Bound Analysis in Panel Models with Correlated Random Effects

Victor Chernozhukov   Jinyong Hahn   Whitney Newey
MIT   UCLA   MIT

February 18, 2005

1 Introduction & Motivation

Panel data analysis is viewed an important method of controlling for unobserved individual heterogeneity. By exploiting repeated observations across time for each individual economic agents, we may be able to control for the unobserved heterogeneity that may be possibly correlated with explanatory variables. For linear models, methods of controlling for unobserved heterogeneity are well established. A partial list of references is Amemiya and MaCurdy (1986), Anderson and Hsiao (1982), Bhargava and Sargan (1983), Chamberlain (1982), Hausman and Taylor (1981), and Mundlak (1978). Less is known about how to control for unobserved heterogeneity in nonlinear models. The methods that work for linear models do not carry over in a straightforward way to most nonlinear models because it is impossible to eliminate individual effects by some data transformation, except for a small number of special circumstances discussed by Anderson (1970), Chamberlain (1980), Hausman, Hall, and Griliches (1984), or Wooldridge (1997).

One way of dealing with the unobserved individual effects is to treat each such effect as an unobserved random variable, whose joint distribution with observed explanatory variables is nonparametrically specified. See Chamberlain (1984) for an earlier discussion of such correlated random effects approach. Despite its theoretical appeal, only a limited number of estimators have been developed from such perspective. It is primarily because not many identification results are available for nonlinear panel models with correlated random effects, although there do exist notable exceptions including Honoré (1992, 1993), and Honoré and Kyriazidou (2000a, b). Moreover, Chamberlain's (1992) result on the lack of pint identification for panel probit model was perceived by many to be a pessimistic news for nonlinear panel models with correlated random effects.

This paper develops a bound analysis for nonlinear panel models with correlated random effects. In the last decades, a growing body of literature studied inference where parameters of interest are partially identified, cf. Berk (1961) and Manski (2003). Recent examples include Horowitz and Manski (1995), Manski and Tamer (2002), Mullins (2002), Andrews and Berry (2003), and Chernozhukov, Hong, and Tamer (2003). In the panel literature, Honoré and Tamer (2002) were the first to propose consistent estimation of bounds in dynamic panel models with unknown initial conditions. This paper focuses on the bound identification, estimation, and inference methods for general multinomial panel models with correlated random effects. Conditional on the observed explanatory variables, the model becomes a usual mixture model, and the parametric components are not point identified but are restricted to lie in a set. This builds on the intuition in Chamberlain (1992) and Honoré and Tamer (2002).

We develop a set consistent estimator and associated inference based on the nonparametric maximum likelihood estimation (NPMLE) developed by Kiefer and Wolfowitz (1956). Heckman

and Singer (1984) adopted the NPMLE for a point identified mixture model. Our intuition is that the same procedure can be applied to bound identified models. The inference methods that we propose are based on embedding the partially nonparametric likelihood into a more general non-structrual likelihood, which allows us to provide inferential statements about the finite-dimensional parameters. In order to utilize existing computational algorithm, our analysis is yet confined to models with only discrete explanatory variables. On the other hand, our analysis is applicable to any multinomial panel models including probit models, which could not be treated within the previous framework, e.g. Honoré and Kyriazidou (2000a), that relied on point identification, and more complicated dynamic models considered, e.g., by Wolpin (1987).

The estimator of the bound requires an asymptotic framework where the number of individuals in the sample (n) grows to infinity while the time series dimension (T) is fixed. This is in contrast to the recent proposal suggested by Hahn and Newey (2002), Hahn and Kuersteiner (2003), and Woutersen (2003), which is based on an alternative asymptotic approximation where n and T both grow to infinity. Hahn and Kuersteiner (2003) point out that the alternative asymptotic approximation can be viewed as a higher order approximation when n is fixed and T grows to infinity. In other words, the alternative asymptotics based proposal is implicitly based on the idea that the parameter of interest is consistently estimated by the usual fixed effects approach as long as T grows to infinity. When T is not sufficiently large, the alternative asymptotic approximation is probably of limited practical value, and the bound analysis is expected to be more plausible. When T is large, then the bound analysis is probably dominated by the convenience of the alternative asymptotics based procedures.

2 Set Identification and Consistent Estimation

We consider a multinomial panel model with correlated random effects and discrete explanatory variables. In particular, we assume that the vector $\mathbf{Y} = (y_{i1}, \ldots, y_{iT})$ of outcome variables can take J possible values $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(J)}$. We also assume that there exists a vector $\mathbf{X} = (x_{i1}, \ldots, x_{iT})$ of explanatory variables, which can take K possible values $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$. Assume that

$$\Pr\left((y_{i1}, \ldots, y_{iT}) = y^{(j)} \middle| \alpha_i, (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} \right) = \mathcal{L}_{(j,k)}\left(\alpha_i, \beta^* \right)$$

for some finite dimensional β^* and some function $\mathcal{L}_{(j,k)}$. Let $\Lambda^*_{(k)}$ denote the unknown conditional distribution of α_i given $(x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)}$. We then have

$$\Pr\left((y_{i1}, \ldots, y_{iT}) = y^{(j)} \middle| (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} \right) = \int \mathcal{L}_{(j,k)}\left(\alpha, \beta^* \right) \Lambda^*_{(k)}\left(d\alpha \right)$$

Our objective is to estimate β^*.

An example of the above model is a very simple Probit model

$$\Pr\left(y_{it} = 1 \,|\, x_{i1}, x_{i2}, \alpha_i\right) = \Phi\left(\alpha_i + \gamma \cdot d_i + x_{it}'\theta\right)$$

where we assume that y_{it}'s are i.i.d. over time conditional on $(x_{i1}, x_{i2}, \alpha_i)$. Here, d_i denotes the time dummy which is equal to one in the second period. If x is a scalar such that $(x_{i1}, x_{i2}) = (0, 0)$, or $(0, 1)$ with probability one, we have

$$\Pr\left(y_{i1} = 1, y_{i2} = 1 \,|\, (x_{i1}, x_{i2}) = (0, 0), \alpha_i\right) = \Phi\left(\alpha_i\right)\Phi\left(\alpha_i + \gamma\right)$$
$$\Pr\left(y_{i1} = 1, y_{i2} = 1 \,|\, (x_{i1}, x_{i2}) = (0, 1), \alpha_i\right) = \Phi\left(\alpha_i\right)\Phi\left(\alpha_i + \gamma + \theta\right)$$

where $\Phi\left(\cdot\right)$ is the CDF of $\mathcal{N}\left(0, 1\right)$. Let $\mathbf{x}^{(1)} = (0, 0)$ and $\mathbf{x}^{(2)} = (0, 1)$. We then have

$$\Pr\left(y_{i1} = 1, y_{i2} = 1 \,|\, (x_{i1}, x_{i2}) = (0, 0)\right) = \int \Phi\left(\alpha\right)\Phi\left(\alpha + \gamma\right)\Lambda_{(1)}^*\left(d\alpha\right)$$
$$\Pr\left(y_{i1} = 1, y_{i2} = 1 \,|\, (x_{i1}, x_{i2}) = (0, 1)\right) = \int \Phi\left(\alpha\right)\Phi\left(\alpha + \gamma + \theta\right)\Lambda_{(2)}^*\left(d\alpha\right)$$

It is not difficult to see that dynamic logit model considered by Honoré and Kyriazidou (2000) also belongs to the class of models considered in this paper. Consistent point estimation is difficult for both cases because the semiparametric information bound is equal to zero when time dummies are included. See Chamberlain (1992) and Hahn (2001). It is therefore of interest to pursue a bound analysis even for this apparently simple model.

Letting $Q^* \equiv \left(\Lambda_{(1)}^*, \ldots, \Lambda_{(K)}^*\right)$, we can write the individual log likelihood compactly as $L\left(y_i, x_i; \beta, Q\right)$. Due to the usual argument based on Jensen's inequality, we can see that $\left(\beta^*, Q^*\right)$ is such that

$$E\left[L\left(y_i, x_i; \beta, Q\right)\right] \leq E\left[L\left(y_i, x_i; \beta^*, Q^*\right)\right]$$

for every (β, Q). This implies that

$$\sup_Q E\left[L\left(y_i, x_i; \beta, Q\right)\right] \leq \sup_Q E\left[L\left(y_i, x_i; \beta^*, Q\right)\right]$$

for every β. Therefore, if we define B to be the set of β's that maximizes $\sup_Q E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$, i.e.,

$$B \equiv \left\{\beta : \sup_Q E\left[L\left(y_i, x_i; \beta, Q\right)\right] \geq \sup_Q E\left[L\left(y_i, x_i; \beta', Q\right)\right], \quad \forall \beta'\right\}$$

we can easily see that $\beta^* \in B$. In other words, β^* is bound identified by the set B.

Condition 1 *(i) $\mathcal{L}_{(j,k)}\left(\alpha, \beta\right)$ is continuous in (α, β) for all (j, k); (ii) $\beta^* \in \mathbb{B}$ for some compact $\mathbb{B}$; and (iii) α_i has a support contained in a compact set $\mathbb{C}$.*

It is natural to estimate B by the the level set of the finite-sample profile likelihood

$$B_n = \left\{ \beta : \sup_Q \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, x_i; \beta, Q\right) \geq \sup_\beta \sup_Q \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, x_i; \beta, Q\right) - \epsilon_n \right\}$$

where $\epsilon_n > 0$ is the cut-off parameter that shrinks to zero as a function of the sample size, following Manski and Tamer (2002). Similar approach was adopted by Honoré and Tamer (2002).

The parameter is chosen so that

Condition 2 $\epsilon_n \propto n^{-1/2} a_n$ *for some* $a_n \to \infty$ *and* $n^{-1/2} a_n \to 0$.

This choice of the cut-off is not sufficiently precise to be useful in practice. A more useful choice of ϵ_n is provided in the next Section 4.

Characterization and calculation of

$$\sup_Q \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, x_i; \beta, Q\right)$$

for fixed β can be done by using results established by Lindsay (1983a, 1983b, 1995). In econometric literature, Heckman and Singer's (1984) estimator is the best known example that applies such results. We discuss some salient features of Lindsay's results in Section B.

Theorem 1 *Under Conditions 1 and 2, we have*

$$d_H\left(B_n, B\right) = o_p(1),$$

where d_H *is the Hausdorff distance between sets*

$$d_H\left(B_n, B\right) = \max \left[\sup_{b_n \in B_n} \inf_{b \in B} |b_n - b|, \sup_{b \in B} \inf_{b_n \in B_n} |b_n - b| \right]$$

Proof. See Section B. ∎

3   Some Aspects of Computation

3.1   Characterization of Nonparametric MLE

Throughout this section and appendix, we will assume for simplicity of notation a simple probit model, where x_{it} is a scalar and takes following value: $(x_{i1}, x_{i2}) = (0, 0)$. The proof of more general case follows identically after an appropriate change of notation. Note that the likelihood

equal to

$$
\begin{aligned}
L\left(y_{i1}, y_{i2}; \beta, Q\right) \;=\;\; & y_{i1}y_{i2}\log\left(\int \Phi\left(\alpha\right)\Phi\left(\alpha+\beta\right)Q\left(d\alpha\right)\right)\\
& + y_{i1}\left(1-y_{i2}\right)\log\left(\int \Phi\left(\alpha\right)\left(1-\Phi\left(\alpha+\beta\right)\right)Q\left(d\alpha\right)\right)\\
& + \left(1-y_{i1}\right)y_{i2}\log\left(\int \left(1-\Phi\left(\alpha\right)\right)\Phi\left(\alpha+\beta\right)Q\left(d\alpha\right)\right)\\
& + \left(1-y_{i1}\right)\left(1-y_{i2}\right)\log\left(\int \left(1-\Phi\left(\alpha\right)\right)\left(1-\Phi\left(\alpha+\beta\right)\right)Q\left(d\alpha\right)\right)
\end{aligned}
$$

Note that $J=4$ and $K=1$ here.

We first note some important features of computation established by Lindsay (1995, Chapter 5). Fix β, and let $\mathcal{L}_{(1)}\left(\beta, \alpha\right) \equiv \Phi\left(\alpha\right)\Phi\left(\alpha+\beta\right)$, $\mathcal{L}_{(2)}\left(\beta, \alpha\right) \equiv \Phi\left(\alpha\right)\left(1-\Phi\left(\alpha+\beta\right)\right)$, $\mathcal{L}_{(3)}\left(\beta, \alpha\right) \equiv \left(1-\Phi\left(\alpha\right)\right)\Phi\left(\alpha+\beta\right)$, and $\mathcal{L}_{(4)}\left(\beta, \alpha\right) \equiv \left(1-\Phi\left(\alpha\right)\right)\left(1-\Phi\left(\alpha+\beta\right)\right)$. Further define $p_1 \equiv \frac{1}{n}\sum_{i=1}^{n} y_{i1}y_{i2}$, $p_2 \equiv \frac{1}{n}\sum_{i=1}^{n} y_{i1}\left(1-y_{i2}\right)$, $p_3 \equiv \frac{1}{n}\sum_{i=1}^{n}\left(1-y_{i1}\right)y_{i2}$, and $p_4 \equiv \frac{1}{n}\sum_{i=1}^{n}\left(1-y_{i1}\right)\left(1-y_{i2}\right)$. We then have

$$
\exp\left(\frac{1}{n}\sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right)\right) = \prod_{j=1}^{J}\left[\int \mathcal{L}_{(j)}\left(\beta, \alpha\right)Q\left(d\alpha\right)\right]^{p_j}
$$

for $J=4$. Fix β, and consider a vector-valued mapping

$$
\alpha \mapsto \mathcal{L}\left(\beta, \alpha\right) \equiv \left(\mathcal{L}_{(1)}\left(\beta, \alpha\right), \mathcal{L}_{(2)}\left(\beta, \alpha\right), \mathcal{L}_{(3)}\left(\beta, \alpha\right), \mathcal{L}_{(4)}\left(\beta, \alpha\right)\right)'
$$

Let $\Gamma\left(\beta\right) \equiv \left\{\mathcal{L}\left(\beta, \alpha\right) : \alpha \in \mathbb{C}\right\}$. Note that, for each β, and $\Gamma\left(\beta\right)$ is a closed and bounded set due to Condition 1. Now, let $\mathcal{M}\left(\beta\right)$ denote the convex hull of $\Gamma\left(\beta\right)$. By Lindsay (1995, Theorem 18, p. 112), it follows that there exists a unique $\widehat{\mathcal{L}}\left(\beta\right)$ on the boundary of $\mathcal{M}\left(\beta\right)$ that maximizes $\sum_{j=1}^{J} p_j \log\left(l_j\right)$ over all $\left(l_1, l_2, l_3, l_4\right) \in \mathcal{M}\left(\beta\right)$. By Lindsay (1995, Theorem 21, p. 116), the solution $\widehat{\mathcal{L}}\left(\beta\right)$ can be represented as

$$
\left(\int \mathcal{L}_{(1)}\left(\beta, \alpha\right)\widehat{Q}\left(d\alpha\right), \int \mathcal{L}_{(2)}\left(\beta, \alpha\right)\widehat{Q}\left(d\alpha\right), \int \mathcal{L}_{(3)}\left(\beta, \alpha\right)\widehat{Q}\left(d\alpha\right), \int \mathcal{L}_{(4)}\left(\beta, \alpha\right)\widehat{Q}\left(d\alpha\right)\right)'
$$

where $\widehat{Q}$ has no more than J points of support. We can therefore conclude that a solution to the problem $\max_{Q \in \mathbb{Q}} \frac{1}{n}\sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right)$, where $\mathbb{Q}$ is a set of probability measures with support in $\mathbb{C}$, is a discrete distribution with no more than J points of support. Repeating the same argument for

$$
\exp\left(E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]\right) = \prod_{j=1}^{J}\left[\int \mathcal{L}_{(j)}\left(\beta, \alpha\right)Q\left(d\alpha\right)\right]^{\pi_j}
$$

$\pi_1 \equiv E\left[y_{i1}y_{i2}\right]$, $\pi_2 \equiv E\left[y_{i1}\left(1-y_{i2}\right)\right]$, $\pi_3 \equiv E\left[\left(1-y_{i1}\right)y_{i2}\right]$, and $\pi_4 \equiv E\left[\left(1-y_{i1}\right)\left(1-y_{i2}\right)\right]$, we can conclude that a solution to $\max_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$ is a discrete distribution with no more than J points of support.

In order to understand the intuition as to why it suffices to consider discrete distributions with no more than J points of support, let

$$\mathcal{P}\left(\beta\right) \equiv \int_{\mathbb{C}} \mathcal{L}\left(\beta, \alpha\right) dQ_0\left(\alpha\right),$$

where Q_0 is the true mixing distribution. In the population, the set of observationally equivalent parameters Θ is given by all (β', Q') that explain observed frequencies:

$$\mathcal{P} = \int \mathcal{L}_{\mathbb{C}}(\beta', \alpha) dQ'\left(\alpha\right)$$

Any element of Θ therefore maximizes the population likelihood. The true value β_0 solves $\mathcal{P} = \mathcal{P}(\beta_0)$ and hence pair (β_0, Q_0) also maximizes the likelihood.

Define $\Gamma\left(\beta\right) \equiv \{\mathcal{L}\left(\beta, \alpha\right) : \alpha \in \mathbb{C}\}$. For each β, $\Gamma\left(\beta\right)$ is a closed and bounded set due to Condition 1. Now, let $\mathcal{M}\left(\beta\right)$ denote the convex hull of $\Gamma\left(\beta\right)$. Note that $\mathcal{P} \in \mathcal{M}\left(\beta_0\right)$. Since any point in $\mathcal{M}\left(\beta_0\right)$ can be written as a convex combination of at most J vectors located in $\Gamma\left(\beta\right)$,[1] it follows that

$$\mathcal{P} = \sum_{j=1}^{J} \pi_j \mathcal{L}\left(\beta_0, \alpha_j\right).$$

where $(\pi_1, ..., \pi_J)$ is on the unit simplex of dimension J. Thus, the mixing distribution with J points of support $(\alpha_1, ...\alpha_J)$ with the above probabilities $(\pi_1, ..., \pi_J)$ solves the maximum likelihood problem in the population.

3.2   Calculation of Marginal Effects

The problem of calculating marginal effects of different kinds can be reduced to calculating the bounds on partial effects that are computed conditional on $x_{it} = x$. For instance, consider computing bounds on the structural partial effects of the form

$$\beta_j \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta\right) dQ_0\left(\alpha\right)\right],$$

where $\beta_j \cdot \phi(\alpha + x'\beta) = \partial\Phi\left(\alpha + x'\beta\right)/\partial x_{(j)}$. The upper and lower bounds on this effects are given by

$$l_j = \min_{(\beta, Q) \in \Theta} \beta_j \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta\right) dQ\left(\alpha\right)\right] \text{ and } u_j = \max_{(\beta, Q) \in \Theta} \beta_j \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta\right) dQ\left(\alpha\right)\right].$$

It can be shown that it suffices to consider only discrete distributions Q for calculation of l_j and u_j. We will focus on the upper bound u_j; an analogous argument applies to the lower bound l_j.

Let (β^u, Q^u) denote some maximizing parameters such that

$$u_j = \beta_j^u \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta^u\right) dQ^u\left(\alpha\right)\right]$$

---

[1]This seems obvious, but there might be a name for this – Caratheodory's theorem? - V.

The main claim is that for any β^u there exists another discrete mixing distribution Q_L^u with at most $J + 1$ points of support that also solves this equation.

Note that, for any $\epsilon > 0$ we can find a distribution $Q_N^u \in \Theta$ with a large number $N \gg J$ of support points $(\alpha_1, ..., \alpha_N)$ such that

$$u_j - \epsilon < \beta_j^u \cdot \left[\int_{\mathbb{C}} \phi(\alpha + x'\beta^u) dQ_N^u(\alpha) \right] \leq u_j.$$

Our goal is to show that given such Q_N^u, it suffices to allocate its mass over only at most $J + 1$ points of support. Indeed, consider the problem of allocating $(\pi_1, ..., \pi_N)$ among $(\alpha_1, ..., \alpha_N)$ in order to solve

$$\max_{(\pi_1, ..., \pi_N)} \beta_j^u \cdot \left[\sum_{j=1}^{N} \phi\left(\alpha_j + x'\beta^u\right) \pi_j \right]$$

subject to the constraints:

$$\pi_j \geq 0, \quad j = 1, \ldots, N$$
$$\sum_{j=1}^{J} \pi_j \mathcal{L}\left(\beta_0, \alpha_j\right) = \mathcal{P},$$
$$\sum_{j=1}^{J} \pi_j = 1.$$

This a linear program of the form

$$\max_{\pi \in \mathbb{R}^N} c'\pi \quad \text{such that} \quad \pi \geq 0, \quad A\pi = b, \quad 1'y = 1,$$

and any basic feasible solution to this program has N active constraints, of which at most $\text{rank}(A) + 1$ can be equality constraints. This means that at least $N - \text{rank}(A) - 1$ of active constraints are the form $\pi_j = 0$.[2] Hence a basic solution to this linear programming problem will have at least $N - (J + 1)$ zeroes, that is at most $J + 1$ strictly positive π_j's. Thus, we have shown that given the original Q_N^u with $N \gg J$ points of support there exists a distribution $Q_L^u \in \Theta$ with just $J + 1$ points of support such that

$$u_j - \epsilon < \beta_j^u \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta^u\right) dQ_N^u(\alpha) \right] \leq \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta^u\right) dQ_L^u(\alpha) \right] \leq u_j.$$

This construction works for every $\epsilon > 0$.

The final claim is that there exists a distribution $Q_L^u \in \Theta$ with $J + 1$ points of support $(\alpha_1, ..., \alpha_{J+1})$ such that

$$u_j = \beta_j^u \cdot \left[\int_{\mathbb{C}} \phi\left(\alpha + x'\beta^u\right) dQ_L^u(\alpha) \right].$$

---

[2] See, e.g., Theorem 2.3 and Definition 2.9 (ii) in Bertsimas and Tsitsiklis (1997).

Suppose otherwise, then it must be that

$$u_j > u_j - \epsilon \geq \beta_j^u \cdot \left[\int_{\mathbb{C}} \phi(\alpha + x'\beta^u) dQ_L^u(\alpha) \right],$$

for some $\epsilon > 0$ and for *all* Q_L^u with $J+1$ points of support. This immediately gives a contradiction to the previous step where we have shown that, for any $\epsilon > 0$, u_j and the right hand side can be brought close to each other by strictly less than ϵ.

4   Inference

Theorem 1 does not provide any practical guidance on the choice of the cut-off level ϵ_n. It is also desirable that the choice of the cut-off ϵ_n is tied to inferential statements, which appear to pose special challenges in this setting. In this subsection we propose to base inference on the inversion of the nonparametric likelihood ratio, embedding the previous semi-parametric likelihood in a more general nonparametric family. The approach provides conservative inferences about β or its components.

To simplify presentation of ideas define the following model-implied probabilities:

$$
\begin{aligned}
p_{jk}(\beta, Q) &\equiv \Pr\left((y_{i1}, \ldots, y_{iT}) = y^{(j)}, (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} | \beta, Q \right) \\
&= \int \mathcal{L}_{(j,k)}(\alpha, \beta) \Lambda_{(k)}(d\alpha) \times \Pr\left((x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} \right)
\end{aligned}
$$

and it is convenient to denote

$$\mathcal{P}(\beta, Q) \equiv \{ p_{jk}(\beta, Q), j = 1, ..., J, k = 1, ..., K \}.$$

From the proof of Theorem 1, it follows that the model-implied probabilities coincide with the true choice probabilities for the true β^* and some (generally non-unique) pseudo-true Q^*:

$$p_{jk} \equiv \Pr\left((y_{i1}, \ldots, y_{iT}) = y^{(j)}, (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} \right) = p_{jk}(\beta^*, Q^*).$$

Consider also the empirical probabilities

$$\hat{p}_{jk} \equiv \frac{1}{n} \sum_{i=1}^{n} 1\left((y_{i1}, \ldots, y_{iT}) = y^{(j)}, (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)} \right).$$

The nonparametric log-likelihood ratio evaluated at $\mathcal{P}' = \left\{ p'_{jk}, j = 1, ..., J, k = 1, ..., K \right\}$ takes the form

$$LR\left(\mathcal{P}' \right) = n \sum_{j,k} \hat{p}_{jk} \ln \hat{p}_{jk} - n \sum_{j,k} p'_{jk} \ln p'_{jk}.$$

The quantity of especial interest is this one:

$$LR\left(\mathcal{P}\right) = n \sum_{j,k} \hat{p}_{jk} \ln \hat{p}_{jk} - n \sum_{j,k} p_{jk} \ln p_{jk} \tag{1}$$

and its α-quantile is given by

$$c_\alpha\left(\mathcal{P}\right) = \inf_c \left\{ c : P\{LR\left(\mathcal{P}\right) \leq c\} \geq \alpha \right\}.$$

The joint confidence bound for (β^*, Q^*) is then given by

$$I_\alpha\left(\left(\beta^*, Q^*\right)\right) = \left\{ (\beta, Q) \in (B, \mathbb{Q}_L) : LR\left(\mathcal{P}\left(\beta, Q\right)\right) \leq c_\alpha \right\},$$

where $\mathbb{Q}_L$ is the subset of discrete distributions that, conditional on $(x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)}$, have J support points in $\mathbb{C}$. The quantile $c_\alpha\left(\mathcal{P}\right)$ is asymptotically pivotal by the classical Pearson's argument $LR\left(\mathcal{P}\right) \Rightarrow \chi^2\left(\left(J-1\right)K\right)$, hence we have that $c_\alpha\left(\mathcal{P}\right)$ can be consistently estimated by the $\alpha-$ quantile of $\chi^2\left(\left(J-1\right)K\right)$ variable, denoted as $\hat{c}_\alpha$, and the approximate confidence region is then given by

$$\hat{I}_\alpha\left(\left(\beta^*, Q^*\right)\right) = \left\{ (\beta, Q) : LR\left(\mathcal{P}\left(\beta, Q\right)\right) \leq c_\alpha \right\}.$$

The preceding argument established the following result.

Theorem 2 *Under Condition 1, we have*

$$P\{(\beta^*, Q^*) \in \hat{I}_\alpha((\beta^*, Q^*))\} \to \alpha$$

as $n \to \infty$

Theorem 2 also leads to a precise choice of the cut-off level needed to insure consistent estimation in the previous section. One such choice is given by

$$\epsilon_n = \hat{c}_{\alpha_n},$$

where the significance level α_n should tend to 1 such that the α_n -th quantile of $\chi^2\left(\left(J-1\right)K\right)$ variable satisfies Condition 2 as $n \to \infty$ slowly enough. This choice makes the estimating region B_n in Section 1 coincide with desired confidence region of probability level α_n. In practice, α_n may be set equal to some conventional value such as .90 or .95.

9

5 Sieve Estimation

The method proposed in the previous sections critically hinges on the multinomial structure on the distribution of $(y_{i1}, \ldots, y_{iT})$. As such, it is not expected to go beyond multinomial models. In this section, we propose an alternative estimator based on the method of sieves. The method of sieves cannot avoid some degree of arbitrariness for a given finite sample, and thus may be deemed inferior to the Lindsay-type method when dealing with a multinomial model. On the other hand, the method of sieves is immediately generalizable to models such as panel sample selection models.

In order to simplify notation and technical argument, we present the alternative method in the context of multinomial models. Recall that

$$\Pr\left(\left(y_{i1}, \ldots, y_{iT}\right) = y^{(j)} \middle| (x_{i1}, \ldots, x_{iT}) = \mathbf{x}^{(k)}\right) = \int \mathcal{L}_{(j,k)}\left(\alpha, \beta^*\right) \Lambda_{(k)}^*\left(d\alpha\right)$$

and we wrote the individual log likelihood compactly as $L\left(y_i, x_i; \beta, Q\right)$, where $Q^* \equiv \left(\Lambda_{(1)}^*, \ldots, \Lambda_{(K)}^*\right)$, we can write the individual log likelihood compactly as $L\left(y_i, x_i; \beta, Q\right)$. Also recall that the set B is such that

$$B \equiv \left\{\beta : \sup_Q E\left[L\left(y_i, x_i; \beta, Q\right)\right] \geq \sup_Q E\left[L\left(y_i, x_i; \beta', Q\right)\right], \ \forall \beta'\right\}.$$

Maximization of the sample analog of $E\left[L\left(y_i, x_i; \beta, Q\right)\right]$ over all possible distributions Q may be difficult for arbitrary models. It may therefore be useful to consider the method of sieves, and estimate the set B by the the level set of the finite-sample profile likelihood

$$B_{p,n} = \left\{\beta : \sup_{Q \in \mathbb{Q}_n} \frac{1}{n} \sum_{i=1}^n L\left(y_i, x_i; \beta, Q\right) \geq \sup_\beta \sup_{Q \in \mathbb{Q}_n} \frac{1}{n} \sum_{i=1}^n L\left(y_i, x_i; \beta, Q\right) - \epsilon_n\right\}$$

where $\mathbb{Q}_n$ denotes the approximating set, and $\epsilon_n > 0$ is some cut-off parameter. We assume the following high-level assumption on the approximating set $\mathbb{Q}_n$.

Condition 3 $\sup_{\beta \in \mathbb{B}, Q \in \mathbb{Q}_n} \sup \left|\int \mathcal{L}_{(j,k)}\left(\alpha, \beta\right) \Lambda_{(k)}\left(d\alpha\right) - \int \mathcal{L}_{(j,k)}\left(\alpha, \beta\right) \Lambda_{(k)}^*\left(d\alpha\right)\right| = O\left(\sigma_n\right)$ for some $\sigma_n = o\left(1\right)$.

The parameter is now chosen so that

Condition 4 $\epsilon_n \propto \eta_n a_n$ for some $a_n \to \infty$ such that $a_n = (\eta_n)$, where $\eta_n \equiv \max\left(\sigma_n, n^{-1/2}\right)$.

It can be shown that the sieve estimator $B_{p,n}$ is also consistent:

Theorem 3 *Under Conditions 1, 3, and 4, we have $d_H\left(B_{p,n}, B\right) = o_p\left(1\right)$.*

Proof. Identical to Section B, except that Lemma 2 is replaced by Lemma 3. ∎

10

6 Possible Extensions

Our analysis is yet confined to models with only discrete explanatory variables. It would be interesting to extend the analysis to models with continuous explanatory variables. It may be possible to come up with a sieve-type modification. We expect to obtain a consistent estimator of the bound by applying the NPMLE combined with increasing number of partitions of the support of the explanatory variables, but we do not yet have any proof. Empirical likelihood based method should work in a straightforward manner if the panel model of interest is characterized by a set of moment restrictions instead of a likelihood. We may be able to improve the finite-sample property of our confidence region by using Bartlett type corrections.

Appendix

A Some Lemmas

It would be nice to have uniform consistency of $\frac{1}{n}\sum_{i=1}^{n} L(y_{i1}, y_{i2}; \beta, Q)$ for establishing consistency of B_n.

Lemma 1 *Under Condition 1,*

$$\sup_{\beta \in \mathbb{B}, Q \in \mathbb{Q}} \left| \frac{1}{n}\sum_{i=1}^{n} L(y_{i1}, y_{i2}; \beta, Q) - E[L(y_{i1}, y_{i2}; \beta, Q)] \right| = O_{p^*}\left(\frac{1}{\sqrt{n}}\right)$$

Here, $\mathbb{Q}$ denotes the collection of distributions with support contained in contained in a compact set $\mathbb{C}$.

Proof. Note that

$$\frac{1}{n}\sum_{i=1}^{n} L(y_{i1}, y_{i2}; \beta, Q)$$

$$= \left[\frac{1}{n}\sum_{i=1}^{n} y_{i1}y_{i2}\right] \cdot \log\left(\int \Phi(\alpha)\,\Phi(\alpha+\beta)\,Q(d\alpha)\right)$$

$$+ \left[\frac{1}{n}\sum_{i=1}^{n} y_{i1}(1-y_{i2})\right] \cdot \log\left(\int \Phi(\alpha)(1-\Phi(\alpha+\beta))\,Q(d\alpha)\right)$$

$$+ \left[\frac{1}{n}\sum_{i=1}^{n} (1-y_{i1})y_{i2}\right] \cdot \log\left(\int (1-\Phi(\alpha))\,\Phi(\alpha+\beta)\,Q(d\alpha)\right)$$

$$+ \left[\frac{1}{n}\sum_{i=1}^{n} (1-y_{i1})(1-y_{i2})\right] \cdot \log\left(\int (1-\Phi(\alpha))(1-\Phi(\alpha+\beta))\,Q(d\alpha)\right)$$

and

$$E[L(y_{i1}, y_{i2}; \beta, Q)]$$

$$= E[y_{i1}y_{i2}] \cdot \log\left(\int \Phi(\alpha)\,\Phi(\alpha+\beta)\,Q(d\alpha)\right)$$

$$+ E[y_{i1}(1-y_{i2})] \cdot \log\left(\int \Phi(\alpha)(1-\Phi(\alpha+\beta))\,Q(d\alpha)\right)$$

$$+ E[(1-y_{i1})y_{i2}] \cdot \log\left(\int (1-\Phi(\alpha))\,\Phi(\alpha+\beta)\,Q(d\alpha)\right)$$

$$+ E[(1-y_{i1})(1-y_{i2})] \cdot \log\left(\int (1-\Phi(\alpha))(1-\Phi(\alpha+\beta))\,Q(d\alpha)\right)$$

Further note that $\frac{1}{n}\sum_{i=1}^{n}y_{i1}y_{i2} = E[y_{i1}y_{i2}] + O_p\left(\frac{1}{\sqrt{n}}\right)$, etc. Therefore, the requisite uniform convergence with rate $O_p\left(\frac{1}{\sqrt{n}}\right)$

$$\Delta_n = \sup_{\beta\in\mathbb{B}, Q\in\mathbb{Q}}\left|\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q) - E[L(y_{i1}, y_{i2}; \beta, Q)]\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

follows, provided

$$\left|\log\left(\int\Phi(\alpha)\Phi(\alpha+\beta)Q(d\alpha)\right)\right|, \quad \left|\log\left(\int\Phi(\alpha)(1-\Phi(\alpha+\beta))Q(d\alpha)\right)\right|,$$

$$\left|\log\left(\int(1-\Phi(\alpha))\Phi(\alpha+\beta)Q(d\alpha)\right)\right|, \quad \left|\log\left(\int(1-\Phi(\alpha))(1-\Phi(\alpha+\beta))Q(d\alpha)\right)\right|$$

are bounded, which in turn is implied by Condition 1. ∎

From Lemma 1, we obtain one-sided uniform convergence

Lemma 2 *Under Condition 1,*

$$\sup_{\beta\in\mathbb{B}}\left|\sup_{Q\in\mathbb{Q}}\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q) - \sup_{Q\in\mathbb{Q}}E[L(y_{i1}, y_{i2}; \beta, Q)]\right| = O_{p^*}\left(\frac{1}{\sqrt{n}}\right)$$

Proof. Define

$$Q^*(\beta)\in\operatorname*{argsup}_{Q\in\mathbb{Q}}\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q), \ \ Q^{\#}(\beta)\in\operatorname*{argsup}_{Q\in\mathbb{Q}}E[L(y_{i1}, y_{i2}; \beta, Q)].$$

By definition of $Q^*(\beta)$ and $Q^{\#}(\beta)$, we have uniformly in β and for all n,

$$\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q^{\#}(\beta)) - E\left[L\left(y_{i1}, y_{i2}; \beta, Q^{\#}(\beta)\right)\right]$$

$$\leq \ \frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q^*(\beta)) - E\left[L\left(y_{i1}, y_{i2}; \beta, Q^{\#}(\beta)\right)\right]$$

$$\leq \ \frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q^*(\beta)) - E[L(y_{i1}, y_{i2}; \beta, Q^*(\beta))]$$

Hence

$$\left|\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q^*(\beta)) - E\left[L\left(y_{i1}, y_{i2}; \beta, Q^{\#}(\beta)\right)\right]\right| \leq 2\Delta_n = O_{p^*}\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in β, where Δ_n was defined in (2). Because $\Delta_n = O_p\left(\frac{1}{\sqrt{n}}\right)$, we obtain the desired result. ∎

Lemma 3 *Under Conditions 1 and 3,*

$$\sup_{\beta\in\mathbb{B}}\left|\sup_{Q\in\mathbb{Q}_n}\frac{1}{n}\sum_{i=1}^{n}L(y_{i1}, y_{i2}; \beta, Q) - \sup_{Q\in\mathbb{Q}}E[L(y_{i1}, y_{i2}; \beta, Q)]\right| = O_{p^*}(\eta_n)$$

13

Proof. Because of Lemma 1, it suffices to show that

$$\sup_{\beta \in \mathbb{B}} \left| \sup_{Q \in \mathbb{Q}_n} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right| = O\left(\eta_n\right)$$

which follows from Condition 3. ∎

Lemma 4 *Under Condition 1,* $\max_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$ *is continuous in* β.

Proof. By the discussion in Section B, we can see that the problem

$$\max_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$$

can be rewritten as

$$\max_{\substack{\left(\alpha^{(1)}, \dots, \alpha^{(4)}\right) \in \mathbb{C} \\ \left(p^{(1)}, \dots, p^{(4)}\right) \in \mathbb{S}}} \sum_{j=1}^{J} \pi_j \log \left[\sum_{k=1}^{J} \mathcal{L}_{(j)}\left(\beta, \alpha^{(k)}\right) p^{(k)}\right],$$

where $J = 4$ and $\mathbb{S}$ denotes the unit simplex in $\mathbb{R}^J$. Here, $\left(\alpha^{(1)}, \dots, \alpha^{(J)}\right)$ and $\left(p^{(1)}, \dots, p^{(J)}\right)$ characterize a discrete distribution with no more than J points of support. Because the objective function is continuous in $\left(\beta, \alpha^{(1)}, \dots, \alpha^{(J)}, p^{(1)}, \dots, p^{(J)}\right)$, and because $\mathbb{C} \times \mathbb{S}$ is compact, we can apply the Theorem of the Maximum (e.g. Stokey and Lucas 1989, Theorem 3.6), and obtain the desired conclusion. ∎

B   Proof of Theorem 1

We now turn to the proof of Theorem 1.

PART 1: The first part of the proof modifies slightly the argument of Manski and Tamer (2002) for the present context. Define

$$\begin{aligned}
\bar{L}_n^* &\equiv \sup_{\beta \in \mathbb{B}} \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right), \\
L_n^* &\equiv \inf_{\beta \in B} \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right), \\
L^* &\equiv \sup_{\beta \in \mathbb{B}} \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] = \sup_{\beta \in B} \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right], \\
\Delta_n &\equiv \sup_{\beta \in \mathbb{B}, Q \in \mathbb{Q}} \left| \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right|.
\end{aligned} \tag{2}$$

Note that $\sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$ is constant over B by definition, which implies that

$$L^* = \inf_{\beta \in B} \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$$

14

Therefore, we obtain

$$
\begin{aligned}
|L_n^* - L^*| &= \left| \inf_{\beta \in B} \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - \inf_{\beta \in B} \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right| \\
&\leq \sup_{\beta \in B} \left| \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right| \\
&\leq \sup_{\beta \in \mathbb{B}, Q \in \mathbb{Q}} \left| \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right| = \Delta_n
\end{aligned}
$$

Also note that

$$
\left| \bar{L}_n^* - L^* \right| = \left| \sup_{\beta \in \mathbb{B}} \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - \sup_{\beta \in \mathbb{B}} \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right| \leq \Delta_n
$$

It follows that

$$
\left| \bar{L}_n^* - L_n^* \right| \leq \left| \bar{L}_n^* - L^* \right| + \left| L_n^* - L^* \right| \leq \Delta_n + \Delta_n = 2\Delta_n.
$$

Suppose now that $b \in B$. Note that

$$
\bar{L}_n^* - \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; b, Q\right) \leq \bar{L}_n^* - \inf_{\beta \in B} \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) = \bar{L}_n^* - L_n^*
$$

Therefore, if $\epsilon_n > \bar{L}_n^* - L_n^*$, then we have $\bar{L}_n^* - \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; b, Q\right) \leq \epsilon_n$, or

$$
b \in B_n
$$

by definition of B_n. In other words, $\epsilon_n > \bar{L}_n^* - L_n^*$, then $\epsilon_n > \bar{L}_n^* - L_n^*$, $\inf_{b_n \in B_n} |b_n - b| = 0$. Because the choice of b was arbitrary, we can conclude that

$$
\sup_{b \in B} \inf_{b_n \in B_n} |b_n - b| = 0
$$

if $\epsilon_n > \bar{L}_n^* - L_n^*$. Because $\epsilon_n > 2\Delta_n$ with probability converging to one due to Lemma 2 and choice of ϵ_n, it follows that $\sup_{b \in B} \inf_{b_n \in B_n} |b_n - b| = 0$ with probability converging to one.[3]

PART 2: Define

$$
B(\epsilon) \equiv \left\{ \beta : L^* - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \leq \epsilon \right\}
$$

It suffices to show that $B_n \subseteq B(\epsilon)$ with probability converging to one. This is because it would imply $\inf_{b \in B} |b_n - b| < \delta(\epsilon)$ for $(b_n \in B_n)$, which implies

$$
\sup_{b_n \in B_n} \inf_{b \in B} |b_n - b| < \delta(\epsilon),
$$

---

[3] The "probability" here actually means the inner probability. We ignore such measure theoretic subtlety in this paper.

with probability converging to one. Here $\delta(\epsilon)$ that can be made arbitrarily small by making ϵ sufficiently small by continuity of $\sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right]$ in β, which was established in Lemma 4. This would prove that $\sup_{b_n \in B_n} \inf_{b \in B} |b_n - b| = o_p(1)$.

It remains to show that, for any $\epsilon > 0$, we have $B_n \subseteq B(\epsilon)$ with probability converging to one. For this purpose it suffices to show that

$$\sup_{\beta \in B_n} \left[L^* - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right] \leq \epsilon.$$

Note that

$$\left| \sup_{\beta \in B_n} \left(L^* - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right) - \sup_{\beta \in B_n} \left(\bar{L}_n^* - \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) \right) \right|$$

$$\leq \sup_{\beta \in B_n} \left| \left(L^* - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right) - \left(\bar{L}_n^* - \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) \right) \right|$$

$$\leq \left| L^* - \bar{L}_n^* \right| + \sup_{\beta \in B_n} \left| \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right|$$

$$\leq 2\Delta_n.$$

By definition of the level set B_n, we have

$$\sup_{\beta \in B_n} \left[\bar{L}_n^* - \sup_{Q \in \mathbb{Q}} \frac{1}{n} \sum_{i=1}^{n} L\left(y_{i1}, y_{i2}; \beta, Q\right) \right] \leq \epsilon_n.$$

It follows that

$$\sup_{\beta \in B_n} \left[L^* - \sup_{Q \in \mathbb{Q}} E\left[L\left(y_{i1}, y_{i2}; \beta, Q\right)\right] \right] \leq \epsilon_n + 2\Delta_n$$

By Lemma 1 and choice of ϵ_n, we have $\epsilon_n + 2\Delta_n < \epsilon$ with probability converging to one, which shows the requisite claim.

References

[1] ALVAREZ, J., AND M. ARELLANO (1998), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", *unpublished manuscript.*

[2] AMEMIYA, T., AND T.E. MACURDY (1986), "Instrumental Variables Estimation of an Error-Component Model," *Econometrica*, 54, 869-880.

[3] ANDERSON, E. (1970), "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283-301.

[4] ANDERSON, T., AND C. HSIAO (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

[5] ANDREWS, D., AND S. BERRY (2003), "On Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria", *unpublished manuscript.*

[6] BERTSIMAS, D., AND TSITSIKLIS, J. N. (1997), *Introduction to Linear Optimization*, Athena Scientific, Belmont, Massachusetts.

[7] BHARGAVA, A., AND J.D. SARGAN (1983), "Estimating Dynamic Random Effects Models from Panel Data," *Econometrica*, 51, 1635-1659.

[8] CHAMBERLAIN, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225-238.

[9] CHAMBERLAIN, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5-46.

[10] CHAMBERLAIN, G. (1984), "Panel Data," in Z. GRILICHES AND M. INTRILIGATOR eds *Handbook of Econometrics*. Amsterdam: North-Holland.

[11] CHAMBERLAIN, G. (1992), "Binary Response Models for Panel Data: Identification and Information", *unpublished manuscript.*

[12] CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2003), "Parameter Set Inference in a Class of Econometric Models", *unpublished manuscript.*

[13] HAHN, J., AND G. KUERSTEINER (2001), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica*, forthcoming.

[14] HAHN, J. (2001), "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects", *Econometric Theory* 17, 913 - 932.

[15] HAHN, J., AND G. KUERSTEINER (2003), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", *unpublished manuscript.*

[16] HAHN, J., AND W. NEWEY (2002), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *unpublished manuscript.*

[17] HAUSMAN, J.A. AND W.E. TAYLOR (1981), "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

[18] HECKMAN, J.J., AND B. SINGER (1984), "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica* 52, 271 - 320.

[19] HONORÉ, B.E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", *Econometrica* 60, 533 - 565.

[20] HONORÉ, B.E. (1993), "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables", *Journal of Econometrics* 59, 35 - 61.

[21] HONORÉ, B.E., AND E. KYRIAZIDOU (2000a), "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica* 68, 839 - 874.

[22] HONORÉ, B.E., AND E. KYRIAZIDOU (2000b), "Estimation of Tobit-type models with individual specific effects", *Econometric Reviews* 19, 341 - 366.

[23] HONORÉ, B.E., AND E. TAMER (2002), "Bounds on Parameters in Dynamic Discrete Choice Models", *unpublished manuscript.*

[24] HOROWITZ, J., AND C. MANSKI (1995), "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica* 63, 281 - 302.

[25] KIEFER, J., AND J. WOLFOWITZ (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters", *Annals of Mathematical Statistics* 27, 886 - 906.

[26] LINDSAY, B.G. (1983a), "The Geometry of Mixture Likelihoods: A General Theory", *Annals of Statistics* 11, 86 - 94.

[27] LINDSAY, B.G. (1983b), "The Geometry of Mixture Likelihoods, Part II: The Exponential Family", *Annals of Statistics* 11, 783 - 792.

[28] LINDSAY, B.G. (1995), *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 5, IMS: Hayward.

[29] LINDSAY, B.G., AND M.L. LESPERANCE (1995), "A Review of Semiparametric Mixture Models", *Journal of Statistical Planning and Inference* 47, 29 - 39.

[30] MANSKI, C. (1990), "Nonparametric Bounds on Treatment Effects", *American Economic Review Papers and Proceedings* 80, 319 - 323.

[31] MANSKI, C. (2003), *Partial Identification of Probability Distributions*, Springer-Verlag: New York.

[32] MANSKI, C.F., AND E. TAMER (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", Econometrica 70, 519 - 546.

[33] MULLINS, C. (2002), "Identification and Estimation with Contaminated Data: When Do Covariate Data Sharpen Inference?", *unpublished manuscript*.

[34] NEYMAN, J., AND E.L. SCOTT, (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.

[35] STOKEY, N.L., AND R.E. LUCAS (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press: Cambridge.

[36] WOLPIN, K. I. (1987), "Estimating a Structural Job Search Model: the Transition from School to Work", *Econometrica* 55, 801 - 818.

[37] WOOLDRIDGE, J.M. (1997), "Multiplicative Panel Data Models without the Strict Exogeneity Assumption", *Econometric Theory* 13, 667 - 678.

[38] WOUTERSEN, T.M. (2002), "Robustness Against Incidental Parameters," *unpublished manuscript*.