arXiv:2005.05942v4 [econ.EM] 19 Dec 2021

Moment Conditions for Dynamic Panel Logit Models with Fixed Effects*

Bo E. Honoré‡ Martin Weidner§

December 2021

Abstract

This paper builds on Bonhomme (2012) to develop a method to systematically construct moment conditions for dynamic panel data logit models with fixed effects. After introducing the moment conditions obtained in this way, we explore their implications for identification and estimation of the model parameters that are common to all individuals, and we find that those common model parameters are estimable at root-n rate for many more dynamic panel logit models than has been appreciated by the existing literature. In the case where the model contains one lagged variable, the moment conditions in Kitazawa (2013, 2016) are transformations of a subset of ours. A GMM estimator that is based on the moment conditions is shown to perform well in Monte Carlo simulations and in an empirical illustration to labor force participation.

*We thank Stephane Bonhomme, Geert Dhaene, Chris Muris, Shakeeb Khan, Yoshitsugu Kitazawa, and Whitney Newey 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 (Grant Number SES-1530741), by the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice (grant numbers RES-589-28-0001, RES-589-28-0002 and ES/P008909/1), 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 College London, m.weidner@ucl.ac.uk

1 Introduction

This paper revisits the problem of estimating the parameters of a binary logit model with fixed effects and lagged dependent variables with short panels. It is notoriously difficult to estimate the parameters of even standard nonlinear panel data models if one wants to allow for an individual–specific constant, akin to a fixed effect in a linear panel data model. This is especially true for models that include lagged outcomes as explanatory variables.

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 a sufficient statistic for the fixed effects.[1] Manski (1987) showed that it is possible to identify and consistently estimate a semiparametric version of the same 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.

A number of papers have attempted to relax the assumption that the explanatory variables are strictly exogenous by including lagged dependent variables. This is motivated by the desire to distinguish between state dependence and unobserved heterogeneity (see e.g. Heckman 1978, 1981c,b,a). 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 effect, but which may 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

[1] Andersen (1970) studied the asymptotic properties of the resulting estimator. Hausman, Hall, and Griliches (1984) used a similar strategy to estimate fixed effects panel data Poisson regression models.

to logit models that have both lagged dependent variables and strictly exogenous explanatory variables. However, as shown in Honoré and Kyriazidou (2000) and D'Addio and Honoré (2010), this approach does apply if one is also willing to condition on the vector of covariates being equal across certain time periods. The resulting estimator 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.[2] This motivates the estimation approach based on a logistic distribution assumption that we follow in the current paper.

A recent paper, Honoré and Kyriazidou (2019), considers identification of a panel data logit model with fixed effects and two lagged outcomes as explanatory variables. This model was also considered by Chamberlain (1985), who showed that if one conditions on a sufficient statistic for the fixed effect, the resulting conditional likelihood depends on the coefficient on the second lag, but not on the coefficient on the first lag. Despite this, the numerical calculations in Honoré and Kyriazidou (2019) found that both parameters are identified. Needless to say, it is possible that the parameters are indeed not identified, but that the identified set is so small that it appears numerically as if it is point identified. The other natural explanation (other than the possibility that the calculations in Honoré and Kyriazidou (2019) are incorrect) is that the parameters are indeed point identified, but not through a conditional likelihood approach.

This paper's contribution is to explore a method for systematically deriving moment conditions for panel data logit models with fixed effects that are not based on conditioning on a sufficient statistic for the fixed effect. We do this following a line of argument suggested in Bonhomme (2012). We give sufficient conditions for these mo-

[2]Or, alternatively, change the model as in Bartolucci and Nigro (2010) and Al-Sadoon, Li, and Pesaran (2017).

ment conditions to actually identify the parameters of the model. Subject to standard regularity conditions, the usual GMM theory then delivers root-n asymptotic normality of the estimator.[3]

Transformations of some of the moment conditions for the dynamic panel logit model presented in this paper have previously been derived by Kitazawa (2013, 2016), who also used them to estimate the model parameters at root-n rate via GMM.[4] The way in which we derive and present our moment conditions and the way we explore the implications for identification are quite different from Kitazawa's. Using our methods we also find various new moment conditions for panel logit AR(1), if the number of time periods is large enough, and for panel logit AR(p) models with $p \geq 2$.

Our results relate to a larger literature on estimating nonlinear panel data models with fixed effects and shorter panels.[5] Censored regression was studied by Honoré (1992) for the static model and 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 shorter panels include Abrevaya (1999) and Abrevaya (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

[3]Hahn (2001) showed that root-n estimation is sometimes not possible in dynamic logit models with fixed effects. By contrast, our result show that sometimes root-n estimation is possible in such models. However, those results are not inconsistent, see our discussion in Section 2.1.3.

[4]See also Kitazawa (2017) for an implementation of his moment conditions. We were not aware of the papers by Yoshitsugu Kitazawa when we wrote the initial draft of our paper, and we are thankful to him for pointing out the connection.

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

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 constructing moment conditions in dynamic panel data logit models by considering a logit AR(1) model. We discuss the derivation of the resulting moment conditions, their relation to the literature, and their implications for identification and estimation in some detail. Section 3 demonstrates how the approach can be used to construct moment conditions for panel logit AR(p) models of order p larger than one. Sections 4 and 5 present Monte Carlo evidence and an empirical illustration to labor force participation. Section 6 concludes the paper.

2 The fixed effect logit AR(1) model

We start by investigating a panel data logit model with fixed effects, one lagged dependent variable, and a set of strictly exogenous explanatory variables. The aim is to construct moment conditions based on a fixed (small) number of time periods.

We observe outcomes $Y_{it} \in \{0,1\}$ and strictly exogenous regressors $X_{it} \in \mathbb{R}^K$ for individuals $i = 1, \ldots, n$ over time periods $t = 0, \ldots, T$. The total number of time periods for which outcomes are observed is $T_{\text{obs}} = T+1$. We assume that the distribution of Y_{it}, conditional on the regressors $X_i = (X_{i1}, \ldots, X_{iT})$ and past outcomes $Y_i^{t-1} = (Y_{i,t-1}, Y_{i,t-2}, \ldots)$, is described by the logistic single index model

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

for $i \in \{1, \ldots, n\}$ and $t \in \{1, 2, \ldots, T\}$. Here $A_i \in \mathbb{R}$ are unobserved fixed effects, while $\beta \in \mathbb{R}^K$ and $\gamma \in \mathbb{R}$ are unknown parameters that are common to all individuals. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})$, $X_i = (X_{i1}, \ldots, X_{iT})$, and let the true model parameters be denoted by β_0 and γ_0. In the following, all probabilistic statements are for the model distribution generated under β_0 and γ_0. For example, $\Pr\big(Y_i = y_i \,\big|\, Y_{i0} = y_{i0},\, X_i =$

$x_i,\ A_i = \alpha_i\big) = p_{y_{i0}}(y_i, x_i, \beta_0, \gamma_0, \alpha_i)$, where

$$p_{y_{i0}}(y_i, x_i, \beta, \gamma, \alpha_i) := \prod_{t=1}^{T} \frac{1}{1 + \exp\left[(1 - 2y_{it})\left(x_{it}'\beta + y_{i,t-1}\gamma + \alpha_i\right)\right]}. \tag{2}$$

We drop the index i until we discuss estimation in Section 2.5, that is, instead of Y_i, X_i, A_i we just write Y, X, A for the corresponding random variables. The joint distribution of $(X, A) \in \mathbb{R}^{K \times T} \times \mathbb{R}$ is left unrestricted.

Our first goal is to obtain moment conditions for β and γ in model (1) that are valid independent of the realization of the fixed effects A. That is, we want to find moment functions $m_{y_0}(y, x, \beta, \gamma)$ such that

$$\mathbb{E}\left[m_{y_0}(Y, X, \beta_0, \gamma_0) \,\middle|\, Y_0 = y_0,\ X = x,\ A = \alpha\right] = 0, \qquad \text{for all} \quad \alpha \in \mathbb{R}. \tag{3}$$

In addition, the key to root-n consistent estimation of β and γ is that our moment functions in this section will also be valid for all $x \in \mathbb{R}^{K \times T}$, implying the unconditional moments $\mathbb{E}\left[\mathbb{1}\left(Y_0 = y_0\right) g(X)\, m_{y_0}(Y, X, \beta_0, \gamma_0)\right] = 0$, where $\mathbb{1}(.)$ is the indicator function, and $g(x)$ can be any functions of the regressors such that the expectation is well-defined. Subject to identification conditions, these unconditional moments can then be used to estimate β and γ via GMM at the root-n rate.

2.1 Moment conditions for $T = 3$

We first consider the model (1) with $T = 3$. In most applications, this corresponds to a total of four time periods: three for which the models is assumed to apply, plus the initial condition, y_0. It turns out that $T = 3$ is the smallest number of time periods for which we can derive moment conditions for model (1) — our discussion in Section 2.2 shows that it is not possible to derive moment conditions when $T = 2$.

Let $x_{ts} = x_t - x_s$. For $y_0 = 0$, we define

$$m_0^{(a)}(y, x, \beta, \gamma) = \begin{cases} \exp(x_{12}'\beta) & \text{if } y = (0,1,0), \\ \exp(x_{13}'\beta - \gamma) & \text{if } y = (0,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ \exp(x_{32}'\beta) - 1 & \text{if } y = (1,1,0), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_0^{(b)}(y, x, \beta, \gamma) = \begin{cases} \exp(x_{23}'\beta) - 1 & \text{if } y = (0,0,1), \\ -1 & \text{if } (y_1, y_2) = (0,1), \\ \exp(x_{31}'\beta) & \text{if } y = (1,0,0), \\ \exp(\gamma + x_{21}'\beta) & \text{if } y = (1,0,1), \\ 0 & \text{otherwise.} \end{cases}$$

For $y_0 = 1$, we define

$$m_1^{(a)}(y, x, \beta, \gamma) = \begin{cases} \exp(x_{12}'\beta + \gamma) & \text{if } y = (0,1,0), \\ \exp(x_{13}'\beta) & \text{if } y = (0,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ \exp(x_{32}'\beta) - 1 & \text{if } y = (1,1,0), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_1^{(b)}(y, x, \beta, \gamma) = \begin{cases} \exp(x_{23}'\beta) - 1 & \text{if } y = (0,0,1), \\ -1 & \text{if } (y_1, y_2) = (0,1), \\ \exp(x_{31}'\beta - \gamma) & \text{if } y = (1,0,0), \\ \exp(x_{21}'\beta) & \text{if } y = (1,0,1), \\ 0 & \text{otherwise.} \end{cases}$$

If Y_t is generated according to (1), then $Z_t = 1 - Y_t$ is also generated according to (1), but with X_t replaced by $-X_t$ and A replaced by $A - \gamma$. The pairs $\left(m_0^{(a)}, m_0^{(b)}\right)$ and $\left(m_1^{(a)}, m_1^{(b)}\right)$ are symmetric in the sense that $m_1^{(a)}(y, x, \beta, \gamma) = m_0^{(b)}(1 - y, -x, \beta, \gamma)$ and $m_1^{(b)}(y, x, \beta, \gamma) = m_0^{(a)}(1 - y, -x, \beta, \gamma)$. See also the proof of Lemma 1.

The following lemma establishes that the functions defined above satisfy the moment conditions (3).

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

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

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

$$\sum_{y\in\{0,1\}^3} p_{y_0}(y, x, \beta_0, \gamma_0, \alpha)\, m^{(a/b)}_{y_0}(y, x, \beta_0, \gamma_0) = 0. \tag{4}$$

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

2.1.1 Derivation of the moment functions

Once the above expressions for $m^{(a/b)}_{y_0}(y, x, \beta, \gamma)$ are available, Lemma 1 provides a formal justification of the moment conditions, and no further proof is required. But it may still be of interest to see how the expressions for $m^{(a/b)}_{y_0}(y, x, \beta, \gamma)$ can be obtained in the first place, and in the following we explain this informally.

Recall that the model probabilities are given in (2). Here, we use them for $T = 3$, which gives eight probabilities. We define vectors in $\mathbb{R}^8$ for the model probabilities and for the candidate moment functions:

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

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 are the same in the probability vector $\mathbf{p}(\alpha) = \mathbf{p}_{y_0}(x,\beta,\gamma,\alpha)$ and in the moment vector $\mathbf{m} = \mathbf{m}_{y_0}(x,\beta,\gamma)$. The probability vector $\mathbf{p}(\alpha)$ as a function of α is given by the model specification. We say that 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 Lemma 1, because $\mathbf{m}'\,\mathbf{p}(\alpha)$ is equal to $\mathbb{E}\left[m_{y_0}(Y,X,\beta_0,\gamma_0)\,\middle|\,Y_0 = y_0,\ X = x,\ A = \alpha\right]$.

This algebraic formulation of the problem of finding valid moment conditions in panel models with fixed effects is at the heart of the functional differencing method in Bonhomme (2012), and what follows is very much in the spirit of functional differencing. However, while Bonhomme (2012) provides a general computational method for estimating many different panel data models, our goal here is to obtain explicit algebraic formulas for moment conditions for one specific model, namely the dynamic logit model.

Any valid moment vector also satisfies $\lim_{\alpha\to\pm\infty} \mathbf{m}'\,\mathbf{p}(\alpha) = 0$. 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 $\mathbf{e}_8' \, \mathbf{m} = 0$.

Next, for concrete numerical values of x, β, γ, and for a finite set of concrete fixed effect values $\alpha_1 < \alpha_2 \ldots < \alpha_Q$, $Q \geq 8$, one can construct the $Q \times 8$ matrix $\mathbf{L} = [\mathbf{p}(\alpha_1), \ldots, \mathbf{p}(\alpha_Q)]'$, and numerically calculate the nullspace of $\mathbf{L}$. This is the set of vectors $\mathbf{m}$ that satisfy $\mathbf{L}\,\mathbf{m} = 0$. Obviously, any valid moment vector is an element of this nullspace. By experimenting with concrete numerical parameter and regressor values one can easily form the following hypothesis:

(2) The nullspace of $\mathbf{L}$ is always two dimensional, implying that the set of valid moment vectors $\mathbf{m}$ is two-dimensional.

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

Motivated by hypothesis (2), we want to find two linearly independent moment vectors $\mathbf{m}^{(a)}$ and $\mathbf{m}^{(b)}$ for each $y_0 \in \{0, 1\}$. To distinguish $\mathbf{m}^{(a)}$ and $\mathbf{m}^{(b)}$ from each other, we impose the condition $\mathbf{e}_2' \, \mathbf{m}^{(a)} = 0$ for the first vector and the condition $\mathbf{e}_7' \, \mathbf{m}^{(b)} = 0$ 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 $\mathbf{e}_5' \, \mathbf{m}^{(a)} = -1$ and $\mathbf{e}_4' \, \mathbf{m}^{(b)} = -1$. Together with the conditions in (1), we have so far specified four affine restrictions on each of the vectors $\mathbf{m}^{(a)}, \mathbf{m}^{(b)} \in \mathbb{R}^8$. To define $\mathbf{m}^{(a)}$ and $\mathbf{m}^{(b)}$ uniquely, we require four more affine conditions each, and for this we choose four values α_q and impose the orthogonality between $\mathbf{p}(\alpha_q)$ and $\mathbf{m}^{(a/b)}$. Thus, motivated by (1) and (2), we need to solve the following two linear systems of equations:

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

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

If it is indeed possible to find such moment functions $\mathbf{m}^{(a/b)}$, then it must be possible for the four values $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$ to be chosen arbitrarily without affecting the solutions $\mathbf{m}^{(a/b)}$. For example, $\alpha_q = q$ would be a valid choice. We have also made other arbitrary choices ($\mathbf{e}_2' \, \mathbf{m}^{(a)} = 0$, $\mathbf{e}_7' \, \mathbf{m}^{(a)} = 0$, and the normalizations) in order to obtain unique definitions for $\mathbf{m}^{(a/b)}$ that are algebraically convenient. The two-dimensional span of the vectors $\mathbf{m}^{(a)}$ and $\mathbf{m}^{(b)}$ and the potential of the moment conditions to identify and estimate β and γ is not affected by these choices.

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

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

Plugging the analytical expression for $\mathbf{p}(\alpha) = \mathbf{p}(x, \beta, \gamma, \alpha)$ into the definitions $\mathbf{B}^{(a)}$ and $\mathbf{B}^{(b)}$, we thus obtain analytical expressions for $\mathbf{m}^{(a)} = \mathbf{m}^{(a)}_{y_0}(x, \beta, \gamma)$ and $\mathbf{m}^{(b)} = \mathbf{m}^{(b)}_{y_0}(x, \beta, \gamma)$. The results are reported at the beginning of Section 2.1. Doing those calculations by hand would be cumbersome, but on a modern computer algebra system (e.g. Mathematica), the calculations and simplifications of the expressions for $\mathbf{m}^{(a)}$ and $\mathbf{m}^{(b)}$ take about one second each. Once the analytical expressions for $m^{(a/b)}$ at the beginning of Section 2.1 are available, verifying that those moment functions satisfy the linear systems (a) and (b) above is straightforward and can easily be done by hand. In particular, the key result $\mathbf{p}'(\alpha) \, \mathbf{m}^{(a/b)} = 0$ is exactly the statement of Lemma 1.

Here, we have described the derivation of the moment conditions for panel AR(1) logit models with $T = 3$. All other moment conditions for panel logit models that we provide in this paper are derived in a similar fashion.

2.1.2 Relation to Kitazawa (2013, 2016)

Kitazawa (2013) defines

$$\begin{aligned}
U_t &= y_t + (1-y_t)y_{t+1} - (1-y_t)y_{t+1}\exp(-\beta\Delta x_{t+1}) - \delta y_{t-1}(1-y_t)y_{t+1}\exp(-\beta\Delta x_{t+1}),\\
\hbar U_t &= U_t - y_{t-1} - \tanh\left[\frac{-\gamma y_{t-2} + \beta(\Delta x_t + \Delta x_{t+1})}{2}\right](U_t + y_{t-1} - 2U_t y_{t-1}),\\
\Upsilon_t &= y_t y_{t+1} + y_t(1-y_{t+1})\exp(\beta\Delta x_{t+1}) + \delta(1-y_{t-1})y_t(1-y_{t+1})\exp(\beta\Delta x_{t+1}),\\
\hbar\Upsilon_t &= \Upsilon_t - y_{t-1} - \tanh\left[\frac{\gamma(1-y_{t-2}) + \beta(\Delta x_t + \Delta x_{t+1})}{2}\right](\Upsilon_t + y_{t-1} - 2\Upsilon_t y_{t-1}), \qquad (5)
\end{aligned}$$

where $\delta = e^{\gamma} - 1$ and $\Delta x_t = x_t - x_{t-1}$. He shows that, for $t \in \{2, \dots, T-1\}$,[6] the functions $\hbar U_t$ and $\hbar\Upsilon_t$ are valid moment functions, in the sense of (3). 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 show that

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

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 $T = 3$. For AR(1) models with $T > 3$, and for AR(p) with $p \geq 2$ we obtain moment functions below that have not been derived previously in the literature.

[6] This is written here in our conventions for t and T. In his conventions, valid moment functions are obtained for $t \in \{3, \dots, T-1\}$.

2.1.3 Relation to other existing results

Honoré and Kyriazidou (2000) observe that if $x_2 = x_3$, then the conditional likelihood function which conditions on $Y = y_0$, $Y_3 = y_3$, and $Y_1 + Y_2 = 1$ can be used to estimate (β, γ). They then use this observation to construct an estimator that uses kernel weights based on $x_2 - x_3$. In this section, we note that the first order conditions for this conditional likelihood when $x_2 = x_3$ are linear transformations of the moment functions found above. In this sense, the moment conditions above can be thought of as an extension of Honoré and Kyriazidou (2000) to the case where $x_2 \neq x_3$.

We first observe that when $x_2 = x_3$,

$$
\begin{aligned}
\frac{m_0^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta - \gamma\right)\, m_0^{(b)}(y,x,\beta,\gamma)}{\exp\left(-\gamma\right) - 1} &= \begin{cases} -\exp\left(x_{12}'\beta\right) & \text{if } y = (0,1,0), \\ 1 & \text{if } y = (1,0,0), \\ 0 & \text{otherwise}, \end{cases} \\
\frac{m_0^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta\right)\, m_0^{(b)}(y,x,\beta,\gamma)}{\exp(\gamma) - 1} &= \begin{cases} -\exp\left(x_{12}'\beta - \gamma\right) & \text{if } y = (0,1,1), \\ 1 & \text{if } y = (1,0,1), \\ 0 & \text{otherwise}. \end{cases}
\end{aligned}
\tag{6}
$$

Since we know from Lemma 1 that those linear combinations are valid moment conditions, we conclude that conditional on $y_0 = 0$ and $x_2 = x_3$, the probability ratio of the events $y = (1,0,0)$ and $y = (0,1,0)$ is equal to $\exp\left(x_{12}'\beta_0\right)$, and the probability ratio of the events $y = (1,0,1)$ and $y = (0,1,1)$ is equal to $\exp\left(x_{12}'\beta_0 - \gamma_0\right)$, independent of α. Analogous results hold for $y_0 = 1$.[7] Observing that those probability ratios do not depend on α is equivalent to the finding in Honoré and Kyriazidou (2000) that the

[7]For $y_0 = 1$ and $x_2 = x_3$ we have

$$
\begin{aligned}
\frac{m_1^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta\right)\, m_1^{(b)}(y,x,\beta,\gamma)}{\exp\left(-\gamma\right) - 1} &= \begin{cases} -\exp\left(x_{12}'\beta + \gamma\right) & \text{if } y = (0,1,0), \\ 1 & \text{if } y = (1,0,0), \\ 0 & \text{otherwise}, \end{cases} \\
\frac{m_1^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta + \gamma\right)\, m_1^{(b)}(y,x,\beta,\gamma)}{\exp\left(\gamma\right) - 1} &= \begin{cases} -\exp\left(x_{12}'\beta\right) & \text{if } y = (0,1,1), \\ 1 & \text{if } y = (1,0,1), \\ 0 & \text{otherwise}. \end{cases}
\end{aligned}
$$

conditional log-likelihood

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

does not depend on α, for $x=(x_1,x_2,x_2)$. The corresponding first order conditions read

$$\frac{\partial \ell_{0,0}(y,x,\beta,\gamma)}{\partial \gamma} = 0,$$

$$\frac{\partial \ell_{0,0}(y,x,\beta,\gamma)}{\partial \beta} = \frac{x_{12}}{1+\exp\left(x_{12}'\beta\right)} \left[\frac{m_0^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta-\gamma\right)\, m_0^{(b)}(y,x,\beta,\gamma)}{\exp\left(-\gamma\right)-1}\right], \tag{7}$$

and

$$\begin{pmatrix} \dfrac{\partial \ell_{0,1}(y,x,\beta,\gamma)}{\partial \gamma} \\ \dfrac{\partial \ell_{0,1}(y,x,\beta,\gamma)}{\partial \beta} \end{pmatrix} = \begin{pmatrix} -1 \\ x_{12} \end{pmatrix} \frac{1}{1+\exp\left(x_{12}'\beta-\gamma\right)} \times \left[\frac{m_0^{(a)}(y,x,\beta,\gamma) + \exp\left(x_{12}'\beta\right)\, m_0^{(b)}(y,x,\beta,\gamma)}{\exp(\gamma)-1}\right], \tag{8}$$

and analogously for $y_0 = 1$. 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 (1) 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 $(\beta_2, \beta_3, \gamma)$.[8] Thus, our moment

[8]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.

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. Section 2.4 provides formal results on point-identification of β and γ from our moment conditions.

The numerical calculations in Honoré and Tamer (2006) suggest that if the only explanatory variable in (1) is a time trend, then the model is not point identified and the identified regions appears to be a connected set. The moment conditions in 2.1 show that the latter is false. For example, for $x_t = t$ and $y_0 = 0$ the conditions $0 = \mathbb{E}\left[m_{y_0}^{(a/b)}(Y, X, \beta, \gamma) \,\middle|\, Y_0 = y_0,\, X = x\right]$ become

$$0 = e^{-\beta}\, p_0(0,1,0) + e^{-2\beta-\gamma}\, p_0(0,1,1) - \sum_{y_3 \in \{0,1\}} p_0(1,0,y_3) + \left(e^{\beta} - 1\right) p_0(1,1,0),$$

$$0 = \left(e^{-\beta} - 1\right) p_0(0,0,1) - \sum_{y_3 \in \{0,1\}} p_0(0,1,y_3) + e^{2\beta}\, p_0(1,0,0) + e^{\beta+\gamma}\, p_0(1,0,1),$$

where $p_{y_0}(y_1, y_2, y_3) = \Pr(Y = (y_1, y_2, y_3) \,|\, Y_0 = y_0)$. Reparameterizing to $(\tau, \theta) = (\exp(\gamma + \beta), \exp(\beta))$, these conditions become

$$\tau = \frac{p_0(0,1,1)}{-p_0(0,1,0) + \theta \sum_{y_3} p_0(1,0,y_3) - \left(\theta^2 - \theta\right) p_0(1,1,0)},$$

$$\tau = \frac{-\left(\theta^{-1} - 1\right) p_0(0,0,1) + \sum_{y_3} p_0(0,1) - \theta^2\, p_0(1,0,0)}{p_0(1,0,1)}. \qquad (9)$$

Assuming that model (1) holds one finds that the first expression for τ in (9) is a strictly convex function of θ,[9] while the second expression for τ is a strictly concave function of θ. This implies that there are at most two solutions to the moment conditions based on

[9] To show this strict convexity one needs use that $\tau > 0$, which implies $-p_0(0,1,0) + \theta \sum_{y_3} p_0(1,0,y_3) - \left(\theta^2 - \theta\right) p_0(1,1,0) > 0$.

$y_0 = 0$ (the true parameter values being one of them).[10] Of course, the initial condition $y_0 = 1$ will give two additional moment conditions, which will also help to identify the true parameter values.

The only situation in which the two equations in (9) have only one solution is when the derivative of τ with respect to θ (evaluated at the true parameter values) is the same in both expressions of (9). This happens when $\gamma_0 = \beta_0 = 0$ (so $\tau_0 = \theta_0 = 1$). In that case, γ_0 and β_0 are point identified from the two moment conditions $0 = \mathbb{E}\left[m^{(a/b)}_{y_0}(Y, X, \beta, \gamma) \,\middle|\, Y_0 = y_0,\, X = x\right]$, but the "Hessian" of the resulting GMM estimator is singular.

2.2 Impossibility of moment conditions when $T = 2$

Here, we argue that it is not possible to derive moment conditions for model (1) 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 $m_{y_0}(y, x, \beta, \gamma)$ would satisfy

$$\sum_{y \in \{0,1\}^2} p_{y_0}(y, x, \beta, \gamma, \alpha)\; m_{y_0}(y, x, \beta, \gamma) = 0, \tag{10}$$

for all $\alpha \in \mathbb{R}$. In the limit $\alpha \to \infty$ the model probabilities become zero, except for $p_{y_0}((1, 1), x, \beta, \gamma, \alpha) \to 1$. This implies $m_{y_0}((1, 1), x, \beta, \gamma) = 0$. Analogously, in the limit $\alpha \to -\infty$ we have $p_{y_0}((0, 0), x, \beta, \gamma, \alpha) \to 1$, which implies $m_{y_0}((0, 0), x, \beta, \gamma) = 0$. Thus, only $m_{y_0}((0, 1), x, \beta, \gamma)$ and $m_{y_0}((1, 0), x, \beta, \gamma)$ can be non-zero, and (10) therefore implies that

$$\begin{aligned}\frac{m_{y_0}((0,1), x, \beta, \gamma)}{m_{y_0}((1,0), x, \beta, \gamma)} &= -\frac{p_{y_0}((1,0), x, \beta, \gamma, \alpha)}{p_{y_0}((0,1), x, \beta, \gamma, \alpha)} \\ &= -\exp\left((x_1 - x_2)'\,\beta + \gamma y_0\right) \frac{1 + \exp\left(x_2'\beta + \alpha\right)}{1 + \exp\left(x_2'\beta + \gamma + \alpha\right)}.\end{aligned} \tag{11}$$

[10] We have investigated the "false" solution to the moment conditions using numerical calculations like those in Honoré and Tamer (2006) and Honoré and Kyriazidou (2019). Preliminary results suggest that for that parameter vector, it is impossible to find a distribution of the individual specific effects, A, which will produce the observed probabilities. We plan to investigate this further in future research.

Unless $\gamma = 0$, the right hand side of (11) 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 (11) yields the moment conditions implied by Rasch (1960a)'s conditional likelihood.

2.3 Generalizations of the moment conditions to $T > 3$

The discussion in Section 2.2 demonstrates that it is not possible to apply our approach to the case $T = 2$. We now show how the moment functions for $T = 3$ in Section 2.1 generalize to more than three time periods (after the initial y_0).

Define the single index for time period t as $z_t(y_0, y, x, \beta, \gamma) = x_t' \beta + y_{t-1} \gamma$, and define the corresponding pairwise differences $z_{ts}(y_0, y, x, \beta, \gamma) = z_t(y_0, y, x, \beta, \gamma) - z_s(y_0, y, x, \beta, \gamma)$. Then, for triples of time periods $t, s, r \in \{1, 2, \ldots, T\}$ with $t < s < r$, we define the moment functions

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

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

For $T = 3$ and $(t, s, r) = (1, 2, 3)$, these moment functions are exactly those introduced in Section 2.1.

Proposition 1 *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* $y_0 \in \{0,1\}$, $x \in \mathbb{R}^{K\times T}$, $\alpha \in \mathbb{R}$, *and* $w : \{0,1\}^{t-1} \to \mathbb{R}$ *that*

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

The proof is given in Appendix A.2.1. Because Proposition 1 holds for all functions $w : \{0,1\}^{t-1} \to \mathbb{R}$ we can equivalently write its conclusion as

$$\mathbb{E}\left[m_{y_0}^{(a)(t,s,r)}(Y,X,\beta_0,\gamma_0)\,\middle|\,(Y_0,\ldots,Y_{t-1}) = (y_0,\ldots,y_{t-1}),\; X = x,\; A = \alpha\right] = 0,$$
$$\mathbb{E}\left[m_{y_0}^{(b)(t,s,r)}(Y,X,\beta_0,\gamma_0)\,\middle|\,(Y_0,\ldots,Y_{t-1}) = (y_0,\ldots,y_{t-1}),\; X = x,\; A = \alpha\right] = 0,$$

that is, the moment functions are valid conditional on all outcomes Y_τ with $\tau < t$.

Unbalanced panels: The only regressor and outcome values that enter into the moment functions $m_{y_0}^{(a)(t,s,r)}$ and $m_{y_0}^{(b)(t,s,r)}$ are (x_t, x_s, x_r) and $(y_{t-1}, y_t, y_{s-1}, y_s, y_{r-1}, y_t)$.[11] Thus, as long as those variables are observed we can evaluate $m_{y_0}^{(a)(t,s,r)}$ and $m_{y_0}^{(b)(t,s,r)}$. 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 5 we discuss how to combine the moment functions for unbalanced panels.

Linear dependence and completeness: Proposition 1 states that the functions $\mathfrak{m}_{y_0}(y,x,\beta,\gamma) = w(y_1,\ldots,y_{t-1})\, m_{y_0}^{(a/b)(t,s,r)}(y,x,\beta,\gamma)$ are valid moment functions, in the sense of equation (3). We conjecture that for $\gamma \neq 0$ (and arbitrary y_0, x and β) any such valid moment function for panel logit AR(1) models with $T \geq 3$ can be written as

$$\mathfrak{m}_{y_0}(y,x,\beta,\gamma) = \sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}\Big[w_{y_0}^{(a)}(t,s,y_1,\ldots,y_{t-1},x,\beta,\gamma)\, m_{y_0}^{(a)(t,s,T)}(y,x,\beta,\gamma)$$

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

$$+ w_{y_0}^{(b)}(t,s,y_1,\ldots,y_{t-1},x,\beta,\gamma)\, m_{y_0}^{(b)(t,s,T)}(y,x,\beta,\gamma)\Big], \quad (12)$$

with weights $w_{y_0}^{(a/b)}(t,s,y_1,\ldots,y_{t-1},x,\beta,\gamma) \in \mathbb{R}$ that are *uniquely* determined by the function $\mathfrak{m}_{y_0}(\cdot,x,\beta,\gamma)$. For given values of y_0, x, β and γ, the set of valid moment functions is a linear subspace of the 2^T-dimensional vector space of real-valued functions over the outcomes $y \in \{0,1\}^T$. The conjecture in equation (12) states that a basis for this linear subspace can be obtained from the moment functions in Proposition 1 with $r = T$. An analytical proof of this conjecture is currently beyond our reach. However, we have verified the conjecture numerically for many different combinations of $T \geq 3$, $y_0 \in \{0,1\}$, $x \in \mathbb{R}^{K\times T}$, $\beta \in \mathbb{R}^K$ and $\gamma \neq 0$, and we believe that it holds for all such values.

To give a concrete example, consider the moment function $m_{y_0}^{(a)(1,2,3)}(y,x,\beta,\gamma)$ for $T = 4$. According to (12) it should be possible to express this moment function in terms of the "basis functions" $m_{y_0}^{(a/b)(t,s,4)}(y,x,\beta,\gamma)$, with $1 \leq t < s \leq 4$. Indeed, for $\gamma \neq 0$ one can show that[12]

$$\begin{aligned}
&(e^{\gamma}-1)\, m_{y_0}^{(a)(1,2,3)}(y,x,\beta,\gamma) = \left(1+e^{\gamma}-e^{x_{43}'\beta}-e^{\gamma+x_{34}'\beta}\right) m_{y_0}^{(a)(1,2,4)}(y,x,\beta,\gamma)\\
&+\left(e^{\gamma+x_{34}'\beta}-1\right) m_{y_0}^{(a)(1,3,4)}(y,x,\beta,\gamma) + e^{y_0\gamma}\left(e^{x_{14}'\beta}-e^{x_{13}'\beta}\right) m_{y_0}^{(b)(1,3,4)}(y,x,\beta,\gamma)\\
&+e^{(1-y_1)(y_0\gamma+x_{14}'\beta)}\left[\left(e^{x_{42}'\beta}-e^{x_{32}'\beta}\right) m_{y_0}^{(a)(2,3,4)}(y,x,\beta,\gamma)+\left(e^{x_{43}'\beta}-1\right) m_{y_0}^{(b)(2,3,4)}(y,x,\beta,\gamma)\right],
\end{aligned}$$

and an analogous result holds for $m_{y_0}^{(b)(1,2,3)}(y,x,\beta,\gamma)$. Thus, for $T = 4$ and fixed $y_0 \in \{0,1\}$ we find that from the ten valid moment functions provided by Proposition 1 only eight are linearly independent.[13]

For general $T \geq 3$, assuming that the conjecture is correct, the total number of independent moment conditions is equal to the number of basis elements on the right

[12] $m_{y_0}^{(b)(1,2,4)}(y,x,\beta,\gamma)$ could also appear in the expression for $m_{y_0}^{(a)(1,2,3)}(y,x,\beta,\gamma)$ according to (12), but it happens to have a coefficient of zero in that linear combination.

[13] Both $m_{y_0}^{(a)}$ and $m_{y_0}^{(b)}$ give five moment functions each for $T = 4$, namely $m_{y_0}^{(a/b)(1,2,3)}(y,x,\beta,\gamma)$, $m_{y_0}^{(a/b)(1,2,4)}(y,x,\beta,\gamma)$, $m_{y_0}^{(a/b)(1,3,4)}(y,x,\beta,\gamma)$, $\mathbb{1}(y_1 = 0)\, m_{y_0}^{(a/b)(2,3,4)}(y,x,\beta,\gamma)$ and $\mathbb{1}(y_1 = 1)\, m_{y_0}^{(a/b)(2,3,4)}(y,x,\beta,\gamma)$, but $m_{y_0}^{(a/b)(1,2,3)}(y,x,\beta,\gamma)$ can be expressed in terms of the others.

hand side of (12), which we calculate to be

$$\underbrace{2}_{m^{(a)}\,\&\,m^{(b)}} \times \underbrace{\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}}_{\text{allowed } t,s \text{ values}} \underbrace{2^{t-1}}_{\substack{\text{number of linearly}\\ \text{independent functions}\\ w(y_1,\ldots,y_{t-1})}} = \sum_{t=1}^{T-2} 2^t\,(T-t-1) = 2^T - 2T. \tag{13}$$

Thus, the dimension of the valid moment function subspace is $2T$ smaller than the total dimension of $\{0,1\}^T \to \mathbb{R}$. This means that the condition in equation (3) imposes $2T$ linear restrictions on the set of valid moment functions.[14]

Thus, we find that the number of available moment conditions for panel logit AR(1) models is completely independent from the value of the explanatory variables $x \in \mathbb{R}^{K\times T}$ that is conditioned on. By contrast, for panel logit AR(2) models we find in Section 3 that the number of available moments depends on the value of $x \in \mathbb{R}^{K\times T}$, for example, more moments are generally available for $x = 0$ than for $x \neq 0$. We find it remarkable that this is not the case for AR(1) models.

Notice also that the moment condition count in (13) is consistent with known results for panel logit AR(1) models without regressors (i.e. $x = 0$, or equivalently $\beta = 0$) and $\gamma \neq 0$. From Cox (1958) it is known that the distribution of Y conditional on $(Y_0, Y_T, \sum_{t=1}^{T-1} Y_t)$ is independent of the fixed effects A. For given values $Y_0 = y_0$, $Y_T = y_T$, and $\sum_{t=1}^{T-1} Y_t = q$ there are $\binom{T-1}{q}$ possible outcomes $y \in \{0,1\}^T$ in that conditioning set, and the independence of their conditional distribution from A implies that the ratios of all their pairwise probabilities is independent of A, which corresponds to $\binom{T-1}{q} - 1$ linearly independent restrictions on the model probabilities. Each such restriction corresponds to one valid moment condition in the sense of (3).[15] For a fixed

[14]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 $2^T - T - 1$. Those moments correspond to the first order conditions of the conditional likelihood in Rasch (1960b) and Andersen (1970).

[15]Not all of those moment conditions directly contain information on the parameter γ. Let $y, \widetilde{y} \in \{0,1\}^T$ be two outcome values with $y_T = \widetilde{y}_T$ and $\sum_{t=1}^{T-1} y_t = \sum_{t=1}^{T-1} \widetilde{y}_t$. Then we have

$$p_{y_0}(y,0,0,\gamma,\alpha) - \exp\left(\gamma \sum_{t=1}^{T} (y_{t-1}\, y_t - \widetilde{y}_{t-1}\, \widetilde{y}_t)\right) p_{y_0}(\widetilde{y},0,0,\gamma,\alpha) = 0,$$

which provides a valid moment function in the sense of (3), but that moment function only depends initial condition y_0 the total number of moment conditions is therefore equal to

$$\underbrace{2}_{y_T \in \{0,1\}} \times \underbrace{\sum_{q=0}^{T-1}}_{\substack{\text{all possible values} \\ \text{for } q=\sum_{t=1}^{T-1} Y_t}} \underbrace{\left[\binom{T-1}{q} - 1 \right]}_{\substack{\text{number of linearly} \\ \text{independent restrictions} \\ \text{obtained from that} \\ \text{conditioning set}}} = 2^T - 2T,$$

which verifies (13) in the case that $x = 0$. For $T = 3$ we have made the relation between our moment functions and the conditional likelihood approach explicit in Section 2.1.3 above (note that $x = 0$ is a special case of $x_2 = x_3$).

Finally, the moment functions of Kitazawa (2013) given in equation (5) are obtained from our moment functions when considering consecutive time periods (t, s, r). In that case, for $t \in \{2, \ldots, T-1\}$, we have

$$\hbar U_t = \left\{ \tanh \left[\frac{-\gamma y_{t-2} + \beta(\Delta x_t + \Delta x_{t+1})}{2} \right] - 1 \right\} m_{y_0}^{(b)(t-1,t,t+1)}(y, x, \beta, \gamma),$$
$$\hbar \Upsilon_t = \left\{ \tanh \left[\frac{\gamma(1 - y_{t-2}) + \beta(\Delta x_t + \Delta x_{t+1})}{2} \right] + 1 \right\} m_{y_0}^{(a)(t-1,t,t+1)}(y, x, \beta, \gamma).$$

However, only considering consecutive values (t, s, r) only provides $2^{T-1} - 2$ linearly independent moment functions for each initial condition,[16] out of $2^T - 2T$ available in total. For example, for $T = 4$ and fixed initial conditions, only six out of eight available linearly independent moment conditions can be obtained in that way. Thus, for $T > 3$ Kitazawa (2013) only provides a subset of the available moments.

on the parameter γ if $\sum_{t=1}^{T} y_{t-1} y_t \neq \sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{y}_t$. However, for our comparison with (13) that is not relevant, because we are just counting the dimension of the space of valid moment functions, independent from their parameter dependence. Also, moment functions that do not depend on the parameter can still be useful for model testing, or to construct efficient estimators from other moment conditions that contain the parameter.

[16] Those moment functions have the form $w(y_1, \ldots, y_{t-2})\, m_{y_0}^{(a/b)(t-1,t,t+1)}(y, x, \beta, \gamma)$, that is, for given t we obtain 2^{t-2} linearly independent functions through the choice of $w(y_1, \ldots, y_{t-2})$, times two for $m_{y_0}^{(a)}$ and $m_{y_0}^{(b)}$. The total number is therefore $2 \times \sum_{t=2}^{T-1} 2^{t-2} = 2^{T-1} - 2$, for each initial condition.

2.4 Identification

This section shows that our moment conditions for the panel logit AR(1) model 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

$$g_s(\beta, \gamma) = 0, \qquad \text{for all } s \in \{-,+\}^K, \tag{14}$$

has at most one solution.

To explain the lemma, consider the case $K = 1$[17] 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. The lemma states that under these conditions the system of two equations $g_\pm(\beta, \gamma) = 0$ can have at most one solution in β and γ. The proof is by contradiction: Let there be two solutions $g_\pm(\beta_1, \gamma_1) = 0$ and $g_\pm(\beta_2, \gamma_2) = 0$ with $(\beta_1, \gamma_1) \neq (\beta_2, \gamma_2)$, and (without loss of generality) $\gamma_1 \leq \gamma_2$. Then, for $\beta_1 \leq \beta_2$ we have $g_+(\beta_1, \gamma_1) < g_+(\beta_2, \gamma_2)$ by the monotonicity assumptions, while for $\beta_1 > \beta_2$ we have $g_-(\beta_1, \gamma_1) < g_-(\beta_2, \gamma_2)$ by the monotonicity assumptions. In both cases we have a contradiction to $g_\pm(\beta_1, \gamma_1) = g_\pm(\beta_2, \gamma_2) = 0$. We conclude that there cannot be two such solutions. The general proof for $K \in \mathbb{N}_0$ is analogous, and provided in Appendix A.2.2.

[17] 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.

Lemma 2 can be used to provide sufficient conditions for point identification in panel logit AR(1) models with $T = 3$ from the moment conditions in Section 2.1. For that purpose we define the sets

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

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 $m^{(a/b)}_{y_0}(y, x, \beta, \gamma)$ from Section 2.1 have particular monotonicity properties in the parameters β_k. For example, for $m^{(b)}_0$ we have, for all $\beta \in \mathbb{R}^K$ and $\gamma \in \mathbb{R}$, that

$$\frac{\partial \, \mathbb{E}\left[m^{(b)}_0(Y, X, \beta, \gamma) \,\middle|\, Y_0 = 0, \; X = x \right]}{\partial \beta} \begin{cases} > 0, & \text{for } x \in \mathcal{X}_{k,+} \, , \\ < 0, & \text{for } x \in \mathcal{X}_{k,-} \, . \end{cases} \tag{15}$$

This is because the parameter β appears in $m^{(b)}_0(y, x, \beta, \gamma)$ only through $\exp(x'_{23}\beta)$, $\exp(x'_{23}\beta)$ and $\exp(x'_{31}\beta)$; $x \in \mathcal{X}_{k,+}$ (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). Analogous monotonicity properties hold for the other moment functions $m^{(a/b)}_{y_0}(y, x, \beta, \gamma)$.

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

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

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

Because $\mathcal{X}_s$ is the intersection of the sets $\mathcal{X}_{k,\pm}$ the expected moment functions $\overline{m}^{(a)}_{y_0,s}(\beta,\gamma)$ and $\overline{m}^{(a)}_{y_0,s}(\beta,\gamma)$ have particular monotonicity properties with respect to all the elements of β. For example, (15) implies that $\overline{m}^{(b)}_{0,s}(\beta,\gamma)$ 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 $y_0 \in \{0,1\}$ and $\xi \in \{a,b\}$. Let the outcomes $Y = (Y_1,Y_2,Y_3)$ be generated from model (1) 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 $\overline{m}^{(\xi)}_{y_0,s}(\beta,\gamma)$ is well-defined.[18] *Then, the solution to*

$$\overline{m}^{(\xi)}_{y_0,s}(\beta,\gamma) = 0 \qquad \text{for all } s \in \{-,+\}^K \tag{16}$$

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

Proof. Consider $y_0 = 0$ and $\xi = b$. Using the definition of the moment function $m^{(b)}_0(y,x,\beta,\gamma)$ in Section 2.1 and the distribution of $Y|X,A$ implied by model (1) we find

$$\begin{aligned}
\frac{\partial\,\overline{m}^{(b)}_{0,s}(\beta,\gamma)}{\partial\gamma} &= \mathbb{E}\left[\frac{\partial\, m^{(b)}_0(Y,X,\beta,\gamma)}{\partial\gamma} \,\middle|\, Y_0 = 0,\ X \in \mathcal{X}_s \right] \\
&= \mathbb{E}\left[\frac{\partial \exp\left(\gamma + X_{21}'\beta\right)}{\partial\gamma} \Pr\left[Y = (1,0,1) \,\middle|\, Y_0 = 0, X, A\right] \,\middle|\, Y_0 = 0,\ X \in \mathcal{X}_s \right] \\
&= \mathbb{E}\left[\exp\left(\gamma + X_{21}'\beta\right)\, p_0((1,0,1),X,\beta_0,\gamma_0,A) \mid Y_0 = 0,\ X \in \mathcal{X}_s\right] > 0,
\end{aligned}$$

where in the last step (to conclude that the expression is positive) we used that $\exp\left(\gamma + x_{21}'\beta\right)\, p_0((1,0,1),x,\beta_0,\gamma_0,\alpha) > 0$ for all $x \in \mathbb{R}^{K\times 3}$ and $\alpha \in \mathbb{R}$.[19] We have

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

[19] Notice that we have introduced the domain of A to be $\mathbb{R}$. If we had introduced the domain of A

thus shown that $\overline{m}_{0,s}^{(b)}(\beta,\gamma)$ is strictly increasing in γ. Analogously, one can show that $\overline{m}_{0,s}^{(b)}(\beta,\gamma)$ 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 (15) above.

We can therefore apply Lemma 2 with $g_s(\beta,\gamma)$ equal to $\overline{m}_{0,s}^{(b)}(\beta,\gamma)$ to find that the system of equations in (16) 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 $y_0 \in \{0,1\}$ and $\xi \in \{a,b\}$ we can analogously apply Lemma 2 with $g_s(\beta,\gamma)$ equal to $\overline{m}_{0,s}^{(a)}(-\beta,-\gamma)$, $\overline{m}_{1,s}^{(a)}(-\beta,\gamma)$, and $\overline{m}_{1,s}^{(b)}(\beta,-\gamma)$. ■

Notice that only one of the moment functions $m_{y_0}^{(a)}$ or $m_{y_0}^{(b)}$ 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 $\Pr(Y_0 = y_0,\ X \in \mathcal{X}_s) > 0$, 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 Borel measure*. By contrast, the conditional likelihood approach in Honoré and Kyriazidou (2000) conditions on the set $x_2 = x_3$, which has zero Borel measure, and therefore also often zero probability measure, implying that the resulting estimates for β and γ usually converge at a rate smaller than root-n. In our approach here, using the sample analogs of the moment conditions $\overline{m}_{y_0,s}^{(a/b)}(\beta,\gamma) = 0$ 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 (or from Proposition 1 for $T > 3$), resulting in potentially much more efficient estimators for β and γ, and the following subsection describes how we implement such

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 have $p_0((1,0,1), X, \beta_0, \gamma_0, A) = 0$.

estimators in practice. Nevertheless, from a theoretical perspective, we believe that the new identification result in Theorem 1 is interesting, because it comprises a significant improvement over existing results for dynamic panel logit models with explanatory variables.

2.5 Estimation

For estimation, we suggest normalizing all moment functions such that we have $\sup_{y,x,\beta,\gamma} |\widetilde{m}(y,x,\beta,\gamma)| < \infty$. For example, rescaled versions of our $T = 3$ moment functions in Section 2.1 are given by

$$\widetilde{m}_0^{(a)}(y,x,\beta,\gamma) = \frac{m_0^{(a)}(y,x,\beta,\gamma)}{1+\exp(x_{12}'\beta)+\exp(x_{13}'\beta-\gamma)+\exp(x_{32}'\beta)},$$

$$\widetilde{m}_0^{(b)}(y,x,\beta,\gamma) = \frac{m_0^{(b)}(y,x,\beta,\gamma)}{1+\exp(x_{23}'\beta)+\exp(x_{31}'\beta)+\exp(\gamma+x_{21}'\beta)},$$

$$\widetilde{m}_1^{(a)}(y,x,\beta,\gamma) = \frac{m_1^{(a)}(y,x,\beta,\gamma)}{1+\exp(x_{12}'\beta+\gamma)+\exp(x_{13}'\beta)+\exp(x_{32}'\beta)},$$

$$\widetilde{m}_1^{(b)}(y,x,\beta,\gamma) = \frac{m_1^{(b)}(y,x,\beta,\gamma)}{1+\exp(x_{23}'\beta)+\exp(x_{31}'\beta-\gamma)+\exp(x_{21}'\beta)}.$$

Here, each moment function was 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 automatically rescaled in this way. See the factors $1/[1+\exp(x_{12}'\beta)]$ and $1/[1+\exp(x_{12}'\beta-\gamma)]$ in equations (7) and (8) above, and compare this to the entries in the corresponding moment functions in (6).

These rescaled moment functions are valid conditional on any realization of the regressors. We can form unconditional moment functions by multiplying with arbitrary powers of the regressor values. For example, for $T = 3$,

$$M_{y_0}(y, x, \beta, \gamma) = \big(1,\ x_{12}',\ x_{23}',\ x_{13}'\big)' \otimes \begin{pmatrix} \widetilde{m}_{y_0}^{(a)}(y, x, \beta, \gamma) \\ \widetilde{m}_{y_0}^{(b)}(y, x, \beta, \gamma) \end{pmatrix},$$

where $\otimes$ is the tensor product. We can choose to estimate the parameters β and γ separately for each initial condition $y_0 \in \{0, 1\}$ or jointly. If we choose to combine the information from both initial conditions, then the combined moment function reads

$$M(y_0, y, x, \beta, \gamma) = \begin{pmatrix} \mathbb{1}(y_0 = 0)\, M_0(y, x, \beta, \gamma) \\ \mathbb{1}(y_0 = 1)\, M_1(y, x, \beta, \gamma) \end{pmatrix}.$$

The corresponding GMM estimator is given by

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

where W is a symmetric positive-definite weight matrix. In the Monte Carlo simulations and the empirical example in Sections 4 and 5, 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.[20]

Under standard regularity conditions we have

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

with $\Omega = \mathrm{Var}[m(Y_{i,0}, Y_i, X_i, \beta_0, \gamma_0)]$ and $G = \mathbb{E}\left[\frac{\partial m(Y_{i,0}, Y_i, X_i, \beta_0, \gamma_0)}{\partial \beta'}, \frac{\partial m(Y_{i,0}, Y_i, X_i, \beta_0, \gamma_0)}{\partial \gamma'} \right]$. Section 5 explains how we implement the GMM estimator when $T > 3$, and, in partic-

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

ular, how to deal with unbalanced panels.

In general, some arbitrary choices are required when translating our conditional moment conditions in Lemma 1 (or Proposition 1 for $T > 3$) into a GMM estimator based on unconditional moments. We do not claim that the particular GMM estimator described here has any optimality properties, but it performs well in our Monte Carlo simulations even for relatively small sample sizes, see Section 4.

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

In this section, we consider logit AR(p) models. Specifically, we generalize the model in (1) to

$$\Pr\left(Y_{it} = 1 \,\middle|\, Y_i^{t-1}, X_i, A_i, \beta, \gamma\right) = \frac{\exp\left(X_{it}'\,\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}{1 + \exp\left(X_{it}'\,\beta + \sum_{\ell=1}^{p} Y_{i,t-\ell}\,\gamma_\ell + A_i\right)}, \tag{17}$$

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 (2), we define

$$p_{y_i^{(0)}}(y_i, x_i, \beta, \gamma, \alpha_i) = \prod_{t=1}^{T} \frac{\exp\left(x_{it}'\,\beta + \sum_{\ell=1}^{p} y_{i,t-\ell}\,\gamma_\ell + \alpha_i\right)}{1 + \exp\left(x_{it}'\,\beta + \sum_{\ell=1}^{p} y_{i,t-\ell}\,\gamma_\ell + \alpha_i\right)}, \tag{18}$$

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

The main goal of this section is to show that, as was the case for the AR(1) model in Section 2, there are many more moment conditions available that can be used to identify and estimate the common parameters β and γ in fixed effect logit AR(p) models with $p > 1$ than was previously known.

Our main focus is again on moment conditions that are applicable to all values of the parameters and all realizations of the regressors. By numeric calculations with various concrete values of the regressors and parameters, we found 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 such general moment conditions, and for all $p \in \{0, \ldots, 6\}$ and $T \in \{2 + p, \ldots, 8\}$ we have verified numerically that the number of linearly independent moment conditions available for each initial condition $y^{(0)}$ is equal to[21]

$$\ell = 2^T - (T + 1 - p)\, 2^p. \tag{19}$$

Recall that the moment functions can be interpreted as vectors in the 2^T-dimensional vector space of functions over the outcomes $y \in \{0, 1\}^T$. Equation (19) states that for fixed values of $y^{(0)}$, x, β, γ the probabilities $\{p_{y^{(0)}}(y, x, \beta, \gamma, \alpha) \,:\, \alpha \in \mathbb{R}\}$ only span a subspace of dimension $(T + 1 - p)\, 2^p$ in that vector space, leaving $2^T - (T + 1 - p)\, 2^p$ dimensions for its orthogonal complement. We consider it remarkable that for fixed p the dimension of the span of $p_{y^{(0)}}(y, x, \beta, \gamma, \alpha)$ only grows linearly in T, while the dimension of the vector space itself grows exponentially.

Equation (13) above is a special case of (19) for $p = 1$, and we have already provided analytic formulas for all those AR(1) moment functions in Section 2.3. In this section (and the appendix) we provide analytic formulas for the general moment functions when $p = 2$ and $T \in \{4, 5\}$ (where $\ell = 4$ and $\ell = 16$), and for $p = 3$ and $T = 5$ (where $\ell = 8$). Thus, for dynamic fixed effect logit models with $T \leq 5$, all the possible moment conditions that are applicable to general parameter and regressor values are available in this paper.

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. 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

[21] We believe that this formula for the number of linearly independent general moments holds for all integers p, T with $T \geq 2 + p$, but a general proof of this conjecture is beyond the scope of this paper.

each initial condition.[22] More importantly, there are additional moment conditions for special realizations of the regressors that are valid for all possible parameter values, and they can provide identifying information for the parameters β and γ even when $T < 2 + p$ (see Section 3.3 and 3.4 below).[23]

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

Consider model (17) with $p = 2$ and $T = 4$. Again write $z_t(y_0, y, x, \beta, \gamma) = x_t' \beta + y_{t-1} \gamma_1 + y_{t-2} \gamma_2$ for the single index that describes how the parameters β and γ enter into the model at time period t, and let $z_{ts}(y_0, y, x, \beta, \gamma) = z_t(y_0, y, x, \beta, \gamma) - z_s(y_0, y, x, \beta, \gamma)$ be its time-differences. To save space we drop the arguments of the differences in the following formulas and write z_{ts} instead of $z_{ts}(y_0, y, x, \beta, \gamma)$. 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

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

There are three additional moment functions $m^{(b,2,4)}_{y^{(0)}}(y, x, \beta, \gamma)$, $m^{(c,2,4)}_{y^{(0)}}(y, x, \beta, \gamma)$ and $m^{(d,2,4)}_{y^{(0)}}(y, x, \beta, \gamma)$. They are provided in Appendix A.3.1.

Here, the additional superscripts indicate $p = 2$ and $T = 4$. In these moment functions most of the dependence on the parameters β and γ is hidden in the single index notation, but for $y^{(0)} = (0, 0)$ we provide equivalent formulas for the moment

[22] 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 $\gamma_2 \neq 0$.

[23] 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 $x \in \mathbb{R}^{K \times T}$.

functions with explicit dependence on β and γ in Appendix A.3.2.

Notice that for the $T = 3$ moment conditions in Section 2.3 it was possible to absorb all the dependence on the parameters β and γ into the single index differences z_{ts}, while here for $T = 4$ there still is a remaining explicit dependence on the parameter γ_1. Another difference to Section 2.3 is that the moment conditions here do not easily generalize to non-consecutive sets of time periods. For $p = 2$ and $T > 4$ one cannot just replace the time periods $1, 2, 3, 4$ by general time periods $t < s < r < q$ and obtain valid moment conditions from the above (of course, just shifting all the time periods $1, 2, 3, 4$ by a constant is allowed since the model is unchanged under time-shifts). Thus, the structure of the moment conditions for AR(2) models is more complicated than for AR(1) models, or at least we do not understand their general structure equally well yet. Nevertheless, the moment conditions here generalize those for the AR(1) case, because they provide valid moment conditions for general regressors and for any realization of the individual specific effects.

Lemma 3 *If the outcomes $Y = (Y_1, Y_2, Y_3, Y_4)$ are generated from model (17) with $p = 2$, $T = 4$ and true parameters β_0 and γ_0, then we have for all $y^{(0)} \in \{0, 1\}^2$, $x \in \mathbb{R}^{K\times4}$, $\alpha \in \mathbb{R}$, and $\xi \in \{a, b, c, d\}$ that*

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

The proof of the lemma is discussed in Appendix A.3.2. Analogous to our discussion in Section 2.5 one can apply GMM to those moment conditions. Under suitable regularity conditions, this allows us to estimate the parameters β and γ of the panel AR(2) logit model at root-n-rate.

For the derivation of these moment functions we again followed the approach described in Section 2.1.1. To keep the analytical 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 for the parameters and regressors and following the approach in Section 2.1.1 to calculate numerical solutions for the moment functions when imposing various constraints and normalizations on them). For example, for $m_{y^{(0)}}^{(a,2,4)}$ one can form the conjecture that there should exist a valid moment function of the form

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

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

$$\sum_{y \in \{0,1\}^4} p_{y^{(0)}}(y, x, \beta_0, \gamma_0, \alpha_q)\, m_{y^{(0)}}^{(a,2,4)}(y, x, \beta_0, \gamma_0) = 0, \quad q \in \{1,2,3,4,5\}, \tag{20}$$

where we can choose five mutually different values $\alpha_q \in \mathbb{R}$ arbitrarily (the solution for $m_{y^{(0)}}^{(a,2,4)}$ will not depend on that choice). The derivation of the other moment functions is analogous.

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 initial conditions, making the total required number of observations $T_{\text{obs}} = 8$. Let $z_t(y_0, y, x, \beta, \gamma) = x_t' \beta + y_{t-1} \gamma_1 + y_{t-2} \gamma_2 + y_{t-3} \gamma_3$, and let $z_{ts}(y_0, y, x, \beta, \gamma) = z_t(y_0, y, x, \beta, \gamma) - z_s(y_0, y, x, \beta, \gamma)$. 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

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

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 A.3.3. We have the following lemma.

Lemma 4 *If the outcomes $Y=(Y_1,Y_2,Y_3,Y_4,Y_5)$ are generated from model (17) 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 3. In fact, the structure and derivation of $m_{y^{(0)}}^{(a,3,5)}(y,x,\beta,\gamma)$ is very similar to that of of $m_{y^{(0)}}^{(a,2,4)}(y,x,\beta,\gamma)$. When plugging in the definition of the single index into the elements of $m_{y^{(0)}}^{(a,3,5)}$ one finds that $m_{(y_{-2},y_{-1},y_0)}^{(a,3,5)}(y)=\exp\left[x_{51}'\beta-y_0\gamma_1-(y_{-1}-1)\gamma_2-y_{-2}\gamma_3\right]$ for both $(y_1,y_2,y_3,y_4)=(1,0,0,0)$ and $(y_1,y_2,y_3,y_4)=(1,0,0,1)$, that is, those entries of $m_{y^{(0)}}^{(a,3,5)}$ are actually identical. Thus, as soon as one has correctly conjectured the structure of $m_{y^{(0)}}^{(a,3,5)}$, then analogous to (20) we just need to solve a system of seven equations in seven unknowns to derive $m_{y^{(0)}}^{(a,3,5)}$. The derivation of all the other moment functions in Lemma 4 also requires solving a system in seven unknowns.

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 4 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 3.1. Using the results presented so far 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 (19). See Appendix A.3.5 for details.

3.3 AR(2) models with $T = 3$ and no explanatory variables.

Here, we consider the AR(2) logit model with fixed effects and no explanatory variables. This corresponds to model (17) with $p = 2$ and $x = 0$ (no parameters β). Honoré and Kyriazidou (2019) study this model and their numerical calculations suggested that the common parameters in such a model are point identified for $T = 3$ (i.e. $T_{\text{obs}} = 5$), but neither an analytical proof of identification nor an estimation method for the parameters $\gamma = (\gamma_1, \gamma_2)$ have been known so far. Here, we present moment conditions for this model that allow for identification and root-n estimation of the parameters γ. We keep this discussion short, because the moment conditions here are actually a special case of those in the following subsection.

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

$$\mathbb{E}\left[m_{y^{(0)}}(Y, \gamma_0) \,\middle|\, Y^{(0)} = y^{(0)},\, A = \alpha \right] = 0, \tag{21}$$

with moment functions given by

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

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

The moment conditions in (21) for $y^{(0)}=(0,0)$ and $y^{(0)}=(1,1)$ are strictly monotone in γ_1 and do not depend on γ_2 — they therefore each identify the parameter γ_1 uniquely. The moment conditions for $y^{(0)}=(0,1)$ and $y^{(0)}=(1,0)$ are strictly monotone in γ_2, and they therefore each identify the parameter γ_2 for known value of γ_1. This partly solves the puzzle presented in Honoré and Kyriazidou (2019). A GMM estimator based on those moment will be root-n consistent under standard regularity conditions.

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

Consider model (17) 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 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[24]

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

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

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

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

where 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

[24]The moment functions here are closely related to those for AR(1) models in Section 2. For example, $m_{(0_p)}$ and $m_{(1,0_{p-1})}$ can be written in one formula as

$$m_{y_0}(y, x, \beta, \gamma) = \begin{cases} \exp\left[z_{12}(y_0, y, x, \beta, \gamma)\right] & \text{if } (y_1, y_2, y_3) = (0,1,0), \\ \exp\left[z_{13}(y_0, y, x, \beta, \gamma)\right] & \text{if } (y_1, y_2, y_3) = (0,1,1), \\ -1 & \text{if } (y_1, y_2) = (1,0), \\ 0 & \text{otherwise}, \end{cases}$$

which is very similar to the formula for $m_{y_0}^{(a)(t,s,r)}(y, x, \beta, \gamma)$ in Section 2.3. However, here we find it convenient to write out the dependence on the parameters β and γ explicitly.

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 5 *If the outcomes $Y = (Y_1, Y_2, Y_3)$ are generated from model (17) with $p \geq 2$, $T = 3$, and true parameters β_0 and γ_0, then we have for all $y^{(0)} \in \big\{0_p, (0, 1_{p-1}), (1, 0_{p-1}), 1_p\big\}$, $(x_1, x_2) \in \mathbb{R}^{K\times 2}$, 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 A.3.6.

Identification: We now want to provide identification results for the parameters β and γ using the moment conditions in Lemma 5, 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 A.3.7).

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 2 *Let the outcomes $Y = (Y_1, Y_2, Y_3)$ be generated from model (17) 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*

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

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 2. 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 3.4 for $p=2$. More generally, for suitable discrete regressors one may also estimate β and γ at root-n rate.

4 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.

4.1 Fixed effects logit AR(1)

We consider first the fixed effect logit AR(1) model in equation 1. 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 the appendix. 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 the appendix 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 $K = 10$)[25].

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.

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

4.2 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 the appendix 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 2.5, 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 3.1 by the sum of the absolute values of the possible seven non-zero terms in the moment conditions in Sections 3.1 and A.3.2 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 the appendix 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 $K = 10$)

5 Empirical illustration

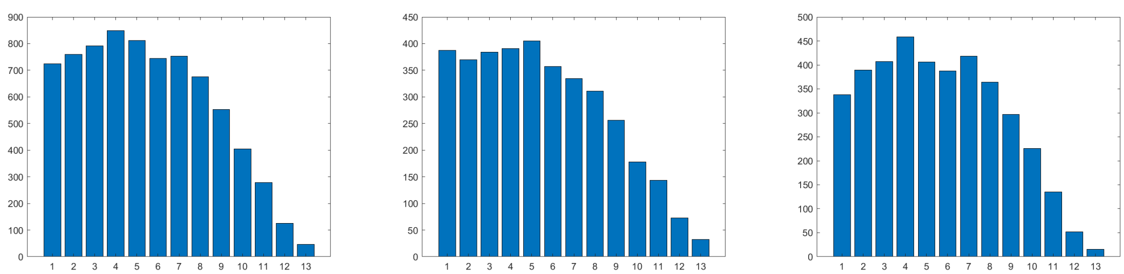
In this section, we illustrate the proposed GMM approach using data from the National Longitudinal Survey of Youth 1997[26] (NLSY97) covering the years 1997 to 2010. The

[26] 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

dependent variable is a binary variable indicating employment status by whether the respondent reported working ≥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 has (Children), a dummy variable for being married (Married), a transformation[27] 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 per individual in each of the three samples.

Figure 1: Histogram of Number of Observations Per Individual.



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

5.1 Estimation of the AR(1) model

The highly unbalanced nature of the panel makes it potentially important to consider weighing each observation according to the number of observations for that person. Specifically, if one uses the moment functions in Section 2.3 to all triplets (t, s, r), then observations with $T = T_i$ time periods will contribute $\binom{T_i}{3}$ terms to the sample analog of the moment. We use the following heuristic argument to choose the weights by mimicking the implicit weights used by the deviations from means estimator of a static linear panel data model. That estimator can be estimated by a linear regression of the difference of all $(y_{it} - y_{is})$ on the corresponding $(x_{it} - x_{is})$, where the differences for

not knowing the answer to the question), no other entries had missings imposed upon them.

[27]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.

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.

individual i are given weights $(T_i - 1)$ divided by the number of differences for that individual, $\binom{T_i}{2}$. We therefore weigh the triplets (t, s, r) for an observation with T_i time periods by $(T_i - 1) / \binom{T_i}{3}$. In addition to weighing the moment functions according to T_i, we also scale them as in Section 2.5.

Table 1 reports the estimation results. As expected, and consistent with the Monte Carlo results, 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 "'overfitted", leading to a downward bias in the estimated state dependence. Moreover, the standard logit estimator 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.

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.

5.2 Estimation of the AR(2) model

To estimate the AR(2) version of the model, we apply the moment conditions in Section 3.1 to all consecutive sequences of six outcomes (treating the first two as initial conditions). The moment functions are scaled as in the Monte Carlo Section. 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.

6 Conclusion

We have presented moment conditions for short panel data logit models with lagged dependent variables and strictly exogenous explanatory variables, and we have provided sufficient conditions for a finite number of these moments to identify the parameter of interest. A GMM estimator based on the moment conditions performs well in practice. Methodologically, the paper builds on the functional differencing insights of Bonhomme (2012), and our results here illustrate the power of this approach for identification and estimation of non-linear panel data models. Using the computational approach taken here to apply the functional differencing method in other non-linear panel models is an interesting topic for future research.

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., 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.

D'Addio, A. C., and B. E. Honoré (2010): "Duration dependence and timevarying variables in discrete time duration models," Brazilian Review of Econometrics, 30(2), 487–527.

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

Fernández-Val, I., and M. Weidner (2016): "Individual and time effects in nonlinear 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., and E. T. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," Econometrica, 74(3), 611–629.

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

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.

——— (2017): "DFEL-RTN, a set of TSP programs for root-N consistent estimations of dynamic fixed effects logit models," Discussion paper, No. 81, Kyushu Sangyo University Faculty of Economics.

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.

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.

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.

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 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

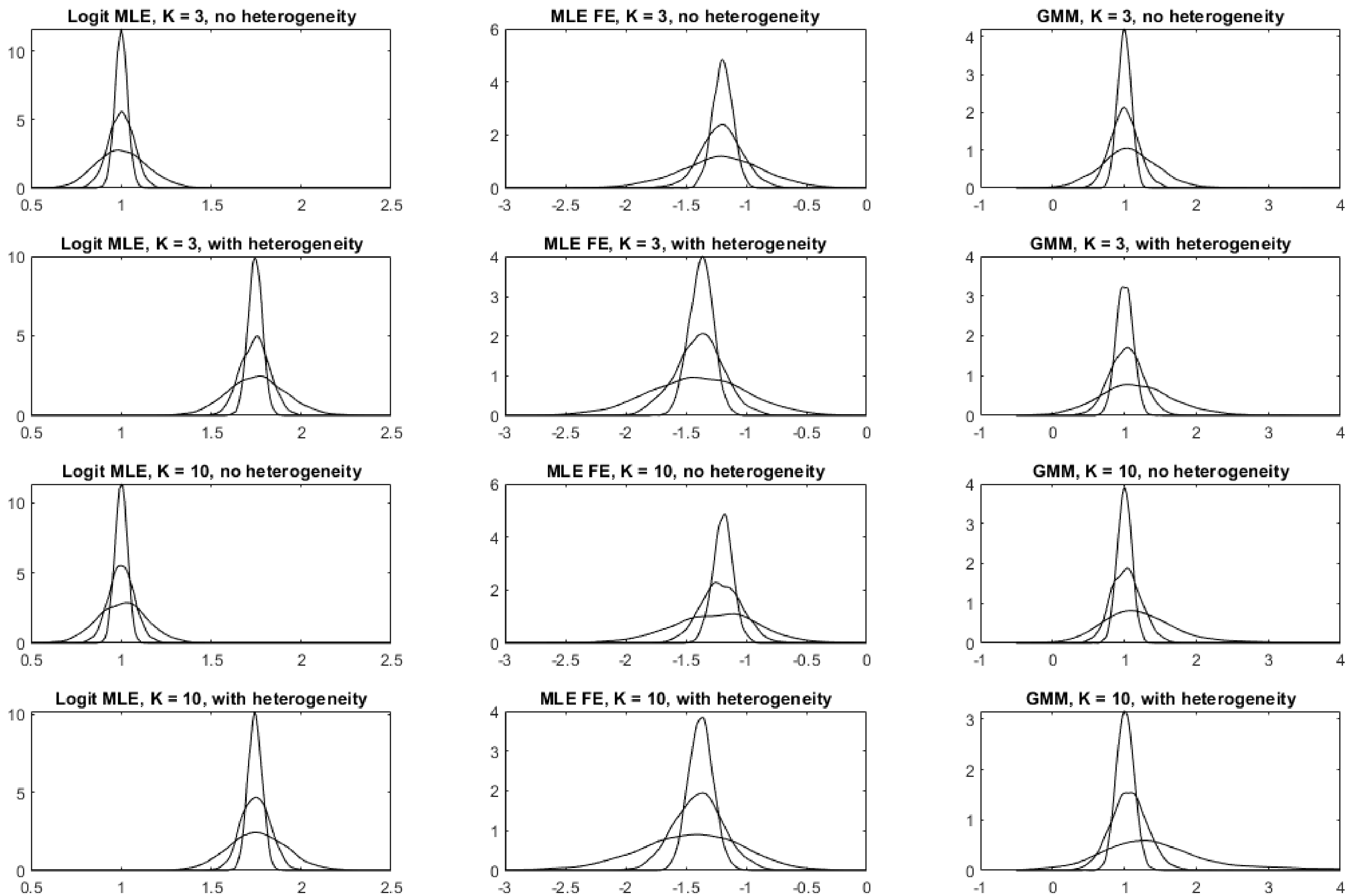
	$n = 500$					$n = 2000$					$n = 8000$				
	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3	γ_1	γ_2	β_1	β_2	β_3
True	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00	1.00	0.50	1.00	1.00	0.00
Bias	−2.87	−1.86	0.10	0.07	−0.00	−2.81	−1.83	0.06	0.05	0.00	−2.79	−1.82	0.05	0.05	−0.00
MAE	2.87	1.86	0.33	0.16	0.14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



A.2 Omitted Proofs for Section 2

A.2.1 Proof of Lemma 1

By plugging in the definition of the model probabilities $p_{y_0}(y, x, \beta_0, \gamma_0, \alpha)$ and the moment functions $m_{y_0}^{(a/b)}(y, x, \beta_0, \gamma_0)$ we want to verify equation (4). For $m_0^{(a)}$ we obtain

$$\sum_{y \in \{0,1\}^3} p_0(y, x, \beta_0, \gamma_0, \alpha)\, m_0^{(a)}(y, x, \beta_0, \gamma_0) = \frac{\exp(\zeta_1 - \zeta_2)}{[1+\exp(\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]}$$
$$+ \frac{\exp(\zeta_1 - \zeta_3 - \gamma_0)}{[1+\exp(\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(-\zeta_3-\gamma_0)]} - \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]}$$
$$+ \frac{\exp[(\zeta_3+\gamma_0) - (\zeta_2+\gamma_0)] - 1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2-\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]},$$

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

$$\sum_{y \in \{0,1\}^3} p_0(y, x, \beta_0, \gamma_0, \alpha)\, m_0^{(a)}(y, x, \beta_0, \gamma_0) = \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]}$$
$$+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2)]\,[1+\exp(\zeta_3+\gamma_0)]} - \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]}$$
$$+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]\,[1+\exp(-\zeta_3-\gamma_0)]}$$
$$- \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2-\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]},$$

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

$$\frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_3+\gamma_0)]} - \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(-\zeta_2-\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]}$$

$$= \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]},$$

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

$$\begin{aligned}
\sum_{y\in\{0,1\}^3} p_0(y,x,\beta_0,\gamma_0,\alpha)\, m_0^{(a)}(y,x,\beta_0,\gamma_0) &= -\frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]} \\
&+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]\,[1+\exp(-\zeta_3-\gamma_0)]} \\
&+ \frac{1}{[1+\exp(-\zeta_1)]\,[1+\exp(\zeta_2+\gamma_0)]\,[1+\exp(\zeta_3+\gamma_0)]} \\
&= 0,
\end{aligned}$$

where we used that $-1+1/\,[1+\exp(-\zeta_3-\gamma_0)]+1/\,[1+\exp(\zeta_3+\gamma_0)] = 0$. We have thus explicitly shown the statement of Lemma 1 for $m_0^{(a)}$.

The results for $m_0^{(b)}$ and $m_1^{(a/b)}$ can be derived analogously. However, once the result for $m_0^{(a)}$ is derived, then there is actually no need for any further calculation. Instead, it suffices to notice that the model probabilities $p_{y_0}(y,x,\beta,\gamma,\alpha)$ 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 $m_0^{(a)}$ gives the moment function $m_1^{(b)}$. Thus, by applying this symmetry transformation to our known result $\sum_{y\in\{0,1\}^3} p_0(y,x,\beta_0,\gamma_0,\alpha)$ $m_0^{(a)}(y,x,\beta_0,\gamma_0) = 0$ we obtain

$$\sum_{y\in\{0,1\}^3} p_1(y,x,\beta_0,\gamma_0,\alpha)\, m_1^{(b)}(y,x,\beta_0,\gamma_0) = 0.$$

Furthermore, the model probabilities $p_{y_0}(y,x,\beta,\gamma,\alpha)$ are also unchanged under the transformation

- $y_0 \rightarrow 1-y_0\,, \quad x_1'\beta \rightarrow x_1'\beta + (2y_0-1)\gamma\,, \quad y_{t-1}$ 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 $p_{y_0}(y,x,\beta,\gamma,\alpha)$ as a function of the product $x_t'\beta$, instead of x_t and β individually. Applying this transformation to $m_0^{(a)}$ gives the moment function $m_1^{(a)}$, and applying it to $m_1^{(b)}$ gives

$m_0^{(b)}$. Thus, by applying this symmetry transformation to our known results for $m_0^{(a)}$ and $m_1^{(b)}$ we obtain

$$\sum_{y\in\{0,1\}^3} p_1(y,x,\beta_0,\gamma_0,\alpha)\; m_1^{(a)}(y,x,\beta_0,\gamma_0) = 0,$$
$$\sum_{y\in\{0,1\}^3} p_0(y,x,\beta_0,\gamma_0,\alpha)\; m_0^{(b)}(y,x,\beta_0,\gamma_0) = 0.$$

■

Proof of Proposition 1

We first state and prove a useful lemma (Lemma 6 below), which will then be used to prove the proposition. In order to state this lemma we need to introduce some notation. Let $\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3 \in \{0,1\}$, $W \in \mathcal{W}$, and $V \in \mathcal{V}$ be random variables, where $\mathcal{W}$ and $\mathcal{V}$ are measurable spaces. Let $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Y}_3)$, and let $p(\widetilde{y}, w, v) \in [0,\infty)$ describe the joint distribution of $(\widetilde{Y}, W, V)$, that is, for all measurable subsets $\mathcal{Y}_* \subset \{0,1\}^3$, $\mathcal{W}_* \subset \mathcal{W}$ and $\mathcal{V}_* \subset \mathcal{V}$ we have

$$\Pr\left(\widetilde{Y} \in \mathcal{Y}_* \;\;\&\;\; W \in \mathcal{W}_* \;\;\&\;\; V \in \mathcal{V}_*\right) = \sum_{\widetilde{y}\in\mathcal{Y}_*} \int_{w\in\mathcal{W}_*} \int_{v\in\mathcal{V}_*} p(\widetilde{y}, w, v)\, \mu(dw)\, \nu(dv),$$

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

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

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 $p_1(\widetilde{y}_1)$, $p_2(\widetilde{y}_2 \,|\, w)$, $p_3(\widetilde{y}_3 \,|\, v)$ we assume logistic models:

$$p_1(\widetilde{y}_1) = \Lambda\left[(2\,\widetilde{y}_1 - 1)\,\pi_1\right], \quad p_2(\widetilde{y}_2 \,|\, w) = \Lambda\left[(2\,\widetilde{y}_2 - 1)\,\pi_2(w)\right], \quad p_3(\widetilde{y}_3 \,|\, v) = \Lambda\left[(2\,\widetilde{y}_3 - 1)\,\pi_3(v)\right], \tag{23}$$

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 $\pi_2 : \mathcal{W} \to \mathbb{R}$ and $\pi_3 : \mathcal{V} \to \mathbb{R}$ are functions. The only real assumption that we impose on $f(w \,|\, \widetilde{y}_1)$ and $g(v \,|\, \widetilde{y}_2, w)$ is that

$$g(v \,|\, 1, w) = g(v \,|\, 1), \tag{24}$$

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

$$\int_{w \in \mathcal{W}} f(w \,|\, \widetilde{y}_1)\, \mu(dw) = 1, \qquad\qquad \int_{v \in \mathcal{V}} g(v \,|\, \widetilde{y}_2, w)\, \nu(dv) = 1. \tag{25}$$

Notice that if we would strengthen (24) to $g(v \,|\, \widetilde{y}_2, w) = g(v \,|\, \widetilde{y}_2)$, for $\widetilde{y}_2 \in \{0, 1\}$, then (22) is equivalent to assuming a Markov chain

$$\widetilde{Y}_1 \xrightarrow{f} W \xrightarrow{p_2} \widetilde{Y}_2 \xrightarrow{g} V \xrightarrow{p_3} \widetilde{Y}_3,$$

but since we only impose (24) we also allow for dependence of W and V, conditional on $\widetilde{Y}_2 = 0$. Finally, we define $m : \{0, 1\}^3 \times \mathcal{W} \times \mathcal{V} \to \mathbb{R}$ by

$$m(\widetilde{y}, w, v) := \begin{cases} \exp\left[\pi_1 - \pi_2(w)\right] & \text{if } \widetilde{y} = (0, 1, 0), \\ \exp\left[\pi_1 - \pi_3(v)\right] & \text{if } \widetilde{y} = (0, 1, 1), \\ -1 & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1, 0), \\ \exp\left[\pi_3(v) - \pi_2(w)\right] - 1 & \text{if } \widetilde{y} = (1, 1, 0), \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Lemma 6 *Let $\pi_1 \in \mathbb{R}$, $\pi_2 : \mathcal{W} \to \mathbb{R}$ and $\pi_3 : \mathcal{V} \to \mathbb{R}$. Let the random variables $\widetilde{Y} \in \{0, 1\}^3$, $W \in \mathcal{W}$, $V \in \mathcal{V}$ be such that their distributions satisfy (22), (23), (24), (25), and let $m : \{0, 1\}^3 \times \mathcal{W} \times \mathcal{V} \to \mathbb{R}$ be defined by (26). Then we have*

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

Proof. Define

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

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 (23) and (26) we find that

$$\begin{aligned} h(\widetilde{y}_1, \widetilde{y}_2, w, v) &= \begin{cases} \Lambda(\pi_1)\,\Lambda[-\pi_3(v)] & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (0,1), \\ -p_2(\widetilde{y}_2 \,|\, w)\, p_1(\widetilde{y}_1) & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,0), \\ \Lambda(\pi_1)\,\Lambda[\pi_3(v)] \; - p_2(\widetilde{y}_2 \,|\, w)\, p_1(\widetilde{y}_1) & \text{if } (\widetilde{y}_1, \widetilde{y}_2) = (1,1), \\ 0 & \text{otherwise,} \end{cases} \\ &= \underbrace{\mathbb{1}\{\widetilde{y}_2 = 1\} \;\Lambda(\pi_1)\,\Lambda[(2\widetilde{y}_1 - 1)\pi_3(v)]}_{=:\, h_1(\widetilde{y}_1, \widetilde{y}_2, v)} \; \underbrace{- \mathbb{1}\{\widetilde{y}_1 = 1\} \; p_2(\widetilde{y}_2 \,|\, w)\, p_1(\widetilde{y}_1)}_{=:\, h_2(\widetilde{y}_1, \widetilde{y}_2, w)}, \end{aligned}$$

where we have decomposed $h(\widetilde{y}_1, \widetilde{y}_2, w, v)$ into the sum of $h_1(\widetilde{y}_1, \widetilde{y}_2, v)$, which does not depend on w, and of $h_2(\widetilde{y}_1, \widetilde{y}_2, w)$, which does not depend on v. Notice that the term $\Lambda[(2\widetilde{y}_1 - 1)\pi_3(v)]$ in $h_1(\widetilde{y}_1, \widetilde{y}_2, v)$ is identical to $p_3(\widetilde{y}_3 \,|\, v)$, but with $\widetilde{y}_3$ replaced by $\widetilde{y}_1$. Also using (24) and (25) we find

$$\begin{aligned} &\sum_{\widetilde{y}_1 \in \{0,1\}} \sum_{\widetilde{y}_2 \in \{0,1\}} \int_{w \in \mathcal{V}} \int_{v \in \mathcal{V}} h_1(\widetilde{y}_1, \widetilde{y}_2, v) \;\; g(v \,|\, \widetilde{y}_2, w) \;\; f(w \,|\, \widetilde{y}_1)\, \mu(dw)\, \nu(dv) \\ &= \Lambda(\pi_1) \sum_{\widetilde{y}_1 \in \{0,1\}} \int_{w \in \mathcal{V}} \int_{v \in \mathcal{V}} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(v)] \;\; g(v \,|\, 1) \;\; f(w \,|\, \widetilde{y}_1) \;\; \mu(dw)\, \nu(dv) \\ &= \Lambda(\pi_1) \sum_{\widetilde{y}_1 \in \{0,1\}} \int_{v \in \mathcal{V}} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(v)] \;\; g(v \,|\, 1) \; \underbrace{\int_{w \in \mathcal{V}} f(w \,|\, \widetilde{y}_1)\, \mu(dw)}_{=1} \; \nu(dv) \\ &= \Lambda(\pi_1) \int_{v \in \mathcal{V}} \underbrace{\sum_{\widetilde{y}_1 \in \{0,1\}} \Lambda[(2\widetilde{y}_1 - 1)\pi_3(v)]}_{=1} \;\; g(v \,|\, 1)\, \nu(dv) \\ &= \Lambda(\pi_1) \underbrace{\int_{v \in \mathcal{V}} g(v \,|\, 1)\, \nu(dv)}_{=1} \\ &= \Lambda(\pi_1). \end{aligned}$$

Similarly, we calculate

$$
\begin{aligned}
&\sum_{\widetilde{y}_1\in\{0,1\}}\sum_{\widetilde{y}_2\in\{0,1\}}\int_{w\in\mathcal{V}}\int_{v\in\mathcal{V}} h_2(\widetilde{y}_1,\widetilde{y}_2,w)\;\; g(v\,|\,\widetilde{y}_2,w)\;\; f(w\,|\,\widetilde{y}_1)\,\mu(dw)\,\nu(dv) \\
&=\sum_{\widetilde{y}_1\in\{0,1\}}\sum_{\widetilde{y}_2\in\{0,1\}}\int_{w\in\mathcal{V}} h_2(\widetilde{y}_1,\widetilde{y}_2,w)\underbrace{\int_{v\in\mathcal{V}} g(v\,|\,\widetilde{y}_2,w)\,\nu(dv)}_{=1} f(w\,|\,\widetilde{y}_1)\,\mu(dw) \\
&=-\,p_1(1)\int_{w\in\mathcal{V}}\underbrace{\sum_{\widetilde{y}_2\in\{0,1\}} p_2(\widetilde{y}_2\,|\,w)}_{=1}\;\; f(w\,|\,1)\,\mu(dw) \\
&=-\,p_1(1)\underbrace{\int_{w\in\mathcal{V}} f(w\,|\,1)\,\mu(dw)}_{=1} \\
&=-\,\Lambda(\pi_1).
\end{aligned}
$$

Combining the results in the last two displays gives

$$
\sum_{\widetilde{y}_1\in\{0,1\}}\sum_{\widetilde{y}_2\in\{0,1\}}\int_{w\in\mathcal{V}}\int_{v\in\mathcal{V}} h(\widetilde{y}_1,\widetilde{y}_2,w,v)\;\; g(v\,|\,\widetilde{y}_2,w)\;\; f(w\,|\,\widetilde{y}_1)\,\mu(dw)\,\nu(dv)=0,
$$

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

Using this lemma we now **prove Proposition 1**, which is the the main goal of this subsection. Let

$$
\widetilde{Y}_1=Y_t, \qquad \widetilde{Y}_2=Y_s, \qquad \widetilde{Y}_3=Y_r, \qquad W=Y_{s-1}, \qquad V=Y_{r-1},
$$

and

$$
\pi_1=X_t'\,\beta+Y_{t-1}\,\gamma, \qquad \pi_2(W)=X_s'\,\beta+Y_{s-1}\,\gamma, \qquad \pi_3(V)=X_r'\,\beta+Y_{r-1}\,\gamma.
$$

Then, the moment function $m_{y_0}^{(a)(t,s,r)}(y,x,\beta,\gamma)$ defined in the main text is exactly equal to the moment function $m(\widetilde{y},w,v)$ defined in (26). Furthermore, the assumptions of Proposition 1 guarantee that conditional on (Y^{t-1}, X, A) the distribution of $\widetilde{Y}_1$, $\widetilde{Y}_2$, $\widetilde{Y}_3$, W and V satisfies the assumptions of Lemma 6. We can therefore apply Lemma 6 to find that

$$\mathbb{E}\left[m_{y_0}^{(a)(t,s,r)}(Y, X, \beta_0, \gamma_0) \,\middle|\, Y^{t-1},\, X,\, A\right] = 0.$$

By applying the law of iterated expectations we thus find that

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

The result for $m_{y_0}^{(b)(t,s,r)}(y, x, \beta, \gamma)$ follows from this by applying an appropriate symmetry transformation $(Y_t \leftrightarrow 1 - Y_t)$.

A.2.2 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 (14). ■

A.3 Additional Material and Omitted Proofs for Section 3

A.3.1 Moment functions for the AR(2) model with $T = 4$

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

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

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

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

A.3.2 Proof of Lemma 3

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 (18) (for autoregressive order $p=2$) to verify by direct calculation that

$$\sum_{y\in\{0,1\}^4} p_{(0,0)}(y,x,\beta_0,\gamma_0,\alpha)\, m^{(\xi,2,4)}_{(0,0)}(y,x,\beta_0,\gamma_0) = 0, \tag{27}$$

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 $z_t(y_0, y, x, \beta, \gamma) = x_t'\beta + y_{t-1}\gamma$ 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 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 (27) 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 $y^{(0)} \in \{0,1\}^2$. ■

A.3.3 Moment functions for the AR(3) model with $T = 5$

In Section 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:

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

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

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

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

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

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

and

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

A.3.4 Proof of Lemma 4

This proof is analogous to the proof of Lemma 3.

A.3.5 AR(2) models with $T = 5$

Consider model (17) 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 3.1 and 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

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

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

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

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:

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

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

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 (19).

A.3.6 Proof of Lemma 5

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 (17), 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 $p \geq 4$. ■

A.3.7 Results for AR(p) model with $p \geq 3$, $T = 4$ and $x_3 = x_4$

In Section 3.4 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 3.4 by one time period. For $x = (x_1, x_2, x_3, x_3)$, we have[28]

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

The second valid moment function is obtained from $m_{(0,0)}^{(d,2,4)}(y, x, \beta, \gamma)$ for $p = 2$ from Section 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 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 7 *If the outcomes $Y = (Y_1, Y_2, Y_3, Y_4)$ are generated from model (17) 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 $\xi \in \{a, b, c\}$*

[28]Obviously, we will not get any new identifying information from that moment function, beyond what was already discussed in Section 3.4, 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 5. ∎

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 3 *Let the outcomes* $Y = (Y_1, Y_2, Y_3, Y_4)$ *be generated from* (17) *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 2 we have provided conditions under which all the parameters β, γ of a panel logit AR(3) model are point-identified.