Identification of Dynamic Panel Logit Models with Fixed Effects

Christopher Dobronyi[*] Jiaying Gu[†] Kyoo il Kim[‡]

University of Toronto *University of Toronto* *Michigan State University*

April 15, 2021

Abstract

We show that the identification problem for a class of dynamic panel logit models with fixed effects has a connection to the *truncated moment problem* in mathematics. We use this connection to show that the sharp identified set of the structural parameters is characterized by a set of moment equality and inequality conditions. This result provides sharp bounds in models where moment equality conditions do not exist or do not point identify the parameters. We also show that the sharp identifying content of the non-parametric latent distribution of the fixed effects is characterized by a vector of its generalized moments, and that the number of moments grows linearly in T. This final result lets us point identify, or sharply bound, specific classes of functionals, without solving an optimization problem with respect to the latent distribution.

Keywords: Stieltjes Truncated Moment Problem, Dynamic Panel Logit Model, Fixed Effects, Moment Inequalities, Functionals of Latent Distribution

[*]Chris Dobronyi, Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, M5S3G7, Canada. Email: christopher.dobronyi@mail.utoronto.ca.

[†]Jiaying Gu, Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, M5S3G7, Canada. Email: jiaying.gu@utoronto.ca.

[‡]Kyoo il Kim, Department of Economics, Michigan State University, 486 W. Circle Dr, East Lansing, MI 48864, USA. Email: kyookim@msu.edu.

1 Introduction

We study the identification of a class of dynamic panel logit models with fixed effects (e.g., Chamberlain (1985), Honoré and Kyriazidou (2000), Hahn (2001), Bonhomme (2012), Honoré and Weidner (2020)). We show that the sharp identified set of the structural parameters is characterized by a set of moment equality and inequality conditions. These conditions render either point or set identification. Whether we can achieve point identification depends on the model specification and the number of observable time periods. Our approach to deriving the moment equality conditions is analytic and easy to implement. Our moment inequality conditions provide sharp bounds in models without moment equality conditions, and sharpen the identified set when moment equality conditions generate multiple solutions.[1] We also characterize the sharp identifying content of the non-parametric latent distribution to be a finite vector of its *generalized* moments. This last result allows us to identify a class of functionals, including average marginal effects and some other counterfactual parameters, which involve the latent distribution of the fixed effects. In particular, we characterize conditions when functionals of the latent distribution can be point identified even though the distribution itself is not point identified. We also characterize conditions when functionals can be sharply bounded without solving an optimization problem with respect to the latent distribution.

The dynamic panel logit model is an indispensable empirical tool for modeling repeated choices made by households, firms and individual consumers. It is commonly used, in part, because it can easily account for *permanent unobserved heterogeneity*, letting us distinguish between *true dynamics*, induced by lagged choice dependence, and *spurious dynamics*, as a result of individual heterogeneity (Heckman (1981a)). The longitudinal perspective of the data provides opportunities and challenges for such models with unobserved heterogeneity. The challenges are mainly due to the well known incidental parameter problem and the initial condition problem when the number of time periods T is fixed. The incidental parameter problem makes it difficult to consistently estimate structural parameters that capture the

[1] In our example of the dynamic panel logit model with a time trend or time dummies, we have two roots from all the moment equality conditions, one of them a false root. We can easily rule out this false root by examining the sign of a linear combination of the observed choice probabilities.

true dynamics. When T is fixed, it is generally not possible to treat the individual fixed effects as parameters and estimate them consistently in nonlinear panel models, which then contaminates the estimation of structural parameters (Neyman and Scott (1948)). The initial condition problem states that the joint distribution of the initial value of the choices and the unobserved heterogeneity is not nonparametrically point identified, making it hard to estimate counterfactual parameters, like the average marginal effect, and other functionals of the distribution of the fixed effects (e.g., Heckman (1981b) and Wooldridge (2005a)).

There are two common approaches to deal with these challenges.[2] The random effect approach places a restriction on the joint distribution of the initial condition and the unobserved heterogeneity through a parametric distribution, or a finite mixture model assumption (e.g., Chamberlain (1980) and Wooldridge (2005b)). If these assumptions are satisfied, then structural parameters and functionals of the latent distribution can be point identified and consistently estimated. However, if they are not satisfied, we can obtain misleading results.

Conversely, the fixed effect approach is entirely non-parametric with respect to the unobserved heterogeneity, but it often focuses only on the identification and estimation of the structural parameters. While this approach avoids imposing restrictive assumptions, its use can be limited since researchers frequently care about counterfactual parameters.

In the fixed effect approach, the structural parameters can often be identified and consistently estimated using the *conditional maximum likelihood method*, pioneered by Andersen (1970) and Chamberlain (1985). This method involves finding a minimally sufficient statistic for the fixed effect, and constructing a partial likelihood that conditions on this statistic. By the definition of sufficiency, this partial likelihood no longer depends on the fixed effects. If this partial likelihood depends on the structural parameters, then we obtain moment equality conditions that can be used for identification and estimation. Aguirregabiria, Gu, and Luo (2020) extend this approach to structural dynamic logit models in which agents make forward-looking choices. This method is easy to implement, but it does not always result in useful moment equality conditions, and even if it does, it can fail to find *all* of the relevant

[2] For a more complete survey of the literature, we refer the readers to Arellano and Honoré (2001).

moment conditions.[3]

In a recent paper by Honoré and Weidner (2020), the authors apply the functional differencing approach in Bonhomme (2012) in order to find new moment equality conditions for the structural parameters (in addition to those found using the sufficient statistics approach) in the AR(1) dynamic panel logit model with covariates.[4] They also find moment equality conditions in models for which the sufficient statistics approach provides no moment conditions, as in the AR(2) dynamic panel logit model. The functional differencing approach is able to find more moment equalities than the sufficient statistics approach because it searches for them (numerically) using the *full likelihood*. This result leads to an intriguing question: Is the factorization of the full likelihood through sufficiency the optimal one for identifying the structural parameters?

We propose a new formulation of the full likelihood. This formulation reveals a polynomial structure for the fixed effects in the dynamic panel logit models. It paves the way for an algebraic approach for constructing *all* moment equality conditions for the structural parameters. This approach involves finding the basis of the left null space of a matrix that only depends on the structural parameters. When the left null space is of zero dimension, the model does not provide any moment equality conditions. Else, the resulting moment equality conditions coincide with the set of moment equality conditions derived using the functional differencing approach. Thus, our results complement Honoré and Weidner (2020) by providing an algebraic approach for deducing the number and the form of the moment equality conditions for the structural parameters. Since we take an analytical approach, we can generate these moment conditions without the need for a numerical search.

Our formulation also reveals that, in addition to moment equalities, there are informative

[3]There is one exception. If the likelihood of the sufficient statistics no longer depends on the structural parameters, then the conditional maximum likelihood method guarantees to utilize all relevant information on the structural parameters. In most cases, as in the dynamic panel logit model, this condition is not satisfied.

[4]For the AR(1) model, Kitazawa (2021) also derives a set of moment equality conditions using a transformation method. Neither Honoré and Weidner (2020) or Kitazawa (2021) consider moment inequalities, and they refer to moment equality conditions as moment conditions. We make the distinction here between moment equality and inequality conditions for our analysis. Bajari, Hahn, Hong, and Ridder (2011) also use a functional differencing approach to generate moment equality conditions in static discrete games with complete or incomplete information. The latent variable in their context is the equilibrium selection mechanism when multiple equilibria exist.

moment inequalities. These moment inequalities are derived using a rich set of mathematical results on the *truncated moment problem*, dating back to Chebyshev (1874). In particular, we show that the resulting moment equality and inequality conditions characterize the *sharp identified set* of the structural parameters. Using this result, we are able to (i) construct the sharp identified set of the structural parameters when moment equality conditions are not available, as in the AR(1) dynamic panel logit model with two periods, and (ii) rule out false roots in models where moment equality conditions solve for multiple roots, as in the AR(1) dynamic panel logit model with only a time trend covariate (see Section 2.1.3 in Honoré and Weidner (2020)). We further show that the model with time dummies presents a similar feature. In each of these models, the moment equality conditions pin down a finite set of points as potential candidates for the parameter values, and the moment inequality conditions rule out the false roots, rendering point identification.

The literature on dynamic discrete choice models proposes ways to obtain sharp identified set of structural parameters and functionals of the latent distribution via optimization. For instance, the linear programming approach in Honoré and Tamer (2006), or the quadratic programming approach in Chernozhukov, Fernández-Val, Hahn, and Newey (2013), can be applied for this purpose. In particular, optimization can be used to search for the existence of a probability measure for the fixed effects that rationalizes the population choice probabilities for a given value of the structural parameters. This collects the set of all parameters and latent distributions that produce an observationally equivalent model. After solving for this set, we can construct sharp bounds for any integrable functional of the latent distribution of the fixed effects.

The optimization approaches, described above, are widely applicable, and can be used for models beyond those with a logistic error assumption. However, they are challenging from a practical point-of-view, because identification is characterized using an *infinite dimensional existence problem*. By focusing on dynamic panel logit models, we are able to characterize the sharp identified set for the structural parameters using moment equality and inequality conditions. Our approach is attractive because it converts the *infinite dimensional existence problem* into a set of moment conditions, and therefore avoids optimization with respect to the latent distribution. Importantly, we also show that, the sharp identifying content of the latent distribution is a finite vector of generalized moments and that the length of this vector grows only linearly in T.[5] Using this result, for a class of functionals, we show it is possible to profile out the latent distribution and convert the infinite dimensional optimization problem to a finite dimensional one to obtain sharp bounds.

As an example of the results described above, we show that the average marginal effect of the dynamic panel logit model is a linear combination of the vector of generalized moments, which we can learn directly from the data as soon as the structural parameters are point identified. This explains why it is possible to point identify the average marginal effect even when the latent distribution itself is not point identified. This echoes the important findings in Aguirregabiria and Carro (2020), who are the first to show that the average marginal effect is point identified in a class of dynamic panel logit models. We generalize their results by providing conditions on functionals, under which these functionals are point identified and provide examples of other counterfactual parameters that satisfy these conditions. Moreover, while Aguirregabiria and Carro (2020) restrict their attention to models in which the structural parameters are point identified (in order to make use of a sequential identification argument), we are able to consider models in which the structural parameters are only partially identified, and sharply bound functionals using a *finite* dimensional optimization problem.

The rest of the paper is organized as follows. Section 2 introduces the identification problem, and provides a simple example in order to illustrate our approach. Section 3 discusses our general results. Section 4 provides a collection of informative examples. These examples include the AR(1) and AR(2) dynamic panel logit models with and without a time trend, time dummies, and covariates. Section 5 concludes. Technical details, proofs and algebraic derivations are gathered in the Appendix.

[5] Both the conditional maximum likelihood approach in Chamberlain (1985) and the functional differencing approach by Bonhomme (2012) aim at differencing out the fixed effects to derive moment conditions for the structural parameters, and do not consider the identification of the latent distribution.

2 Dynamic Panel Logit Model with Fixed Effects

As our baseline, we consider a dynamic panel logit model with one lag, covariates, and fixed effects, defined by:

$$Y_{it} = \mathbb{1}\{\alpha_i + \beta Y_{it-1} + X_{it}'\gamma \geq \epsilon_{it}\}, \tag{2.1}$$

where we observe outcomes $Y_{it} \in \{0, 1\}$ and the covariates X_{it} for all individuals $i = 1, \ldots, n$ over time periods $t = 0, 1, \ldots, T$. The latent variable α_i is the permanent unobserved heterogeneity, allowed to have a nonparametric distribution that depends on the initial choice Y_{i0} and the covariates $X_i = (X_{i1}, \ldots, X_{iT})$. Covariates are strictly exogenous with respect to the error term ϵ_{it} and have discrete support $\mathcal{X}$ with cardinality $|\mathcal{X}|$. The error term is independently and identically distributed with respect to a standard logistic distribution. Since we focus on identification, we drop the individual index i for the rest of the paper, unless explicitly needed. Throughout, our identification analysis is conditional on Y_0 taking a fixed value y_0.[6]

Let $\theta = \{\beta, \gamma\}$ denote the structural parameters, and let $\mathcal{Y}$ denote the set containing all possible choice history, $\mathcal{Y} := \{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^J\}$, for which $J = 2^T$. Then, the likelihood of the choice history $\boldsymbol{y}^j$ conditional on $(y_0, \boldsymbol{x}, \alpha)$ equals:

$$\mathbb{P}((Y_1, \ldots, Y_T) = \boldsymbol{y}^j \mid Y_0 = y_0, X = \boldsymbol{x}, \alpha) := \mathcal{L}_j(\alpha, \theta, \boldsymbol{x}, y_0) = \prod_{t=1}^{T} \frac{\exp(\alpha + \beta y_{t-1} + \gamma' x_t)^{y_t}}{1 + \exp(\alpha + \beta y_{t-1} + \gamma' x_t)}, \tag{2.2}$$

where $\boldsymbol{y}^j = (y_1, \ldots, y_T)$ and $\boldsymbol{x} = (x_1, \ldots, x_T)$. Integrating out the fixed effects leads to the population choice probability, denoted as $\mathcal{P}_j = \mathbb{P}((Y_1, \ldots, Y_T) = \boldsymbol{y}^j | Y_0 = y_0, X = \boldsymbol{x})$. We further denote the vector $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \{\mathcal{P}_1, \ldots, \mathcal{P}_J\}$. When $\gamma = 0$, this model reduces to the model in Chamberlain (1985). This model can be generalized to incorporate more than one lag (see Section 4).

[6] By fixing the initial choice at y_0, we mute the variation of Y_0 so that it becomes clear what is the identifying content of the model for each fixed values of y_0. When there is indeed variation in Y_0, it can provide additional identifying constraints for the structural parameters because they do not change for different values of y_0. For instance, we will take an intersection of the identified set of the structural parameters constructed using $y_0 = 0$ and $y_0 = 1$. Our results can also be applied to cases where Y_0 is not observed, but is assumed to have a degenerate distribution. This may be a reasonable assumption in certain empirical applications.

2.1 Identification Analysis

For exposition we consider the one covariate case. The identification analysis easily extends to the multiple covariates with additional notation. Define $A = \exp(\alpha)$, $B = \exp(\beta)$, and $C = \exp(\gamma)$, and let $Q(A \mid y_0, \boldsymbol{x})$ denote the conditional distribution of A with support $\mathcal{A} = [0, \infty)$. The vector $\mathcal{P}_{\boldsymbol{x}}$ is assumed to be observed for our identification analysis. Furthermore, let $\boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0)$ denote the vector that stacks $\mathcal{L}_j(A, \theta, \boldsymbol{x}, y_0)$, for $j = 1, \ldots, J$. For each given θ, y_0, and $\boldsymbol{x}$, the identified set of the latent distribution of the fixed effects is the set of probability measures Q on $\mathcal{A}$ that rationalize the population choice probability $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$:

$$\mathcal{Q}(\theta, y_0, \boldsymbol{x}) = \{Q : \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_{\mathcal{A}} \boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) dQ(A|y_0, \boldsymbol{x})\}. \tag{2.3}$$

Definition 2.1 (Identified Set)**.** *The identified set of the structural parameter θ is defined by:*

$$\Theta^* = \{\theta : \mathcal{Q}(\theta, y_0, \boldsymbol{x}) \neq \emptyset, \ \textit{for all } \boldsymbol{x} \in \mathcal{X}\}.$$

Moreover, the joint identified set of the structural parameter θ and the latent distribution is defined by:

$$\mathcal{I}^*(y_0, \boldsymbol{x}) = \{(\theta, Q) : \theta \in \Theta^* \ \textit{and } Q \in \mathcal{Q}(\theta, y_0, \boldsymbol{x})\}.$$

If Θ^* is a singleton set, then θ is point identified, and the true distribution of the fixed effects, denoted as $Q_0(A \mid y_0, \boldsymbol{x})$, is known to be a member of $\mathcal{Q}(\theta_0, y_0, \boldsymbol{x})$ where θ_0 denotes the true value of θ. The question of whether a point θ belongs to the identified set Θ^* can be viewed as an *infinite dimensional existence problem*—it asks whether there exists a probability measure Q such that the observed vector of choice probabilities can be rationalized by the model given θ. We now show that we can reduce this *infinite dimensional existence problem* to a *finite* set of moment equality and inequality conditions, for each $\boldsymbol{x} \in \mathcal{X}$.

We first present this result for the simple case in which $T = 2$ and $\gamma = 0$ to establish intuition. This simple case reveals a fundamental connection between our identification problem and the *truncated moment problem* in mathematics. We use this simple case to motivate our general identification results in Section 3.

This simple case is also interesting by itself. Honoré and Weidner (2020) have shown that there are no moment equality conditions. We show that the model still provides information about the structural parameters through a finite set of moment inequalities that define the sharp identified set. To the best of our knowledge, our paper is the first to establish this sharp identification result for this model.

2.2 Simple Case: Two Time Periods without Covariates

Consider the simple case described above in which $T = 2$ and $\gamma = 0$. For exposition, we fix $y_0 = 0$.[7] The likelihood vector can be denoted $\boldsymbol{\mathcal{L}}(A, \beta)$ and written as:[8]

$$\boldsymbol{\mathcal{L}}(A, \beta) = G(\beta) \begin{pmatrix} 1 \\ A \\ A^2 \\ A^3 \end{pmatrix} \frac{1}{g(A, \beta)}, \tag{2.4}$$

with $g(A, \beta) = (1 + A)^2(1 + AB)$ and the matrix $G(\beta)$ is defined by:

$$G(\beta) = \begin{pmatrix} 1 & B & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & B & 0 \\ 0 & 0 & B & B \end{pmatrix}.$$

Therefore, in this model, we can write:

$$\boldsymbol{\mathcal{P}} = G(\beta) \int_{\mathcal{A}} \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}' d\bar{Q}(A \mid \beta), \tag{2.5}$$

where $\boldsymbol{\mathcal{P}}$ is the vector of population choice probabilities and $d\bar{Q}(A \mid \beta) = \frac{1}{g(A,\beta)} dQ(A)$. Since $1/g(A, \beta)$ is bounded for all $A \in \mathcal{A}$, the measure $\bar{Q}(A \mid \beta)$ is a finite positive Borel measure on $\mathcal{A}$. It is easy to check that $G(\beta)$ is of full rank unless $\beta = 0$ (which we rule out).[9]

There are several features of the formulation in (2.4) that are worth mentioning. First, the choice of $g(A, \beta)$ is natural. Since $\mathcal{L}_j(A, \beta)$ is a ratio of polynomials of A, we can choose

[7] The same results derived can be extended to the case where $y_0 = 1$.

[8] The elements in the set $\mathcal{Y}$ is ordered as: $\{(0, 0), (1, 0), (0, 1), (1, 1)\}$. See details in Appendix A.2.

[9] If $\beta = 0$, then we have a static logit model, also known as the Rasch model by Rasch (1961). Identification of this model is well understood. See for instance Cressie and Holland (1983).

$g(A, \beta)$ to be a polynomial of A of the smallest degree for which the product $\boldsymbol{\mathcal{L}}_j(A, \beta) g(A, \beta)$ is a polynomial of A, for all $j = 1, \ldots, J$. Doing so leads to a matrix $G(\beta)$ that does not depend on the fixed effects. Second, the functions in the vector $\{1/g, A/g, A^2/g, A^3/g\}$ are linearly independent with support $\mathcal{A}$. We prove this result in Lemma 3.1.

Since $G(\beta)$ is of full rank, we obtain:

$$\boldsymbol{r}(\beta) = (r_0(\beta), \ldots, r_3(\beta))' := G(\beta)^{-1} \boldsymbol{\mathcal{P}} = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}' d\bar{Q}(A \mid \beta).$$

Notice that, for a given β, the vector $\boldsymbol{r}(\beta)$ is observed (since $\boldsymbol{\mathcal{P}}$ denotes the observed vector of choice probabilities and $G(\beta)^{-1}$ is a known matrix). The right-hand side of this equation is a vector of moments up to order 3 with respect to the measure $\bar{Q}(A \mid \beta)$. Hence, the question of whether a particular value of β belongs to the identified set translates into the question of whether there exists a finite positive Borel measure for which the vector $\boldsymbol{r}(\beta)$ can be written as a vector of moments up to order 3. This result reveals the fundamental connection to the *truncated moment problem*, which we alluded to in the previous section. In particular, one of the questions studied in the literature on the *truncated moment problem* is whether there exists a Borel measure that rationalizes a finite sequence of numbers as its moments.

To this end, we define the moment space of any positive Borel measure μ on $[0, \infty)$ to be:

$$\mathcal{M}_K = \left\{ \boldsymbol{c} \in \mathbb{R}^{K+1} : \text{ there exists } \mu \text{ such that } c_k = \int_0^{+\infty} A^k d\mu(A), \text{ for all } k = 0, 1, \ldots, K \right\}.$$

With this definition, we can write $\Theta^* = \{\beta : \boldsymbol{r}(\beta) \in \mathcal{M}_3\}$ in this simple case. The unique geometric structure of the moment space $\mathcal{M}_K$ leads to the following theorem.

Theorem 2.1. *For the dynamic logit model in (2.1) with $T = 2$ and $\gamma = 0$, the value $\beta \in \Theta^*$ if and only if $\sum_{j=0}^{3} \eta_j r_j(\beta) \geq 0$, for every non-trivial real-valued sequence of coefficients $\{\eta_j\}_{j=0}^{3}$ such that $\sum_{j=0}^{3} \eta_j A^j \geq 0$, for all $A \in [0, \infty)$.*

Theorem 2.1 is Theorem 9.1 in Karlin and Studden (1966), applied to our context. The key insight is that the dual cone[10] of the moment space, which is itself a convex cone, can be identified as the space of non-negative polynomials of A up to degree K. Now we discuss the

[10]For a convex cone $\mathcal{C}$ contained in $\mathbb{R}^d$, the dual cone $\mathcal{C}^+ = \{\boldsymbol{\lambda} \in \mathbb{R}^d \mid \boldsymbol{\lambda}' \boldsymbol{c} \geq 0 \text{ for all } \boldsymbol{c} \in \mathcal{C}\}$.

implication of Theorem 2.1. Since every non-negative polynomial of A with an odd degree has a represention with the form:[11]

$$\sum_{j=0}^{2m+1} \eta_j A^j = Af^2(A) + q^2(A),$$

for all $A \in [0, \infty)$, where $f(A)$ and $q(A)$ are polynomials. In our context, we have $m = 1$, and $f(A)$ and $q(A)$ are polynomials of at most degree 1. Therefore, we can write $f(A) = \xi_0 + \xi_1 A$ and $q(A) = \lambda_0 + \lambda_1 A$, for any coefficients (ξ_0, ξ_1) and (λ_0, λ_1) in $\mathbb{R}^2$ such that:

$$\sum_{j=0}^{3} \eta_j A^j = A(\xi_0 + \xi_1 A)^2 + (\lambda_0 + \lambda_1 A)^2 \geq 0.$$

Retrieving the coefficient η_j yields a constraint on the vector $\boldsymbol{r}(\beta)$:

$$\begin{pmatrix} \lambda_0 & \lambda_1 \end{pmatrix} \begin{pmatrix} r_0(\beta) & r_1(\beta) \\ r_1(\beta) & r_2(\beta) \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \lambda_1 \end{pmatrix} + \begin{pmatrix} \xi_0 & \xi_1 \end{pmatrix} \begin{pmatrix} r_1(\beta) & r_2(\beta) \\ r_2(\beta) & r_3(\beta) \end{pmatrix} \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} \geq 0,$$

for any coefficients (λ_0, λ_1) and (ξ_0, ξ_1) in $\mathbb{R}^2$.

This last condition, as we show later in Theorems 2.2 and 2.3 (in a more general form), boils down to checking the non-negativity of the two matrices above, defined using the elements of $\boldsymbol{r}(\beta)$. Since the non-negativity of a square matrix is equivalent to all of its principal minors being non-negative, and $\boldsymbol{r}(\beta)$ is a linear combination of choice probabilities with coefficients depend on β, we obtain *moment inequalities* for β. In fact, these moment inequalities characterize *all* of the information about β implied by this model, and the sharp bounds for β implied by these moment inequalities can be derived *analytically* (see Sections 3 and 4.2).

This model has a polynomial structure with respect to the fixed effects that we exploit for identification. It turns out, this structure is present in the dynamic panel logit model with and without covariates, for any finite T. In Section 3, we formally prove this result, and provide the following additional generalizations:

---

[11] Every non-negative polynomial up to degree K, $\sum_{j=0}^{K} \eta_j A^j$, over $[0, \infty)$ can be written in the form: (i) if $K = 2m+1$ (odd case), then $\sum_{j=0}^{2m+1} \eta_j A^j = Af^2(A) + q^2(A)$ where $f(A)$ and $q(A)$ are polynomial functions of A up to order m. If $K = 2m$ (even case), then $\sum_{j=0}^{2m} \eta_j A^j = f^2(A) + Aq^2(A)$ where $f(A)$ are polynomials of A of at most order m and $q(A)$ is a polynomial of A up to order $m - 1$. See Karlin and Studden (1966). We are in the odd case here because we have moments up to order 3.

1. When $T > 2$, we obtain moment equalities, in addition to moment inequalities. The number of available moment equalities is determined by the dimension of the left null space of $G(\beta)$. We can find these moment equalities by deriving its basis. A detailed formulation of this result is provided in Theorem 3.1, and its connection to the results in Honoré and Weidner (2020) is provided in Section 3.1.1.

2. Our moment space characterization leads to a sharp characterization of the latent distribution of the fixed effects that is useful for identifying functionals of Q, including some important counterfactual parameters. The identifying content of Q is provided in Theorem 3.3; its application to the identification of functionals is provided in Section 3.2.

2.3 Results on the Truncated Moment Problem

The polynomial reformulation in (2.4), alongside the characterization of the identified set Θ^* through the moment space $\mathcal{M}_3$, brings us into the *moment problem*. The *moment problem* poses the question: Is a sequence of real numbers equal to the sequence of moments associated with some finite Borel measure supported on some set $\mathbb{K}$? When the sequence of real numbers is an infinite sequence, we have the *full moment problem*. Else, we have the *truncated moment problem*. When $\mathbb{K} = [0, \infty)$, as in our context, the truncated moment problem becomes the *truncated Stieltjes moment problem*. We refer the readers to the authoritative treatments of the moment problem in Karlin and Studden (1966) and Krein and Nudelman (1977).

In our context, the finite sequence of real numbers is the vector $\boldsymbol{r}(\beta)$, and we have the following general theorem. For any $m \times n$ matrix $\boldsymbol{A}$, define $\text{Range}(\boldsymbol{A}) = \{\boldsymbol{A}\boldsymbol{u}, \boldsymbol{u} \in \mathbb{R}^n\}$. The symbol $\succeq 0$ represents a square matrix being positive semidefinite.

Theorem 2.2 (Truncated Stieltjes Moment Problem (odd case))**.** *Let* $\boldsymbol{c} = \{c_0, c_1, \ldots, c_m\} \in \mathbb{R}^{m+1}$ *denote a finite dimensional vector. If* m *is odd (i.e.* $m = 2k+1$*), define the matrix*

$$H_k(\boldsymbol{c}) = \begin{pmatrix} c_0 & c_1 & \cdots & c_k \\ c_1 & c_2 & \cdots & c_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_k & c_{k+1} & \cdots & c_{2k} \end{pmatrix} \quad \text{and} \quad B_k(\boldsymbol{c}) = \begin{pmatrix} c_1 & c_2 & \cdots & c_{k+1} \\ c_2 & c_3 & \cdots & c_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k+1} & c_{k+2} & \cdots & c_{2k+1} \end{pmatrix}$$

Then $\boldsymbol{c} \in \mathcal{M}_{2k+1}$ *if and only if* $H_k(\boldsymbol{c}) \succeq 0$, $B_k(\boldsymbol{c}) \succeq 0$ *and* $\{c_{k+1}, c_{k+2}, \ldots, c_{2k+1}\}$ *is in Range*$(H_k(\boldsymbol{c}))$.

Theorem 2.3 (Truncated Stieltjes Moment Problem (even case)). *Let* $\boldsymbol{c} = \{c_0, c_1, \ldots, c_m\} \in \mathbb{R}^{m+1}$ *denote a finite dimensional vector. If* m *is even (i.e.* $m = 2k$*), define the matrix*

$$H_k(\boldsymbol{c}) = \begin{pmatrix} c_0 & c_1 & \cdots & c_k \\ c_1 & c_2 & \cdots & c_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_k & c_{k+1} & \cdots & c_{2k} \end{pmatrix} \quad \text{and} \quad B_{k-1}(\boldsymbol{c}) = \begin{pmatrix} c_1 & c_2 & \cdots & c_k \\ c_2 & c_3 & \cdots & c_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_k & c_{k+1} & \cdots & c_{2k-1} \end{pmatrix}$$

Then $\boldsymbol{c} \in \mathcal{M}_{2k}$ *if and only if* $H_k(\boldsymbol{c}) \succeq 0$, $B_{k-1}(\boldsymbol{c}) \succeq 0$ *and* $\{c_{k+1}, c_{k+2}, \ldots, c_{2k}\}$ *is in the Range*$(B_{k-1}(\boldsymbol{c}))$.

Proof. See Curto and Fialkow (1991). ■

In Theorems 2.2 and 2.3, $H_k(\boldsymbol{c})$ is a *Hankel matrix*, and $B_k(\boldsymbol{c})$ and $B_{k-1}(\boldsymbol{c})$ are localized verions of $H_k(\boldsymbol{c})$ (in which the moments are shifted by one index). The condition on the range is only relevant when $H_k(\boldsymbol{c})$ is singular. Hankel matrices have a recursive structure, as shown by Curto and Fialkow (1991). Therefore, if a Hankel matrix is singular and positive, then its entries are uniquely recursively determined by elements in a non-singular sub-Hankel matrix (of a smaller dimension). This is a feature of the results on the existence of a representing measure in the *truncated moment problem*. See an explicit discussion of the condition on the range for model (2.1) with $T = 2$ and $\gamma = 0$ in Section 4.2.

3 General Results for the AR(1) Model

With the results from dynamic panel logit model in which $T = 2$ and $\gamma = 0$ in hands, we now generalize. For model (2.1) with $T \geq 2$ and $\theta = (\beta, \gamma)$, there exists a polynomial function

$g(A, \theta, \boldsymbol{x}, y_0)$ of the order $2T - 1$ such that,

$$\boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) = G(\theta, \boldsymbol{x}) \begin{pmatrix} 1 \\ A \\ \vdots \\ A^{2T-1} \end{pmatrix} \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)}, \tag{3.1}$$

where $G(\theta, \boldsymbol{x})$ is a matrix of dimension $2^T \times 2T$. See Appendix A.6 for an explicit formulation of $g(A, \theta, \boldsymbol{x}, y_0)$. Then the population choice probabilities have the following representation:

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_{\mathcal{A}} \boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) dQ(A|y_0, \boldsymbol{x}) = G(\theta, \boldsymbol{x}) \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' d\bar{Q}(A|y_0, \theta, \boldsymbol{x}) \tag{3.2}$$

where $d\bar{Q}(A|y_0, \theta, \boldsymbol{x}) = \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A|y_0, \boldsymbol{x})$. Since $1/g(A, \theta, \boldsymbol{x}, y_0)$ is bounded everywhere on $\mathcal{A}$, the measure $\bar{Q}(A|y_0, \theta, \boldsymbol{x})$ is a positive Borel measure supported on $\mathcal{A}$. We further define the set of vectors:

$$\boldsymbol{M}_{\boldsymbol{x}} = \{\boldsymbol{v}_{\boldsymbol{x}} \in \mathbb{R}^{2^T} : \boldsymbol{v}'_{\boldsymbol{x}} G(\theta, \boldsymbol{x}) = 0\}, \tag{3.3}$$

where the vector $\boldsymbol{v}_{\boldsymbol{x}}$ is implicitly a function of the parameter θ and $\boldsymbol{x}$. By construction, $\boldsymbol{v}_{\boldsymbol{x}}$ is a member of the left null space of $G(\theta, \boldsymbol{x})$. Also, both $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ and $G(\theta, \boldsymbol{x})$ (and hence $\boldsymbol{v}_{\boldsymbol{x}}$) depend on y_0, and all these terms are defined given a fixed value of y_0. Since we take a conditional approach, for ease of notation, we suppress their dependence on y_0 unless we derive them separately for $y_0 = 0$ and $y_0 = 1$.

3.1 Identification of Structural Parameters

We now state out main result:

Theorem 3.1. *If $G(\theta, \boldsymbol{x})$ is full rank, then $\theta \in \Theta^*$ if and only if the following conditions hold:*

(a) For all $\boldsymbol{x} \in \mathcal{X}$, we have $\boldsymbol{v}'_{\boldsymbol{x}} \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = 0$ for all $\boldsymbol{v}_{\boldsymbol{x}} \in \boldsymbol{M}_{\boldsymbol{x}}$.

(b) For all $\boldsymbol{x} \in \mathcal{X}$, we have $\boldsymbol{r}(\theta, \boldsymbol{x}) \in \mathcal{M}_{2T-1}$, where $\boldsymbol{r}(\theta, \boldsymbol{x}) = H(\theta, \boldsymbol{x}) \boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ and $H(\theta, \boldsymbol{x})$ is a matrix of dimension $2T \times 2^T$ such that $H(\theta, \boldsymbol{x}) G(\theta, \boldsymbol{x}) = I_{2T}$.

The requirement on $G(\theta, \boldsymbol{x})$ being full rank can be checked in each specific model. Theorem 3.1 shows that identification of the structural parameters θ can be formulated as the problem of *moment equalities* (from condition (a)) and *moment inequalities* (from condition (b)). The moment equality conditions take the form $\mathbb{E}[\boldsymbol{v}_{\boldsymbol{x}}' \mathbb{Y} | Y_0 = y_0, X = \boldsymbol{x}] = \boldsymbol{v}_{\boldsymbol{x}}' \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = 0$, for each $\boldsymbol{x} \in \mathcal{X}$, where $\mathbb{Y}$ is a vector of length 2^T with its elements being $\mathbb{1}\{(Y_1, \ldots, Y_T) = \boldsymbol{y}^j\}$ for $\boldsymbol{y}^j \in \mathcal{Y}$ (this is due to the fact that $\boldsymbol{v}_{\boldsymbol{x}}' G(\theta, \boldsymbol{x}) = 0$ implies $\boldsymbol{v}_{\boldsymbol{x}}' \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = 0$ given our representation in (3.2)). The form of these moment equalities can be found as the basis of the left null space of $G(\theta, \boldsymbol{x})$, defined in $\boldsymbol{M}_{\boldsymbol{x}}$.[12] The moment inequalities arise from the condition that $\boldsymbol{r}(\theta, \boldsymbol{x})$ has to be a moment sequence, and the fact that $\boldsymbol{r}(\theta, \boldsymbol{x})$ is a linear combination of the elements in $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$. Through the necessary and sufficient conditions of the *truncated moment problem* in Theorem 2.2 or Theorem 2.3, we impose non-negativity on the two Hankel matrices formed from elements in $\boldsymbol{r}(\theta, \boldsymbol{x})$, which is equivalent to all of its principal minors being non-negative. These constraints on $\boldsymbol{r}(\theta, \boldsymbol{x})$ lead to moment inequality conditions for θ. For the construction of $\boldsymbol{r}(\theta, \boldsymbol{x})$, in principle, the matrix $H(\theta, \boldsymbol{x})$ can always be constructed as $(G(\theta, \boldsymbol{x})' G(\theta, \boldsymbol{x}))^{-1} G(\theta, \boldsymbol{x})'$. In Appendix A.3, we discuss an alternative procedure.

3.1.1 Relationship with Results in Honoré and Weidner (2020)

We now make a connection to the results in Honoré and Weidner (2020). In this paper, the authors numerically search for moment equality conditions for θ using the functional differencing approach in Bonhomme (2012). To be precise, let us define the set:

$$\kappa_{\boldsymbol{x}} = \{\boldsymbol{m}_{\boldsymbol{x}} \in \mathbb{R}^{2^T} : \boldsymbol{m}_{\boldsymbol{x}}' \boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) = 0, \forall A \in \mathcal{A}\}. \tag{3.4}$$

Then, the moment conditions in Honoré and Weidner (2020) consist of, for each $\boldsymbol{x} \in \mathcal{X}$, $\mathbb{E}[\boldsymbol{m}_{\boldsymbol{x}}' \mathbb{Y} | Y_0 = y_0, X = \boldsymbol{x}] = \boldsymbol{m}_{\boldsymbol{x}}' \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = 0$ for all $\boldsymbol{m}_{\boldsymbol{x}} \in \kappa_{\boldsymbol{x}}$. We now prove that the set $\boldsymbol{M}_{\boldsymbol{x}}$ defined in (3.3) coincides with the set $\kappa_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{X}$.

Lemma 3.1. *The functions in* $\mathcal{V}_{\theta, \boldsymbol{x}, y_0}(A) = \left\{ \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)}, \frac{A}{g(A, \theta, \boldsymbol{x}, y_0)}, \cdots, \frac{A^{2T-1}}{g(A, \theta, \boldsymbol{x}, y_0)} \right\}$ *are linearly*

[12] An earlier working paper version of Buchinsky, Hahn, and Kim (2010) shows that using Johnson (2004) characterization of semiparametric information bound for discrete choice models one can derive the moment equality conditions for a few examples of dynamic panel logit models with fixed effects with $T \leq 3$. They, however, do not consider a more general dynamic panel logit specification like model (2.1) with covariates, or the moment inequalities that we develop in this paper.

independent.

Given the formulation in (3.1) for $\boldsymbol{\mathcal{L}}$, the vectors in the set $\kappa_{\boldsymbol{x}}$ are those such that $\boldsymbol{m}_{\boldsymbol{x}}' G(\theta, \boldsymbol{x})$ are orthogonal to the vector $\mathcal{V}_{\theta,\boldsymbol{x},y_0}(A)$ for all $A \in \mathcal{A}$. Lemma 3.1 states that the set of functions in $\mathcal{V}_{\theta,\boldsymbol{x},y_0}$ spans $\mathbb{R}^{2T}$, which implies that $\boldsymbol{m}_{\boldsymbol{x}}' G(\theta, \boldsymbol{x}) = 0$ for all $\boldsymbol{m}_{\boldsymbol{x}} \in \kappa_{\boldsymbol{x}}$. This observation leads to the next Theorem.

Theorem 3.2. $\boldsymbol{M}_{\boldsymbol{x}} = \kappa_{\boldsymbol{x}}$ *for all* $\boldsymbol{x} \in \mathcal{X}$.

Theorem 3.2 proves that the moment equality conditions found by our approach coincide with the moment conditions in Honoré and Weidner (2020). Our moment equality conditions can always be written as linear combinations of the moment conditions in Honoré and Weidner (2020), and vice versa.

Honoré and Weidner (2020) find moment conditions numerically by applying the functional differencing method. Given the equivalence in Theorem 3.2, our results provide an algebraic foundation for their results. In addition, they use numerical evidence to conjecture that there are $2^T - 2T$ moment conditions, for each given $\boldsymbol{x}$. This result can be easily verified in our framework. Provided that $G(\theta, \boldsymbol{x})$ is of full rank (a condition that can be verified in each specific model), the left null space of $G(\theta, \boldsymbol{x})$ is of dimension $2^T - 2T$, for any $T \geq 2$. As a result, we expect to have $2^T - 2T$ linearly independent vectors that form a basis for this space, where each vector serves as a moment equality condition for identifying θ. This feature can be used to explain why there are no moment equality conditions available when $T = 2$, an impossibility result established in Honoré and Weidner (2020). Specifically, when $T = 2$, $G(\theta, \boldsymbol{x})$ is a 4×4 full rank square matrix, implying that its left null space is of zero dimension, and that there does not exist a vector $\boldsymbol{v}_{\boldsymbol{x}}$ for which $\boldsymbol{v}_{\boldsymbol{x}}' G(\theta, \boldsymbol{x}) = 0$ (with the exception of the null vector). In this case, all of the identifying content of θ is contained in the moment inequality conditions that we characterize in Theorem 3.1. Moreover, the moment equality conditions, if they exist, are often nonlinear in θ or $\exp(\theta)$. Hence, even if we have K moment equality conditions with K being greater or equal to the number of parameters θ, there is, in general, no guarantee that we will have point identification. We discuss two such cases in Section 4 for model (2.1): one with a time trend variable, and one with time dummies, for which we illustrate the moment inequality conditions become informative.

3.1.2 Decomposition of the Degree of Freedom in $\mathcal{P}_x$

We know that the vector $\boldsymbol{\mathcal{P}_x}$ is in the $2^T - 1$ unit simplex. Therefore, it has $2^T - 1$ degree of freedom. By definition the function $g(A, \theta, \boldsymbol{x}, y_0)$ is a polynomial function of A of degree $2T-1$. This feature implies that there exists a vector of constant $\boldsymbol{c} = \{c_0, \ldots, c_{2T-1}\}$, which possibly depends on $\{\theta, \boldsymbol{x}, y_0\}$, such that $\sum_{j=0}^{2T-1} c_j A^j / g(A, \theta, \boldsymbol{x}, y_0) = 1$. Combined with the result in Lemma 3.1, we can conclude that the vector $\boldsymbol{r}(\theta, \boldsymbol{x})$ has $2T-1$ degrees of freedom. This result implies that $\mathcal{P}_{\boldsymbol{x}}$ has the remaining $(2^T - 1) - (2T - 1) = 2^T - 2T$ degrees of freedom, and that these degrees of freedom are fully exploited by the moment equalities implied by condition (a) in Theorem 3.1. Among all of the examples that we consider in this paper, we are always able to construct a matrix $G(\theta, \boldsymbol{x})$ with full column rank. However, our characterization and decomposition can generalize to models for which $G(\theta, \boldsymbol{x})$ is not of full rank. See the discussion in Appendix A.3.

3.1.3 Practical Considerations

We characterize the sharp identified set for θ using moment equalities and inequalities. For small T, these moment conditions can be found analytically, as we do for model (2.1), with $T = 2$ and $\gamma = 0$, in Section 4. For large T, it may be difficult to analytically derive the moment inequalities. Fortunately, in practice, we do not need the form of these inequalities to construct Θ^*. Theorem 3.1 offers a simple algorithm for constructing this set: When the moment equalities have a unique solution θ^* for θ, we obtain $\Theta^* = \{\theta^*\}$,[13] and when multiple distinct solutions exist, we can simply check whether the moment inequalities are satisfied at these solutions (instead of checking them for the whole parameter space) to construct Θ^*.

For each candidate value for θ (i.e., those that satisfy the moment equality conditions, if such conditions exist), checking the moment inequalities implied by Theorems 3.1, 2.2, and 2.3, boils down to checking the non-negativity of two square matrices, and checking whether a subvector of the moment sequence belongs to the range of a matrix when the Hankel matrices are singular. Therefore, checking the moment inequalities only involves

[13] In this case, moment inequalities can be used to check the validity of the model. Since Θ^* is non-empty (as it contains θ_0), if the moment inequalities are not satisfied at θ^*, then we can conclude that the model is wrong.

simple linear algebra operations.

3.2 Identification of Functionals of Unobserved Heterogeneity

As mentioned in Section 2.2, we can also deduce the identifying content of the latent distribution, and that this content is characterized by a finite vector of generalized moments. We now formalize this result.

Formulation (3.1) reveals that we have an equivalent definition of the set in (2.3) as

$$\mathcal{Q}(\theta, y_0, \boldsymbol{x}) = \{Q : \boldsymbol{r}(\theta, \boldsymbol{x}) = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ\}, \tag{3.5}$$

where $\boldsymbol{r}(\theta, \boldsymbol{x}) = H(\theta, \boldsymbol{x})\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ is the transformation of the population probability vector $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ defined in Theorem 3.1. We can use this definition to deduce the result below.

Theorem 3.3. *For each $\boldsymbol{x} \in \mathcal{X}$ and each value of $\theta \in \Theta^*$, the sharp identifying content of the latent distribution is the vector of generalized moments, $\mathbb{E}_Q[A^j/g(A, \theta, \boldsymbol{x}, y_0)]$ for $j = 1, 2, \ldots, 2T-1$ for all $Q \in \mathcal{Q}(\theta, y_0, \boldsymbol{x})$.*

Remark 3.1. *By definition, the function $g(A, \theta, \boldsymbol{x}, y_0)$ is a polynomial function of A of degree $2T-1$ (see Appendix A.6 for its explicit form). Therefore, there is a linear relationship between a non-zero constant and $\{1/g, A/g, \cdots, A^{2T-1}/g\}$, for all $A \in \mathcal{A}$ (i.e., these are $2T$ linearly independent functions that reside in $2T-1$ dimensional vector space). This result explains why, in Theorem 3.3, the index j starts at 1 instead of 0. In particular, moments $\mathbb{E}_Q[A^j/g]$, $j = 1, 2, \ldots, 2T-1$ determine the 'zero' moment $\mathbb{E}_Q[1/g]$. This result can also be seen from the fact that choice probabilities sum to one: $1 = \mathbf{1}'\mathcal{P}_{\boldsymbol{x}}$. Indeed, this property implies $1 = \mathbf{1}'\mathcal{P}_{\boldsymbol{x}} = \mathbf{1}'G(\theta, \boldsymbol{x})\mathbb{E}_Q[V(A)/g]$, where $V(A) = (1, A, \ldots, A^{2T-1})'$. Note that, $\mathbf{1}'G(\theta, \boldsymbol{x})$ is a known $1 \times 2T$ vector given θ.*

Now, suppose that we are interested in a functional:

$$\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})] = \int_{\mathcal{A}} \psi(A, \theta_0, \boldsymbol{x}) dQ_0(A|y_0, \boldsymbol{x}),$$

for some function ψ and value $\boldsymbol{x} \in \mathcal{X}$, where θ_0 is the true value of the structural parameter, and $Q_0(A|y_0, \boldsymbol{x})$ is the true latent distribution.

For the interpretation of the next result, recall that the identified set $\mathcal{I}^*(y_0, \boldsymbol{x})$, defined in Definition 2.1, is the set of pairs $\{\theta, Q\}$ such that $\theta \in \Theta^*$ and $Q \in \mathcal{Q}(\theta, y_0, \boldsymbol{x})$, where $\mathcal{Q}(\theta, y_0, \boldsymbol{x})$ is now equivalently defined in (3.5).

Definition 3.1. *The sharp identified set for* $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})]$ *is* $[\ell b(\boldsymbol{x}), ub(\boldsymbol{x})]$ *where:*

$$\ell b(\boldsymbol{x}) = \inf_{\{\theta,Q\} \in \mathcal{I}^*(y_0,\boldsymbol{x})} \int_{\mathcal{A}} \psi(A, \theta, \boldsymbol{x}) dQ(A),$$
$$ub(\boldsymbol{x}) = \sup_{\{\theta,Q\} \in \mathcal{I}^*(y_0,\boldsymbol{x})} \int_{\mathcal{A}} \psi(A, \theta, \boldsymbol{x}) dQ(A).$$

This bounding problem is not always tractable because, although we have characterized the identified set Θ^* for θ using a finite set of moment conditions, the bounds in Definition 3.1 involve optimizing over a set of distribution, and this set can contain infinitely many elements. However, Theorem 3.3 immediately suggests that when the parameters of interest can be represented as a linear combination of the generalized moments, we can construct sharp bound without needing to optimize with respect to the latent distribution Q. In a sense, we can profile out the distribution Q and obtain the sharp bound by solving a finite dimensional optimization problem with respect to θ. It turns out, a number of interesting parameters fall into this category.

To present the formal result, we distinguish between two cases: The case in which the identified set Θ^* is a singleton (so that θ is point identified), and the case in which the identified set Θ^* has more than one element.

3.2.1 When θ is Point Identified

When $\Theta^* = \{\theta_0\}$, we know by Definition 2.1 the true distribution $Q_0(A|y_0, \boldsymbol{x})$ is a member of $\mathcal{Q}(\theta_0, y_0, \boldsymbol{x})$, defined in (3.5). Then, $\boldsymbol{r}(\theta_0, \boldsymbol{x}) = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' \frac{1}{g(A,\theta_0,\boldsymbol{x},y_0)} dQ_0(A|y_0, \boldsymbol{x})$, which equals $H(\theta_0, \boldsymbol{x})\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ by Theorem 3.1. Since θ_0 is point identified, so will be $\boldsymbol{r}(\theta_0, \boldsymbol{x})$, which is a linear combination of $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$. We then obtain the following result:

Theorem 3.4. *For the model specified in* (2.1)*, if* θ *is point identified and the product* $\psi(A, \theta_0, \boldsymbol{x})g(A, \theta_0, \boldsymbol{x}, y_0)$ *is a polynomial function of* A *with a degree that is no larger than*

$2T-1$ *such that:*

$$\psi(A,\theta_0,\boldsymbol{x})g(A,\theta_0,\boldsymbol{x},y_0)=\sum_{j=0}^{2T-1}\eta_j(\theta_0,\boldsymbol{x})A^j,$$

for some vector $\boldsymbol{\eta}(\theta_0,\boldsymbol{x})=(\eta_0(\theta_0,\boldsymbol{x}),\eta_1(\theta_0,\boldsymbol{x}),\ldots,\eta_{2T-1}(\theta_0,\boldsymbol{x}))$*, then* $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A,\theta_0,\boldsymbol{x})]$ *is point identified and equal to* $\boldsymbol{\eta}(\theta_0,\boldsymbol{x})'\boldsymbol{r}(\theta_0,\boldsymbol{x})$*.*

3.2.2 When θ is Partially Identified

When Θ^* consists of a set of values of θ, then for each $\theta\in\Theta^*$, all probability measure in the set $\mathcal{Q}(\theta,y_0,\boldsymbol{x})$ produces the same first $2T-1$ generalized moments $\boldsymbol{r}(\theta,\boldsymbol{x})$ as $H(\theta,\boldsymbol{x})\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ by Theorem 3.1. This then leads to the characterization of the sharp bound of $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A,\theta_0,\boldsymbol{x})]$.

Theorem 3.5. *For the model specified in* (2.1)*, if* Θ^* *is a non-singleton set and the product* $\psi(A,\theta_0,\boldsymbol{x})g(A,\theta_0,\boldsymbol{x},y_0)$ *is a polynomial function of* A *with a degree that is no larger than* $2T-1$ *such that:*

$$\psi(A,\theta_0,\boldsymbol{x})g(A,\theta_0,\boldsymbol{x},y_0)=\sum_{j=0}^{2T-1}\eta_j(\theta_0,\boldsymbol{x})A^j,$$

for some vector $\boldsymbol{\eta}(\theta_0,\boldsymbol{x})=(\eta_0(\theta_0,\boldsymbol{x}),\eta_1(\theta_0,\boldsymbol{x}),\ldots,\eta_{2T-1}(\theta_0,\boldsymbol{x}))$*, then the sharp bound for* $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A,\theta_0,\boldsymbol{x})]$ *is* $[\ell b(\boldsymbol{x}),ub(\boldsymbol{x})]$ *where:*

$$\ell b(\boldsymbol{x})=\inf_{\theta\in\Theta^*}\boldsymbol{\eta}(\theta,\boldsymbol{x})'\boldsymbol{r}(\theta,\boldsymbol{x}),$$

$$ub(\boldsymbol{x})=\sup_{\theta\in\Theta^*}\boldsymbol{\eta}(\theta,\boldsymbol{x})'\boldsymbol{r}(\theta,\boldsymbol{x}).$$

Theorem 3.4 and 3.5 provide conditions under which a functional of $Q_0(A|y_0,\boldsymbol{x})$ can be point identified or sharply bounded directly from a linear combination of the vector of generalized moment $\boldsymbol{r}(\theta,\boldsymbol{x})$, itself a linear combination of the observed choice probability $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$, without the need to optimize with respect to the distribution Q. For Theorem 3.5, the optimization is with respect to the *finite* dimensional parameters and the feasible regions are defined by a finite set of moment (in)equality conditions.

We now give three examples for which Theorem 3.4 or 3.5 becomes applicable.

3.2.3 Average Marginal Effects

For model (2.1) without covariates, denote the transition probability conditional on A as:

$$\Pi_k(A,\beta) = \frac{A(\exp(\beta))^k}{1+A(\exp(\beta))^k}, \quad k = \{0,1\}.$$

The average marginal effect (AME) is then defined as $AME_{y_0} = \int \Pi_1(A,\beta_0) - \Pi_0(A,\beta_0) dQ_0(A|y_0)$.

Proposition 3.1. *For model (2.1) with $\gamma = 0$, for $k = \{0,1\}$, $\Pi_k(A,\beta)g(A,\beta,y_0)$ is a polynomial function of A with a degree that is no larger than $2T-1$.*

With Proposition 3.1, we can now apply Theorem 3.4 and 3.5 and show that the aggregated transition probabilities, $\int \Pi_0(A,\beta_0)dQ_0(A|y_0)$ and $\int \Pi_1(A,\beta_0)dQ_0(A|y_0)$, are point identified as soon as $T \geq 3$. When $T = 2$, they can be sharply bounded. Since the average marginal effect is just the difference of them, the same conclusion applies. This result echoes those obtained in Aguirregabiria and Carro (2020), who are the first to point out that the AME is point identified for dynamic panel logit model with one lag. Because Aguirregabiria and Carro (2020) take a sequential identification approach to establish the identification results on AME, they require that the structural parameters are point identified. Our result extends to the cases where structural parameters are partially identified. The generalized moments $\boldsymbol{r}(\theta,\boldsymbol{x})$, discussed in Theorem 3.3, provides sharp identifying information on the latent distribution, which in turn leads to sharp bound on the AME. In Section 4.1, we provide a detailed derivation and numerical illustration for point identification of the AME for the model (2.1) with $T = 3$ and $\gamma = 0$. In Section 4.2, we discuss sharp bounds of the AME for the same model with $T = 2$.

For model (2.1) with covariates, transition probability needs to be defined conditional on certain values of the covariates. For exposition let's consider the case with one covariate. For fixed values $\boldsymbol{x} \in \mathcal{X}$ and $\tilde{x} \in \mathbb{R}$, denote:

$$\Pi_{k,\tilde{x},\boldsymbol{x}}(A,\theta) = \mathbb{P}(Y_{iT+1} = 1|Y_{iT} = k, X_{T+1} = \tilde{x}, X_{1:T} = \boldsymbol{x}, A, \theta), \quad k = \{0,1\}.$$

This quantity is interpreted as the transition probability from period T to period $T+1$ for all individuals with their fixed effects taking value A, and their covariates taking value

$\{\boldsymbol{x}, \tilde{x}\} = \{x_1, \ldots, x_T, \tilde{x}\}$ from period 1 to $T+1$. Note the vector $\boldsymbol{x}$ can be any value in the support $\mathcal{X}$ (i.e., they do not necessarily need to take the same value as required in Aguirregabiria and Carro (2020) for one of their definitions of the AME).

Proposition 3.2. *For the model (2.1) with covariates, consider the transition probability with fixed values $\boldsymbol{x} \in \mathcal{X}$ and $\tilde{x}$. If $\tilde{x} \in \{x_2, \ldots, x_T\}$, then for $k = \{0, 1\}$, $\Pi_{k,\tilde{x},\boldsymbol{x}}(A, \theta) g(A, \theta, \boldsymbol{x}, y_0)$ is a polynomial function of A with a degree that is no larger than $2T - 1$.*

With this result, Theorem 3.4 or 3.5 leads to either point identification or partial identification of the AME, defined as:

$$AME_{\tilde{x},\boldsymbol{x}} = \int_{\mathcal{A}} \{\Pi_{1,\tilde{x},\boldsymbol{x}}(A, \theta_0) - \Pi_{0,\tilde{x},\boldsymbol{x}}(A, \theta_0)\} dQ_0(A|y_0, \boldsymbol{x}).$$

Example 1. *Consider the model (2.1) with $T = 3$ and one covariate, and let $y_0 = 0$. In this model θ is point identified, and hence we know the value of θ_0 (see Section 4.4). Without loss of generality let $\tilde{x} = x_3$, then $g(A, \theta_0, \boldsymbol{x}, y_0) = (1 + AB_0 C_0^{x_2})(1 + AB_0 C_0^{x_3})(1 + AC_0^{x_1})(1 + AC_0^{x_2})(1 + AC_0^{x_3})$. We then have:*

$$\begin{aligned}
&(\Pi_{1,\tilde{x},\boldsymbol{x}}(A, \theta_0) - \Pi_{0,\tilde{x},\boldsymbol{x}}(A, \theta_0)) g(A, \theta_0, \boldsymbol{x}, y_0) \\
&= (B_0 - 1) \begin{pmatrix} 1 & A & A^2 & A^3 & A^4 & A^5 \end{pmatrix} \boldsymbol{\eta}(\theta_0, \boldsymbol{x}, \tilde{x}),
\end{aligned}$$

where $\begin{pmatrix} 0, & C_0^{x_3}, & C_0^{x_3}(C_0^{x_1} + C_0^{x_2} + B_0 C_0^{x_2}), & C_0^{x_2+x_3}(C_0^{x_1} + B_0 C_0^{x_1} + B_0 C_0^{x_2}), & B_0 C_0^{x_1+2x_2+x_3}, & 0 \end{pmatrix}'$ is denoted by $\boldsymbol{\eta}(\theta_0, \boldsymbol{x}, \tilde{x})$. Applying Theorem 3.4, the average marginal effect is point identified as:

$$AME_{\tilde{x},\boldsymbol{x}} = (\exp(\beta_0) - 1) \boldsymbol{\eta}(\theta_0, \boldsymbol{x}, \tilde{x})' \boldsymbol{r}(\theta_0, \boldsymbol{x}),$$

with $\boldsymbol{r}(\theta, \boldsymbol{x}) = H(\theta, \boldsymbol{x}) \mathcal{P}_{\boldsymbol{x}}$ defined as in Theorem 3.1. See more details for this example in Section 4.4.

In Section 4.3, we also consider the model (2.1) with $T = 2$ and covariates, for which we derive the sharp bound of the AME in a specific numerical example.

3.2.4 Posterior Mean of the Fixed Effects

In addition to the average marginal effect, researchers may also be interested in the posterior mean of a function of the fixed effects conditional on the observed choice history. This can be useful to infer the degree of heterogeneity of individuals, conditional on them making certain sequence of choices. For simplicity, we focus on model (2.1) without covariates. For all $\boldsymbol{y}^j \in \mathcal{Y}$, let's define:

$$\varphi(A, \theta, \boldsymbol{y}^j) = A\mathcal{L}_j(A, \theta, y_0).$$

Proposition 3.3. *For model (2.1) with $\gamma = 0$, for all $\boldsymbol{y}^j \in \mathcal{Y}\backslash\{1, 1, \ldots, 1\}$, $\varphi(A, \theta, \boldsymbol{y}^j)g(A, \theta, y_0)$ is a polynomial function of A with a degree that is no larger than $2T - 1$.*

Applying Theorem 3.4 or 3.5, we again can point identify or construct sharp bound for the following quantity:

$$\mathbb{E}[A|\boldsymbol{y}^j] = \frac{\int_{\mathcal{A}} A\mathcal{L}_j(A, \theta_0, y_0)dQ_0(A|y_0)}{\int_{\mathcal{A}} \mathcal{L}_j(A, \theta_0, y_0)dQ_0(A|y_0)} = \frac{\int_{\mathcal{A}} \varphi(A, \theta_0, \boldsymbol{y}^j)dQ_0(A|y_0)}{\mathcal{P}_j},$$

for all sequence of choices $\boldsymbol{y}^j \in \mathcal{Y} \setminus \{1, 1, \ldots, 1\}$. This quantity has the interpretation as the posterior mean of $\exp(\alpha)$ conditional on a particular choice history $\boldsymbol{y}^j$.

Example 2. *For model (2.1) with $\gamma = 0$, $y_0 = 0$ and $T = 3$. Consider the sequence of choices, $\boldsymbol{y} = \{0, 1, 0\}$, and let the observed choice probability be denoted as $\mathbb{P}_0(0, 1, 0)$, then we can show that:*

$$\mathbb{E}[\exp(\alpha)|\boldsymbol{y}] = \frac{1}{\mathbb{P}_0(0, 1, 0)} \begin{pmatrix} 0 & 0 & 1 & B_0 + 1 & B_0 & 0 \end{pmatrix} \boldsymbol{r}(\beta_0).$$

This conditional mean is point identified since $B_0 = \exp(\beta_0)$, and $\boldsymbol{r}(\beta_0) = H(\beta_0)\boldsymbol{\mathcal{P}}$, are both point identified given Theorem 3.1, and $\mathbb{P}_0(0, 1, 0)$, which is directly observed from data. See more details in Section 4.1 and a specific form of $H(\beta_0)$ in Appendix A.4.

3.2.5 Counterfactual Choice Probability with No Dynamics

Here we consider, as a counterfactual parameter of interest, the choice probability where there is no dynamics in the model, keeping everything else unchanged. This parameter contains information on how much of the choice persistence can be explained by the fixed effects, when compared with the observed probability of the choice history under lagged choice dependence (i.e., $\beta \neq 0$).

For model (2.1) without covariates, we denote, for any sequence of choices $\boldsymbol{y}$,

$$\psi(A, \boldsymbol{y}) = \frac{A^{\sum_{t=1}^{T} y_t}}{(1+A)^T}, \tag{3.6}$$

and for the same model one covariate, we denote:

$$\psi(A, \boldsymbol{y}, \boldsymbol{x}) = \frac{A^{\sum_{t=1}^{T} y_t} \prod_{t=1}^{T} C^{x_t y_t}}{\prod_{t=1}^{T}(1 + AC^{x_t})}. \tag{3.7}$$

Both quantities correspond to the counterfactual choice probability of $\boldsymbol{y} \in \mathcal{Y}$, conditional on the fixed effects taking value A. For the case with covariates, we further condition on the value of the covariates being $\boldsymbol{x} \in \mathcal{X}$.

Proposition 3.4. *For the model (2.1) without covariates, if $y_0 = 0$, $\psi(A, \boldsymbol{y})g(A, \beta, y_0)$ is a polynomial function of A with a degree that is no larger than $2T - 1$. For the model with covariates, if $y_0 = 0$, $\psi(A, \boldsymbol{y}, \boldsymbol{x})g(A, \theta, \boldsymbol{x}, y_0)$ is a polynomial function of A with a degree that is no larger than $2T - 1$ for all $\boldsymbol{y} \in \mathcal{Y}$ and $\boldsymbol{x} \in \mathcal{X}$.*

Applying Theorem 3.4 or 3.5, we can point identify or construct sharp bound for the counterfactual choice probabilities, defined respectively as:

$$\text{no covariate:} \quad \mathcal{P}^\star(\boldsymbol{y}) = \int_{\mathcal{A}} \psi(A, \boldsymbol{y}) dQ_0(A|y_0),$$
$$\text{with covariate:} \quad \mathcal{P}^\star_{\boldsymbol{x}}(\boldsymbol{y}) = \int_{\mathcal{A}} \psi(A, \boldsymbol{y}, \boldsymbol{x}) dQ_0(A|y_0, \boldsymbol{x}),$$

for any $\boldsymbol{y} \in \mathcal{Y}$ and $\boldsymbol{x} \in \mathcal{X}$.

Example 3. *For model (2.1) with $T = 3$ and one covariate, we point identify the structural parameters, and hence $B_0 = \exp(\beta_0)$ and $C_0 = \exp(\gamma_0)$ are known. For $y_0 = 0$, let's consider the choice history $\boldsymbol{y} = \{1, 1, 1\}$. We can show:*

$$\psi(A, \boldsymbol{y}, \boldsymbol{x})g(A, \theta_0, \boldsymbol{x}, y_0) = A^3 C_0^{x_1+x_2+x_3} + A^4 B_0 C_0^{x_1+x_2+x_3}[C_0^{x_3} + C_0^{x_2} + B_0 C_0^{x_2+x_3}],$$

and therefore the counterfactual choice probability,

$$\mathcal{P}^\star_{\boldsymbol{x}}(1,1,1) = C_0^{x_1+x_2+x_3} \begin{pmatrix} 0 & 0 & 0 & 1 & B_0 C_0^{x_3} + C_0^{x_2} + B_0 C_0^{x_2+x_3} & 0 \end{pmatrix} \boldsymbol{r}(\theta_0, \boldsymbol{x}),$$

is point identified. Denote the observed conditional choice probability as $\mathbb{P}_{0,\boldsymbol{x}}(1,1,1)$*, we can identify* $\mathbb{P}_{0,\boldsymbol{x}}(1,1,1) - \mathcal{P}^{\star}_{\boldsymbol{x}}(1,1,1)$*, which measures how much the state dependence contributes to the persistent choice.*

3.3 Connection with the CMLE Approach

The conditional maximum likelihood approach uses sufficient statistics to factorize the likelihood into a component that depends on the fixed effects, and a component that does not. For instance, in model (2.1) without covariates, we can write:

$$\mathcal{P}(\boldsymbol{y} \mid y_0, \theta) = \mathcal{P}(\boldsymbol{y} \mid S(\boldsymbol{y}), \theta) \int_{\mathcal{A}} \mathcal{P}(S(\boldsymbol{y}) \mid A, \theta) dQ(A \mid y_0), \tag{3.8}$$

for each $\boldsymbol{y} = \{y_1, \ldots, y_T\} \in \{0,1\}^T$, where $S(\boldsymbol{y})$ is a sufficient statistic for A. In this model, we can use $S(\boldsymbol{y}) = \{y_0, \sum_{t=1}^{T-1} y_t, y_T\}$. After factorizing the likelihood, we can derive moment equality conditions using the first component in (3.8). This procedure leads to point identification in this particular model as soon as $T \geq 3$. The *conditional maximum likelihood estimator* (CMLE) solves the empirical counterpart of the system of equations implied by these moment conditions.

While this procedure leads to a $\sqrt{n}$-consitent estimator, the second component in (3.8) depends on θ, leaving us with an interesting question: Is there useful information in the second component that is being thrown away by the CMLE approach? This question is related to an open puzzle. Specifically, Hahn (2001) shows that the CMLE does not achieve the semiparametric efficiency bound (when $T = 3$), suggesting that there might exist an estimator that is asymptotically more efficient, but no such estimator has been found to date.[14]

In this section, we use our framework to revisit this puzzle. In particular, we apply our methodology to the second component in (3.8) in order to determine whether it contains any useful information about θ.

Consider model (2.1) without covariates given $T \geq 3$ and $y_0 = 0$.[15] In this model, the support of $S(\boldsymbol{y})$ contains $2T$ points since $\sum_{t=1}^{T-1} y_t \in \{0, 1, \ldots, T-1\}$ and $y_T \in \{0,1\}$. Label

[14] Hahn (2001) also makes use of the polynomial structure of the dynamic panel logit model in his analysis of the semiparametric efficiency for the CMLE.

[15] We fix $y_0 = 0$ to be consistent with Hahn (2001).

these points as $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{2T}\}$. The likelihood associated with the sufficient statistic $S(\boldsymbol{y})$ is:

$$\boldsymbol{\mathcal{L}}^S(A, \theta, y_0) = \Big(\mathcal{P}(S(\boldsymbol{y}) = \boldsymbol{s}_1 \mid A, y_0), \quad \cdots, \quad \mathcal{P}(S(\boldsymbol{y}) = \boldsymbol{s}_{2T} \mid A, y_0)\Big)'.$$

As done in (3.1) for our benchmark model, the second component in (3.8) can be reformulated:

$$\boldsymbol{\mathcal{P}}^S = \int_{\mathcal{A}} \boldsymbol{\mathcal{L}}^S(A, \theta, y_0) dQ(A \mid y_0) = \tilde{G}(\theta) \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots A^{2T-1} \end{pmatrix} \frac{1}{\tilde{g}(A, \theta, y_0)} dQ(A \mid y_0),$$

where $\boldsymbol{\mathcal{P}}^S$ denotes the vector of length $2T$ with elements corresponding to the probabilities of $S(\boldsymbol{y})$ taking certain values on its support, and $\tilde{g}(A, \theta, y_0)$ denotes a polynomial function of A of degree $2T-1$. The crucial observation is that the matrix $\tilde{G}(\theta)$ is a square $2T \times 2T$ matrix of full rank. Therefore, the second component of (3.8) does not produce any moment equality conditions. All of the information about θ in this component must be in the form of moment inequality conditions. This result suggests that there does not exist an asymptotically more efficient estimator than the CMLE for this particular model.[16]

The above discussion of model (2.1) without covariates may leave the reader with the impression that the CMLE approach always factors the model in the appropriate way, and that focusing on the component of the likelihood conditional on the sufficient statistic will lead to an asymptotically efficient estimator. However, when covariates X are introduced, the only sufficient statistic is the original vector of choice histories such that $S(\boldsymbol{y}) = \boldsymbol{y}$, unless we impose a support restriction on X, as in Honoré and Kyriazidou (2000).[17] Hence, the sufficient statistic of the fixed effects does not reduce the 2^T vector $\mathcal{P}_{\boldsymbol{x}}$ to the lower dimensional subspace where it lives. Fortunately, we can still apply our methodology in order to achieve the necessary reduction in dimension.

[16] The constructions of $\tilde{G}(\theta)$ and $\tilde{g}(A, \theta, y_0)$ are available upon request. Given our finding that there does not exist additional moment equality conditions from the likelihood associated with the sufficient statistic, we revisit Theorem 1 in Hahn (2001), and can confirm indeed the CMLE is semiparametrically efficient for this model.

[17] For $T = 3$, if we impose $x_2 = x_3$ as in Honoré and Kyriazidou (2000), then we get back the same sufficient statistics as the model without covariates.

4 Examples

In this section, we consider several examples. We present two-period and three-period dynamic panel logit models both with and without covariates, and generalize to any T in the appendix. As special cases, we explicitly discuss the dynamic panel logit model with a time trend, as well as the model with time dummies. At last we discuss extensions to the AR(2) model.

4.1 Three Periods without Covariates

We first consider the three period dynamic panel logit model without covariates. This model reduces to the model treated in Chamberlain (1985), hence we know β is point identified. We use this setting to first illustrate how we construct the moment equalities in Theorem 3.1. We then discuss point identification of the average marginal effects using Theorem 3.4. The likelihood function for a given choice history $\boldsymbol{y}^j \in \mathcal{Y}$ fixing $y_0 = 0$ is:

$$\mathcal{L}_j(A, \beta, y_0) = \prod_{t=1}^{3} \frac{\exp(\alpha + \beta y_{t-1})^{y_t}}{1 + \exp(\alpha + \beta y_{t-1})} = \frac{A^{y_1+y_2+y_3} B^{y_2 y_1 + y_3 y_2}}{(1+A)(1+AB^{y_1})(1+AB^{y_2})},$$

where $A = \exp(\alpha)$ and $B = \exp(\beta)$. After integrating out the fixed effect, we obtain:[18]

$$\boldsymbol{\mathcal{P}} = \int_0^\infty G(\beta) \begin{pmatrix} 1 & A & \cdots & A^5 \end{pmatrix}' \frac{1}{g(A, \beta, y_0)} dQ(A \mid y_0),$$

where the matrix $G(\beta)$ and the form of $g(A, \beta, y_0)$ are given in Appendix A.4.

The left null space of the matrix $G(\beta)$ is spanned by the following two vectors:[19]

$$v_1 = \begin{pmatrix} 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}'$$
$$v_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -B & 1 & 0 \end{pmatrix}'$$

The moment equality condition implied by the second vector in this basis directly provides the point identification of β:

$$\beta_0 = \log\left(\mathbb{P}_0(0,1,1)\right) - \log\left(\mathbb{P}_0(1,0,1)\right).$$

[18] The vector $\boldsymbol{\mathcal{P}}$ has elements $\mathbb{P}((Y_1, Y_2, Y_3) = \boldsymbol{y} | Y_0 = y_0)$ with $\boldsymbol{y} \in \mathcal{Y}$. The elements in the set $\mathcal{Y}$ are ordered as: $\{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\}$.

[19] To construct this basis, we assume that $B \neq 1$.

Now that β is point identified, we apply Theorem 3.4 to show that the average marginal effect is point identified. The AME is defined as:[20]

$$\text{AME}_{y_0} = \int_0^\infty \frac{AB_0}{1+AB_0} dQ_0(A|y_0) - \int_0^\infty \frac{A}{1+A} dQ_0(A|y_0),$$

where $B_0 = \exp(\beta_0)$. When $y_0 = 0$, the AME can therefore be written:

$$\text{AME}_0 = (B_0 - 1) \int_0^\infty A(1+A)^2(1+AB_0) \frac{1}{g(A, \beta_0, y_0)} dQ(A \mid y_0).$$

There is a similar expression if $y_0 = 1$. It is clear that this expression is a linear combination of the generalized moments defined in Theorem 3.3. Since β is point identified, the coefficients of this linear combination are known, and the AME is point identified. With some algebra, we can show that $\text{AME}_0 = (B_0 - 1)(\mathbb{P}_0(0,1,0) + \mathbb{P}_0(1,0,1))$. See details in Appendix A.4.

4.2 Two Periods without Covariates

Next, we consider the two-period model without covariates. This is the example elaborated in Section 2 where the only identification conditions for the structural parameters are given by moment inequalities. Taking the formulation in (2.5), and the matrix $H(\beta)$ to be the inverse of $G(\beta)$, we obtain:

$$H(\beta) = G^{-1}(\beta) = \frac{1}{B-1} \begin{pmatrix} B-1 & -B^2 & B & 0 \\ 0 & B & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & \frac{B-1}{B} \end{pmatrix}.$$

Now, let $\boldsymbol{\mathcal{P}} = (p_0, p_1, p_2, p_3)'$ and we obtain:

$$\boldsymbol{r}(\beta) = H(\beta)\boldsymbol{\mathcal{P}} = \frac{1}{B-1} \begin{pmatrix} (B-1)p_0 - B^2 p_1 + B p_2 \\ B p_1 - p_2 \\ -p_1 + p_2 \\ p_1 - p_2 + \frac{B-1}{B} p_3 \end{pmatrix}, \tag{4.1}$$

[20] We define the AME conditioning on Y_0. If the researcher wants to learn the AME without conditioning on Y_0 and if Y_0 is observed, then we can integrate it out.

for every $B \neq 1$. By Theorem 3.1 and Theorem 2.2, a value of β is in the identified set Θ^* if, and only if, when evaluated at β, the following two (Hankel) matrices are non-negative:

$$H_1(\boldsymbol{r}(\beta)) = \begin{pmatrix} r_0(\beta) & r_1(\beta) \\ r_1(\beta) & r_2(\beta) \end{pmatrix} \text{ and } B_1(\boldsymbol{r}(\beta)) = \begin{pmatrix} r_1(\beta) & r_2(\beta) \\ r_2(\beta) & r_3(\beta) \end{pmatrix}$$

together with the range condition which states $\{r_2(\beta), r_3(\beta)\}$ is in $\text{Range}(H_1(\boldsymbol{r}(\beta)))$.

The range condition deserves some extra explanation. We distinguish two cases. The first case is when $H_1(\boldsymbol{r}(\beta))$ is singular (i.e., $\det(H_1(\boldsymbol{r}(\beta))) = 0$). In this case, there exists a constant $c_0 > 0$ such that $r_j(\beta) = c_0 r_{j-1}(\beta)$ for $j = 1, 2$. For $\{r_2(\beta), r_3(\beta)\}$ to be in $\text{Range}(H_1(\boldsymbol{r}(\beta)))$, we would have $r_3(\beta) = c_0 r_2(\beta)$. Since we can rule out $\boldsymbol{r}(\beta) = 0$ for any $\beta \neq 0$, then the non-negativity of $H_1(\boldsymbol{r}(\beta))$ and $B_1(\boldsymbol{r}(\beta))$ implies that $r_0(\beta) > 0$. To summarize, the constraints on $\boldsymbol{r}(\beta)$ in this case include: $r_0(\beta) > 0$, $c_0 > 0$ and $r_j(\beta) = c_0 r_{j-1}(\beta)$ for $j = 1, 2, 3$. These conditions impose very strong restrictions on $\boldsymbol{r}(\beta)$. Yet if $\boldsymbol{r}(\beta)$ indeed is a moment sequence with $\det(H_1(\boldsymbol{r}(\beta))) = 0$, then we can only have a degenerate distribution since $\det(H_1(\boldsymbol{r}(\beta))) = r_0 r_2 - r_1^2$ has the interpretation as the variance of a probability measure.[21] If the variance is zero, then the measure must be degenerate.

The second case to consider is when $H_1(\boldsymbol{r}(\beta))$ is non-singular. In this case the range condition always holds, because we can solve the system of equations $H_1(\boldsymbol{r}(\beta))\boldsymbol{w} = (r_2, r_3)'$ and get a unique $\boldsymbol{w}$ which represents $(r_2, r_3)'$ as a linear combination of columns of $H_1(\boldsymbol{r}(\beta))$. It is also easy to show that in this case $r_j > 0$ for all $j = 0, 1, 2, 3$. To summarize, the conditions for $\boldsymbol{r}(\beta)$ to be a moment sequence include: $r_j > 0$, $r_0 r_2 - r_1^2 > 0$, $r_1 r_3 - r_2^2 \geq 0$. These constraints on $\boldsymbol{r}(\beta)$ map into constraints on the original vector $\boldsymbol{\mathcal{P}}$ which then lead to the following theorem.

Proposition 4.1. *For dynamic logit model defined in* (2.1) *with $T = 2$ and $\gamma = 0$, (i) the sign of β_0 is identified, and (ii) the following two cases define the sharp identified set Θ^*:*

[21] Recall $r_j(\beta) = \int A^j d\bar{Q}(A, \beta)$ with $\bar{Q}(A, \beta)$ being some finite positive Borel measure and it can be made into a probability measure, denoted as $\tilde{Q}(A, \beta)$ by dividing $\bar{Q}(A, \beta)$ by r_0. Then consider $r_0 r_2 - r_1^2 = r_0^2(\frac{r_2}{r_0} - (\frac{r_1}{r_0})^2) = r_0^2(\int A^2 d\tilde{Q}(A, \beta) - (\int A d\tilde{Q}(A, \beta))^2)$. Since $r_0 > 0$ and the only possibility for $r_0 r_2 - r_1^2 = 0$ is $\tilde{Q}(A, \beta)$ having zero variance, which means $\tilde{Q}(A, \beta)$ is a degenerate probability measure.

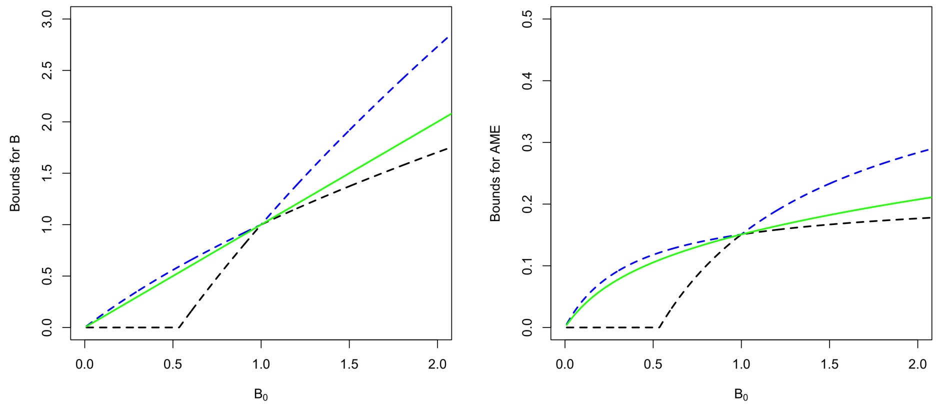

Figure 1: We illustrate the bounds for both the structural parameter as well as the average marginal effect imposed by the moment inequalities in the short panel without covariates as we vary B_0. For each values of β_0 ranging from log(0.01) to log(2), the data generating process assumes Q to be discrete with equal mass at -2 and 1 and $y_0 = 0$. Green solid line illustrates the true value B_0 and AME; blue dotted line illustrates the upper bound; black dotted line illustrates the lower bound.

1. *If $\beta_0 > 0$, then $\beta \in \Theta^*$ if, and only if:*

$$\frac{q_0 + \sqrt{q_0^2 - 4p_1p_2p_3(p_1 - p_2 + p_3)}}{2p_1(p_1 - p_2 + p_3)} \leq B \leq \frac{q_1 + \sqrt{q_1^2 + 4p_1p_2(p_0p_1 - p_0p_2 - p_2^2)}}{2p_1p_2}; \quad (4.2)$$

2. *If $\beta_0 < 0$, then $\beta \in \Theta^*$ if, and only if:*

$$\max\left\{0, \frac{q_1 - \sqrt{q_1^2 + 4p_1p_2(p_0p_1 - p_0p_2 - p_2^2)}}{2p_1p_2}\right\} \leq B \leq \frac{q_0 - \sqrt{q_0^2 - 4p_1p_2p_3(p_1 - p_2 + p_3)}}{2p_1(p_1 - p_2 + p_3)}; \quad (4.3)$$

where $B = \exp(\beta)$, $q_0 = p_1^2 - p_1p_2 + p_1p_3 + p_2p_3$ and $q_1 = p_0p_2 - p_0p_1 + p_1p_2 + p_2^2$.

The proof is given in the Appendix A.1. These bounds are often quite narrow, suggesting that moment inequality conditions are informative about β. We illustrate these bounds in the left panel of Figure 1.

The average marginal effect for this model can be defined as:

$$AME_{y_0} = \int \frac{AB_0}{1 + AB_0} - \frac{A}{1 + A} dQ_0(A|y_0).$$

It is easy to verify that when $y_0 = 0$, $g(A, \beta, y_0)\left\{\frac{AB}{1+AB} - \frac{A}{1+A}\right\} = \boldsymbol{\eta}(\beta)' \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}'$ with $\boldsymbol{\eta}(\beta)' = \begin{pmatrix} 0 & (B-1) & (B-1) & 0 \end{pmatrix}$, and hence the sharp bound for AME_0 is characterized, by applying Theorem 3.4, as $\left[\inf_{\beta\in\Theta^*}\boldsymbol{\eta}(\beta)'\boldsymbol{r}(\beta), \sup_{\beta\in\Theta^*}\boldsymbol{\eta}(\beta)'\boldsymbol{r}(\beta)\right]$. Given the form of $\boldsymbol{r}(\beta)$ in (4.1), $\boldsymbol{\eta}(\beta)'\boldsymbol{r}(\beta) = (B-1)p_1$. This then leads to the sharp bound for AME_0 to take an extremely simple form as $\left[\underline{\mathrm{B}}-1, \bar{B}-1\right]p_1$, where $\underline{\mathrm{B}}$ and $\bar{B}$ being the lower and upper bound of the sharp bound for $\exp(\beta)$. These bounds are illustrated in the right panel of Figure 1.

4.3 Two Periods with a Covariate

Now consider the generalization of the two periods model by adding one covariate (i.e., $\gamma \neq 0$) and denote $C = \exp(\gamma)$. Maintain the assumption that $y_0 = 0$ and thus choosing $g(A, \theta, \boldsymbol{x}, y_0) = (1 + AC^{x_1})(1 + AC^{x_2})(1 + ABC^{x_2})$, we obtain that:[22]

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_0^\infty G(\theta, \boldsymbol{x}) \begin{pmatrix} 1 & A & \cdots & A^3 \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A \mid y_0, \boldsymbol{x}),$$

with the matrix $G(\theta, \boldsymbol{x})$ defined by:

$$G(\theta, \boldsymbol{x}) = \begin{pmatrix} 1 & BC^{x_2} & 0 & 0 \\ 0 & C^{x_1} & C^{x_1+x_2} & 0 \\ 0 & C^{x_2} & BC^{2x_2} & 0 \\ 0 & 0 & BC^{x_1+x_2} & BC^{x_1+2x_2} \end{pmatrix}.$$

When $\beta \neq 0$, the matrix $G(\theta, \boldsymbol{x})$ is of full rank for any value of $\beta, \gamma \neq 0$ and for any $\{x_1, x_2\} \in \mathbb{R}^2$. Therefore, the left null space of the matrix $G(\theta, \boldsymbol{x})$ is of zero dimension and there exists no moment equality conditions. We then use Theorem 3.1 to construct sharp identified set for $\{\beta, \gamma\}$ in this setting. In particular, consider $H(\theta, \boldsymbol{x}) = G^{-1}(\theta, \boldsymbol{x})$, which leads to the transformed probability vector being:

$$\boldsymbol{r}(\theta, \boldsymbol{x}) = G^{-1}(\theta, \boldsymbol{x})\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \frac{1}{B-1} \begin{pmatrix} (B-1)p_0 - B^2C^{x_2-x_1}p_1 + Bp_2 \\ BC^{-x_1}p_1 - C^{-x_2}p_2 \\ -C^{-(x_1+x_2)}p_1 + C^{-2x_2}p_2 \\ C^{-(x_1+2x_2)}p_1 - C^{-3x_2}p_2 + \frac{B-1}{B}C^{-(x_1+2x_2)}p_3 \end{pmatrix}, \quad (4.4)$$

[22] The elements in the set $\mathcal{Y}$ are ordered as: $\{(0,0), (1,0), (0,1), (1,1)\}$ and we denote $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = (p_0(\boldsymbol{x}), p_1(\boldsymbol{x}), p_2(\boldsymbol{x}), p_3(\boldsymbol{x}))'$.

for every $B, C \neq 1$. Applying Theorem 2.2 and Theorem 3.1, we get all the moment inequalities for each values of $\boldsymbol{x}$. The identified set for θ is the set that takes the intersection of these inequalities for all $\boldsymbol{x} \in \mathcal{X}$.

In Figure 2, we illustrate the bounds implied by this model given a specific choice for $\{\beta_0, \gamma_0\}$ and Q_0 under the assumption that $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2\} = \{(1,0),(0,0)\}$. In this figure, we see that the sharp identified set Θ^* is rather small, even with only two possible values in the support $\mathcal{X}$.

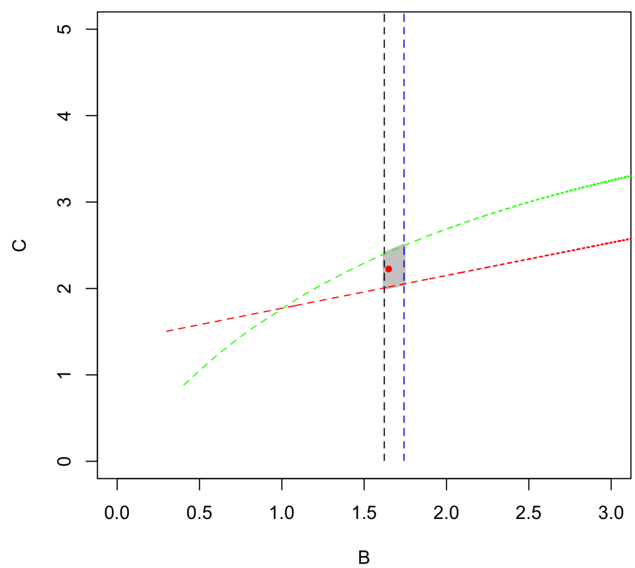

Figure 2: We illustrate the *binding* constraints imposed by the moment inequalities given $\mathcal{X} = \{(1,0),(0,0)\}$. For this figure, we assume that $Q_0(A|\boldsymbol{x}, y_0)$ is discrete with equal mass at -2 and 1 if $\boldsymbol{x} = (1,0)$ and is discrete with equal mass at -1 and -2 if $\boldsymbol{x} = (0,0)$, and that $(\beta_0, \gamma_0) = (0.50, 0.80)$. The shaded region is the sharp identified set; the red point illustrates the true parameters; blue dotted line illustrates the upper bound imposed by $r_0(\beta)r_2(\beta) - r_1(\beta)^2 \geq 0$ and the black dotted line illustrates the lower bound imposed by $r_1(\beta)r_3(\beta) - r_2(\beta)^2 \geq 0$ given $\boldsymbol{x} = (0,0)$; red dotted line corresponds to the lower bound imposed by $r_0(\theta)r_2(\theta) - r_1(\theta)^2 \geq 0$ and the green dotted line depicts the upper bound imposed by $r_1(\theta)r_3(\theta) - r_2(\theta)^2 \geq 0$ given $\boldsymbol{x} = (1,0)$. Constraints like $r_j \geq 0$ for all j are not binding and are not plotted for better visualization.

.

Now we consider the bound for the average marginal effect in this model. Fix $\tilde{x} = 0$, $y_0 = 0$, and consider the same data generating process used to generate bounds in Figure 2 for the identified set of θ, then the average marginal effect can be specified as, for $j = \{1, 2\}$,

$$AME_{\tilde{x},\boldsymbol{x}_j,y_0} = \int \Big\{ \frac{AB_0}{1+AB_0} - \frac{A}{1+A} \Big\} dQ_0(A|y_0, \boldsymbol{x}_j).$$

It is easy to verify that,

$$\Big\{ \frac{AB}{1+AB} - \frac{A}{1+A} \Big\} g(A, \theta, \boldsymbol{x}_j, y_0) = \boldsymbol{\eta}(\theta, \boldsymbol{x}_j)' \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}',$$

with $\boldsymbol{\eta}(\theta, \boldsymbol{x}_1)' = \begin{pmatrix} 0 & (B-1) & (B-1)C & 0 \end{pmatrix}$ and $\boldsymbol{\eta}(\theta, \boldsymbol{x}_2)' = \begin{pmatrix} 0 & (B-1) & (B-1) & 0 \end{pmatrix}$. Then we have, using $\boldsymbol{r}(\theta, \boldsymbol{x})$ derived in (4.4),

$$\begin{aligned} \boldsymbol{\eta}(\theta, \boldsymbol{x}_1)' \boldsymbol{r}(\theta, \boldsymbol{x}_1) &= (\frac{B}{C} - 1) p_1(\boldsymbol{x}_1) + (C-1) p_2(\boldsymbol{x}_1), \\ \boldsymbol{\eta}(\theta, \boldsymbol{x}_2)' \boldsymbol{r}(\theta, \boldsymbol{x}_2) &= (B-1) p_1(\boldsymbol{x}_2). \end{aligned}$$

Now we can apply Theorem 3.5 to bound the AME. In the particular data generating process that generates Figure 2, the true value for $AME_{\tilde{x},\boldsymbol{x}_1,y_0}$ is 0.0749, and we obtain a bound with $[0.0655, 0.0934]$. The true value of $AME_{\tilde{x},\boldsymbol{x}_2,y_0}$ is 0.0859, and we obtain a bound as $[0.0828, 0.0979]$. Both sharp bounds are very informative and very simple to construct. The sharp bound for AME can be directly mapped from the identified set Θ^*.

4.4 Three Periods with a Covariate

Let us now extend the example above to a model with three time periods. In doing so, we will be able to compare our results with some existing results in the literature. We again have the representation of the vector of population choice probability[23] for $\boldsymbol{x} \in \mathcal{X}$ as:

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_0^\infty G(\theta, \boldsymbol{x}) \begin{pmatrix} 1 & A & \cdots & A^5 \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A \mid \theta, y_0, \boldsymbol{x}),$$

with $g(A, \theta, \boldsymbol{x}, y_0) = (1 + AC^{x_1})(1 + AC^{x_2})(1 + AC^{x_3})(1 + ABC^{x_2})(1 + ABC^{x_3})$ when $y_0 = 0$. For the complete form of $G(\theta, \boldsymbol{x})$ see Appendix A.5.

[23] The vector $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ has elements $\mathbb{P}((Y_1, \ldots, Y_T) = \boldsymbol{y} | Y_0 = y_0, X = \boldsymbol{x})$ with $\boldsymbol{y} \in \mathcal{Y}$. The elements in the set $\mathcal{Y}$ are ordered as: $\{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\}$.

The left null space of the matrix $G(\theta, \boldsymbol{x})$ is spanned by the following two vectors:[24]

$$v_1 = \begin{pmatrix} 0 \\ C^{x_{i3}}(B-1) \\ C^{x_{i1}}(1 - BC^{x_{i3}-x_{i2}}) \\ C^{x_{i1}}(1 - C^{x_{i2}-x_{i3}}) \\ BC^{x_{i3}}(1 - C^{x_{i3}-x_{i2}}) \\ B(C^{x_{i3}} - C^{x_{i2}}) \\ 0 \\ 0 \end{pmatrix} \text{ and } v_2 = \begin{pmatrix} 0 \\ C^{x_{i3}-x_{i1}}(BC^{x_{i2}} - C^{x_{i3}}) \\ C^{x_{i3}}(1-B) \\ C^{x_{i3}} - C^{x_{i2}} \\ -BC^{x_{i3}-x_{i1}}(C^{x_{i3}} - C^{x_{i2}}) \\ 0 \\ C^{x_{i3}} - C^{x_{i2}} \\ 0 \end{pmatrix}. \tag{4.5}$$

In Appendix A.5, we show linear combination of v_1 and v_2 leads to moment conditions given by Honoré and Weidner (2020) for the same model. Note that if we were to impose the restriction $x_2 = x_3$, this basis would reduce to the moment conditions derived in Honoré and Kyriazidou (2000). In particular, under the condition $x_2 = x_3$, the rank of $G(\theta, \boldsymbol{x})$ is still six and the basis for the left null space of $G(\theta, \boldsymbol{x})$ leads to the following two vectors:

$$v_1 = \begin{pmatrix} 0 & -C^{x_{i2}-x_{i1}} & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}'$$
$$v_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -BC^{x_{i2}-x_{i1}} & 1 & 0 \end{pmatrix}'$$

Whenever $x_1 \neq x_2$, the moment equalities have a unique solution:

$$\beta = \log(p_1) - \log(p_2) - \log(p_5) + \log(p_6) \text{ and } \gamma = \frac{\log(p_1) - \log(p_2)}{x_1 - x_2},$$

where we label $\mathcal{P}_{\boldsymbol{x}} = (p_0, \dots, p_7)'$ and $p_j = p_j(\boldsymbol{x}) = p_j(x_1, x_2, x_2)$ under $x_3 = x_2$. However, as pointed out in Honoré and Weidner (2020), the assumption $x_2 = x_3$ is not needed to derive useful moment equality conditions for the structural parameters.

4.5 Three Periods with a Time Trend

Now we consider the special case in which the only covariate in the three-period model is a time trend variable, hence the support of X, given by $\mathcal{X}$ contains a single element $\{1, 2, 3\}$.

[24] To construct this basis, we assume that $C \neq 1$ and $x_2 \neq x_3$.

Under this specification, the likelihood function equals:

$$\mathcal{L}_j(A,\theta,y_0)=\prod_{t=1}^{3}\frac{\exp(\alpha+\beta y_{t-1}+\gamma t)^{y_t}}{1+\exp(\alpha+\beta y_{t-1}+\gamma t)}=\frac{A^{\sum_{t=1}^3 y_t}B^{\sum_{t=1}^3 y_t y_{t-1}}C^{\sum_{t=1}^3 t y_{it}}}{\prod_{t=1}^3(1+AB^{y_{t-1}}C^t)},$$

where $A=\exp(\alpha)$, $B=\exp(\beta)$, and $C=\exp(\gamma)$. Let $\theta=\{\beta,\gamma\}$ and the matrix $G(\theta)$ is given by:

$$\begin{pmatrix}
1 & BC^2(1+C) & B^2C^5 & 0 & 0 & 0\\
0 & C & C^3(1+BC) & BC^6 & 0 & 0\\
0 & C^2 & C^4(B+C) & BC^7 & 0 & 0\\
0 & C^3 & BC^5(1+C) & B^2C^8 & 0 & 0\\
0 & 0 & BC^3 & BC^5(1+C) & BC^8 & 0\\
0 & 0 & C^4 & C^6(1+BC) & BC^9 & 0\\
0 & 0 & BC^5 & BC^7(B+C) & B^2C^{10} & 0\\
0 & 0 & 0 & B^2C^6 & B^2C^8(1+C) & B^2C^{11}
\end{pmatrix},$$

with $g(A,\boldsymbol{x},\theta,y_0)=(1+AC)(1+AC^2)(1+AC^3)(1+ABC^2)(1+ABC^3)$ when $y_0=0$.

The left null space of the matrix $G(\theta)$ is spanned by the following two vectors:[25]

$$v_1=\begin{pmatrix}0 & C^3(B-1) & C(1-BC) & (C-1) & -BC^3(C-1) & BC^2(C-1) & 0 & 0\end{pmatrix}'$$
$$v_2=\begin{pmatrix}0 & C^2(B-C) & C(1-B) & (C-1) & -BC^2(C-1) & 0 & (C-1) & 0\end{pmatrix}'$$

It is easy to show that this basis cannot point identify (β,γ). The moment equalities implied by this basis have two non-trivial roots: one at the true parameters (β_0,γ_0), and one at a *false root* (see Figure 3 for illustration). This model is also analyzed by Honoré and Weidner (2020) and they lead to the same conclusion about false roots and hence the structural parameters remain under-identified for a fixed value of y_0.[26] Here we show that we can easily rule out the false root using moment inequalities and hence render point identification. This is the main point we address in this example for which the literature using only moment

[25] To construct this basis, we assume that $C\neq 1$.

[26] Honoré and Weidner (2020) propose to use the variation of y_0 to resolve this issue. It is true that we obtains two more moment conditions when $y_0=1$. Combining the four moment conditions allows one to point identify all structural parameters. However, our point here is that even if we fix $y_0=0$, we can already point identify the structural parameters by using the information from the moment inequalities. Later, we will also show an example (time dummy) where even by finding all moment equality conditions with variation in y_0, we still obtain multiple solutions to the system of moment equalities. But the information in the moment inequalities allows us to rule out false roots, and hence yields point identification.

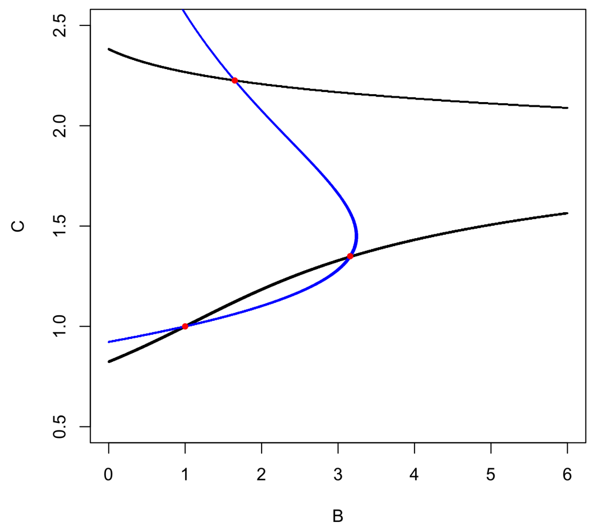

Figure 3: Black illustrates the curve on which the first moment equality holds; blue illustrates the curve on which the second moment equality holds. For this figure, we assume that Q is discrete with equal mass at -2 and 1, and that $(\beta_0, \gamma_0) = (0.50, 0.80)$. There are three solutions: the trivial root $B = C = 1$, the correct root, and the false root. Notice that the trivial root is assumed away in the construction of the moment equalities, leaving us with two roots.

equality conditions has not obtained point identification. To illustrate, we consider a specific example. In particular, assume that Q_0 is discrete with equal mass at -2 and 1, and that $(\beta_0, \gamma_0) = (0.50, 0.80)$. The same DGP is used to generate Figure 3. This specification yields the following population choice probabilities:

$$\boldsymbol{\mathcal{P}} \simeq \Big(0.0924, 0.0226, 0.0458, 0.1424, 0.0257, 0.0508, 0.1743, 0.4456\Big)'.$$

In this example, the moment equalities produce two non-trivial roots: one at the correct location $\theta_0 = (0.50, 0.80)$, and another roughly located at $\tilde{\theta} = (1.15, 0.30)$. We can rule out the false root by checking the non-negativity of the (Hankel) matrices in Theorem 2.2. In fact, in this particular example, it is sufficient for us to check only the sign of the second element of the transformed probability $\boldsymbol{r}(\theta)$. Indeed, by Theorem 2.2, it must be non-negative. Intuitively, it must be non-negative because it is the mean of a finite positive Borel measure. When we check these values, we find:

$$r_1(\theta_0) \simeq 0.01 \;\text{ and }\; r_1(\tilde{\theta}) \simeq -0.24,$$

where $r_1(\theta)$ denotes the second element of $\boldsymbol{r}(\theta)$. Therefore, ruling out the false root here is as simple as checking the sign of one transformed probability.

4.6 Three Periods with Time Dummies

We now consider a more complex example: the three-period dynamic panel logit model with time dummies. This model is characterized by $\gamma x_t = \gamma_t$, for $t = 1, 2, 3$. For simplicity, define $\gamma = \gamma_2$ and $\delta = \gamma_3$ (and $\gamma_1 = 0$ for normalization) and denote $\theta = \{\beta, \gamma, \delta\}$. Under this specification, the likelihood becomes:

$$\mathcal{L}_j(A, \theta, y_0) = \prod_{t=1}^{3} \frac{\exp(\alpha + \beta y_{t-1} + \gamma_t)^{y_t}}{1 + \exp(\alpha + \beta y_{t-1} + \gamma_t)} = \frac{A^{\sum_{t=1}^{3} y_t} B^{\sum_{t=1}^{3} y_t y_{t-1}} C^{y_2} D^{y_3}}{\prod_{t=1}^{3}(1 + AB^{y_{t-1}} C^{\mathbb{1}\{t=2\}} D^{\mathbb{1}\{t=3\}})},$$

where $A = \exp(\alpha)$, $B = \exp(\beta)$, $C = \exp(\gamma)$, and $D = \exp(\delta)$. The information content of this model was studied by Hahn (2001) and an earlier working paper version of Buchinsky, Hahn, and Kim (2010). They, however, considered only moment equality conditions from which they found zero information on β in the sense of Chamberlain (1992). Now using our approach we will show that when we combine both moment equality and inequality conditions we have a sharp bound on β, and our numerical illustration shows it can be even a singleton (i.e., point identification).[27]

For this model we now have three parameters to consider. After integrating out the fixed effect, we obtain:[28]

$$\boldsymbol{\mathcal{P}} = \int_0^\infty G(\theta) \begin{pmatrix} 1 & A & \cdots & A^5 \end{pmatrix}' \frac{1}{g(A, \theta, y_0)} dQ(A \mid y_0),$$

where the form of $G(\theta)$ and $g(A, \theta, y_0)$ is included in Appendix A.7.

Again as before, we focus on $y_0 = 0$. The left null space of the matrix $G(\theta)$ is spanned by the following two vectors:

$$v_1 \;=\; \begin{pmatrix} 0, & -BCD + BD^2, & -BCD, & -BCD, & C, & BD, & 0, & 0 \end{pmatrix}' \tag{4.6}$$

[27] We reproduce some of their results in Appendix A.7, which have never been published, so to compare with our new results.

[28] The vector $\boldsymbol{\mathcal{P}}$ has elements $\mathbb{P}((Y_1, \ldots, Y_T) = \boldsymbol{y} | Y_0 = y_0)$ with $\boldsymbol{y} \in \mathcal{Y}$. The elements in the set $\mathcal{Y}$ are ordered as: $\{(1,1,1), (1,1,0), (1,0,1), (1,0,0), (0,1,1), (0,1,0), (0,0,1), (0,0,0)\}$. We further label the entries of the vector $\boldsymbol{\mathcal{P}}$ to be $\{p_1, p_2, \ldots, p_8\}$.

$$v_2 = \Big(0,\ \ CD-D^2,\ \ CD-BCD,\ \ CD-D^2,\ \ D-C/B,\ \ 0,\ \ -C+D,\ \ 0\Big)'.$$

Let us now characterize the set of all solutions to the moment equality conditions defined through (4.6), and then discuss whether there is any additional information contained in the moment inequalities. To start, we consider a linear combination of v_1 and v_2:

$$\begin{aligned} v_1 &= \Big(0,\ \ BD(D-C),\ \ -BCD,\ \ -BCD,\ \ C,\ \ BD,\ \ 0,\ \ 0\Big)' \\ \frac{(v_1+Bv_2)}{B} &= \Big(0,\ \ 0,\ \ -BCD,\ \ -D^2,\ \ D,\ \ D,\ \ -C+D,\ \ 0\Big)' \end{aligned}$$

The moment equality implied by the second vector above yields:

$$B = \frac{-D^2p_4 + D(p_5+p_6) + (-C+D)p_7}{CDp_3}. \tag{4.7}$$

Therefore, there exists a deterministic relationship between B and (C,D) given $\boldsymbol{\mathcal{P}}$. Consequently, the identification problem can be effectively reduced from a three parameter problem to a two parameter problem. The moment equality implied by the first vector remains. This moment equality, $v_1'\boldsymbol{\mathcal{P}} = 0$, can be written as:

$$\big\{(-CD+D^2)p_2 - CD(p_3+p_4) + Dp_6\big\}\big\{-D^2p_4 + D(p_5+p_6) + (-C+D)p_7\big\} + C^2Dp_3p_5 = 0. \tag{4.8}$$

This result implies that we can solve for the sharp identified set by finding all of the solutions (C^*, D^*) to (4.8), which is a polynomial functions of C and D. We then use (4.7) to deduce B^* given each solution (C^*, D^*), and then we remove false solutions by checking the moment inequalities induced by $\boldsymbol{r}(\theta) \in \mathcal{M}_5$, as stated in Theorem 3.1 and Theorem 2.2 at every solution. Details on how to construct $\boldsymbol{r}(\theta)$ is included in the Appendix A.7.

We illustrate the power of moment inequalities in Figure 4. In the left panel, we see a curve in the positive orthant $\mathbb{R}^2_+$. This curve contains the solutions (C^*, D^*) to the moment equality in (4.8) under a specific choice of Q_0 and $(\beta_0, \gamma_0, \delta_0)$. We see that there are an uncountable number of non-trivial solutions from just the moment equality conditions. Interestingly and importantly, we find numerically that every false root is ruled out using our moment inequalities induced by $\boldsymbol{r}(\theta) \in \mathcal{M}_5$, which illustrates the informativeness of these moment inequalities for this model.

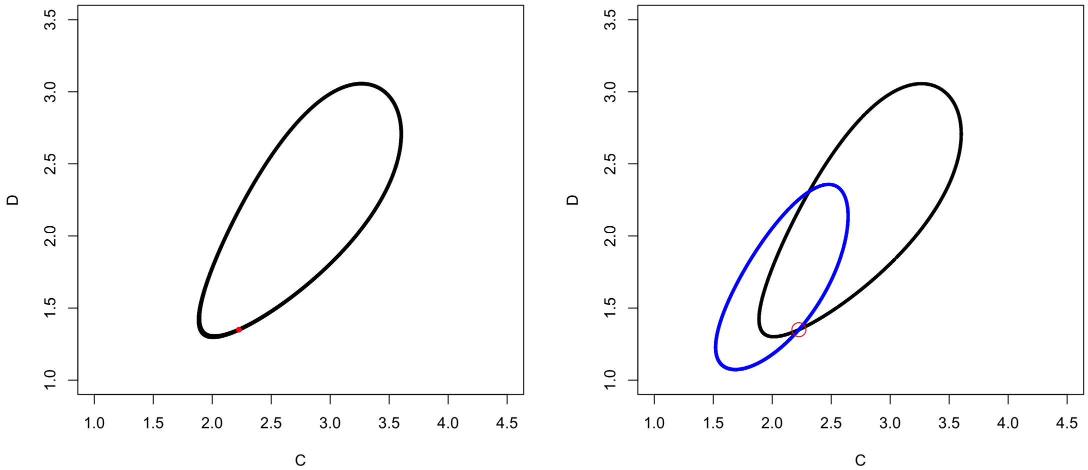

Figure 4: The left figure: Black illustrates the curve on which the moment equality in (4.8) holds, derived when $y_0 = 0$; the red point denotes the true solution. The right figure: the added blue curve illustrates the set of values of (C, D) on which the moment equality in (A.18) holds, derived when $y_0 = 1$ in Appendix A.7. The red circled point is again the true value. For both figures, we assume that Q_0 is discrete with equal mass at -2 and 1, and that $(\beta_0, \gamma_0, \delta_0) = (0.50, 0.80, 0.30)$.

Moreover, we now investigate whether using variation in Y_0 will provide point identification for θ. As shown in Appendix A.7, for $y_0 = 1$, we find another moment equality condition that involves C and D, and again B can be written as a deterministic function of (C, D). So we have two moment conditions for two unknown parameters (C, D) if Y_0 indeed has variation.

However, since these moment equality conditions are polynomial functions of two variables, C and D, it is in general not clear how many real valued roots can be solved from these two moment equalities, neither a conclusive judgement on whether we can find a unique real valued solution. Indeed, using the same Q_0 in the previous example, we find that the two moment equality conditions on (C, D) lead to *two* roots, one of them a false root. This is illustrated in the right panel of Figure 4. In this case, we just need to use our moment inequality conditions to check for these two candidate points, and we can rule out the false root. In particular, the false root takes value $(B, C, D) \approx (1.646, 2.312, 2.308)$. Applying Theorem 2.2, we find that the second element of $\boldsymbol{r}(\theta)$ takes values -0.179 for $y_0 = 0$ and -0.146 for $y_0 = 1$ at this parameter value, which concludes that this value of θ can not generate a vector $\boldsymbol{r}(\theta)$ that belongs to the moment space. Therefore, it can not be in the identified set. In this example, combining the moment equality and inequality conditions allows us to achieve point identification of θ, whether the variation of Y_0 is available or not.

4.7 AR(2) Model with Three Periods without Covariates

Now we extend the model (2.1) to the dynamic panel logit model with two lags, specified as:

$$Y_{it} = \mathbb{1}\{\alpha_i + \beta_1 Y_{it-1} + \beta_2 Y_{it-2} \geq \epsilon_{it}\}.$$

Assume we observe $(Y_{-1}, Y_0, Y_1, Y_2, Y_3)$. For exposition, we fix $(y_{-1}, y_0) = (0,0)$. Similar analysis can be done with any value of $(y_{-1}, y_0) \in \{0,1\}^2$. Variation of (Y_{-1}, Y_0) allows extra identifying constraint on the structural parameters since each value of $\{y_{-1}, y_0\}$ yields a set of moment conditions. Here by fixing y_{-1} and y_0 at certain values, we consider the situation that the researchers do not have access for such variation (i.e., (Y_{-1}, Y_0) only takes certain fixed value in the population).

Denoting again $A = \exp(\alpha)$, $B_1 = \exp(\beta_1)$ and $B_2 = \exp(\beta_2)$, we can represent the likelihood of choice history (for general T) as:

$$\mathcal{L}_j(A, \theta, y_{-1}, y_0) = A^{\sum_{t=1}^T y_t} B_1^{\sum_{t=1}^T y_t y_{t-1}} B_2^{\sum_{t=1}^T y_t y_{t-2}} \sigma_A(0,0)^{m_1} \sigma_A(0,1)^{m_2} \sigma_A(1,0)^{m_3} \sigma_A(1,1)^{m_4}$$

with $\sigma_A(0,0) = \frac{1}{1+A}$, $\sigma_A(0,1) = \frac{1}{1+AB_1}$, $\sigma_A(1,0) = \frac{1}{1+AB_2}$ and $\sigma_A(1,1) = \frac{1}{1+AB_1B_2}$ and $m_1 = \sum_{t=1}^T (1-y_{t-2})(1-y_{t-1})$, $m_2 = \sum_{t=1}^T (1-y_{t-2}) y_{t-1}$, $m_3 = \sum_{t=1}^T y_{t-2}(1-y_{t-1})$ and $m_4 = \sum_{t=1}^T y_{t-1} y_{t-2}$. With $T = 3$ and $\{y_{-1}, y_0\} = \{0, 0\}$, we have $m_1 \in \{1,2,3\}$ and $m_j \in \{0,1\}$, for $j = 2,3,4$. By picking $g(A, \beta_1, \beta_2, y_{-1}, y_0) = (1+A)^3(1+AB_1)(1+AB_2)(1+AB_1B_2)$, $\mathcal{L}_j(A, \theta, y_{-1}, y_0) g(A, \beta_1, \beta_2, y_{-1}, y_0)$ becomes a polynomial function of A with a degree no larger than 6 for all possible choice history in the set $\mathcal{Y}$. Therefore, we again have the formulation that:

$$\boldsymbol{\mathcal{P}} = G(\theta) \int_0^{+\infty} (1, A, \ldots, A^6)' \frac{1}{g(A, \theta, y_{-1}, y_0)} dQ(A|y_{-1}, y_0),$$

with $G(\theta)$ being a 8×7 matrix of full column rank provided $\beta_1, \beta_2 \neq 0$. The particular form of $G(\theta)$ is available upon request.[29] The left null space of $G(\theta)$ has dimension equals to one,

[29] We use a symbolic toolbox in Matlab to derive the form of G and H for all our examples. These code will be available in one of the authors' website.

hence we expect one moment equality condition. Indeed, it takes the form:

$$-B_1\mathbb{P}_0(1,0,0) + B_1\mathbb{P}_0(0,1,0) - B_1\mathbb{P}_0(1,0,1) + \mathbb{P}_0(0,1,1) = 0,$$

where $\mathbb{P}_0(y_1, y_2, y_3)$ denotes the choice probability of the choice history $(0, 0, y_1, y_2, y_3)$. Clearly β_1 is point identified from this moment equality. We now demonstrate the information on β_2 through moment inequalities. Construct $H(\theta)$ such that $H(\theta)G(\theta)$ is an identify matrix of dimension 7. Now let $\boldsymbol{r}(\theta) = H(\theta)\boldsymbol{\mathcal{P}}$. The identified set of θ can be constructed as:

$$\Theta^* = \{\theta = \{\beta_1, \beta_2\} : \exp(\beta_1)(-\mathbb{P}_0(1,0,0)+\mathbb{P}_0(0,1,0)-\mathbb{P}_0(1,0,1))+\mathbb{P}_0(0,1,1) = 0, \boldsymbol{r}(\theta) \in \mathcal{M}_6\}.$$

Applying Theorem 3.1 and Theorem 2.3 leads to the moment inequalities.

Honoré and Weidner (2020) show that as we vary the initial values $\{y_{-1}, y_0\}$, we get more moment conditions such that both β_1 and β_2 may be point identified. Our results do not contradict with theirs. When we indeed have variation in the initial choices, we can take the intersection of all the implied restrictions on the structure parameters, including both the moment equalities and inequalities. For the case with $T = 3$, we indeed point identify θ. However, our result becomes useful in situations where there is no or only limited variation of $\{Y_{-1}, Y_0\}$ in the population, and hence the structural parameters are only partially identified.

We can also consider the identification of average marginal effect for this model. Denote the transition probability, conditional on A, as:

$$\Pi_{k_1,k_2}(A,\theta) := \mathbb{P}(Y_{t+1} = 1 | Y_t = k_1, Y_{t-1} = k_2, A, \theta) = \frac{AB_1^{k_1}B_2^{k_2}}{1 + AB_1^{k_1}B_2^{k_2}}, \quad \{k_1, k_2\} \in \{0,1\}^2.$$

For any $\{k_1, k_2, \tilde{k}_1, \tilde{k}_2\} \in \{0,1\}^4$, the average marginal effect can be defined as:

$$AME_{y_{-1},y_0}(k_1, k_2, \tilde{k}_1, \tilde{k}_2) = \int \Pi_{k_1,k_2}(A,\theta) - \Pi_{\tilde{k}_1,\tilde{k}_2}(A,\theta) dQ_0(A|y_{-1}, y_0).$$

It is easy to verify that $\Pi_{k_1,k_2}(A,\theta)g(A,\theta,y_{-1},y_0)$ is a polynomial function of A with a degree that is no larger than 6. For instance, consider $k_1 = k_2 = 0$, then $\Pi_{0,0}(A,\theta)g(A,\theta,y_{-1},y_0) = A(1+A)^2(1+AB_1)(1+AB_2)(1+AB_1B_2)$. Applying Theorem 3.4, we can then construct the sharp bound for the average marginal effect from the sharp identified set Θ^*.

4.8 AR(2) Model with Three Periods and a Covariate

Introducing one covariate in the AR(2) model leads to:

$$Y_{it} = \mathbb{1}\{\alpha_i + \beta_1 Y_{it-1} + \beta_2 Y_{it-2} + \gamma X_{it} \geq \epsilon_{it}\}.$$

We restrict attention to one covariate for ease of notation, but the framework easily extends to multiple regressors. For each value of $\boldsymbol{x} \in \mathcal{X}$, again denoting $A = \exp(\alpha)$, $B_1 = \exp(\beta_1)$, $B_2 = \exp(\beta_2)$ and $C = \exp(\gamma)$ and fixing $\{y_{-1}, y_0\} = \{0, 0\}$, we have the following representation of the likelihood of a choice history $\boldsymbol{y}^j$:

$$\mathcal{L}_j(A, \theta, \boldsymbol{x}, y_{-1}, y_0) = A^{\sum_{t=1}^T y_t} B_1^{\sum_{t=1}^T y_t y_{t-1}} B_2^{\sum_{t=1}^T y_t y_{t-2}} C^{\sum_{t=1}^T y_t x_t} / \prod_{t=1}^{T} (1 + A B_1^{y_{t-1}} B_2^{y_{t-2}} C^{x_t}).$$

Take $g(A, \theta, \boldsymbol{x}, y_{-1}, y_0) = \prod_{t=1}^T (1 + AC^{x_t}) \prod_{t=2}^T (1 + AB_1 C^{x_t}) \prod_{t=3}^T (1 + AB_2 C^{x_t})$, such that $\mathcal{L}_j(A, \theta, \boldsymbol{x}, y_{-1}, y_0) g(A, \theta, \boldsymbol{x}, y_{-1}, y_0)$ is a polynomial function of A for all j. When $T = 3$, it is a polynomial function with a degree that is no larger than 7. Therefore, we again have the formulation:

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G(\theta, \boldsymbol{x}) \int_0^{+\infty} \begin{pmatrix} 1 & A & \cdots & A^7 \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_{-1}, y_0)} dQ(A | \theta, \boldsymbol{x}, y_{-1}, y_0).$$

When $\theta \neq 0$ (all elements), the matrix $G(\theta, \boldsymbol{x})$ is of dimension 8×8 and has full rank. Therefore, the left null space is of zero dimension unless $x_2 = x_3$ as observed by Honoré and Weidner (2020). When $x_2 = x_3$, the rank of the matrix $G(\theta)$ is 7, and hence the left null space is of dimension one, yielding one moment equality for each distinct values of $\boldsymbol{x}$ under the assumption $x_2 = x_3$. When it is not possible to impose the equality restriction on the covariate values, we can again consider partial identification of the structural parameters using the moment inequalities due to Theorem 3.1. Our approach will be useful in this particular setting because there exists no moment equality unless $x_2 = x_3$. This allows us to permit regressors that vary over time (i.e., time trend, age variable, or time dummies).

5 Concluding Remarks

We characterize the sharp identified set for the structural parameters in a class of dynamic panel logit models. By reformulating the identification problem as a *truncated moment problem*, we show that all information on the structural parameters can be characterized by a set of moment equality and inequality conditions. We use this result to identify sharp bounds in models where structural parameters are not point identified, rule out false roots in models that cannot be identified using only moment equalities. We then characterize the identifying content of the latent distribution of fixed effects and show that we can only learn a finite vector of generalized moment of the latent distribution. Nevertheless, we provide conditions for a class of functionals of the latent distribution that can be point identified as soon as the structural parameters are point identified. We also discuss cases where functionals can be sharply bounded by only solving a very simple finite dimensional optimization problem. We illustrate the usefulness of our results using a series of examples.

The connection to the truncated moment problem is due to the polynomial structure of the logit distribution with respect to the fixed effects. Any other model that enjoys a similar polynomial structure can make use of our results. Our analytical approach to find moment equality conditions may be generalized to models with multi-dimensional fixed effects. These include the multinomial panel logit model and the bivariate models in which we consider choices of multiple interactive individuals (i.e., those considered in Honoré and Kyriazidou (2019), Honoré and de Paula (2021) and Aguirregabiria, Gu, and Mira (2021)). In these more complicated models, however, it can be challenging to generalize the results on the equivalence between model constraints on the generalized moment vector and a set of moment inequalities. This is due to the fact that the sum of square representation of non-negative polynomial functions only holds for the one-dimensional case. Nevertheless, the connection of the identification problem to the truncated moment problem may still be useful. We leave to future research on the applicability of our results to other models.

References

Aguirregabiria, V., and J. M. Carro (2020): "Identification of Average Marginal Effects in Fixed Effects Dynamic Discrete Choice Models," *Working Paper, University of Toronto.*

Aguirregabiria, V., J. Gu, and Y. Luo (2020): "Sufficient statistics for unobserved heterogeneity in structural dynamic logit models," *Journal of Econometrics, forthcoming.*

Aguirregabiria, V., J. Gu, and P. Mira (2021): "Identification of Structural Parameters in Dynamic Discrete Choice Games with Fixed Effects Unobserved Heterogeneity," *Working Paper.*

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., and B. Honoré (2001): "Panel data models: some recent developments," in *Handbook of econometrics*, vol. 5, pp. 3229–3296. Elsevier.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011): "A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models," *International Economic Review*, 52(3), 807–824.

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

Buchinsky, M., J. Hahn, and K. i. Kim (2010): "Semiparametric information bound of dynamic discrete choice models," *Economics Letters*, 108(2), 109–112.

Chamberlain, G. (1980): "Analysis of covariance with qualitative data," *The Review of Economic Studies*, 47(1), 225–238.

——— (1985): "Heterogeneity, duration dependence and omitted variable bias," *Longitudinal Analysis of Labor Market Data. Cambridge University Press New York.*

——— (1992): "Efficiency bounds for semiparametric regression," *Econometrica*, 60, 567–596.

Chebyshev, P. L. (1874): *Sur les valeurs limites des intégrales*. Imprimerie de Gauthier-Villars.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

Cressie, N., and P. W. Holland (1983): "Characterizing the manifest probabilities of latent trait models," *Psychometrika*, 48(1), 129–141.

Curto, R. E., and L. A. Fialkow (1991): "Recursiveness, positivity and truncated moment problems," *Houston Journal of Mathematics*, 17, 603–635.

Hahn, J. (2001): "The information bound of a dynamic panel logit model with fixed effects," *Econometric Theory*, 17, 913–932.

Heckman, J. J. (1981a): "Heterogeneity and state dependence," in *Studies in Labor Markets*, pp. 91–140. University of Chicago Press.

——— (1981b): "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in *In Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–195. Manski CF, McFadden D (eds). MIT Press: Cambridge, MA.

Honoré, B. E., and A. de Paula (2021): "Identification in Simple Binary Outcome Panel Data Models," *Cemmap Working Paper CWP14/21.*

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

——— (2019): "Panel vector autoregressions with binary data," in *Panel Data Econometrics*, pp. 197–223. Elsevier.

Honoré, B. E., and E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74(3), 611–629.

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

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

Karlin, S., and W. J. Studden (1966): *Tchebycheff systems: With applications in analysis and statistics*, vol. 15. Interscience Publishers.

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

Krein, M., and A. Nudelman (1977): "The Markov moment problem and extremal problems, Transl. Math," *Monographs, American Math. Soc., Providence*, 50.

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

Rasch, G. (1961): "On general laws and the meaning of measurement in psychology," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 4, pp. 321–333.

Wooldridge, J. M. (2005a): "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models," *Review of Economics and Statistics*, 87(2), 385–390.

——— (2005b): "Simple solutions to the initial conditions problem for dynamic nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

A Online Appendix

A.1 Proofs of Lemmas, Theorems, and Propositions

Proof of Theorem 3.1

It suffices to prove the theorem for a specific value of $\boldsymbol{x} \in \mathcal{X}$. To show necessity, we fix a pair $(\theta, Q) \in \mathcal{I}^*(y_0, \boldsymbol{x})$ defined in Definition 2.1 and we show that conditions (a) and (b) have to hold. In particular, we know from (3.1) that

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G(\theta, \boldsymbol{x}) \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A).$$

with $g(A, \theta, \boldsymbol{x}, y_0)$ specified in Appendix A.6. Then it is easy to verify condition (a) by the definition of the set $\boldsymbol{M}_{\boldsymbol{x}}$. Condition (b) can be verified by the fact that $1/g(A, \theta, \boldsymbol{x}, y_0)$ is bounded on the support $\mathcal{A}$ as well as the fact that we can always construct $H(\theta, \boldsymbol{x}) = (G(\theta, \boldsymbol{x})' G(\theta, \boldsymbol{x}))^{-1} G(\theta, \boldsymbol{x})'$ such that we can find a finite positive Borel measure μ supported on $\mathcal{A}$ with $d\mu(A) = \frac{1}{g(A, \beta, \boldsymbol{x}, y_0)} dQ(A)$ whose total mass and the first $2T-1$ moments are represented by $\boldsymbol{r}(\theta, \boldsymbol{x})$.

To show sufficiency, fix an arbitrary pair $(\theta, \boldsymbol{r}(\theta, \boldsymbol{x}))$ that satisfies the conditions (a) and (b) in Theorem 3.1. If there exists no moment equality condition, then this $(\theta, \boldsymbol{r}(\theta, \boldsymbol{x}))$ satisfies the condition (b) only. We will show we can always construct a probability measure Q supported on $\mathcal{A}$, given $(\theta, y_0, \boldsymbol{x})$, consistent with $\boldsymbol{r}(\theta, \boldsymbol{x})$, a vector of moments with respect to a Borel measure μ. In particular, we show that the constructed Q can generate the generalized moments, defined later, such that they are made identical to $\boldsymbol{r}(\theta, \boldsymbol{x})$. For this constructed Q, we also show the following (A.1) holds

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_{\mathcal{A}} \boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) dQ(A) = G(\theta, \boldsymbol{x}) \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A). \tag{A.1}$$

This implies $(\theta, Q) \in \mathcal{I}^*(y_0, \boldsymbol{x})$.

For ease of notation, below we suppress the dependence on these values $(\theta, y_0, \boldsymbol{x})$ in $Q, G, H, \boldsymbol{r}$ and g. First, note that the logit model (A.1) implies $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ has the following representation as $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G \times \boldsymbol{c}$ where $\boldsymbol{c} \in \mathcal{M}_{2T-1}$, and hence $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ is a linear projection on G. We will show (1) $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G \times \boldsymbol{r}$ if $\boldsymbol{r} \in \mathcal{M}_{2T-1}$, $\boldsymbol{r} = H\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$, and $HG = I$, and (2) we can construct a probability measure Q supported on $\mathcal{A}$ such that it generates a set of the generalized moments that match $\boldsymbol{r}$, and hence this Q and θ, satisfying the conditions (a) and (b), generate the model (A.1), concluding this (θ, Q) must be in $\mathcal{I}^*(y_0, \boldsymbol{x})$.

For (1), note that $\boldsymbol{r} = H\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = HG \times \boldsymbol{c}$ for some $\boldsymbol{c} \in \mathcal{M}_{2T-1}$ since $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ is a linear projection on G. It then follows that $\boldsymbol{r} = HG \times \boldsymbol{c} = \boldsymbol{c}$ since $HG = I$, and hence $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G \times \boldsymbol{r}$.

For (2), let Q follow a discrete distribution supported on $2T$ distinct values, denoted as $\{a_1, \ldots, a_{2T}\}$ with a probability measure $\pi_j = \mathbb{P}(A = a_j)$ such that $\sum_{j=1}^{2T} \pi_j = 1$. Note we are not fixing these values, any set of distinct support points work for our construction. Without loss of generality let $0 < a_1 < a_2 < \ldots < a_{2T} < \infty$, then we show we can recover π_j's such that Q generates a set of generalized moments $\int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' / g(A) dQ$ such that they have identical values to the vector $\boldsymbol{r}$. We write a linear system of equations

$$A^g \overline{\pi} \equiv \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1/g(a_1) & a_2/g(a_2) & \cdots & a_{2T}/g(a_{2T}) \\ \vdots & \vdots & \vdots & \vdots \\ a_1^{2T-1}/g(a_1) & a_2^{2T-1}/g(a_2) & \cdots & a_{2T}^{2T-1}/g(a_{2T}) \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{2T} \end{pmatrix} = \begin{pmatrix} 1 \\ r_1 \\ \cdots \\ r_{2T-1} \end{pmatrix} \equiv \overline{\boldsymbol{r}}, \quad \text{(A.2)}$$

where r_j denotes the $j+1$-th element in the vector $\boldsymbol{r}$, the j-th moment with respect to a Borel measure μ. Here note that the zero-th moment $E_Q[1/g(A)]$ is redundant because the probabilities sum to one

$$1 = \mathbf{1}'\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \mathbf{1}'G \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' / g(A) dQ$$

for any Q, and in the system of equations we replace it with the condition $\sum_{j=1}^{2T} \pi_j = 1$, so that the resulting π_j's construct a proper distribution Q.

We know that A^g is nonsingular (a similar argument to Lemma 3.1), and we obtain $\overline{\pi} = (A^g)^{-1}\overline{\boldsymbol{r}}$. Finally, note that $1 = \mathbf{1}'\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \mathbf{1}'G \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' / g(A) dQ = \mathbf{1}'G\boldsymbol{r}$ implies $E_Q[1/g(A)] = r_0$ given Q matches the $2T-1$ generalized moments with $(r_1, \ldots, r_{2T-1})$ in (A.2). This concludes (2).

Proof of Lemma 3.1.

The set of functions in $\mathcal{V}_{\theta,\boldsymbol{x},y_0}(A)$ are linearly independent if $\sum_{j=1}^{2T} c_{j-1}A^{j-1}/g(A,\theta,\boldsymbol{x},y_0) = 0$ holds only with constants $(c_0,\ldots,c_{2T-1}) = 0$. Therefore, to prove the claim, it suffices to show that the determinant of the following $2T \times 2T$ matrix:

$$\begin{pmatrix} \frac{1}{g(a_0)} & \frac{1}{g(a_1)} & \cdots & \frac{1}{g(a_{2T-1})} \\ \frac{a_0}{g(a_0)} & \frac{a_1}{g(a_1)} & \cdots & \frac{a_{2T-1}}{g(a_{2T-1})} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{a_0^{2T-1}}{g(a_0)} & \frac{a_1^{2T-1}}{g(a_1)} & \cdots & \frac{a_{2T-1}^{2T-1}}{g(a_{2T-1})} \end{pmatrix}$$

where we write $g(a) = g(A = a,\theta,\boldsymbol{x},y_0)$ given $(\theta,\boldsymbol{x},y_0)$ for ease of notation, is non-zero for some distinct set of points $a_0,a_1,\ldots,a_{2T-1}$. Now take any distinct set of points such that $0 < a_0 < a_1 < \ldots < a_{2T-1} < \infty$, the determinant of the above matrix can be written as $\left(\prod_{j=0}^{2T-1}\frac{1}{g(a_j)}\right)\prod_{0\le s<u\le 2T-1}(a_u - a_s)$, which is not equal to zero by construction.

Proof of Theorem 3.2

Given the representation (3.1) we have that $\boldsymbol{\mathcal{L}}(A,\theta,\boldsymbol{x},y_0) = G(\theta,\boldsymbol{x})\mathcal{V}_{\theta,\boldsymbol{x},y_0}(A)$ for each $A \in \mathcal{A}$. The vectors in set $\kappa_{\boldsymbol{x}}$ satisfy that $\boldsymbol{m}_{\boldsymbol{x}}'G(\theta,\boldsymbol{x})$ is orthogonal to the vector $\mathcal{V}_{\theta,\boldsymbol{x},y_0}(A)$ for all $A \in \mathcal{A}$. Since the functions in $\mathcal{V}_{\theta,\boldsymbol{x},y_0}$ are linearly independent as shown in Lemma 3.1, the set of vectors $\mathcal{V}_{\theta,\boldsymbol{x},y_0}(A)$ spans $\mathbb{R}^{2T}$, hence the only vector of length $2T$ that can be orthogonal to $\mathbb{R}^{2T}$ is the null vector, i.e. $\boldsymbol{m}_{\boldsymbol{x}}'G(\theta,\boldsymbol{x}) = 0$, which precisely defined the set $\boldsymbol{M}_{\boldsymbol{x}}$.

Proof of Theorem 3.3

For ease of notation, given $(\theta,y_0,\boldsymbol{x})$, we write the choice probabilities, through (3.1), as

$$\boldsymbol{\mathcal{P}} = G\int_{\mathcal{A}} D(A)dQ(A) \equiv GD, \tag{A.3}$$

where Q is a probability distribution supported on $\mathcal{A}$, G is a $2^T \times 2T$ matrix, D is a $2T \times 1$ vector, and $D(A) = (1/g, A/g,\ldots,A^{2T-1}/g)'$ in our representation. In this case $\boldsymbol{\mathcal{P}}$ can be spanned by only (at most) $2T$ number of linearly independent vectors that span D since D

is a $2T \times 1$ vector. We then have

$$\boldsymbol{\mathcal{P}} = G \sum_{l=1}^{2T} \pi_l D(a_l) \equiv G \int_{\mathcal{A}} D(A) d\tilde{Q}(A), \tag{A.4}$$

for some distinct values $\{a_1, \ldots, a_{2T}\}$ in the support $\mathcal{A}$ and weights $\{\pi_1, \ldots, \pi_{2T}\}$ that sums up to 1 such that $D(a_l)$'s are linearly independent. Such a set exists because of Lemma 3.1. This implies for any given $Q(A)$ in (A.3) and a finite vector $\boldsymbol{\mathcal{P}}$ we can always construct an equivalent model to (A.3) using a finite mixture $\tilde{Q}$. Now note that, following a similar construction to the system of equations (A.2), once we know $2T-1$ generalized moments of $Q(A)$, we can find a discrete distribution $\tilde{Q}$ that satisfies (A.4). Therefore knowing $2T-1$ generalized moments of $Q(A)$ from (A.3) exhausts all information of $Q(A)$ we can learn from the model (A.3).

Proof of Theorem 3.4

The parameter of interest $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})]$ can be represented as

$$\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})] = \sum_{j=0}^{2T-1} \eta_j(\theta_0, \boldsymbol{x}) \mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[A^j / g(A, \theta_0, \boldsymbol{x}, y_0)].$$

Since θ is point identified, then we know θ_0. From Theorem 3.3, we know all measures in the set $\mathcal{Q}(\theta_0, y_0, \boldsymbol{x})$ defined in (3.5) have the same vector of generalized moments. Since $Q_0(A|y_0, \boldsymbol{x}) \in \mathcal{Q}(\theta_0, y_0, \boldsymbol{x})$ by construction, we have $r_j(\theta_0, \boldsymbol{x}) = \mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[A^j / g(A, \theta_0, \boldsymbol{x}, y_0)]$, for $j = 0, \ldots, 2T-1$, which are observed quantities: given θ_0, $\boldsymbol{r}(\theta_0, \boldsymbol{x}) = H(\theta_0, \boldsymbol{x}) \mathcal{P}_{\boldsymbol{x}}$. Therefore $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})] = \boldsymbol{\eta}(\theta_0, \boldsymbol{x})' \boldsymbol{r}(\theta_0, \boldsymbol{x})$.

Proof of Theorem 3.5

We show the argument for sharp lower bound since the argument for the sharp upper bound is the same. By Definition 3.1, the lower bound of $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A, \theta_0, \boldsymbol{x})]$ can be written as

$$\begin{aligned} \ell b(\boldsymbol{x}) &= \inf_{\theta \in \Theta^*, Q \in \mathcal{Q}(\theta, y_0, \boldsymbol{x})} \int_{\mathcal{A}} \psi(A, \theta, \boldsymbol{x}) dQ(A) \\ &= \inf_{\theta \in \Theta^*, Q \in \mathcal{Q}(\theta, y_0, \boldsymbol{x})} \int_{\mathcal{A}} \boldsymbol{\eta}(\theta, \boldsymbol{x})' \begin{pmatrix} 1 & A & \ldots, A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A) \end{aligned}$$

$$= \inf_{\theta \in \Theta^*} \boldsymbol{\eta}(\theta, \boldsymbol{x})' \boldsymbol{r}(\theta, \boldsymbol{x}),$$

where the second equality is due to the fact that $\psi(A, \theta, \boldsymbol{x}) g(A, \theta, \boldsymbol{x}, y_0)$ can be represented as a polynomial funciton of A up to degree $2T - 1$ with coefficients $\boldsymbol{\eta}(\theta, \boldsymbol{x})$. The last equality is due to the fact that all measures in the set $\mathcal{Q}(\theta, y_0, \boldsymbol{x})$, defined in (3.5), have the same vector of generalized moments, represented as $\boldsymbol{r}(\theta, \boldsymbol{x})$, which we observe for each given $\theta \in \Theta^*$.

Proof of Proposition 3.1

Using the form of $g(A, \beta, y_0)$ in Appendix A.6, for $k = 0$, we have $\Pi_0(A, \beta) g(A, \beta, y_0) = A(1 + A)^{T-1-y_0}(1 + AB)^{T-1+y_0}$, which is a polynomial function of A of degree $2T - 1$. For $k = 1$, we have $\Pi_1(A, \beta) g(A, \beta, y_0) = AB(1 + AB)^{T-2+y_0}(1 + A)^{T-y_0}$, which is again a polynomial function of A of degree $2T - 1$.

Proof of Proposition 3.2

We discuss four cases. If $k = 0$ and $y_0 = 0$, then

$$\frac{AB^k C^{\tilde{x}}}{1 + AB^k C^{\tilde{x}}} g(A, \theta, \boldsymbol{x}, y_0) = \frac{AC^{\tilde{x}}}{(1 + AC^{\tilde{x}})} \prod_{t=2}^{T} (1 + ABC^{x_t}) \prod_{t=1}^{T} (1 + AC^{x_t}).$$

Since $\tilde{x} \in \{x_2, \dots, x_T\}$, the right hand side is a polynomial function of A up to degree $2T - 1$. If $k = 0$ and $y_0 = 1$, then

$$\frac{AB^k C^{\tilde{x}}}{1 + AB^k C^{\tilde{x}}} g(A, \theta, \boldsymbol{x}, y_0) = \frac{AC^{\tilde{x}}}{(1 + AC^{\tilde{x}})} \prod_{t=1}^{T} (1 + ABC^{x_t}) \prod_{t=2}^{T} (1 + AC^{x_t}),$$

which is again a polynomial function of A up to degree $2T - 1$. Similar argument applies for the case $k = 1, y_0 = 0$ and $k = 1, y_0 = 1$. We omit their forms for brevity.

Proof of Proposition 3.3

Note that

$$\mathcal{L}_j(A, \beta, y_0) = \frac{A^{n^{11}+n^{01}} B^{n^{11}}}{(1 + AB)^{n^{11}+n^{10}} (1 + A)^{n^{01}+n^{00}}},$$

with $n^{kj} = \sum_{t=1}^{T} 1\{y_{t-1} = k, y_t = j\}$ for $k, j \in \{0, 1\}$. Since $\max_{\boldsymbol{y}^j \in \mathcal{Y} \setminus \{1,\ldots,1\}} \left\{n^{11} + n^{01}\right\} \leq T - 1$ and the denominator of $\mathcal{L}_j(A, \beta, y_0)$ is always a polynomial function of A of degree T, then $\mathcal{L}_j(A, \beta, y_0)g(A, \beta, y_0)$ is polynomial of A up to degree $2T - 2$, which implies that $A\mathcal{L}_j(A, \beta, y_0)g(A, \beta, y_0)$ is a polynomial of A up tp degree $2T - 1$.

Proof of Proposition 3.4

For the model (2.1) without covariates, given $g(A, \beta, y_0) = (1 + AB)^{T-1+y_0}(1 + A)^{T-y_0}$, we can verify

$$\psi(A, \boldsymbol{y})g(A, \beta, y_0) = A^{\sum_t y_t}(1 + AB)^{T-1+y_0}(1 + A)^{-y_0}.$$

Since $\sum_t y_t \in [0, T]$, when $y_0 = 0$, it is a polynomial function of A up to degree $2T-1$. For the model (2.1) with covariates, given $g(A, \theta, \boldsymbol{x}, y_0) = \prod_{t=2-y_0}^{T}(1 + ABC^{x_t}) \prod_{t=1+y_0}^{T}(1 + AC^{x_t})$, we can verify

$$\psi(A, \boldsymbol{y}, \boldsymbol{x})g(A, \theta, \boldsymbol{x}, y_0) = A^{\sum_{t=1}^{T} y_t} \prod_{t=1}^{T} C^{x_t y_t} \prod_{t=2-y_0}^{T} (1 + ABC^{x_t})(1 + AC^{x_1})^{-y_0}.$$

Then when $y_0 = 0$, it is a polynomial function of A up to degree $2T - 1$ for any $\boldsymbol{x} \in \mathcal{X}$.

Proof of Proposition 4.1

We discuss the non-singular case when $\det(H_1(\boldsymbol{r}(\beta))) > 0$ and comment on the singular case where $\det(H_1(\boldsymbol{r}(\beta))) = 0$ at last.

For the non-singular case, as discussed in Section 4.2, the identifying condition is $r_j > 0$ for $j = 0, \ldots, 3$, $r_0 r_2 - r_1^2 > 0$ and $r_1 r_3 - r_2^2 \geq 0$. The sign of β_0 is identified because $r_2 > 0$ if and only if $B - 1$ has the same sign as $B_0 - 1$. This result follows directly from the fact that $B_0 > 1$ implies $p_2 > p_1$, and $B_0 < 1$ implies $p_2 < p_1$. Since the sign of $B_0 - 1$ is identified, we can, without loss of generality, restrict our attention to two distinct cases: (a) B and B_0 larger than 1, and (b) B and B_0 smaller than 1.

Since in each case, B and B_0 have the same sign, the argument above implies $r_2 > 0$. Furthermore, the condition $r_1 r_3 - r_2^2 \geq 0$ implies $r_1, r_3 > 0$. Similarly, $r_1 > 0$ and $r_0 r_2 - r_1^2 > 0$ implies $r_0 > 0$. Consequently, we only need to check $r_0 r_2 - r_1^2 > 0$ and $r_1 r_3 - r_2^2 \geq 0$.

Consider case (a), B and B_0 are larger than 1. In this case $r_0r_2 - r_1^2 > 0$ if and only if:

$$-B^2p_1p_2 + B(p_0p_2 - p_0p_1 + p_1p_2 + p_2^2) + (p_0p_1 - p_0p_2 - p_2^2) > 0. \tag{A.5}$$

This expression is a quadratic equation with a discriminant equal to:

$$(p_0p_2 - p_0p_1 + p_1p_2 + p_2^2)^2 + 4p_1p_2(p_0p_1 - p_0p_2 - p_2^2) > 0.$$

This discriminant is strictly positive because $p_0, p_1, p_2 \neq 0$ and $p_2 \neq p_1$. Therefore, the quadratic equation in (A.5) has two distinct real-valued roots, and the quadratic formula implies that its roots have the form:

$$\frac{(p_0p_2 - p_0p_1 + p_1p_2 + p_2^2) \pm \sqrt{(p_0p_2 - p_0p_1 + p_1p_2 + p_2^2)^2 + 4p_1p_2(p_0p_1 - p_0p_2 - p_2^2)}}{2p_1p_2}. \tag{A.6}$$

Since the quadratic equation in (A.5) defines a parabola that opens down, the parameter B must be between these roots. Similarly, in this case, $r_1r_3 - r_2^2 \geq 0$ if and only if:

$$B^2p_1(p_1 - p_2 + p_3) - B(p_1^2 - p_1p_2 + p_1p_3 + p_2p_3) + p_2p_3 \geq 0. \tag{A.7}$$

The discriminant of this quadratic equation equals:

$$(p_1^2 - p_1p_2 + p_1p_3 + p_2p_3)^2 - 4p_1p_2p_3(p_1 - p_2 + p_3) > 0.$$

This discriminant is again strictly positive because $p_0, p_1, p_2 \neq 0$ and $p_2 \neq p_1$. Therefore, the quadratic equation in (A.7) has two distinct real-valued roots, and the quadratic formula implies that its roots have the form:

$$\frac{(p_1^2 - p_1p_2 + p_1p_3 + p_2p_3) \pm \sqrt{(p_1^2 - p_1p_2 + p_1p_3 + p_2p_3)^2 - 4p_1p_2p_3(p_1 - p_2 + p_3)}}{2p_1(p_1 - p_2 + p_3)}. \tag{A.8}$$

Since the quadratic equation in (A.7) defines a parabola that opens up, the parameter B cannot be between these roots. Finally, not all of these roots are needed: In particular, we can show that (i) the smaller root in (A.6) is equal to 1, (ii) the smaller root in (A.8) is smaller than 1, and (iii) the larger root in (A.8) is larger than 1. Together, these results imply the bounds in (4.2).

Define $C_0 = p_0(p_2 - p_1) + p_2(p_1 + p_2)$ and $D_0 = 4p_1p_2(p_0p_1 - p_0p_2 - p_2^2)$. To see that the smaller root in (A.6) is equal to 1, notice that, this root is equal to:

$$\frac{C_0 - \sqrt{C_0^2 + D_0}}{2p_1p_2},$$

and that this root is equal to 1 if and only if $C_0 - 2p_1p_2 = \sqrt{C_0^2 + D_0}$. Indeed, we can show that the left-hand side of this equality is strictly positive whenever $B_0 > 1$. Consequently, this root is equal to 1 if and only if:

$$-4p_1p_2C_0 + 4p_1^2p_2^2 = D_0.$$

It is easy to verify that this equality always holds by simply plugging in C_0 and D_0. Furthermore, it can be shown that $B_0 > 1$ implies $p_3 > p_2$, which, in turn, implies that the smaller root in (A.8) is smaller than 1. It is, therefore, left to show that the larger root in (A.8) is larger than 1. To see this result, let us define $C_1 = p_1D_1 + p_2p_3$, where $D_1 = p_1 - p_2 + p_3$, and assume that this root is, in fact, no larger than 1 such that:

$$\frac{C_1 + \sqrt{C_1^2 - 4p_1p_2p_3D_1}}{2p_1D_1} \leq 1.$$

This inequality leads to a contradiction because it holds if and only if:

$$\sqrt{C_1^2 - 4p_1p_2p_3D_1} \leq 2p_1D_1 - C_1 = p_1D + p_2p_3 = (p_1 + p_3)(p_1 - p_2) < 0.$$

Last we know the identified set is not empty because it has to contain the true points B_0. Therefore, we can rule out the case where the larger root of in (A.8) is located to the right of the larger root in (A.6), which will render an empty set for the identified set.

Consider case (b) where both B and B_0 are smaller than 1. In this case, $r_0r_2 - r_1^2 > 0$ if and only if

$$-B^2p_1p_2 + B(p_0p_2 - p_0p_1 + p_1p_2 + p_2^2) + (p_0p_1 - p_0p_2 - p_2^2) < 0. \tag{A.9}$$

and $r_1r_3 - r_2^2 \geq 0$ if and only if:

$$B^2p_1(p_1 - p_2 + p_3) - B(p_1^2 - p_1p_2 + p_1p_3 + p_2p_3) + p_2p_3 \leq 0. \tag{A.10}$$

We know that these quadratic equations have two distinct real-valued roots each, with the forms in (A.6) and (A.8). Because the quadratic equation in (A.9) defines a parabola that opens down, the parameter B cannot be between the roots in (A.6). Similarly, because the quadratic equation in (A.10) defines a parabola that opens up, the parameter B must be between the roots in (A.8).

Like before, not all of these roots are needed: It can be shown that $B_0 < 1$ implies $p_3 < p_2$, which, in turn, implies that the larger root in (A.8) is larger than 1. We can, therefore, ignore this root. Moreover, since $r_0 r_2 - r_1^2 > 0$ and $r_1 r_3 - r_2^2 \geq 0$ implies $B < 1$, it must be the case that these bounds, when considered together, yield an upper bound no larger than 1, implying that we can also ignore the larger root in (A.6).

Lastly for the singular case where $\det(H_1(\boldsymbol{r}(\beta))) = 0$, the condition for $B \in \Theta^*$ is that there exists $c_0 > 0$ such that $r_j = c_0 r_{j-1}$ for $j = 1, 2, 3$. This implies that $\det(B_1(\boldsymbol{r}(\beta))) = 0$. When $B_0 > 1$, there are only two possibilities for the identification condition to hold, either B equals to the larger root of (A.6) or the larger root of (A.8). Likewise, under the singular case if $B_0 < 1$, then there are only two possibilities for the identification condition to hold, either B equals to the smaller root of (A.6) or the smaller root of (A.8).

A.2 Details of Formulation in (2.4)

For model (2.1) with $T = 2$ and $\gamma = 0$, we have

$$\boldsymbol{\mathcal{L}}(A, \beta, y_0) = \begin{pmatrix} \mathbb{P}((Y_1, Y_2) = (0,0) \mid Y_0 = y_0, \alpha) \\ \mathbb{P}((Y_1, Y_2) = (1,0) \mid Y_0 = y_0, \alpha) \\ \mathbb{P}((Y_1, Y_2) = (0,1) \mid Y_0 = y_0, \alpha) \\ \mathbb{P}((Y_1, Y_2) = (1,1) \mid Y_0 = y_0, \alpha) \end{pmatrix} = \begin{pmatrix} \frac{1}{(1+AB)^{y_0}(1+A)^{2-y_0}} \\ \frac{AB^{y_0}}{(1+AB)^{1+y_0}(1+A)^{1-y_0}} \\ \frac{A}{(1+AB)^{y_0}(1+A)^{2-y_0}} \\ \frac{A^2 B^{y_0+1}}{(1+AB)^{1+y_0}(1+A)^{1-y_0}} \end{pmatrix}.$$

By choosing $g(A, \beta, y_0) = (1 + A)^{2-y_0}(1 + AB)^{1+y_0}$, we get:

$$\boldsymbol{\mathcal{L}}(A, \beta, y_0) = \begin{pmatrix} 1 & B & 0 & 0 \\ 0 & B^{y_0} & B^{y_0} & 0 \\ 0 & 1 & B & 0 \\ 0 & 0 & B^{y_0+1} & B^{y_0+1} \end{pmatrix} \begin{pmatrix} 1 \\ A \\ A^2 \\ A^3 \end{pmatrix} \frac{1}{g(A, \beta, y_0)}.$$

Now take the integral with respect to A and evaluate at $y_0 = 0$, and we get the expression in (2.4) as well as the form of the matrix $G(\beta)$.

A.3 Additional Discussion of Theorem 3.1

A.3.1 Alternative Construction of $H(\theta, x)$

In principle, the matrix $H(\theta, \boldsymbol{x})$ can always be constructed by $(G(\theta, \boldsymbol{x})'G(\theta, \boldsymbol{x}))^{-1}G(\theta, \boldsymbol{x})'$ to fulfill condition (b) in Theorem 3.1. However, sometimes it is more convenient to consider an alternative construction for $H(\theta, \boldsymbol{x})$ via the following procedure especially for $T > 2$.

Device a matrix H_0 of dimension $2T \times 2^T$ boolean matrix which picks $2T$ rows out of 2^T rows of the matrix $G(\theta, \boldsymbol{x})$, denote it as the reduced square matrix $\tilde{G}(\theta, \boldsymbol{x}) := H_0 G(\theta, \boldsymbol{x})$. Since $G(\theta, \boldsymbol{x})$ is full rank, so is the reduced $\tilde{G}(\theta, \boldsymbol{x})$. Perform LU factorization and write $\tilde{G}(\theta, \boldsymbol{x}) = LU$ and let $H(\theta, \boldsymbol{x}) = U^{-1}L^{-1}H_0$. This is more convenient because both U and L are upper or lower traingular matrices, which are easier to invert symbolically. Depending on the set of $2T$ rows one picks, there can exist multiple $H(\theta, \boldsymbol{x})$. That said, we do not need to check all possible choices of $H(\theta, \boldsymbol{x})$. To see this, given (3.1), we always have the relationship, $\boldsymbol{\mathcal{P}}_x = G(\theta, \boldsymbol{x})D(\theta, \boldsymbol{x})$, where $D(\theta, \boldsymbol{x}) = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A|y_0, \boldsymbol{x})$ is the generalized vector of moments of A. Choosing any H such that $HG(\theta, \boldsymbol{x}) = I_{2T}$ leads to $\boldsymbol{r}(\theta, \boldsymbol{x}) = H\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = D(\theta, \boldsymbol{x})$. We may have multiple form of H that fulfills the condition $HG(\theta, \boldsymbol{x}) = I_{2T}$, which will just lead to different expression of $H\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ in terms of the elements in $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$. This is possible because some equations are redundant in the system $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G(\theta, \boldsymbol{x})D(\theta, \boldsymbol{x})$.

A.3.2 Generalization to Situations Without the Full Rank Condition

For this discussion, we suppress the possible dependence of $G, \boldsymbol{\mathcal{P}}$, and g on $(\theta, y_0, \boldsymbol{x})$. If G is not of full column rank, then we can always write it as $G = G_0 C$ where G_0 is comprised of linearly independent column vectors (i.e., a basis), and reformulate the choice probability equation (When G has full column rank we set C to be an identity matrix). Specifically, for the purpose of discussion, let G_0 be a 2^T by $2T - m$ matrix with $0 < m < 2T$ (so the rank of G equals to $2T - m$), and hence C is a $2T - m$ by $2T$ matrix. Let the vector of polynomials in A be $V(A) = (1, A, \dots, A^{2T-1})'$ and the vector of the generalized moments be $V_g = \mathbb{E}_Q[V(A)/g(A)]$. Then the choice probability is given by $\boldsymbol{\mathcal{P}} = GV_g = G_0CV_g$ similar to (3.2), and we can rewrite this as $\boldsymbol{\mathcal{P}} = G_0\tilde{V_g}$ where $\tilde{V_g} = C\mathbb{E}_Q[V(A)/g(A)] = \mathbb{E}_Q[CV(A)/g(A)]$. By construction $CV(A)$ is another vector of polynomials in A.

Given this formulation, we can also decompose the degrees of freedom in $\boldsymbol{\mathcal{P}}$. The left null space of G_0 is of dimension $2^T-(2T-m)$ for any $T \geq 2$. Therefore, we can find $2^T-(2T-m)$ linearly independent vectors that form a basis for this space, where each vector serves as a moment equality condition for identifying θ. Note that this decomposition is now equal to the number of rows in $\boldsymbol{\mathcal{P}}$ (i.e., 2^T) minus the number of rows in $\tilde{V_g}$ (i.e., $2T - m$).

A.4 Details for Section 4.1: AR(1) with Three Periods and No Covariates

The choice probability for model (2.1) with $T = 3$ and $\gamma = 0$ can be written as

$$\boldsymbol{\mathcal{P}} = \int_0^{+\infty} G(\beta) \begin{pmatrix} 1 & A & \cdots & A^5 \end{pmatrix}' \frac{1}{g(A, \beta, y_0)} dQ(A|y_0),$$

with $G(\beta)$ taking the form:

$$G(\beta) = \begin{pmatrix} 1 & 2B & B^2 & 0 & 0 & 0 \\ 0 & 1 & 1+B & B & 0 & 0 \\ 0 & 1 & 1+B & B & 0 & 0 \\ 0 & 1 & 2B & B^2 & 0 & 0 \\ 0 & 0 & B & 2B & B & 0 \\ 0 & 0 & 1 & 1+B & B & 0 \\ 0 & 0 & B & B(1+B) & B^2 & 0 \\ 0 & 0 & 0 & B^2 & 2B^2 & B^2 \end{pmatrix},$$

and $g(A, \beta, y_0) = (1 + A)^3(1 + AB)^2$ for $y_0 = 0$.

To construct $\boldsymbol{r}(\theta)$, consider the matrix $H(\beta)$ to be

$$\frac{1}{(B-1)^2}\begin{pmatrix} (B-1)^2 & 0 & -B^2(2B-3) & B(B-2) & B^3 & -B^3 & 0 & 0 \\ 0 & 0 & B(B-2) & 1 & -B^2 & B^2 & 0 & 0 \\ 0 & 0 & 1 & -1 & B & -B & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1/B & B-2 & 0 & 0 \\ 0 & 0 & -1 & 1 & \frac{B-2}{B} & 3-2B & 0 & \frac{(B-1)^2}{B^2} \end{pmatrix},$$

and we can verify $H(\beta)G(\beta) = I_6$.

For the average marginal effect, we can verify

$$\begin{aligned} \text{AME}_0 &= (B_0-1)\int_0^\infty A(1+A)^2(1+AB_0)\frac{1}{g(A,\beta_0,y_0)}dQ(A \mid y_0) \\ &= \begin{pmatrix} 0 & B_0-1 & (2+B_0)(B_0-1) & (1+2B_0)(B_0-1) & B_0(B_0-1) & 0 \end{pmatrix} \boldsymbol{r}(\beta_0) \\ &= (B_0-1)(\mathbb{P}_0(0,1,0)+\mathbb{P}_0(1,0,1)) \end{aligned}$$

Note that since there are multiple forms of $H(\beta)$ that satisfy the requirement $H(\beta)G(\beta) = I_6$, we may have different representation of AME_0. For example, we can also verify that $AME_0 = (B_0-1)\mathbb{P}_0(0,1,0) + \frac{B_0-1}{B_0}\mathbb{P}_0(0,1,1)$. But their values have to be the same if the model is correct, since we should have the same value for the generalized moments $\boldsymbol{r}(\theta)$ as long as $H(\beta)G(\beta) = I_6$.

A.5 Details for Section 4.4: AR(1) with Three Periods and a Covariate

For the example with three periods and one covariate in Section 4.4, the likelihood has the following form:

$$\mathcal{L}_j(A,\theta,\boldsymbol{x},y_0) = \prod_{t=1}^{3}\frac{\exp(\alpha+\beta y_{t-1}+\gamma x_t)^{y_t}}{1+\exp(\alpha+\beta y_{t-1}+\gamma x_t)} = \frac{A^{\sum_{t=1}^3 y_t}B^{\sum_{t=1}^3 y_t y_{t-1}}C^{\sum_{t=1}^3 x_t y_t}}{\prod_{t=1}^3(1+AB^{y_{t-1}}C^{x_t})},$$

where $A = \exp(\alpha)$, $B = \exp(\beta)$, and $C = \exp(\gamma)$.

When $y_0 = 0$, the matrix $G(\theta, \boldsymbol{x})$ has the following form:

$$\begin{pmatrix}
1 & B(C^{x_2}+C^{x_3}) & B^2C^{x_2+x_3} & 0 & 0 & 0 \\
0 & C^{x_1} & C^{x_1}(C^{x_2}+BC^{x_3}) & BC^{x_1+x_2+x_3} & 0 & 0 \\
0 & C^{x_2} & C^{x_2}(BC^{x_2}+C^{x_3}) & BC^{2x_2+x_3} & 0 & 0 \\
0 & C^{x_3} & BC^{x_3}(C^{x_2}+C^{x_3}) & B^2C^{x_2+2x_3} & 0 & 0 \\
0 & 0 & BC^{x_1+x_2} & BC^{x_1+x_2}(C^{x_2}+C^{x_3}) & BC^{x_1+2x_2+x_3} & 0 \\
0 & 0 & C^{x_1+x_3} & C^{x_1+x_3}(C^{x_2}+BC^{x_3}) & BC^{x_1+x_2+2x_3} & 0 \\
0 & 0 & BC^{x_2+x_3} & BC^{x_2+x_3}(BC^{x_2}+C^{x_3}) & B^2C^{2x_2+2x_3} & 0 \\
0 & 0 & 0 & B^2C^{x_1+x_2+x_3} & B^2C^{x_1+x_2+x_3}(C^{x_2}+C^{x_3}) & B^2C^{x_1+2x_2+2x_3}
\end{pmatrix}.$$

Linear combinations of v_1 and v_2 derived in Section 4.4 lead to moment conditions derived in Honoré and Weidner (2020) for the same model. In particular, for $y_0 = 0$, we have

$$\frac{1}{B(C^{x_2}-C^{x_3})}v_1 - \frac{C^{x_1}}{BC^{x_3}(C^{x_2}-C^{x_3})}v_2 = \begin{pmatrix}0, & -1, & C^{x_1-x_2}, & 0, & C^{x_3-x_2}-1, & -1, & C^{x_1-x_3}/B, & 0\end{pmatrix}' \tag{A.11}$$

$$-\frac{C^{x_2}}{C^{x_1}(C^{x_2}-C^{x_3})}v_1 + \frac{1}{C^{x_2}-C^{x_3}}v_2 = \begin{pmatrix}0, & C^{x_3-x_1}, & -1, & C^{x_2-x_3}-1, & 0, & BC^{x_2-x_1}, & -1, & 0\end{pmatrix}' \tag{A.12}$$

and hence these vectors of coefficients lead to the same moment conditions in Honoré and Weidner (2020) when $y_0 = 0$. Similar derivation can be made for $y_0 = 1$.

A.6 AR(1) Model with General T

We first consider the dynamic logit model with T periods without covariates. The likelihood function for $\boldsymbol{y}^j = \{y_1, \ldots, y_T\}$ can be represented by

$$\mathcal{L}_j(A, \beta, y_0) = \frac{A^{n^{11}+n^{01}}B^{n^{11}}}{(1+AB)^{n^{11}+n^{10}}(1+A)^{n^{01}+n^{00}}},$$

with $n^{kj} = \sum_{t=1}^{T} 1\{y_{t-1} = k, y_t = j\}$ for $k, j \in \{0, 1\}$. Since $\max_{\boldsymbol{y}^j \in \mathcal{Y}}\{n^{11} + n^{10}\} = \max_{\boldsymbol{y}^j \in \mathcal{Y}} \sum_{t=1}^{T} 1\{y_{t-1} = 1\}$ and $\max_{\boldsymbol{y}^j \in \mathcal{Y}}\{n^{01} + n^{00}\} = \max_{\boldsymbol{y}^j \in \mathcal{Y}} \sum_{t=1}^{T} 1\{y_{t-1} = 0\}$, we can take $g(A, \beta, y_0) = (1 + AB)^{T-1+y_0}(1 + A)^{T-y_0}$. With this choice of $g(A, \beta, y_0)$, we can construct the matrix $G(\beta)$ of dimension $2^T \times 2T$ such that:

$$\boldsymbol{\mathcal{P}} = G(\beta) \int_0^\infty \begin{pmatrix}1 & A & \cdots & A^{2T-1}\end{pmatrix}' \frac{1}{g(A, \beta, y_0)} dQ(A|y_0).$$

For model (2.1) with covariates, we allow the distribution of A to depend arbitrarily on the covariates, hence all the analysis is conditioned on $X = \boldsymbol{x} = \{x_1, x_2, \ldots, x_T\}$ (note we do not need to assume $\boldsymbol{x}$ to be a vector taking the same values, it can take any value in the support $\mathcal{X} \subset \mathbb{R}^T$). Here the derivation is made for one covariate. It can be easily generalized to multiple covariates with additional notation. The likelihood of $\boldsymbol{y}^j = \{y_1, y_2, \ldots, y_T\} \in \mathcal{Y}$ is then

$$\mathcal{L}_j(A, \theta, \boldsymbol{x}, y_0) = \frac{A^{\sum_{t=1}^T y_t} B^{\sum_{t=1}^T y_t y_{t-1}} C^{\sum_{t=1}^T x_t y_t}}{\prod_{t=1}^T (1 + AB^{y_{t-1}} C^{x_t})} = \frac{A^{\sum_{t=1}^T y_t} B^{\sum_{t=1}^T y_t y_{t-1}} C^{\sum_{t=1}^T x_t y_t}}{(1 + ABC^{x_1})^{y_0} (1 + AC^{x_1})^{1-y_0} \prod_{t=2}^T (1 + AB^{y_t - 1} C^{x_t})}$$

with $B = \exp(\beta)$ and $C = \exp(\gamma)$. By taking $g(A, \theta, \boldsymbol{x}, y_0) = \prod_{t=2-y_0}^T (1 + ABC^{x_t}) \prod_{t=1+y_0}^T (1 + AC^{x_t})$, we can construct $G(\theta, \boldsymbol{x})$ of dimension $2^T \times 2T$ such that:

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = G(\theta, \boldsymbol{x}) \int_0^\infty \begin{pmatrix} 1 & A & \cdots & A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ(A|\boldsymbol{x}, y_0).$$

A.7 Details for Section 4.6

A.7.1 Information on β by Chamberlain (1992)

As in Section 4.6, we fix $y_0 = 0$. For the two basis vectors displayed in (4.6), we now show that there is no information on β using the Chamberlain (1992) approach. We begin by writing down the GMM representation induced by the moment equalities. Because of redundancy, we only need to consider the first 7 elements of $\mathcal{Y}$. Without loss of generality, let $\tilde{\mathbb{Y}}$ denote the first 7 elements of the vector $\mathbb{Y}$ with elements $\mathbb{1}\{(Y_1, \ldots, Y_T) = \boldsymbol{y}\}$ for $\boldsymbol{y} \in \mathcal{Y}$ and denote the corresponding probabilities as $\tilde{\boldsymbol{\mathcal{P}}}$. The moment restriction has the form:

$$\begin{aligned} \mathbb{E}[v_1^{*\prime} \tilde{\mathbb{Y}} | Y_0 = y_0] &= 0 \\ \mathbb{E}[v_2^{*\prime} \tilde{\mathbb{Y}} | Y_0 = y_0] &= 0, \end{aligned} \tag{A.13}$$

where v_1^* and v_2^* denote the first 7 elements of v_1 and v_2 defined in (4.6), respectively. By Chamberlain (1992)' argument on semiparametric information, we know that the information for (B, C, D) implied by (A.13) is given by $\Delta' \Omega^{-1} \Delta$, where:

$$\Omega = \begin{pmatrix} v_1^{*\prime} \\ v_2^{*\prime} \end{pmatrix} \left(\text{diag}\left(\tilde{\boldsymbol{\mathcal{P}}} \right) - \tilde{\boldsymbol{\mathcal{P}}} \tilde{\boldsymbol{\mathcal{P}}}' \right) (v_1^* \; v_2^*) \quad \text{and} \quad \Delta = \begin{bmatrix} \tilde{\boldsymbol{\mathcal{P}}}' \boldsymbol{V}_1 \\ \tilde{\boldsymbol{\mathcal{P}}}' \boldsymbol{V}_2 \end{bmatrix},$$

in which we make use of the notation:

$$\boldsymbol{V}_1 = \begin{pmatrix} 0 & 0 & 0 \\ D(D-C) & -BD & B(2D-C) \\ -CD & -BD & -BC \\ -CD & -BD & -BC \\ 0 & 1 & 0 \\ D & 0 & B \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \boldsymbol{V}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & D & C-2D \\ -CD & D-BD & C-BC \\ 0 & D & C-2D \\ \frac{1}{B^2}C & -\frac{1}{B} & 1 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Now, notice that, Δ can be decomposed into $[\Delta_1, \Delta_2, \Delta_3]$, where:

$$\Delta_1 = \begin{pmatrix} -CD(p_2+p_3+p_4) + D(Dp_2+p_6) \\ -CDp_3 + \frac{1}{B^2}Cp_5 \end{pmatrix}, \quad \Delta_2 = \begin{pmatrix} p_5 - BD(p_2+p_3+p_4) \\ -p_7 + D(p_2+p_3+p_4-Bp_3) - \frac{1}{B}p_5 \end{pmatrix},$$

$$\text{and } \Delta_3 = \begin{pmatrix} -BC(p_2+p_3+p_4) + B(2Dp_2+p_6) \\ p_5+p_7+C(p_2+p_3+p_4) - 2D(p_2+p_4) - BCp_3 \end{pmatrix}. \tag{A.14}$$

Therefore, with a little bit of algebra, it can be shown that, if we take the vector:

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = [\Delta_2 \ \Delta_3]^{-1} \Delta_1,$$

then we must obtain the following equality:

$$\Delta \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \end{pmatrix}' = 0.$$

This result implies that the partial information for β contained in the moment equalities implied by (4.6) must be equal to zero, as the partial information for B implied by $\Delta'\Omega^{-1}\Delta$ is characterized by the minimum:

$$\min_{\lambda_1,\lambda_2} \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \end{pmatrix} \Delta'\Omega^{-1}\Delta \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \end{pmatrix}',$$

which must equal zero.

A.7.2 Detailed Construction for the Time Dummy Model

In this part, we explicitly let results depend on y_0. For the time dummy case considered in Section 4.6, let

$$\boldsymbol{\mathcal{P}}^{y_0} = \begin{pmatrix} \mathbb{P}((Y_1,Y_2,Y_3)=(1,1,1)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(1,1,0)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(1,0,1)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(1,0,0)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(0,1,1)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(0,1,0)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(0,0,1)|Y_0=y_0) \\ \mathbb{P}((Y_1,Y_2,Y_3)=(0,0,0)|Y_0=y_0) \end{pmatrix} := \begin{pmatrix} p_1^{y_0} \\ p_2^{y_0} \\ p_3^{y_0} \\ p_4^{y_0} \\ p_5^{y_0} \\ p_6^{y_0} \\ p_7^{y_0} \\ p_8^{y_0} \end{pmatrix}.$$

When $y_0 = 0$, the matrix $G(\theta)$, now denoted as $G^0(\theta)$ to explicitly reflect its dependency on y_0, is defined by:

$$G^0(\theta) = \begin{pmatrix} 0 & 0 & 0 & B^2CD & B^2CD(C+D) & B^2C^2D^2 \\ 0 & 0 & BC & BC(C+D) & BC^2D & 0 \\ 0 & 0 & D & CD+BD^2 & BCD^2 & 0 \\ 0 & 1 & C+BD & BCD & 0 & 0 \\ 0 & 0 & BCD & BCD(BC+D) & B^2C^2D^2 & 0 \\ 0 & C & C(BC+D) & BC^2D & 0 & 0 \\ 0 & D & BD(C+D) & B^2CD^2 & 0 & 0 \\ 1 & B(C+D) & B^2CD & 0 & 0 & 0 \end{pmatrix},$$

and $g(A, \theta, y_0) = (1 + A)(1 + AC)(1 + AD)(1 + ABC)(1 + ABD)$.

The vector $\boldsymbol{r}^0(\theta)$ in this case can be constructed from $H^0(\theta)\boldsymbol{\mathcal{P}}^0$ where $H^0(\theta)G^0(\theta) = I_6$.

For instance consider $H^0(\theta)$ to be

$$\begin{pmatrix} 0 & \frac{B^2(C^2+D^2+CD)}{C(B-1)} & \frac{B^2CD-(C+D-BC)B^2(C+D)}{(B-1)(D-C)} & 0 & -\frac{(BD-C-D)B(C+D)+BCD}{D(B-1)(D-C)} & 0 & -\frac{B(C+D)}{D} & 1 \\ 0 & -\frac{B(C+D)}{C(B-1)} & \frac{BC+BD-B^2C}{(B-1)(D-C)} & 0 & \frac{BD-C-D}{D(B-1)(D-C)} & 0 & 1/D & 0 \\ 0 & \frac{1}{C(B-1)} & -\frac{1}{(B-1)(D-C)} & 0 & \frac{1}{BD(B-1)(D-C)} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{D(B-1)(D-C)} & 0 & -\frac{1}{BCD(B-1)(D-C)} & 0 & 0 & 0 \\ 0 & \frac{-1}{BC^2D(B-1)} & -\frac{1}{D^2(B-1)(D-C)} & 0 & \frac{1}{BC^2D(B-1)(D-C)} & 0 & 0 & 0 \\ \frac{1}{B^2C^2D^2} & \frac{D+C}{BC^3D^2(B-1)} & \frac{1}{D^3(B-1)(D-C)} & 0 & \frac{-1}{BC^3D(B-1)(D-C)} & 0 & 0 & 0 \end{pmatrix}.$$

The basis vectors of the left null space of $G^0(\theta)$ reduce to the moment condition for (C, D), as discussed in Section 4.6,

$$0 = (-CD + D^2)p_2^0 - CD(p_3^0 + p_4^0) + Dp_6^0 + \frac{C^2Dp_3^0p_5^0}{-D^2p_4^0 + D(p_5^0 + p_6^0) + (-C + D)p_7^0} \quad \text{(A.15)}$$

and B has a deterministic relationship with (C, D) as

$$B = \frac{-D^2p_4^0 + D(p_5^0 + p_6^0) + (-C + D)p_7^0}{CDp_3^0}. \quad \text{(A.16)}$$

Also, the moment inequality conditions are imposed through $\boldsymbol{r}^0(\theta) = H^0(\theta)\boldsymbol{\mathcal{P}}^0 \in \mathcal{M}_5$.

When $y_0 = 1$, we can make a similar derivation and have $G^1(\theta)$ as

$$\begin{pmatrix} 0 & 0 & 0 & B^3CD & B^3CD(C+D) & B^3C^2D^2 \\ 0 & 0 & B^2C & B^2C(C+D) & B^2C^2D & 0 \\ 0 & 0 & BD & BCD + B^2D^2 & B^2CD^2 & 0 \\ 0 & B & B(C+BD) & B^2CD & 0 & 0 \\ 0 & 0 & BCD & BCD(BC+D) & B^2C^2D^2 & 0 \\ 0 & C & C(BC+D) & BC^2D & 0 & 0 \\ 0 & D & BD(C+D) & B^2CD^2 & 0 & 0 \\ 1 & B(C+D) & B^2CD & 0 & 0 & 0 \end{pmatrix},$$

and the left null space of the matrix $G^1(\theta)$ is spanned by the following two vectors:

$$v_1 = \begin{pmatrix} 0 & -(C-D)/B & -C/B & -C/B & C/BD & 1 & 0 & 0 \end{pmatrix}'$$
$$v_2 = \begin{pmatrix} 0 & -D/B & -(CD-BCD)/B(C-D) & -D/B & (C-BD)/B(C-D) & 0 & 1 & 0 \end{pmatrix}'$$

or equivalently

$$v_1 = \begin{bmatrix} 0 \\ -(C-D)/B \\ -C/B \\ -C/B \\ C/BD \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad -v_1 D + v_2(C-D) = \begin{bmatrix} 0 \\ 0 \\ CD \\ D^2/B \\ -D \\ -D \\ C-D \\ 0 \end{bmatrix}.$$

From $0 = Bv_1' \boldsymbol{\mathcal{P}}^1$ it follows that

$$B = \frac{(C-D)p_2^1 + C(p_3^1 + p_4^1) - Cp_5^1/D}{p_6^1}. \tag{A.17}$$

Also we obtain

$$\begin{aligned} 0 = (-v_1 D + v_2(C-D))' \mathcal{P}^1 &= CDp_3^1 + \frac{D^2 p_4^1}{B} - D(p_5^1 + p_6^1) + (C-D)p_7^1 \\ &= CDp_3^1 - D(p_5^1 + p_6^1) + (C-D)p_7^1 + \frac{D^3 p_4^1 p_6^1}{(C-D)Dp_2^1 + CD(p_3^1 + p_4^1) - Cp_5^1} \end{aligned} \tag{A.18}$$

where the third equality is obtained from (A.17).

Here to construct the vector of generalized moments $\boldsymbol{r}^1(\theta)$, we can take $H^1(\theta)$ as:

$$\begin{pmatrix} 0 & \frac{CBD-B(C+D)^2}{C(1-B)} & \frac{BCD-\{(C+D)-BC\}B(C+D)}{(B-1)(D-C)} & 0 & -\frac{\{BD-C-D\}B(C+D)+BCD}{D(B-1)(D-C)} & 0 & -\frac{B(C+D)}{D} & 1 \\ 0 & \frac{(C+D)}{C(1-B)} & \frac{(C+D)-BC}{(B-1)(D-C)} & 0 & \frac{BD-C-D}{D(B-1)(D-C)} & 0 & 1/D & 0 \\ 0 & \frac{-1}{BC(1-B)} & \frac{-1}{B(B-1)(D-C)} & 0 & \frac{1}{BD(B-1)(D-C)} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{BD(B-1)(D-C)} & 0 & -\frac{1}{BCD(B-1)(D-C)} & 0 & 0 & 0 \\ 0 & \frac{1}{B^2C^2D(1-B)} & -\frac{1}{BD^2(B-1)(D-C)} & 0 & \frac{1}{BC^2D(B-1)(D-C)} & 0 & 0 & 0 \\ \frac{1}{B^3C^2D^2} & \frac{C+D}{B^2C^3D^2(B-1)} & \frac{1}{BD^3(B-1)(D-C)} & 0 & \frac{1}{BC^3D(1-B)(D-C)} & 0 & 0 & 0 \end{pmatrix},$$

such that $H^1(\theta)G^1(\theta) = I_6$ and $\boldsymbol{r}^1(\theta) = H^1(\theta)\boldsymbol{\mathcal{P}}^1$. The moment inequality is imposed through $\boldsymbol{r}^1(\theta) \in \mathcal{M}_5$.