$$

$$
m^{(d,2,4)}_{(0,0)}(y,x,\beta,\gamma) = \begin{cases}
\exp(x_{12}'\beta) & \text{if } (y_1,y_2,y_3)=(0,1,0),\\
\exp(x_{12}'\beta)\big[1+\exp(x_{23}'\beta-\gamma_1) & \\
\qquad\qquad -\exp(x_{43}'\beta+\gamma_2)\big] & \text{if } y=(0,1,1,0),\\
\exp(x_{14}'\beta-\gamma_1-\gamma_2) & \text{if } y=(0,1,1,1),\\
-1 & \text{if } (y_1,y_2)=(1,0),\\
\exp(x_{42}'\beta+\gamma_2)-1 & \text{if } (y_1,y_2,y_3)=(1,1,0),\\
[\exp(x_{42}'\beta+\gamma_2)-1]\,[1-\exp(x_{43}'\beta)] & \text{if } y=(1,1,1,0),\\
0 & \text{otherwise.}
\end{cases}
$$

Analogous to the proof of Lemma 1, one can now use these explicit expressions together with the definition of the model probabilities in (42) (for autoregressive order $p=2$) to verify by direct calculation that

$$
\sum_{y\in\{0,1\}^4} p_{(0,0)}(y,x,\beta_0,\gamma_0,\alpha)\, m^{(\xi,2,4)}_{(0,0)}(y,x,\beta_0,\gamma_0)=0, \tag{46}
$$

for $\xi\in\{a,b,c,d\}$. This calculation is straightforward, but lengthy, and we have used a computer algebra system (Mathematica) to verify this. Having thus derived the result for the initial conditions $y^{(0)}=(0,0)$, we note that both the model probabilities $p_{y^{(0)}}(y,x,\beta,\gamma,\alpha)$ and the moment functions $m^{(\xi,2,4)}_{y^{(0)}}(y,x,\beta,\gamma)$ are unchanged under the following transformation (with $p=2$ in our case)

(*) $y^{(0)} \to \widetilde{y}^{(0)}$, $\quad y_t$ unchanged for $t\geq 1$,

$x_t'\beta \to x_t'\beta+\sum_{r=t}^{p}(y_{t-r}-\widetilde{y}_{t-r})\gamma_t$ (for $t\leq p$), $\quad x_t'\beta \to x_t'\beta$ (for $t>p$), $\quad \beta\to\beta$,

$$\gamma \to \gamma, \qquad \alpha \to \alpha.$$

Here, we have transformed the initial conditions $y^{(0)}$ into arbitrary alternative initial conditions $\widetilde{y}^{(0)} \in \{0,1\}^p$ and adjusted $x_t'\beta$ such that the single index $z_t(y, y_0, x, \beta, \gamma) = x_t'\,\beta + y_{t-1}\,\gamma$ that enters into the model is unchanged for all $t \in \{1, \ldots, T\}$. Since the moment functions $m_{y^{(0)}}^{(\xi,2,4)}(y, x, \beta, \gamma)$ in Section A.3.1 are defined in terms of the single index $z_{ts} = z_t - z_s$ (and γ is unchanged), they are obviously unchanged under the transformation, and it is easy to see that the model probabilities $p_{y^{(0)}}(y, x, \beta, \gamma, \alpha)$ are unchanged as well. By applying the transformation (*) to (46) we therefore find that

$$\sum_{y \in \{0,1\}^4} p_{y^{(0)}}(y, x, \beta_0, \gamma_0, \alpha)\, m_{y^{(0)}}^{(\xi,2,4)}(y, x, \beta_0, \gamma_0) = 0$$

holds for all initial conditions $y^{(0)} \in \{0,1\}^2$. ■

B.3.4 Moment functions for the AR(3) model with $T = 5$

In Section A.3.2 we already presented $m_{y^{(0)}}^{(a,3,5)}(y, x, \beta, \gamma)$ as a valid moment function for the panel logit AR(3) model and $T = 5$. There are seven additional such moment functions:

$$m_{y^{(0)}}^{(b,3,5)}(y, x, \beta, \gamma) = \begin{cases} \exp(z_{12}) & \text{if } (y_1, y_2, y_3, y_4) = (0,1,0,0), \\ \exp(z_{12}) + \exp(z_{14}) + \exp(z_{12} + z_{34}) & \\ \qquad - \exp(z_{14} + z_{35}) - \exp(z_{12} + z_{54}) & \text{if } y = (0,1,0,1,0), \\ \exp(z_{15})\left(\exp(z_{32}) - \exp(z_{35}) + 1\right) & \text{if } y = (0,1,0,1,1), \\ \exp(-\gamma_1 + \gamma_2 + z_{15}) & \text{if } (y_1, y_2, y_3, y_4) = (0,1,1,0), \\ \exp(\gamma_2 + z_{15}) & \text{if } (y_1, y_2, y_3, y_4) = (0,1,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ \exp(\gamma_1 + z_{52}) - 1 & \text{if } (y_1, y_2, y_3, y_4) = (1,1,0,0), \\ \left(\exp(z_{34}) - \exp(z_{54}) + 1\right) & \\ \qquad \left(\exp(z_{52}) - 1\right) & \text{if } y = (1,1,0,1,0), \\ \exp(z_{32}) - \exp(z_{35}) & \text{if } y = (1,1,0,1,1), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_{y^{(0)}}^{(c,3,5)}(y,x,\beta,\gamma)=\begin{cases}-\exp\left(z_{54}\right)\left(1-\exp\left(z_{25}\right)\right)\left(1-\exp\left(z_{35}\right)\right) & \text{if } y=(0,0,0,1,0),\\ \left(\exp\left(z_{25}\right)-1\right)\left(1-\exp\left(z_{35}\right)\right) & \text{if } y=(0,0,0,1,1),\\ \exp\left(-\gamma_1+\gamma_2+z_{25}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(0,0,1,0),\\ \exp\left(\gamma_2+z_{25}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(0,0,1,1),\\ -1 & \text{if } (y_1,y_2)=(0,1),\\ \exp\left(\gamma_1+z_{51}\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,0,0,0),\\ -\exp\left(z_{21}+z_{34}\right)+\exp\left(z_{51}\right)+\exp\left(z_{21}+z_{54}\right) & \\ \qquad+\exp\left(z_{31}+z_{54}\right)-\exp\left(z_{51}+z_{54}\right) & \text{if } y=(1,0,0,1,0),\\ \exp\left(z_{21}\right)+\exp\left(z_{31}\right)-\exp\left(z_{21}+z_{35}\right) & \text{if } y=(1,0,0,1,1),\\ \exp\left(z_{21}\right) & \text{if } (y_1,y_2,y_3)=(1,0,1),\\ 0 & \text{otherwise,}\end{cases}$$

$$m_{y^{(0)}}^{(d,3,5)}(y,x,\beta,\gamma)=\begin{cases}\exp\left(z_{12}\right) & \text{if } (y_1,y_2,y_3)=(0,1,0,0),\\ \exp\left(z_{12}\right)+\exp\left(z_{13}\right)-\exp\left(z_{12}+z_{53}\right) & \text{if } y=(0,1,1,0,0),\\ \exp\left(z_{15}\right)-\exp\left(z_{12}+z_{43}\right)+\exp\left(z_{12}+z_{45}\right) & \\ \qquad+\exp\left(z_{13}+z_{45}\right)-\exp\left(z_{15}+z_{45}\right) & \text{if } y=(0,1,1,0,1),\\ \exp\left(\gamma_1+z_{15}\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,1,1,1),\\ -1 & \text{if } (y_1,y_2)=(1,0),\\ \exp\left(\gamma_2+z_{52}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(1,1,0,0),\\ \exp\left(-\gamma_1+\gamma_2+z_{52}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(1,1,0,1),\\ \left(\exp\left(z_{52}\right)-1\right)\left(1-\exp\left(z_{53}\right)\right) & \text{if } y=(1,1,1,0,0),\\ -\exp\left(z_{45}\right)\left(1-\exp\left(z_{52}\right)\right)\left(1-\exp\left(z_{53}\right)\right) & \text{if } y=(1,1,1,0,1),\\ 0 & \text{otherwise,}\end{cases}$$

$$m_{y^{(0)}}^{(e,3,5)}(y,x,\beta,\gamma)=\begin{cases}\exp\left(z_{23}\right)-\exp\left(\gamma_1+z_{53}\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,0,1,0),\\ \left(\exp\left(z_{23}\right)-\exp\left(z_{53}\right)\right)\left(\exp\left(z_{34}\right)-\exp\left(z_{54}\right)+1\right) & \text{if } y=(0,0,1,1,0),\\ \exp\left(z_{25}\right)-1 & \text{if } y=(0,0,1,1,1),\\ -1 & \text{if } (y_1,y_2)=(0,1),\\ \exp\left(\gamma_1+\gamma_2+z_{51}\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,0,0,0),\\ \exp\left(\gamma_2+z_{51}\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,0,0,1),\\ \exp\left(\gamma_1+z_{51}\right)\left(-\exp\left(\gamma_1+z_{53}\right)+\exp\left(z_{23}\right)+1\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,0,1,0),\\ \exp\left(z_{51}\right)\Big(\exp\left(z_{23}\right)+\exp\left(z_{24}\right)-\exp\left(z_{53}\right) & \\ -\exp\left(z_{54}\right)-\exp\left(z_{23}+z_{54}\right)+\exp\left(z_{53}+z_{54}\right)+1\Big) & \text{if } y=(1,0,1,1,0),\\ \exp\left(z_{21}\right) & \text{if } y=(1,0,1,1,1),\\ 0 & \text{otherwise,}\end{cases}$$

$$m_{y^{(0)}}^{(f,3,5)}(y,x,\beta,\gamma)=\begin{cases}\exp\left(z_{12}\right) & \text{if } y=(0,1,0,0,0),\\ \exp\left(z_{15}\right)\Big(\exp\left(z_{32}\right)-\exp\left(z_{35}\right)+\exp\left(z_{42}\right) & \\ -\exp\left(z_{45}\right)-\exp\left(z_{32}+z_{45}\right)+\exp\left(z_{35}+z_{45}\right)+1\Big) & \text{if } y=(0,1,0,0,1),\\ \exp\left(\gamma_1+z_{15}\right)\left(-\exp\left(\gamma_1+z_{35}\right)+\exp\left(z_{32}\right)+1\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,1,0,1),\\ \exp\left(\gamma_2+z_{15}\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,1,1,0),\\ \exp\left(\gamma_1+\gamma_2+z_{15}\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,1,1,1),\\ -1 & \text{if } (y_1,y_2)=(1,0),\\ \exp\left(z_{52}\right)-1 & \text{if } y=(1,1,0,0,0),\\ \left(\exp\left(z_{32}\right)-\exp\left(z_{35}\right)\right)\left(e^{z_{43}}-\exp\left(z_{45}\right)+1\right) & \text{if } y=(1,1,0,0,1),\\ \exp\left(z_{32}\right)-\exp\left(\gamma_1+z_{35}\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,1,0,1),\\ 0 & \text{otherwise,}\end{cases}$$

$$m_{y^{(0)}}^{(g,3,5)}(y,x,\beta,\gamma) = \begin{cases} \left(\exp\left(z_{25}\right)-1\right)\left(1-\exp\left(z_{35}\right)\right)\left(1-\exp\left(z_{45}\right)\right) & \text{if } y=(0,0,0,0,1), \\ \left(\exp\left(\gamma_1+z_{25}\right)-1\right)\left(1-\exp\left(\gamma_1+z_{35}\right)\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,0,0,1), \\ \exp\left(\gamma_2+z_{25}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(0,0,1,0), \\ \exp\left(\gamma_1+\gamma_2+z_{25}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(0,0,1,1), \\ -1 & \text{if } (y_1,y_2)=(0,1), \\ \exp\left(z_{51}\right) & \text{if } y=(1,0,0,0,0), \\ \exp\left(z_{21}\right)+\exp\left(z_{31}\right)-\exp\left(z_{21}+z_{35}\right)+\exp\left(z_{41}\right) & \\ \qquad -\exp\left(z_{21}+z_{45}\right)-\exp\left(z_{31}+z_{45}\right) & \\ \qquad +\exp\left(z_{21}+z_{35}+z_{45}\right) & \text{if } y=(1,0,0,0,1), \\ -\exp\left(\gamma_1+z_{21}+z_{35}\right)+\exp\left(z_{21}\right)+\exp\left(z_{31}\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,0,0,1), \\ \exp\left(z_{21}\right) & \text{if } (y_1,y_2,y_3)=(1,0,1), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$m_{y^{(0)}}^{(h,3,5)}(y,x,\beta,\gamma) = \begin{cases} \exp\left(z_{12}\right) & \text{if } (y_1,y_2,y_3)=(0,1,0), \\ -\exp\left(\gamma_1+z_{12}+z_{53}\right)+\exp\left(z_{12}\right)+\exp\left(z_{13}\right) & \text{if } (y_1,y_2,y_3,y_4)=(0,1,1,0), \\ \exp\left(z_{12}\right)+\exp\left(z_{13}\right)+\exp\left(z_{14}\right)-\exp\left(z_{12}+z_{53}\right) & \\ \qquad -\exp\left(z_{12}+z_{54}\right)-\exp\left(z_{13}+z_{54}\right) & \\ \qquad +\exp\left(z_{12}+z_{53}+z_{54}\right) & \text{if } y=(0,1,1,1,0), \\ \exp\left(z_{15}\right) & \text{if } y=(0,1,1,1,1), \\ -1 & \text{if } (y_1,y_2)=(1,0), \\ \exp\left(\gamma_1+\gamma_2+z_{52}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(1,1,0,0), \\ \exp\left(\gamma_2+z_{52}\right)-1 & \text{if } (y_1,y_2,y_3,y_4)=(1,1,0,1), \\ \left(\exp\left(\gamma_1+z_{52}\right)-1\right)\left(1-\exp\left(\gamma_1+z_{53}\right)\right) & \text{if } (y_1,y_2,y_3,y_4)=(1,1,1,0), \\ \left(\exp\left(z_{52}\right)-1\right)\left(1-\exp\left(z_{53}\right)\right)\left(1-\exp\left(z_{54}\right)\right) & \text{if } y=(1,1,1,1,0), \\ 0 & \text{otherwise.} \end{cases}$$

B.3.5 Proof of Lemma 6

This proof is analogous to the proof of Lemma 5.

B.3.6 AR(2) models with $T = 5$

Consider model (14) with $p = 2$ and $T = 5$, where $\gamma \in \mathbb{R}^2$, $y^{(0)} = (y_{-1}, y_0) \in \{0,1\}^2$, $y = (y_1, y_2, y_3, y_4, y_5) \in \{0,1\}^5$, and $x = (x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^{K\times 5}$. Using the results from Section A.3.1 and A.3.2 we can immediately construct valid moment functions for this model as well. Firstly, by using the (time-shifted) AR(2) moments for $T = 4$ we define

$$
\begin{aligned}
m_{y^{(0)}}^{(\xi,2,5)}(y, x, \beta, \gamma) &:= m_{y^{(0)}}^{(\xi,2,4)}((y_1, y_2, y_3, y_4), (x_1, x_2, x_3, x_4), \beta, \gamma),\\
\breve{m}_{y^{(0)}}^{(\xi,2,5)}(y, x, \beta, \gamma) &:= \mathbb{1}\{y_1 = 0\}\; m_{(y_0,y_1)}^{(\xi,2,4)}((y_2, y_3, y_4, y_5), (x_2, x_3, x_4, x_5), \beta, \gamma),\\
\widetilde{m}_{y^{(0)}}^{(\xi,2,5)}(y, x, \beta, \gamma) &:= \mathbb{1}\{y_1 = 1\}\; m_{(y_0,y_1)}^{(\xi,2,4)}((y_2, y_3, y_4, y_5), (x_2, x_3, x_4, x_5), \beta, \gamma),
\end{aligned}
$$

where $\xi \in \{a, b, c, d\}$. These are twelve valid moment functions for the AR(2) model with $T = 5$, because the model probabilities are invariant under time-shifts. Secondly, we can use our AR(3) moments with $T = 5$ to define

$$
\ddot{m}_{y^{(0)}}^{(\xi,2,5)}(y, x, \beta, \gamma) := m_{(0,y^{(0)})}^{(\xi,3,5)}(y, x, \beta, (\gamma_1, \gamma_2, 0)),
$$

where $\xi \in \{a, b, c, d, e, f, g, h\}$. We thus obtain eight valid moment functions for the AR(2) model with $T = 5$, because the AR(2) model is a special case of the AR(3) model with $\gamma_3 = 0$.

These are twenty valid moment functions in total for $p = 2$ and $T = 5$. However, not all of them are linearly independent. One finds four linear dependencies:

$$
\begin{aligned}
& e^{y_0\gamma_1 + y_{-1}\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(c,2,5)} - \breve{m}_{y^{(0)}}^{(a,2,5)}\right) + e^{x_{25}'\beta + (y_0-1)\gamma_1 + (y_{-1}+y_0)\gamma_2}\; \breve{m}_{y^{(0)}}^{(a,2,5)}\\
&\qquad + e^{\gamma_1 + x_{51}'\beta}\left(\ddot{m}_{y^{(0)}}^{(b,2,5)} - \widetilde{m}_{y^{(0)}}^{(a,2,5)}\right) + e^{\gamma_1 + x_{21}'\beta + y_0\gamma_2}\, \widetilde{m}_{y^{(0)}}^{(a,2,5)} = 0,\\
& e^{x_{25}'\beta + (y_0+1)\gamma_1 + (y_{-1}+y_0-1)\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(a,2,5)} - \breve{m}_{y^{(0)}}^{(b,2,5)}\right) + e^{(y_0+1)\gamma_1 + y_{-1}\gamma_2}\; \breve{m}_{y^{(0)}}^{(b,2,5)}\\
&\qquad + e^{\gamma_1 + x_{21}'\beta + y_0\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(d,2,5)} - \widetilde{m}_{y^{(0)}}^{(b,2,5)}\right) + e^{\gamma_2 + x_{51}'\beta}\widetilde{m}_{y^{(0)}}^{(b,2,5)} = 0,\\
& e^{y_0\gamma_1 + y_{-1}\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(g,2,5)} - \breve{m}_{y^{(0)}}^{(c,2,5)}\right) + e^{x_{25}'\beta + y_0\gamma_1 + (y_{-1}+y_0)\gamma_2}\; \breve{m}_{y^{(0)}}^{(c,2,5)}\\
&\qquad + e^{x_{51}'\beta}\left(\ddot{m}_{y^{(0)}}^{(f,2,5)} - \widetilde{m}_{y^{(0)}}^{(c,2,5)}\right) + e^{\gamma_1 + x_{21}'\beta + y_0\gamma_2}\, \widetilde{m}_{y^{(0)}}^{(c,2,5)} = 0,\\
& e^{x_{25}'\beta + (y_0-1)\gamma_1 + (y_{-1}+y_0-1)\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(e,2,5)} - \breve{m}_{y^{(0)}}^{(d,2,5)}\right) + e^{y_0\gamma_1 + y_{-1}\gamma_2}\; \breve{m}_{y^{(0)}}^{(d,2,5)}
\end{aligned}
$$

$$+e^{x_{21}'\beta+y_0\gamma_2}\left(\ddot{m}_{y^{(0)}}^{(h,2,5)}-\widetilde{m}_{y^{(0)}}^{(d,2,5)}\right)+e^{\gamma_2+x_{51}'\beta}\,\widetilde{m}_{y^{(0)}}^{(d,2,5)}=0,$$

where the arguments (y, x, β, γ) on all the moment functions were omitted. Using those relations we can, for example, express all the $\widetilde{m}_{y^{(0)}}^{(\xi,2,5)}$, $\xi \in \{a, b, c, d\}$, in terms of the other sixteen moment functions. Thus, by dropping all the $\widetilde{m}_{y^{(0)}}^{(\xi,2,5)}$ we obtain one possible set of irreducible moment conditions for the AR(2) model at $T = 5$. The total number of linearly independent moment conditions available for $p = 2$ and $T = 5$ is therefore equal to $\ell = 16$, in agreement with equation (43).

B.3.7 Proof of Lemma 7

Analogous to the Proof of Lemma 1 one can verify by direct calculation that for AR(p) model with $p \in \{2, 3\}$ we have

$$\sum_{y\in\{0,1\}^3} p_{y^{(0)}}(y, x, \beta_0, \gamma_0, \alpha)\; m_{y^{(0)}}(y, x, \beta_0, \gamma_0) = 0,$$

for all $x = (x_1, x_2, x_2)$ and $y^{(0)} \in \left\{0_p, (0, 1_{p-1}), (1, 0_{p-1}), 1_p\right\}$. Thus, the statement of the lemma is true for $p \in \{2, 3\}$. Next, using the definition of $p_{y^{(0)}}(y, x, \beta, \gamma, \alpha)$ in (14), one can verify that the model probabilities for $p \geq 4$ can be expressed in terms of the probabilities for the AR(3) model as follows:

$$\begin{aligned}
p_{(y_*,0_{p-1})}(y, x, \beta, \gamma, \alpha) &= p_{(y_*,0,0)}(y, x, \beta, (\gamma_1, \gamma_2, \gamma_p), \alpha),\\
p_{(y_*,1_{p-1})}(y, x, \beta, \gamma, \alpha) &= p_{(y_*,1,1)}\left(y, x, \beta, (\gamma_1, \gamma_2, \gamma_p), \alpha + \sum_{r=3}^{p-1}\gamma_r\right),
\end{aligned}$$

where $y_* \in \{0, 1\}$ denotes the value of the first observed outcome y_{t_0} for time period $t_0 = 1 - p$. Thus, since the lemma holds for $p = 3$ and for all values of α, and since the moment functions for $p \geq 4$ are obtained from those for $p = 3$ by replacing γ_3 by γ_p, we conclude that the lemma also holds for $p \geq 4$. ■

B.3.8 Results for AR(p) model with $p \geq 3$, $T = 4$ and $x_3 = x_4$

In Section B.3.1 we obtained identification results for the parameters β, γ_1, and γ_p for AR(p) models with $p \geq 3$. Here, we explain how γ_2 and γ_{p-2} can also be identified if data for $T = 4$

(i.e., $T_{\text{obs}} = 4 + p$) time periods are available. We consider moment conditions that are valid conditional on $X_3 = X_4$. With this, three valid moment functions are available for the p-vectors of initial conditions $y^{(0)} = (y_{t_0}, \ldots, y_0)$ that are constant over their last $p-2$ elements. No moment conditions are available for other initial conditions. For $y^{(0)} = 0_p$, the first of these three valid moment functions is simply obtained by shifting the corresponding moment function for $T = 3$ in Section B.3.1 by one time period. For $x = (x_1, x_2, x_3, x_3)$, we have[29]

$$m_{0_p}^{(a,p,4)}(y, x, \beta, \gamma) = \mathbb{1}(y_1 = 0)\, m_{(0_p)}((y_2, y_3, y_4), (x_2, x_3, x_3), \beta, \gamma).$$

The second valid moment function is obtained from $m_{(0,0)}^{(d,2,4)}(y, x, \beta, \gamma)$ for $p = 2$ from Section A.3.1. We have

$$m_{0_p}^{(b,p,4)}(y, x, \beta, \gamma) = m_{(0,0)}^{(d,2,4)}(y, (x_1, x_2, x_3, x_3), \beta, (\gamma_1, \gamma_2)),$$

but there are some simplifications to this moment function here since $x_3 = x_4$. None of the other moment functions from Section A.3.1 (and none of their linear combinations) can be lifted to become a moment function for $p \geq 3$; only $m_{(0,0)}^{(d,2,4)}$. Finally, a third valid moment function for $p \geq 3$, $T = 4$, $x_3 = x_4$, and $y^{(0)} = 0_p$ is given by

$$m_{0_p}^{(c,p,4)}(y, x, \beta, \gamma) = \begin{cases} -\exp(\gamma_1) & \text{if } y = (0,0,1,0), \\ -1 & \text{if } y = (0,0,1,1), \\ \exp(x_{32}'\beta)\,[\exp(\gamma_1) - \exp(\gamma_2)] & \text{if } y = (0,1,0,0), \\ \exp(\gamma_1 - \gamma_2) - 1 & \text{if } y = (0,1,0,1), \\ -1 & \text{if } (y_1, y_2, y_3) = (0,1,1), \\ \exp(x_{31}'\beta + \gamma_2) & \text{if } (y_1, y_2, y_3) = (1,0,0), \\ \exp(x_{21}'\beta + \gamma_1) & \text{if } (y_1, y_2, y_3) = (1,0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 8 *If the outcomes $Y = (Y_1, Y_2, Y_3, Y_4)$ are generated from model (14) with $p \geq 3$, $T = 4$ and true parameters β_0 and γ_0, then we have for all $(x_1, x_2, x_3) \in \mathbb{R}^{K \times 3}$, $\alpha \in \mathbb{R}$, and $\xi \in \{a, b, c\}$*

[29] Obviously, we will not get any new identifying information from that moment function, beyond what was already discussed in Section B.3.1, but we list it here for completeness.

that

$$\mathbb{E}\left[m_{0_p}^{(\xi,p,4)}(Y,X,\beta_0,\gamma_0)\,\middle|\,Y^{(0)}=0_p,\ X=(x_1,x_2,x_3,x_3),\ A=\alpha\right]=0.$$

Proof. Analogous to the proof of Lemma 7. ■

Notice that the moment functions $m_{0_p}^{(b,p,4)}$ and $m_{0_p}^{(c,p,4)}$ contain the parameter γ_2. Similarly, one can obtain valid moment functions for all of the eight initial conditions $y^{(0)} \in \{0_p, (1,0_{p-1}), (1,1,0_{p-2}), (0,1,0_{p-2}), (0,0,1_{p-2}), (1,0,1_{p-2}), (0,1_{p-1}), 1_p\}$, and some of these also contain the parameter γ_{p-1}. However, for $p \geq 5$ none of these moment functions contain the parameters $\gamma_3, \ldots, \gamma_{p-2}$. One requires $T \geq 5$ to identifythose parameters using moment conditions like the ones in this paper. We will not discuss this here.

One example of a moment function that features the parameters γ_{p-2} is given by

$$m_{(0,1,0_{p-2})}^{(c,p,4)}(y,x,\beta,\gamma)=\begin{cases} -\exp(\gamma_1) & \text{if } y=(0,0,1,0),\\ -1 & \text{if } y=(0,0,1,1),\\ \exp(x_{32}'\beta-\gamma_p)\left[\exp(\gamma_1)-\exp(\gamma_2)\right] & \text{if } y=(0,1,0,0),\\ \exp(\gamma_1-\gamma_2)-1 & \text{if } y=(0,1,0,1),\\ -1 & \text{if } (y_1,y_2,y_3)=(0,1,1),\\ \exp(x_{31}'\beta+\gamma_2-\gamma_{p-1}) & \text{if } (y_1,y_2,y_3)=(1,0,0),\\ \exp(x_{21}'\beta+\gamma_1-\gamma_{p-1}+\gamma_p) & \text{if } (y_1,y_2,y_3)=(1,0,1),\\ 0 & \text{otherwise,} \end{cases}$$

which is a valid moment function for $p \geq 3$, $T=4$, $x_3=x_4$ and $y^{(0)}=(0,1,0_{p-2})$.

Notice that for $p=3$ we have $\gamma_2=\gamma_{p-1}$, which leads to a simplification in $m_{(0,1,0_{p-2})}^{(c,p,4)}$, since $\gamma_2=\gamma_{p-1}$ drops out of the entry for $(y_1,y_2,y_3)=(1,0,0)$. All other elements of this moment function then are either independent of $\gamma_2=\gamma_{p-1}$ or are strictly decreasing in $\gamma_2=\gamma_{p-1}$. Thus, for $p=3$ the moment function $m_{(0,1,0_{p-2})}^{(c,p,4)}$ can be used to identify $\gamma_2=\gamma_{p-1}$ uniquely. The following theorem formalizes this result.

Theorem 4 *Let the outcomes $Y=(Y_1,Y_2,Y_3,Y_4)$ be generated from (14) with $p=3$, $T=4$, and* *true parameters* β_0 *and* γ_0*. For all* $\epsilon > 0$*, assume that*

$$\Pr\left(Y^{(0)} = (0,1,0),\ \|X_3 - X_4\| \leq \epsilon\right) > 0.$$

Assume that the expectation in the following display is well-defined. Then, we have

$$\mathbb{E}\left[m^{(c,p,4)}_{(0,1,0)}(Y, X, \beta_0, (\gamma_{0,1}, \gamma_2, \gamma_{0,3}))\,\middle|\, Y^{(0)} = (0,1,0),\ X_3 = X_4\right] = 0$$

if and only if $\gamma_2 = \gamma_{0,2}$*. Thus, if the parameters* β*,* γ_1*, and* γ_3 *are point-identified, then* γ_2 *is also point-identified under the assumptions provided here.*

Proof. The result follows since $\mathbb{E}\left[m^{(c,p,4)}_{(0,1,0)}(Y, X, \beta_0, (\gamma_{0,1}, \gamma_2, \gamma_{0,3}))\,\middle|\, Y^{(0)} = (0,1,0),\ X_3 = X_4\right]$ is strictly decreasing in γ_2, and is equal to zero at $\gamma_2 = \gamma_{0,2}$. ■

Thus, together with the results in Theorem 3 we have provided conditions under which all the parameters β, γ of a panel logit AR(3) model are point-identified.