Semiparametric Identification and Estimation of Multinomial Discrete Choice Models using Error Symmetry

Arthur Lewbel[†] Jin Yan[‡] Yu Zhou[§]

Original February 2019, revised December 2021

Abstract

We provide a new method to point identify and estimate cross-sectional multinomial choice models, using conditional error symmetry. Our model nests common random coefficient specifications (without having to specify which regressors have random coefficients), and more generally allows for arbitrary heteroskedasticity on most regressors, unknown error distribution, and does not require a "large support" (such as identification at infinity) assumption. We propose an estimator that minimizes the squared differences of the estimated error density at pairs of symmetric points about the origin. Our estimator is root N consistent and asymptotically normal, making statistical inference straightforward.

1 Introduction

Traditional multinomial choice models, such as multinomial logit (MNL) and multinomial probit (MNP), e.g., McFadden (1974), assume homoskedastic errors. However, in reality substantial

We thank Songnian Chen, Jeremy Fox, Bruce Hansen, Shakeeb Khan, Brendan Kline, Rosa Matzkin, Byoung Park, Jack Porter, Andres Santos, Xiaoxia Shi, Matthew Shum, Christopher Taber, Yuanyuan Wan and Hanghui

unobserved heterogeneity is common, e.g., Heckman (2001). We provide a new method to point identify preference parameters in cross-sectional multinomial choice models in the presence of general unobserved individual heterogeneity. Our identi…cation is semiparametric, in that we do not specify the joint distribution of the latent errors, and we allow for arbitrary heteroskedasticity with respect to most regressors, including possible random coe¢ cients. We propose a corresponding and Tang (2016) in contrast ...nd that symmetry, when combined with conditional independence

Using the analogy principle, we construct a corresponding estimator that minimizes the squared di¤erences of the estimated error densities at each data point with its corresponding symmetry point. We show this minimum distance estimator is root N consistent and asymptotically normal. Computing the objective function of our estimator does not entail either numerical integration or deconvolution techniques, which are often required by random coe¢ cients models. Moreover, our estimator does not require specifying which covariates, if any, have random coe¢ cients, and is no more or less complicated regardless of how many covariates have random coe¢ cients, or any other more complicated forms of heteroskedasticity.

Many methods have been developed for identifying and estimating utility function parameters with cross-sectional multinomial choice data. Many of those methods assume independence between the covariates and error terms, ruling out the possibility of individual heterogeneity such as random coe¢ cients (Ruuluydence the general multinomial choice case, we show root-N consistency and asymptotic normality of our estimator, and we provide proofs for all of our theorems.

2 The Model and Identification

2.1 The Random Utility Framework

To simplify notation and presentation of our results, for the main text of this paper we restrict attention to the case of three choices, with the relative utility of the outside option, denoted $j = 0$, normalized to equal zero. General results for an arbitrary number of multinomial choices, and single outcome like y_0, because the choice of any one outcome depends on the utilities of all of the outcomes.

an absolutely continuous density function, $f_{"_1"_2}(t_1;t_2\ X)$, which is centrally symmetric about the origin, i.e.,

$$f_{"_1"_2}(t_1;t_2\ X) = f_{"_1"_2}(\quad t_1;\quad t_2\ X);$$

for any vector $(t_1;t_2)$ $_{12}$

I1, this yields the equations

$$\frac{@^2E(y_0 \quad z = z \; ; X = X \;)}{@z_1 @z_2} = \frac{@^2 \Pr(y_0 = 1 \quad z = z \; ; X = X \;)}{@z_1 @z_2} \tag{4}$$

$$= f_{"_1"_2} \quad z_1 \quad x_1^0 \;{}^o ; \quad z_2 \quad x_2^0 \;{}^o \quad X = X \qquad (\quad 1)^2 ;$$

and

$$\frac{@^2E(y_0 \quad z = \quad z \quad 2X \quad ; X = X \;)}{@z_1 @z_2} = \frac{@^2 \Pr(y_0 = 1 \quad z = \quad z \quad 2X \quad ; X = X \;)}{@z_1 @z_2} \tag{5}$$

$$= f_{"_1"_2} \; z_1 + 2x_1^0 \quad x_1^0 \;{}^o ; z_2 + 2x_2^0 \quad x_2^0 \;{}^o \quad X = X \qquad (\quad 1)^2 :$$

The left sides of equations (4) and (5) are both identi…ed, and can be estimated as nonparametric regression derivatives, given a value of . If $=$ ^{o}, then by the symmetry Assumption I2, the right sides of equations (4) and (5) are equal to each other. De…ne the function $d_0(\; ; z \; ; X \;)$ as have positive probability measure for any in the parameter space other than ^{o}. Assumptions I4 and I5 give one set of conditions that su¢ ce.Assumption I4 provides a subset of the support of covariates with positive measure on which the function $d_0(\; ; z \; ; X \;)$ can be identi…ed, while Assumption I5 ensures that symmetry points are unique.

Given these assumptions we obtain identi…cation as follows. All proofs are in the Supplementary Appendix.

Theorem 2.1 If Assumption I hold, then the parameter vector ^{o} is point identi…ed by De…nition 2.1.

2.3.1 Discussion

Theorem 2.1 used expectations of y_0. Additional identifying information (resulting in more e¢ -cient associated estimators) can similarly be obtained from y_1 and y_2. Details are in the supplemental appendix.

The conditional independence between z and " in Assumption I1 is known as a distributional exclusion restriction (Powell, 1994, p. 2484). This allows for interpersonal heteroskedasticity on a subset of covariates: Higher moments of " can depend (in unknown ways) on X , but not z. Assumption I2 is our error symmetry restriction. Without loss of generality we assume that the point of symmetry is the origin, because any nonzero term could be absorbed into the intercept of the utility index as discussed in equation (1).

Assumption I3 assumes a compact parameter space, which is a standard assumption for many nonlinear models, including semiparametric multinomial discrete choice models. Assumption I4(a)

Assumption I5(a) ensures that the error density functions in (4) and (5) are evaluated at interior points of their support. Assumption I5(b) requires that the error density function has a unique (local) symmetry point over a subset of its support, $\mathcal{E}_{\varepsilon}(X)$. This does not rule out densities having flat sections, but it does limit the range of any such flat sections.

2.4 An Alternative Identification Strategy

Existing binary choice estimators that make use of latent error symmetry (e.g. Chen (2000) and Chen, Khan and Tang (2016) are based on the error distribution function rather than on the error density function as in Theorem 2.1. To illustrate, take a simple binary choice model where $y = I(z + a + v \geq 0)$. If v is a symmetric random variable around zero and $v \perp z$, then

$$E(y \mid z = c) = \Pr(-v \leq c + a) = \Pr(v \leq c + a) = E(1 - y \mid z = -c - 2a) \tag{8}$$

The constant a is identified by equating the above two expectations, which only requires estimation of the conditional mean of y and not its derivatives. This immediately extends to identification of covariate coefficients instead of just a constant.

We could have similarly based identification and estimation of our multinomial β on the distribution instead of the density of the errors, and thereby only required nonparametric regressions and not their derivatives for estimation. However, unlike the binary choice case, identification and estimation using the distribution instead of the density of the errors becomes complicated and clumsy in the multinomial setting. This is because, in the binary choice case, error symmetry just equates two conditional expectations, corresponding to two error intervals, while for multinomial choice, one must equate error rectangles.

To see the issue, begin again from equation (3). Let $[a; b]$ be a rectangle in the support of ε. Point $a = (a_1; a_2)$ is the lower left vertex of this rectangle and $b = (b_1; b_2)$ is the upper right vertex. By central symmetry, the probability of ε being in the rectangle $[a; b] = [a_1; b_1] \times [a_2; b_2]$ is the same as the probability of ε being in the rectangle $[-b; -a] = [-b_1; -a_1] \times [-b_2; -a_2]$. This then implies

$$\int_{[a,b]} f_{\varepsilon_1\varepsilon_2}(t_1,t_2 \mid X)\,dt = \int_{[a,b]} f_{\varepsilon_1\varepsilon_2}(-t_1,-t_2 \mid X)\,dt = \int_{[-b,-a]} f_{\varepsilon_1\varepsilon_2}(t_1,t_2 \mid X)\,dt, \qquad (9)$$

where the first equality in (9) holds by Assumption A2 and the second one holds by changing of variables.[7]

The integrals on both sides of equation (9) can be computed using the conditional distribution function of ε, which in turn is obtained from the conditional expectation of y_0. For example, consider the left-hand side integral:

$$\begin{aligned}
\textstyle\int_{[a,b]} f_{\varepsilon}(t \mid X)\,dt &= \Pr(a \le \varepsilon \le b \mid X) \\
&= \Pr(a_1 \le \varepsilon_1 \le b_1, a_2 \le \varepsilon_2 \le b_2 \mid X) \\
&= \Pr(\varepsilon_1 \le b_1, \varepsilon_2 \le b_2 \mid X) - \Pr(\varepsilon_1 < a_1, \varepsilon_2 \le b_2 \mid X) \\
&\quad - \Pr(\varepsilon_1 \le b_1, \varepsilon_2 < a_2 \mid X) + \Pr(\varepsilon_1 < a_1, \varepsilon_2 < a_2 \mid X).
\end{aligned} \qquad (10)$$

We prefer to identify and estimate by matching each point in the data using densities, rather than by matching rectangles using distributions, for many reasons. First, equating error distribution rectangles involves more tuning parameters, since rectangles need to be chosen. Second, matching densities only requires finding enough points $(z = z; X = X)$ in the data that have matches $(z = z\ 2X; X = X)$ that lie in the support of the covariates. In contrast, each matching rectangle requires finding an entire range of covariates that lie in the support and has a range of matches that also lie entirely in the support. Third, to gain efficiency we will later create more moments by replacing y_0 with different choices y_j. When matching density points, the same covariate values (points) that work for any one choice j will also work for any other choice. The same is not true for matching distribution rectangles, because for rectangles each match entails pairs of observations rather than individual observations. Finally, the computation cost of estimation is lower for equating error densities than for distribution rectangles. For a sample of size N, we compute error densities at $2N$ points, while in contrast, using rectangles would entail computing the error distribution at $N(N\ 1)2^J$ points.

3 A Minimum Distance Estimator and its Asymptotic Properties

3.1 Population Objective Functions for Estimation

Given the identification strategy described in Section 2, we develop a minimum distance estimator (hereafter, MD estimator) for ^{o} using the identifying restriction $d_0(^o; z; X) = 0$, where d_0 is defined by equation (6). Note that the function $d_0(; z; X)$ is well defined if both points $(z; X)$ and $(z\ 2X; X)$ are in the interior of the support of covariates, $_{(z;X)}$. For this reason, we only wish to evaluate the function $d_0(; z; X)$

the range of these values. Define functions $\psi_0(\cdot)$ and $\varphi_0(\cdot)$ by

$$\psi_0\left(z;X;\beta;\gamma\right) \equiv \varphi_0(z;X)\,\varphi_0\left(z - 2X\beta;X\right)\,\varphi_0\left(z - 2X\gamma;X\right); \tag{11}$$

and

$$\varphi_0(z;X) \equiv 1\left(|z| \le c_1\right) \cdot 1\left(|X| \le C_2\right). \tag{12}$$

Here the absolute value of a vector or matrix, $|\cdot|$, is defined as the corresponding vector or matrix

3.2 An Estimator

We now provide an estimator for function $d_0(\ ; z_n; X_n)$ in (13) as

$$\hat{d}_{0;\ n}(\ ; z_n; X_n) \quad \hat{\varphi}^{(2)}_{o;\ n}(z_n; X_n) \quad \hat{\varphi}^{(2)}_{cs;\ n}(z_n; X_n;\ ): \tag{14}$$

where $\hat{\varphi}^{(2)}_{o;\ n}(z_n; X_n)$ and $\hat{\varphi}^{(2)}_{cs;\ n}(z_n; X_n;\ )$ are leave-one-out, Nadaraya-Watson nonparametric regression kernel estimators for the derivatives on the right hand side of equation (6) (see the supplemental appendix for details). By replacing the expectation in $Q_0(\ )$ with its sample mean and replacing the function $d_0(\ ; z_n; X_n)$

$Q_{Nj}(\ )$ over each choice. We also extend all our results to multnomial choice with an arbitrary number of choices, instead of just three as above.

4 Monte Carlo Experiments

In this section, we use Monte Carlo experiments to study the …nite-sample properties of the minimum distance (MD) estimator proposed above. We consider four data generating processes (DGPs). In each DGP, individual n's utility from alternative j, u_{nj}, is speci…ed as

$$u_{nj} = z_{nj} + x_{nj}\ _{n} + "_{nj} \quad \text{for } n = 1; 2; :::; N \text{ and } j = 0; 1; 2: \tag{17}$$

$_{n}{}^{n}$

Table 1: Monte Carlo Results of estimating ^{o} (True Parameter $^{o} = 0.2$)

DGP	N	MNP		MD (y_0)		MD (y_0, y_1, y_2)	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
1	1000	-0.0012	0.0435	0.0216	0.2368	-0.0017	0.1337
	2000	-0.0010	0.0307	0.0055	0.1355	-0.0078	0.0788
2	1000	0.5656	0.5833	0.1047	0.3521	-0.0392	0.3048
	2000	0.5627	0.5714	0.0543	0.2308	-0.0289	0.1747
3	1000	-0.0013	0.0454	0.0317	0.2220	0.0015	0.1417
	2000	-0.0017	0.0319	0.0158	0.1301	-0.0051	0.0812
4	1000	-0.7512	0.7718	-0.0054	0.3765	-0.0748	0.3550
	2000	-0.7481	0.7585	0.0180	0.2616	-0.0343	0.2149

covariates. Under all four DGP's our MD estimator remains consistent.

Table 1 reports the bias and root mean square error (RMSE) of each estimator in our simulations. The first set of columns reports the MNP estimator, the second reports our MD estimator using only y_0, while the third uses observations of all choices y_0, y_1, and y_2 (MNP also uses observations of all choices).

Under DGP 1, the MD estimators have small finite sample bias, and RMSEs two to four times larger than that of the correctly specified efficient MNP estimator. Under DGP 2, the bias of the misspecified MNP estimator is around three times the true parameter value, and this bias remains as the sample size is doubled. In contrast, the bias and RMSE of the MD estimators are much smaller than the MNP estimator, and they decrease sharply as the sample size increases. In DGP 3, the random coefficients MNP is correctly specified, and so performs better than the MD estimators in terms of bias and RMSE. However, in DGP 4 where the random component is heterogeneous, the bias of MNP is almost four times the true parameter value and does not vanish as sample size grows. In contrast, the bias of the MD estimators is still relatively small.[9] In all the DGPs, in terms of RMSE, the MD estimator using y_0, y_1, and y_2 performs better than

[9] We speculate that the bias in the MD estimators might be further reduced by a bandwidth search, and/or using local linear estimation for the first stage choice probabilities.

References

[1] Ahn, H., Ichimura, H., Powell, J.L., and Ruud, P. (2018): “Simple Estimators for Invertible Index Models,” Journal of Business and Economic Statistics, 36, 1-10.

[2] Berry, S. and Haile P.A. (2010): “Nonparametric Identi…cation of Multinomial Choice Demand Models with Heterogeneous Consumers,” Cowles Foundation Discussion Paper #1718.

[3] Berry, S., Levinsohn, J., and Pakes, A. (1995): “Automobile Prices in Market,” Econometrica, 63, 841-890.

[4] Berry, S., Levinsohn, J., and Pakes, A. (2004): “Di¤erentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market,” Journal of Political Economy, 112(1), 68-105.

[5] Blundell, R. and Powel, J.L. (2004): “Endogeneity in Semiparametric Binary Response Models,” Review of Economic Studies, 71, 655–679.

[6] Chen, S. (2000): “E¢ cient Estimation of Binary Choice Models under Symmetry,” Journal of Econometrics, 96, 183-199.

[7] Chen, S. and Z8):S.8(Y80.004(an7)-334(In)28.018en,)-(onomicor)-334282.9ametric Ztudmation o(79n)-35(45

[10] Delaigle, A. and Hall, P. (2016): "Methodology for Non-parametric Deconvolution When the Error Distribution is Unknown," Journal of the Royal Statistical Society, Series B, 78, 231–252.

[11] Dong, Y. and Lewbel, A. (2011): "Nonparametric Identification of a Binary Random Factor in Cross Section Data," Journal of Econometrics, 163, 163-171.

[12] Fox, J.T. (2007): "Semiparametric Estimation of Multinomial Discrete-choice Models Using a Subset of Choices," RAND Journal of Economics, 38, 1002-1019.

[13] Fox, J.T. and Gandhi, A. (2016): "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," RAND Journal of Economics, 47, 118-139.

[14] Goolsbee, A. and Petrin. A (2004): "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," Econometrica, 72, 351-381.

[15] Hausman, J.A. and Wise, D.A. (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, 46, 403-426.

[16] Heckman, J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture" Journal of Political Economy, 109, 673-748.

[17] Honoré, B. E., Kyriazidou, E., and Udry, C. (1997): "Estimation of Type 3 Tobit Models using Symmetric Trimming and Pairwise Comparisons," Journal of Econometrics, 76, 107-128.

[18] Khan, S., Ouyang, F., and Tamer, E. (2019): " Inference in Semiparametric Multinomial Response Models," Working Paper.

[19] Kumbhakar, S.C. and Lovell, C.A.K. (2000): Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

[20] Lee, L. (1995): "Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models," Journal of Econometrics, 65, 381-428.

[21] Lewbel, A. (1997): "Constructing Instruments for Regressions With Measurement Error When no Additional Data are Available, with An Application to Patents and R&D," Econometrica, 65, 1201-1213.

[22] Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables," Journal of Econometrics, 97, 145-177.

[23] Manski, C.F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics, 3, 205-228.

[24] Manski CF. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," Journal of Econometrics, 27, 313-333.

[25] Manski, C.F. (1988): "Identification of Binary Response Models," Journal of the American Statistical Association, 83, 729-738.

[26] McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in Fron-

[31] Powell, J.L. and Ruud, P.A. (2008): "Simple Estimators for Semiparametric Multinomial Choice Models," Working Paper.

[32] Ruud, P.A. (1986): "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," Journal of Econometrics, 32, 157-187. Cambridge University Press.

[33] Serfling, R. (2006): "Multivariate Symmetry and Asymmetry," Encyclopedia of Statistical Sciences. John Wiley & Sons, Inc.

[34] Shi, X., Shum, M., and Song, W. (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," Econometrica, 86, 737-761.

[35] Train, K. (2009): Discrete Choice Methods with Simulation.

[36] Yan, J. (2013): "A Smoothed Maximum Score Estimator for Multinomial Discrete Choice Models," Working Paper.

[37] Yan, J. and Yoo, H. (2019): "Semiparametric Estimation of the Random Utility Model with Rank-ordered Choice Data," Journal of Econometrics, 211, 414-438.

[38] Zhou, Y. (2021): "Identification and Estimation of of Entry Games under the Symmetry of Unobservables," Working Paper, NYU, Shanghai.

Supplementary Appendix: Semiparametric Identi…cation and Estimation of Multinomial Discrete Choice Models using Error Symmetry

Arthur Lewbel[y] Jin Yan[z] Yu Zhou[x]

Original February 2019, revised December 2021

we only require that differences z and X be observed, and regularity conditions (e.g., continuity of z) are only be imposed on z_j and x_j, not on $\bar{z}_j$ and $\bar{x}_j$. In addition to these covariates, our identification only requires that y_0 be observed, not the entire vector of outcomes y. This is possible because z provides information about the other outcomes. Nevertheless, the associated estimators will be more efficient by observing and making use of more the elements of y, since each additional outcome y_j one observes provides additional overidentifying information.

Assumption I1 immediately implies

$$\Pr(y_0 = 1 \mid z; X) = F_{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_J}\left(z_1 \;\; x_1^{0} \;\; {}^{o}; \; z_2 \;\; x_2^{0} \;\; {}^{o}; \ldots; \; z_J \;\; x_J^{0} \;\; {}^{o} \mid z; X \right) \tag{S.A.5}$$

$$= F_{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_J}\left(z_1 \;\; x_1^{0} \;\; {}^{o}; \; z \right.$$

Based on Assumptions I1 and I2, we have that if $\theta = \theta^o$, then $d_0(\theta; z; X) = 0$. Given some regularity conditions, setting the function d_0 equal to zero at a collection of values of z and X provides enough equations to point identify θ^o. The proof of Theorem S.A.1 is provided in Section S.C.

Theorem S.A.1 *If Assumptions* I *hold, then the parameter vector* $\theta^o \in \Theta$ *is point identified by Definition* 1.

S.A.2 Identification Using Multiple Choices

In Section A.1, we identified the parameter vector θ^o using only derivatives of the conditional mean of y_0. Here we illustrate that identification can be achieved using the conditional mean of y_j for any $j \in J$. Later we will increase efficiency of estimation by combining the identifying moments based on each of the observed choices y_j.

We now introduce some additional notation. For each $j \in J$, define $X^{(j)}$ as the matrix that consists of differenced covariate vectors $x_k - x_j$ for all $k \in J$ and $k \neq j$. For example, when $1 < j < J$, $X^{(j)} \equiv (x_0 - x_j; \ldots; x_{j-1} - x_j; x_{j+1} - x_j; \ldots; x_J - x_j)' \in \mathbb{R}^{J \times q}$. By this notation, we have $X^{(0)} \equiv (x_1 - x_0; \ldots; x_J - x_0) = X$. In the same fashion, define $z^{(j)} \in \mathbb{R}^J$ as the vector

Proposition S.A.2 *If Assumption I2 holds, then for every* $j \in J$ *and almost every* $X^{(j)} \in S_{X^{(j)}}$, *the conditional distribution function* $F_{\varepsilon^{(j)}}(t^{(j)} \mid X^{(j)})$ *admits an absolutely continuous density function,* $f_{\varepsilon^{(j)}}(t^{(j)} \mid X^{(j)})$*, which is centrally symmetric about the origin*, *i.e.*,

$$f_{\varepsilon^{(j)}}(t^{(j)} \mid X^{(j)}) = f_{\varepsilon^{(j)}}(-t^{(j)} \mid X^{(j)}); \qquad \text{(S.A.13)}$$

for any vector $t^{(j)} \in S_{\varepsilon^{(j)}}(X^{(j)})$ *where* $S_{\varepsilon^{(j)}}(X^{(j)}) \subseteq \mathbb{R}^J$.

of the left-hand sides of (S.A.15) and (S.A.16), that is,

$$d_j(\theta; z^{(j)}, X^{(j)}) \equiv \partial^J E(y_j \mid z^{(j)} = z^{(j)}, X^{(j)} = X^{(j)})/\partial z_1^{(j)} \cdots \partial z_J^{(j)} \tag{S.A.17}$$
$$\qquad - \partial^J E(y_j \mid z^{(j)} = -z^{(j)} - 2X^{(j)}\theta, X^{(j)} = X^{(j)})/\partial z_1^{(j)} \cdots \partial z_J^{(j)},$$

which always equals zero when $\theta = \theta^o$ and may be non-zero when $\theta \neq \theta^o$.

Then, analogous to Definition 1, define

$$D_j(\theta) \equiv \Big\{ (z^{(j)}, X^{(j)}) \in \text{int}\left(S_{(z^{(j)}, X^{(j)})}\right) \cap (-z^{(j)} - 2X^{(j)}\theta, X^{(j)}) \in \text{int}\left(S_{(z^{(j)}, X^{(j)})}\right) : d_j(\theta; z^{(j)}, X^{(j)}) \neq 0 \Big\}. \tag{S.A.18}$$

Recall that there is a one-to-one correspondence, respectively, between $X^{(j)}$ and X, $z^{(j)}$ and z, and $\varepsilon^{(j)}$ and ε. For every $(z, X) \in \text{int}(S_{(z,X)})$ such that $(-z - 2X\theta, X) \in \text{int}(S_{(z,X)})$, we immediately have $(z^{(j)}, X^{(j)}) \in \text{int}(S_{(z^{(j)}, X^{(j)})})$ and $(-z^{(j)} - 2X^{(j)}\theta, X^{(j)}) \in \text{int}(S_{(z^{(j)}, X^{(j)})})$, as well as $d_j(\theta; z^{(j)}, X^{(j)}) = 0$ if and only if $d_j(\theta; z, X) = 0$. Therefore, we can also use the choice probability of any alternative in the choice set to achieve identification.

S.A.3 Individual Heterogeneity and Random Coefficient

Our identifying assumptions do not refer specifically to random coefficients. Here we provide sufficient conditions for our key identification assumptions I1 and I2 to hold when unobserved

$_n$ = $_n$ [illegible] Our symmetry assumption implies that ^{o} will also be the mean coe¢ cients, as long as these [illegible] exist, but we dont impose this existence.[2]

We can rewr[illegible] utility function (S.A.19) as $u_{nj} = z_{nj} + x_{nj}^0 \; ^o + "_{nj}$ where $"_{nj} = (\; _{nj} + x_{nj}^0 \; _n)$ for $j = 1; : : : ; J$. [illegible] $_n = ("_{n1}; : : : ; "_{nJ})$ is often called the composite error in the presence of random coe¢ cients [illegible] Theorem 2.1, if this composite error vector $"_n$ satis…es Assumptions I1 (Exclusion Restrictio[illegible] 2 (Central Symmetry), then ^{o} is point identi…ed under the regularity conditions given by [illegible]ptions I3-I5. We now give su¢ cient conditions for I1 and I2 to hold with random coe¢ ci[illegible]

Assumption RC [illegible]

RC1: Condi[illegible]n almost every $X \; 2 \; S_X$, the covariate vector z is independent of $(\; ; \;)$; ^{o} and the co[illegible]al distribution function of

to know exactly which covariates have random coefficients and which do not. Last, our model does not require thin tails or unimodality, unlike, e.g., normal random coefficient MNP models.[4]

One restriction we do impose is that we require one covariate in each choice, z_j, not have a random coefficient. Setting the coefficient of some covariate z equal to one is often a natural, economically meaningful normalization. For example, utility of choices are typically modeled as benefits minus costs. Benefits may be subjective and so vary heterogeneously as in random coefficients, while costs are often objective and fixed. In these cases z would be a cost measure. Examples are willingness to pay studies where the benefits equal the willingness to pay, and consumer choice applications where z_j is the price of choice j. (See e.g., Bliemer and Rose 2013 for more discussion and examples.[5]) Nevertheless, we could also assume that, before normalizing, the variable z has a random coefficient, provided that the random coefficient is the same for all choices and is positive (this latter restriction is a special case of the hemisphere condition required by semiparametric binary choice random coefficient estimators. See, e.g., Gautier and Kitamura 2013). This restriction is needed because we can't allow renormalizations that would change any individual's relative ranking of utilities. Note that in this case, we require our symmetry condition to hold after renormalization, not before.

for $j = 0; \ldots; J$, where d_j is defined as the same as equation (S.A.17).

For each j, the function $d_j(\ ; z^{(j)}; X^{(j)})$ is well defined if both points $(z^{(j)}; X^{(j)})$ and $(z^{(j)} \quad 2X^{(j)} \ ; X^{(j)})$ are in the interior of the support of covariates, $S_{(z^{(j)}; X^{(j)})}$. For this reason, we only wish to evaluate the function $d_j(\ ; z^{(j)}; X^{(j)})$ at such points. This can be achieved by multiplying each function $d_j(\ ; z^{(j)}; X^{(j)})$ by a trimming function of the form

$$_j\left(z^{(j)}; X^{(j)}; \bar{\ }; \underline{\ }\right) \quad \&_j\left(z^{(j)}; X^{(j)}\right) \&_j\left(z^{(j)} \quad 2X^{(j)}\bar{\ }; X^{(j)}\right) \&_j\left(z^{(j)} \quad 2X^{(j)}\underline{\ }; X^{(j)}\right);$$

where $X^{(j)}\bar{\ }$ ($X^{(j)}\underline{\ }$) gives the upper (lower) bound value that the index $X^{(j)}$ can take. A simple choice for the function $\&_j(\ )$ is $\&_j\left(z^{(j)}; X^{(j)}\right) \quad 1 \ |z^{(j)}| \quad c_1^{(j)} \quad 1 \ |X^{(j)}| \quad C_2^{(j)}$, where the absolute value of a vector or matrix, $|\ |$, is defined as the corresponding vector or matrix of the absolute values of each element, $c_1^{(j)} \in R^J$ is a vector of trimming constants for the covariate vector $z^{(j)}$, and $C_2^{(j)} \in R^{J \ q}$ is a matrix of trimming constants for the covariate matrix $X^{(j)}$ such that $\left(c_1^{(j)}; C_2^{(j)}\right)$ is in the interior of the support of covariates $S_{(z^{(j)}; X^{(j)})}$. Denote $S^{Tr}_{z^{(j)}}\left(X^{(j)}; \bar{\ }; \underline{\ }\right)$ as the largest set of values $z^{(j)}$ given $\bar{\ }$, $\underline{\ }$, and $X^{(j)}$, such that $S^{Tr}_{z^{(j)}}\left(X^{(j)}; \bar{\ }; \underline{\ }\right) \quad \text{int}\left(S_{z^{(j)}}\left(X^{(j)}\right)\right)$. We describe the regularity conditions on the trimming function in Assumption TR.

Assumption TR. The trimming function $_j\left(z^{(j)}; X^{(j)}; \bar{\ }; \underline{\ }\right)$ is strictly positive and bounded on $S^{Tr}_{z^{(j)}}\left(X^{(j)}; \bar{\ }; \underline{\ }\right) \quad \text{int}\left(S_{X^{(j)}}\right)$, and is equal to zero on its complementary set for $j = 0; \ldots; J$.

Theorem S.B.1 *If Assumptions I and TR hold, then (i)* $Q_j(\ ) \quad 0$ *for any* $\quad \in \quad$ *and (ii)* $Q_j(\ ) = 0$ *if and only if* $\ = \ ^o$.

Theorem S.B.1 shows identification based on the population objective function. Proofs is available at authors'webpage.

S.B.2 MD Estimator and Regularity Conditions

Next, we derive the sample objective function based on population objective function and the asymptotic properties of the MD estimator. To ease notation, we denote the conditional means

$$E\left[y_j \mid z^{(j)} = z_n^{(j)}; X^{(j)} = X_n^{(j)} \right] \quad \varphi_j\left(z_n^{(j)}; X_n^{(j)} \right) \quad \varphi_{j;o}\left(z_n^{(j)}; X_n^{(j)} \right);$$

$$E\left[y_j \mid z^{(j)} = z_n^{(j)} \quad 2X_n^{(j)}; X^{(j)} = X_n^{(j)} \right] \quad \varphi_j\left(z_n^{(j)} \quad 2X_n^{(j)}; X_n^{(j)} \right) \quad \varphi_{j;cs}\left(z_n^{(j)}; X_n^{(j)}; \quad \right);$$

and function

$$d_j\left(\; ; z_n^{(j)}; X_n^{(j)} \right) \quad \frac{\partial E\left[y_j \mid z^{(j)} = z_n^{(j)}; X^{(j)} = X_n^{(j)} \right]}{\partial z_1^{(j)} \quad \partial z_J^{(j)}} \quad \frac{\partial E\left[y_j \mid z^{(j)} = z_n^{(j)} \quad 2X_n^{(j)}; X^{(j)} = X_n^{(j)} \right]}{\partial z_1^{(j)} \quad \partial z_J^{(j)}} \tag{S.B.2}$$

$$\varphi_{j;o}^{(J)}\left(z_n^{(j)}; X_n^{(j)} \right) \quad \varphi_{j;cs}^{(J)}\left(z_n^{(j)}; X_n^{(j)}; \quad \right):$$

where $\varphi_{j;o}^{(J)}\left(z_n^{(j)}; X_n^{(j)} \right) \quad \partial^J \varphi_{j;o}\left(z_n^{(j)}; X_n^{(j)} \right) = \partial z_1^{(j)} \quad \partial z_J^{(j)}$ and $\varphi_{j;cs}^{(J)}\left(z_n^{(j)}; X_n^{(j)}; \quad \right)$ is defined in the similar way as $\varphi_{j;o}^{(J)}\left(z_n^{(j)}; X_n^{(j)} \right)$. Now, consider a leave-one-out (LOO) Nadaraya-Watson (NW) estimator for $\varphi_{j;o;\ n}^{(J)}$ as $\hat{\varphi}_{j;o;\ n}^{(J)}(\; ; \;) = \frac{\frac{1}{N\ 1} \sum_{m=1;m\neq n}^{N} y_{mj} K_{h_z}^{(J)}\left(z_m^{(j)} \right) K_{h_X}\left(X_m^{(j)} \right)}{\frac{1}{N\ 1} \sum_{m=1;m\neq n}^{N} K_{h_z}\left(z_m^{(j)} \right) K_{h_X}\left(X_m^{(j)} \right)}$, where

$K_{h_z}\left(z_m^{(j)} \right) = \prod_{l=1}^{J} h_{z_l}^{1} k\left(h_{z_l}^{1} \quad z_{ml}^{(j)} \right)$; and $K_{h_X}\left(X_m^{(j)} \right) = \prod_{l=1}^{J} \prod_{r=1}^{q} h_{x_{lr}}^{1} k\left(h_{x_{lr}}^{1} \quad x_{mlr}^{(j)} \right)$.

The properties of the kernel function k and those of the bandwidth $h_z \quad (h_{z_1}; \quad ; h_{z_J})^0$ and h_X ec-11s/T-2 /TT52z

By replacing the expectation in $Q_j(\theta)$ with its sample mean and function $d_j(\theta; z_n^{(j)}; X_n^{(j)})$ with its LOO estimator $\hat{d}_{j;-n}\left(\theta; z_n^{(j)}; X_n^{(j)}\right)$, we define the MD estimator

$$\hat{\theta} \in \arg\min_{\theta \in \Theta} Q_{Nj}(\theta);$$

$$\text{where} \quad Q_{Nj}(\theta) = \frac{1}{2N}\sum_{n=1}^{N}\left[\omega_j\left(z_n^{(j)}; X_n^{(j)}\right)\hat{d}_{j;-n}\left(\theta; z_n^{(j)}; X_n^{(j)}\right)\right]^2.$$

We denote the gradient of the objective function as $q_{Nj}(\theta) = \nabla_\theta Q_{Nj}(\theta)$ and the Hessian matrix of the objective function as $H_{Nj}(\theta) = \nabla_{\theta\theta'} Q_{Nj}(\theta)$. The smoothness of the objective function suggests the first-order condition (FOC): $q_{Nj}\left(\hat{\theta}\right) = 0_q$. Applying the standard first-order Taylor expansion to $q_{Nj}\left(\hat{\theta}\right)$ around the true parameter vector θ^o yields $q_{Nj}\left(\hat{\theta}\right) = q_{Nj}(\theta^o) + H_{Nj}\left(\tilde{\theta}\right)\left(\hat{\theta} - \theta^o\right)$, where $\tilde{\theta}$ is a vector between the MD estimator $\hat{\theta}$ and the true parameter vector θ^o. Then the influence function will be given by

$$\hat{\theta} - \theta^o = -\left[H_{Nj}\left(\tilde{\theta}\right)\right]^{-1} q_{Nj}(\theta^o). \tag{S.B.4}$$

We will show that $H_{Nj}\left(\tilde{\theta}\right) \to_p H_j(\theta^o)$, where

$$H_j(\theta^o) = E\left\{\omega_j^2\left(z_n^{(j)}; X_n^{(j)}\right)\nabla_\theta d_j\left(\theta^o; z_n^{(j)}; X_n^{(j)}\right)\left[\nabla_\theta d_j\left(\theta^o; z_n^{(j)}; X_n^{(j)}\right)\right]'\right\}, \tag{S.B.5}$$

and $\sqrt{N} q_{Nj}(\theta^o) \to_d N(0_q; \Omega_j)$, where Ω_j is the probability limit of the variance-covariance matrix of $q_{Nj}(\theta^o)$. To obtain these properties, we assume the following regularity conditions.

Assumption E.

E1: $\{(y_n; z_n; X_n)$, for $n = 1; \ldots; N\}$ is a random sample drawn from the infinite population distribution.

E2: The following smoothness conditions hold: (a) The density function $f_j\left(z^{(j)}; X^{(j)}\right)$ is continuous in the components of $z^{(j)}$ for all $z^{(j)} \in S_z^{Tr}\left(X^{(j)}; \beta; \lambda\right)$ and $X^{(j)} \in \text{int}(S_X)$. In addition, $f_j\left(z^{(j)}; X^{(j)}\right)$ is bounded away from zero uniformly over its support. (b) Functions $f_j\left(z^{(j)}; X^{(j)}\right)$, $g_j\left(z^{(j)}; X^{(j)}\right)$ and $\varphi_j\left(z^{(j)}; X^{(j)}\right)$ are s ($s \geq J + 1$) times continuously differentiable in the components of $z^{(j)}$ for all $z^{(j)} \in S_z^{Tr}\left(X^{(j)}; \beta; \lambda\right)$ and have bounded derivatives.

E3: The kernel function k is an l-th ($l \geq 1$) order bias-reducing kernel that satisfies (a) $k(u) = k(-u)$ for any u in the support of schitz conditions: for somem $z^{(j)};\cdot$

$$\frac{@^{2}\,'_{j;o}\left(z^{(j)}+t;\cdot\right)}{@z_{1}^{(j)}\;@z_{J}^{(j)}} \quad \frac{@^{2}\,'_{j;o}\left(z^{(j)};\cdot\right)}{@z_{1}^{(j)}\;@z_{J}^{(j)}} < m \quad z^{(j)};\cdot \;\; ktk \; ; \quad \frac{@^{2}\,'_{j;cs}\left(z^{(j)}+t;\cdot\right)}{@z_{1}^{(j)}\;@z_{J}^{(j)}} \quad @^{2}\,'_{j;cs}\left(z^{(j)};\cdot\right)$$

0.7 7019 -6.982 Td (1)Tj /TT12 7.97 Tf 9.995 2.542 Td ( )Tj /TT5 7.97 Tf [(@)-58.996(z)]TJ /TT10 5.978 Tf 16.265 4.24 Td (()Tj.424 -0.7

(b) (Asymptotic Normality) *The MD estimator is asymptotically normal, i.e.,*

$$\sqrt{N}\ \hat{}\ \ {}^{o}\ \to_{d} N\ \ 0_q; H_j^{1}\ \ {}_j H_j^{1}$$

where matrix ${}_j$ E $t_{nj} t_{nj}^{0}$ *and* H_j

would be equal, while under the alternative there must exist symmetric points where the densities are not equal. Also under the null, our estimator is consistent. So a test could be constructed based on the di¤erence in error density estimates at many symmetry points (other than those used for estimation), using our estimated parameters to construct symmetry points. More general speci…cation tests could also be constructed, using the fact that our parameters are over identi…ed when more than one choice j is observed.

S.C Proof of Identi…cation

Proof of Theorem S.A.1: First, we show that $D_0(\ ^o)$ is a set of measure zero. If not, assume that there is a point $(z\ ; X\ )$ in set $D_0(\ ^o)$. By de…nition in equation (8), both points $(z\ ; X\ )$ and $(\ z\ \ 2X\ \ ^o; X\ )$ are in set $\mathrm{int}(S_{(z\ ;X\ )})$. By Assumptions I1, I2, and equations (5)-(7), we have function

$$d_0(\ ^o; z\ ; X\ ) = (\ 1)^{\ J}[f_{"}(\ z\ \ X\ \ ^o j X\ = X\ )\ \ f_{"}(z\ + X\ \ ^o j X\ = X\ )] = 0\ ;$$

which is a contradiction with de…nition in equation (8).

Next, we prove that $\Pr[(z\ ; X\ ) 2 D_0(\ )] > 0$ for any $\ 6= \ ^o$, where $\ 2\ $ and parameter space $\ $ satis…es Assumption I3. Denote the set $X(\ )\ \ fX\ \ 2 S_X\ j X\ (\ \ \ ^o) 6= 0g$, which is a collection of covariate values at which $X\ \ 6= X\ \ ^o$. By Assumption I4(a) and the fact $\ \ ^o 6= 0_q$, $X(\ )$ is a subset in the support of S_X with positive measure, that is,

$$\Pr[X\ \ 2 X(\ )] > 0: \qquad \text{(S.C.1)}$$

Recall that we use X_c and X_d, respectively, to denote the continuous and discrete covariates in X. We de…ne the interior of the support of X as $\mathrm{int}(S_X)$
$(X_c; X_d) 2 S_{(X_c;X_d)} j X_c 2 \mathrm{int}(S_{X_c}(X_d)) ; X_d 2 S_{X_d}$. De…ne

$$\tilde{S}_{(z\ ;X\ )}(\ )\ \ \Big\{ (z\ ; X\ ) 2 S_{(z\ ;X\ )}\ \ z\ \ 2 \tilde{S}_z(X\ ); X\ \ 2 X(\ ) \setminus \mathrm{int}(S_X) \Big\}\ ; \qquad \text{(S.C.2)}$$

where $\tilde{S}_z(X)$ satisfies Assumption I4(c). By construction, set $\tilde{S}_{(z;X)}(\ )$ is a Lebesgue measurable subset of $int(S_{(z;X)})$. Next we construct a subset in the support of covariates $(z;X)$ as follows:

$$\tilde{D}_0(\ ) \quad \{ (z;X) \in \tilde{S}_{(z;X)}(

S.D Monte Carlo Details

As discussed in the paper, our Monte Carlo design includes 4 data generating processes (DGPs). Details of the distribution of each DGP are provided in Table 1.

Table 1: Designs of the Data Generating Processes (DGPs)

DGP	Distribution of $_n$	Distribution of $"_{nj}$
1	$_n = 0{:}2$	$"_{nj} = {}_{nj}$
2	$_n = 0{:}2$	$"_{nj} = \frac{1}{2} e^{2x_{nj}} {}_{nj}$,
3	$_n = 0{:}2 + {}_n$ where $_n = \frac{1}{2}\#_n$	$"_{nj} = \frac{1}{2} {}_{nj}$
4	$_n = 0{:}2 + {}_n$ where $_n = (e^{x_{n1}} + e^{x_{n2}}) \; \#_n$	$"_{nj} = \frac{1}{2} {}_{nj}$

Note: both $\#_n$ and $_{nj}$ are standard normal random varaibles, and they are independent of each other and all the covariates, and i.i.d. across the subscripted dimension(s).

For the MD estimator, we consider both the case where the researcher only observes whether the outside option (i.e., alternative 0) is chosen, and so just minimizes $Q_{N0}(\;)$, and the case where the researcher also observes which alternative is chosen by each decision maker, and so minimizes the sum of $Q_{Nj}(\;)$ for $j = 0; 1; 2$. In all DGPs, each covariate z_{nj} is a continuous uniform random variable over the interval $[\, 9; \; 9]$ and x_{nj} is a binary variable that takes value of 2 or $\; 2$ with equal probability for $j = 1; 2$. The covariates of alternative 0 are $z_{n0} = 0$ and $x_{n0} = 0$. All the observed covariates are independent of each other and are independent, identically distributed across the subscripted dimension(s).

We use a grid search to compute our MD estimator over a parameter space of $[0{:}8; \; 0{:}8]$ with the bin width of $0{:}05$. In the estimation of choice probabilities we apply a truncated normal density for the kernel function $k_h(\,)$ with bandwidth $h_j = sd(z_{nj})N^{(\, 1=22)}$, where $j = 1; 2$. Our are $O(h^s)$ with s J + 1 and $O(Nh^{2(J}$

Online Supplemental Appendix to: Semiparametric Identi…cation and Estimation of Multinomial Discrete Choice Models using Error Symmetry

Arthur Lewbel [y] Jin Yan [z] Yu Zhou [x]

Original February 2019, revised December 2021

S.D Proofs Regarding Estimation

In this section, we provide the proofs of Theorems S.B.1-S.B.3 in Section S.B of the Supplementary Appendix and their related lemmas. Speci…cally, Section S.D.1 provides the proof of Theorem S.B.1 on the population sample objective function; Section S.D.2 collects preliminary lemmas needed for the asympotic properties of the MD estimator de…ned in Section S.B.2; Section S.D.3 provides the proofs of Theorem S.B.2, the consistency of the MD estimator; and Section S.D.4 gives the proofs of Theorem S.B.3, the asymptotic linearity and normality of the estimator and related lemmas. Throughout this appendix, we use the same notations and acronyms de…ned in the main text.

S.D.1 Proof of the Population Objective Function

Proof of Theorem S.B.1 : Part (i) can be shown directly from the quadratic form of the population objective function. We will explicitly prove that Part (ii) holds. To show the existence of a minimizer, recall the population objective function

$$Q_j(\;) \quad 1$$

mentary Appendix Section S.A.1; and the second term equals to zero since $d_j \quad {}^{o}; z_n^{(j)}; X_n^{(j)} \quad = 0$.

Q.E.D.

Proof of Lemma S.D.2: The proofs for three terms are similar. We wil focus on the proof for $\hat{g}_j\left(z_n^{(j)};X_n^{(j)}\right)$. Other terms can be done in a similar fashion. First, by the fact that the outcome come variables By the fact that the outcome variables are binary and function f_j is bounded away from zero, applying the results of Lemma B.1 and Lemma B.2 in Newey (1994) gives the first equality in each equation. Second, the second equality follows from Asssumption 10 using Lemma 8.10 in Newey and McFadden (1994). Q.E.D.

Define $\hat{f}_j^{(t)}\left(z^{(j)};X^{(j)}\right) = \partial^t \hat{f}_j / \partial z_{1;(t)} \cdots \partial z_{t;(t)}$ be the derivative with respect to $z_{(t)}^{(j)}$, where $z_{(t)}^{(j)} = \left(z_{1;(t)}^{(j)}; \cdots; \partial z_{t;(t)}^{(j)}\right)$ be any t-element of $z^{(j)}$. Similarly, we can define $f_j^{(t)}\left(z^{(j)};X^{(j)}\right)$, $\hat{g}_j^{(t)}\left(z^{(j)};X^{(j)}\right)$, $g_j^{(t)}\left(z^{(j)};X^{(j)}\right)$, $\hat{\varphi}_j^{(t)}\left(z^{(j)};X^{(j)}\right)$ and $\varphi_j^{(t)}\left(z^{(j)};X^{(j)}\right)$:

Lemma S.D.3 Under Assumptions E2-E5, for $t = 1;\dots;J$,

$$\sup_{z_n^{(j)};X_n^{(j)} \in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}} \left|\hat{f}_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right) - f_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right)\right| = O_p\left(\sqrt{\frac{\ln N}{N h_N^{J+2t+Jq}}} + h_N^s\right) = o_p\left(N^{-1/4}\right)$$

$$\sup_{z_n^{(j)};X_n^{(j)} \in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}} \left|\hat{g}_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right) - g_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right)\right| = O_p\left(\sqrt{\frac{\ln N}{N h_N^{J+2t+Jq}}} + h_N^s\right) = o_p\left(N^{-1/4}\right)$$

$$\sup_{z_n^{(j)};X_n^{(j)} \in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}} \left|\hat{\varphi}_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right) - \varphi_j^{(t)}\left(z_n^{(j)};X_n^{(j)}\right)\right| = O_p\left(\sqrt{\frac{\ln N}{N h_N^{J+2t+Jq}}} + h_N^s\right) = o_p\left(N^{-1/4}\right)$$

Proof of Lemma S.D.3: The proof follows the same method used in Lemma S.D.2. Q.E.D.

Lemma S.D.4 Under Assumptions E2-E5, for $t = 1; \ldots; J$,

$$\sup_{z_n^{(j)}; X_n^{(j)} \in S^{Tr}_{(z^{(j)}; X^{(j)})}} \left| \nabla \hat{f}_j^{(t)}\left(z_n^{(j)}; X_n^{(j)}\right) - \nabla f_j^{(t)}\left(z_n^{(j)}; X_n^{(j)}\right) \right|$$

$$= O_p\left(\sqrt{\frac{\ln N}{\quad}}\right)$$

Condition (4) following Hong and Tamer (2003). We first introduce an infeasible sample objective function $\tilde{Q}_{Nj}(\theta)$, defined as

$$\tilde{Q}_{Nj}(\theta) = \frac{1}{2N}\sum_{n=1}^{N}\left[\psi_j\left(z_n^{(j)};X_n^{(j)}\right) - d_j\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right]^2 .$$

Following the triangle inequality, we have

$$\left|\hat{Q}_{Nj}(\theta) - Q_j(\theta)\right| \leq \left|\hat{Q}_{Nj}(\theta) - \tilde{Q}_{Nj}(\theta)\right| + \left|\tilde{Q}_{Nj}(\theta) - Q_j(\theta)\right| . \tag{S.D.3}$$

Then, it is sufficient to show that the two terms on the right side of (S.D.3) go to zero uniformly, that is, (i) $\sup_{\theta\in\Theta}\left|\hat{Q}_{Nj}(\theta) - \tilde{Q}_{Nj}(\theta)\right| = o_p(1)$ and (ii) $\sup_{\theta\in\Theta}\left|\tilde{Q}_{Nj}(\theta) - Q_j(\theta)\right| = o_p(1)$.

For Part (i), we observe that

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\hat{Q}_{Nj}(\theta) - \tilde{Q}_{Nj}(\theta)\right| \qquad \text{(S.D.4)}\\
&= \sup_{\theta\in\Theta}\left|\frac{1}{2N}\sum_{n=1}^{N}\psi_j^2\left(z_n^{(j)};X_n^{(j)}\right)\left[\hat{d}_{j,n}^2\left(\theta;z_n^{(j)};X_n^{(j)}\right) - d_j^2\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right]\right|\\
&= \sup_{\theta\in\Theta}\left|\frac{1}{2N}\sum_{n=1}^{N}\psi_j^2\left(z_n^{(j)};X_n^{(j)}\right)\left(\hat{d}_{j,n}\left(\theta;z_n^{(j)};X_n^{(j)}\right) + d_j\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right)\right.\\
&\qquad\left.\left(\hat{d}_{j,n}\left(\theta;z_n^{(j)};X_n^{(j)}\right) - d_j\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right)\right|\\
&\leq C\sup_{\theta\in\Theta}\sup_{\left(z_n^{(j)};X_n^{(j)}\right)\in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}}\left|\hat{d}_{j,n}\left(\theta;z_n^{(j)};X_n^{(j)}\right) - d_j\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right| = o_p(1) .
\end{aligned}$$

The first equality in (S.D.4) follows from definition and direct calculation. The second equality holds by factorization. The next inequality is satisfied by the fact that functions ψ_j and d_j are boundedQEXS. The last equality follows the fact that

$$\sup_{\theta\in\Theta}\sup_{z\equiv\left(z_n^{(j)};X_n^{(j)}\right)\in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}}\left|\hat{d}_{j,n}\left(\theta;z_n^{(j)};X_n^{(j)}\right) - d_j\left(\theta;z_n^{(j)};X_n^{(j)}\right)\right|$$

is bounded by the product of a constant and the derivative functions shown by Lemma S.D.3.

Part (ii) holds by showing pointwise convergence and stochastic equicontinuity. By the Law of Large Numbers (LLN), we can directly obtain the pointwise convergence of $Q_{Nj}(\theta)$ to $Q_j(\theta)$. Next we can conclude the uniformity by showing stochastic equicontinuity, that is,

$$\sup_{\theta^{(1)};\theta^{(2)}\in\Theta;\|\theta^{(1)}-\theta^{(2)}\|\le\delta}\left|Q_{Nj}\left(\theta^{(1)}\right)-Q_{Nj}\left(\theta^{(2)}\right)\right| = o_p(1):$$

Following Andrews (1994), the stochastic equicontinuity can be shown by verifying that $Q_{Nj}(\theta)$ is the type II class of function, satisfying the Lipschitz condition $\left|Q_{Nj}\left(\theta^{(1)}\right)-Q_{Nj}\left(\theta^{(2)}\right)\right| \le C\|\theta^{(1)}-\theta^{(2)}\|$. We verify that this holds from the continuity of the quadratic form of the objective function and the continuity of the kernel derivative functions with bounded second derivatives. Q.E.D.

S.D.4 Asymptotic Linearity and Normality of the MD Estimator

In this section, we first show the lemmas that contribute to the proof of Theorem S.B.3

Lemma S.D.5 Under Assumptions I,TR and E, $H_{Nj}(\tilde{\theta}) \to_p H_j$, where

$$H_j = E\left[\psi_j^2\left(z_n^{(j)};X_n^{(j)}\right)\nabla d_j\left(\theta^o;z_n^{(j)};X_n^{(j)}\right)\left[\nabla d_j\left(\theta^o;z_n^{(j)};X_n^{(j)}\right)\right]^0\right]$$

Proof of Lemma S.D.5: To show the desired result, we first show that the following results hold:

(i) $H_{Nj}(\tilde{\theta}) = H_{Nj;1}(\tilde{\theta}) + H_{Nj;2}(\tilde{\theta})$; where

$$H_{Nj;1}(\tilde{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\psi_j^2\left(z_n^{(j)};X_n^{(j)}\right)\hat{d}_{j;n}\left(\tilde{\theta};z_n^{(j)};X_n^{(j)}\right)\left[\nabla'\hat{d}_{j;n}\left(\tilde{\theta};z_n^{(j)};X_n^{(j)}\right)\right];$$

$$H_{Nj;2}(\tilde{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\psi_j^2\left(z_n^{(j)};X_n^{(j)}\right)\nabla\hat{d}_{j;n}\left(\tilde{\theta};z_n^{(j)};X_n^{(j)}\right)\left[\nabla\hat{d}_{j;n}\left(\tilde{\theta};z_n^{(j)};X_n^{(j)}\right)\right]^0;$$

(ii) $H_{Nj;1}(\tilde{\theta}) = o_p(1)$; and (iii) $H_{Nj;2}(\tilde{\theta}) \to_p H_j$.

The decomposition in Part (i) follows from direct calculation. For Part (ii), observe that

$$H_{Nj;1}\ \tilde{} = \left[H_{Nj;1}\ \tilde{} \quad H_{Nj;1}\left({}^{o}\right) \right] + H_{Nj;1}\left({}^{o}\right) = o_p(1)$$

Given that $\tilde{}$ lies between ^{o} and $\tilde{}$, we get that $\tilde{}$ is uniformly consistent, and by applying the Delta method for the continuity of the choice probability, we obtain that $H_{Nj;1}\ \tilde{} \quad H_{Nj;1}\left({}^{o}\right) = o_p(1)$. Next, $H_{NjNj;1}($

Next we calculate

$$q_{Nj;2}\,(\ {}^{o}) \qquad \text{(S.D.7)}$$

$$= \frac{1}{N}\sum_{n=1}^{N} {}_{j}^{2}\; z_n^{(j)};X_n^{(j)} \Big['\hat{}^{(J)}_{j;o;\ n}\; z_n^{(j)};X_n^{(j)} \quad '^{(J)}_{j;o}\; z_n^{(j)};X_n^{(j)}$$

$$+\hat{}^{(J)}_{j;cs;\ n}\; z_n^{(j)};X_n^{(j)};\ {}^{o} \quad '^{(J)}_{j;cs}\; z_n^{(j)};X_n^{(j)};\ {}^{o}\Big]$$

$$\Big[\ r\ \hat{d}_{j;\ n}\ {}^{o}; z_n^{(j)};X_n^{(j)} \quad r\ d_j\ {}^{o}; z_n^{(j)};X_n^{(j)}\Big]$$

$$= A_1 + A_2 + A_3 + A_4$$

where

$$A_1 = \frac{1}{N}\sum_{n=1}^{N}\Big[{}_{j}\; z_n^{(j)};X_n^{(j)} \quad '\hat{}^{(J)}_{j;o;\ n}\; z_n^{(j)};X_n^{(j)} \quad '^{(J)}_{j;o}\; z_n^{(j)};X_n^{(j)}\Big]$$

$$\sum_{j=1}^{J} {}_{j}\; z_n^{(j)};X_n^{(j)} \quad r\ '\hat{}^{(J)}_{j;o;\ n;(j\ )}\; z_n^{(j)};X_n^{(j)} \quad r\ '^{(J)}_{j;o;(j\ )}\; z_n^{(j)};X_n^{(j)} \qquad x_{n;j}^{(j)}$$

$$A \quad \frac{1}{N}\sum_{n=1}^{N}\Big[{}_{j}\; z_n^{(j)};X_n^{(j)} \quad '\hat{}^{(J)}_{j;o;\ n}\; z_n^{(j)};X_n^{(j)} \quad '^{(J)}_{j;o}\; z_n^{(j)};X_n^{(j)}\Big]$$

$$\sum_{j=1}^{J} {}_{j}\; z_n^{(j)};X_n^{(j)} \quad r$$

X.)75 Td (()Tj /TT21 Td (z)(j)Tj /TT3-249 T3()Tj /TT2 7.97 Tf (j)Tj /TT3 7.97 Tf 7.177^

j

where j represents the choice of j product and (j) represents the derivatives with respect to j index. For A_1, we have

$$\frac{1}{N}\sum_{n=1}^{N}\left[\left|\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right)-\varphi^{(J)}_{j;o}\left(z_n^{(j)};X_n^{(j)}\right)\right|\right]^2 \leq \left(\sup_{\left(z_n^{(j)};X_n^{(j)}\right)\in S^{Tr}_{\left(z^{(j)};X^{(j)}\right)}}\left|\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right)-\varphi^{(J)}_{j;o}\left(z_n^{(j)};X_n^{(j)}\right)\right|\right)^2 = o_p\left(N^{1/2}\right)$$

and $z_n^{(j)}$ [illegible]

where $_N(!_m;!_n) = {}_{N;o}(!_m;!_n) \quad {}_{N;cs}(!_m;!_n)$ with

$$_{N;o}(!_m;!_n) = {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \quad y_{mj} \quad ' {}_{j;o} \; z_n^{(j)}; X_n^{(j)}$$

$$K_{h_z}^{(J)} \; z_m^{(j)} \quad z_n^{(j)} \; K_{h_X} \; X_m^{(j)} \quad X_n^{(j)} \; f_j^{\,1} \; z_n^{(j)}; X_n^{(j)} \;\; ;$$

$$_{N;cs}(!_m;!_n) = {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \quad y_{mj} \quad ' {}_{j;cs} \; z_n^{(j)}; X_n^{(j)}; \;^o$$

$$K_{h_z}^{(J)} \; z_m^{(j)} \quad z_n^{(j)} \quad 2X_n^{(j)} \;^o \quad K_{h_X} \; X_m^{(j)} \quad X_n^{(j)} \; f_j^{\,1} \quad z_n^{(j)} \quad 2X_n^{(j)} \;^o; X_n^{(j)} \;\; :$$

where $K_{h_z}^{(J)} \; z_m^{(j)} = \prod_{l=1}^{J} h_N^{2J} \; k^{(1)} \; h_{z_l}^{1} \; z_{ml}^{(j)}$ where $k^{(1)}$ is the …rst derivative of kernel function.

Proof of Lemma S.D.7: We …rst observe that

$$q_{Nj;1}(\;^o) = \frac{1}{N}\sum_{n=1}^{N} {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \; \hat{d}_{j;\,n} \;^o; z_n^{(j)}; X_n^{(j)} \qquad \text{(S.D.9)}$$

$$= \frac{1}{N}\sum_{n=1}^{N} {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \; \Big[\hat{'}^{(J)}_{j;o;\,n} \; z_n^{(j)}; X_n^{(j)} \quad \hat{'}^{(J)}_{j;cs;\,n} \; z_n^{(j)}; X_n^{(j)}; \;^o \Big] \;\; :$$

$$= \frac{1}{N}\sum_{n=1}^{N} {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \; \Big\{\Big[\hat{'}^{(J)}_{j;o;\,n} \; z_n^{(j)}; X_n^{(j)} \quad '^{(J)}_{j;o} \; z_n^{(j)}; X_n^{(j)} \Big]$$

$$\Big[\hat{'}^{(J)}_{j;cs;\,n} \; z_n^{(j)}; X_n^{(j)}; \;^o \quad '^{(J)}_{j;cs} \; z_n^{(j)}; X_n^{(j)}; \;^o \Big]\Big\}$$

$$= \frac{1}{N}\sum_{n=1}^{N} {}_j \; z_n^{(j)}; X_n^{(j)}; \;^o \; \Big\{\Big[\hat{'}^{(J)}_{j;o;\,n} \; z_n^{(j)}; X_n^{(j)} \quad E \; \hat{'}^{(J)}_{j;o;\,n} \; z_n^{(j)}; X_n^{(j)} \quad z_n^{(j)}; X_n^{(j)} \Big]$$

$$+ \; E \Big[\hat{'}^{(J)}_{j;o;\,n} \; z_n^{(j)}; X_n^{(j)} \quad z_n^{(j)}; X_n^{(j)} \Big] \quad '^{(J)}_{j;o} \; z_n^{(j)}; X_n^{(j)} \Big]$$

$$\Big[\hat{'}^{(J)}_{j;cs;\,n} \; z_n^{(j)}; X_n^{(j)}; \;^o \quad E \; \hat{'}^{(J)}_{j;cs;\,n} \; z_n^{(j)}; X_n^{(j)}; \;^o \quad z_n^{(j)}; X_n^{(j)} \Big]$$

$$+ E \Big[\hat{'}^{(J)}_{j;cs;\,n} \; z_n^{(j)}; X_n^{(j)}; \;^o \; X$$

The second, third and fourth equalities follows from adding and substracting terms. The last equality holds by the fact that

$$\sup_{z_n^{(j)};X_n^{(j)} \in S^{Tr}\left(z^{(j)};X^{(j)}\right)} E\left[\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right) \left(z_n^{(j)};X_n^{(j)}\right)\right] - \varphi^{(J)}_{j;o}\left(z_n^{(j)};X_n^{(j)}\right) = O(h^s) \quad \text{h489 3.91}$$

$$\sup_{z_n^{(j)};X_n^{(j)} \in S^{Tr}\left(z^{(j)};X^{(j)}\right)} E\left[\hat{\varphi}^{(J)}_{j;cs;n}\left(z_n^{(j)};X_n^{(j)};o\right) \left(z_n^{(j)};X_n^{(j)}\right)\right] - \varphi^{(J)}_{j;cs}\left(z_n^{(j)};X_n^{(j)};o\right) = O(h^s)$$

Next, to derive $\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right) - E\left[\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right)\left(z_n^{(j)};X_n^{(j)}\right)\right]$, we observe that

$$\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)};o\right) - E\left[\hat{\varphi}^{(J)}_{j;o;n}\left(z_n^{(j)};X_n^{(j)}\right) \ldots\right]$$ 1f 10.909 0 Td (E)TTd 749 10840 6.51j -12.109 as

Since each term is of order $O_p\left(N^{-1/4}\right)$, thus $R_{o;1}$ is of order $O_p\left(N^{-1/2}\right)$. Denoting

$$_{N;o;1}(\omega_m;\omega_n) = {}_j\left(z_n^{(j)};X_n^{(j)};{}^o\right)$$

$$f_j^{-1}\left(y_{mj} - \varphi_{j;o}^{(J)}\left(z_n^{(j)};X_n^{(j)}\right)\right) K_{h_z}^{(J)}\left(z_m^{(j)} - z_n^{(j)}\right) K_{h_X}\left(X_m^{(j)} - X_n^{(j)}\right)$$

will give the first term of $_{N;1}(\omega_m;\omega_n)$.

In addition, to derive $\hat{\varphi}_{j;cs;n}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right) - E\left[\hat{\varphi}_{j;cs;n}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right) \mid z_n^{(j)};X_n^{(j)}\right]$, we observe that

$$\hat{\varphi}_{j;cs;n}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right) - E\left[\hat{\varphi}_{j;cs;n}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right) \mid z_n^{(j)};X_n^{(j)}\right]$$

$$= \hat{f}_j^{-1} y_{mj} K_{h_z}^{(J)}\left(z_m^{(j)} - z_n^{(j)} - 2\,{}^oX_n^{(j)}\right) K_{h_X}\left(X_m^{(j)} - X_n^{(j)}\right)$$

$$- E\left[\hat{\varphi}_{j;cs;n}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right) \mid z_n^{(j)};X_n^{(j)}\right]$$

$$= f_j^{-1}\frac{1}{N}\sum_{m=1;m\neq n}^{N}\left(y_{mj} - \varphi_{j;cs}^{(J)}\left(z_n^{(j)};X_n^{(j)};{}^o\right)\right)$$

$$K_{h_z}^{(J)}\left(z_m^{(j)} - z_n^{(j)} - 2\,{}^oX_n^{(j)}\right) K_{h_X}\left(X_m^{(j)} - X_n^{(j)}\right) + R_{cs;1}$$

where the second equality holds by the same argument for $\hat{f}_j^{-1}$, and $R_{cs;1}$ collects the higher order terms from the decompostion of $\hat{f}_j^{-1}$, with the order of $O_p\left(N^{-1/2}\right)$, by the same argument as above. Denoting

$$_{N;cs;1}(\omega_m;\omega_n) = {}_j\left(z_n^{(j)};X_n^{(j)};{}^o\right)$$

$$f_j^{-1}\left(y_{mj} - \varphi_{j;cs}^{(J)}\right) K_{h_z}^{(J)}\left(z_m^{(j)} - z_n^{(j)} - 2\,{}^oX_n^{(j)}\right) K_{h_X}\left(X_m^{(j)} - X_n^{(j)}\right)$$

will give the second term of $_{N;1}(\omega_m;\omega_n)$.

Combining all the terms gives the desired results. Q.E.D.

Lemma S.D.8 Under Assumptions E2-E5,

$$\frac{1}{N(N-1)}\sum_{m=1}^{N}\sum_{n=1;n\neq m}^{N} \;_N(\omega_m;\omega_n) = \frac{1}{N}\sum_{m=1}^{N} t_{mj} + o_p\left(N^{-1/2}\right)$$

and

$$N^{-1/2}\sum_{m=1}^{N} t_{mj} \to_d N\left(0_q;\ \;_j\right);$$

where $t_{mj} = (\;_{mj;o} \quad \;_{mj;cs})\ \partial\ \;_j\ z_m^{(j)}; X_m^{(j)};\ ^{o} = \partial z_1^{(j)} \quad \partial z^{(j)}\partial$ y.909 Tf 1 ... 2(()/TT10 10. Tf 9.061 -1.637 Td (j)Tj /TT. 34T

$$
\begin{aligned}
& E\left[\left\| \xi_{N;o}(\omega_m;\omega_n) \right\|^2 \right] \qquad \text{(S.D.10)} \\
&= \int \left[K^{(J)}_{h_z}\left(z_m^{(j)} - z_n^{(j)} \right) K_{h_X}\left(X_m^{(j)} - X_n^{(j)} \right) \right]^2 \\
&\quad \left[\varphi_j\left(z_m^{(j)}; X_m^{(j)} \right) + \varphi_j^2\left(z_n^{(j)}; X_n^{(j)} \right) - 2\varphi_j\left(z_m^{(j)}; X_m^{(j)} \right) \varphi_j\left(z_n^{(j)}; X_n^{(j)} \right) \right] \\
&\quad f_j\left(z_m^{(j)}; X_m^{(j)} \right) f_j\left(z_n^{(j)}; X_n^{(j)} \right) f_j^{-1}\left(z_n^{(j)}; X_n^{(j)} \right) \varepsilon_j^2\left(z_n^{(j)}; X_n^{(j)}; \theta^o \right) dz_m^{(j)} dX_m^{(j)} dz_n^{(j)} dX_n^{(j)} \\
&= \int \left[K^{(J)}_{h_z}\left(u_z^{(j)} \right) K_{h_X}\left(u_X^{(j)} \right) \right]^2 \\
&\quad \left[\varphi_j\left(z_m^{(j)}; X_m^{(j)} \right) + \varphi_j^2\left(z_m^{(j)} - u_z^{(j)} h_z; X_m^{(j)} - u_X^{(j)} h_X \right) \right. \\
&\quad \left. - 2\varphi_j\left(z_m^{(j)}; X_m^{(j)} \right) \varphi_j\left(z_m^{(j)} - u_z^{(j)} h_z; X_m^{(j)} - u_X^{(j)} h_X \right) \right] \\
&\quad f_j\left(z_m^{(j)}; X_m^{(j)} \right) \varepsilon_j^2\left(z_m^{(j)} - u_z^{(j)} h_z; X_m^{(j)} - u_X^{(j)} h_X; \theta^o \right) dz_m^{(j)} dX_m^{(j)} du_z^{(j)} du_X^{(j)} \\
&= O\left(h_N^{2J - J - Jq} \right) = O\left(N \left(N h_N^{2J + J + rfJq} \right)^{-1} \right) = o(N);
\end{aligned}
$$

where the first equality in (S.D.10) follows from definitions; the second equality holds using a change of variables; and the third equality is satisfied by Assumptions E3 and E4. The desired result then follows from Assumption E4. Similarly, we can show $E\left[\left\| \xi_{N;cs}(\omega_m;\omega_n) \right\|^2 \right] = o(N)$.

Next we show that the second term in $\hat{\Delta}_n$ contributes to the asymptotic linearity and normality, while the first and third terms are asymptotically negligible. In sum, we show that (i) $E[\xi_N(\omega_m;\omega_n)] = E[r_{N1}(\omega_m)] = E[r_{N2}(\omega_n)] = o\left(N^{-1/2}\right)$, (ii) $\frac{1}{N}\sum_{n=1}^{N}\left(r_{N2}(\omega_n) - E[r_{N2}(\omega_n)]\right) = $ N78(=)]TJ 5.584 $o_p\left(N^{-1/2}\right)$; and (iii) $\frac{1}{N}\sum_{m=1}^{N}\left(r_{N1}(\omega_m) - E[r_{N1}(\omega_m)]\right) = N^{-1}\sum_{m=1}^{N} t_{mj}$; where $N^{-1/2}\sum_{=1}^{N}$ 1.636 Td [(]))-2

$$
\begin{aligned}
E[\psi_{N;o}(\omega_m;\omega_n)] &= E[E[\psi_{N;o}(\omega_m;\omega_n)|\omega_n]] \\
&= E\Big[E\Big[\delta_j\big(z_n^{(j)};X_n^{(j)};\theta_o\big)\big(y_{mj}-\varphi_{j;o}\big(z_n^{(j)};X_n^{(j)}\big)\big) \\
&\quad K_{h_z}^{(J)}\big(z_m^{(j)}-z_n^{(j)}\big)K_{h_X}\big(X_m^{(j)}-X_n^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big)\Big|\omega_n\Big]\Big] \\
&= E\Big[\int\delta_j\big(z_n^{(j)};X_n^{(j)};\theta_o\big)\Big(\varphi_{j;o}\big(z_n^{(j)}+u_z^{(j)}h_z;X_n^{(j)}+\Lambda u_X^{(j)}h_X\big)-\varphi_{j;o}\big(z_n^{(j)};X_n^{(j)}\big)\Big) \\
&\quad K_{h_z}^{(J)}\big(u_z^{(j)}\big)K_{h_X}\big(u_X^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big) \\
&\quad f_j^{-1}\big(z_n^{(j)}+u_z^{(j)}h_z;X_n^{(j)}+\Lambda u_X^{(j)}h_X\big)\,du_z^{(j)}du_X^{(j)}\Big|\omega_n\Big] \\
&= E\left[\int\delta_j\big(z_n^{(j)};X_n^{(j)};\theta_o\big)\frac{\partial\Big(\varphi_{j;o}\big(z_n^{(j)}+u_z^{(j)}h_z;X_n^{(j)}+\Lambda u_X^{(j)}h_X\big)-\varphi_{j;o}\big(z_n^{(j)};X_n^{(j)}\big)\Big)}{\partial u_z^{(j)}}\right. \\
&\quad \left. K_{h_z}\big(u_z^{(j)}\big)K_{h_X}\big(u_X^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big)\Big|\omega_n\right] \\
&= O(h_N^s).
\end{aligned}
$$

In addition, we can show that $E[\psi_{N;cs}(\omega_m;\omega_n)] = E[E[\psi_{N;cs}(\omega_m;\omega_n)|\omega_n]] = O(h_N^s)$. Then it implies that

$$
E[\psi_N(\omega_m;\omega_n)] = E[\psi_{N;o}(\omega_m;\omega_n)] - E[\psi_{N;cs}(\omega_m;\omega_n)] = O(h_N^s).
$$

Second, to show Part (ii) holds, by direct calculation, we have

$$r_{N2;o}(\omega_n) = E[\,{}_{N;o}(\omega_m;\omega_n)\,|\,\omega_n] \tag{S.D.11}$$

$$= E\Big[\,{}_j\big(z_n^{(j)};X_n^{(j)};{}_o\big)\big(y_{mj} - \varphi_{j;o}(z_n^{(j)};X_n^{(j)})\big)$$

$$K_{h_z}^{(J)}\big(z_m^{(j)} - z_n^{(j)}\big)K_{h_X}\big(X_m^{(j)} - X_n^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big)\,\Big|\,\omega_n\Big]$$

$$= E\Big[\,{}_j\big(z_n^{(j)};X_n^{(j)};{}_o\big)\big(\varphi_{j;o}(z_m^{(j)};X_m^{(j)}) - \varphi_{j;o}(z_n^{(j)};X_n^{(j)})\big)$$

$$K_{h_z}^{(J)}\big(z_m^{(j)} - z_n^{(j)}\big)K_{h_X}\big(X_m^{(j)} - X_n^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big)\,\Big|\,\omega_n\Big]$$

$$= \int {}_j\big(z_n^{(j)};X_n^{(j)};{}_o\big)\frac{\partial\big(\varphi_{j;o}(z_n^{(j)} + u_z^{(j)}h_z;X_n^{(j)} + \zeta u_X^{(j)}h_X) - \varphi_{j;o}(z_n^{(j)};X_n^{(j)})\big)}{\partial u_z^{(j)}}$$

$$K_{h_z}\big(u_z^{(j)}\big)K_{h_X}\big(u_X^{(j)}\big)f_j^{-1}\big(z_n^{(j)};X_n^{(j)}\big)$$

$$f_j\big(z_n^{(j)} + u_z^{(j)}h_z;X_n^{(j)} + \zeta u_X^{(j)}h_X\big)\,du_z^{(j)}du_X^{(j)}$$

$$= O(h_N^s) = o\big(N^{-1/2}\big):$$

The last second equality follows from integration by parts and a Taylor expansion. We therefore get that $\frac{1}{N}\sum_{n=1}^{N}(r_{N2}(\omega_n) - E[r_{N2}(\omega_n)]) = o_p\big(N^{-1/2}\big)$.

where

$$r_o(\omega_m) = \int \frac{\partial^J \ell_j\left(z_m^{(j)}; X_m^{(j)}; \theta^o\right) \varphi_j\left(z_m^{(j)}; X_m^{(j)}\right)}{\partial z_1^{(j)} \cdots \partial z_J^{(j)}} K_{h_z}(u)\, du$$

probability to zero. Note that

$$\frac{1}{\sqrt{N}}\sum_{m=1}^{N}\left(r_o(\omega_m) - E[r_o(\omega_m)]\right) \tag{S.D.15}$$

$$= \frac{1}{\sqrt{N}}\sum_{m=1}^{N}\left(\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)\frac{\partial^J\varphi_{j;o}\left(z_m^{(j)};X_m^{(j)}\right)}{\partial z_1^{(j)}\cdots\partial z_J^{(j)}}\right.$$

$$\left. - E\left[\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)\frac{\partial^J\varphi_{j;o}\left(z_m^{(j)};X_m^{(j)}\right)}{\partial z_1^{(j)}\cdots\partial z_J^{(j)}}\right]\right)$$

$$- \frac{1}{\sqrt{N}}\sum_{m=1}^{N} y_{mj}\,\varphi_{j;o}\left(z_m^{(j)};X_m^{(j)}\right)\frac{\partial^J\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)}{\partial z_1^{(j)}\cdots\partial z_J^{(j)}}$$

$$+ E\left[y_{mj}\,\varphi_{j;o}\left(z_m^{(j)};X_m^{(j)}\right)\frac{\partial^J\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)}{\partial z_1^{(j)}\cdots\partial z_J^{(j)}}\right]$$

and

$$\frac{1}{\sqrt{N}}\sum_{m=1}^{N}\left(r_{cs}(\omega_m) - E[r_{cs}(\omega_m)]\right) \tag{S.D.16}$$

$$= \frac{1}{\sqrt{N}}\sum_{m=1}^{N}\left(\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)\frac{\partial^J\varphi_{j;cs}\left(z_m^{(j)};X_m^{(j)};\theta_o\right)}{\partial z_1^{(j)}\cdots\partial z_J^{(j)}}\right.$$

$$- E\left[\psi_j\left(z_m^{(j)};X_m^{(j)};\theta_o\right)\partial^J\varphi_{j;cs;}\right.$$

Then

$$\frac{1}{\sqrt{N}} \sum_{m=1}^{N} (r\,(\omega_m$$

Hong, H. and Tamer E. (2003): "Inference in Censored Models with Endogenous Regressors," Econometrica, 71(3), 905-932.

Powell, J.L., Stock, J., and Stocker, T. (1989): "Semiparametric Estimation of Index Models," Econometrica, 57, 1403-1430.

Newey, W.K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," Econometric Theory, 10(2), 233-253

Newey, W.K. and McFadden, D. (1994): "Large Sample Estimation and Hypothesis Testing," Handbook of Econometrics, Vol. 4, 2111-2245.