Inference on Semiparametric Multinomial Response Models

Shakeeb Khan[y]
Boston College

Fu Ouyang[z]
University of Queensland

Elie Tamer[x]
Harvard University

May 12, 2019

Abstract

In this paper we explore inference on regression coefficients in semi parametric multinomial response models. We consider cross sectional, and both static and dynamic panel settings where we focus throughout on point inference under sufficient conditions for point identification. The approach to identification uses a matching insight throughout all three models and relies on variation in regressors: with cross section data, we match across individuals while with panel data we match within individuals over time. Across models, IIA is not assumed as the unobserved errors across choices are allowed to be arbitrarily correlated. For the cross sectional model estimation is based on a localized rank objective function, analogous to that used in Abrevaya, Hausman, and Khan (2010), and presents a generalization of existing approaches. In panel data settings rates of convergence are shown to exhibit a curse of dimensionality in the number of alternatives. The results for the dynamic panel data model generalizes the work of Honoré and Kyriazidou (2000) to cover the multinomial case. A simulation study establishes adequate finite sample properties of our new procedures and we apply our estimators to a scanner panel data set.

Keywords: Multinomial choice, Rank Estimation, Adaptive Inference, Dynamic Panel Data.

JEL: C22,C23,C25.

We are grateful for helpful comments received from conference participants at the 2016 Panel Data Work Shop at University of Amsterdam, the 22nd Panel Data Conference in Perth, Australia, the 2016 Australasian Meetings of the Econometric Society at UTS, and seminar participants at various institutions.

[y]Email: shakeeb.khan@bc.edu.

[z]Email: f.ouyang@uq.edu.au.

[x]Email: elietamer@fas.harvard.edu.

1 Introduction

Many important economic decisions involve households' or firms' choice among qualitative or discrete alternatives. Examples are individuals' choice among transportation alternatives, family sizes, residential locations, brands of automobile, health plans etc. The theory of discrete choice is designed to model these kinds of choice settings and panels. Throughout we relax the IIA property by allowing for arbitrary correlation in the proofs of many of the theorems stated in the paper.

2 Semiparametric Multinomial Choice

We consider the standard multinomial choice model where the dependent variable takes one of $J + 1$ mutually exclusive and exhaustive alternatives (numbered from $j = 0$ to $j = J$). Specifically, for individual i, alternative j is assumed to have an unobservable indirect utility y_{ij} for that individual. The alternative with the highest indirect utility is assumed chosen. Thus the observed variable y_{ij} has the form

$$y_{ij} = 1[y_{ij} > y_{ik} \text{ for all } k \neq j]$$

with the convention that $y_{ij} = 0$ indicates choice of alternative j is not made by agent i.

$x_1 = \tilde{x}_1$ *and* $x_2 = \tilde{x}_2$ *. This objective function does not get us information about since with the matching,* z *drops out. But, once* x_1 *and* x_2 *are "known", then one can use another rank procedure in a second step where we condition on* $x_1 = x_2$*. The choice probability for choice 1 for*

In addition, assume a random sample of observations of the vector $(y_i; x_i); i = 1; 2; ::: n$, $G_{1n}^{(1)}(b)$

Then define

$$_2(\ _i;\ \ ) = \ (y_i^{(1)}; x_{i1};\ ; x_{i2})\ \ f_2(x_{i2})$$

where $f_2(\ )$ denotes the density function of x_{i2}. The function $_2(\ ;\ )$ will characterize the limiting distribution of the maximizer of 2.4. We have the limiting distribution theorem theorem, whose proof follows from identical arguments to those used in Abrevaya et al. (2010), and is based on the following regularity conditions:

KWR1 The parameter space B is compact.

KWR2 Random sampling of $(y_i^{(1)}; x_{i1}; y_i^{(2)}; x_{i2})$

KWR3 $_{i1};\ _{i2}$ is distributed independently of $x_{i1}; x_{i2}$, and has density function which is positive on R^2.

KWR4 Conditional on x_{i2}, x_{i1} has rank p.

KWR5 For all in a neighborhood of $_0$ and all $_i$ in its support, $(\ _i;\ )$ is twice continuously differentiable with respect to .

KWR6 The $p \ \ p$ matrix $\nabla_{2}\ _2(\ _i;\ _0)$ is invertible, where $\nabla_{2}\ _2(\ ;\ )$ denotes the second derivative of $_2(\ ;\ )$ with respect to its second argument.

KWR7 The $p \ \ 1$ vector $\nabla_{1}\ _2(\ _i;\ _0)$ has finite second moment, where $\nabla_{1}\ _2(\ ;\ )$ denotes the first derivative of $_2(\ ;\ )$ with respect to its second argument.

KWR8 The density function of x_{2i}, $f_2(\ )$ is ℓ times continuously differentiable with bounded ℓ^{th} derivative, where ℓ is an even integer satisfying $\ell > p/2$.

KWR9 The kernel function $K(\ )$ is of order ℓ and h_n satisfies $\sqrt{n}h_n^{\ell} \to 0$ and $nh_n^p \to \infty$.

Theorem 2.2. *Under Assumptions KWR1-KWR9,*

$$\sqrt{n}(\hat{\ } \ \ _0) \Rightarrow N(0; V^{\ 1}\ \ V^{\ 1})$$

where $V = \frac{1}{2}E[\nabla_{2}\ _2(\ _i;\ _0)$ *and* $\ = \ E[\nabla_{1}\ _2(\ _i;\ _0)\nabla_{1}\ _2(\ _i;\ _0)']$.

3 Panel Data Multinomial Choice

3.1 Static Multinomial Choice

Paralleling the increase in popularity of estimating multinomial response models in applied work is the estimation of panel data models. The increased availability of longitudinal panel data sets has presented new opportunities for econometricians to control for individual unobserved heterogeneity across agents. In linear panel data models, unobserved additive individual-speci c heterogeneity, if assumed constant over time (i.e., " xed effects"), can be controlled for when estimating the slope parameters by rst differencing the observations.

Discrete panel data models have received a great deal of interest in both the econometrics and statistics literature, beginning with the seminal paper of Andersen (1970). For a review of the early work on this model see Chamberlain (

among others. The literature on multinomial choice for panel data is more limited. Recent results include Shi et al. (2018) and Pakes and Porter (2014). The latter is concerned with partial identification. The former achieves point identification. For recent work on partial identification in binary dynamic panel data models under weak assumptions, see also Khan et al. (2019).

Here we propose point identification results under similar weak conditions as ones used in Manski (1987). To illustrate our identification results, assume $T = 2; J = 2$ (So the choice set is $\{0; 1; 2\}$, with 2 time periods) w.l.o.g. and as before impose normalization that $y_{i0t} \equiv 0$ for $t = 1; 2$.

Our identification strategy will be analogous to the cross-sectional case, but now we match and do our comparisons *within* individuals over time as opposed to pairs of individuals. As we will show the analogy is not perfect as we have to condition on "switchers", in a way similar to the estimation of the conditional Logit model in Andersen (1970)

Notation: (i) y

Theorem 3.1. $_0$ *is point identified relative to all* $b \in B \setminus \{ _0\}$. *Let* b

$$
\begin{aligned}
y_{i0t} &= 0 \\
y_{i1t} &= x^{0}_{i1t}\ {}_{0} + \ {}_{0}1[y_{i1(t\ \ 1)} = 1] + \ {}_{i1}\ \ {}_{i1t} \\
y_{i2t} &= x^{0}_{i2t}\ {}_{0} + \ {}_{i2}\ \ {}_{i1t}
\end{aligned}
$$

In this model, the parameters of interest are $_0$ and $_0$. Identi cation is more complicated in dynamic models, even for binary choice. For example, Chamberlain (1985) shows that $_0$ is not identi ed when there are 3 time periods, $t = 0; 1; 2$.[8] Honoré and Kyriazidou (2000) show point identi cation [9] of $_0$ and $_0$ when there are 4 time periods, $t = 0; 1; 2; 3$. Their identi cation is based on conditioning on the subset of the population whose regressors do not change in periods 2 and 3. Finally, Khan et al. (2019) derive sharp bounds for coef cients in dynamic binary choice models with xed effects under weak conditions (allowing for time trends, time dummies, etc).

Our identi cation strategy for the dynamic multinomial choice model is based on conditioning on the subpopulation whose regressors are time-invariant in different manners, depending on which alternative they are associated with. Speci cally, in the three choices, four time periods setting above we condition on the subpopulation whose regressor values for choice 2 do not change in period 1, 2 and 3 and whose regressor values for choice 1 do not change over time in period 2 and 3.

After such conditioning, the problem reduces to identifying parameters in a dynamic binary choice model, for which existing methods can be applied. For example, if the post conditioning model is a dynamic Logit, which would arise if we begin with a dynamic multinomial Logit, we can use the method proposed in Honoré and Kyriazidou (2000), which is valid for four time periods. An attractive feature of their procedure is that when the covariates are discrete, the estimator will converge at the parametric rate with a limiting normal distribution, so conducting inference is relatively easy. We demonstrate both methods for the dynamic multinomial choice model considered here.

For the dynamic multinomial Logit model, we use the following conditional likeli-

[8] But $_0$ is identi ed if $\ _0 = 0$.

[9] Their point identi cation result requires further restrictions on the serial behavior of the exogenous regressors that rules out, among other things, time trends as regressors. Our identi cation result for the dynamic multinomial choice imposes similar restrictions and so also does not allow for time trends as regressors.

for all $\in 2$; and (iii) $x \equiv (x^{0}_{2(12)}; x^{0}_{2(23)}; x^{0}_{1(23)})^{0}$ and the event $f\ x = 0g$. Here we deliberately keep the notation as close as possible to Honoré and Kyriazidou (2000). Then, we outline the regularity conditions for point identification and consistency of our semiparametric estimator based on the objective function (3.3).

DP1 $f(y_i; x_i)g^{n}_{i=1}$ is a random sample of n observations, where $y_i \equiv (y^{0}_{i0}; y^{0}_{i1}; y^{0}_{i2}; y^{0}_{i3})^{0}$ and $x_i \equiv (x^{0}_{i1}; x^{0}_{i2}; x^{0}_{i3})^{0}$

4 Simulation Study

$$y_{i0} = 0$$
$$y_{ij} = x_{ij}^{(1)} + \beta_1 x_{ij}^{(2)} + \beta_2 x_{ij}^{(3)} - \epsilon_{ij};\ j = 1; 2$$

where $x_{ij}^{(1)}; x_{ij}^{(2)}; x_{ij}^{(3)}$ denote the 3 components of the vector x_{ij}, $\beta_1 = \beta_2 = 1$, $x_{i1}^{(1)} \stackrel{iid}{\sim} N(0;1)$, $x_{i2}^{(1)} \stackrel{iid}{\sim}$ Bino(1; 0:5), $x_{ij}^{(k)} \stackrel{iid}{\sim}$ Bino(1; 0:5) for all $j \in \{1; 2\}$ and $k \in \{2; 3\}$, and

$$(\epsilon_{i1}; \epsilon_{i2}) \stackrel{iid}{\sim} MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0{:}5 \\ 0{:}5 & 1 \end{pmatrix}\right)$$

Table 1 reports the results for this benchmark design.

Table 1: (Design 1) 3 Choices, 3 Regressors, 2 Parameters

	β_1				β_2			
	Mean	RMSE	Median	MAD	Mean	RMSE	Median	MAD
N = 250	0.0161	0.4706	0.0104	0.3430	0.0182	0.4798	-0.0135	0.3383
N = 500	0.0418	0.3726	0.0190	0.2297	0.0428	0.3684	0.0224	0.2222
N = 1000	0.0138	0.2619	0.0022	0.1562	0.0098	0.2577	-0.0022	0.1585

As our cross-sectional estimator is "localized" (matching covariates associated with $J - 1$ alternatives), one may worried about that the dimensionality of the design (both in the regressor space and choice space) may have a large effect on the results in Monte Carlo studies. In order to investigate the finite sample performance of the proposed estimator in higher dimensional, more complicated designs, we consider the following two modifications of the benchmark design:

Design 2: We keep the choice set and error distribution unchanged, while add two regressors to the benchmark design. Specifically, we consider the DGP with latent utility functions:

$$y_{i0} = 0$$
$$y_{ij} = x_{ij}^{(1)} + \beta_1 x_{ij}^{(2)} + \beta_2 x_{ij}^{(3)} + \beta_3 x_{ij}^{(4)} + \beta_4 x_{ij}^{(5)} - \epsilon_{ij};\ j = 1; 2$$

where $\beta_1 = \beta_2 = 1$, $\beta_3 = \beta_4 = 0$, $x_{i1}^{(1)} \stackrel{iid}{\sim} N(0;1)$, and all other regressors are iid Bino(1; 0:5). Note that the DGP is the same as for the benchmark case and the only difference is that two additional regressors are included in the estimation.

Design 3: We keep the latent utility fun17 shrink at the parametric rate. This seems true regardless of the number of regressors, though as expected performance for each sample size deteriorates with the number of regressors. However, that is not the case as we increase the size of the choice set. As seen in Table 3, with five choices, the finite sample performance is relatively poor, and furthermore, does not improve with larger sample sizes as well as it did in the other designs. Thus it appears to us that for this model the adversarial effects of dimensionality lie in the choice dimension and not as much in the regressor dimension.[16]

We then turn to examine the finite sample properties of the maximum score estimators for panel data multinomial choice models. We start from the static panel case and consider the design (Design 4) with choice set $\{0;1;2\}$ and a panel of two time period ($T=2$). The latent utility functions for individual i in time period $t \in \{1;2\}$ are

$$
\begin{aligned}
y^*_{i0t} &= 0 \\
y^*_{ijt} &= x^{(1)}_{ijt} + \beta_1 x^{(2)}_{ijt} + \beta_2 x^{(3)}_{ijt} + \alpha_{ij} + \epsilon_{ijt}; \; j = 1;2
\end{aligned}
$$

where $\beta_1 = \beta_2 = 1$, $x^{(1)}_{i1t} \overset{iid}{\sim} N(0;1)$ for all t, all other regressors are iid Bino(1; 0.5), and

$$
(\epsilon_{i11}; \epsilon_{i21}; \epsilon_{i12}; \epsilon_{i22}) \overset{iid}{\sim} MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}\right)
$$

The fixed effects are generated as $\alpha_{i1} = T^{-1}\sum_{t=1}^{T} x_{i1t}$ and $\alpha_{i2} = T^{-1}\sum_{t=1}^{T} x_{i2t} - 0.5$. In Table 4 and 5, we report respectively the results for this static panel design using one-step and two-step maximum score estimators.

Table 4: (Design 4) 3 Choices, 3 Regressors, 2 Parameters, 2 Periods

	β_1	β_2

Table 5: (Design 4, Two-step) 3 Choices, 3 Regressors, 2 Parameters, 2 Periods

	1				2			
	Mean	RMSE	Median	MAD	Mean	RMSE	Median	MAD
$N = 500$	-0.0539	0.6008	-0.0580	0.5144	-0.0562	0.5895	-0.0513	0.4842
$N = 1000$	-0.0413	0.5978	-0.0479	0.5134	-0.0497	0.5732	-0.0477	0.4717
$N = 2000$	-0.0252	0.5557	-0.0356	0.4311	0.0154	0.5632	0.0014	0.4465
$N = 5000$	0.0329	0.4930	-0.0033	0.3598	0.0017	0.4928	-0.0203	0.3568
$N = 10000$	0.0256	0.4438	0.0065	0.3149	0.0389	0.4415	0.0100	0.3131

Our dynamic panel design (Design 5) has the same choice set as the static design but four time periods ($T = 3$, $t \in \{0;1;2;3\}$). The latent utility functions are

$$\begin{aligned} y_{i0t} &= 0; \ t = 0;1;2;3 \\ y_{ij0} &= x^{(1)}_{ij0} + \ x^{(2)}_{ij0} + \ {}_{ij} \ \ {}_{ij0}; \ j = 1;2 \\ y_{i1t} &= x^{(1)}_{i1t} + \ x^{(2)}_{i1t} + \ y_{i1(t \ 1)} + \ {}_{i1} \ \ {}_{i1t}; \ t = 1;2;3 \\ y_{i2t} &= x^{(1)}_{i2t} + \ x^{(2)}_{i2t} + \ {}_{i2} \ \ {}_{i2t}; \ t = 1;2;3 \end{aligned}$$

where $(\ ; \ ) = (1;0.5)$, $y_{i1(t \ 1)} = 1[u_{i1(t \ 1)} > \max\{0; u_{i2(t \ 1)}\}]$, $x^{(1)}_{i1t} \overset{iid}{\ } N(0;1)$ for all t, all other regressors are iid Bino(1; 0:5), and

$$(\ {}_{i1t}; \ {}_{i2t}) \quad MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

independent across i and over time. We use the same way to generate the fixed effects as the static design. One-115.G0 67( )]ı2

be expected as the logit estimator is based on iid type 1 extreme value errors.

Table 8: (Design 5, Logit) 3 Choices, 3 Regressors, 2 Parameters, 4 Periods

	Mean	RMSE	Median	MAD	Mean	RMSE	Median	MAD
N = 500	0.6760	0.8963	0.6036	0.3433	1.5346	3.6605	0.4860	2.0067
N = 1000	0.6535	0.7731	0.6337	0.2569	1.8503	3.8376	2.0596	2.3813
N = 2000								

$x^{(2)}_{ijt}$: "display", 0-1 valued.

$x^{(3)}_{ijt}$: "feature", 0-1valued.

Note that the data set is an unbalanced panel with $n = 136$ households and T varying with i ($\min\{T_i\} = 14$, $\max\{T_i\} = 77$).

As we can see, the results are strikingly different. For the parametric estimators for multinomial Probit relative coefficients for display and feature are 0.1226 and 0.9608, respectively. For multinomial Logit, they are 0.2150 and 1.1829. In each parametric setting, the coefficient (ratio) on display is not significantly different from 0 at the 95% level, whereas the coefficient on feature is significantly positive. For our semiparametric estimates, the results are coefficient estimates of (0:3331; 0:3081) where as above $y_{i1t} = 1[y_{it} = 1]$.

Employing each of our two estimators for the dynamic model, our estimation results were $(\hat{b}_1; \hat{b}_2; b) = (1:5041; 1:4408; 0:5710)$ with criterion function = 3:118645 for the semi-parametric estimator, and $(\hat{b}_1; \hat{b}_2; b) = (0:1274; 1:6865; 0:6185)$ for the Logit. Note for both the semiparametric and Logit estimates the first two estimated coefficients are very different when compared to the static model, indicating the dynamic specification may be relevant for this data set, and ignoring this aspect can lead to misspecification. This point is consistent with the estimated coefficient on lagged choice being quite different from zero, indicating “persistence” in consumer behavior for this product.

Table 12: Parametric and Semiparametric Estimates for Dynamic Panel Data Model

	1	2	
Semiparametric	0.6024	1.2716	0.4005
Conditional Logit	0.8270	2.2931	1.2091

6 Conclusions

In this paper we proposed new estimation procedures for semiparametric multin1o3cclusions

References

ABREVAYA

HAN, A. K. (1987): "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35, 357–362.

HANDEL, B. R. (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *American Economic Review*, 103, 2643–82.

HECKMAN, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–960.

——— (1981): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden, MIT Press.

——— (1991): "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, 75–79.

HODERLEIN, S. AND H. WHITE (2012): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *Journal of Econometrics*, 168, 300–314.

HONORÉ, B. AND E. KYRIAZIDOU (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

ILLANES, G. (2016): "Switching Costs in Pension Plan Choice," *Unpublished manuscript*.

JAIN, D., N. VILCASSIM, AND P. CHINTAGUNTA (1994): "A Random-Coefficients Logit Brand-Choice Model Applied to Panel data," *Journal of Business Economics and Statistics*, 12, 317–328.

KETCHAM, J. D., C. LUCARELLI, AND C. A. POWERS (2015): "Paying attention or paying too much in Medicare Part D," *American Economic Review*, 105, 204–33.

KHAN, S., M. PONOMAREVA, AND E. TAMER (2016): "Identification in Panel Data Models with Endogenous Censoring," *Journal of Econometrics*, 194, 57–75.

——— (2019): "Identification of Dynamic Panel Binary Response Models," Working Paper.

KHAN, S. AND E. TAMER (2018): "Discussion of Simple Estimators for Invertible Index Models by H. Ahn, H. Ichimura, J. Powell, and P. Ruud," *Journal of Business & Economic Statistics*, 36, 11–15.

KIM, J. AND D. POLLARD (1990): "Cube root asymptotics," *The Annals of Statistics*, 18, 191–219.

KLEIN, R. AND R. SPADY (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387–421.

LEE, L.-F. (1995): "Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models," *Journal of Econometrics*, 65, 381–428.

MANSKI

SHI, X., M. SHUM, AND W. SONG (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models using Cyclic Monotonicity," *Econometrica*, 86, 737–761.

A Static and Dynamic Panel Data Estimators

A.1 Static Panel Data Estimators

A.1.1 Consistency

Let denote the event $x_{2(12)} = 0$. To simplify notation, we define $z_1 = x_{2(12)}$, $z_2 = y_{1(12)}$, $z_3 = x_{1(12)}$,

$$Q(b) = f(0)E[\ (b)|\ ]$$

$$Q_n(b) = \frac{1}{nh_n^p}\sum_{i=1}^{n} K(z_{1i}/h_n)\ _i(b)$$

and

$$'(\ ) = f(\ )E[\ (b)|z_1 = \ ]$$

In what follows, we focus on the case where $B \quad \{b \in R^p : b_1 = 1\}$. The case with $B \quad \{b \in R^p : b_1 = \ 1\}$ is symmetric.

Lemma A.1. *Under Assumptions SP3 - SP6,* $Q(\ _0) > Q(b)$ *for all* $b \in B \setminus \{\ _0\}$*.*

Proof of Lemma A.1. Denote $Z_b = \{z_3 : \mathrm{sgn}(z_3^0 b) \neq \mathrm{sgn}(z_3^0\ _0)\}$ for all $b \in B \setminus \{\ _0\}$. Note that $P(\tilde{z}_3^0\tilde{b} = \tilde{z}_3^0\tilde{\ }_0|\ ) < 1$ by Assumption SP5, and $P(z_3^{(1)} \in N|\tilde{z}_3;\ ) > 0$ by Assumption SP4, where $N = \{\ \tilde{z}_3^0\tilde{b} < z_3^{(1)} < \ \tilde{z}_3^0\tilde{\ }_0\} \cup \{\ \tilde{z}_3^0\tilde{\ }_0 < z_3^{(1)} < \ \tilde{z}_3^0\tilde{b}\}$. Therefore,

$$P(Z_b|\ ) = P(z_3^{(1)} \in N|\tilde{z}_3^0\tilde{b} \neq \tilde{z}_3^0\tilde{\ }_0;\ )\ P(\tilde{z}_3^0\tilde{b} \neq \tilde{z}_3^0\tilde{\ }_0|\ ) > 0$$

Then, we have

$$\begin{aligned}
& Q(\ _0) \quad Q(b) \\
&= f(0)E[z_2(\mathrm{sgn}(z_3^0\ _0) \quad \mathrm{sgn}(z_3^0 b))|\ ] \\
&= 2f(0)\int_{Z_b} \mathrm{sgn}(z_3^0\ _0)E[z_2|z_3;\ ]\, dF_{z_3|} \\
&= 2f(0)\int_{Z_b} \mathrm{sgn}(z_3^0\ _0)E[E[z_2|x;\ ;\ ]\,|z_3;\ ]\, dF_{z_3|} \\
&= 2f(0)\int_{Z_b} E[\mathrm{sgn}(z_3^0\ _0)(P(y_{11} = 1|x;\ ;\ ) \quad P(y_{12} = 1|x;\ ;\ ))\,|z_3;\ ]\, dF_{z_3|}
\end{aligned}$$

Next, note that by definition,

$$P(y_{11} = 1|x;\ ;\ ) = P(x_{11}^0\ _0 + \ _1 \quad _{11} > \max\{0; x_{21}^0\ _0 + \ _2 \quad _{21}\}|x;\ ;\ )$$

and

$$P(y_{12} = 1|x;\ ;\ ) = \ P(x_{12}'\ _0 + \ _1 \ \ _{12} > \max\{0; x_{22}'\ _0 + \ _2 \ \ _{22}\}|x;\ ;\ )$$

Hence, by Assumption SP3, we have $\mathrm{sgn}(P(y_{11} = 1|x;\ ;\ ) \quad P(y_{12} = 1|x;\ ;\ )) = \mathrm{sgn}(z_3'\ _0)$. Furthermore, $P(y_{11} = 1|x;\ ;\ ) = \ P(y_{12} = 1|x;\ ;\ )$

1 . Then, as $\mathcal{F}_b$ is Euclidean for the constant envelope 1 (see Example 2.11 in Pakes and Pollard (1989)), $\mathcal{F}$ is Euclidean for the constant envelope $\sup_{v \in R^p} |K(v)| < 1$. Next, note that by Assumptions SP6 and SP8(ii),

$$\begin{aligned}\sup_{\mathcal{F}_n} E|K(z_1/h_n)\ \ (b)| &= \sup_{\mathcal{F}_n} \int E[|K(z_1/h_n)\ \ (b)||z_1]f(z_1)dz_1 \\ &= \sup_{\mathcal{F}_n} h_n^p \int K(v)E[|\ \ (b)||z_1 = vh_n]f(vh_n)dv \\ &\ \ \ \sup_{\mathcal{F}_n} h_n^p \int K(v)f(vh_n)dv = O(h_n^p)\end{aligned}$$

Then, under Assumption SP9(ii), applying Lemma 5 in Honoré and Kyriazidou (2000) yields

$$\sup_{\mathcal{F}_n} h_n^p |Q_n(b)\ \ \ EQ_n(b)| = O_p\left(\sqrt{\frac{h_n^p \log n}{n}}\right) = o_p(h_n^p)$$

As the final step, we show that $\sup_{b \in B} |Q(b)\ \ \ EQ_n(b)| = o(1)$. Notice that by Assumptions SP7, SP8(ii), SP8(iii), and SP9(i),

$$\begin{aligned}\sup_{b \in B} |Q(b)\ \ \ EQ_n(b)| &= \sup_{b \in B} |\varphi\ (0)\ \ \ h_n^{\ p} \int K(z_1/h_n)\varphi\ (z_1)dz_1| \\ &= \sup_{b \in B} |\varphi\ (0)\ \ \ h_n^{\ p} \int K(z_1/h_n)[\varphi\ (0) + \varphi^{(1)}(\ \ )^0 z_1]dz_1|\end{aligned}$$

where $\psi_n(z) = 2h_n^{-(J-1)p}K(z_1/h_n)z_2$. By definition and change of variables, we have

$$E[\psi_n(z)^2|z_3] = 4h_n^{-(J-1)p}\int K(v)^2 f_{z_1|z_2\neq 0,z_3}(vh_n)P(z_2\neq 0|z_3)dv$$

almost surely for all n. Under Assumptions SP3, SP6', and SP8'(i), there exist some $c_1, c_2 > 0$ such that $c_1 < h_n^{(J-1)p}E[\psi_n(z)^2|z_3] < c_2$ almost surely. Then, using the same argument in Seo and Otsu (2018) (Section B.1 of the supplementary material), we have for all $b_1, b_2 \in B$,

$$\begin{aligned}
&h_n^{(J-1)p/2}\|g_{n,b_1}(z) - g_{n,b_2}(z)\|_2\\
&= E[h_n^{(J-1)p}\psi_n(z)^2(1[z_3'b_1>0] - 1[z_3'b_2>0])^2]^{1/2}\\
&= E[h_n^{(J-1)p}E[\psi_n(z)^2|z_3](1[z_3'b_1>0] - 1[z_3'b_2>0])^2]^{1/2}\\
&\geq c_1^{1/2}E|1[z_3'b_1>0] - 1[z_3'b_2>0]| \gtrsim |b_1 - b_2|_2 \qquad (A.1)
\end{aligned}$$

where $\|\cdot\|_2$ denotes the $L_2(P)$ norm. Similarly, we can obtain

$$\begin{aligned}
&h_n^{(J-1)p}E[\sup_{b\in B:|b-\beta|_2<\varepsilon}|g_{n,b}(z) - g_{n,\beta}(z)|^2]\\
&= E[h_n^{(J-1)p}E[|\psi_n(z)|^2|z_3]\sup_{b\in B:|b-\beta|_2<\varepsilon}|1[z_3'b>0] - 1[z_3'\beta>0)|^2]\\
&\leq c_2E\sup_{b\in B:|b-\beta|_2<\varepsilon}|1[z_3'b>0] - 1[z_3'\beta>0]| \leq c_2'\varepsilon \qquad (A.2)
\end{aligned}$$

for some $c_2' > 0$, sufficiently large n, and all β in a neighborhood of β_0.

Next, note that under Assumptions SP8'(ii)-(iv), SP7', and SP9'(iii), we have

$$\begin{aligned}
E[g_{n,b}(z)] &= \int K(v)E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = vh_n]f_{z_1}(vh_n)dv\\
&= f_{z_1}(0)E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = 0]\\
&\quad + h_n^2\int K(v)v'\frac{\partial^2 f_{z_1}(\zeta)E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = \zeta]}{\partial\zeta\partial\zeta'}\Big|_{\zeta=\bar{v}}v\,dv\\
&= f_{z_1}(0)E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = 0] + o((nh_n^{(J-1)p})^{-2/3}) \qquad (A.3)
\end{aligned}$$

where $\bar{v}$ is a point on the line joining 0 and vh_n, and the second equality follows from the dominated convergence theorem and mean value theorem.

Denote $Z_b = \{z_3 : \mathrm{sgn}(z_3'b) \neq \mathrm{sgn}(z_3'\beta_0)\}$ for all $b \in B\setminus\{\beta_0\}$. Following similar argument used in the proof of Lemma A.4,

$$\begin{aligned}
E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = 0] &= 2\int_{Z_b}\mathrm{sgn}(z_3'\beta_0)E[z_2|z_3, z_1 = 0]dF_{z_3|z_1=0}\\
&= 2\int_{Z_b}|E[z_2|z_3, z_1 = 0]|dF_{z_3|z_1=0} > 0
\end{aligned}$$

Therefore, applying the same argument as Kim and Pollard (1990) pp. 214-215 yields

$$\frac{\partial}{\partial b} E[z_2(\mathrm{sgn}(z_3'b)|z_1 = 0]|_{b=\beta_0} = 0 \tag{A.4}$$

and

$$\frac{\partial^2 E[z_2(\mathrm{sgn}(z_3'b) - \mathrm{sgn}(z_3'\beta_0))|z_1 = 0]}{\partial b \partial b'}$$
$$= \int 1[z_3'\beta_0 = 0] \left(\frac{\partial}{\partial z_3} E[z_2|z_3; z_1 = 0]\right)' \beta_0 z_3 z_3' f_{z_3|z_1=0}(z_3) d\sigma_0 \tag{A.5}$$

where σ_0 is the surface measure on the boundary of $\{z_3 : z_3'\beta_0 \geq 0\}$.

Putting (A.3), (A.4), and (A.5) together, we have

$$E[g_{n,\beta}(z)] = 1$$

A.2 Dynamic Panel Data Estimators

Here, we only establish regularity conditions and prove consistency of our dynamic panel data estimator, as consistency for the static model follows as a special case.

Consider the events:

$$A = \{y_{10} = d_0; y_{11} = 1; y_{12} = 0; y_{13} = d_3\}$$
$$B = \{y_{10} = d_0; y_{11} = 0; y_{12} = 1; y_{13} = d_3\}$$

where d_0 and d_3 are either 0 or 1. In what follows, denote $z = (x'_{1(12)}; y_{1(03)})'$.

Lemma A.2. *Under Assumption DP3,* $\mathrm{sgn}(P(A|x; \; ; \;) \quad P(B|x; \; ; \;)) = \quad \mathrm{sgn}(z' \; _0)$.

Proof of Lemma B.1. By Assmption DP3, we have

$$P(A|x; \; ; \;) = P(y_{10} = 1|x; \; ; \;)^{d_0}(1 \quad P(y_{10} = 1|x; \; ; \;))^{1 \; d_0}$$
$$P(x'_{11} \; _0 + \; _0 d_0 + \; _1 \quad _{11} > \max\{x'_{21} \; _0 + \; _2 \quad _{21}; 0\}|x; \; ; \;)$$
$$(1 \quad P(x'_{12} \; _0 + \; _0 + \; _1 \quad _{12} > \max\{x'_{21} \; _0 + \; _2 \quad _{22}; 0\}|x; \; ; \;))$$
$$P(x'_{12} \; _0 + \; _1 \quad _{13} > \max\{x'_{21} \; _0 + \; _2 \quad _{23}; 0\}|x; \; ; \;)^{d_3}$$
$$(1 \quad P(x'_{12} \; _0 + \; _1 \quad _{13} > \max\{x'_{21} \; _0 + \; _2 \quad _{23}; 0\}|x; \; ; \;))^{1 \; d_3}$$

and similarly,

$$P(B|x; \; ; \;) = P(y_{10} = 1|x; \; ; \;)^{d_0}(1 \quad P(y_{10} = 1|x; \; ; \;))^{1 \; d_0}$$
$$(1 \quad P(x'_{11} \; _0 + \; _0 d_0 + \; _1 \quad _{11} > \max\{x'_{21} \; _0 + \; _2 \quad _{21}; 0\}|x; \; ; \;))$$
$$P(x'_{12} \; _0 + \; _1 \quad _{12} > \max\{x'_{21} \; _0 + \; _2 \quad _{22}; 0\}|x; \; ; \;)$$
$$P(x'_{12} \; _0 + \; _0 + \; _1 \quad _{13} > \max\{x'_{21} \; _0 + \; _2 \quad _{23}; 0\}|x; \; ; \;)^{d_3}$$
$$(1 \quad P(x'_{12} \; _0 + \; _0 + \; _1 \quad _{13} > \max\{x'_{21} \; _0 + \; _2 \quad _{23}; 0\}|x; \; ; \;))^{1 \; d_3}$$

It is not hard to verify that

$$\frac{P(A|x; \; ; \;)}{P(B|x; \; ; \;)} > 1 \quad , \quad x'_{11} \; _0 + \; _0 d_0 > x'_{12} \; _0 + \; _0 d_3$$

for each of the 4 cases corresponding to the values of d_0 and d_3. Then, the desired result follows. □

In what follows, we focus on the case where $\quad \{ \quad = (b', g)' \; 2 \; \mathbb{R}^{p+1} : b_1 = 1\}$. The case with $\quad \{ \quad = (b', g)' \; 2 \; \mathbb{R}^{p+1} : b_1 = \quad 1\}$ is symmetric.

Lemma A.3. *Under Assumptions DP3 - DP5,* $P(\mathrm{sgn}(z^0\;) \neq \mathrm{sgn}(z^0\;{}_0)|\;) > 0$ *for all* $\;2$ $n\, f\;{}_0 g.$

Proof of Lemma B.2. To prove the statement in the lemma, it suffices to show that for all $\;2\;\; n\, f\;{}_0 g$, (i) $P(\tilde{z}^{0}\neq \tilde{z}^{0}\;{}_0|\;) > 0$, and (ii) $P(x^{(1)}_{1(12)}\; 2\; N|\tilde{x}_{1(12)};y_{10} = d_0; y_{13} = d_3;\;) > 0$ for all $(d_0; d_3)\; 2\; f0;1g^2$ and for any proper interval N on the real line.

(i) If $g = \;{}_0$, then $P(\tilde{z}^{0} = \tilde{z}^{0}\;{}_0|\;) = P(\tilde{x}^0_{1(12)}(\tilde{b}\;\;\tilde{\;}_0) = 0|\;) < 1$ by DP5. For the case with $g \neq \;{}_0$,

$$P(\tilde{z}^{0} = \tilde{z}^{0}\;{}_0|\;) = \sum_{d_0 2 f\, 0;1 g} \int P((g\;\;{}_0)y_{13} = (\;{}_0\;\; g)d_0 + \tilde{x}^0_{1(12)}(\tilde{\;}_0\;\;\tilde{b})|y_{10} = d_0;\tilde{x}_{1(12)};\;)\; P(y_{10} = d_0|\tilde{x}_{1(12)};\;)\; dF_{\tilde{x}_{1(12)}|}$$

By Assumption DP3, $P((g\;\;{}_0)y_{13} = (\;{}_0\;\; g)d_0 + \tilde{x}^0_{1(12)}(\tilde{\;}_0\;\;\tilde{b})|y_{10} = d_0;\tilde{x}_{1(12)};\;) < 1$ for all $d_0\; 2\; f0;1g$, g= d0025 d Tf 53.01 Tf 6.083 -1.794 Td [(0252 h/F22.9552 Tf 12.425 0 Td [(d)]TJ/F70 7.9701 Tf 6.083 -1.794 Td [(0252 $\tilde{x}_{1(12)} < 1$

Note that the expectation above is strictly positive for almost all z since $P(A|x; ;\ )$ $P(B|x; ;\ ) = 0$ if and only if $\mathrm{sgn}(z^{0}\ _{0}) = 0$ which is an event having zero probability measure under Assumption DP4. It then follows from Lemma A.3 and Assumption DP6 that $Q(\ _{0})\quad Q(\ ) > 0$ for all $\ 2\quad n\{\ _{0}\}$. □

To simplify notation, we de ne

$$Q_n(\ ) = \frac{1}{nh_n^{3p}} \sum_{i=1}^{n} K(\ x_i=h_n)\ _i(\ )$$

and

$$(\ ) = f(\ )E[\ (\ )|x_{2(12)} = x_{2(23)} = x_{1(23)} = \ ]$$

Proof of Theorem 3.3The proof proceeds by verifying the following conditions for Theorem 9.6.1 in Amemiya (1985): (C1) is a compact set, (C2)$Q_n(\ )$ is a measurable function for all $\ 2\ $, (C3) $Q_n(\ )$ converges in probability to a nonstochastic function $Q(\ )$ uniformly in $\ 2\ $, (C4) $Q(\ )$ is continuous in and is uniquely maximized at $\ _0$.

The compactness of is satis ed by construction. Condition (C2) holds trivially. Lemma B.3 above has shown that the identi cation condition in Tf 5.6-(.)]TJ -33ldsn

Nolan and Pollard (1987), F_h is Euclidean for the constant envelope $\sup_{v \in R^{3p}} |K(v)| < \infty$. Furthermore, as F is Euclidean for the constant envelope $\sup_{\in} |\;(\;)| = 1$ (see Example 2.11 in Pakes and Pollard (1989)), F is Euclidean for the constant envelope $\sup_{v \in R^{3k}} |K(v)| < \infty$. Next, note that by Assumptions DP6 and DP8(ii),

$$
\begin{aligned}
\sup_{F_n} E|K(\;x/h_n)\;(\;)| &= \sup_{F_n} \int [E|K(\;x/h_n)\;(\;)||\;x] f(\;x) d\;x \\
&= \sup_{F_n} h_n^{3p} \int K(v)[E|\;(\;)||\;x = vh_n] f(vh_n) dv \\
&\quad \sup_{F_n} h_n^{3p} \int K(v) f(vh_n) dv = O(h_n^{3p})
\end{aligned}
$$

Then, under Assumption DP9(ii), we obtain by applying Lemma 5 in Honoré and Kyriazidou (2000) that

$$
\sup_{F_n} h_n^{3p} |Q_n(\;) \quad EQ_n(\;)| = O_p\left(\sqrt{\frac{h_n^{3p}\log n}{n}}\right) = o_p(h_n^{3p})
$$

Next, we show that $\sup_{\in} |EQ_n(\;) \quad Q(\;)| = o(1)$. Notice that by Assumptions DP7, DP8(ii), DP8(iii), and DP9(i),

$$
\begin{aligned}
\sup_{\in} |EQ_n(\;) \quad Q(\;)| &= \sup_{\in} \left|\frac{1}{h_n^{3p}} \int K(\;x/h_n)\;(\;x) d\;x \quad (0)\right| \\
&= \sup_{\in} \left|\frac{1}{h_n^{3p}} \int K(\;x/h_n)[\;(0) + \;^{(1)}(\;)^{0}\;x] d\;x \quad (0)\right| \\
&= \sup_{\in} \left|\;(0) \int K(v)dv + h_n \int K(v)\;^{(1)}(v_n)^{0} v dv \quad (0)\right| \\
&= \sup_{\in} \left|h_n \int K(v)\;^{(1)}(v_n)^{0} v dv\right| \\
&\quad h_n \int K(v)|\;^{(1)}(v_n)|_1 |v|_1 dv \\
&= O(h_n) = o(1)
\end{aligned}
$$

where $|\;|_1$ denotes the l_1 norm of a vector. Therefore,

$$
\sup_{\in} |Q_n(\;) \quad Q(\;)| \quad \sup_{F_n} |Q_n(\;) \quad EQ_n(\;)| + \sup_{\in} |EQ_n(\;) \quad Q(\;)| = o_p(1)
$$

and the desired result follows. □